



버전 16

선형 모형 적합

“진정 무엇인가를 발견하는 여행은 새로운 풍경을 바라보는
것이 아니라 새로운 눈을 가지는 데 있다.”

Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

16.0

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020–2021. JMP® 16 선행 모형 적합. Cary, NC: SAS Institute Inc.

JMP® 16 선행 모형 적합

Copyright © 2020-2021, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

March 2021

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to <http://support.sas.com/thirdpartylicenses>.

JMP 활용하기

JMP를 처음 사용하든 오랫동안 사용해왔든, JMP와 관련해서 배워야 하는 것은 언제나 있습니다.

JMP.com을 방문하면 다음과 같은 유용한 자료를 볼 수 있습니다.

- JMP 시작 방법에 대한 라이브 및 녹화 웹 캐스트
- 새로운 기능과 고급 기법에 대한 비디오 데모 및 웹 캐스트
- JMP 교육 과정 등록에 대한 상세 정보
- 현지에서 개최되는 세미나 일정
- 다른 사용자들이 JMP를 사용하는 방법을 보여주는 성공 사례
- JMP 사용자 커뮤니티, 추가 기능 및 스크립트 예를 포함한 사용자용 리소스, 포럼, 블로그, 컨퍼런스 정보 등

<https://www.jmp.com/getstarted>

목차

선형 모형 적합

1	JMP 알아보기	15
	설명서 및 추가 리소스	
	JMP 설명서에 사용되는 서식 규칙	17
	JMP 도움말	18
	JMP 설명서 라이브러리	18
	JMP 학습을 위한 추가 리소스	24
	JMP 자습서	24
	샘플 데이터 테이블	24
	통계 및 JSL 용어에 대해 알아보기	25
	JMP 팁 및 힌트 알아보기	25
	JMP 툴팁	25
	JMP 사용자 커뮤니티	25
	무료 온라인 Statistical Thinking 교육 과정	26
	JMP 새로운 사용자를 위한 입문 키트	26
	Statistics Knowledge 포털	26
	JMP 교육 과정	26
	사용자가 작성한 JMP 설명서	26
	JMP 시작하기 창	26
	JMP 기술 지원	27
2	모형 규격	29
	선형 모형 지정	
	모형 적합 플랫폼 개요	31
	모형 적합을 사용한 회귀 분석의 예	32
	모형 적합 플랫폼 시작	34
	모형 적합 시작 창 요소	35
	모형 효과 생성	37
	적합 분석법	43
	모형 규격 옵션	45
	결측값 정보화	46
	타당도 검사	47
	모형 규격 및 모형 적합의 예	47

단순 선형 회귀	48
X 다항식 (k 차)	49
X 및 Z 다항식 (k 차)	50
다중 선형 회귀	50
일원 분산 분석	51
이원 분산 분석	52
교호작용이 있는 이원 분산 분석	53
삼원 완전 요인	55
공분산 분석, 동일한 기울기	57
공분산 분석, 서로 다른 기울기	58
2 요인 내포 임의 효과 모형	60
3 요인 완전 내포 임의 효과 모형	62
단순 분할구 또는 반복 측정 모형	63
2 요인 반응 표면 모형	64
매듭이 지정된 스플라인 효과	66
3 표준 최소 제곱 모형	69
모형의 공통 클래스 분석	
예 : 표준 최소 제곱 사용	72
표준 최소 제곱 분석법 시작	74
모형 적합 시작 창	75
모형 적합 시작 창의 표준 최소 제곱 옵션	76
표준 최소 제곱의 검증	78
결측값	78
최소 제곱 적합 보고서	79
단일 반응과 다중 반응의 비교	79
강조와 관련된 보고서 구조	79
특수 보고서	79
최소 제곱 적합 옵션	83
적합 그룹 옵션	83
반응 옵션	84
회귀 보고서	85
적합 요약	85
분산 분석	86
모수 추정값	87
효과 검정	88
효과 상세 정보	89
적합 결여	104
추정값	105
예측 표현식 표시	107

정렬된 추정값	108
확장 추정값	111
표시자 파라미터화 추정값	113
순차 검정	114
사용자 검정	115
다중 비교	117
기울기 비교	130
결합 요인 검정	130
역추정 예측	131
Cox 혼합물	135
모수 검정력	136
추정값 상관계수	138
명목형 효과에 대한 코딩	139
효과 선별	140
척도화된 추정값 및 연속형 항 코딩	141
그림 옵션	142
정규 확률도 보고서	146
Bayes 그림 보고서	147
파레토도 보고서	149
요인 프로파일링	150
프로파일러	151
교호작용 그림	151
등고선 프로파일러	153
혼합물 프로파일러	154
입방체도	155
Y의 Box-Cox 변환	156
표면 프로파일러	158
행 진단	159
레버리지 그림	161
Press	164
열 저장	165
예측 계산식	167
효과 요약 보고서	168
혼합 및 임의 효과 모형 보고서 및 옵션	172
혼합 모형 및 임의 효과 모형	172
REML(제한 최대 가능도) 방법	176
EMS(전통적) 모형 적합 보고서	181
모형 항 사이에 선형 종속성이 있는 모형	182
특이성 상세 정보	183
모수 추정값 보고서	184

효과 검정 보고서	185
예	185
표준 최소 제곱 분석법에 대한 통계 상세 정보	187
강조 규칙	187
사용자 검정 예에 대한 상세 정보	188
추정값 상관계수	189
레버리지 그림 상세 정보	189
Kackar-Harville 수정	192
검정력 분석	193
4 표준 최소 제곱 예	203
모형의 공통 클래스 분석	
일원 분산 분석 예	205
동일한 기울기의 공분산 분석 예	207
서로 다른 기울기의 공분산 분석 예	209
반응 표면 모형 예	212
전체 반응 표면 모형 적합	212
모형 축소	213
반응 표면 보고서 검토	214
예측 프로파일러를 사용하여 임계점 찾기	216
등고선 프로파일러를 사용하여 표면 보기	217
분할구 설계 예	219
임의 효과 모수 추정 예	223
매듭이 지정된 스플라인 효과 예	225
활성 요인에 대한 Bayes 그림 예	227
5 단계별 회귀 모형	229
변수 선택을 사용하여 모형 찾기	
단계별 회귀 개요	231
예 : 단계별 회귀 사용	231
단계별 보고서	233
단계별 플랫폼 옵션	233
단계별 회귀 제어 패널	234
현재 추정값 보고서	240
단계 기록 보고서	241
교차, 교호작용 또는 다항식 항이 있는 모형	242
결합 규칙의 예	242
범주형 및 순서형 효과가 있는 모형	244
계층 항 생성	244
명목형 항이 있는 모형의 예	245

계층 항에 대한 제한 규칙의 예	249
이항 및 순서형 로지스틱 단계별 회귀 수행	251
예 : 로지스틱 단계별 회귀 사용	252
모든 가능 모형 옵션	253
예 : 모든 가능 모형 옵션 사용	254
모형 평균화 옵션	255
예 : 모형 평균화 옵션 사용	256
단계별 회귀의 검증 옵션	257
단계별 회귀에서 두 개 또는 세 개의 값이 있는 검증 데이터 집합	257
단계별 회귀의 K 폴드 교차 검증	260
6 일반화 회귀 모형	263
변수 선택 기법을 사용하여 모형 생성	
일반화 회귀 분석법 개요	265
일반화 회귀의 예	266
일반화 회귀 분석법 시작	269
분포	270
일반화 회귀 보고서 창	277
일반화 회귀 보고서 옵션	278
모형 시작 제어판	278
반응 분포	279
추정 방법 옵션	279
고급 제어	283
검증 방법 옵션	286
조기 중지	287
시작	287
모형 적합 보고서	287
회귀 그림	288
모형 요약	288
추정 상세 정보	291
해 경로	291
중심화 및 척도화된 예측 변수에 대한 모수 추정값	294
원래 예측 변수에 대한 모수 추정값	295
활성 모수 추정값	296
효과 검정	297
모형 적합 옵션	297
일반화 회귀 분석법에 대한 통계 상세 정보	305
추정 방법에 대한 통계 상세 정보	306
고급 제어에 대한 통계 상세 정보	308
분포에 대한 통계 상세 정보	309

7	일반화 회귀 예	315
	정규화 기법을 사용하여 모형 생성	
	Poisson 일반화 회귀 예	317
	이항 일반화 회귀 예	319
	영과잉 Poisson 회귀 예	321
	가로형 데이터에 대한 일반화 회귀 예	323
8	혼합 모형	327
	평균 및 공분산 공동 모델링	
	혼합 모형 분석법 개요	330
	예 : 혼합 모형 사용	330
	혼합 모형 분석법 시작	335
	모형 적합 시작 창	335
	데이터 형식	340
	혼합 적합 보고서	341
	적합 통계량	343
	임의 효과 공분산 모수 추정값	345
	고정 효과 모수 추정값	345
	반복 효과 공분산 모수 추정값	346
	랜덤 계수	347
	임의 효과 예측	347
	고정 효과 검정	347
	순차 검정	348
	다중 비교	348
	기울기 비교	349
	주변 모형 추론	349
	실제값 대 예측값 그림	349
	잔차 그림	349
	Profilers	350
	조건부 모형 추론	351
	실제값 대 조건부 예측값 그림	351
	조건부 잔차 그림	351
	조건부 프로파일러	352
	변동도	352
	공분산 및 상관 행렬	353
	열 저장	354
	혼합 모형 분석법의 추가 예	354
	반복 측정 예	355
	분할구 예	370
	공간 예 : 균일도 시험	376

상관관계가 있는 반응의 예	384
혼합 모형 분석법에 대한 통계 상세 정보	390
수렴 스코어 검정	391
랜덤 계수 모형	392
반복 측정	393
반복 공분산 구조	393
공간 및 시간 변동	398
Kackar-Harville 수정	401
9 다변량 반응 모형	403
MANOVA 를 사용하여 관계 적합	
다중 반응 모형의 예	405
MANOVA 보고서	407
MANOVA 적합 옵션	407
반응 지정	408
반응 선택 옵션	408
사용자 검정 옵션	409
모형 효과 옵션	409
다변량 검정	413
확장된 다변량 보고서	414
다변량 검정 비교	415
단변량 검정 및 구형성 검정	415
반복 측정을 사용하는 다변량 모형	416
반복 측정 예	417
복합 다변량 모형의 예	418
판별 분석	421
판별 저장 옵션의 예	421
MANOVA 분석법에 대한 통계 상세 정보	422
다변량 검정	422
근사 F- 검정	423
정준 상세 정보	424
10 로그 선형 분산 모형	427
반응의 분산 및 평균 모델링	
로그 선형 분산 모형 개요	429
산포 효과	429
모형 규격	429
참고	430
로그 선형 분산 분석법 시작	430
예: 로그 선형 분산 사용	431

로그 선형 보고서	433
로그 선형 플랫폼 옵션	434
열 저장	435
행 진단	435
잔차 검토	435
적합 모형 프로파일링	436
적합 모형 프로파일링의 예	436
11 로지스틱 회귀 모형	439
명목형 또는 순서형 반응에 대한 회귀 모형 적합	
명목형 및 순서형 로지스틱 분석법 개요	441
명목형 로지스틱 회귀	441
순서형 로지스틱 회귀	441
로지스틱 회귀 모형을 적합시키는 기타 JMP 플랫폼	442
로지스틱 회귀의 예	442
명목형 로지스틱 회귀의 예	442
순서형 로지스틱 회귀의 예	444
명목형 및 순서형 로지스틱 분석법 시작	446
로지스틱 회귀 모형의 검증	447
로지스틱 적합 보고서	448
전체 모형 검증	449
적합 상세 정보	450
적합 결여 검증	450
로지스틱 적합 플랫폼 옵션	451
명목형 및 순서형 적합 옵션	451
명목형 적합 옵션	453
순서형 적합 옵션	454
로지스틱 회귀의 추가 예	455
역추정 예측의 예	455
명목형 로지스틱 모형에 대한 효과 요약의 사용 예	457
2 차 순서형 로지스틱 모형의 예	459
예 : 여러 열의 개수 쌓기	462
명목형 및 순서형 로지스틱 분석법에 대한 통계 상세 정보	463
로지스틱 회귀 모형	463
승산비	464
통계적 검정의 관계	465
12 일반화 선형 모형	467
비정규 반응 분포에 대한 모형 적합	
일반화 선형 모형 분석법 개요	469

일반화 선형 모형의 예	469
일반화 선형 모형 분석법 시작	472
일반화 선형 모형 적합 보고서	473
전체 모형 검정	474
일반화 선형 모형 적합 보고서 옵션	475
일반화 선형 모형 분석법의 추가 예	477
대비를 사용하여 변수 수준의 차이 비교	477
오프셋을 사용하는 Poisson 회귀	478
로그 연결을 사용하는 정규 회귀	480
일반화 선형 모형 분석법에 대한 통계 상세 정보	482
모형 선택 및 이탈도	484

A 통계 상세 정보

선형 모형 적합

반응 모형	489
연속형 반응	489
명목형 반응	489
순서형 반응	491
요인 모형	492
연속형 요인	492
명목형 요인	493
순서형 요인	503
빈도	508
일반적인 가정	509
가정된 모형	509
상대적 유의성	509
다중 추론	509
타당도 평가	509
대체 방법	510
핵심 통계적 개념	510
불확도, 통합 개념	511
두 개의 기본 적합 기계	511
가능도, AICc 및 BIC	514
검정력 계산	515
LSN 계산	516
LSV 계산	516
검정력 계산	517
조정 검정력 계산	518
신뢰 한계를 사용한 역추정 예측	519

B 참조 자료 521

C 기술 라이선스 표시 527

1 장

JMP 알아보기 설명서 및 추가 리소스

설명서 서식 규칙, 각 JMP 문서에 대한 설명, 도움말 시스템, 기타 지원 제공 위치 등 JMP 설명서에 대해 알아봅니다.

목차

JMP 설명서에 사용되는 서식 규칙	17
JMP 도움말	18
JMP 설명서 라이브러리	18
JMP 학습을 위한 추가 리소스	24
JMP 자습서	24
샘플 데이터 테이블	24
통계 및 JSL 용어에 대해 알아보기	25
JMP 팁 및 힌트 알아보기	25
JMP 툴팁	25
JMP 사용자 커뮤니티	25
무료 온라인 Statistical Thinking 교육 과정	26
JMP 새로운 사용자를 위한 입문 키트	26
Statistics Knowledge 포털	26
JMP 교육 과정	26
사용자가 작성한 JMP 설명서	26
JMP 시작하기 창	26
JMP 기술 지원	27

JMP 설명서에 사용되는 서식 규칙

설명서 자료가 가리키는 화면 정보를 쉽게 알아볼 수 있도록 다음과 같은 서식 규칙이 사용됩니다.

- 샘플 데이터 테이블 이름, 열 이름, 경로 이름, 파일 이름, 파일 확장자 및 폴더는 **Helvetica**(또는 sans-serif online) 글꼴로 표시됩니다.
- 코드는 **Lucida Sans Typewriter**(또는 monospace online) 글꼴로 표시됩니다.
- 코드 출력은 **Lucida Sans Typewriter** 기울임꼴 (또는 monospace italic online) 글꼴로 표시되고 앞의 코드보다 더 많은 들여쓰기가 적용됩니다.
- **Helvetica bold**(또는 bold sans-serif online) 서식은 작업을 수행하기 위해 사용자가 선택하는 다음과 같은 항목을 나타냅니다.
 - 버튼
 - 체크박스
 - 명령
 - 선택 가능한 목록 이름
 - 메뉴
 - 옵션
 - 탭 이름
 - 텍스트 상자
- 다음 항목은 기울임꼴로 표시됩니다.
 - 중요하거나 JMP 와 관련된 정의가 있는 단어 또는 구
 - 설명서 제목
 - 변수
- JMP Pro 에만 해당되는 기능에는 JMP Pro 아이콘  이 표시됩니다. JMP Pro 기능의 개요는 <https://www.jmp.com/software/pro> 에서 확인할 수 있습니다.

참고 : 참고 섹션에는 특수한 정보와 제한 사항이 표시됩니다.

팁 : 팁 섹션에는 유용한 정보가 표시됩니다.

JMP 도움말

"도움말" 메뉴의 "JMP 도움말"에서는 JMP 기능, 통계적 방법 및 JSL(JMP Scripting Language)에 대한 정보를 검색할 수 있습니다. 다음과 같은 몇 가지 방법으로 JMP 도움말을 열 수 있습니다.

- Windows 에서 **도움말 > JMP 도움말**을 선택하여 JMP 도움말을 검색하고 봅니다.
- Windows 에서 F1 키를 눌러 기본 브라우저로 도움말 시스템을 엽니다.
- 데이터 테이블 또는 보고서 창의 특정 부분에 대한 도움말을 확인합니다. **도구** 메뉴에서 도움말 도구  를 선택한 후 데이터 테이블 또는 보고서 창의 아무 곳이나 클릭하면 해당 영역에 대한 도움말이 표시됩니다.
- JMP 창 내에서 **도움말** 버튼을 클릭합니다.

참고: JMP 도움말은 인터넷이 연결되어 있어야 사용할 수 있습니다. 인터넷이 연결되어 있지 않으면 **도움말 > JMP 설명서 라이브러리**를 선택하여 단일 PDF 파일에서 모든 설명서를 검색할 수 있습니다. 자세한 내용은 "[JMP 설명서 라이브러리](#)"(18 페이지)에서 확인하십시오.

JMP 설명서 라이브러리

도움말 시스템 콘텐츠는 JMP 설명서 라이브러리라는 단일 PDF 파일로도 제공됩니다. 이 파일을 열려면 **도움말 > JMP 설명서 라이브러리**를 선택합니다. JMP 라이브러리에 포함된 각 문서의 개별 PDF 파일을 검색하려면 <https://www.jmp.com/documentation>에서 파일을 다운로드하십시오.

다음 표에서는 JMP 라이브러리에 포함된 각 문서의 용도와 내용을 설명합니다.

문서 제목	문서 용도	문서 내용
JMP 살펴보기	JMP에 익숙하지 않다면 이 설명서부터 시작하십시오.	JMP에 대해 소개하고 데이터 생성 및 분석을 시작하기 위한 정보를 제공합니다. 또한 결과를 공유하는 방법도 알아볼 수 있습니다.
JMP 사용	JMP 데이터 테이블과 기본적인 작업을 수행하는 방법에 대해 알아봅니다.	데이터 가져오기, 열 특성 수정, 데이터 정렬, SAS 연결 등을 비롯하여 JMP의 모든 영역에 걸친 일반적인 JMP 개념 및 기능을 다룹니다.

문서 제목	문서 용도	문서 내용
기본 분석	이 문서를 사용하여 기본적인 분석을 수행합니다.	"분석" 메뉴의 다음 플랫폼에 대해 설명합니다. • 분포 • X로 Y 적합 • 테이블 생성 • 텍스트 탐색기 "분석">"X로 Y 적합"을 통해 이변량 분석, 일원 ANOVA 및 분할 분석을 수행하는 방법을 다룹니다. 붓스트랩을 사용하여 표집 분포에 근사한 값을 산출하는 방법과 시뮬레이션 플랫폼을 사용하여 모수적 재표집을 수행하는 방법도 포함되어 있습니다.
Essential Graphing	데이터에 이상적인 그래프를 찾습니다.	"그래프" 메뉴의 다음 플랫폼에 대해 설명합니다. • 그래프 빌더 • 3D 산점도 • 등고선 그림 • 버블 그림 • 평행 그림 • 셀 그림 • 산점도 행렬 • 삼원 그림 • 트리맵 • 차트 • 중첩 그림 이 설명서에서는 배경 맵 및 사용자 맵을 생성하는 방법도 다룹니다.
Profilers	반응 표면의 횡단면을 볼 수 있게 해주는 대화식 프로파일링 도구의 사용 방법을 알아봅니다.	"그래프" 메뉴에 나열된 모든 프로파일러를 다룹니다. 랜덤 입력을 사용한 시뮬레이션 실행과 함께 잡음 요인 분석이 포함됩니다.

문서 제목	문서 용도	문서 내용
실험 설계 가이드	실험 설계 방법을 알아 보고 적절한 표본 크기 를 결정합니다.	"DOE" 메뉴의 모든 항목을 다룹니다.
선형 모형 적합	모형 적합 플랫폼과 이 플랫폼의 다양한 분석 법에 대해 알아봅니다.	"분석" 메뉴의 모형 적합 플랫폼에서 사용 할 수 있는 다음 분석법에 대해 설명합니다. • 표준 최소 제곱 • 단계별 • 일반화 회귀 • 혼합 모형 • MANOVA • 로그 선형 분산 • 명목형 로지스틱 • 순서형 로지스틱 • 일반화 선형 모형

문서 제목	문서 용도	문서 내용
Predictive and Specialized Modeling	추가 모델링 기법에 대해 알아봅니다.	"분석">"예측 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 신경망 • 파티션 • 붓스트랩 포레스트 • 부스티드 트리 • K 최근접 이웃 • Naive Bayes • 서포트 벡터 머신 • 모형 비교 • 모형 선별 • 검증 열 생성 • 계산식 저장소 "분석">"전문 모델링" 메뉴의 다음 플랫폼에 대해 설명합니다. • 곡선 적합 • 비선형 • 함수 데이터 탐색기 • 가우스 과정 • 시계열 • 매칭 쌍 "분석">"선별" 메뉴의 다음 플랫폼에 대해 설명합니다. • 모델링 유틸리티 • 반응 변수 선별 • 공정 변수 선별 • 예측 변수 선별 • 연관성 분석 • 공정 기록 탐색기

문서 제목	문서 용도	문서 내용
Multivariate Methods	몇 개의 변수를 동시에 분석하기 위한 기법을 알아봅니다.	"분석">"다변량 방법" 메뉴의 다음 플랫폼에 대해 설명합니다. • 다변량 • 주성분 • 판별 • 부분 최소 제곱 • 다중 대응 분석 • 구조 방정식 모형 • 요인 분석 • 다차원 척도법 • 항목 분석 "분석">"군집화" 메뉴의 다음 플랫폼에 대해 설명합니다. • 계층적 군집화 • K 평균 군집화 • 정규 혼합 • 잠재 계층 분석 • 변수 군집화
Quality and Process Methods	공정 평가 및 개선을 위한 도구에 대해 알아봅니다.	"분석">"품질 및 공정" 메뉴의 다음 플랫폼에 대해 설명합니다. • 관리도 빌더 및 개별 관리도 • 측정 시스템 분석 • 계량형 / 계수형 게이지 차트 • 공정 능력 • 모형 기반 다변량 관리도 • 레거시 관리도 • 파레토도 • 다이어그램 • 규격 한계 관리 • OC 곡선

문서 제목	문서 용도	문서 내용
Reliability and Survival Methods	제품 또는 시스템의 신뢰도 평가 및 향상 방법과 사람 및 제품의 생존 데이터 분석 방법을 알아봅니다.	"분석">"신뢰성 및 생존" 메뉴의 다음 플랫폼에 대해 설명합니다. • 수명 분포 • 수명 분포 적합 • 누적 손상 • 재발 분석 • 열화 • 파괴 열화 • 신뢰도 예측 • 신뢰도 성장 • 신뢰도 블록 다이어그램 • 수리 가능 시스템 시뮬레이션 • 생존 • 모수 생존 모형 적합 • 비례 위험 모형 적합
Consumer Research	소비자 선호도를 연구하고 해당 정보를 사용하여 보다 나은 제품 및 서비스를 개발하기 위한 방법을 알아봅니다.	"분석">"소비자 조사" 메뉴의 다음 플랫폼에 대해 설명합니다. • 범주형 • 선택 • 최대차이 • Uplift • 다중 요인 분석
Scripting Guide	강력한 JSL(JMP 스크립트 언어)의 활용 방법을 알아봅니다.	스크립트 작성/디버깅, 데이터 테이블 조작, 표시 상자 생성 및 JMP 응용 프로그램 생성 등의 다양한 주제를 다룹니다.
JSL Syntax Reference	다양한 JSL 함수 및 인수, 그리고 개체 및 표시 상자로 보내는 메시지에 대해 알아봅니다.	JSL 명령의 구문, 예제 및 참고 사항이 포함되어 있습니다.

JMP 학습을 위한 추가 리소스

JMP 도움말 외에도 다음 리소스를 사용하여 JMP에 대해 배울 수 있습니다.

- "JMP 자습서 "
- " 샘플 데이터 테이블 "
- " 통계 및 JSL 용어에 대해 알아보기 "
- "JMP 팁 및 힌트 알아보기 "
- "JMP 툴팁 "
- "JMP 사용자 커뮤니티 "
- " 무료 온라인 Statistical Thinking 교육 과정 "
- "JMP 새로운 사용자를 위한 입문 키트 "
- "Statistics Knowledge 포털 "
- "JMP 교육 과정 "
- " 사용자가 작성한 JMP 설명서 "
- "JMP 시작하기 창 "

JMP 자습서

도움말 > 자습서를 선택하여 JMP 자습서에 액세스할 수 있습니다. **자습서** 메뉴의 첫 번째 항목은 **자습서 디렉터리**입니다. 이 항목을 클릭하면 모든 자습서가 범주별로 그룹화되어 있는 새 창이 열립니다.

JMP에 익숙하지 않다면 **초보자 자습서**부터 시작하십시오. 이 자습서에서는 JMP 인터페이스를 단계별로 안내하며 JMP를 사용하는 데 필요한 기본 사항을 설명합니다.

나머지 자습서는 실험 설계, 표본 평균과 상수의 비교 같은 JMP의 특정 측면을 이해하는 데 유용합니다.

샘플 데이터 테이블

JMP 설명서 모음에 포함된 모든 예에서는 샘플 데이터를 사용합니다. 샘플 데이터 디렉터리를 열려면 **도움말 > 샘플 데이터 라이브러리**를 선택하십시오.

샘플 데이터 테이블의 사전순 목록을 보거나 범주별로 샘플 데이터를 보려면 **도움말 > 샘플 데이터**를 선택하십시오.

샘플 데이터 테이블은 다음 디렉터리에 설치되어 있습니다.

Windows: C:\Program Files\SAS\JMP\16\Samples\Data

macOS: \Library\Application Support\JMP\16\Samples\Data

JMP Pro의 경우에는 JMP 디렉터리가 아니라 JMPPRO 디렉터리에 샘플 데이터가 설치되어 있습니다.

샘플 데이터를 사용한 예를 보려면 **도움말 > 샘플 데이터**를 선택하고 교육 자료 섹션으로 이동하십시오. 교육 자료에 대한 자세한 내용은 <https://jmp.com/tools>에서 확인하십시오.

통계 및 JSL 용어에 대해 알아보기

통계 용어에 대한 도움말을 보려면 "도움말 > 통계 분석 인덱스"를 선택합니다. JSL 스크립트 및 예제에 대한 도움말을 보려면 **도움말 > 스크립트 인덱스**를 선택합니다.

통계 분석 인덱스 통계 용어에 대한 정의를 제공합니다.

스크립트 인덱스 JSL 함수, 개체 및 표시 상자에 대한 정보를 검색할 수 있습니다. "스크립트 인덱스"에서는 샘플 스크립트를 편집 및 실행하고 명령에 대한 도움말을 볼 수도 있습니다.

JMP 팁 및 힌트 알아보기

JMP를 처음 시작할 때는 "오늘의 유익한 정보" 창이 표시됩니다. 이 창에서는 JMP를 사용하기 위한 팁을 제공합니다.

"오늘의 유익한 정보" 기능을 해제하려면 **시작할 때 정보 표시** 체크박스를 선택 해제하십시오. 이 창을 다시 보려면 **도움말 > 오늘의 유익한 정보**를 선택하십시오. 또는 "환경 설정" 창을 사용하여 이 기능을 해제할 수도 있습니다.

JMP 툴팁

JMP에서 다음과 같은 항목을 커서로 가리키면 설명 툴팁 (또는 가리키기 라벨)이 제공됩니다.

- 메뉴 또는 도구 모음 옵션
- 그래프의 라벨
- 보고서 창의 텍스트 결과 (커서를 원 모양으로 움직이면 표시됨)
- 홈 창의 파일 또는 창
- 스크립트 편집기의 코드

팁: Windows의 경우 JMP 환경 설정에서 툴팁을 숨길 수 있습니다. **파일 > 환경 설정 > 일반**을 선택한 후 **메뉴 팁 표시**를 선택 취소합니다. 이 옵션은 macOS에서는 사용할 수 없습니다.

JMP 사용자 커뮤니티

JMP 사용자 커뮤니티에서는 JMP에 대해 알아보고 다른 JMP 사용자와 교류하는 데 도움이 되는 다양한 옵션을 제공합니다. 한 페이지 분량의 가이드, 자습서 및 데모로 구성된 학습 라이브러리부터 시작하는 것이 좋습니다. 다양한 JMP 교육 과정에 등록하여 학습을 계속할 수도 있습니다.

그 밖에도 토론 포럼, 샘플 데이터 및 스크립트 파일 교환, 웹 캐스트 및 소셜 네트워킹 그룹을 비롯한 리소스가 있습니다.

웹 사이트의 JMP 리소스에 액세스하려면 **도움말 > JMP 사용자 커뮤니티**를 선택하거나 <https://community.jmp.com>을 방문하십시오.

무료 온라인 Statistical Thinking 교육 과정

이 무료 온라인 교육 과정에서는 탐색적 데이터 분석, 품질 관리 방법, 상관 및 회귀 등의 항목에 대한 실용적인 통계적 기술을 배울 수 있습니다. 이 교육 과정은 짧은 비디오와 데모, 연습 등으로 구성되어 있습니다. 자세한 내용은 <https://www.jmp.com/statisticalthinking>에서 확인하십시오.

JMP 새로운 사용자를 위한 입문 키트

JMP 새로운 사용자를 위한 입문 키트는 JMP의 기본 사항을 빨리 익힐 수 있도록 돕기 위한 것입니다. 30개의 짧은 데모 비디오 및 작업을 마치면 좀더 편하게 소프트웨어 사용 방법을 익히고 세계 최대 규모의 JMP 사용자 온라인 커뮤니티와 연결할 수 있습니다. 자세한 내용은 <https://www.jmp.com/welcome>에서 확인하십시오.

Statistics Knowledge 포털

Statistics Knowledge 포털에서는 방문자가 확실한 기초를 토대로 통계적 기술을 쌓을 수 있도록 간단한 통계 설명과 함께 명확한 예시 및 그래픽을 제공합니다. 자세한 내용은 <https://www.jmp.com/skp>에서 확인하십시오.

JMP 교육 과정

SAS에서는 숙련된 JMP 전문가 팀의 주도로 다양한 주제에 대한 교육 과정을 제공합니다. 공개 교육, 라이브 웹 교육, 현장 교육 등이 제공되며, 온라인 e-learning 구독을 선택하여 편리한 시간에 학습할 수도 있습니다. 자세한 내용은 <https://www.jmp.com/training>에서 확인하십시오.

사용자가 작성한 JMP 설명서

JMP 웹 사이트에서는 JMP 사용자가 작성한 추가 JMP 사용 설명서가 제공됩니다. 자세한 내용은 <https://www.jmp.com/books>에서 확인하십시오.

JMP 시작하기 창

JMP 또는 데이터 분석에 익숙하지 않다면 먼저 "JMP 시작하기" 창을 살펴보십시오. 이 창에는 옵션이 범주별로 설명되어 있으며 버튼을 클릭하여 옵션을 시작할 수 있습니다. "JMP 시작하기

"창에는 "분석", "그래프", "테이블" 및 "파일" 메뉴에 있는 다양한 옵션이 포함됩니다. 또한 이 창에는 JMP Pro의 기능 및 플랫폼도 나열됩니다.

- "JMP 시작하기" 창을 열려면 **보기** (macOS의 경우 **창**) > **JMP 시작하기**를 선택합니다.
- Windows에서 JMP를 열 때 자동으로 "JMP 시작하기"를 표시하려면 **파일 > 환경 설정 > 일반**을 선택한 후 "초기 JMP 창" 목록에서 **JMP 시작하기**를 선택합니다. macOS에서는 **JMP > 환경 설정 > 초기 JMP 시작하기 창**을 선택합니다.

JMP 기술 지원

JMP 기술 지원은 통계학자 또는 SAS 및 JMP의 교육을 받은 엔지니어가 제공하며, 이들 중 상당수는 통계 또는 기타 기술 분야의 석사 학위를 갖고 있습니다.

기술 지원 전화 번호를 포함한 많은 기술 지원 옵션이 <https://www.jmp.com/support>에서 제공됩니다.

2 장

모형 규격 선형 모형 지정

모형 적합 플랫폼을 사용하면 복잡한 모형을 효율적으로 지정할 수 있습니다. 매크로, 속성 및 변환을 통해 작업이 간소화됩니다. 모형 적합은 다양한 모형과 효과 구조를 적합시키는 방법입니다.

여기에는 다음과 같은 내용이 포함됩니다.

- 단순 및 다중 선형 회귀
- 분산 및 공분산 분석
- 임의 효과, 내포 효과, 혼합 효과, 반복 측정 및 분할구 모형
- 명목형 및 순서형 로지스틱 회귀
- MANOVA(다변량 분산 분석)
- 정준 상관 및 판별 분석
- 로그 선형 분산(평균 및 분산 모델링)
- GLM(일반화 선형 모형)
- 모수 생존 및 비례 위험
- 반응 변수 선별(많은 수의 반응 연구)

JMP PRO JMP Pro에서는 다음 항목도 적합시킬 수 있습니다.

- 다양한 공분산 구조를 가진 혼합 모형
- Elastic Net, Lasso 및 능형 회귀를 포함한 일반화 회귀 모형
- 부분 최소 제곱

모형 적합 플랫폼을 사용하면 원하는 분석법을 선택하여 다양한 유형의 모형을 적합시킬 수 있습니다. 이 장에서는 대부분의 분석법에 공통된 "모형 규격" 창의 요소에 대해 중점적으로 다룹니다.

목차

모형 적합 플랫폼 개요.....	31
모형 적합을 사용한 회귀 분석의 예.....	32
모형 적합 플랫폼 시작.....	34
모형 적합 시작 창 요소.....	35
모형 효과 생성.....	37
적합 분석법.....	43
모형 규격 옵션.....	45
결측값 정보화.....	46
타당도 검사.....	47
모형 규격 및 모형 적합의 예.....	47
단순 선형 회귀.....	48
X 다항식(k 차).....	49
X 및 Z 다항식(k 차).....	50
다중 선형 회귀.....	50
일원 분산 분석.....	51
이원 분산 분석.....	52
교호작용이 있는 이원 분산 분석.....	53
삼원 완전 요인.....	55
공분산 분석, 동일한 기울기.....	57
공분산 분석, 서로 다른 기울기.....	58
2요인 내포 임의 효과 모형.....	60
3요인 완전 내포 임의 효과 모형.....	62
단순 분할구 또는 반복 측정 모형.....	63
2요인 반응 표면 모형.....	64
매듭이 지정된 스플라인 효과.....	66

모형 적합 플랫폼 개요

모형 적합 플랫폼을 사용하면 효과 구조가 복잡한 모형을 효율적으로 지정할 수 있습니다. 이러한 효과 구조는 모형 모수에 대해 선형입니다. 모형을 지정한 후에는 여러 적합 분석법 중에서 적절한 적합 기법을 선택할 수 있습니다. 분석법을 선택하면 해당 분석법과 관련된 선택 항목이 "모형 적합" 창에 제공됩니다. 이 장에서는 대부분의 분석법에 공통된 "모형 규격" 창의 요소에 대해 중점적으로 다룹니다. 모든 분석법에 대한 설명은 "[모형 적합 시작 창 요소](#)"(35 페이지)에서 확인하십시오.

모형 적합은 여러 방법을 사용하여 적합시킬 수 있는 다양한 모형을 지정하는 데 사용할 수 있습니다. 표 2.1에는 모형 적합을 사용하여 정의할 수 있는 몇 가지 일반적인 모형이 나열되어 있습니다. 표에서 X 및 Z 효과는 연속형 모델링 유형의 열을 나타내고 A, B, C 효과는 명목형 또는 순서형 모델링 유형의 열을 나타냅니다.

이러한 모형 효과를 생성하는 클릭 시퀀스, 모형 적합 그림 및 몇 가지 예는 "[모형 규격 및 모형 적합의 예](#)"(47 페이지)에서 확인하십시오.

표 2.1 표준 모형 유형

모형 유형	모형 효과
단순 선형 회귀	X
X 다항식(k 차)	X, X*X,..., X ^k
X 및 Z 다항식(k 차)	X, X*X,..., X ^k , Z, Z*Z,..., Z ^k
다중 선형 회귀	X, Z 및 기타 연속형 열
일원 분산 분석	A
이원 분산 분석	A, B
교호작용이 있는 이원 분산 분석	A, B, A*B
삼원 완전 요인	A, B, C, A*B, A*C, B*C, A*B*C
공분산 분석, 동일한 기울기	A, X
공분산 분석, 서로 다른 기울기	A, X, A*X
2요인 내포 임의 효과 모형	A, B[A]& 랜덤
3요인 완전 내포 임의 효과 모형	A, B[A]& 랜덤, C[A,B]& 랜덤
단순 분할구 또는 반복 측정 모형	A, B[A]& 랜덤, C, C*A
2요인 반응 표면 모형	X&RS, Z&RS, X*X, X*Z, Z*Z

모형 적합을 사용한 회귀 분석의 예

유산소 피트니스 연구의 결과 데이터가 있으며 몇 가지 연속형 변수에서 산소 섭취량을 예측하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Fitness.jmp** 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다. "분석법" 상자가 비어 있습니다.
3. **Oxy** 를 선택하고 **Y** 를 클릭합니다.

연속형 반응을 지정하면 "분석법" 이 기본적으로 "표준 최소 제곱" 으로 설정되지만 다른 분석법을 선택할 수 있습니다. "강조" 는 기본적으로 "효과 레버리지" 로 설정됩니다.

4. **Ctrl** 키를 누르고 **Sex, Age, Weight, Runtime, RunPulse, RstPulse** 및 **MaxPulse** 를 선택합니다. **추가**를 클릭하여 "모형 효과 생성" 목록에 추가합니다. 이 창을 나중에 사용할 수 있게 하려면 **대화상자 열린 채 유지**를 선택할 수 있습니다. "모형 규격" 창이 그림 2.1 에 표시된 것처럼 보입니다.
5. **실행**을 클릭합니다. 그림 2.2 에 보고서의 일부분이 나와 있습니다.

그림 2.1 Fitness 회귀 모형에 대한 모형 규격 창

모형 규격

열 선택: 9개 열
 Name
 Sex
 Age
 Weight
 Oxy
 Runtime
 RunPulse
 RstPulse
 MaxPulse

역할 변수 선택

Y: Oxy (선택적)

가중치: 선택적 숫자
 빈도: 선택적 숫자
 검증: 선택적 숫자
 기준: 선택적

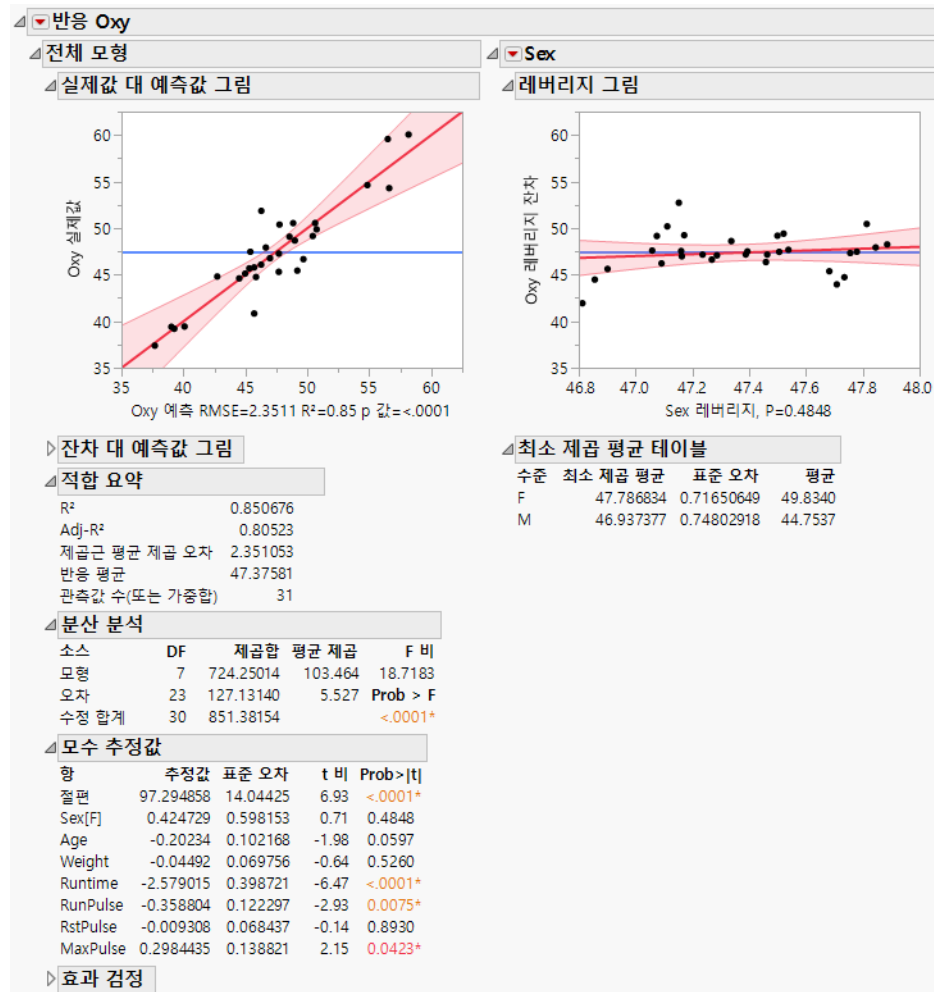
분석법: 표준 최소 제곱
 강조: 효과 레버리지

도움말 실행
 재호출 ☒ 대화상자 열린 채 유지
 제거

모형 효과 생성

추가: Sex, Age, Weight, Runtime, RunPulse, RstPulse, MaxPulse
 교차:
 내포:
 매크로:
 차수: 2
 속성:
 변환:
☐ 절편 없음

그림 2.2 Fitness 데이터에 대한 표준 최소 제곱 보고서 (일부분)



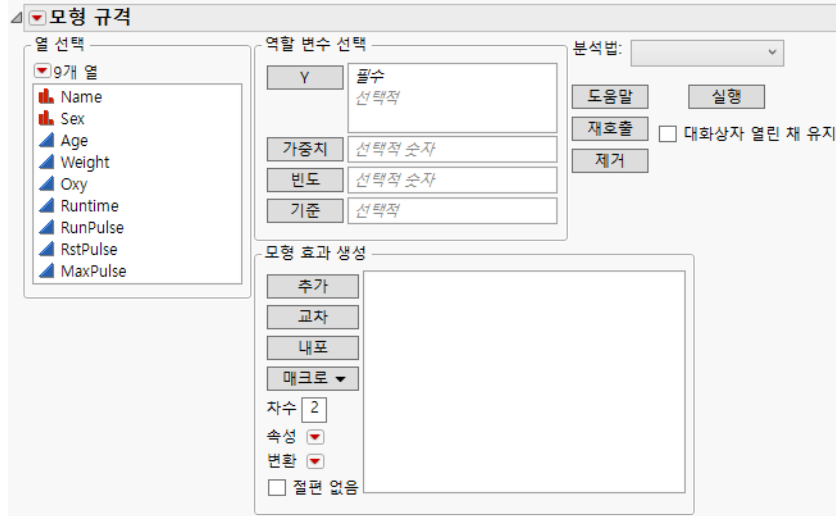
전체 모형에 대한 그림과 보고서는 맨 왼쪽 보고서 열에 나타납니다. 오른쪽 열에는 모형에 지정된 각 효과의 레버리지 그림이 표시됩니다. 공간이 제한되어 그림 2.2에는 Sex 열만 표시되지만 다른 6가지 효과에 대한 열도 보고서에 표시됩니다. 빨간색 삼각형 메뉴에는 보고서 창에 보고서와 그림을 추가하는 기타 옵션이 포함되어 있습니다. 표준 최소 제곱 보고서 창에 대한 자세한 내용은 "표준 최소 제곱 모형" 장의 "최소 제곱 적합 보고서"(79 페이지)에서 확인하십시오.

"모수 추정값" 보고서의 p 값을 살펴보면 Runtime, RunPulse 및 MaxPulse가 산소 섭취량의 유의한 예측 변수임을 알 수 있습니다. 다음 단계에서는 유의하지 않은 예측 변수를 제거하여 모형을 축소할 수 있습니다. 자세한 내용은 "단계별 회귀 모형" 장(229 페이지)에서 확인하십시오.

모형 적합 플랫폼 시작

분석 > 모형 적합을 선택하여 모형 적합 플랫폼을 시작합니다. 그림 2.3에서는 **Fitness.jmp** 샘플 데이터 테이블에 대한 시작 창의 예를 보여 줍니다. 모형 적합 시작 창은 분석법이라고 하는 많은 플랫폼의 공통 시작 창입니다. 자세한 내용은 "**적합 분석법**"(43 페이지)에서 확인하십시오.

그림 2.3 모형 적합 시작 창



참고 : Model 또는 model 스크립트가 있는 데이터 테이블에서 **분석 > 모형 적합**을 선택하면 해당 스크립트를 기반으로 시작 창이 자동으로 채워집니다.

모형 적합 시작 창에는 다음과 같은 주 영역이 있습니다.

- 열 선택 - 현재 데이터 테이블의 열 목록입니다. 현재 데이터 테이블에 제외된 열이 포함되어 있으면 해당 열은 목록에 나타나지 않습니다. "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 **JMP 사용의**에서 확인하십시오.
- 역할 변수 선택 - 표준 JMP 시작 창 버튼이 포함되어 있습니다. 이 버튼에 대한 설명은 "**모형 적합 시작 창 요소**"(35 페이지)에서 확인하십시오.
- 모형 효과 생성 - 모형에 효과를 입력하는 데 사용하는 옵션이 포함되어 있습니다. 자세한 내용은 "**모형 효과 생성**"(37 페이지)에서 확인하십시오.
- 분석법 - 모형 유형을 선택할 수 있는 목록입니다. 분석법을 선택하면 선택한 분석법에 따라 서로 다른 옵션이 나타납니다. 자세한 내용은 "**적합 분석법**"(43 페이지)에서 확인하십시오.

모형 적합 시작 창 요소

다음 모형 적합 시작 창 요소는 대부분의 분석법에 공통적입니다.

모형 규격 "모형 규격" 메뉴에는 다음 옵션이 포함되어 있습니다.

중심화 다항식 모형에 다항식이 포함된 경우 효과를 중심화합니다.

결측값 정보화 효과에 결측값을 수용하는 코딩 시스템을 생성합니다.

유의 수준 설정 모형 보고서에 신뢰 구간의 유의 수준을 설정합니다.

데이터 테이블에 저장 모형 규격을 현재 데이터 테이블의 스크립트에 저장합니다.

스크립트 창에 저장 모형 규격을 스크립트 창에 저장합니다.

SAS 작업 생성 현재 모형 규격에 대한 SAS 코드를 SAS 프로그램 창에 저장합니다.

SAS 로 전송 현재 모형 규격에 대한 SAS 코드를 SAS 세션으로 전송합니다.

수렴 설정 일부 분석법에서 모형 수렴에 사용되는 최대 반복 수 및 수렴 한계를 설정하는 옵션이 포함되어 있습니다.

이러한 옵션에 대한 자세한 내용은 "**모형 규격 옵션**"(45 페이지)에서 확인하십시오.

열 선택 현재 데이터 테이블에서 제외되지 않은 열을 나열합니다.

Y 모형에 대한 하나 이상의 반응 변수 (종속 변수) 를 식별합니다.

참고: 벡터 값이 있는 반응 열은 모형 적합에서 지원되지 않습니다.

가중치 분석을 위해 각 행에 가중치를 할당하는 값이 들어 있는 열입니다. 자세한 내용은 "**가중치**"(36 페이지)에서 확인하십시오.

빈도 분석을 위해 각 행에 빈도를 할당하는 값이 들어 있는 열입니다. 일반적으로 빈도 열의 효과는 데이터 테이블을 확장하여 정수 빈도가 k 인 행을 k 개의 행으로 확장하는 것입니다. 빈도 비율로 지정할 수 있습니다. 자세한 내용은 "**빈도**"(36 페이지)에서 확인하십시오.

JMP PRO 검증 JMP Pro에서는 일부 분석법에 대해 검증 열을 입력할 수 있습니다. 각 분석법에서 검증 열을 처리하는 방법에 대한 자세한 내용은 해당 분석법 관련 장에서 확인하십시오. "열 선택" 목록에서 아무 열도 선택하지 않은 상태로 "검증" 버튼을 클릭하면 데이터 테이블에 검증 열을 추가할 수 있습니다. "검증 열 생성" 유틸리티에 대한 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오. JMP 모델링 플랫폼에서 검증 열이 사용되는 방법에 대한 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오.

기준 변수의 각 수준에 대한 개별 분석으로 구성된 보고서를 생성하는 열을 식별합니다. 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 분석이 생성됩니다.

추가 모형에 효과를 추가합니다. 자세한 내용은 "**추가**"(37 페이지)에서 확인하십시오.

교차 둘 이상의 변수를 교차시켜 교호작용 및 다항식 효과를 생성합니다. 자세한 내용은 "**교차**"(37 페이지)에서 확인하십시오.

내포 내포 효과를 생성합니다. 자세한 내용은 "**내포**"(38 페이지)에서 확인하십시오.

매크로 일반적으로 사용되는 모형에 대한 효과를 생성합니다. 자세한 내용은 "**매크로**"(39 페이지)에서 확인하십시오.

차수 매크로를 사용하여 생성된 요인 또는 다항식 효과가 있는 모형에 지정된 차수를 적용합니다. 자세한 내용은 "**매크로**"(39 페이지)의 **특정 차수까지의 요인** 및 **특정 차수까지의 다항식**에서 확인하십시오.

속성 모형 효과에 속성을 적용합니다. 이러한 속성은 효과를 처리하는 방법을 결정합니다. 자세한 내용은 "**속성**"(40 페이지)에서 확인하십시오.

변환 선택한 연속형 효과 또는 Y 열을 변환합니다. 자세한 내용은 "**변환**"(41 페이지)에서 확인하십시오.

절편 없음 모형에서 절편 항을 제외합니다.

분석법 적합 방법론을 지정합니다. 자세한 내용은 "**모형 적합 시작 창 요소**"(35 페이지)에서 확인하십시오. 선택한 분석법에 따라 다른 옵션이 나타납니다.

목표 수준 (특정 분석법에서 Y가 이항 변수이고 명목형 모델링 유형일 때만 사용 가능) 확률을 모델링할 수준을 지정합니다. 기본값은 수준 순서에 따라 두 수준 중 더 높은 값입니다.

도움말 모형 적합 시작 창의 도움말 항목으로 이동합니다.

재호출 최근에 실행한 모형 규격으로 시작 창을 채웁니다.

제거 선택한 변수를 할당된 역할에서 제거합니다. 또는 효과를 두 번 클릭하거나 효과를 선택하고 Backspace 키를 누릅니다.

실행 지정된 모형 및 분석법에 대한 보고서 창을 생성합니다.

대화상자 열린 채 유지 분석을 실행한 후 언제든지 분석을 변경하고 다시 실행할 수 있도록 시작 창을 열린 채로 유지합니다.

빈도

빈도 텍스트 상자에 입력된 빈도 변수는 대부분의 모형 적합 분석법에서 지원됩니다. 일반적으로 빈도는 다음과 같은 방식으로 해석됩니다. 빈도가 f 인 행이 있다고 가정해 보겠습니다. 그러면 계산된 결과는 각각 빈도가 1인 해당 행의 f 개 복사본을 포함하는 데이터 테이블의 결과와 동일합니다.

빈도 값이 0이거나 결측인 행은 분석에서 제외됩니다. 빈도 값이 음수인 행은 중도절단된 관측값에 대해서만 허용되며 그렇지 않으면 분석에서 제외됩니다. 음수 빈도 값을 중도절단된 관측값과 함께 사용하면 절단된 분포를 적합시킬 수 있습니다.

빈도 값은 정수일 필요가 없습니다. 값이 정수가 아닌 열을 포함하여 빈도 열을 처리하는 방법을 설명하는 기술 상세 정보는 "**빈도**"(508 페이지)에서 확인하십시오.

가중치

가중치 변수는 관측값의 분산이 서로 다른 경우에 유용할 수 있습니다. 예를 들어 각 행이 사전 요약된 평균으로 구성된 데이터에 대해 회귀 모델링을 수행할 때 이러한 현상이 발생할 수 있습니다. 여기서 관측값을 더 많이 포함하는 행(분산이 더 작음)이 관측값을 더 적게 포함하는 행(분산

이 더 큼)보다 손실 함수에 더 크게 기여해야 합니다. 적절하게 정의된 가중치를 사용하여 이를 확인할 수 있습니다.

가중치 변수는 많은 모형 적합 분석법에서 지원됩니다. 가중치 변수를 지원하는 각 분석법은 다음 방법 중 하나를 사용합니다.

- 분산 척도
- 빈도 대칭

분산 척도

최소 제곱 또는 최대 가능도 정규 이론을 사용하여 추정을 수행할 때 주어진 행의 가중치 w 는 손실 함수에 대한 해당 행의 기여도를 $w^{-1/2}$ 로 조정합니다.

가중치 변수는 추정값과 표준 오차에 영향을 줍니다. 그러나 빈도 변수와 달리 가설 검정에 사용되는 자유도에는 영향을 주지 않습니다.

가중치 값이 음수이거나 0인 행은 분석에서 제외됩니다.

빈도 대칭

명목형 로지스틱 및 순서형 로지스틱 분석법에서 가중치 변수는 빈도 변수인 것처럼 처리됩니다. 이러한 분석법에서 가중치 변수와 빈도 변수가 모두 지정되면 각 관측값의 빈도 값이 가중치 값과 빈도 값의 곱과 동일한 것으로 처리합니다.

모형 효과 생성

이 섹션에서는 모형에 효과를 쉽게 입력하기 위해 사용할 수 있는 옵션에 대해 설명합니다. 이러한 옵션을 사용하여 특정 유형의 모형을 생성하는 방법에 대한 예는 "[모형 규격 및 모형 적합의 예](#)"(47 페이지)에서 확인하십시오.

추가

모형에 효과를 추가합니다. 이러한 효과는 "열 선택" 목록에서 직접 추가하거나 해당 목록에서 선택한 후 "매크로" 또는 "속성"을 사용하여 수정할 수 있습니다. "교차" 또는 "내포"를 사용하여 효과를 생성하고 추가하거나 수정할 수도 있습니다. 효과에 포함된 변수의 모델링 유형과 효과에 할당된 속성에 따라 해당 효과가 모형에서 처리되는 방식이 결정됩니다.

참고 : "모형 효과 생성" 목록에서 효과를 제거하려면 효과를 두 번 클릭하거나 선택한 후 **제거**를 클릭하거나 Backspace 키 또는 Delete 키를 누르십시오.

교차

교호작용 또는 다항식 효과를 생성합니다. "열 선택" 목록에서 두 개 이상의 변수를 선택하고 **교차**를 클릭합니다. 아니면 "열 선택" 목록에서 하나 이상의 변수를 선택하고 "모형 효과 생성" 목록에서 하나 이상의 효과를 선택한 후 **교차**를 클릭합니다.

교차효과를 파라미터화하고 코드화하는 방법에 대한 자세한 내용은 "[통계 상세 정보](#)"(487 페이지)에서 확인하십시오.

참고 : 최대 10 개의 열을 결합하는 효과를 교차 및 내포로 생성할 수 있습니다.

교차효과의 예

제품 코팅에 염료를 도포해야 한다고 가정해 보겠습니다. 염료 pH(Dye pH)와 염료 농도(Dye Concentration) 둘 다 코팅 색상에 영향을 주는 것 같습니다. 이러한 요인 효과를 알아보기 위해 염료 pH 및 염료 농도가 각각 높은 수준과 낮은 수준으로 설정된 실험을 설계합니다. 염료 pH가 색상에 미치는 영향은 염료 농도의 수준이 낮을 때보다 높을 때 더 뚜렷하게 나타날 수 있습니다. 이를 **교호작용**이라고 합니다. 가능한 이 교호작용을 모델링하려면 "모형 효과 생성" 목록에서 교차 항(Dye pH * Dye Concentration)을 포함합니다. 이렇게 하면 JMP에서 교호작용을 검정할 수 있습니다.

내포

내포 효과를 생성합니다. 한 효과(B)의 수준이 다른 효과(A)의 단일 수준 내에서만 발생하는 경우 B가 A 안에 내포된다고 합니다. 일반적으로 B[A] 표기가 사용되며 이는 "B가 A 안에 내포되어 있다"는 의미입니다. 내포는 계층 관계를 정의합니다. A를 **외부 효과**라고 하고 B를 **내부 효과**라고 합니다. 내포된 항은 범주형이어야 합니다.

참고 : 내포 항은 외부에서 내부 순으로 지정해야 합니다. 예를 들어 B가 A 안에 내포되고 C가 B 안에 내포된 경우 모형은 A, B[A], C[B,A]((A, B[A], C[A,B]와 같음)로 지정됩니다. 최대 10 개의 열을 결합하는 효과를 교차 및 내포로 생성할 수 있습니다.

내포 효과의 예

두 학교의 각 수학 교사를 내포의 예로 들어보겠습니다. 한 학교에는 세 명의 수학 교사가 있고 다른 학교에는 두 명의 수학 교사가 있습니다. 각 학교의 각 교사는 중복되지 않는 학생 그룹으로 구성된 2~3 개의 학급을 가르칩니다. 이 예에서 학급(C)은 교사(B)에 내포되고 교사(B)는 학교(A)에 내포됩니다. "모형 적합" 시작 창에서 다음과 같이 효과를 입력합니다.

1. A와 B를 모두 "모형 효과 생성" 패널에 추가합니다.
2. "모형 효과 생성" 패널에서 B를 선택합니다.
3. "열 선택" 목록에서 A를 선택합니다.
4. **내포**를 클릭합니다. 그러면 B가 B[A] 효과로 변환됩니다.
5. C를 "모형 효과 생성" 패널에 추가합니다.
6. "모형 효과 생성" 패널에서 C를 선택합니다.
7. "열 선택" 목록에서 A와 B를 선택합니다.
8. **내포**를 클릭합니다. 그러면 C가 C[A, B] 효과로 변환됩니다.

매크로

"매크로" 목록에서 일반적으로 사용되는 모형에 대한 효과를 자동으로 생성하는 옵션을 선택하고 "모형 효과 생성" 목록에 입력합니다.

완전 요인 "열 선택" 목록에서 선택한 열에 대한 모든 주효과 및 교호작용을 생성합니다. 이러한 효과는 "열 선택" 목록에서 주효과가 나열된 순서에 따라 입력됩니다. 대체 순서에 대한 내용은 이 테이블의 **정렬된 요인 배치**를 참조하십시오.

특정 차수까지의 요인 모든 주효과 및 지정된 차수까지의 교호작용을 생성합니다. **매크로** 버튼 아래의 **차수** 상자에 차수를 지정합니다.

정렬된 요인 배치 **완전 요인** 옵션과 동일한 효과 집합을 생성하지만 차수 순으로 나열됩니다. 모든 주효과가 먼저 나열되고 그 뒤에 모든 이원 교호작용, 모든 삼원 교호작용 등이 나열됩니다.

반응 표면 주효과, 이원 교호작용 및 2 차 항을 생성합니다. 선택한 주효과에는 **RS** 로 표시된 반응 표면 속성이 제공됩니다. **RS** 속성이 주효과에 적용되고 "표준 최소 제곱" 분석법을 선택하면 반응 표면 보고서가 제공됩니다. 이 보고서에는 반응 표면의 임계값 및 형태에 대한 정보가 포함됩니다.

"속성"(40 페이지)의 "반응 표면 효과" 및 **실험 설계 가이드**의 도 확인하십시오.

혼합물 반응 표면 주효과 및 이원 교호작용을 생성합니다. 주효과에는 반응 표면 (**RS**) 및 혼합물 (**혼합물**) 속성이 표시됩니다. "표준 최소 제곱" 분석법을 사용할 경우 **혼합물** 속성은 혼합물 모형을 적합시킵니다. **RS** 속성은 혼합물 모형과 관련된 반응 표면 보고서를 생성합니다.

"속성"(40 페이지)의 "혼합물 효과" 및 **실험 설계 가이드**의 도 확인하십시오.

특정 차수까지의 다항식 주효과 및 지정된 차수까지의 다항식 항을 생성합니다. **매크로** 버튼 아래의 **차수** 상자에 차수를 지정합니다.

Scheffe 입방 혼합물 실험에 대한 반응 표면을 지정하는 데 유용한 주효과, 교호작용 및 Scheffe 입방 항을 생성합니다. 이 매크로는 완전 입방 모형을 생성합니다.

혼합물에 대해 3 차 다항식 모형을 적합시키는 경우 $X1 \times X1 \times X2$ 와 같은 짝수 거듭제곱 항은 추정할 수 없으므로 포함하지 않아야 합니다. 그러나 표면의 완전 다항식 규격에는 **Scheffe 입방** 항이라고 하는 $X1 \times X2 \times (X1 - X2)$ 형식의 항이 포함될 수 있습니다.

차수 상자에 3 을 입력한 후 **혼합물 반응 표면** 매크로 명령을 선택하면 Scheffe 입방 항도 포함됩니다.

JMP PRO 그룹화된 회귀변수 보고서의 여러 부분에서 하나의 효과로 처리되는 연속형 요인 집합에 대한 단일 효과를 생성합니다. "효과 요약"에서 그룹화된 효과를 추가하거나 제거할 수 있습니다. 그룹의 각 요인에 대한 개별 그림과 검정이 아니라 각 그룹 효과에 대해 하나의 레버리지 그림과 하나의 효과 검정을 나타냅니다. 이 옵션은 데이터 테이블에 범주형 효과 수준을 나타내는 표시자 열이 포함되어 있을 때 유용할 수 있습니다.

참고 : 회귀변수 그룹에 포함된 비연속형 요인은 분석에 포함되지 않습니다.

속성

"속성" 목록에서는 "모형 효과 생성" 목록에서 선택한 효과에 할당할 수 있는 속성을 선택합니다.

임의 효과 효과에 "랜덤" 속성을 할당합니다. 임의 효과에 대한 자세한 내용은 "표준 최소 제곱 모형" 장의 **"임의 효과 및 적합 방법 지정"**(175 페이지)에서 확인하십시오.

반응 표면 효과 효과에 **RS** 속성을 할당합니다. 관련된 모형 항이 "모형 효과 생성" 목록에 포함되어 있어야 합니다. "매크로" 목록의 "반응 표면" 옵션은 이러한 항을 자동으로 생성하고 주효과에 RS 속성을 할당합니다. 반응 표면 보고서를 생성하려는 경우 교호작용 및 다항식 항에는 반응 표면 속성을 할당할 필요가 없습니다. 주효과에만 이 속성을 할당하면 됩니다.

로그 분산 효과 효과에 "로그 분산" 속성을 할당합니다. 이 속성은 효과가 반응의 분산 모형에 포함된다는 것을 나타냅니다.

모형의 효과를 반응의 평균과 분산 모두에 포함하려면 효과를 두 번 지정해야 합니다. 탭 형식 인터페이스에서 "평균 효과" 탭과 "분산 효과" 탭에 모두 표시되어야 합니다. 아니면 "평균 효과" 탭에서 두 번 입력할 수 있으며, 한 번은 "로그 분산 효과" 속성을 사용하고 또 한 번은 "로그 분산 효과" 속성을 사용하지 않습니다.

혼합물 효과 주효과에 "혼합물" 속성을 할당합니다. 이 속성은 혼합물과 관련된 주효과를 지정하는 데 사용됩니다. "매크로" 목록의 "혼합물 반응 표면" 옵션은 선택한 효과에 혼합물 속성을 자동으로 할당하고 가능한 경우 반응 표면 보고서를 제공합니다.

제외된 효과 효과에 "제외됨" 속성을 할당합니다. 이 속성은 모형 적합에서 해당 효과를 제외합니다. 그러나 이 효과는 적합 결여 검정을 위해 관측값을 그룹화하는 데 사용됩니다. "표준 최소 제곱" 분석법을 사용할 경우 이 효과에 대한 최소 제곱 평균 테이블이 제공됩니다.

매듭이 지정된 스플라인 효과 연속형 주효과에 "매듭 지정" 속성을 할당합니다. 이 속성은 효과에 대한 3 차 스플라인을 모형 규격에 암시적으로 추가합니다. 자세한 내용은 **"매듭이 지정된 스플라인 효과"**(40 페이지)에서 확인하십시오.

매듭이 지정된 스플라인 효과

매듭이 지정된 스플라인은 예측 변수의 유연한 함수를 사용하여 반응 Y 를 적합시키는 데 사용됩니다. 단일 예측 변수 X 를 예로 들겠습니다. "매듭이 지정된 스플라인 효과"가 X 에 할당되고 k 개의 매듭이 지정되면 $k - 2$ 개의 추가 효과가 예측 변수 집합에 암시적으로 추가됩니다. 이러한 각 효과는 매듭에 의해 세그먼트가 정의되는 조각별 3 차 다항식 스플라인입니다. 자세한 내용은 Stone and Koo(1985)를 참조하십시오.

스플라인 수는 사용자가 직접 지정한 매듭 수에 의해 결정됩니다. 스플라인과 관련된 계수는 분석법에 사용되는 방법을 기반으로 추정됩니다.

매듭 배치는 문헌의 지침을 따릅니다. 특히 100 개 이하의 점이 있는 경우 첫 번째 매듭과 마지막 매듭은 각각 다섯 번째로 작은 점과 가장 큰 점입니다. 그렇지 않은 경우 첫 번째 매듭과 마지막 매듭은 각각 매듭이 5 개 이하일 때는 0.05 분위수와 0.95 분위수에 배치되고, 매듭이 5 개보다 많을 때는 0.025 분위수와 0.975 분위수에 배치됩니다. 기본 매듭 수는 관측값이 30 개보다 많으면 5 개이고 관측값이 30 개 미만이면 3 개입니다.

참고 : 매듭이 지정된 스플라인은 주효과 연속형 항에 대해서만 구현됩니다.

변환

"변환" 옵션은 선택한 Y 열 또는 "모형 효과 생성" 텍스트 상자에서 선택한 주효과를 변환합니다.

참고 : " 열 선택 " 목록에서 마우스 오른쪽 버튼을 클릭하고 " 변환 " 을 선택하여 열을 변환할 수도 있습니다. 변환된 열에 대한 참조가 " 열 선택 " 목록에 나타납니다. 그런 다음 데이터 테이블 열과 마찬가지로 모형 적합 창의 열을 사용할 수 있습니다. 자세한 내용은 JMP 사용의 에서 확인하십시오.

없음 적용된 변환 옵션을 제거합니다.

로그 선택한 변수에 자연 로그 변환을 적용합니다.

Sqrt 선택한 변수의 값 제곱근을 적용합니다.

제곱 선택한 변수의 값 제곱을 적용합니다.

역수 변수 X 에 $1/X$ 변환을 적용합니다.

지수 선택한 변수에 지수 변환을 적용합니다.

Arrhenius 변수 T(섭씨 온도) 에 아레니우스 변환을 적용합니다.

$$X = \frac{11604.5181215503}{T + 273.15}$$

이 식은 아레니우스 관계식의 구성 요소로, 여기에 활성화 에너지를 곱하게 됩니다.

아레니우스 역 변수 X 에 아레니우스 역 변환을 적용합니다.

$$T = \frac{11604.5181215503}{X} - 273.15$$

Logit 선택한 열에 대한 로지스틱 함수의 역함수를 계산합니다 (여기서 p 는 0 ~ 1 사이).

$$\text{로짓}(p) = \log\left(\frac{p}{1-p}\right)$$

로지스틱 선택한 열에 대한 로지스틱 (Squish 및 Logist 라고도 함) 함수를 계산합니다 (결과 범위 0 ~ 1).

$$\text{로지스틱}(x) = \frac{1}{(1 + e^{-x})}$$

로지스틱 백분율 선택한 열에 대한 로짓을 백분율로 계산합니다 (여기서 pct 는 0 ~ 100 범위의 백분율).

$$\text{LogitPct}(pct) = \log \left(\frac{\left(\frac{pct}{100} \right)}{1 - \left(\frac{pct}{100} \right)} \right)$$

LogisticPct 선택한 열에 대한 로지스틱 (또는 Logist) 을 백분율로 계산합니다 (결과 범위 0 ~ 100).

$$\text{LogisticPct}(x) = \frac{100}{(1 + e^{-x})}$$

절편 없음

절편 항이 없는 모형을 적합시키려면 "절편 없음"을 선택합니다. 특정 모델링 구조에는 절편 모형이 필요하지 않습니다. 이러한 경우 "절편 없음" 상자가 기본으로 선택되어 있습니다.

모형 효과 생성 탭

다음 분석법의 경우 탭 형식 인터페이스를 사용하여 모형 효과를 입력할 수 있습니다.

참고 : 첫 번째 (주) 탭에서 효과에 적용된 속성은 모형에서 효과가 처리되는 방법을 결정합니다. 모형을 실행한 후 보고서의 빨간색 삼각형 메뉴에서 " 모형 대화상자 " 를 요청하면 해당 효과가 적절한 탭에 나타납니다.

표준 최소 제곱 다음과 같이 모형 효과를 입력합니다.

고정 효과 탭 고정 효과로 모델링할 효과를 입력합니다. 고정 효과는 특정 처리 수준이 관심 대상인 효과입니다. 전체 처리 수준에 대해 평균 반응을 비교할 수 있습니다.

임의 효과 탭 임의 효과로 모델링할 효과를 입력합니다. 임의 효과는 수준이 더 큰 모집단의 랜덤 표본으로 간주되는 효과입니다. 이 효과로 인한 반응의 변동을 추정할 수 있습니다.

혼합 모형 다음과 같이 모형 효과를 입력합니다.

고정 효과 탭 고정 효과로 모델링할 효과를 입력합니다. 이 테이블에서 "표준 최소 제곱" 을 참조하십시오.

임의 효과 탭 임의 효과로 모델링할 효과를 입력합니다. 분산 성분 모형 및 랜덤 계수 모형에 사용합니다.

반복 구조 탭 반복 효과에 대한 공분산 구조를 선택하려면 사용합니다.

로그 선형 분산 다음과 같이 모형 효과를 입력합니다.

평균 효과 탭 기대값을 모델링할 효과를 입력합니다.

분산 효과 탭 분산을 모델링할 효과를 입력합니다.

효과의 기대값과 분산을 모두 모델링하려면 두 탭에 모두 입력해야 합니다.

모수 생존 다음과 같이 모형 효과를 입력합니다.

위치 효과 탭 위치 모수 μ 또는 형상 모수 (Weibull 분포의 경우) 를 모델링하는 데 사용할 효과를 입력합니다.

척도 효과 탭 척도 모수를 모델링하는 데 사용할 효과를 입력합니다.

적합 분석법

모형 적합 시작 창에서 분석법을 지정하여 적합 및 분석 방법을 선택합니다. JMP는 하나 이상의 반응과 입력된 요인을 기반으로 원하는 분석법에 대한 초기 컨텍스트 기반 추측을 수행하지만 "분석법" 메뉴에서 이 선택을 변경할 수 있습니다.

다음과 같은 적합 분석법을 사용할 수 있습니다.

표준 최소 제곱 반응이 연속형인 모형을 적합시킵니다. 회귀, 분산 분석, 공분산 분석, 혼합 모형 및 설계된 실험 분석 기법이 포함됩니다. 자세한 내용은 "**표준 최소 제곱 모형**" 장 (69 페이지) 및 "**표준 최소 제곱에 대한 강조 옵션**" (77 페이지) 에서 확인하십시오.

단계별 표준 최소 제곱 및 순서형 로지스틱 분석 (또는 이항 반응을 포함하는 명목형) 에 대한 변수 선택을 용이하게 합니다. 연속형 반응의 경우 교차 검증, p 값, BIC 및 AICc 기준이 제공됩니다. 모든 가능 모형을 적합시키고 모형 평균화를 위한 옵션도 제공됩니다. 로지스틱 적합의 경우 p 값, BIC 및 AICc 기준이 제공됩니다. 자세한 내용은 "**단계별 회귀 모형**" 장 (229 페이지) 에서 확인하십시오.

JMP PRO 일반화 회귀 정규화 (벌점이라고도 함) 회귀 기법을 사용하여 일반화 선형 모형을 적합시킵니다. 정규화 기법에는 능형 회귀, Lasso 회귀, 적응형 Lasso 회귀, Elastic Net 및 적응형 Elastic Net 이 포함됩니다. 반응 분포에는 정규, 이항, Poisson, 영과잉 Poisson, 음이항, 영과잉 음이항 및 감마 분포가 포함됩니다. 자세한 내용은 "**일반화 회귀 모형**" 장 (263 페이지) 및 "**분포**" (270 페이지) 에서 확인하십시오.

JMP PRO 혼합 모형 복소 공분산 구조를 가진 연속형 반응에 대해 다양한 선형 모형을 적합시킵니다. 다음과 같은 상황에서 사용됩니다.

- 분할구 실험
- 랜덤 계수 모형
- 반복 측정 설계
- 공간 데이터
- 상관관계가 있는 반응 데이터

자세한 내용은 "**혼합 모형**" 장 (327 페이지) 에서 확인하십시오.

MANOVA 여러 개의 연속형 Y 변수를 포함하는 모형을 적합시킵니다. 다변량 분산 분석, 반복 측정, 판별 분석 및 정준 상관 기법이 포함됩니다. 자세한 내용은 "**다변량 반응 모형**" 장 (403 페이지) 에서 확인하십시오.

로그 선형 분산 연속형 Y 변수의 경우 평균과 분산 모두에 대한 모형을 생성합니다. 두 모형에 대해 서로 다른 효과 집합을 지정할 수 있습니다. 자세한 내용은 "[로그 선형 분산 모형](#)" 장 (427 페이지) 에서 확인하십시오.

명목형 로지스틱 명목형 반응에 대해 로지스틱 회귀 모형을 적합시킵니다. 자세한 내용은 "[로지스틱 회귀 모형](#)" 장 (439 페이지) 에서 확인하십시오.

순서형 로지스틱 순서형 반응에 대해 로지스틱 회귀 모형을 적합시킵니다. 자세한 내용은 "[로지스틱 회귀 모형](#)" 장 (439 페이지) 에서 확인하십시오.

비례 위험 중도절단을 고려하여 설명 변수가 생존 시간에 미치는 영향을 평가하기 위해 준모수 회귀 모형 (Cox 비례 위험 모형) 을 적합시킵니다.

분석 > 신뢰성 및 생존 > 비례 위험 모형 적합을 선택하여 이 분석법을 시작할 수도 있습니다. 자세한 내용은 *Reliability and Survival Methods* 의 에서 확인하십시오.

모수 생존 생존 시간에 대해 일반 선형 회귀 모형을 적합시킵니다. 생존 시간을 하나 이상의 설명 변수에 대한 함수로 표현할 수 있는 경우 이 옵션을 사용합니다. 다양한 생존 분포 및 중도절단을 고려합니다.

분석 > 신뢰성 및 생존 > 모수 생존 모형 적합을 선택하여 이 분석법을 시작할 수도 있습니다. 자세한 내용은 *Reliability and Survival Methods* 의 에서 확인하십시오.

일반화 선형 모형 다양한 분포 및 연결 함수를 사용하여 일반화 선형 모형을 적합시킵니다. 로지스틱, Poisson 및 지수 회귀 기법이 포함됩니다. 자세한 내용은 "[일반화 선형 모형](#)" 장 (467 페이지) 에서 확인하십시오.

JMP PRO 부분 최소 제곱 잠재 요인을 사용하여 하나 이상의 Y 변수에 대해 모형을 적합시킵니다. 설명 변수 (X 변수) 의 상관관계가 높거나 관측값보다 X 변수가 더 많은 경우에 이 방법으로 모형을 적합시킬 수 있습니다.

분석 > 다변량 방법 > 부분 최소 제곱을 선택하여 부분 최소 제곱 분석을 시작할 수도 있습니다. 자세한 내용은 *Multivariate Methods* 의 에서 확인하십시오.

반응 변수 선별 여러 반응에 대해 선형 모형 효과의 검정 수행 과정을 자동화합니다. 검정 결과와 요약 통계량은 데이터 테이블과 그림에 제공됩니다. FDR(False Discovery Rate) 방법은 유의성이 잘못 선언되지 않도록 보호합니다. 로버스트 추정 방법은 이상치에 대한 검정 민감도를 줄여 줍니다.

참고 : 이 분석법은 연속형 반응만 허용합니다. **분석 > 선별 > 반응 변수 선별**을 선택하여 개별 요인에 대한 반응 변수 선별을 사용할 수도 있습니다. 이 플랫폼은 범주형 반응을 지원하고 동등성 검정 및 실제적 유의성 검정도 제공합니다. 자세한 내용은 *Predictive and Specialized Modeling* 의 에서 확인하십시오.

모형 규격 옵션

"모형 규격"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

중심화 다항식 차수가 1 보다 큰 효과에 포함된 연속형 항을 평균으로 중심화합니다. 이 옵션은 효과에 포함된 항에 "혼합물 효과" 속성이 할당되거나 "혼합물" 열 특성이 있는 경우를 제외하고 기본적으로 선택됩니다. "코딩" 열 특성이 있는 항은 지정된 상한값과 하한값 사이의 중간으로 중심화됩니다.

중심화는 회귀 계수를 더 쉽게 해석하고 낮은 차수 효과와 높은 차수 효과 사이의 공선성을 줄이는 데 유용합니다.

결측값 정보화 결측값에 대한 코딩 시스템을 제공합니다. 이 시스템을 사용하면 결측값이 있어도 예측 모형을 추정할 수 있습니다. 이 옵션은 결측 데이터에 정보가 포함된 경우 유용합니다. 자세한 내용은 "**결측값 정보화**"(46 페이지)에서 확인하십시오.

이 옵션은 표준 최소 제곱, 단계별, 일반화 회귀, MANOVA, 로그 선형 분산, 명목형 로지스틱, 순서형 로지스틱, 비례 위험, 모수 생존, 일반화 선형 모형 및 반응 변수 선별 분석법에서 사용할 수 있습니다.

유의 수준 설정 모형 적합 분석에 신뢰 구간의 유의 수준을 설정합니다. 기본 유의 수준은 0.05입니다.

오차 규격 (임의 효과가 없는 경우 표준 최소 제곱 분석법에만 사용 가능) "최소 제곱 적합" 보고서에서 표준 오차 및 검정에 사용되는 오차 분산과 오차 자유도를 지정합니다. "오차 규격"을 변경해도 "스튜던트화 잔차" 그림과 "Box Cox 변환" 보고서는 영향을 받지 않습니다. "오차 규격"이 "순수 오차" 또는 "지정됨"인 경우 "분산 분석" 보고서에 추가 열이 나타납니다. 자세한 내용은 "**분산 분석**"(86 페이지)에서 확인하십시오.

기본 추정값 모형의 표준 제곱근 평균 제곱 오차와 오차 자유도를 사용하여 모든 검정 및 표준 오차를 계산합니다.

순수 오차 "적합 결여" 보고서의 순수 오차 평균 제곱과 관련 자유도를 사용하여 모든 검정 및 표준 오차를 계산합니다. 자세한 내용은 "**적합 결여**"(104 페이지)에서 확인하십시오.

주의: 순수 오차 자유도가 1 이면 검정력이 약하고 신뢰 한계가 크다는 경고 메시지가 표시됩니다.

지정됨 사용자가 지정한 오차 분산 값과 오차 자유도를 사용하여 모든 검정 및 표준 오차를 계산합니다.

데이터 테이블에 저장 모형 적합 시작 창 규격을 데이터 테이블에 연결된 스크립트로 저장합니다. 스크립트 이름은 "Model" 입니다. 테이블에 "Model" 이라는 스크립트가 포함된 경우 **분석 > 모형 적합**을 선택하면 이 스크립트가 시작 창을 자동으로 채웁니다. 필요한 경우 스크립트 이름을 바꿀 수 있습니다.

JSL 스크립트에 대한 자세한 내용은 **Scripting Guide** 의 에서 확인하십시오.

스크립트 창에 저장 모형 적합 시작 창 규격을 스크립트 창에 복사합니다. 스크립트 창을 저장한 후 언제든지 스크립트를 실행하여 모형을 다시 생성할 수 있습니다.

SAS 작업 생성 스크립트 창에서 SAS 로 현재 분석 및 데이터 테이블을 다시 생성할 수 있는 SAS 프로그램을 생성합니다. 프로그램이 생성되면 SAS 로 코드를 전송하는 몇 가지 옵션을 사용할 수 있습니다.

1. 코드를 복사하여 SAS 프로그램 편집기에 붙여 넣습니다. 이 방법은 이전 버전의 SAS(8.2 이전 버전)를 실행 중인 경우에 유용합니다.
2. **편집 > SAS 로 전송**을 선택합니다.
3. 파일을 저장한 후 두 번 클릭하여 SAS 의 로컬 복사본에서 엽니다. 이 방법은 SAS ODS 옵션을 활용하려는 경우에 유용합니다 (예: SAS 코드에서 HTML 또는 PDF 출력 생성).

자세한 내용은 JMP 사용의 에서 확인하십시오.

SAS 로 전송 코드를 SAS 로 전송하고 JMP 에서 결과를 표시합니다. SAS 서버에 연결되어 있지 않으면 연결 과정을 안내하는 메시지가 표시됩니다.

자세한 내용은 JMP 사용의 에서 확인하십시오.

수렴 설정 "수렴 설정" 메뉴에는 다음 옵션이 포함되어 있습니다.

최대 반복 수 모형 적합에 사용되는 최대 반복 수를 지정합니다. 최대 반복 수의 기본값은 100 입니다. 모형이 쉽게 수렴되지 않는 경우 반복 수를 늘릴 수 있습니다. 데이터 집합이 매우 크거나 모형이 복잡한 경우에는 반복 수를 제한할 수 있습니다.

수렴 한계 모형 적합에 대한 수렴 한계를 지정합니다. 모형이 쉽게 수렴되지 않는 경우 수렴 한계를 늘릴 수 있습니다. 수렴 한계의 기본값은 0.00000001 입니다.

참고 : "수렴 설정" 메뉴는 특정 분석법에 대해서만 나타납니다. "표준 최소 제곱" 분석법을 사용할 경우 임의의 효과가 존재하고 시작 창에서 "방법"에 "REML"이 선택된 경우에만 "수렴 설정" 메뉴가 나타납니다.

결측값 정보화

"결측값 정보화" 옵션은 결측값이 있어도 예측 모형을 추정할 수 있는 코딩 시스템을 생성합니다. 연속형 모형 효과와 범주형 모형 효과를 모두 코드화합니다.

연속형 효과

연속형 주효과에 결측값이 있으면 새 설계 행렬 열이 생성됩니다. 이 열은 주효과 열이 결측이면 값이 1이고 결측이 아니면 값이 0인 표시 변수입니다. 또한 연속형 주효과의 결측값은 분석에 포함된 행의 비결측값 평균으로 대체됩니다. 이 평균은 모수 추정값의 해석력을 유지하는 중립 값입니다.

표시 변수와 관련된 모수는 결측값 그룹화를 통해 예측된 반응과 공변량이 평균으로 설정된 경우 예측된 반응 간의 차이를 추정합니다.

고차 효과의 경우 공변량의 결측값이 공변량 평균으로 대체됩니다. 이렇게 하면 "중심화 다항식"이 선택되었다는(기본 설정) 가정하에 결측값이 있는 행의 고차 효과가 0이 됩니다. "중심화 다항식"은 다항식에 포함된 개별 항을 평균으로 중심화하기 때문입니다.

"효과 검정" 보고서에서 결측값이 있는 각 연속형 주효과의 모수 수=2입니다. "모수 추정값" 보고서에서 결측값이 있는 연속형 주효과의 모수에는 <열 이름> 또는 평균(결측된 경우)이라는 라벨이 지정되고, 표시자 모수에는 <열 이름> 결측됨이라는 라벨이 지정됩니다. 데이터 테이블에 저장하는 예측 계산식은 이러한 모형 모수에 해당하는 표현식으로 제공됩니다.

범주형 효과

명목형 또는 순서형 주효과에 결측값이 있는 경우 결측값은 해당 효과의 개별 수준으로 코드화됩니다. 따라서 "효과 검정" 보고서에서 결측값이 있는 각 범주형 주효과에 하나의 추가 모수가 있습니다.

"모수 추정값" 보고서에서 명목형 효과의 모수에는 <열 이름>[]이라는 라벨이 지정됩니다. 순서형 효과의 모수에는 <열 이름>[-x]라는 라벨이 지정되며, 여기서 x는 순서 값이 가장 높은 수준을 나타냅니다.

연속형 효과와 마찬가지로 데이터 테이블에 저장하는 예측 계산식은 모형 모수에 해당하는 표현식으로 제공됩니다.

코딩 테이블

"표준 최소 제곱" 분석법을 사용하는 경우 **열 저장 > 코딩 테이블 저장**을 선택하여 결측값 정보화 모형에 사용된 설계 행렬 열을 볼 수 있습니다.

타당도 검사

모형 적합은 모형을 검사하여 계층에 중복 효과 또는 결측 효과와 같은 오류가 있는지 확인합니다. 경고 메시지가 표시되면 **계속**을 클릭하여 적합을 진행하거나 **취소**를 클릭하여 적합 과정을 중지할 수 있습니다.

모형 규격 및 모형 적합의 예

이 섹션에서는 모형 적합 플랫폼의 "모형 효과 생성" 패널을 사용하여 지정할 수 있는 다양한 모형 유형의 효과를 입력하기 위한 템플릿을 제공합니다.

- 모형 효과 X와 Z는 연속형 열을 나타냅니다.
- 모형 효과 A, B, C는 명목형 또는 순서형 열을 나타냅니다.

또한 대부분의 모형에 대해 모형 적합 보기가 시각적으로 표시됩니다.

- ["단순 선형 회귀"](#)

- "X 다항식 (k 차)"
- "X 및 Z 다항식 (k 차)"
- "다중 선형 회귀"
- "일원 분산 분석"
- "이원 분산 분석"
- "교호작용이 있는 이원 분산 분석"
- "삼원 완전 요인"
- "공분산 분석, 동일한 기울기"
- "공분산 분석, 서로 다른 기울기"
- "2 요인 내포 임의 효과 모형"
- "3 요인 완전 내포 임의 효과 모형"
- "단순 분할구 또는 반복 측정 모형"
- "2 요인 반응 표면 모형"
- "매듭이 지정된 스플라인 효과"

단순 선형 회귀

입력할 효과 : X

1. "열 선택" 목록에서 X 를 선택합니다.
2. **추가**를 클릭합니다.

단순 선형 회귀 모형의 예

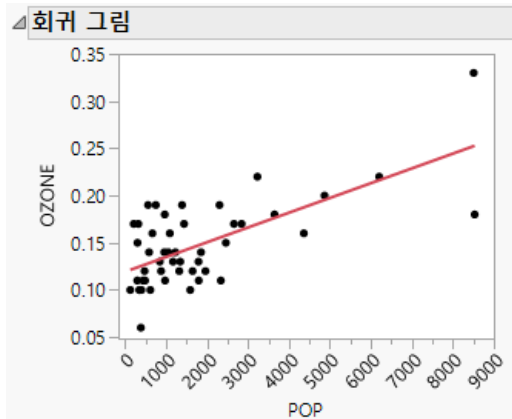
지정된 도시의 인구 (천 단위) 와 오존 수준 사이의 관계에 관심이 있습니다 . Ozone 은 관심 반응 변수이고 POP 는 연속형 모형 효과입니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Polycity.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. "열 선택" 목록에서 Ozone 을 선택하고 Y 를 클릭합니다 .
4. "열 선택" 목록에서 POP 를 선택합니다 .
5. **추가**를 클릭합니다 .

그림 2.4 단순 선형 회귀에 대한 모형 적합 창

실행을 클릭하면 회귀 그림을 포함한 다양한 결과를 보여 주는 "최소 제곱 적합" 보고서가 나타납니다. "회귀 그림"에는 데이터와 해당 데이터에 대한 단순 선형 회귀 모형 적합이 표시됩니다.

그림 2.5 단순 선형 회귀에 대한 모형 적합



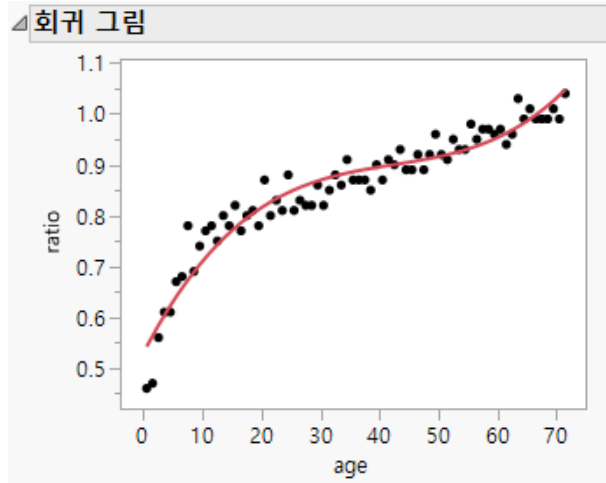
X 다항식 (k 차)

입력할 효과 : $X, X^2, \dots, X^k$

1. 차수 텍스트 상자에 k 를 입력합니다.
2. "열 선택" 목록에서 X 를 선택합니다.
3. 매크로 > 특정 차수까지의 다항식을 선택합니다.

그림 2.6에서는 Growth.jmp 샘플 데이터 테이블의 데이터와 해당 데이터에 대한 3차 다항식 모형 적합 그림을 보여 줍니다. 이것은 **Bivariate** 데이터 테이블 스크립트를 실행할 때 생성되는 적합 중 하나입니다.

그림 2.6 한 변수에서 3차 다항식에 대한 모형 적합



X 및 Z 다항식 (k 차)

입력할 효과 : $X, X^2, \dots, X^k, Z, Z^2, \dots, Z^k$

1. **차수** 텍스트 상자에 k 를 입력합니다.
2. "열 선택" 목록에서 X 와 Z 를 선택합니다.
3. **매크로 > 특정 차수까지의 다항식**을 선택합니다.

다중 선형 회귀

입력할 효과 : 선택한 열

1. "열 선택" 목록에서 관심 있는 연속형 효과를 선택합니다.
2. **추가**를 클릭합니다.

그림 2.7에서는 Grandfather Clocks.jmp 샘플 데이터 테이블의 데이터와 해당 데이터에 대한 다중 선형 회귀 적합이 포함된 표면 프로파일러 그림을 보여 줍니다. 모형 효과는 **Age**와 **Bidders**입니다. 반응은 **Price**입니다. **Fit Model with Surface Profiler Plot** 데이터 테이블 스크립트를 실행하여 그림을 확인할 수 있습니다.

그림 2.7 예측 변수가 두 개인 다중 선형 회귀 모형에 대한 모형 적합



다중 선형 회귀 모형의 예

여러 예측 변수가 있는 예는 "[모형 적합을 사용한 회귀 분석의 예](#)"(32 페이지)에서 확인하십시오.

일원 분산 분석

입력할 효과 : A

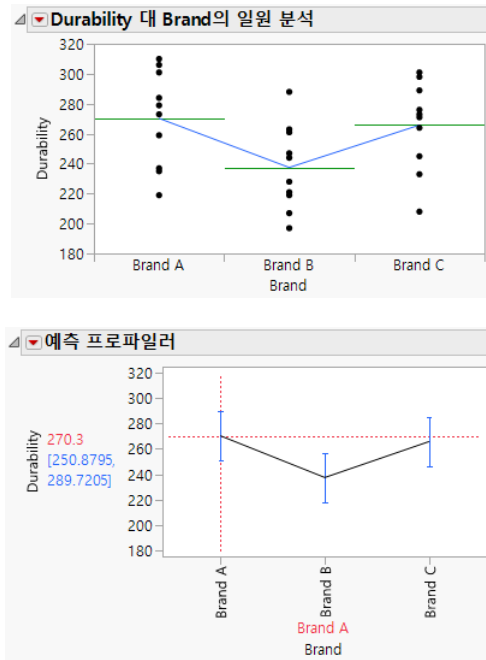
1. "열 선택" 목록에서 명목형 또는 순서형 효과 하나 (A) 를 선택합니다.
2. 추가를 클릭합니다.

Golf Balls.jmp 샘플 데이터 테이블을 예로 들어보겠습니다. Brand(브랜드)에 따라 Durability(내구성)가 다른지 여부에 관심이 있습니다. 그림 2.8에서는 두 개의 그림을 보여 줍니다.

첫 번째 그림은 "X로 Y 적합"을 사용하여 생성되며 브랜드별 데이터를 보여 줍니다. 각 브랜드의 평균에 수평선이 표시되고 선분이 평균을 연결합니다. 이 그림을 생성하려면 Golf Balls.jmp 샘플 데이터 테이블에서 **Oneway: Durability by Brand** 스크립트를 실행합니다.

두 번째 그림은 "모형 적합"을 사용하여 생성되는 프로파일러 그림입니다. 이 그림은 각 브랜드에 대한 예측 반응을 선분으로 연결하여 보여 줍니다. 이 그림을 생성하려면 Golf Balls.jmp 샘플 데이터 테이블에서 **Fit Model: Durability by Brand** 스크립트를 실행합니다. 빨간색 수직 파선을 관심 브랜드로 드래그합니다. 그러면 빨간색 수평 파선이 업데이트되어 예측 반응에서 세로 축과 교차합니다.

그림 2.8 일원 분산 분석에 대한 모형 적합



이원 분산 분석

입력할 효과 : A, B

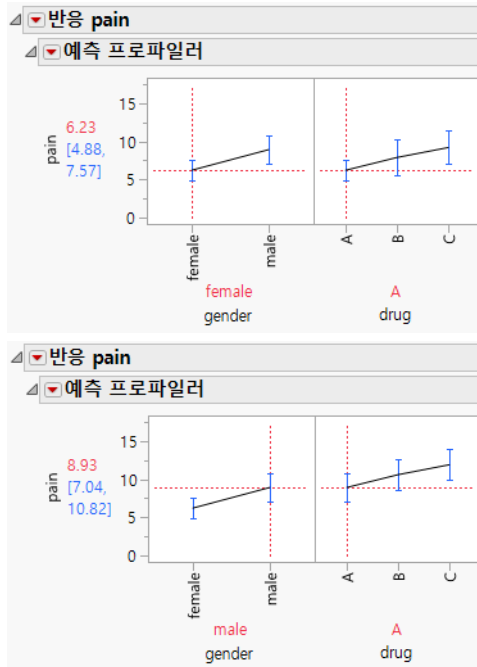
1. "열 선택" 목록에서 명목형 또는 순서형 효과 두 개 (A, B) 를 선택합니다.
2. **추가**를 클릭합니다.

그림 2.9에서는 Analgesics.jmp 샘플 데이터 테이블의 데이터 적합에 대한 두 개의 프로파일러 그림을 보여 줍니다. 모형 효과는 **gender**와 **drug**입니다. 반응은 **pain**입니다. 이 그림을 생성하려면 **분석 > 모형 적합**을 선택하고 **pain**을 Y로, **gender**와 **drug**를 모형 효과로 선택한 후 **실행**을 클릭합니다. 그런 다음 "반응 pain"의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 프로파일러**를 선택합니다.

각 그림의 선분은 빨간색 수직 파선으로 정의된 설정의 예측값을 연결합니다. 다른 설정에 대한 예측을 확인하려면 이 선분을 이동합니다.

그림 2.9의 위쪽 그림에는 여성에 대한 예측이 표시되고 아래쪽 그림에는 남성에 대한 예측이 표시됩니다. 세 가지 약물의 상대적 효과는 전체 **gender** 수준에서 일관적입니다. 이는 모형에 교호작용 항이 없기 때문입니다. 교호작용이 있는 예는 그림 2.11에서 확인하십시오.

그림 2.9 교호작용이 없는 이원 분산 분석에 대한 모형 적합



교호작용이 있는 이원 분산 분석

입력할 효과 : A, B, A*B

1. " 열 선택 " 목록에서 명목형 또는 순서형 효과 두 개 (A, B) 를 선택합니다 .
2. 매크로 > 완전 요인을 선택합니다 .

또는

1. " 열 선택 " 목록에서 명목형 또는 순서형 효과 두 개 (A, B) 를 선택합니다 .
2. 추가를 클릭합니다 .
3. " 열 선택 " 목록에서 A 와 B 를 다시 선택하고 교차를 클릭합니다 .

교호작용이 있는 이원 분산 분석의 예

팝콘 산출량 데이터가 있고 팝콘 유형과 배치 크기가 산출량에 영향을 주는지 여부에 관심이 있습니다 .

1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Popcorn.jmp 를 엽니다 .
2. 분석 > 모형 적합을 선택합니다 .
3. " 열 선택 " 목록에서 yield 를 선택하고 Y 를 클릭합니다 .
4. " 열 선택 " 목록에서 popcorn 과 batch 를 선택합니다 .

5. 매크로 > 완전 요인을 선택합니다.

그림 2.10 교호작용이 있는 이원 분산 분석에 대한 모형 적합 창

모형 규격

열 선택: 5개 열
 popcorn
 oil amt
 batch
 yield
 trial

역할 변수 선택: Y (yield) 선택적

분석법: 표준 최소 제곱
 강조: 효과 레버리지

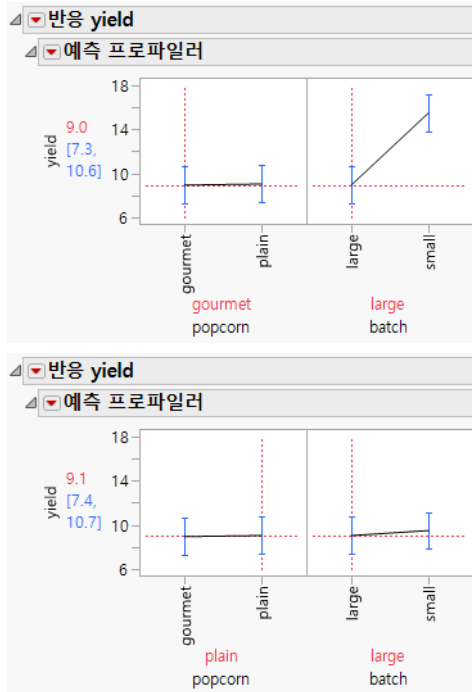
도움말 실행
 재호출 ☐ 대화상자 열린 채 유지
 제거

모형 효과 생성
 추가: popcorn, batch, popcorn*batch
 교차
 내포
 매크로
 자수: 2
 속성: 속성
 변환: 변환
☒ 결편 없음

그림 2.11에서는 이 예에 대한 적합 프로파일러 그림을 보여 줍니다. 이 그림을 생성하려면 그림 2.10에 표시된 "모형 적합" 창에서 **실행**을 클릭합니다. 그런 다음 "반응 yield"의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 프로파일러**를 선택합니다.

popcorn이 위쪽 그림에는 gourmet로 설정되고 아래쪽 그림에는 plain으로 설정되어 있습니다. batch 설정에 대한 예측값이 popcorn 유형에 따라 어떻게 달라지는지 확인하십시오. 이것은 popcorn과 batch 간 교호작용의 결과입니다.

그림 2.11 교호작용이 있는 이원 분산 분석에 대한 모형 적합



삼원 완전 요인

입력할 효과 : A, B, C, A*B, A*C, B*C, A*B*C

1. "열 선택" 목록에서 명목형 또는 순서형 효과 세 개 (A, B, C) 를 선택합니다.
2. 매크로 > 완전 요인을 선택합니다.

삼원 완전 요인의 예

Speed(속도), Angle(각도), Material(재료) 또는 이러한 요인의 교호작용이 절삭 공구의 Wear(마모) 에 영향을 주는지 여부에 관심이 있습니다.

1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Tool Wear.jmp 를 엽니다.
2. 분석 > 모형 적합을 선택합니다.
3. "열 선택" 목록에서 Wear 를 선택하고 Y 를 클릭합니다.
4. "열 선택" 목록에서 Speed, Angle, Material 을 선택합니다.
5. 매크로 > 완전 요인을 선택합니다.

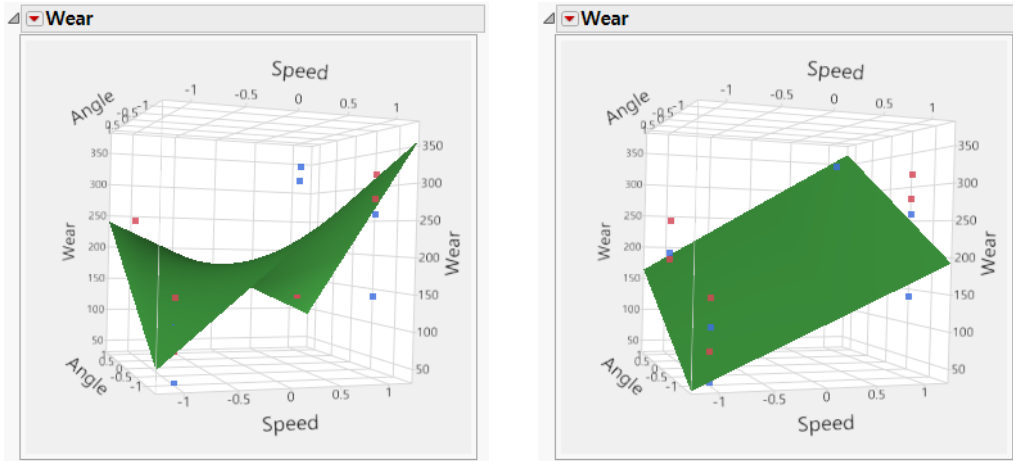
그림 2.12 삼원 완전 요인에 대한 모형 적합 창

그림 2.13의 표면 프로파일러 그림에서는 두 가지 연속형 효과, 즉 Speed와 Angle 측면에서 Wear에 대한 예측 반응을 보여 줍니다. 왼쪽 그림에는 Material이 A일 때의 예측 반응이 표시되고 오른쪽 그림에는 Material이 B일 때의 예측 반응이 표시됩니다. Material이 A인 점은 빨간색으로 표시되고 Material이 B인 점은 파란색으로 표시됩니다. Material 수준에 따른 반응 표면 형태의 차이는 삼원 교호작용의 결과입니다.

표면 프로파일러 그림을 생성하려면 그림 2.12에 표시된 "모형 적합" 창에서 **실행**을 클릭합니다. 그런 다음 "반응 Wear"의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 표면 프로파일러**를 선택합니다. 그림에 점을 추가하려면 "표면 모양" 패널을 열고 **실제값**을 클릭합니다. 점을 더 크게 표시하려면 그림에서 마우스 오른쪽 버튼을 클릭하고 **설정**을 선택한 후 **표식 크기**를 조정합니다.

Material A와 B 둘 다에 대한 그림을 표시하려면 "독립 변수" 패널에서 Material로 표시된 슬라이더를 사용하여 Material A에는 0을 설정하고 Material B에는 1을 설정합니다. 테이블에는 표면 프로파일러 그림을 생성하는 두 개의 데이터 테이블 스크립트(**Prediction and Surface Profilers, Surface Profilers for Two Materials**)가 포함되어 있습니다.

그림 2.13 삼원 완전 요인 설계에 대한 모형 적합 - Material A(왼쪽), Material B(오른쪽)



공분산 분석, 동일한 기울기

X를 공변량으로 사용하여 A의 효과를 검정하는 데 관심이 있습니다. 반응에 대한 X의 효과가 A의 수준에 따라 달라지지 않는다고 믿을 만한 이유가 있다고 가정해 보겠습니다.

입력할 효과 : A, X

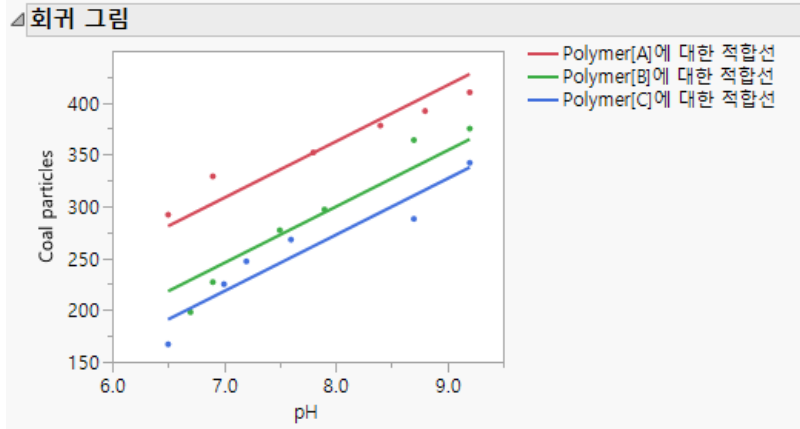
1. "열 선택" 목록에서 명목형 또는 순서형 효과 하나(A)와 연속형 효과 하나(X)를 선택합니다.
2. **추가**를 클릭합니다 .

그림 2.14에서는 Cleansing.jmp 샘플 데이터 테이블의 데이터를 보여 줍니다. 세척 탱크에서 어떤 Polymer(폴리머)가 Coal particles(석탄 입자)를 가장 많이 제거하는지에 관심이 있습니다. 그러나 탱크의 pH도 제거에 영향을 주는 것으로 의심됩니다. 이 그림은 공분산 분석 적합을 보여 줍니다. 여기서 pH와 Coal particles를 연결하는 기울기는 전체 Polymer 수준에서 동일하다고 가정합니다.

다음 단계에 따라 그림 2.14의 그림을 생성할 수 있습니다. "모형 적합" 창에서 Coal particles를 **Y**로 입력하고 pH와 Polymer 둘 다 **모형 효과 생성** 상자에 입력합니다. **실행**을 클릭합니다. "최소 제곱 적합" 보고서에 "회귀 그림"이 나타납니다. 점에 색상을 지정하려면 **행 > 열 값에 따른 색상 또는 표식**을 선택하고, **열 값에 따른 표식** 목록에서 Polymer를 선택한 후 **확인**을 클릭합니다.

그러나 더 완전한 분석은 pH와 Polymer가 Coal particles에 미치는 영향에서 상호 작용한다는 것을 나타냅니다. 해당 모형 적합은 "**공분산 분석, 서로 다른 기울기**"(58 페이지)에 나와 있습니다.

그림 2.14 공분산 분석, 동일한 기울기에 대한 모형 적합



공분산 분석, 서로 다른 기울기

여기서도 X를 공변량으로 사용하여 A의 효과를 검정하는 데 관심이 있습니다. 그러나 반응에 대한 X의 효과가 A의 수준에 따라 달라질 가능성을 허용하도록 모형을 생성합니다.

입력할 효과 : A, X, A*X

1. "열 선택" 목록에서 명목형 또는 순서형 효과 하나(A)와 연속형 효과 하나(X)를 선택합니다.
2. 매크로 > 완전 요인을 선택합니다.

또는

1. "열 선택" 목록에서 명목형 또는 순서형 효과 하나(A)와 연속형 효과 하나(X)를 선택합니다.
2. 추가를 클릭합니다.
3. "열 선택" 목록에서 A 와 X 를 다시 선택하고 교차를 클릭합니다.

공분산 분석, 서로 다른 기울기의 예

세 가지 폴리머 (Polymer) 중 석탄 입자 제거 (Coal particles) 에 영향을 미치는 종류가 있는지 여부에 관심이 있습니다. 탱크 pH 는 폴리머의 탱크 세척력에 영향을 줄 수 있으므로 공변량으로 포함됩니다. 세 가지 Polymer 유형에 대해 pH 와 Coal particles 간의 관계를 모델링할 때 기울기가 다를 수 있다는 점을 고려합니다.

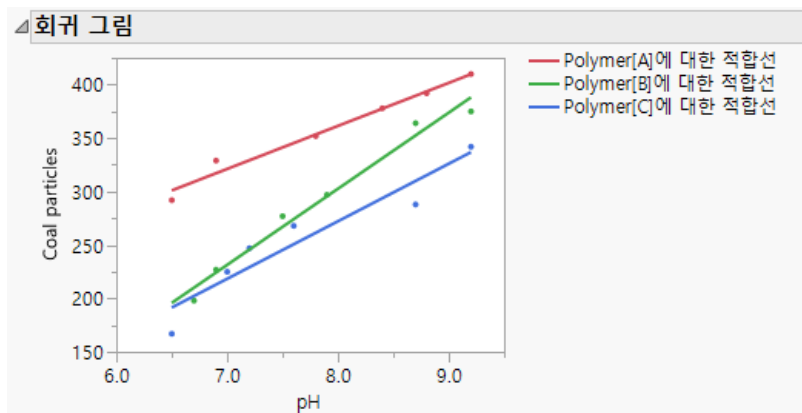
1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Cleansing.jmp 를 엽니다.
2. 분석 > 모형 적합을 선택합니다.
3. "열 선택" 목록에서 Coal particles 를 선택하고 Y 를 클릭합니다.
4. "열 선택" 목록에서 pH 와 Polymer 를 선택합니다.
5. 매크로 > 완전 요인을 선택합니다.

그림 2.15 공분산 분석, 서로 다른 기울기에 대한 모형 적합 창

실행을 클릭하면 "최소 제곱 적합" 보고서가 나타납니다. "효과 검정" 보고서는 pH와 Polymer 간의 교호작용이 유의하므로 모형에 포함되어야 함을 나타냅니다.

보고서에 제공된 회귀 그림이 그림 2.16에 나와 있습니다. 이 그림은 점과 모형 적합을 보여 줍니다. 교호작용으로 인해 pH와 Coal particles를 연결하는 선의 기울기가 Polymer에 따라 달라질 수 있습니다. 이 교호작용에 상관없이 관심 범위에 대해 Polymer A가 일관되게 가장 높은 제거율을 보인다는 점에 유의하십시오. 그림 2.16에 표시된 것처럼 점에 색상을 지정하려면 **행 > 열 값에 따른 색상 또는 표식**을 선택하고, **열 값에 따른 표식** 목록에서 Polymer를 선택한 후 **확인**을 클릭합니다.

그림 2.16 공분산 분석, 서로 다른 기울기에 대한 모형 적합



2 요인 내포 임의 효과 모형

두 가지 요인 A와 B가 있지만 B가 A에 내포된 모형을 예로 들어보겠습니다. 내포 효과가 고정 효과로 처리되는 경우도 있지만 대부분의 상황에서 내포 효과는 임의 효과로 처리됩니다. 따라서 아래 설명된 모형에서는 내포 효과가 임의 효과로 입력됩니다.

입력할 효과 : A, B[A]& 랜덤

1. "열 선택" 목록에서 명목형 또는 순서형 효과 두 개 (A, B) 를 선택합니다.
2. **추가**를 클릭합니다.
3. A 안에 B 내포 : "모형 효과 생성" 목록에서 B 를 선택하고 "열 선택" 목록에서 A 를 선택합니다. 두 효과가 강조 표시되어야 합니다.
4. **내포**를 클릭합니다.
5. "모형 효과 생성" 목록에서 B[A] 가 강조 표시된 상태로 **속성 > 임의 효과**를 선택합니다.

2 요인 내포 임의 효과 모형의 예

측정 시스템 분석 연구의 일부로 무작위 선택된 24개의 부품을 측정합니다. 이 부품은 일반적으로 이러한 부품을 측정하는 6명의 작업자에게 균등하게 분배됩니다. 각 작업자는 할당된 4개 부품 각각에 대해 세 가지 측정을 독립적으로 수행합니다.

한 명의 작업자가 측정한 부품은 해당 작업자에 의해서만 측정되므로 **Part** 는 **Operator** 에 내포됩니다. 부품은 랜덤 생산 표본이므로 **Part** 가 임의 효과로 간주됩니다. 이러한 특정 작업자 6 명에게 관심이 있으므로 **Operator** 는 고정 효과로 처리됩니다. 적절한 모형을 지정합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Variability Data/2 Factors Nested.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. "열 선택" 목록에서 Y 를 선택하고 **Y** 를 클릭합니다.
4. "열 선택" 목록에서 **Operator** 와 **Part** 를 선택합니다.
5. **추가**를 클릭합니다.
6. **Operator** 안에 **Part** 내포 : "모형 효과 생성" 목록에서 **Part** 를 선택합니다. "열 선택" 목록에서 **Operator** 를 선택합니다. 두 효과가 강조 표시되어야 합니다.
7. **내포**를 클릭합니다.
8. "모형 효과 생성" 목록에서 Part[Operator] 가 강조 표시된 상태로 **속성 > 임의 효과**를 선택합니다.

그림 2.17 2 요인 내포 임의 효과 모형에 대한 모형 적합 창

9. **실행**을 클릭하여 "최소 제곱 적합" 보고서를 생성합니다.

그림 2.18에서는 두 개의 그림을 보여 줍니다. 첫 번째 그림은 4개 부품 각각에 대한 각 Operator의 세 가지 측정값을 보여 주는 계량형 차트입니다. 수평 선분은 각 Operator의 평균 측정값을 표시합니다.

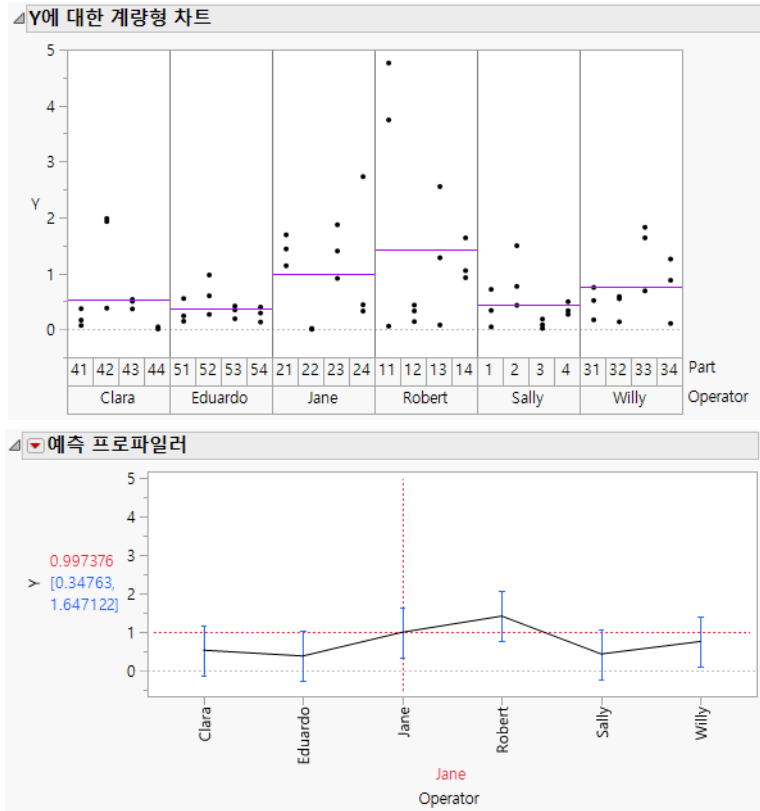
그림 2.18의 계량형 차트를 생성하려면 2 Factors Nested.jmp 샘플 데이터 테이블에서 **Variability Chart - Nested** 데이터 테이블 스크립트를 실행합니다. "계량형 게이지"의 빨간색 삼각형을 클릭한 후 **범위 막대 표시**를 선택 취소하고 **그룹 평균 표시**를 선택합니다.

두 번째 그림은 "최소 제곱 적합" 보고서에서 Operator에 대한 "예측 프로파일러" 그림입니다. 이 그림은 각 작업자에 대한 예측 반응을 보여 줍니다. Jane에 설정된 빨간색 수직 파선은 Jane의 예측 반응이 0.997임을 나타냅니다. "예측 프로파일러" 그림에 제공된 모형 예측과 "계량형 차트"의 원시 데이터 간 대응 관계를 확인할 수 있습니다.

예측 프로파일러 그림을 생성하려면 "반응 Y"의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 프로파일러**를 선택합니다.

이러한 그림은 각 Operator에 대한 예측 측정값을 모델링하는 방법을 보여 줍니다. 그러나 작업자의 부품 측정 방식이 다른지 여부에만 관심이 있는 것은 아닙니다. 부품 측정 자체의 변동성에도 관심이 있으므로 Part와 관련된 분산 성분을 추정해야 합니다.

그림 2.18 2 요인 내포 임의 효과 모형에 대한 모형 적합



3 요인 완전 내포 임의 효과 모형

세 가지 요인 A, B, C가 있지만 B는 A에 내포되고 C는 A와 B 둘 다에 내포된 모형을 예로 들어 보겠습니다. 또한 B와 C를 임의 효과로 간주합니다.

입력할 효과 : A, B[A]& 랜덤, C[A,B]& 랜덤

1. "열 선택" 목록에서 명목형 또는 순서형 효과 세 개 (A, B, C) 를 선택합니다.
2. **추가**를 클릭합니다.
3. A 안에 B 내포 : "모형 효과 생성" 목록에서 B를 선택하고 "열 선택" 목록에서 A를 선택합니다. 두 효과가 강조 표시되어야 합니다.
4. **내포**를 클릭합니다.
5. A와 B 안에 C 내포 : "모형 효과 생성" 목록에서 C를 선택하고 "열 선택" 목록에서 A와 B를 선택합니다. 세 효과가 강조 표시되어야 합니다.
6. **내포**를 클릭합니다.

7. " 모형 효과 생성 " 목록에서 B[A] 와 C[A,B] 가 강조 표시된 상태로 **속성 > 임의 효과**를 선택합니다.

단순 분할구 또는 반복 측정 모형

여기서 A는 주구 변수, B[A]는 주구 ID, C는 분할구 또는 반복 측정 변수입니다.

입력할 효과 : A, B[A]& 랜덤 , C, C*A

1. " 열 선택 " 목록에서 명목형 또는 순서형 효과 두 개 (A, B) 를 선택합니다.
2. **추가**를 클릭합니다.
3. A 안에 B 내포 : " 모형 효과 생성 " 목록에서 B 를 선택하고 " 열 선택 " 목록에서 A 를 선택합니다. 두 효과가 강조 표시되어야 합니다.
4. **내포**를 클릭합니다.
5. " 모형 효과 생성 " 목록에서 B[A] 를 선택합니다.
6. **속성 > 임의 효과**를 선택합니다.
7. " 열 선택 " 목록에서 세 번째 명목형 또는 순서형 효과 (C) 를 선택합니다.
8. **추가**를 클릭합니다.
9. " 모형 효과 생성 " 목록에서 C 를 선택하고 " 열 선택 " 목록에서 A 를 클릭합니다. 두 효과 모두 강조 표시되어야 합니다.
10. **교차**를 클릭합니다.

단순 반복 측정 모형의 예

사계절 중 각 계절마다 6 마리 동물 각각의 이동 거리에 대한 데이터가 있습니다. 두 가지 종 (species) 이 있으며 동물 식별자는 species 에 내포되어 있습니다. 6 마리의 동물은 더 큰 종 개체군을 대표하므로 subject 를 임의 효과로 처리하기로 합니다. 각 동물에 대해 반복 측정이 있다는 사실을 고려하여 반응 (miles) 을 species 와 season 의 함수로 모델링하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Animals.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. " 열 선택 " 목록에서 miles 를 선택하고 **Y** 를 클릭합니다.
4. " 열 선택 " 목록에서 species 와 subject 를 선택합니다.
5. **추가**를 클릭합니다.
6. species 안에 subject 내포 : " 모형 효과 생성 " 목록에서 subject 를 선택합니다. " 열 선택 " 목록에서 species 를 선택합니다. 두 효과가 강조 표시되어야 합니다.
7. **내포**를 클릭합니다.
8. " 모형 효과 생성 " 목록에서 subject[species] 를 선택합니다.
9. **속성 > 임의 효과**를 선택합니다.

10. "열 선택" 목록에서 **season** 을 선택합니다.
11. **추가**를 클릭합니다.
12. "모형 효과 생성" 목록에서 **season** 을 선택합니다. "열 선택" 목록에서 **species** 를 클릭합니다. 두 효과 모두 강조 표시되어야 합니다.
13. **교차**를 클릭합니다.

그림 2.19 단순 반복 측정 모형에 대한 모형 적합 창

2 요인 반응 표면 모형

입력할 효과 : $X \& RS$, $Z \& RS$, $X * X$, $X * Z$, $Z * Z$

1. "열 선택" 목록에서 연속형 효과 두 개 (X , Z) 를 선택합니다.
2. **매크로 > 반응 표면**을 선택합니다.

2 요인 반응 표면 모형의 예

Odor Control Original.jmp 샘플 데이터 테이블에서 반응 표면을 적합시켜 **response**, **odor** 을 **temp** 및 **gl ratio** 의 함수로 모델링하려고 합니다. **Response Surface** 데이터 테이블 스크립트와 같이 **ht** 를 포함할 수도 있지만 이 예에서는 포함하지 않습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Odor Control Original.jmp** 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. "열 선택" 목록에서 **odor** 을 선택하고 **Y** 를 클릭합니다.
4. "열 선택" 목록에서 **temp** 와 **gl ratio** 를 선택합니다.

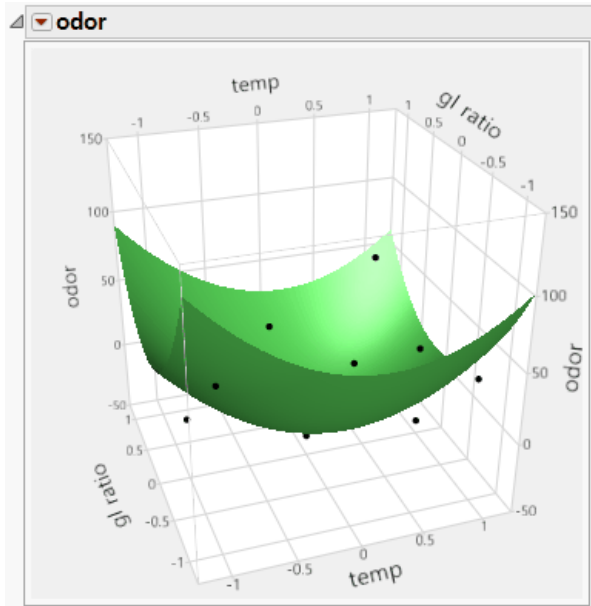
5. 매크로 > 반응 표면을 선택합니다.

그림 2.20 2 요인 반응 표면 모형에 대한 모형 적합 창

6. 실행을 클릭합니다.

그림 2.21에서는 Odor Control Original.jmp 샘플 데이터 테이블의 데이터와 해당 데이터에 대한 2차 반응 표면 적합이 포함된 표면 프로파일러 그림을 보여 줍니다. 이 그림을 생성하려면 "반응 odor"의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 표면 프로파일러**를 선택합니다. 점을 표시하려면 표시 아이콘을 클릭하여 **표면 모양** 패널을 열고 **실제값**을 클릭합니다. 점을 더 크게 표시하려면 그림에서 마우스 오른쪽 버튼을 클릭하고 **설정**을 선택한 후 **표식 크기**를 조정합니다.

그림 2.21 두 변수에서 2 차 다항식에 대한 모형 적합



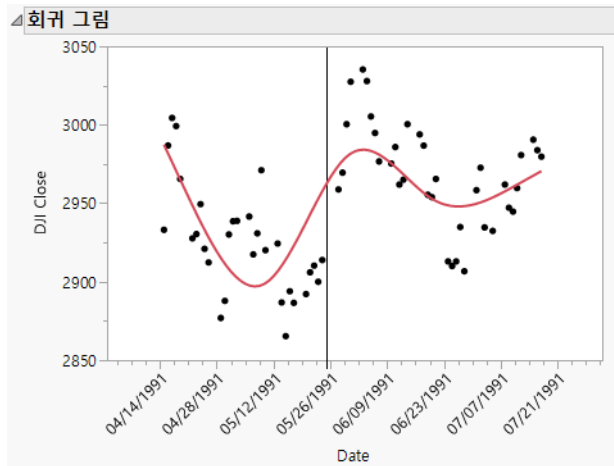
매듭이 지정된 스플라인 효과

입력할 효과 : X& 매듭 지정

1. " 열 선택 " 목록에서 연속형 효과 X 를 선택합니다 .
2. **추가**를 클릭합니다 .
3. " 모형 효과 생성 " 목록에서 X 를 선택합니다 .
4. **속성 > 매듭이 지정된 스플라인 효과**를 선택합니다 .
5. 메뉴가 나타나면 매듭 수를 지정하거나 기본값을 적용합니다 .
6. **확인**을 클릭합니다 .

그림 2.22에서는 XYZ Stock Averages (plots).jmp 샘플 데이터 테이블의 데이터에 대한 모형 적합이 포함된 회귀 그림을 보여 줍니다. 여기서 **Date**에는 매듭이 지정된 스플라인 효과가 할당되고 5개의 매듭이 지정되었습니다. DJI Close가 반응입니다.

그림 2.22 5 개의 매듭이 있는 매듭 지정 스플라인에 대한 모형 적합

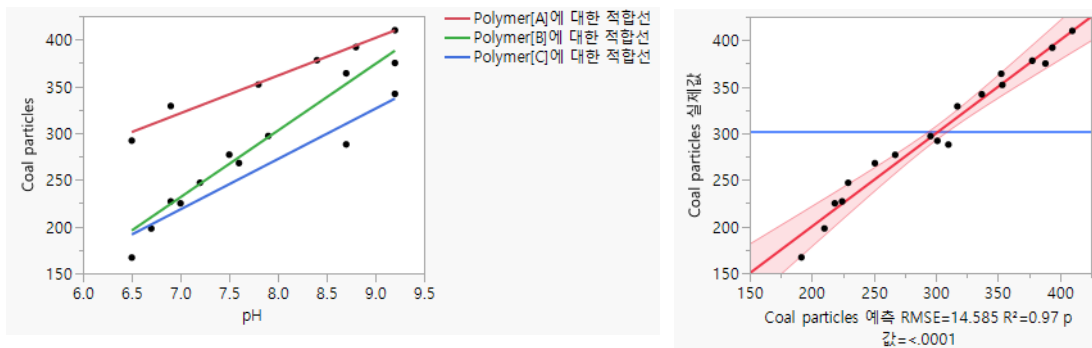


표준 최소 제곱 모형 모형의 공통 클래스 분석

모형 적합 플랫폼 내의 표준 최소 제곱 분석법은 광범위한 표준 모형을 적합시킵니다. 이러한 모형에는 설계된 실험을 분석하는 데 일반적으로 사용되는 모형뿐만 아니라 회귀, 분산 분석, 공분산 분석 및 혼합 모형이 포함됩니다. 표준 최소 제곱 분석법을 사용하면 최소 제곱을 사용하거나, 임의 효과의 경우 REML(제한 최대 가능도)을 사용하여 연속형 반응 데이터에 대한 선형 모형을 생성할 수 있습니다.

분석 결과는 프로파일러, 등고선 그림 및 표면 그림과 같은 강력한 동적 시각화 도구로 지원됩니다 (Profilers 참조). 이러한 시각적 표시는 모형에 대한 관심을 가지게 하거나 모형에 대한 이해를 보완하고 지원합니다. 이를 통해 여러 반응을 동시에 최적화하고 잡음 효과를 탐색할 수 있습니다.

그림 3.1 표준 최소 제곱 그림의 예



목차

예 : 표준 최소 제공 사용	72
표준 최소 제공 분석법 시작	74
모형 적합 시작 창	75
모형 적합 시작 창의 표준 최소 제공 옵션	76
표준 최소 제공의 검증	78
결측값	78
최소 제공 적합 보고서	79
단일 반응과 다중 반응의 비교	79
강조와 관련된 보고서 구조	79
특수 보고서	79
최소 제공 적합 옵션	83
적합 그룹 옵션	83
반응 옵션	84
회귀 보고서	85
적합 요약	85
분산 분석	86
모수 추정값	87
효과 검정	88
효과 상세 정보	89
적합 결여	104
추정값	105
예측 표현식 표시	107
정렬된 추정값	108
확장 추정값	111
표시자 파라미터화 추정값	113
순차 검정	114
사용자 검정	115
다중 비교	117
기울기 비교	130
결합 요인 검정	130
역추정 예측	131
Cox 혼합물	135
모수 검정력	136
추정값 상관계수	138
명목형 효과에 대한 코딩	139
효과 선별	140
척도화된 추정값 및 연속형 항 코딩	141
그림 옵션	142

정규 확률도 보고서.....	146
Bayes 그림 보고서	147
파레토도 보고서	149
요인 프로파일링	150
프로파일러.....	151
교호작용 그림	151
등고선 프로파일러	153
혼합물 프로파일러	154
입방체도.....	155
Y의 Box-Cox 변환.....	156
표면 프로파일러	158
행 진단	159
레버리지 그림	161
Press.....	164
열 저장	165
예측 계산식	167
효과 요약 보고서	168
혼합 및 임의 효과 모형 보고서 및 옵션	172
혼합 모형 및 임의 효과 모형	172
REML(제한 최대 가능도) 방법	176
EMS(전통적) 모형 적합 보고서.....	181
모형 항 사이에 선형 종속성이 있는 모형	182
특이성 상세 정보.....	183
모수 추정값 보고서.....	184
효과 검정 보고서.....	185
예.....	185
표준 최소 제곱 분석법에 대한 통계 상세 정보.....	187
강조 규칙	187
사용자 검정 예에 대한 상세 정보.....	188
추정값 상관계수	189
레버리지 그림 상세 정보	189
Kackar-Harville 수정.....	192
검정력 분석	193

예 : 표준 최소 제곱 사용

질병 치료에 약물이 미치는 영향에 대한 연구에서 30명의 환자를 무작위로 10명씩 세 그룹으로 나누었습니다. 이 중 두 그룹에는 약물(Drug a, Drug d)이 투여되고 세 번째 그룹에는 플라시보(Drug f)가 처리되었습니다. 각 환자에 대해 전처리 측정(x) 및 후처리 측정(y)이 수행되었습니다. 전처리 점수 x 는 환자 간의 질병 단계 차이를 설명하기 위해 공변량으로 포함됩니다. 이 예는 Snedecor and Cochran 연구 자료(1967, p. 422)에서 가져왔습니다.

세 가지 약물(Drug) 그룹에 차이가 있는지 확인하려고 합니다. y 반응, Drug, x 모형 효과 및 Drug와 x 의 교호작용으로 모형을 생성합니다. 교호작용은 질병 단계에 따라 약물 효과에 차이가 있는 상황을 설명할 수 있습니다. 모형 적합 창과 다양한 분석법에 대한 배경 정보는 "[모형 규격](#)" 장(29 페이지)에서 확인하십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Drug.jmp를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. **Y**를 선택하고 **Y**를 클릭합니다.

이 열을 **Y**로 추가하면 적합 **분석법**이 "표준 최소 제곱"으로 설정됩니다. "효과 레버리지"가 선택된 **강조** 옵션이 추가되며 원하는 경우 변경할 수 있습니다.

4. Drug와 x 를 선택합니다. "열 선택" 목록에 이 두 효과가 강조 표시된 상태에서 **매크로**를 클릭하고 **완전 요인**을 선택합니다. 매크로는 이 두 효과 및 효과의 이원 교호작용을 "모형 효과 생성" 목록에 추가합니다.

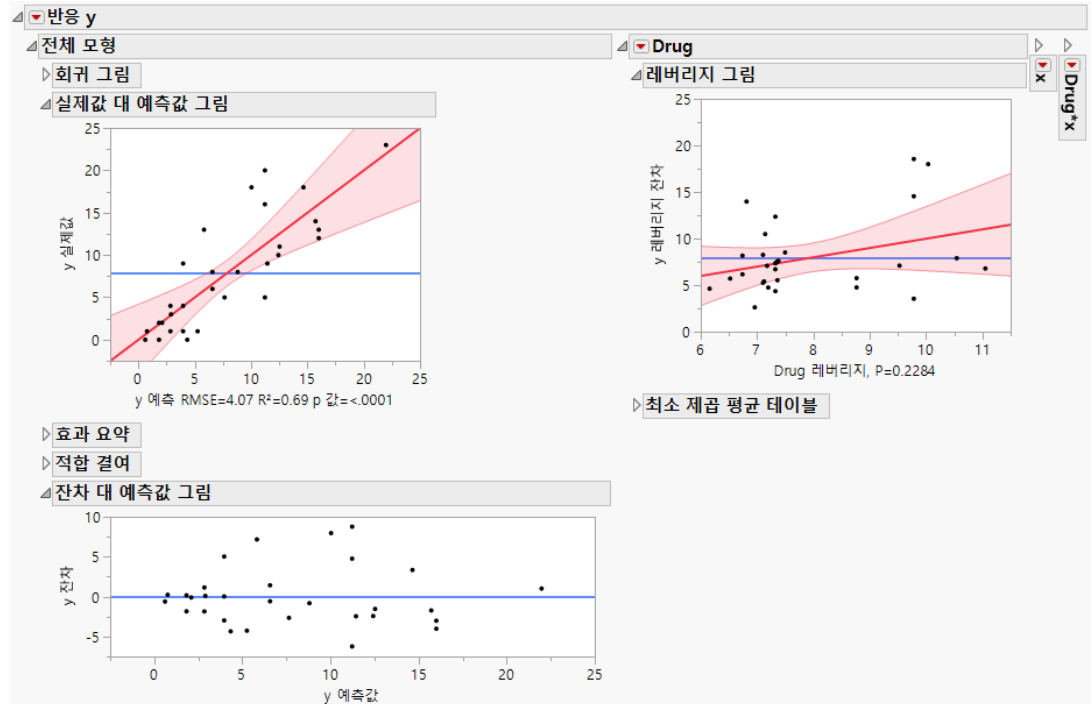
그림 3.2 완료된 모형 적합 시작 창

The screenshot shows the 'Model Specifier' dialog box in JMP. The 'Y' variable is 'y'. The 'X' variables are 'Drug' and 'x'. The 'Model' list contains 'Drug', 'x', and 'Drug*x'. The 'Analysis Method' is 'Standard Least Squares'. The 'Emphasis' is 'Effect Leverage'. The 'Number of Terms' is 2. The 'Save' checkbox is checked. The 'Delete' checkbox is unchecked.

5. **실행**을 클릭합니다.

그림 3.3에 "최소 제곱 적합" 보고서가 표시되어 있습니다. 포함된 보고서 중 일부는 공간을 고려하여 닫혀 있습니다. "실제값 대 예측값 그림", "잔차 대 예측값 그림" 및 "레버리지 그림"은 모형 적합 및 기본 가정에 대해 불일치를 나타내지 않습니다.

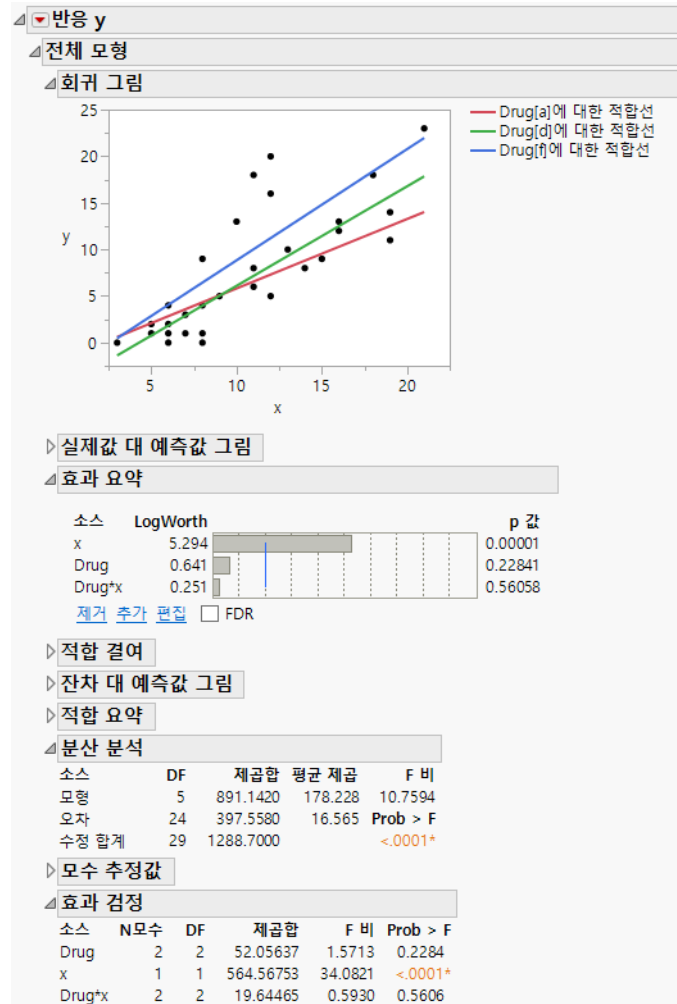
그림 3.3 모형 적합 평가를 위한 그림을 보여 주는 최소 제곱 적합 보고서



모형 적합에 명백한 문제가 없으므로 통계 검정을 해석할 수 있습니다. 그림 3.4에서는 관련 보고서를 보여 줍니다. 전체 모형은 "분산 분석" 보고서에 표시된 대로 유의합니다.

"회귀 그림"에 따르면 Drug와 전처리 측정 x가 상호 작용하지만 "효과 검정" 보고서의 "Prob > F" 값은 이러한 결론을 지지하지 않습니다. 또한 "효과 검정" 보고서에서는 x가 y를 설명하는 데 유의하지만 Drug는 유의하지 않음을 보여 줍니다. 이 연구를 수행하여 세 그룹 간의 차이를 감지하지 못했습니다. 그러나 Drug에 효과가 없다는 결론을 내릴 수는 없습니다. 약물의 효과는 서로 다를 수 있지만 연구 규모가 그 차이를 감지할 만큼 크지 않았습니다.

그림 3.4 유의성 평가를 위한 보고서를 표시하는 최소 제곱 적합 보고서



표준 최소 제곱 분석법 시작

표준 최소 제곱은 모형 적합 시작 창에서 선택할 수 있는 여러 분석 기법 중 하나입니다.

이 섹션에서는 모형 적합 시작 창에서 표준 최소 제곱을 적합 방법으로 선택하는 방법을 설명합니다. 이 선택 항목과 관련된 옵션에 대한 내용도 다룹니다. "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

모형 적합 시작 창

모형 적합 시작 창에서 고정 효과와 임의 효과가 모두 있는 모형을 지정할 수 있습니다. 옵션은 사용자가 지정한 모형의 특성에 따라 다릅니다.

고정 효과만

표준 최소 제곱 분석법을 사용하여 모형을 적합시키려면 **분석 > 모형 적합**을 선택한 후 "분석법" 목록에서 **표준 최소 제곱**을 선택합니다. Y 목록에 하나 이상의 연속형 변수를 입력하면 분석법이 기본적으로 "표준 최소 제곱"으로 설정됩니다. 그러나 연속형 Y 변수에 대해 다른 항목을 선택할 수 있습니다. 표준 최소 제곱 적합에 대해 고정 효과만 지정하면 "모형 적합" 시작 창이 그림 3.5에 표시된 것과 같이 나타납니다. 이 예에서는 Big Class.jmp 샘플 데이터 테이블을 사용한 시작 창을 보여 줍니다.

그림 3.5 고정 효과 모형에 대한 모형 적합 시작 창

"분석법" 목록에서 "표준 최소 제곱" 분석법이 선택되면 "강조" 옵션도 나타납니다. "강조" 옵션은 초기 보고서 창에 제공되는 보고서를 제어합니다. JMP는 포함된 모형 효과를 기반으로 사용자에게 필요한 보고서를 추론합니다. 그러나 초기 보고서의 일부로 표시되지 않은 보고서도 기본 보고서의 빨간색 삼각형 메뉴에서 적절한 옵션을 선택하여 표시할 수 있습니다.

각 "강조" 옵션에서 사용할 수 있는 보고서에 대한 자세한 내용은 **"표준 최소 제곱에 대한 강조 옵션"**(77 페이지)에서 확인하십시오.

임의 효과

지정된 모형에 하나 이상의 임의 효과가 포함된 경우 "모형 적합" 시작 창에서 추가 옵션을 사용할 수 있습니다. **Machine.jsp** 샘플 데이터 테이블을 예로 들어보겠습니다. 무작위로 선택된 6명의 작업자가 각각 3대의 기계에서 작업을 수행하고 출력이 평가됩니다. 6명의 특정 작업자에 대한 평가가 다른지 여부를 판단하는 것보다 직원 전체의 평가에서 변동을 추정하는 데 관심이 있습니다. 이를 위해서는 모형을 지정할 때 **person**과 **machine*person**을 임의 효과로 처리해야 합니다.

이 모형에 대한 모형 적합 시작 창이 그림 3.6에 나와 있습니다. **person**에 임의 효과 속성이 적용되면 "방법" 옵션 및 분산 성분과 관련된 두 가지 옵션이 "모형 적합" 시작 창에 나타납니다.

그림 3.6 임의 효과를 포함하는 모형에 대한 모형 적합 시작 창

모형 적합 시작 창의 표준 최소 제곱 옵션

다음 모형 적합 시작 창 옵션은 표준 최소 제곱 분석법과 관련이 있습니다.

강조 초기 보고서 창에 나타나는 보고서 및 그림 유형을 제어합니다. 자세한 내용은 "[표준 최소 제곱에 대한 강조 옵션](#)"(77 페이지)에서 확인하십시오.

방법 (임의 효과가 지정된 경우에만 표시됨) 다음 방법 중 하나를 사용하여 모형을 추정합니다.

REML 자세한 내용은 "[REML 분산 성분 추정값](#)"(178 페이지)에서 확인하십시오.

EMS 기대 평균 제곱을 나타내며 적률 방법이라고도 합니다. 자세한 내용은 "[EMS\(전통적\) 모형 적합 보고서](#)"(181 페이지)에서 확인하십시오.

비유계 분산 성분 ("방법"에 "REML"이 선택된 경우에만 표시됨) 분산 성분 추정값이 음수가 될 수 있도록 허용합니다. 이 옵션은 기본적으로 선택되어 있습니다. 분산 추정값을 0으로

제한하면 고정 효과 검정에서 편향이 발생할 수 있으므로 고정 효과에 관심이 있다면 이 옵션이 선택된 상태여야 합니다. 자세한 내용은 "[음의 분산](#)"(175 페이지)에서 확인하십시오.

분산 성분만 추정 ("방법"에 "REML"이 선택된 경우에만 표시됨) 분산 성분 추정값만 표시하는 보고서를 제공합니다. 자세한 내용은 "[분산 성분만 추정](#)"(176 페이지)에서 확인하십시오.

별도 적합 (Y 변수가 여러 개이고 임의 효과가 없는 모형에만 표시됨) 비결측인 모든 행을 사용하여 각 Y 변수에 대해 별도의 모형을 적합시킵니다. 자세한 내용은 "[결측값](#)"(78 페이지)에서 확인하십시오.

표준 최소 제곱에 대한 강조 옵션

"강조" 목록의 세 옵션은 표준 최소 제곱 분석법에 대한 초기 보고서의 일부로 표시되는 그림 및 보고서 유형을 제어합니다. 아래 설명을 참조하십시오. JMP에서는 데이터 테이블의 행 수, "모형 효과 생성" 목록에 입력된 효과 수 및 효과에 적용된 속성을 기반으로 기본 강조 옵션이 선택됩니다. 사용자가 필요에 따라 이 강조 선택을 변경할 수 있습니다. JMP에서 강조가 선택되는 방법에 대한 자세한 내용은 "[강조 규칙](#)"(187 페이지)에서 확인하십시오.

초기 보고서가 열리면 플랫폼 보고서 창의 빨간색 삼각형 메뉴에서 다른 보고서와 그림을 추가할 수 있습니다.

다음 강조 옵션을 사용할 수 있습니다.

효과 레버리지 모형 적합 상세 정보가 포함된 보고서 및 레버리지 그림과 잔차 그림을 표시합니다. 이 옵션은 모형 적합에 중점을 두는 경우에 유용합니다.

효과 레버리지는 가장 포괄적인 옵션입니다. 이 강조 옵션은 보고서를 전체 모형 관련 보고서와 개별 모형 효과 관련 보고서로 나눕니다. 전체 모형 보고서는 보고서 창의 왼쪽에 "전체 모형"이라는 제목 아래에 표시되고, 효과 보고서는 오른쪽에 표시됩니다.

효과 선별 정렬되거나 척도화된 모수 추정값 보고서와 그래프(해당하는 경우), 예측 프로파일러 및 모형 적합 상세 정보가 포함된 보고서를 표시합니다. 이 옵션은 효과가 많이 있고 선별 설계와 같이 초기에 활성 효과를 찾는 데 중점을 두는 경우에 유용합니다.

"효과 선별"을 선택하면 Box-Cox 변환이 계산됩니다. 추정된 λ 에 대한 신뢰 구간에 1이 포함되지 않으면 "Box-Cox 변환" 보고서가 나타납니다. 자세한 내용은 "[Y의 Box-Cox 변환](#)"(156 페이지)에서 확인하십시오.

최소 보고서 회귀 그림 및 모형 적합 상세 정보가 포함된 보고서만 표시합니다. 이 강조 옵션은 임의 효과 속성이 모형 효과에 적용될 때 기본값입니다.

이 옵션은 최소한의 상세 정보와 가장 간략한 정보를 제공합니다. 빨간색 삼각형 메뉴에서 관심 있는 특정 보고서를 요청할 수 있습니다.

모든 강조 옵션에 대해 표시되는 보고서 또는 그림을 변경하려면 플랫폼 환경 설정을 사용하십시오. **파일 > 환경 설정 > 플랫폼 > 최소 제곱 적합**으로 이동하여 **설정** 체크박스를 사용합니다.

- 옵션이 보고서에 표시되지 않도록 하려면 옵션 옆의 **설정**을 선택하고 해당 옵션은 선택하지 않습니다.
- 옵션이 보고서에 표시되도록 하려면 **설정**을 선택하고 해당 옵션을 선택합니다.

JMP[®] PRO 표준 최소 제곱의 검증

JMP Pro에서는 "모형 적합" 창에 검증 열을 지정할 수 있습니다. 검증 열은 숫자 데이터 유형이어야 하고 두 개 이상의 구분되는 값을 포함해야 합니다.

- 열에 두 개의 값이 포함된 경우 더 작은 값이 훈련 데이터 집합을 정의하고 더 큰 값이 검증 데이터 집합을 정의합니다.
- 열에 세 개의 값이 포함된 경우 작은 값부터 순서대로 훈련 데이터 집합, 검증 데이터 집합, 테스트 데이터 집합을 각각 정의합니다.
- 열에 네 개 이상의 구분되는 값이 포함된 경우에는 검증 열이 무시됩니다.

JMP Pro의 표준 최소 제곱 분석법은 검증 열 사용을 지원합니다. 검증 열을 입력하면 "교차 검증" 보고서가 제공됩니다. 자세한 내용은 "[교차 검증 보고서](#)"(82 페이지)에서 확인하십시오.

검증 열을 입력하면 검증 데이터 집합과 테스트 데이터 집합의 관측값이 보고서의 그림에서 각각 'v'와 't'로 표시됩니다.

JMP 모델링 플랫폼에서 검증 열이 사용되는 방법에 대한 자세한 내용은 **Predictive and Specialized Modeling**의 에서 확인하십시오.

결측값

기본적으로 Y 또는 모형 효과에 대한 결측값이 있는 행은 분석에서 제외됩니다.

참고: JMP에서는 "모형 적합" 창의 "모형 규격" 빨간색 삼각형 메뉴에 "결측값 정보화" 옵션을 제공합니다. "결측값 정보화" 옵션을 사용하면 모형 효과가 결측인 행을 사용하여 모형을 적합시킬 수 있습니다. 자세한 내용은 "모형 규격" 창의 "[결측값 정보화](#)"(46 페이지)에서 확인하십시오.

모형에 임의 효과가 포함된 경우 Y 값은 기본적으로 별도로 적합됩니다. 개별 보고서가 "적합 그룹" 보고서에 나타납니다.

모형에 고정 효과만 포함되어 있고 다음 조건에 해당되는 경우를 가정해 보겠습니다.

- 둘 이상의 Y 반응을 지정했습니다.
- 이 중 일부 Y 반응에 결측값이 있습니다.
- "모형 적합" 시작 창에서 "별도 적합" 옵션을 선택하지 않았습니다.

그러면 다음 옵션 중 하나를 선택하라는 메시지가 표시됩니다.

- 별도 적합 - 해당하는 특정 Y에 대해 비결측인 모든 행을 사용하여 각 Y를 적합시킵니다.
- 함께 적합 - 모든 Y 변수에 대해 비결측인 행만 사용하여 각 Y를 적합시킵니다.

"별도 적합"을 선택하면 "적합 그룹" 보고서에 Y 변수에 대한 개별 보고서가 포함됩니다. "적합 그룹"의 빨간색 삼각형 메뉴에서 프로파일러를 선택하여 동일한 프로파일러에서 모든 Y 변수를

볼 수 있습니다. 아니면 개별 Y 변수 보고서에서 프로파일러를 선택하여 프로파일러에서 해당 변수만 볼 수도 있습니다.

"함께 적합"을 선택하면 "최소 제곱 적합" 보고서에 각 Y 변수에 대한 개별 보고서가 포함됩니다. 그러나 보고서의 일부는 모든 Y 변수에 대해 결합됩니다(예: 효과 요약 및 프로파일러).

최소 제곱 적합 보고서

표준 최소 제곱 분석법을 사용하여 모형을 적합시키면 "최소 제곱 적합" 보고서가 생성됩니다. 이 보고서의 내용은 데이터 특성과 "모형 적합" 시작 창에서 선택한 옵션에 따라 결정됩니다.

팁: 기본적으로 나타나지 않는 보고서를 항상 표시하려면 **파일 > 환경 설정 > 플랫폼 > 최소 제곱 적합**을 사용하여 해당 보고서를 선택하십시오.

단일 반응과 다중 반응의 비교

단일 반응 변수 Y를 적합시키는 경우 "최소 제곱 적합" 창에서 "반응 Y"라는 제목의 보고서에 여러 상세 보고서가 구성됩니다. 여러 반응을 적합시키는 경우에는 일반적으로 "최소 제곱 적합"이라는 제목의 보고서에 개별 반응에 대한 보고서가 구성됩니다. 그러나 결측 반응 데이터가 있을 때 "별도 적합" 옵션을 선택하면 개별 반응에 대한 보고서가 "적합 그룹"이라는 제목의 보고서에 구성됩니다.

강조와 관련된 보고서 구조

"모형 적합" 시작 창에서 "효과 레버리지" 강조를 선택하면 지정된 반응에 대한 보고서가 열 형태로 배열됩니다. 왼쪽 열은 모형과 관련된 추가 보고서를 포함하는 "전체 모형" 보고서로 구성됩니다. 모형의 각 효과에 대한 보고서는 "전체 모형" 보고서의 오른쪽 열에 표시됩니다.

"모형 적합" 시작 창에서 "효과 선별" 또는 "최소 보고서" 강조를 선택하면 각 반응에 대한 모든 보고서가 왼쪽 열에 배열됩니다.

특수 보고서

이 섹션에서는 효과 속성과 관련하여 사용자가 선택한 옵션 또는 데이터 구조에 따라 사용할 수 있는 보고서에 대해 설명합니다.

특이성 상세 정보

모형 효과 사이에 선형 종속성이 있는 경우 "특이성 상세 정보" 보고서가 반응 보고서 제목 아래에 첫 번째 보고서로 나타납니다. 이 보고서에는 모형 항이 충족하는 선형 함수 테이블이 포함됨

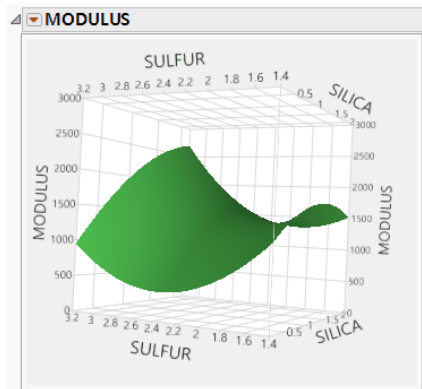
정준 곡률 보고서

2차 모수 추정값 행렬의 고유값과 고유 벡터는 곡률 유형을 결정합니다. 고유 벡터는 가장 큰 곡률과 가장 작은 곡률의 방향을 포함하여 표면의 주곡률 방향을 보여 줍니다.

고유값은 "정준 곡률" 테이블의 첫 번째 행에 제공됩니다.

- 고유값이 음수이면 반응 표면이 최대값에서 아래쪽으로 곡선을 이룹니다.
- 고유값이 양수이면 표면 모양이 최소값에서 위쪽으로 곡선을 이룹니다.
- 양수 고유값과 음수 고유값이 모두 있으면 표면이 안장 모양이 되어 한쪽은 위로, 다른 쪽은 아래로 각각 곡선을 이룹니다. Tiretread.jmp 샘플 데이터 테이블을 사용한 예는 그림 3.9 에서 확인하십시오.

그림 3.9 안장 모양 표면을 사용한 표면 프로파일러 그림



고유값 아래에 나열된 고유 벡터는 주 축의 방향을 보여 줍니다. 고유값의 절대값이 클수록 관련된 방향으로 반응 표면의 곡률이 커집니다. 경우에 따라 0 고유값이 발생합니다. 이 고유값은 적합된 표면이 해당 고유 벡터가 설명하는 방향을 따라 편평하다는 의미입니다.

참고 : 20 개가 넘는 요인으로 구성된 반응 표면 모형의 경우 반응 표면 보고서가 표시되지 않습니다. 오류 메시지 또는 경고가 제공되지 않습니다. 반응 표면 설계에 대한 자세한 내용은 실험 설계 가이드의 에서 확인하십시오.

혼합 및 임의 효과 모형 보고서

"모형 적합" 시작 창에서 임의 효과를 지정하면 "방법" 목록이 나타납니다. 이 목록에는 "REML(권장)"과 "EMS(전통적)"라는 두 가지 적합 방법이 제공됩니다. 선택한 모형과 방법에 따라 "열 저장" 및 "프로파일러" 옵션과 추가 보고서가 표시됩니다.

REML 방법 보고서에 대한 자세한 내용은 "REML(제한 최대 가능도) 방법"(176 페이지)에서 확인하십시오. EMS 방법 보고서에 대한 자세한 내용은 "EMS(전통적) 모형 적합 보고서"(181 페이지)에서 확인하십시오.

교차 검증 보고서

JMP[®] PRO "모형 적합" 시작 창에 검증 열을 입력하면 "교차 검증" 보고서가 표시됩니다. 이 보고서에는 검증에 사용된 각 데이터 집합에 대한 다음 정보가 제공됩니다.

소스 데이터 집합을 훈련 데이터 집합, 검증 데이터 집합 또는 테스트 데이터 집합으로 식별합니다.

R² 훈련 데이터 집합을 사용하여 파생된 모형과 관련하여 주어진 데이터 집합의 관측값에 대해 계산된 R² 값입니다. 훈련 데이터 집합의 경우 이 값은 일반적인 R² 값입니다.

각 훈련, 검증 및 테스트 데이터 집합의 R² 값은 다음과 같이 계산됩니다.

- 주어진 데이터 집합의 각 관측값에 대해 예측 오차를 계산합니다. 이 값은 실제 반응과 훈련 데이터 집합 모형에서 예측한 반응 간의 차이입니다.
- 예측 오차를 제공하고 합하여 SSE_{Source} 를 구합니다. 여기서 서브스크립트 *Source* 는 훈련, 검증 또는 테스트 데이터 집합을 나타냅니다.
- *Source* 데이터 집합의 관측값에 대한 실제 반응과 평균 간의 차이를 제공하고 합합니다. 이 값을 SST_{Source} 로 나타냅니다.
- *Source* 데이터 집합의 R² 은 다음과 같습니다.

$$R^2_{Source} = 1 - \frac{SSE_{Source}}{SST_{Source}}$$

참고 : 검증 데이터 집합과 테스트 데이터 집합의 R² 값은 음수일 수 있습니다.

RASE 평균 제곱 예측 오차의 제곱근입니다. 각 훈련, 검증 및 테스트 데이터 집합의 RASE 는 다음과 같이 계산됩니다.

- 주어진 데이터 집합의 각 관측값에 대해 예측 오차를 계산합니다. 이 값은 실제 반응과 훈련 데이터 집합 모형에서 예측한 반응 간의 차이입니다.
- 예측 오차를 제공하고 합하여 SSE_{Source} 를 구합니다. 여기서 서브스크립트 *Source* 는 훈련, 검증 또는 테스트 데이터 집합을 나타냅니다.
- 관측값 수를 n 으로 나타냅니다.
- RASE 는 다음과 같습니다.

$$RASE_{Source} = \sqrt{\frac{SSE_{Source}}{n}}$$

빈도 소스 데이터 집합의 관측값 수입니다.

최소 제곱 적합 옵션

Y가 두 개 이상 있고 결측 반응 값이 없거나, 결측값이 있는 Y가 두 개 이상 있을 수 있습니다. "함께 적합"을 선택하면 반응이 "최소 제곱 적합" 보고서에 그룹화됩니다. 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

프로파일러 모든 반응을 하나의 프로파일러에 표시합니다. 모든 반응에 대한 모형 항의 효과를 동시에 보고 여러 최적화를 수행할 수 있습니다. 자세한 내용은 "**요인 프로파일링**"(150 페이지)에서 확인하십시오.

모형 대화상자 현재 분석을 위해 완료된 시작 창을 표시합니다.

효과 요약 모형에서 효과를 추가하거나 제거할 수 있는 대화식 "효과 요약" 보고서를 표시합니다. 자세한 내용은 "**효과 요약 보고서**"(168 페이지)에서 확인하십시오.

다음 옵션에 대한 자세한 내용은 **JMP 사용**의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

적합 그룹 옵션

Y가 두 개 이상 있고 "별도 적합"을 선택하면 반응이 "적합 그룹" 보고서에 그룹화됩니다. 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

프로파일러 모든 반응을 하나의 프로파일러에 표시합니다. 모든 반응에 대한 모형 항의 효과를 동시에 보고 여러 최적화를 수행할 수 있습니다. 자세한 내용은 "**프로파일러**"(151 페이지)에서 확인하십시오.

등고선 프로파일러 모든 반응을 하나의 등고선 프로파일러에 표시합니다. 모든 반응에 대한 모형 항의 효과를 동시에 탐색할 수 있습니다.

표면 프로파일러 각 반응에 대한 별도의 표면 프로파일러 보고서를 표시합니다.

행으로 배열 플랫폼 분석 보고서를 지정된 수의 행으로 재배열합니다. 이 옵션은 창 또는 출력 페이지에 더 많은 보고서가 들어가도록 보고서를 배열하는 데 주로 사용됩니다.

적합도를 기준으로 정렬 적합 유의성 (R^2)을 기준으로 보고서를 정렬합니다. 이 옵션은 R^2 이 플랫폼 수준으로 표시되는 플랫폼에만 사용할 수 있습니다. 예를 들어 하나의 시작 창에서 수

백 개의 일원 분석을 생성하면 적합 그룹에 분석이 나타나고, 가장 강한 관계가 먼저 나타나도록 정렬할 수 있습니다.

다음 옵션에 대한 자세한 내용은 **JMP 사용**의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

반응 옵션

반응에 대한 빨간색 삼각형 메뉴를 사용하면 필요에 따라 보고서를 사용자 정의할 수 있습니다.

회귀 보고서 기본 보고서와 보고서 옵션을 제공합니다. 자세한 내용은 "**회귀 보고서**"(85 페이지)에서 확인하십시오.

추정값 모수 추정값과 관련된 추가 분석을 위한 옵션을 제공합니다. 자세한 내용은 "**추정값**"(105 페이지)에서 확인하십시오.

효과 선별 유의한 효과를 식별하기 위한 보고서 및 그림을 제공합니다. 자세한 내용은 "**효과 선별**"(140 페이지)에서 확인하십시오.

요인 프로파일링 반응이 모형 항과 어떻게 관련되어 있는지 검토하기 위한 프로파일러, 교호작용 및 입방체도를 제공합니다. Box-Cox 변환 적합을 위한 그림과 보고서도 제공합니다. 자세한 내용은 "**요인 프로파일링**"(150 페이지)에서 확인하십시오.

행 진단 잔차를 검토하기 위한 그림 및 보고서를 제공합니다. 또한 PRESS 통계량을 보고하고 Durbin-Watson 검정을 제공합니다. 자세한 내용은 "**행 진단**"(159 페이지)에서 확인하십시오.

열 저장 결과를 별도의 데이터 테이블에 저장하는 "코딩 테이블 저장"을 제외하고 모형 결과를 데이터 테이블의 열로 저장합니다. 자세한 내용은 "**열 저장**"(165 페이지)에서 확인하십시오.

모형 대화상자 현재 분석을 위해 완료된 "모형 적합" 시작 창을 표시합니다.

효과 요약 모형에서 효과를 추가하거나 제거할 수 있는 대화식 "효과 요약" 보고서를 표시합니다. 자세한 내용은 "**효과 요약 보고서**"(168 페이지)에서 확인하십시오.

다음 옵션에 대한 자세한 내용은 **JMP 사용**의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

회귀 보고서

"회귀 보고서" 메뉴는 모형 적합, 효과 유의성 및 모형 모수에 대한 요약 정보를 제공합니다.

적합 요약 모형 적합 요약을 표시하거나 숨깁니다. 자세한 내용은 "**적합 요약**"(85 페이지)에서 확인하십시오.

분산 분석 적합 모형을 단순 평균 모형과 비교하기 위한 계산을 표시하거나 숨깁니다. 자세한 내용은 "**분산 분석**"(86 페이지)에서 확인하십시오.

모수 추정값 모수 추정값 및 각 모수가 0 이라는 가설에 대한 t -검정이 포함된 보고서를 표시하거나 숨깁니다. 자세한 내용은 "**모수 추정값**"(87 페이지)에서 확인하십시오.

효과 검정 모형의 고정 효과에 대한 검정을 표시하거나 숨깁니다. 자세한 내용은 "**효과 검정**"(88 페이지)에서 확인하십시오.

효과 상세 정보 개별 효과에 대한 상세 정보, 그림 및 검정이 포함된 보고서를 표시하거나 숨깁니다. 자세한 내용은 "**효과 상세 정보**"(89 페이지)에서 확인하십시오.

"효과 레버리지" 강조 옵션을 선택하면 "최소 제곱 적합" 보고서 창 맨 위에 각 효과의 개별 보고서가 제공됩니다. 이 보고서에는 레버리지 그림과 효과 상세 정보 옵션이 포함됩니다. 자세한 내용은 "**레버리지 그림**"(161 페이지)에서 확인하십시오.

적합 결여 검정을 수행할 수 있을 때 모형에 적절한 효과가 있는지 평가하는 검정을 표시하거나 숨깁니다. 자세한 내용은 "**적합 결여**"(104 페이지)에서 확인하십시오.

모든 신뢰 구간 표시 다음 통계량에 대한 신뢰 구간을 표시하거나 숨깁니다.

- "모수 추정값" 보고서의 모수 추정값
- "최소 제곱 평균" 테이블의 최소 제곱 평균

AICc "적합 요약" 보고서에 AICc 및 BIC 값을 표시하거나 숨깁니다. 자세한 내용은 "통계 상세 정보" 부록의 "**가능도, AICc 및 BIC**"(514 페이지)에서 확인하십시오.

적합 요약

"적합 요약" 보고서에는 R^2 계산과 AICc 및 BIC 값과 같은 상세 정보가 제공됩니다.

R^2 랜덤 오차가 아니라 모형에 기인할 수 있는 반응의 변동 비율을 추정합니다. R^2 (다중 결정 계수라고도 함)은 해당 "분산 분석" 테이블의 값을 사용하여 다음과 같이 계산됩니다.

$$\frac{\text{Sum of Squares(Model)}}{\text{Sum of Squares(C. Total)}}$$

R^2 이 1 에 가까울수록 R^2 이 0 에 가까울 때보다 데이터에 더 잘 적합된다는 것을 나타냅니다. R^2 이 0 에 가까우면 모형이 반응 평균보다 더 나은 반응 예측 변수가 아님을 나타냅니다.

참고 : R^2 값이 낮으면 설명되지 않은 변동을 설명하는 변수가 모형에 없는 것일 수 있습니다. 하지만 데이터의 내재 변동 범위가 큰 경우에는 유용한 회귀 모형이라도 R^2 값이 낮을 수 있습니다. 일반적인 R^2 값에 대해 알아보려면 연구 영역의 문헌을 읽어 보십시오.

Adj- R^2 모형의 모수 수에 대한 R^2 통계량을 조정합니다. Adj- R^2 을 사용하면 모수 수가 다른 모형을 쉽게 비교할 수 있습니다. 계산에 자유도가 사용됩니다. Adj- R^2 은 해당 "분산 분석" 테이블의 값을 사용하여 다음과 같이 계산됩니다.

$$1 - \frac{\text{Mean Square(Error)}}{\text{제곱합(수정 합계)/DF(수정 합계)}}$$

제곱근 평균 제곱 오차 랜덤 오차의 표준편차를 추정합니다. 이 값은 "분산 분석" 보고서에 표시된 오차 평균 제곱의 제곱근입니다.

참고 : 제곱근 평균 제곱 오차는 일반적으로 *RMSE* 라고 합니다.

반응 평균 반응 값의 전체 평균을 표시합니다.

관측값 수 (또는 가중합) 모형에 사용된 관측값 수를 표시합니다.

- 이 값은 결측값, 제외된 행, 가중치 또는 빈도 역할에 할당된 열이 없는 경우 데이터 테이블의 행 수와 동일합니다.
- 이 값은 가중치 역할에 할당된 열이 있는 경우 가중치 열에 포함된 양수 값의 합입니다.
- 이 값은 빈도 역할에 할당된 열이 있는 경우 빈도 열에 포함된 양수 값의 합입니다.

AICc (" 회귀 보고서 " 메뉴에서 "AICc" 옵션을 선택했거나, " 최소 제곱 적합 " 환경 설정에서 "AICc"를 설정한 경우에만 표시됨) 수정된 AICc(Akaike 정보 기준) 값을 표시합니다. 자세한 내용은 "통계 상세 정보" 부록의 "**가능도, AICc 및 BIC**"(514 페이지)에서 확인하십시오.

BIC (" 회귀 보고서 " 메뉴에서 "AICc" 옵션을 선택했거나, " 최소 제곱 적합 " 환경 설정에서 "AICc"를 설정한 경우에만 표시됨) BIC(Bayes 정보 기준) 값을 표시합니다. 자세한 내용은 "통계 상세 정보" 부록의 "**가능도, AICc 및 BIC**"(514 페이지)에서 확인하십시오.

분산 분석

"분산 분석" 보고서에는 적합 모형을 모든 예측값이 반응 평균과 동일한 모형과 비교하기 위한 계산이 제공됩니다.

참고 : " 모형 적합 " 시작 창에 빈도 또는 가중치 변수를 입력하면 "**빈도**"(36 페이지) 및 "**가중치**"(36 페이지)에 나오는 설명과 일치하도록 "분산 분석" 보고서의 항목이 조정됩니다.

"분산 분석" 보고서에는 다음과 같은 열이 포함됩니다.

소스 세 가지 변동 소스이며 모형, 오차 및 수정 합계 중 하나입니다.

DF 각 변동 소스에 대한 관련 DF(자유도)입니다. 수정 합계 DF는 항상 관측값 수에서 1을 뺀 값이며 다음과 같이 모형 자유도와 오차 자유도로 분할됩니다.

- 모형 DF는 모형 적합에 사용되는 모수(절편 제외)의 수입니다.
- 오차 DF는 수정 합계 DF와 모형 DF의 차이입니다.

제곱합 각 변동 소스에 대한 관련 SS(제곱합)입니다.

- 총(수정 합계) SS는 반응 값과 표본 평균 간의 차이에 대한 제곱합입니다. 이 값은 반응 값의 총 변동을 나타냅니다.
- 오차 SS는 적합된 값과 실제값 간의 차이에 대한 제곱합입니다. 이 값은 적합 모형에 의해 설명되지 않은 변동을 나타냅니다.
- 모형 SS는 수정 합계 SS와 오차 SS 간의 차이입니다. 이 값은 모형에 의해 설명된 변동을 나타냅니다.

평균 제곱 모형 및 오차 변동 소스에 대한 평균 제곱 통계량입니다. 각 평균 제곱은 제곱합을 해당 자유도로 나눈 것입니다.

참고: 오차에 대한 평균 제곱의 제곱근은 "적합 요약" 보고서의 RMSE와 동일합니다.

사용된 MSE ("오차 규격"이 "순수 오차" 또는 "지정됨"인 경우에만 표시됨) 기본 오차 규격이 선택되지 않은 경우 사용되는 평균 제곱 오차입니다. 이 값은 평균 제곱 열의 평균 제곱 오차 대신 F비를 계산하는 데 사용됩니다.

사용된 DFE ("오차 규격"이 "순수 오차" 또는 "지정됨"인 경우에만 표시됨) 기본 오차 규격이 선택되지 않은 경우 사용되는 오차 자유도입니다. 이 값은 DF 열의 오차 DF 대신 Prob > F 값을 계산하는 데 사용됩니다.

F 비 모형 평균 제곱을 오차 평균 제곱으로 나눈 것입니다. F 비는 모형이 모든 예측값과 반응 평균이 동일한 모형과 유의하게 다른지 여부를 검정하기 위한 검정 통계량입니다.

Prob > F 검정의 p 값입니다. Prob > F 값은 절편을 제외한 모든 모수가 0일 때 관측된 것과 동일한 F비를 얻을 확률을 측정합니다. Prob > F 값이 작으면 관측된 F비의 발생 확률이 낮음을 나타냅니다. 이러한 값은 모형에 하나 이상의 유의 효과가 있다는 증거로 간주됩니다.

모수 추정값

"모수 추정값" 보고서에는 모형 모수의 추정값이 표시되고, 각 모수에 대해 모수가 0이라는 가설에 대한 t -검정이 제공됩니다.

참고: 모형 항 사이에 선형 종속성이 있는 경우에도 가능하면 추정값을 구하고 검정합니다. 이러한 추정값에는 "편향" 또는 "영점화됨"이라는 라벨이 지정됩니다. 자세한 내용은 "[모형 항 사이에 선형 종속성이 있는 모형](#)"(182 페이지)에서 확인하십시오.

항 추정된 모수에 해당하는 모형 항입니다. "모형 적합" 시작 창에서 "절편 없음" 옵션을 선택한 경우를 제외하고 첫 번째 항은 항상 절편입니다. 연속형 효과는 데이터 테이블 열 이름과 함께 나타냅니다. 고차 항의 일부인 연속형 열은 중심화될 수 있습니다. 명목형 또는 순서형

효과는 대괄호 안에 수준 값이 포함된 상태로 나타납니다. 명목형 및 순서형 항의 코딩에 대한 자세한 내용은 " [명목형 효과에 대한 코딩](#) "(139 페이지) 및 " 통계 상세 정보 " 부록의 " [요인 모형](#) "(492 페이지) 에서 확인하십시오.

추정값 각 항에 대한 모수 추정값입니다. 이 값은 모형 계수의 추정값입니다. 모형 항 사이에 선형 종속성이 있으면 " 편향 " 또는 " 영점화됨 " 이라는 라벨이 지정될 수 있습니다. 자세한 내용은 " [모형 항 사이에 선형 종속성이 있는 모형](#) "(182 페이지) 에서 확인하십시오.

표준 오차 추정된 각 모수에 대한 표준 오차 추정값입니다.

t 비 각 모수의 실제 값이 0 인지 여부를 검정합니다. *t* 비는 표준 오차에 대한 추정값의 비율입니다. 모형에 대한 일반적인 가정이 주어지면 *t* 비는 귀무가설 하에서 스튜던트 *t* 분포를 따릅니다.

Prob>|t| 실제 모수 값이 0 이라는 검정에 반대되는 양측 대립가설에 대한 *p* 값입니다.

95% 하한 모수 추정값에 대한 95% 신뢰 하한입니다. 이 열은 회귀 보고서 > 모든 신뢰 구간 표시 옵션을 선택했거나, 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 95% 하한을 선택한 경우에만 나타납니다.

95% 상한 모수 추정값에 대한 95% 신뢰 상한입니다. 이 열은 회귀 보고서 > 모든 신뢰 구간 표시 옵션을 선택했거나, 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 95% 상한을 선택한 경우에만 나타납니다.

표준화 베타 모든 항이 평균 0, 분산 1로 표준화된 회귀 모형의 모수 추정값입니다. 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 표준화 베타를 선택한 경우에만 나타납니다.

VIF 모형의 각 항에 대한 분산 팽창 계수입니다. VIF 값이 높으면 모형의 항 사이에 공선성 문제가 있음을 나타냅니다.

i 번째 항 x_i 의 VIF 는 다음과 같이 정의됩니다.

$$VIF_i = \frac{1}{1 - R_i^2}$$

여기서 R_i^2 은 다른 설명 변수의 함수로 나타낸 x_i 회귀에 대한 R^2 (다중 결정 계수) 입니다. 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > VIF 를 선택한 경우에만 나타납니다.

설계 표준 오차 모수 추정값의 상대 분산 제곱근입니다 (Goos and Jones 2011, p. 25).

$$\sqrt{\text{diag}(X'X)^{-1}}$$

이러한 값은 RMSE 로 나눈 표준 오차입니다. 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 설계 표준 오차를 선택한 경우에만 나타납니다.

효과 검정

"효과 검정" 보고서는 모형에 고정 효과가 있는 경우에만 나타납니다. 주어진 효과에 대한 효과 검정은 해당 효과와 관련된 모든 모수가 0이라는 귀무가설을 검정합니다. 단일 연속형 설명 변수의 경우 효과에 하나의 모수만 있을 수 있습니다. 이 경우 검정은 "모수 추정값" 보고서의 해당 항에 대한 *t*-검정과 동등합니다. 명목형 또는 순서형 효과에는 수준 수에 따라 관련된 모수가 여

러 개 있을 수 있습니다. 이러한 효과에 대한 효과 검정은 관련된 모든 모수가 0인지 여부를 검정합니다.

다음 사항에 유의하십시오.

- 가능한 경우 항이 선형 종속성과 관련된 효과에 대해 효과 검정이 수행됩니다. 자세한 내용은 "[모형 항 사이에 선형 종속성이 있는 모형](#)"(182 페이지)에서 확인하십시오.
- 파라미터화 및 특이성 처리는 SAS GLM 프로시저와 다릅니다. 파라미터화 및 특이성 처리에 대한 자세한 내용은 "통계 상세 정보" 부록의 "[요인 모형](#)"(492 페이지)에서 확인하십시오.

"효과 검정" 보고서에는 다음 열이 포함됩니다.

소스 모형의 효과입니다.

모수 수 효과와 관련된 모수의 수입니다. 연속형 효과에는 하나의 모수가 있습니다. 명목형 또는 순서형 효과에 대한 모수의 수는 수준 수에서 1을 뺀 값입니다. 교차 효과에 대한 모수의 수는 각 개별 효과에 대한 모수의 수를 곱한 값입니다.

DF 효과 검정에 대한 자유도입니다. 일반적으로 모수 수와 DF는 동일합니다. 예측 변수 사이에 선형 종속성이 있는 경우에는 값이 다를 수 있습니다. 이러한 경우 DF는 모수 수보다 작을 수 있으며 이는 효과와 관련된 하나 이상의 모수를 검정할 수 없음을 나타냅니다. DF가 모수 수보다 작은 경우에는 항상 LostDFs가 보고서 줄 오른쪽에 나타납니다. 오차의 자유도가 있는 경우 검정이 수행됩니다. 자세한 내용은 "[효과 검정 보고서](#)"(185 페이지)에서 확인하십시오.

제곱합 효과가 0이라는 가설에 대한 제곱합입니다.

F 비 효과가 0인지 검정하기 위한 F 통계량입니다. "F 비"는 효과의 평균 제곱을 오차 평균 제곱으로 나눈 비율입니다. 효과의 평균 제곱은 효과 제곱합을 자유도로 나눈 값입니다.

Prob > F 효과 검정의 p 값입니다.

평균 제곱 효과의 평균 제곱, 즉 효과에 대한 제곱합을 DF로 나눈 값입니다.

참고: 보고서를 마우스 오른쪽 버튼으로 클릭하고 **열 > 평균 제곱**을 선택한 경우에만 나타납니다.

효과 상세 정보

"효과 상세 정보" 보고서에는 개별 효과에 대한 상세 정보, 그림 및 검정이 제공됩니다. 이 보고서는 "모형 적합" 시작 창에서 선택한 강조에 따라 별도의 보고서로 구성됩니다.

- 효과 레버리지 강조: "전체 모형" 보고서 오른쪽에 있는 "최소 제곱 적합" 보고서의 위에 각 효과의 개별 보고서가 있습니다. 이 경우 보고서에는 효과에 대한 레버리지 그림이 포함됩니다.
- 효과 선택 또는 최소 보고서 강조: "효과 상세 정보" 보고서가 제공되지만 처음에 닫혀 있습니다. 표시 아이콘을 클릭하면 보고서가 표시됩니다.

보고서의 초기 콘텐츠는 최소 제곱 평균 테이블입니다. 효과의 특성에 따라 이 테이블이 적절하지 않을 수 있으며 초기 기본 보고서에 콘텐츠가 아무 것도 표시되지 않을 수 있습니다. 그러나 특정 빨간색 삼각형 옵션을 사용할 수 있습니다.

효과 옵션 테이블

효과 이름 옆의 빨간색 삼각형 메뉴는 다음 옵션을 제공합니다. 특정 모델링 유형의 경우 일부 옵션은 적절하지 않을 수 있으므로 사용할 수 없습니다.

최소 제곱 평균 테이블 효과를 검정할 때 비교되는 통계량을 표시합니다. 자세한 내용은 "[최소 제곱 평균 테이블](#)"(91 페이지)에서 확인하십시오.

연속형 효과에는 이 옵션을 사용할 수 없습니다.

최소 제곱 평균 그림 명목형 및 순서형 효과에 대한 최소 제곱 평균 그림을 표시합니다. 교호작용 효과의 경우 이 옵션은 "[최소 제곱 평균 그림 옵션](#)" 창을 표시합니다. 자세한 내용은 "[최소 제곱 평균 그림](#)"(92 페이지)에서 확인하십시오.

최소 제곱 평균 대비 ... 명목형 효과, 순서형 효과 및 교호작용의 수준을 비교하기 위해 대비를 지정하고 검정할 수 있는 "[대비 지정](#)" 창을 표시합니다. 자세한 내용은 "[최소 제곱 평균 대비 ...](#)"(95 페이지)에서 확인하십시오.

최소 제곱 평균 스튜던트 t 스튜던트 t 검정을 사용하여 최소 제곱 평균의 쌍별 비교를 위한 검정 및 신뢰 구간을 표시합니다. 자세한 내용은 "[최소 제곱 평균 스튜던트 t 및 최소 제곱 평균 Tukey HSD](#)"(98 페이지)에서 확인하십시오.

참고: 유의 수준은 개별 비교에 적용되고 모든 비교에 전체적으로 적용되지 않습니다. 비교 모음의 오차율은 개별 검정의 오차율보다 큽니다.

최소 제곱 평균 Tukey HSD Tukey-Kramer HSD(Honestly Significant Difference) 검정 (Tukey 1953, Kramer 1956)을 사용하여 최소 제곱 평균의 쌍별 비교를 위한 검정 및 신뢰 구간을 표시합니다. 자세한 내용은 "[최소 제곱 평균 스튜던트 t 및 최소 제곱 평균 Tukey HSD](#)"(98 페이지)에서 확인하십시오.

참고: 유의 수준은 쌍별 비교 모음에 적용됩니다. 유의 수준은 표본 크기가 같으면 정확하고 표본 크기가 다르면 약간 작습니다 (Hayter 1984).

최소 제곱 평균 Dunnett 지정하는 대조 수준에 대해 쌍별 비교를 위한 검정 및 신뢰 구간을 표시합니다. 또한 검정 결과 그림도 제공합니다. 자세한 내용은 "[최소 제곱 평균 Dunnett](#)"(101 페이지)에서 확인하십시오.

조각 검정 교호작용에 있는 각 열의 각 수준에 대해 교호작용에 있는 다른 분류 열의 모든 수준 간 쌍별 비교를 공동으로 검정합니다. 자세한 내용은 "[조각 검정](#)"(102 페이지)에서 확인하십시오.

참고: 명목형 및 순서형 효과를 포함하는 교호작용에만 사용할 수 있습니다.

검정력 분석 효과 검정에 대한 검정력을 분석할 수 있는 "검정력 상세 정보" 보고서를 표시합니다. 자세한 내용은 "**검정력 분석**"(103 페이지)에서 확인하십시오.

최소 제곱 평균 테이블

최소 제곱 평균은 다른 모형 요인이 **중립** 값으로 설정된 범주형 효과의 수준에 대해 모형에서 예측한 값입니다. 연속형 효과의 중립 값은 표본 평균으로 정의됩니다. 관심 효과와 연관이 없는 명목형 효과의 중립 값은 해당 효과에 대한 계수의 평균입니다. 연관이 없는 순서형 효과의 중립 값은 값 순서화에서 효과의 첫 번째 수준으로 정의됩니다.

최소 제곱 평균은 **조정 평균** 또는 **모집단 주변 평균**이라고도 합니다. 모형에 다른 효과가 있는 경우 최소 제곱 평균이 단순 평균과 다를 수 있습니다. 사실 최소 제곱 평균은 표본 평균보다 서로 더 가까운 것이 일반적입니다. 이러한 예측이 수행되는 중립 값의 특성 때문에 이런 상황이 발생합니다.

최소 제곱 평균은 다른 모형 요인의 특정 값에 대한 예측이므로 서로 비교할 수 있습니다. 효과를 검정할 때 최소 제곱 평균을 사용하여 비교합니다. 최소 제곱 평균에 대한 자세한 내용은 "통계 상세 정보" 부록의 "**명목형 요인에 대한 최소 제곱 평균**"(496 페이지) 및 "**순서형 최소 제곱 평균**"(506 페이지)에서 확인하십시오.

주요 효과의 경우 최소 제곱 평균 테이블에 표본 평균도 포함됩니다(그림 3.10).

최소 제곱 평균 테이블의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. **weight** 를 선택하고 **Y** 를 클릭합니다.
4. **age, sex** 및 **height** 를 선택하고 **추가**를 클릭합니다.
5. **강조** 목록에서 **효과 선별**을 선택합니다.
6. **실행**을 클릭합니다.
7. "효과 상세 정보" 보고서는 "최소 제곱 적합" 보고서의 아래쪽 근처에 나타나고 처음에는 닫혀 있습니다. "효과 상세 정보" 보고서 옆의 표시 아이콘을 클릭하여 보고서를 표시합니다.

그림 3.10에 표시된 것처럼 "효과 상세 정보" 보고서에는 세 가지 효과에 대한 보고서가 각각 표시됩니다. "최소 제곱 평균 테이블"은 **age** 및 **sex**에 제공되지만 연속형 효과인 **height**에는 제공되지 않습니다. 최소 제곱 평균이 표본 평균과 어떻게 다른지 확인하십시오.

제거 지정된 효과에 대한 최소 제곱 평균 그림 보고서를 제거합니다.

최소 제곱 평균 그림 옵션

교호작용 효과의 빨간색 삼각형 메뉴에서 " 최소 제곱 평균 그림 " 옵션을 선택하면 " 최소 제곱 평균 그림 옵션 " 창이 나타납니다.

그림 3.11 최소 제곱 평균 그림 옵션 창

☐ 교호작용 그림 생성

중첩을 위한 항 선택

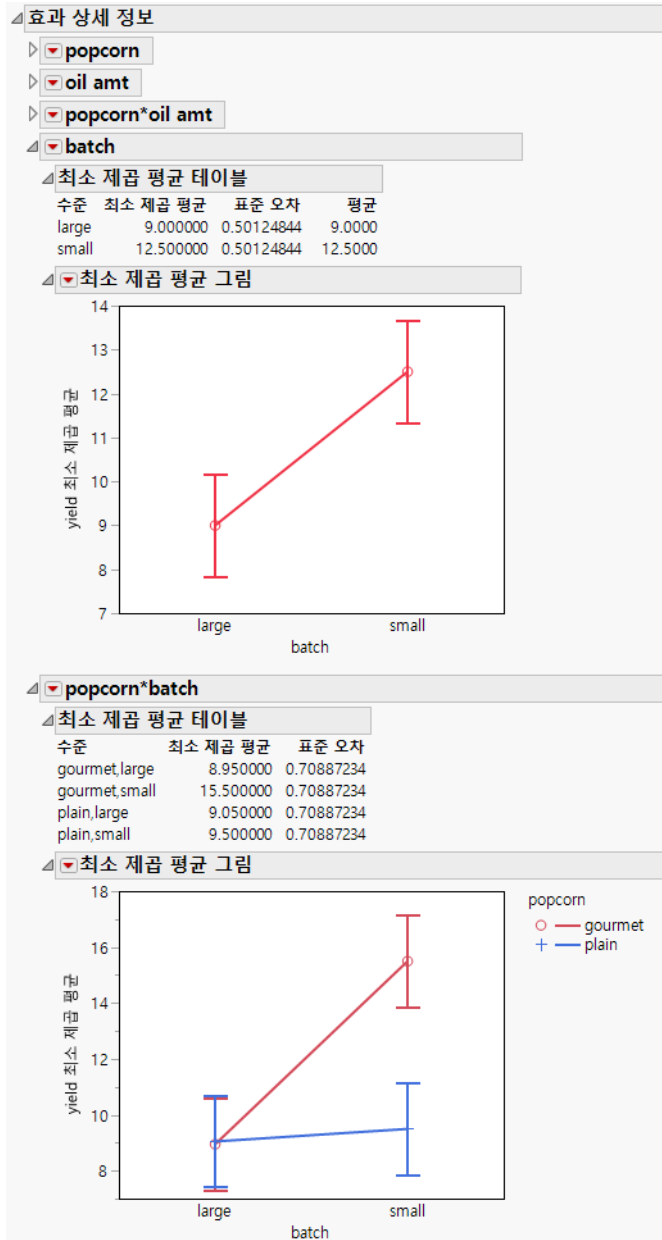
popcorn
batch

확인 취소

창에서 아무 것도 선택하지 않고 "확인"을 클릭하면 최소 제곱 평균 그림 하나가 나타납니다. 그림의 가로 축은 각 조합에 대해 별도의 효과를 얻기 위해 내포된 요인의 수준으로 구성됩니다. 교호작용 그림을 생성하려면 "교호작용 그림 생성" 옆의 상자를 선택합니다. "중첩을 위한 항 선택" 옵션을 사용하면 교호작용 그림에서 중첩 변수로 표시되는 효과를 선택할 수 있습니다.

삼원 교호작용 항의 경우 교호작용 그림에 대한 중첩 변수를 선택하면 두 번째 옵션 패널이 나타납니다. 두 번째 패널에서 아무 것도 선택하지 않고 "확인"을 클릭하면 교호작용 그림 하나가 나타납니다. 아니면 두 번째 옵션 패널을 사용하여 "별도 그림을 위한 항 선택"에서 선택한 효과의 각 수준에 대한 별도의 교호작용 그림을 생성할 수도 있습니다.

그림 3.12 두 효과에 대한 최소 제곱 평균 테이블 및 최소 제곱 평균 그림



최소 제곱 평균 그림의 예

그림 3.12 에 나오는 보고서를 생성하려면 다음 단계를 수행하십시오 .

1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Popcorn.jmp 를 엽니다 .

2. **분석 > 모형 적합**을 선택합니다.
3. **Yield**를 선택하고 **Y**를 클릭합니다.
4. **popcorn, oil amt** 및 **batch**를 선택하고 **매크로 > 완전 요인**을 클릭합니다. 그러면 "강조"가 "효과 선별"로 변경됩니다.
5. **실행**을 클릭합니다.
6. "효과 상세 정보" 표시 아이콘을 클릭하여 7개 모형 효과의 상세 정보를 표시합니다.
7. "batch"의 빨간색 삼각형을 클릭하고 **최소 제곱 평균 그림**을 선택합니다.
8. "popcorn*batch"의 빨간색 삼각형을 클릭하고 **최소 제곱 평균 그림**을 선택합니다. "최소 제곱 평균 그림 옵션" 창이 나타납니다.
9. "최소 제곱 평균 그림 옵션" 창에서 "교호작용 그림 생성" 옆의 상자를 클릭합니다.
10. "중첩을 위한 항 선택"에서 **popcorn**을 선택합니다.
11. **확인**을 클릭합니다.
12. popcorn*batch에 대한 그림에서 요인을 전치하려면 8단계 및 9단계를 반복합니다.
13. "중첩을 위한 항 선택"에서 **batch**를 선택하고 **확인**을 클릭합니다.

그림 3.13 요인이 전치된 교호작용의 최소 제곱 평균 그림

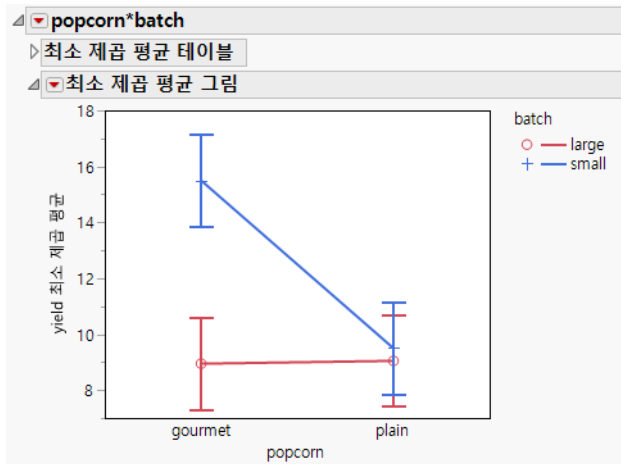


그림 3.13에서는 요인이 전치된 popcorn*batch 교호작용 그림을 보여 줍니다. 이 그림을 그림 3.12의 그림과 비교해 봅니다. 이러한 그림은 동일한 정보를 나타내지만 관심 대상에 따라 하나가 다른 것보다 직관적일 수 있습니다.

최소 제곱 평균 대비 ...

대비는 모수 값의 선형 결합입니다. "대비 지정" 창에서 여러 대비를 지정하고 값이 0인지 여부를 공동으로 검정할 수 있습니다(그림 3.14).

JMP에서는 효과의 최소 제곱 평균에 의해 대비를 구성합니다. 대비의 각 열은 합이 0이 되어 절대값의 합이 2가 되도록 정규화됩니다. 대비에 공변량이 포함된 경우 대비를 검정할 공변량 값을 지정할 수 있습니다.

"대비 지정" 상자에는 효과의 이름과 효과의 수준 이름이 표시됩니다. 대비 값은 처음에 0으로 설정되고 + 및 - 기호를 포함하는 셀 옆에 나타납니다. 수준을 비교하려면 이러한 버튼을 클릭합니다.

+ 또는 - 버튼을 클릭할 때마다 대비 계수는 합이 0이 되고 가능한 경우 절대값의 합이 2가 되도록 정규화됩니다. 추가 수준을 비교하려면 **새 열** 버튼을 클릭합니다. 그러면 새 대비를 정의할 새 열이 나타납니다. 끝나면 **완료**를 클릭합니다. "대비" 보고서가 나타납니다(그림 3.15). 전체 검정은 모든 대비에 대한 결합 F -검정입니다.

참고 : 가능한 최대 대비 수보다 많이 지정하려고 하면 검정이 자동으로 평가됩니다.

"대비" 보고서에는 결합 F -검정에 대한 다음 상세 정보가 제공됩니다.

SS 결합 검정의 제곱합입니다.

NumDF 분자 자유도입니다.

DenDF 분모 자유도입니다.

F 비 SS를 NumDF로 나눈 값을 평균 제곱 오차로 나눈 값의 비율입니다.

Prob > F 유의성 검정의 p 값입니다.

검정 상세 정보 보고서

"검정 상세 정보" 보고서(그림 3.15)에는 검정한 각 대비에 대한 열이 표시됩니다. 이 보고서에는 각 대비에 대해 추정값, 표준 오차, 해당 단일 대비 검정에 대한 t 비, 해당 p 값, 제곱합, 대비 추정값의 신뢰 구간 등이 제공됩니다.

모수 함수 보고서

"모수 함수" 보고서(그림 3.15)에는 지정한 대비가 모형 항의 선형 결합으로 표시됩니다.

최소 제곱 평균 대비의 예

최소 제곱 평균 대비 옵션을 설명하기 위해 처음 두 age 수준을 다음 두 수준과 비교하는 대비를 구성합니다.

다음 단계에 따라 그림 3.14에 표시된 보고서를 생성합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. **weight**를 선택하고 **Y**를 클릭합니다.
4. **age, sex 및 height**를 선택하고 **추가**를 클릭합니다.

5. "열 선택" 목록에서 **age** 를 선택하고 "모형 효과 생성" 목록에서 **height** 를 선택한 후 **교차**를 클릭합니다.
6. **실행**을 클릭합니다.
"최소 제곱 적합" 보고서가 나타납니다.
7. "age" 의 빨간색 삼각형을 클릭하고 **최소 제곱 평균 대비 ...** 를 선택합니다.

그림 3.14 age 에 대한 최소 제곱 평균 대비 지정

대비 지정		
age		
12	0	+
13	0	+
14	0	+
15	0	+
16	0	+
17	0	+

height 62.55

+ 또는 -를 클릭하여 대비 값을 만드십시오.

새 열 완료 도움말

8. age 12 와 13 에 대해 "+" 를 클릭합니다.
9. age 14 와 15 에 대해 "-" 를 클릭합니다.
이 대비는 키 (height) 62.55 에서 예측값을 기반으로 두 연령 (age) 그룹의 평균 체중 (weight) 이 다른지 여부를 검정합니다.
10. 연속형 효과 **height** 옆에 텍스트 상자가 있습니다. 기본값은 연속형 효과의 평균입니다.
11. **완료**를 클릭합니다.
12. "검정 상세 정보" 및 "모수 함수" 보고서를 엽니다.

그림 3.15에 "대비" 보고서가 표시되어 있습니다. 대비 검정은 0.05 수준에서 유의합니다. 평균 키 (height) 62.55에서 12세 및 13세 (age) 어린이의 예측 체중 (weight)이 14세 및 15세 (age) 어린이의 예측 체중 (weight)과 통계적으로 다르다는 결론을 내립니다.

그림 3.15 최소 제곱 평균 대비 보고서

age	
대비	
검정 상세 정보	
12	0.5
13	0.5
14	-0.5
15	-0.5
16	0
17	0
추정값	15.585
표준 오차	6.5334
t 비	2.3854
Prob> t	0.0243
SS	1059.4
95% 하한	2.1792
95% 상한	28.99

SS	NumDF	DenDF	F 비	Prob > F
1059	1	27	5.6900	0.0243*

height	62.55

모수 함수	
모수	
절편	0
age[13-12]	-0.5
age[14-13]	-1
age[15-14]	-0.5
age[16-15]	0
age[17-16]	0
sex[F]	0
height	0
(height-62.55)*age[13-12]	0
(height-62.55)*age[14-13]	0
(height-62.55)*age[15-14]	0
(height-62.55)*age[16-15]	0
(height-62.55)*age[17-16]	0

최소 제곱 평균 스튜던트 t 및 최소 제곱 평균 Tukey HSD

"최소 제곱 평균 스튜던트 t" 및 "최소 제곱 평균 Tukey HSD"(Honestly Significant Difference) 옵션은 모형 효과의 쌍별 비교를 검정합니다.

- 최소 제곱 평균 스튜던트 t 옵션은 일반적인 독립 표본, 등분산 t-검정을 기반으로 합니다. 각 비교는 지정된 유의 수준을 기반으로 합니다. 여러 비교를 수행하여 구한 전체 오차율이 지정된 유의 수준을 초과합니다.
- 최소 제곱 평균 Tukey HSD 옵션은 Tukey HSD 검정을 수행합니다. 이러한 비교를 위한 유의 수준은 전체 쌍별 비교 모음에 적용됩니다. 따라서 최소 제곱 평균 Tukey HSD의 신뢰 구간이 최소 제곱 평균 스튜던트 t의 신뢰 구간보다 넓습니다. 유의 수준은 표본 크기가 같으면 정확하고 표본 크기가 다르면 약간 작습니다 (Hayter 1984).

그림 3.16에서는 Big Class.jmp 샘플 데이터 테이블의 age 효과에 대한 최소 제곱 평균 Tukey 보고서를 보여 줍니다. 이 보고서를 생성하려면 **Fit Model** 데이터 테이블 스크립트를 실행하고 "age"의 빨간색 삼각형을 클릭한 후 **최소 제곱 평균 Tukey HSD**를 선택합니다. 기본적으로 이 보고서에는 교차표 보고서와 연결 문자 보고서가 표시됩니다.

그림 3.16 최소 제곱 평균 Tukey HSD 보고서

최소 제곱 평균 차이 Tukey HSD							
$\alpha = 0.050$ $Q = 3.02917$							
		최소 제곱 평균[j]					
평균[i]-평균[j]	12	13	14	15	16	17	
차이 표준 오차							
차이 CL 하한							
차이 CL 상한							
12		0 13.6856	25.8473	19.7698	10.159	-2.5203	
		0 6.96659	7.24776	8.04937	9.96733	10.5955	
		0 -7.4174	3.89261	-4.613	-20.034	-34.616	
		0 34.7886	47.8019	44.1527	40.3517	29.5754	
13	-13.686	0 12.1617	6.08426	-3.5266	-16.206		
	6.96659	0 6.82786	7.59798	9.69566	10.0859		
	-34.789	0 -8.521	-16.931	-32.896	-46.758		
	7.41737	0 32.8444	29.0998	25.8432	14.346		
14	-25.847	-12.162	0 -6.0774	-15.688	-28.368		
	7.24776	6.82786	0 6.28014	8.57237	8.65355		
	-47.802	-32.844	0 -25.101	-41.655	-54.581		
	-3.8926	8.52105	0 12.9462	10.2789	-2.1545		
15	-19.77	-6.0843	6.07741	0 -9.6108	-22.29		
	8.04937	7.59798	6.28014	0 9.23282	9.21068		
	-44.153	-29.1	-12.946	0 -37.579	-50.191		
	4.61302	16.9313	25.101	0 18.3569	5.61058		
16	-10.159	3.52657	15.6882	9.61083	0 -12.679		
	9.96733	9.69566	8.57237	9.23282	0 10.8916		
	-40.352	-25.843	-10.279	-18.357	0 -45.672		
	20.0337	32.8963	41.6554	37.5786	0 20.3131		
17	2.52026	16.2058	28.3675	22.2901	12.6793	0	
	10.5955	10.0859	8.65355	9.21068	10.8916	0	
	-29.575	-14.346	2.15448	-5.6106	-20.313	0	
	34.6159	46.7577	54.5806	50.1908	45.6716	0	
수준	최소 제곱 평균						
17	A	121.90696					
12	A	119.38671					
16	A B	109.22769					
13	A B	105.70112					
15	A B	99.61686					
14	B	93.53945					

교차표 보고서

두 옵션 모두 교차표 보고서라는 행렬을 표시합니다. 각 셀에는 평균 차이, 차이 표준 오차, 신뢰 하한 및 상한이 포함됩니다. 유의 수준 및 해당 임계값이 행렬 위에 표시됩니다. 기본 유의 수준은 0.05이지만 "모형 적합" 시작 창에서 다른 유의 수준을 지정할 수 있습니다. 통계적으로 차이가 있는 평균 쌍에 해당하는 셀은 빨간색으로 표시됩니다.

연결 문자 보고서

기본적으로 연결 문자 보고서는 교차표 행렬 아래에 나타납니다. 동일한 문자를 공유하거나 이 문자로 연결되는 수준은 통계적으로 차이가 없습니다. 공통 문자로 연결되지 않는 수준은 통계적으로 차이가 있습니다.

그림 3.16에서 17, 12, 16, 13 및 15 수준은 문자 A로 연결됩니다. 연결은 이러한 수준이 0.05 유의 수준에서 차이가 없음을 나타냅니다. 또한 16, 13, 15 및 14 수준은 문자 B로 연결되어 해당 수준이 통계적으로 차이가 없음을 나타냅니다. 그러나 age 17과 14 및 age 12와 14는 공통 문자로 연결되지 않으므로 이 두 쌍의 수준이 통계적으로 차이가 있음을 나타냅니다.

팁 : 연결 문자 보고서를 마우스 오른쪽 버튼으로 클릭하고 " 열 " 을 선택하여 연결 문자 (" 문자 "), 표준 오차 (" 표준 오차 ") 및 신뢰 구간 한계 ("X% 하한 ", "X% 상한 ") 가 포함된 열을 추가할 수 있습니다 . " 문자 " 열에서 연결 문자는 단일 열로 연결됩니다 . 유의 수준과 신뢰 수준은 " 모형 적합 " 시작 창에서 " 유의 수준 설정 " 옵션을 사용하여 지정한 유의 수준에 의해 결정됩니다 .

최소 제곱 평균 스튜던트 t 및 최소 제곱 평균 Tukey HSD 옵션

각 보고서 창에 나타나는 빨간색 삼각형 옵션은 선택적 보고서를 표시하거나 숨깁니다 . 최소 제곱 평균 스튜던트 t 에는 아래 옵션을 모두 사용할 수 있습니다 . 최소 제곱 평균 Tukey HSD 에는 처음 네 개 옵션을 사용할 수 있습니다 . 최소 제곱 평균 스튜던트 t 와 최소 제곱 평균 Tukey HSD 둘 다 교차표 보고서 및 연결 문자 보고서가 기본적으로 표시됩니다 .

교차표 보고서 각 수준 쌍에 대해 평균 차이 , 차이 표준 오차 및 차이의 신뢰 한계를 제공하는 이원 테이블을 표시합니다 . 유의한 차이를 포함하는 셀의 내용이 빨간색으로 강조 표시됩니다 .

연결 문자 보고서 유의한 비교와 유의하지 않은 비교를 연결 문자로 나타냅니다 . 동일한 문자로 연결되지 않은 수준에는 유의한 차이가 있습니다 . 동일한 문자로 연결된 수준에는 유의한 차이가 없습니다 .

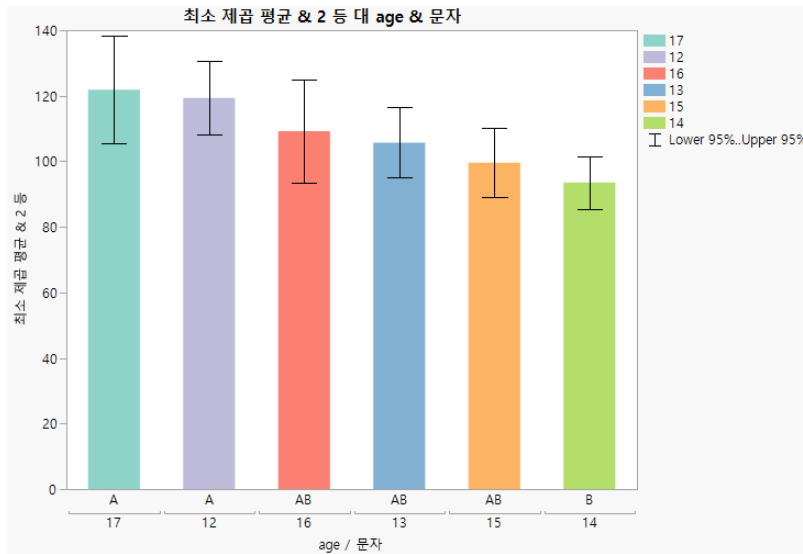
연결 문자 테이블 저장 효과 수준 , 연결 문자 , 최소 제곱 평균 , 표준 오차 및 신뢰 구간을 제공하는 열이 포함된 데이터 테이블을 생성합니다 . 이 테이블에는 신뢰 구간이 위에 겹쳐진 최소 제곱 평균의 막대 차트 (색상으로 구분됨) 를 생성하는 " 막대 차트 " 스크립트가 포함되어 있습니다 . 수준은 최소 제곱 평균이 큰 것부터 순서대로 배열됩니다 .

그림 3.17 에서는 Big Class.jmp 예의 막대 차트를 보여 줍니다 . Fit Model 데이터 테이블 스크립트를 실행하고 "age" 의 빨간색 삼각형을 클릭한 후 **최소 제곱 평균 Tukey HSD** 를 선택합니다 . " 최소 제곱 평균 차이 Tukey HSD " 보고서에서 **연결 문자 테이블 저장** 을 선택합니다 . 데이터 테이블이 나타나면 **막대 차트** 스크립트를 실행합니다 .

정렬된 차이 보고서 차이를 가장 큰 것부터 순서대로 정렬하고 표준 오차 , 신뢰 한계 및 p 값을 제공합니다 . 또한 신뢰 구간이 중첩된 막대 차트에 차이를 표시합니다 .

상세 비교 각 비교에 대한 개별 상세 보고서를 제공합니다 . 이 보고서에는 지정된 비교에 대한 추정 차이 , 표준 오차 , 신뢰 구간 , t 비 , 자유도 및 단측 / 양측 검정의 p 값이 표시됩니다 . 또한 비교에 대한 유의성 검정을 보여 주는 t 분포 그림도 표시됩니다 . 음영 부분의 영역은 양측 검정에 대한 p 값입니다 .

그림 3.17 최소 제곱 평균 차이 HSD 연결 문자 테이블의 막대 차트



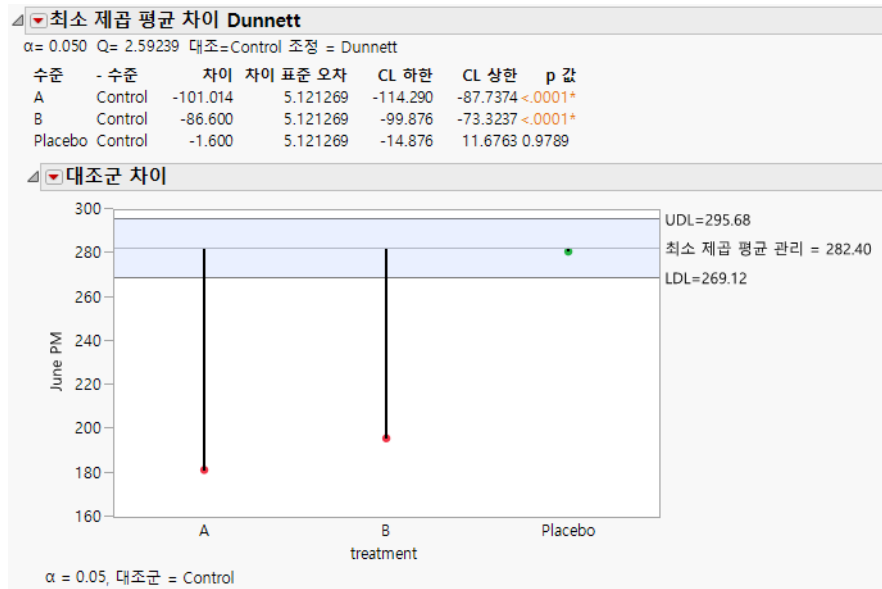
최소 제곱 평균 Dunnett

Dunnett 검정 (Dunnett 1955)에서는 일련의 평균을 대조군 평균과 비교합니다. 오차율은 쌍별 비교 모음에 적용됩니다. 최소 제곱 평균 Dunnett 옵션은 지정된 효과의 수준에 대해 Dunnett 검정을 수행합니다. p 값 및 신뢰 구간 계산에 Hsu 요인 분석 근사가 사용됩니다 (Hsu 1992).

"최소 제곱 평균 Dunnett"을 선택하면 효과의 대조 수준을 입력하는 대화상자가 나타납니다. "최소 제곱 평균 차이 Hsu-Dunnett" 보고서에는 유의 수준, 검정 통계량 (Q) 값 및 대조 수준이 표시 됩니다.

그림 3.18에서는 Cholesterol.jmp 샘플 데이터 테이블의 treatment 효과에 대해 "최소 제곱 평균 Dunnett" 옵션을 사용하여 나타나는 보고서를 보여 줍니다. 여기서 반응은 June PM 이고, Control 이라는 treatment 수준이 대조 수준으로 지정되었습니다.

그림 3.18 최소 제곱 평균 Dunnett 보고서



이 보고서에는 다음 두 가지 옵션이 있습니다.

대조군 차이 보고서 "대조군 차이" 보고서는 기본적으로 표시됩니다. 효과의 각 수준에 대해 대조 수준과 비교되는 수준, 추정 차이, 차이 표준 오차, 신뢰 구간 및 비교를 위한 p 값 등의 정보가 테이블에 표시됩니다.

대조군 차이 차트 대조 수준을 제외한 각 수준에서 점은 해당 수준의 최소 제곱 평균과 대조 수준의 최소 제곱 평균 간 차이를 표시합니다. 결정 상한 및 하한 (UDL, LDL) 이 그림에 표시됩니다. 이 보고서에는 "요약 보고서 표시" 옵션과 "표시 옵션"이 있습니다. "요약 보고서 표시" 옵션은 그림 상세 정보를 제공합니다. "표시 옵션"을 사용하면 그림 모양을 수정할 수 있습니다.

조각 검정

"조각 검정" 옵션은 명목형 또는 순서형 열로 구성된 교호작용 효과에 대해 활성화됩니다. 이 옵션은 교호작용에 포함된 각 명목형 또는 순서형 열의 각 수준에 대해 해당 수준과 관련된 설정의 전체 쌍별 비교를 공동으로 검정하는 보고서를 생성합니다. 검정은 사실상 교호작용의 지정된 "조각" 내의 차이를 검정하는 것입니다.

A의 수준 중 하나가 "Small"인 경우 A*B*C 교호작용에 관심이 있다고 가정해 보겠습니다. A = Small 조각에 대한 "조각 검정" 보고서에서는 A = Small일 때 B*C 수준의 전체 쌍별 비교를 공동으로 검정합니다. 이렇게 하면 교호작용 내에서 수준의 차이를 감지할 수 있습니다.

"조각 검정" 보고서는 "최소 제곱 평균 대비" 보고서와 동일한 형식을 따릅니다. 자세한 내용은 "최소 제곱 평균 대비..."(95 페이지)에서 확인하십시오.

검정력 분석

특정 효과의 F -검정에 대한 후향적 또는 전향적 상세 정보를 얻기 위해 정보를 입력할 수 있는 "검정력 상세 정보 창"을 엽니다.

참고: 필요한 차이를 감지할 수 있도록 충분한 수의 관측값이 연구에 포함되게 하려면 실험을 설계할 때 검정력에 대한 정보를 사용하십시오. 이러한 분석을 전향적 검정력 분석이라고 합니다. DOE 플랫폼을 사용하여 연구를 설계하는 것이 좋습니다. DOE > 표본 크기 및 검정력과 DOE > 설계 평가 둘 다 전향적 검정력 분석에 유용합니다. 표준 최소 제곱을 사용한 전향적 검정력 분석의 예는 "[전향적 검정력 분석](#)"(198 페이지)에서 확인하십시오.

그림 3.19에서는 Big Class.jmp 샘플 데이터 테이블에 대한 "검정력 상세 정보 창"의 예를 보여 줍니다. "검정력 상세 정보 창"을 사용하면 알파(α), 시그마(σ), 델타(δ) 및 개수(연구 크기) 값에 대한 검정력을 탐색할 수 있습니다. 단일 값("시작"만), 두 개의 값("시작", "끝") 또는 시작("시작"), 끝("끝") 및 값 시퀀스의 증분("증분")을 입력합니다. 지정한 값의 가능한 모든 조합에 대해 검정력 계산이 보고됩니다.

그림 3.19 검정력 상세 정보 창

검정력 상세 정보 창

age
각 열에 1개 또는 2개의 값이나 값 시퀀스를 입력하십시오.

	α	σ	δ	개수
시작:	0.050	13.15009	3	20
끝:	.	.	6	60
증분	.	.	1	10

☒ 검정력 계산
☒ 최소 유의 개수 계산
☐ 최소 유의값 계산
☐ 조정 검정력 및 신뢰 구간

완료 취소 도움말

모든 시퀀스 조합에 대해 계산이 수행됩니다.

자세한 내용은 "[검정력 분석](#)"(193 페이지)에서 확인하십시오.

"검정력 상세 정보 창" 보고서에는 다음 열과 옵션이 포함됩니다.

알파 (α) 검정의 유의 수준입니다. 이 값은 0에서 1사이이며 대개 0.05, 0.01 또는 0.10입니다. "모형 적합" 시작 창에서 "유의 수준 설정"을 선택한 후 다른 값을 설정한 경우를 제외하고 첫 번째 행에 표시된 알파의 초기값은 0.05입니다.

시그마 (σ) 모형의 잔차 오차에 대한 추정값입니다. 지침을 위해 첫 번째 행에 표시된 초기값은 RMSE(평균 제곱 오차의 제곱근)입니다.

델타 (δ) 관심 효과 크기입니다. 자세한 내용은 "[효과 크기](#)"(193 페이지)에서 확인하십시오. 첫 번째 행에 표시된 초기값은 가설에 대한 제곱합의 제곱근을 연구에 사용된 관측값 수의 제곱근으로 나눈 값(즉, $\delta = \sqrt{SS/n}$)입니다.

개수 (n) 표본 크기입니다. 첫 번째 행에 표시된 초기값은 현재 연구의 관측값 수입니다.

검정력 계산 검정력을 α, σ, δ 및 n 의 함수로 계산합니다. 검정력은 지정된 σ 및 n 에 대해 α 수준에서 유의한 검정 결과를 확인하여 δ 크기의 차이를 감지할 확률입니다. 자세한 내용은 "통계 상세 정보" 부록의 "**검정력 계산**"(517 페이지)에서 확인하십시오.

최소 유의 개수 계산 지정된 δ 및 σ 에 대해 α 수준에서 유의한 검정 결과를 얻기 위해 필요한 최소 관측값 수를 계산합니다. 자세한 내용은 "통계 상세 정보" 부록의 "**LSN 계산**"(516 페이지)에서 확인하십시오.

최소 유의값 계산 p 값 α 를 생성하는 모수의 선형 함수 또는 모수의 최소 양수 값을 계산합니다. 최소 유의값은 α, σ 및 n 의 함수입니다. 이 옵션은 1 자유도 검정에만 사용할 수 있습니다. 자세한 내용은 "통계 상세 정보" 부록의 "**LSV 계산**"(516 페이지)에서 확인하십시오.

조정 검정력 및 신뢰 구간 후향적 검정력 계산은 F 분포의 비중심성 모수를 추정할 때 표준 오차 및 검정 모수의 추정값을 사용합니다. 조정 검정력은 양의 편향이 제거된 비중심성 모수의 추정값에 기반한 후향적 검정력 계산입니다 (Wright and O'Brien 1988).

조정 검정력의 신뢰 구간은 비중심성 추정값의 신뢰 구간을 기반으로 합니다.

조정 검정력은 표본 추정값을 처리하므로 검정력 및 신뢰 한계는 현재 연구에서 추정된 δ 에 대해서만 계산됩니다. 자세한 내용은 "통계 상세 정보" 부록의 "**조정 검정력 계산**"(518 페이지)에서 확인하십시오.

적합 결여

"적합 결여" 보고서에는 모형이 데이터를 잘 적합시키는지 여부를 평가하는 검정의 상세 정보를 제공합니다. "적합 결여" 보고서는 이 검정을 수행할 수 있을 때만 나타냅니다. 이 검정은 모형에 독립적인 추정값을 사용하여 반응의 분산을 추정할 수 있는지가 관건입니다. 이 추정값을 생성하려면 모형 효과의 반복된 값에서 반응 값을 사용할 수 있어야 합니다. 검정에는 이러한 반복된 관측값을 사용하여 제곱합을 기반으로 순수 오차 추정값을 계산하는 과정이 포함됩니다.

다음 상황에서는 검정 통계량을 계산할 수 없으므로 "적합 결여" 보고서가 나타나지 않습니다.

- X 변수에 대해 반복된 점이 없어 순수 오차 제곱합을 계산할 수 없는 경우
- 모형이 포화되어 추정 모수가 관측값 수만큼 있는 경우. 이러한 모형은 완벽하게 적합되므로 적합 결여를 평가할 수 없습니다.

모형의 오차 제곱합과 순수 오차 제곱합 사이의 차이를 적합 결여 제곱합이라고 합니다. 모형이 적절하지 않은 경우 적합 결여 변동이 순수 오차 변동보다 매우 클 수 있습니다. 예를 들어 예측 변수의 함수 형태가 잘못되었거나, 모형의 교호작용이 충분하지 않거나 올바르지 않을 수 있습니다.

"적합 결여" 보고서에는 다음 열이 포함됩니다.

소스 변동의 소스로는 적합 결여, 순수 오차 및 총 오차의 세 가지가 있습니다.

DF 각 오차 소스의 DF(자유도)입니다.

- 총 오차의 DF는 분산 분석 테이블의 오차 행에 나오는 DF 값과 같습니다. 총 오차 DF는 제곱합 분해를 기반으로 적합 결여의 자유도와 순수 오차의 자유도로 분할됩니다.

- 순수 오차 DF는 반복된 각 관측값 그룹에서 풀링된 것입니다. 일반적으로 g 개의 그룹이 있고 각 그룹이 각 효과에 대해 동일한 설정을 갖는 경우 순수 오차 DF, 즉 DF_{PE} 는 다음과 같이 정의됩니다.

$$DF_{PE} = \sum_{i=1}^g (n_i - 1)$$

여기서 n_i 는 i 번째 그룹의 반복 실험 수입니다.

- 적합 결여 DF는 총 오차 DF와 순수 오차 DF 사이의 차이입니다.

제곱합 각 오차 소스에 대한 관련 SS(제곱합)입니다.

- 총 오차 SS는 해당 분산 분석 테이블의 오차 행(참조)에 나오는 제곱합입니다.
- 순수 오차 SS는 반복된 각 관측값 그룹에 대한 제곱합 값의 총계입니다. 순수 오차 SS를 DF로 나눈 값은 지정된 예측 변수 설정에서 반응의 분산을 추정합니다. 이 추정값은 모형의 영향을 받지 않습니다. 일반적으로 g 개의 그룹이 있고 각 그룹이 각 효과에 대해 동일한 설정을 갖는 경우 순수 오차 SS, 즉 SS_{PE} 는 다음과 같이 정의됩니다.

$$SS_{PE} = \sum_{i=1}^g SS_i$$

여기서 SS_i 는 관측된 각 반응과 i 번째 그룹의 평균 반응 간의 차이에 대한 제곱합입니다.

- 적합 결여 SS는 총 오차 제곱합과 순수 오차 제곱합 사이의 차이입니다.

평균 제곱 소스의 평균 제곱, 즉 제곱합을 DF로 나눈 값입니다. 순수 오차 평균 제곱에 비해 적합 결여 평균 제곱이 크면 모형이 잘 적합되지 않음을 나타냅니다. F 비는 공식 검정을 제공합니다.

F 비 적합 결여의 평균 제곱 대비 순수 오차의 평균 제곱에 대한 비율입니다. F 비는 적합 결여 및 순수 오차 평균 제곱에 의해 추정된 분산이 동일하다는 가설, 즉 "적합 결여 없음"을 나타내는 것으로 해석되는 가설을 검정합니다.

Prob > F 적합 결여 검정의 p 값입니다. p 값이 작으면 적합 결여가 유의함을 나타냅니다.

최대 R² 이러한 효과만 기반으로 한 모형에서 얻을 수 있는 최대 R²입니다. 순수 오차 제곱합은 모형 형태에 따라 변동되지 않습니다. 따라서 이러한 반복된 효과가 있는 모형에서 설명할 수 있는 최대 변동은 다음과 같습니다.

$$\frac{SS(\text{수정 합계}) - SS(\text{순수 오차})}{SS(\text{수정 합계})} = 1 - \frac{SS(\text{순수 오차})}{SS(\text{수정 합계})}$$

이 계산식은 최대 R²을 정의합니다.

추정값

"추정값" 메뉴에서는 모형 모수에 대한 추가 상세 정보를 제공합니다. JMP에서 명목형 및 순서형 효과가 코드화되는 방식을 검토하면 추정값을 이해하는 데 도움이 됩니다. 자세한 내용은 "[사용](#)"

자 검정 예에 대한 상세 정보"(188 페이지), "통계 상세 정보" 부록의 "명목형 요인"(493 페이지) 및 "통계 상세 정보" 부록의 "순서형 요인"(503 페이지)에서 확인하십시오.

모형에 임의 효과가 포함된 경우 다음 중 적절한 옵션만 "추정값" 메뉴에서 사용할 수 있습니다.

"추정값" 메뉴에서는 다음 옵션을 제공합니다.

예측 표현식 표시 추정된 모형에 대한 방정식을 포함하는 "예측 표현식" 보고서를 표시하거나 숨깁니다. 예는 "예측 표현식 표시"(107 페이지)에서 확인하십시오.

정렬된 추정값 선별 상황에서 유용할 수 있는 "정렬된 모수 추정값" 보고서를 표시하거나 숨깁니다. 설계가 포화되지 않은 경우 이 보고서는 절편을 제외한 항이 유의성이 큰 것부터 순서대로 정렬된 "모수 추정값" 보고서입니다. 설계가 포화된 경우에는 유사 t -검정이 제공됩니다. 자세한 내용은 "정렬된 추정값"(108 페이지)에서 확인하십시오.

확장 추정값 (하나 이상의 효과가 연속형이 아닌 경우에만 사용 가능) 명목형 효과의 모든 수준에 대한 모수 추정값을 제공하여 "모수 추정값" 보고서를 확장하는 "확장 추정값" 보고서를 표시하거나 숨깁니다. 자세한 내용은 "확장 추정값"(111 페이지)에서 확인하십시오.

표시자 파라미터화 추정값 (모형 효과 중 명목형 열이 있는 경우에만 사용 가능) 전통적 표시 함수를 사용하여 모형의 명목형 효과가 파라미터화된 모수 추정값을 포함하는 "표시 함수 파라미터화" 보고서를 표시하거나 숨깁니다. 자세한 내용은 "표시자 파라미터화 추정값"(113 페이지)에서 확인하십시오.

순차 검정 모형에 효과가 순차적으로 추가될 때 제곱합이 포함된 "순차 (유형 1) 검정" 보고서를 표시하거나 숨깁니다. 순차 제곱합을 기반으로 F -검정을 수행합니다. 자세한 내용은 "순차 검정"(114 페이지)에서 확인하십시오.

사용자 검정 사용자 가설을 검정할 수 있습니다. 자세한 내용은 "사용자 검정"(115 페이지)에서 확인하십시오.

다중 비교 효과 수준 간의 비교를 지정할 수 있습니다. 이러한 비교에는 단일 효과가 포함되거나, 유연한 사용자 비교를 정의할 수 있습니다. 전체 평균 및 대조군 평균과 비교하거나, Tukey HSD 또는 스튜던트 t 를 사용하여 전체 쌍별 비교를 수행할 수 있습니다. 스튜던트 t 방법을 지정하는 경우 동등성 검정도 수행하여 실제로 중요한 쌍별 차이를 식별할 수 있습니다. 자세한 내용은 "다중 비교"(117 페이지)에서 확인하십시오.

기울기 비교 (명목형 항과 연속형 항이 각각 하나이고 고정 효과에 대한 교호작용 효과가 있는 경우에만 사용 가능) ANCOVA(공분산 분석) 모형에서 교호작용 효과의 각 수준에 대한 기울기를 비교할 수 있는 보고서를 생성합니다. 자세한 내용은 "기울기 비교"(130 페이지)에서 확인하십시오.

결합 요인 검정 (모형에 교호작용이 포함된 경우에만 사용 가능) 모형의 각 주효과에 대해 해당 주효과와 관련된 모든 모수의 결합 검정을 표시하거나 숨깁니다. 자세한 내용은 "결합 요인 검정"(130 페이지)에서 확인하십시오.

역추정 예측 하나 이상의 반응 값에 대해 설명 변수 값을 예측할 수 있습니다. 자세한 내용은 "역추정 예측"(131 페이지)에서 확인하십시오.

Cox 혼합물 (모형에 혼합물 효과가 포함된 경우에만 사용 가능) Cox 혼합물 모형에 대한 모수 추정값을 생성합니다. 이를 사용하여 요인 효과를 도출하고 설계 공간의 참조점을 기준으로

반응 표면 모양을 추정할 수 있습니다. 자세한 내용은 "[Cox 혼합물](#)"(135 페이지)에서 확인하십시오.

모수 검정력 해당 가설 검정과 관련된 검정력 및 기타 상세 정보를 제공하는 열을 "모수 추정값" 보고서에 추가합니다. 자세한 내용은 "[모수 검정력](#)"(136 페이지)에서 확인하십시오.

추정값 상관계수 모형의 모든 모수 추정값에 대한 상관 행렬을 표시하거나 숨깁니다. 자세한 내용은 "[추정값 상관계수](#)"(138 페이지)에서 확인하십시오.

오차 규격 (임의 효과가 없는 경우에만 사용 가능) "최소 제곱 적합" 보고서에서 표준 오차 및 검정에 사용되는 오차 분산과 오차 자유도를 지정합니다. "오차 규격"을 변경해도 "스튜던트화 잔차" 그림과 "[Box-Cox 변환](#)" 보고서는 영향을 받지 않습니다. "오차 규격"이 "순수 오차" 또는 "지정됨"인 경우 "분산 분석" 보고서에 추가 열이 나타납니다. 자세한 내용은 "[분산 분석](#)"(86 페이지)에서 확인하십시오.

기본 추정값 모형의 표준 제곱근 평균 제곱 오차와 오차 자유도를 사용하여 모든 검정 및 표준 오차를 계산합니다.

순수 오차 "적합 결여" 보고서의 순수 오차 평균 제곱과 관련 자유도를 사용하여 모든 검정 및 표준 오차를 계산합니다. 자세한 내용은 "[적합 결여](#)"(104 페이지)에서 확인하십시오.

주의: 순수 오차 자유도가 1 이면 검정력이 약하고 신뢰 한계가 크다는 경고 메시지가 표시됩니다.

지정됨 사용자가 지정한 오차 분산 값과 오차 자유도를 사용하여 모든 검정 및 표준 오차를 계산합니다.

예측 표현식 표시

"예측 표현식 표시" 옵션은 반응을 예측하는 데 사용되는 방정식을 표시합니다. 그림 3.20에서는 Drug.jmp 샘플 데이터 테이블의 예를 보여 줍니다. 이 표현식은 일반적인 JMP 계산식으로 제공됩니다. 예를 들어 $x = 10$ 인 Drug a에 대한 반응을 예측하려면 반응식을 사용하여 $-2.696 - 1.185 + 0.987(10) = 5.99$ 를 계산합니다.

팁: 예측 계산식에 자릿수를 지정하려면 **파일 > 환경 설정 > 테이블**로 이동하여 **기본 필드 너비** 값을 변경합니다.

그림 3.20 예측 표현식

▲ 예측 표현식

$$-2.695772906 + \text{Match}\left(\text{Drug} \begin{pmatrix} \text{"a"} \Rightarrow -1.185036537 \\ \text{"d"} \Rightarrow -1.076065205 \\ \text{"f"} \Rightarrow 2.2611017426 \\ \text{else} \Rightarrow . \end{pmatrix} + 0.9871838111 \cdot x$$

위의 보고서를 생성하려면 다음 단계를 수행하십시오 .

예측 표현식의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Drug.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. Drug 와 x 를 선택하고 **추가**를 클릭합니다 .
5. **실행**을 클릭합니다 .
6. " 반응 y "의 빨간색 삼각형을 클릭하고 **추정값 > 예측 표현식 표시**를 선택합니다 . 그러면 그림 3.20 의 보고서가 나타납니다 .

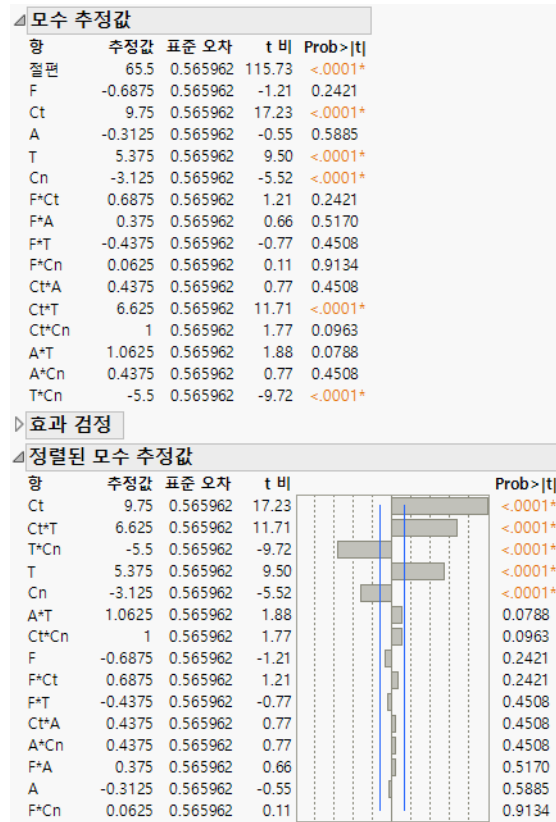
정렬된 추정값

"정렬된 추정값" 옵션은 선별 상황에서 유용한 "모수 추정값" 보고서를 생성합니다. 설계가 포화되지 않은 경우 "정렬된 추정값" 보고서에는 "모수 추정값" 보고서에서 절편을 제외한 항이 유의성이 큰 것부터 순서대로 정렬된 정보가 제공됩니다(그림 3.21의 두 번째 보고서). 설계가 포화된 경우에는 유사 t -검정이 제공됩니다. 이 검정은 Lenth 유사 표준 오차를 기반으로 합니다 (Lenth 1989). 자세한 내용은 "[Lenth PSE](#)"(110 페이지)에서 확인하십시오.

정렬된 추정값 보고서의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Reactor.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. 창 아래쪽 근처의 **차수** 상자에 2 가 표시되어 있는지 확인합니다 .
5. F, Ct, A, T 및 Cn 을 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다 .
6. **실행**을 클릭합니다 .
7. " 반응 Y "의 빨간색 삼각형을 클릭하고 **추정값 > 정렬된 추정값**을 선택합니다 .

그림 3.21 정렬된 모수 추정값



"강조"가 "효과 선별"로 설정되고 모든 효과에 모수가 하나만 있는 경우 "정렬된 모수 추정값" 보고서도 자동으로 나타납니다.

"모수 추정값" 보고서와 "정렬된 모수 추정값" 보고서 (둘 다 그림 3.21에 표시됨)에는 다음과 같은 차이가 있습니다.

- "정렬된 모수 추정값" 보고서에는 절편이 표시되지 않습니다.
- 효과가 t 비의 절대값을 기준으로 정렬되며 가장 유의한 효과가 맨 위에 표시됩니다.
- 막대 차트에는 0.05 유의 수준에 대한 임계값을 나타내는 수직선과 함께 t 비가 표시됩니다.

포화 모형의 정렬된 추정값 보고서

선별 실험에는 자유도가 충분하지 않아 오차를 추정할 수 없는 완전 포화 모형이 포함되는 경우가 많습니다. 이러한 경우 "정렬된 추정값" 보고서(그림 3.21)에서는 상대 표준 오차를 제공하고, Lenth PSE(유사 표준 오차)를 사용하여 t 비와 p 값을 생성합니다. 이러한 값은 이름에 유사라는 라벨이 지정됩니다. 자세한 내용은 "Lenth PSE"(110 페이지) 및 "유사 t 비"(110 페이지)에서 확인하십시오. 노트에 변경 사항이 설명되고 PSE가 표시됩니다.

보고서에는 다음 열이 포함됩니다.

항 관심 계수가 있는 모형 항입니다.

추정값 p 값이 작은 것부터 순서대로 정렬된 모수 추정값이 제공됩니다.

상대 표준 오차 잔차 오차에 대한 자유도가 없는 경우 보고서에 상대 표준 오차가 제공됩니다.
상대 표준 오차는 제공된 평균 제곱 오차를 1로 설정하여 계산됩니다.

유사 t 비 유사 표준 오차를 사용하여 계산된 추정값의 t 비입니다. Lenth PSE 값이 보고서 하단에 표시됩니다.

유사 p 값 $m/3$ 의 DFE(오차 자유도) 값을 사용하여 계산된 p 값입니다. 여기서 m 은 절편을 제외한 모수의 수입니다. DFE 값이 보고서 하단에 표시됩니다.

Lenth PSE

Lenth PSE(유사 표준 오차)는 Lenth(1989)로 인한 잔차 오차의 추정값입니다. 이는 선별 실험에서 비교적 적은 수의 효과가 활성화되는 효과 희소성 원리를 기반으로 합니다. 비활성 효과는 랜덤 잡음을 나타내고 Lenth 추정값의 기반을 만듭니다.

값은 다음과 같이 계산됩니다.

1. 모든 비절편 모수의 절대값을 고려합니다.
2. 절대값이 $3.75 \times$ 추정값 절대값의 중앙값을 초과하는 모든 모수 추정값을 제거합니다.
3. 모수 추정값에 대한 나머지 절대값의 중앙값에 1.5를 곱합니다.

유사 t 비

상대 표준 오차가 같으면 Lenth PSE가 보고서 하단에 표시됩니다. 유사 t 비는 다음과 같이 계산됩니다.

$$\text{유사 } t \text{ 비율} = \frac{\text{추정값}}{\text{PSE}}$$

상대 표준 오차가 같지 않으면 T 척도 Lenth PSE가 계산됩니다. 이 값은 추정값의 PSE를 상대 표준 오차로 나눈 값입니다. 유사 t 비는 다음과 같이 계산됩니다.

$$\text{유사 } t \text{ 비율} = \frac{\text{추정값}}{T \text{ 척도 Lenth PSE} \times \text{상대 표준 오차}}$$

지정된 추정값의 표준 오차를 추정하기 위해 T 척도 Lenth PSE는 추정값의 상대 표준 오차를 곱하여 조정됩니다.

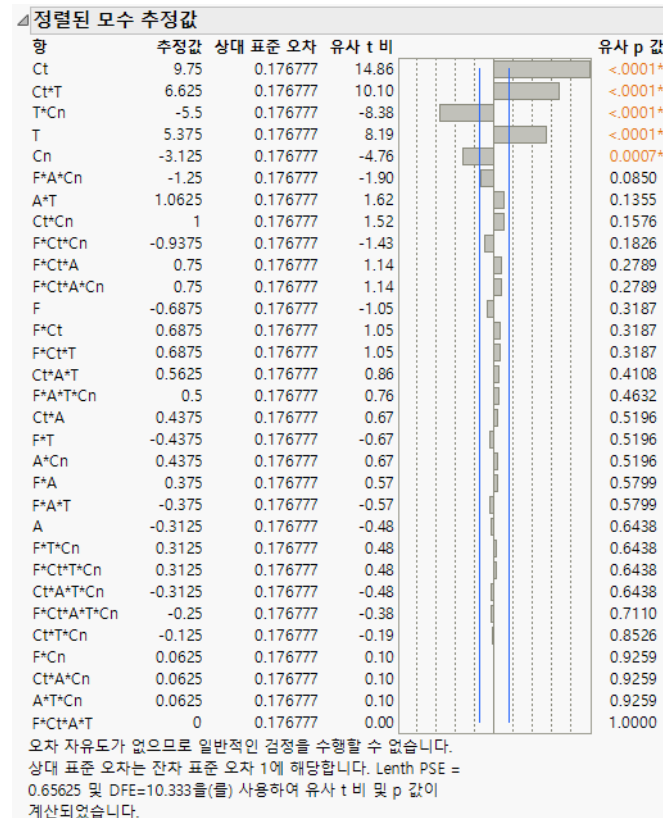
포화 모형의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Reactor.jmp를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. **Y**를 선택하고 **Y**를 클릭합니다.

4. 5 개 열 (F, Ct, A, T 및 Cn) 을 선택합니다 .
5. 매크로 버튼을 클릭하고 **완전 요인**을 선택합니다 .
6. **실행**을 클릭합니다 .
7. " 반응 Y" 의 빨간색 삼각형을 클릭하고 **추정값 > 정렬된 추정값**을 선택합니다 .

사용된 Lenth PSE 및 자유도가 보고서 하단에 제공됩니다 . 보고서에 따르면 유사 p 값을 기 반으로 Ct, Ct*T, T*Cn, T 및 Cn 효과가 매우 유의합니다 .

그림 3.22 포화 모형의 정렬된 모수 추정값 보고서



확장 추정값

모수 추정값을 처리하는 경우 JMP에서 명목형 및 순서형 열이 코드화되는 방식을 이해해야 합니다. 명목형 열을 코드화하는 방법에 대한 자세한 내용은 "[사용자 검정 예에 대한 상세 정보](#)"(188 페이지)에서 확인하십시오. 순서형 열을 코드화하고 모델링하는 방법에 대한 자세한 내용은 "통계 상세 정보" 부록의 "[명목형 요인](#)"(493 페이지) 및 "통계 상세 정보" 부록의 "[순서형 요인](#)"(503 페이지)에서 확인하십시오.

모형에 명목형 항이 있고 전체 추정값 집합에 대한 상세 정보를 확인하려는 경우 "확장 추정값" 옵션을 사용합니다. "확장 추정값" 옵션은 추정값, 표준 오차, t 비 및 p 값을 제공합니다.

확장 추정값 보고서의 예

1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Drug.jmp 를 엽니다.
2. 분석 > 모형 적합을 선택합니다.
3. Y 를 선택하고 Y 를 클릭합니다.
4. Drug 와 x 를 선택하고 추가를 클릭합니다.
5. 실행을 클릭합니다.
6. "반응 y" 의 빨간색 삼각형을 클릭하고 추정값 > 확장 추정값을 선택합니다.

그림 3.23에 "확장 추정값" 보고서와 "모수 추정값" 보고서가 표시되어 있습니다. Drug[f] 항의 추정값은 "확장 추정값" 보고서에 나타납니다. 검정의 귀무가설은 Drug f 그룹의 평균이 전체 평균과 다르지 않다는 것입니다. Drug[f]에 대한 검정이 0.05 수준에서 유의하므로 Drug f 그룹의 평균 반응이 전체 반응과 차이가 있음을 나타냅니다. 자세한 내용은 "확장 추정값에 대한 검정 해석"(112 페이지)에서 확인하십시오.

그림 3.23 모수 추정값과 확장 추정값의 비교

반응 y				
모수 추정값				
항	추정값	표준 오차	t 비	Prob> t
절편	-2.695773	1.911085	-1.41	0.1702
Drug[a]	-1.185037	1.060822	-1.12	0.2742
Drug[d]	-1.076065	1.041298	-1.03	0.3109
x	0.9871838	0.164498	6.00	<.0001*
확장 추정값				
명목형 요인이 모든 수준으로 확장되었습니다.				
항	추정값	표준 오차	t 비	Prob> t
절편	-2.695773	1.911085	-1.41	0.1702
Drug[a]	-1.185037	1.060822	-1.12	0.2742
Drug[d]	-1.076065	1.041298	-1.03	0.3109
Drug[f]	2.2611017	1.093974	2.07	0.0488*
x	0.9871838	0.164498	6.00	<.0001*

확장 추정값에 대한 검정 해석

모형이 n 개 수준을 가진 단일 명목형 요인으로 구성되어 있다고 가정해 보겠습니다. 해당 요인은 $n-1$ 개 수준에 대해 각각 하나씩 $n-1$ 개 표시 변수로 나타냅니다. 이러한 $n-1$ 개 표시 변수 중 하나에 해당하는 모수 추정값은 해당 수준에 대한 평균 반응과 모든 수준에 대한 평균 반응 간의 차이입니다. JMP에서 명목형 변수를 코드화하는 방식 때문에 이렇게 표현됩니다("사용자 검정 예에 대한 상세 정보"(188 페이지) 참조). 모수 추정값은 대개 해당 수준의 효과로 해석됩니다.

Cholesterol.jmp 샘플 데이터 테이블에서 단일 요인 treatment와 반응 June PM을 예로 들어보겠습니다. 항 또는 표시 변수인 treatment[A]와 관련된 모수 추정값은 treatment A에 대한 June PM 평균과 June PM 전체 평균 간의 차이입니다.

명목형 변수의 모든 수준에 대한 효과는 합이 0이 되어야 합니다. 수준 순서에서 마지막 수준, 즉 $-1s$ 로 코드화된 수준의 효과를 가정해 보겠습니다. 이 수준의 효과는 다른 $n-1$ 개 수준에 대한 효과 합의 반대 부호 값입니다. 따라서 마지막 수준의 효과는 다른 $n-1$ 개 수준에 대한 모수 추정값 합의 반대 부호 값입니다.

"추정값" 메뉴의 "확장 추정값" 옵션은 결측 추정값을 계산하고 명목형 열을 포함하는 모든 효과를 검정한 후 텍스트 보고서에 표시합니다. 이러한 효과의 모든 수준에서 추정값의 평균 또는 합이 0인지 확인할 수 있습니다. 특히 이 관계는 이러한 추정값 및 관련 검정이 서로 독립적이지 않음을 나타냅니다.

그림 3.23에 표시된 Drug.jmp 보고서에서 Drug와 관련된 항의 추정값은 공변량 x 를 포함하는 모형을 기반으로 합니다.

참고 :

- Drug[a]에 대한 추정값은 Drug a에 대한 최소 제곱 평균과 y 의 전체 평균 간의 차이입니다.
- "확장 추정값" 보고서에 제공된 Drug[f]에 대한 추정값은 Drug[a] 및 Drug[d]에 대한 추정값 합의 반대 부호 값입니다.
- "확장 추정값" 보고서에 제공된 Drug[f]에 대한 t -검정은 Drug f 그룹에 대한 반응이 전체 평균 반응과 다른지 여부를 검정합니다.
- 명목형 요인이 고차 교호작용에 포함된 경우 "확장 추정값" 보고서가 길어질 수 있습니다. 예를 들어 2 수준 명목형 요인의 5 원 교호작용은 단일 모수 추정값을 생성하지만 부호 변화까지 모두 동일한 $2^5 = 32$ 개의 확장된 효과를 가집니다.

표시자 파라미터화 추정값

이 옵션은 명목형 열이 표시자(SAS GLM) 파라미터화를 사용하여 코드화되고 연속형으로 처리되는 모형에 대한 모수 추정값을 제공하는 "표시 함수 파라미터화" 보고서를 표시합니다. 순서형 열은 일반적인 JMP 코딩 체계를 사용하여 코드화됩니다. SAS GLM 및 JMP 코딩 체계에 대한 설명은 "통계 상세 정보" 부록의 "**요인 모형**"(492페이지)에서 확인하십시오.

JMP 코딩 체계에서 명목형 변수 수준의 표시자에 해당하는 추정값은 같은 수준의 평균 반응과 모든 수준의 평균 반응 간 차이에 대한 추정값입니다. JMP 코딩을 보려면 "표준 최소 제곱"의 빨간색 삼각형 메뉴에서 **열 저장 > 코딩 테이블 저장**을 선택합니다.

표시자 코딩 체계에서 명목형 변수 수준의 표시자에 해당하는 추정값은 같은 수준의 평균 반응과 마지막 수준의 평균 반응 간 차이에 대한 추정값입니다. 마지막 수준은 값 순서 코딩이 가장 높은 수준이며 표시 함수가 모형에 포함되지 않은 수준입니다.

주의 : "표시 함수 파라미터화" 보고서에 제공된 표준 오차 및 t 비는 "모수 추정값" 보고서의 값과 다릅니다. 이는 추정하는 모수가 서로 다르기 때문입니다.

그림 3.24 의 보고서를 생성하려면 "**확장 추정값 보고서의 예**"(112 페이지) 의 단계를 수행하십시오. 여기서 "확장 추정값" 을 선택하는 대신 **표시자 파라미터화 추정값**을 선택합니다.

그림 3.24 표시자 파라미터화 추정값

반응 y				
표시 함수 파라미터화				
항	추정값	표준 오차	t 비	Prob> t
절편	-0.434671	2.471354	-0.18	0.8617
Drug[a]	-3.446138	1.886781	-1.83	0.0793
Drug[d]	-3.337167	1.853866	-1.80	0.0835
x	0.9871838	0.164498	6.00	<.0001*

JMP 코딩 체계는 모형 적합의 일반화 회귀 분석법을 제외하고 JMP 전체에서 명목형 변수에 사용됩니다. 일반화 회귀 분석법의 명목형 변수에 사용되는 코딩 체계에 대한 자세한 내용은 "일반화 회귀 모형" 장의 "**일반화 회귀 분석법 시작**"(269페이지)에서 확인하십시오.

JMP를 사용하는 경우와 SAS GLM 파라미터화를 사용하는 경우 각각 파생된 모형에 차이가 있을 수 있습니다. 일부 모형은 동등합니다. 다른 모형(예: 절편이 없는 모형, 결측 셀이 있는 모형, 명목형 또는 순서형 효과가 있는 모형, 혼합물 모형)은 차이를 나타낼 수 있습니다.

순차 검정

"순차 검정" 보고서에는 모형에 효과가 순차적으로 추가될 때 제곱합 및 검정이 표시됩니다. 입력 순서는 "모형 적합" 시작 창의 "모형 효과 생성" 목록에 나타나는 효과의 순서로 정의됩니다. 그림 3.25 에는 Drug.jmp 샘플 데이터 테이블에 대한 보고서가 표시되어 있습니다.

그림 3.25 순차 검정 보고서

반응 y					
순차(유형 1) 검정					
소스	N모수	DF	Seq SS	F 비	Prob > F
Drug	2	2	293.60000	9.1486	0.0010*
x	1	1	577.89740	36.0145	<.0001*

순차 검정의 기반이 되는 제곱합을 **제1종 제곱합**이라고도 합니다. 이 값은 지정된 효과 입력 순서에 따라 모형을 적합시켜 계산됩니다. 특정 효과를 예로 들겠습니다. 해당 효과 전에 입력한 모든 효과를 포함하는 모형에 대한 모형 제곱합을 계산합니다. 그런 다음 이러한 효과와 지정된 효과를 모두 포함하는 모형에 대한 모형 제곱합을 계산합니다. 지정된 효과에 대한 순차 제곱합은 모형 제곱합의 증가입니다.

그림 3.25에서는 Drug.jmp 샘플 데이터 테이블에 대한 순차 제곱합을 보여 줍니다. "모형 적합" 시작 창에서 Drug를 먼저 입력한 후 x를 입력했습니다. Drug로만 구성된 모형의 제곱합은 293.6입니다. x가 모형에 추가되면 모형 제곱합이 871.4974가 됩니다. 증가한 값 577.8974가 x에 대한 순차 제곱합입니다.

"순차(유형 1) 검정" 보고서에 표시된 검정은 순차 제곱합에 기반한 F-검정이며 **제1종 검정**이라고도 합니다. F비는 지정된 효과를 검정합니다. 여기서 모형은 해당 효과 및 위의 "소스" 열에 나열된 효과만 포함합니다.

순차 제곱합을 합하면 모형 제곱합이 됩니다. 또한 일반적인 모형을 가정했을 때 각 값이 통계적으로 독립적이라는 좋은 특징이 있습니다. 그러나 모형의 항 순서에 따라 값이 달라지므로 적절하지 않은 경우가 많습니다.

다음과 같은 경우 순차 검정이 적절합니다.

- 적절한 순서로 지정된 분산 모형의 균형 분석 (즉, 효과 목록에서 주효과 다음에 이원 교호 작용이 나오는 순서)
- 적절한 순서로 지정된 순수 내포 모형
- 적절한 순서로 지정된 다항 회귀 모형

"모수 추정값" 및 "효과 검정" 보고서에 제공된 검정은 제3종 제곱합을 기반으로 합니다. 여기서 효과에 대한 제곱합은 다른 모든 효과가 모형에 입력된 후 해당 효과에 의해 설명되는 추가 제곱합입니다.

사용자 검정

모형 모수와 관련된 하나 이상의 사용자 가설을 검정하려면 "추정값" 메뉴에서 **사용자 검정**을 선택합니다. 이 창에서 모형 모수에 대한 하나 이상의 선형 함수 또는 대비를 지정할 수 있습니다.

결과에는 각 대비에 대한 개별 검정과 모든 대비에 대한 결합 검정이 포함됩니다. 자세한 내용은 그림 3.26에서 확인하십시오. 개별 대비 보고서에는 모수의 지정된 선형 함수에 대한 추정값과 표준 오차가 제공됩니다. t 비, p 값 및 관련 제곱합도 제공됩니다. 개별 대비 결과 아래에 나오는 모든 대비의 결합 검정에서는 제곱합, 분자 자유도, F 비 및 p 값을 제공합니다.

주의 : 이러한 검정은 잔차 오차를 사용하여 수행됩니다. 모형에 임의 효과가 있고 REML 대신 EMS 를 사용할 경우에는 이러한 검정이 적절하지 않을 수 있습니다.

참고 : 고차 효과에 포함된 효과를 검정하는 경우 사용자 검정 대신 최소 제곱 평균에 대한 검정을 사용하는 것이 좋습니다. 최소 제곱 평균은 다른 모형 효과에 대해 조정됩니다. "효과 상세 정보"에서 최소 제곱 평균 대비를 검정할 수 있습니다.

사용자 검정 보고서 구성 요소

"사용자 검정" 규격 창에는 다음과 같은 구성 요소가 있습니다.

편집 가능한 텍스트 상자 "사용자 검정" 제목 표시줄 아래의 공간은 검정 이름을 입력할 수 있는 편집 가능한 영역입니다.

모수 모형 항을 나열합니다. 항 목록의 오른쪽에는 해당 모수에 대한 0 열이 있습니다. 이 셀에 값을 입력하여 검정을 위한 선형 함수를 지정합니다.

"=" 기호 "모수" 목록의 마지막 행에는 =라벨이 지정됩니다. 이 셀에 상수를 입력하여 각 대비에 대한 규격을 완료합니다.

열 추가 모수의 여러 선형 함수를 공동으로 검정할 수 있도록 0 열을 추가합니다.

완료 " 완료 " 버튼을 클릭하여 검정을 수행합니다 . 그러면 보고서가 변경되어 각 검정 열에 대한 검정 통계량 값 , 표준 오차 및 기타 통계량이 표시됩니다 . 모든 열에 대한 결합 F - 검정은 보고서 하단의 상자에 제공됩니다 .

사용자 검정 보고서 옵션

" 사용자 검정 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

검정력 분석 결합 검정에 대한 검정력 분석을 제공합니다 . 이 옵션은 검정이 수행된 후에만 사용할 수 있습니다 . 자세한 내용은 " **모수 검정력** "(136 페이지) 에서 확인하십시오 .

제거 " 사용자 검정 " 보고서를 제거합니다 .

참고 : 결합 사용자 검정을 여러 번 수행하려면 **추정값 > 사용자 검정**을 반복해서 선택하십시오 .

그림 3.26에서는 Cholesterol.jmp 샘플 데이터 테이블을 사용하여 세 가지 대비가 포함된 규격 창의 예를 보여 줍니다 . 세 가지 검정 모두 상수가 0으로 설정됩니다 . 이 검정에 대한 보고서가 그림 3.27에 표시되어 있습니다 .

사용자 검정의 예

Cholesterol.jmp 샘플 데이터 테이블은 6 개 기간에서 20명의 환자를 대상으로 한 반복 측정을 제공합니다 . 네 개의 **treatment** 그룹이 연구됩니다 . 일반적으로 이 데이터는 모든 반복 측정을 반응으로 사용하여 적절하게 분석해야 합니다 . 이 예에서는 June PM에 대한 반응만 고려합니다 .

세 개의 대비를 검정하려는 경우를 가정해 보겠습니다 . 다음에 대한 평균 반응을 비교하려고 합니다 .

- treatment A 대 treatment B
- treatment A/B 대 대조군
- treatment A/B 대 대조군 / 플라시보 그룹

사용자 검정을 사용하여 이러한 대비를 검정하려면 다음을 수행하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cholesterol.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. June PM 을 선택하고 **Y** 를 클릭합니다 .
4. **treatment** 를 선택하고 **추가**를 클릭합니다 .
5. **실행**을 클릭합니다 .
6. " 반응 June PM" 옆의 빨간색 삼각형을 클릭하고 **추정값 > 사용자 검정**을 선택합니다 .
7. " 사용자 검정 " 규격 보고서에서 **열 추가**를 두 번 클릭하여 두 개의 추가 열을 생성합니다 .
8. 그림 3.26 에 표시된 것과 같이 편집 가능 영역에 검정 이름을 입력하고 세 열에 값을 입력합니다 .

이러한 값, 특히 세 번째 열의 값을 구하는 방법은 "통계 상세 정보" 부록의 "[모수 해석](#)"(493 페이지)에서 확인하십시오.

그림 3.26 세 가지 대비에 대한 사용자 검정 규격 보고서

반응 June PM

사용자 검정

Three joint tests

모수

절편	0	0	0
treatment[A]	1	0.5	0.5
treatment[B]	-1	0.5	0.5
treatment[Control]	0	-1	-0.5
=	0	0	-0.5

가설 검정을 구성하려면 클릭하고 위에 입력하십시오.

완료 열 추가 도움말

9. **완료**를 클릭합니다.

그림 3.27에 표시된 결과는 세 가설 모두 개별적으로 유의할 뿐만 아니라 공동으로도 유의하다는 것을 나타냅니다.

그림 3.27 세 가지 대비에 대한 검정을 보여 주는 사용자 검정 보고서

사용자 검정

Three joint tests

모수

절편	0	0	0
treatment[A]	1	0.5	0.5
treatment[B]	-1	0.5	0.5
treatment[Control]	0	-1	-0.5
=	0	0	-0.5

값	-14.41375873	-93.80687936	-69.65515952
표준 오차	5.1212686663	4.4351487646	3.0026099078
t 비	-2.814489859	-21.15078532	-23.19820478
Prob> t	0.0124630384	4.031938e-13	9.608264e-14
SS	519.39110157	29332.435386	35286.118168

제곱합	42009.051559
분자 DF	3
F 비	213.56323625
Prob > F	4.100821e-13

다중 비교

모형 효과 수준으로 정의된 평균을 비교하는 검정 및 신뢰 수준을 얻으려면 이 옵션을 사용합니다. 다중 비교 방법의 목표는 잘못된 결론에 도달할 확률을 제어하면서 그룹 평균이 다른지 여부를 확인하는 것입니다. "다중 비교" 옵션을 사용하면 그룹 평균을 전체 평균(평균 분석) 및 대조군 평균과 비교할 수 있습니다. Tukey HSD 또는 스튜던트 t 를 사용하여 쌍별 비교를 수행할 수도 있습니다. 스튜던트 t 방법을 지정하는 경우 동등성 검정도 수행하여 실제로 중요한 쌍별 차이를 식별할 수 있습니다.

스튜던트 t 방법은 개별 비교에 대한 오차율만 제어합니다. 따라서 정확한 다중 비교 절차는 아닙니다. 제공된 다른 모든 방법은 관심 있는 모든 비교에 대한 전체 오차율을 제어합니다. 이러한 각 방법은 p 값 및 신뢰 한계를 계산할 때 다중 비교 조정을 사용합니다.

모형에 명목형 및 순서형 효과가 포함된 경우 최소 제곱 평균 추정값을 사용하여 비교를 수행하거나, 사용자 정의 추정값을 사용하여 특정 비교를 정의할 수 있습니다. 모형에 연속형 효과만 포함된 경우에는 사용자 정의 추정값을 사용하여 평균을 비교할 수 있습니다.

팁: 연속형 효과가 비교적 적은 수준으로 구성되어 있다고 가정해 보겠습니다. 최소 제곱 평균 추정값을 사용하여 비교하려는 경우 해당 효과에 순서형 또는 명목형 모델링 유형을 할당하는 것이 좋습니다.

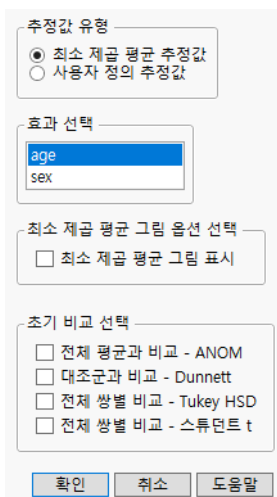
옵션 시작

그림 3.28에서는 "다중 비교" 옵션에 대한 컨트롤 창의 예를 보여 줍니다. 이 예는 **Big Class.jmp** 데이터 테이블을 기반으로 하며 **weight**를 Y 로 지정하고 **age**, **sex** 및 **height**를 모형 효과로 지정합니다. 비교를 위해 최소 제곱 평균 추정값과 사용자 정의 추정값의 두 가지 추정값 유형을 사용할 수 있습니다.

최소 제곱 평균 추정값

이 옵션은 모형에 명목형 또는 순서형 효과가 있는 경우에만 사용할 수 있으며 최소 제곱 평균을 비교합니다. 최소 제곱 평균은 모형에 있는 다른 효과의 일부 중립 값에서 계산된 평균입니다. 최소 제곱 평균의 정의는 "**최소 제곱 평균 테이블**"(91 페이지)에서 확인하십시오. 관심 효과를 선택해야 합니다. 그림 3.28에는 **age**에 대한 최소 제곱 평균 추정값이 지정되어 있습니다. 최소 제곱 평균 그림을 표시하는 옵션이 있습니다. 자세한 내용은 "**최소 제곱 평균 그림 옵션**"(93 페이지)에서 확인하십시오.

그림 3.28 최소 제곱 평균 추정값에 대한 시작 창



The dialog box titled "추정값 유형" (Estimation Type) contains the following options:

- 추정값 유형** (Estimation Type):
 - ☒ 최소 제곱 평균 추정값 (Minimum Squared Mean Estimation)
 - ☐ 사용자 정의 추정값 (Custom Estimation)
- 효과 선택** (Effect Selection):
 - age (selected)
 - sex
- 최소 제곱 평균 그림 옵션 선택** (Minimum Squared Mean Plot Option Selection):
 - ☐ 최소 제곱 평균 그림 표시 (Show Minimum Squared Mean Plot)
- 초기 비교 선택** (Initial Comparison Selection):
 - ☐ 전체 평균과 비교 - ANOM
 - ☐ 대조군과 비교 - Dunnett
 - ☐ 전체 쌍별 비교 - Tukey HSD
 - ☐ 전체 쌍별 비교 - 스튜던트 t

Buttons at the bottom: 확인 (OK), 취소 (Cancel), 도움말 (Help).

사용자 정의 추정값

그림 3.29에서는 사용자 정의 추정값 규격을 보여 줍니다. age의 세 가지 수준과 sex의 두 가지 수준 모두 선택되었습니다. 또한 height의 두 값이 수동으로 입력되었습니다. "추정값 추가" 버튼을 클릭하여 지정된 수준의 가능한 모든 조합이 나열되었습니다. 여기서 추정값을 더 지정하고 "추정값 추가" 버튼을 다시 클릭하여 "비교를 위한 추정값" 목록에 추가할 수 있습니다.

그림 3.29 사용자 정의 추정값에 대한 시작 창

추정값 유형

☐ 최소 제곱 평균 추정값

☒ 사용자 정의 추정값

age 수준 선택

12
13
14
15
16
17

sex 수준 선택

F
M

height

62
68
.
.
.
.

필요한 경우 요인 설정을 선택하고 [추정값 추가] 버튼을 클릭하여 사용자 정의 추정값을 생성하십시오.

비교를 위한 추정값

age	sex	height
12	F	62
12	F	68
12	M	62
12	M	68
14	F	62
14	F	68
14	M	62
14	M	68
17	F	62
17	F	68
17	M	62
17	M	68

초기 비교 선택

☐ 전체 평균과 비교 - ANOM

☐ 대조군과 비교 - Dunnett

☐ 전체 쌍별 비교 - Tukey HSD

☐ 전체 쌍별 비교 - 스튜던트 t

"사용자 정의 추정값"을 사용하는 경우 다음과 같이 모델링 유형에 따라 수준이 지정되지 않은 효과가 설정됩니다.

- 연속형 효과는 효과의 평균으로 설정됩니다.
- 명목형 및 순서형 효과는 값 순서화에서 첫 번째 수준으로 설정됩니다.

참고: 이 섹션에서는 평균이라는 용어를 사용하여 최소 제공 평균 추정값 또는 사용자 정의 추정값을 나타냅니다.

최소 제공 평균 그림 옵션 선택

최소 제공 평균 그림 표시를 선택하면 최소 제공 평균 그림을 나타낼 수 있습니다. 효과가 교호작용 항인 경우 교호작용 그림을 생성할 수도 있습니다. 중첩을 위한 항을 선택합니다. 교호작용 그림을 선택하지 않으면 최소 제공 그림에 효과 항이 내포됩니다. 자세한 내용은 "[최소 제공 평균 그림 옵션](#)"(93 페이지)에서 확인하십시오.

초기 비교 선택

추정값을 지정한 후에는 "초기 비교 선택"에서 옵션을 선택하여 초기 보고서에 표시할 비교 유형을 선택할 수 있습니다. 아니면 옵션을 선택하지 않고 "확인"을 클릭합니다.

전체 평균과 비교 - ANOM 각 효과 최소 제공 평균을 전체 최소 제공 평균과 비교합니다. 평균 분석

대조군과 비교 - Dunnett 각 효과 최소 제공 평균을 대조 수준의 최소 제공 평균과 비교합니다.

전체 쌍별 비교 - Tukey HSD 중복도를 위한 Tukey HSD 조정을 사용하여 효과 최소 제공 평균의 전체 쌍별 비교를 검정합니다.

전체 쌍별 비교 - 스튜던트 t 중복도 조정 없이 효과 최소 제공 평균의 전체 쌍별 비교를 검정합니다.

이러한 각 옵션을 선택하면 맨 위 영역에 해당 보고서에 대한 상세 정보가 표시된 보고서가 열립니다. 이 정보에는 분위수 또는 임계값이 포함됩니다. 실제 다중 비교 절차의 경우 다중 비교 조정에 사용된 방법이 표시됩니다. "사용자 정의 추정값"을 지정한 경우 보고서에는 지정된 추정값에 따라 변하지 않는 효과 목록 및 이러한 효과가 설정된 수준이 표시됩니다. 별도로 지정하지 않으면 연속형 효과는 평균으로 설정됩니다. 명목형 또는 순서형 효과는 값 순서화에서 첫 번째 수준으로 설정됩니다.

"초기 비교 선택" 목록에서 선택하지 않고 "확인"을 클릭하면 "최소 제공 평균 추정값" 테이블 또는 "사용자 정의 추정값" 테이블이 표시된 "다중 비교" 보고서가 열립니다. "다중 비교"의 빨간색 삼각형 메뉴에서 위에 나열된 모든 옵션을 사용할 수 있습니다. 사용 가능한 보고서 및 옵션은 아래에서 설명합니다.

최소 제공 평균 또는 사용자 정의 추정값 보고서

기본적으로 "다중 비교" 옵션은 시작 창에서 선택한 추정값 유형에 따라 "최소 제공 평균 추정값" 보고서 또는 "사용자 정의 추정값" 보고서를 표시합니다. 이 테이블에는 관심 수준의 각 조합에 대해 평균 추정값과 검정 및 신뢰 구간이 제공됩니다. 구체적으로 다음 항목이 이 테이블에 포함됩니다.

범주형 효과의 수준 보고서의 첫 번째 열은 하나 이상의 관심 효과를 식별합니다. 열의 값은 분석 대상 그룹을 지정합니다.

추정값 각 그룹에 대한 평균 추정값입니다.

표준 오차 각 그룹에 대한 평균의 표준 오차입니다.

DF 평균이 0 인지 여부를 검정하는 데 사용할 자유도입니다.

95% 하한 평균의 신뢰 하한입니다. "모형 적합" 창에서 "유의 수준 설정"을 선택하여 신뢰 수준을 변경할 수 있습니다.

95% 상한 평균의 신뢰 상한입니다.

t 비 유의성 검정의 t 비입니다. 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > t 비를 선택한 경우에만 나타납니다.

Prob>|t| 유의성 검정의 p 값입니다. 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > Prob>|t|를 선택한 경우에만 나타납니다.

산술평균 추정값 ("최소 제곱 평균 추정값" 보고서에만 표시됨) 각 그룹에 대한 산술평균의 추정값입니다.

N ("최소 제곱 평균 추정값" 보고서에만 표시됨) 각 그룹에 대한 평균을 계산하는 데 사용되는 관측값 수입니다.

참고: 테이블을 마우스 오른쪽 버튼으로 클릭하고 "열"을 선택하여 t 비 및 p 값을 구할 수 있습니다.

전체 평균과 비교

이 옵션은 지정된 수준에 대한 평균을 해당 수준에 대한 전체 평균과 비교합니다. 전체 평균과의 차이에 대한 신뢰 구간을 보여 주는 테이블과 결정 한계를 보여 주는 차트가 표시됩니다. 비교에 사용된 방법을 ANOM(평균 분석)이라고 합니다(Nelson et al. 2005). ANOM은 전체 평균에 대한 전체 쌍별 비교의 결합 오차율을 제어하는 다중 비교 절차입니다. Lipid Data.jmp 샘플 데이터 테이블에 기반한 보고서는 그림 3.30에서 확인하십시오.

ANOM은 분석 분석과 비슷해 보일 수 있습니다. 그러나 모든 수준에 대한 전체 평균과 다른 평균을 갖는 수준을 식별한다는 점에서 근본적으로 다릅니다. 반면에 분산 분석은 평균 자체의 차이를 검정합니다.

"전체 평균과 비교" 보고서 맨 위에는 다음 내용이 표시됩니다.

분위수 결정 한계를 구성하는 데 사용된 Nelson h 통계량의 값입니다.

조정 DF 결정 한계를 구성하는 데 사용된 자유도입니다.

평균 평균의 평균입니다. 최소 제곱 추정값의 경우 평균의 평균은 그룹 최소 제곱 평균의 가중 평균입니다. 이 가중 평균은 그룹 최소 제곱 평균이 계산되는 중립 설정에서 전체 평균을 나타냅니다.

명확히 말해 평균 최소 제곱 평균은 $L(X'X)^{-1}L'$ 행렬의 대각 항목에 반비례하는 가중치를 사용한 가중 평균입니다. 여기서 L 은 그룹 최소 제곱 평균을 계산하는 데 사용되는 계수의 행렬입니다. 최소 제곱 평균의 기술적 정의에 대한 내용은 SAS Institute Inc. (2020b).

사용자 정의 추정값에 대한 평균의 평균도 비슷하게 정의됩니다. 그러나 이 경우 L 은 추정값을 정의하는 데 사용되는 계수의 행렬입니다.

조정 다음 중 임계값을 구하는 데 사용되는 방법을 설명합니다.

Nelson 정확 임계값과 p 값을 제공합니다. 가능할 때마다 특히 추정값이 상관관계가 없을 때 사용됩니다.

Nelson-Hsu 사용된 Hsu 요인 분석 근사 (Hsu 1992) 를 기반으로 근사 임계값과 p 값을 제공합니다. 정확 값을 구할 수 없을 때 사용됩니다.

Sidak Nelson 과 Nelson-Hsu 둘 다 실패하면 사용됩니다.

기술 상세 정보는 SAS Institute Inc. (2020b).

"전체 평균과 비교" 보고서 메뉴에서 다음 세 가지 옵션을 사용할 수 있습니다.

전체 평균과 차이

이 보고서에는 그룹 평균과 전체 평균의 각 비교에 대해 다음과 같은 상세 정보가 제공됩니다.

- 비교되는 수준
- 차이 - 추정된 차이
- 표준 오차 - 차이의 표준 오차
- 신뢰 구간의 하한 및 상한
- t 비 - "차이" 및 "표준 오차" 열의 비율

전체 평균과 비교 의사결정 차트

이 의사결정 차트에는 각 그룹의 평균에 점이 표시됩니다. 평균의 평균에 수평선을 그립니다. 결정 상한 및 하한이 그림에 표시됩니다. 그룹 평균에 해당하는 점이 이 한계를 벗어난다고 가정해 보겠습니다. 이는 지정된 유의 수준에서 평균 분석 검정을 기반으로 그룹 평균이 전체 평균과 다르다는 것을 나타냅니다. 유의 수준은 차트 아래에 표시됩니다.

"전체 평균과 비교 의사결정 차트" 보고서 메뉴에는 다음 옵션이 있습니다.

요약 보고서 표시 각 그룹에 대한 추정값, 결정 한계 및 한계 초과를 보여 주는 테이블을 생성합니다.

표시 옵션 차트 표시를 제어하는 몇 가지 옵션을 제공합니다.

조정 p 값 계산

p 값이 포함된 열(Prob>|t|)을 "전체 평균과 비교" 보고서에 추가합니다. 불균형 설계에 대한 정확 임계값 및 p 값을 계산하려면 복소 적분이 필요하며 계산이 어려울 수 있습니다. 이러한 분위 수 계산이 실패하면 Sidak 분위수가 계산되지만 p 값은 사용할 수 없습니다.

전체 평균과 비교의 예

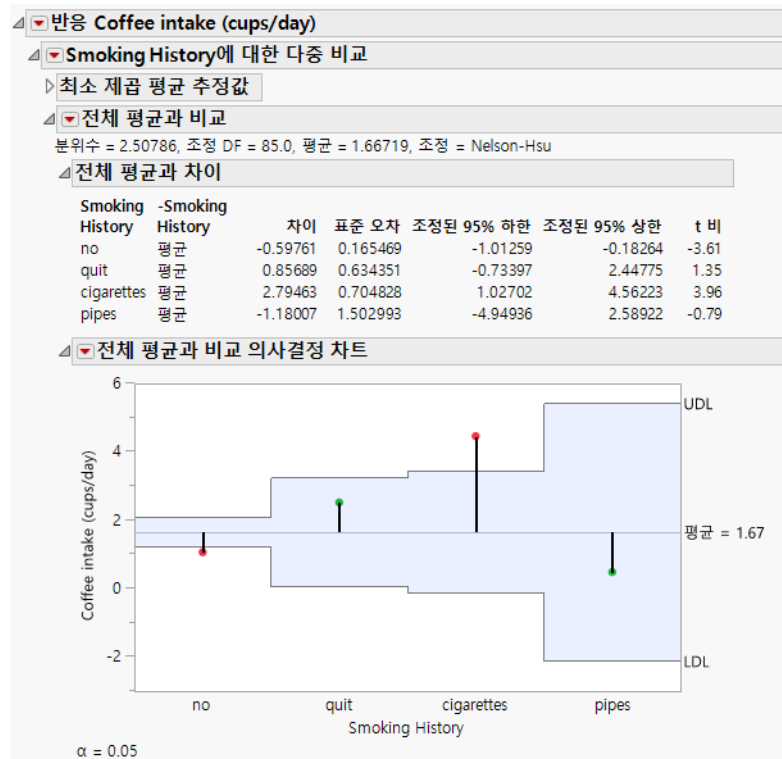
Lipid Data.jmp 샘플 데이터 테이블을 예로 들어보겠습니다. 음주 (Alcohol use) 및 심장 기록 (Heart History) 을 제어하면서 네 가지 흡연 이력 (Smoking History) 범주 중 평균 하루 커피 섭취량 (Coffee intake (cups/day)) 이 전체 평균 커피 섭취량과 다르다는 점에서 특이한 것이 있는지 확

인하려고 합니다 . Coffee intake (cups/day) 를 반응으로 사용하고 Smoking History, Alcohol Use 및 Heart History 를 모형 효과로 사용하여 모형을 지정합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Lipid Data.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Coffee intake (cups/day) 를 선택하고 **Y** 를 클릭합니다 .
4. Smoking History, Alcohol Use 및 Heart History 를 선택하고 **추가**를 클릭합니다 .
5. **실행**을 클릭합니다 .
6. " 반응 Coffee intake (cups/day)" 옆의 빨간색 삼각형을 클릭하고 **추정값 > 다중 비교**를 선택합니다 .
7. " 효과 선택 " 목록에서 Smoking History 를 선택합니다 .
8. " 초기 비교 선택 " 목록에서 **전체 평균과 비교 - ANOM** 을 선택합니다 .
9. **확인**을 클릭합니다 .

그림 3.30에 표시된 결과는 비흡연자와 cigarette 흡연자의 최소 제곱 평균이 커피 섭취량 측면에서 전체 평균과 유의하게 다르다는 것을 나타냅니다.

그림 3.30 평가를 위한 전체 평균과 비교



대조군과 비교

"대조군과 비교 - Dunnett"을 선택하면 대조군을 지정할 수 있는 창이 열립니다. "최소 제곱 평균 추정값"을 선택한 경우 선택한 효과의 모든 수준으로 목록이 구성됩니다. "사용자 정의 추정값"을 선택한 경우에는 지정한 효과 수준의 조합으로 목록이 구성됩니다.

대조군을 선택한 후 "확인"을 클릭하면 "최소 제곱 적합" 보고서에 "대조군과 비교" 보고서가 나타납니다. 이 옵션은 지정된 설정에 대한 평균을 대조군 평균과 비교합니다. 대조군과의 차이에 대한 신뢰 구간을 보여 주는 테이블과 결정 한계를 보여 주는 차트가 표시됩니다. Dunnett 방법이 비교에 사용됩니다. Dunnett 방법은 모든 비교에 대해 오차율을 제어하는 다중 비교 절차입니다(Hsu 1996, Westfall et al. 2011).

p 값과 신뢰 구간을 정확하게 계산할 수 없는 경우 Hsu 요인 분석 근사가 사용됩니다(Hsu 1992). 불균형 설계에 대한 정확 임계값 및 p 값을 계산하려면 복소 적분이 필요하며 계산량이 많을 수 있습니다. 이러한 분위수 계산이 실패하면 Sidak 분위수가 계산됩니다.

지정된 추정값에 대해 변하지 않는 효과 목록 외에도 다음 내용이 "대조군과 비교" 보고서 맨 위에 표시됩니다.

분위수 Dunnett 검정의 임계값입니다.

조정 DF 신뢰 구간을 구성하는 데 사용된 자유도입니다.

대조군 대조군을 정의하는 설정입니다. 이 설정은 단일 효과를 선택한 경우 단일 수준이며, 둘 이상의 효과에 대한 사용자 정의 조합을 지정한 경우에는 수준 조합입니다.

조정 다음 중 임계값을 구하는 데 사용되는 방법입니다.

Dunnett 정확 임계값과 p 값을 제공합니다. 가능할 때마다 특히 추정값이 상관관계가 없을 때 사용됩니다.

Dunnett-Hsu Hsu 요인 분석 근사(Hsu 1992)를 기반으로 근사 임계값과 p 값을 제공합니다. 정확 값을 구할 수 없을 때 사용됩니다.

Sidak Dunnett 과 Dunnett-Hsu 둘 다 실패하면 사용됩니다.

기술 상세 정보는 SAS Institute Inc. (2020b).

"대조군과 비교" 보고서 메뉴에서 다음 세 가지 옵션을 사용할 수 있습니다.

대조군과 차이

이 보고서에는 그룹 평균과 대조군 평균의 각 비교에 대해 다음과 같은 상세 정보가 제공됩니다.

- 비교되는 수준
- 차이 - 추정된 차이
- 표준 오차 - 차이의 표준 오차
- 신뢰 구간의 하한 및 상한
- t 비 - "차이" 및 "표준 오차" 열의 비율

대조군과 비교 의사결정 차트

이 의사결정 차트에는 대조군과 비교되는 각 그룹의 평균에 점이 표시됩니다. 수평선은 대조군의 평균을 나타냅니다. 결정 상한 및 하한이 그림에 표시됩니다. 이 한계를 벗어나는 점은 지정된 유의 수준에서 Dunnett 검정을 기반으로 평균이 대조군 평균과 다른 그룹에 해당합니다. 해당 수준은 차트 아래에 표시됩니다.

"대조군과 비교 의사결정 차트" 보고서 메뉴에는 다음 옵션이 있습니다.

요약 보고서 표시 각 그룹에 대한 추정값, 결정 한계 및 한계 초과를 보여 주는 테이블을 생성합니다.

표시 옵션 차트 표시를 제어하는 몇 가지 옵션을 제공합니다.

조정 p 값 계산

p 값이 포함된 열(Prob>|t|)을 "대조군과 비교" 보고서에 추가합니다. 불균형 설계에 대한 정확한 임계값 및 p 값을 계산하려면 복소 적분이 필요하며 계산이 어려울 수 있습니다. 이러한 분위수 계산이 실패하면 Sidak 분위수가 계산되지만 p 값은 사용할 수 없습니다.

전체 쌍별 비교

"전체 쌍별 비교" 옵션은 "Tukey HSD 전체 쌍별 비교" 또는 "스튜던트 t 전체 쌍별 비교" 보고서를 표시합니다(Hsu 1996, Westfall et al. 2011). Tukey HSD 비교는 유의 수준이 전체 쌍별 비교에 공동으로 적용되도록 구성됩니다. 반면에 스튜던트 t 비교는 유의 수준이 각 개별 비교에 적용됩니다. 스튜던트 t 검정을 사용하여 여러 쌍별 비교를 수행할 때 비교 중 하나가 차이를 잘못 나타낼 위험이 명시된 유의 수준을 훨씬 초과할 수 있습니다.

"Tukey HSD 전체 쌍별 비교" 보고서 맨 위에는 다음 내용이 표시됩니다.

분위수 검정의 임계값입니다. Tukey HSD의 경우 분위수는 $q/(\sqrt{2})$ 입니다. 여기서 q 는 스튜던트화 범위 통계량의 적절한 백분율입니다.

조정 DF 신뢰 구간을 구성하는 데 사용된 자유도입니다.

조정 다음 중 임계값을 구하는 데 사용되는 방법을 설명합니다.

Tukey 정확 임계값과 p 값을 제공합니다. 평균이 상관관계가 없고 분산이 같을 때 또는 분산 균형 설계일 때 사용됩니다.

Tukey-Kramer 근사 임계값과 p 값을 제공합니다. 정확 값을 구할 수 없을 때 사용됩니다.

기술 상세 정보는 SAS Institute Inc. (2020b).

"스튜던트 t 전체 쌍별 비교" 보고서 맨 위에는 분위수(t 검정에 대한 임계값) 및 DF(t 검정에 사용된 자유도)가 표시됩니다.

전체 쌍별 차이 보고서

Tukey HSD 와 스튜던트 t 는 둘 다 전체 수준 쌍을 비교합니다. "전체 쌍별 차이" 보고서에는 각 쌍별 비교에 대해 다음과 같은 정보가 제공됩니다.

- 비교되는 수준
- 차이 - 추정된 평균 차이
- 표준 오차 - 차이의 표준 오차
- t 비 - 차이가 0 인지 여부에 대한 검정의 t 비
- $\text{Prob} > |t|$ - 검정의 p 값
- 평균 차이에 대한 신뢰 구간의 하한 및 상한

전체 쌍별 비교 산점도

이 그림은 차이 그림(diffogram) 또는 평균-평균 산점도라고도 하며, 전체 평균 쌍별 차이에 대한 신뢰 구간을 표시합니다. 예는 그림 3.32에서 확인하십시오. 색상은 유의한 차이를 나타냅니다.

이 그림에는 참조선이 대각선에서 위쪽으로 기울어진 선으로 표시됩니다. 이 선은 두 평균이 동일한 점을 나타냅니다. 각 선분은 쌍별 비교에 대한 신뢰 구간에 해당합니다. 선분에 표시된 점의 좌표는 해당 그룹에 대한 평균입니다. 이러한 점 중 하나를 커서로 가리키면 비교되는 그룹을 식별하고 추정된 차이를 보여 주는 툴팁이 표시됩니다. 선분이 대각선에서 선과 교차하는 경우 평균이 동일할 수 있고 비교가 유의하지 않습니다.

쌍별 비교 산점도에는 다음 옵션이 있습니다.

참조선 표시 산점도의 점에 대한 참조 격자선을 표시합니다. 산점도에 점이 많이 있는 경우 권장되지 않습니다. 점이 많은 경우에는 점을 커서로 가리켜 툴팁 라벨을 보는 것이 더 좋습니다.

전체 쌍별 차이 연결 문자

유의한 비교와 유의하지 않은 비교를 연결 문자로 나타내는 보고서를 표시하려면 이 옵션을 사용합니다. 동일한 문자로 연결되지 않은 수준에는 유의한 차이가 있습니다. 동일한 문자로 연결된 수준에는 유의한 차이가 없습니다.

전체 쌍별 차이 연결 문자 테이블 저장

이 옵션은 효과 수준, 연결 문자, 최소 제곱 평균, 표준 오차 및 신뢰 구간을 포함하는 열이 있는 데이터 테이블을 생성합니다. 데이터 테이블에는 신뢰 구간이 위에 겹쳐진 최소 제곱 평균의 막대 차트(색상으로 구분됨)를 생성하는 "막대 차트" 스크립트가 포함되어 있습니다. 수준은 최소 제곱 평균이 큰 것부터 순서대로 배열됩니다.

동등성 검정

하나 이상의 동등성 검정을 수행하려면 이 옵션을 사용합니다. 동등성 검정은 관심 있는 실제적 차이를 감지하려는 경우에 유용합니다. 지정한 값보다 작은 차이는 실제적으로 동등한 것으로

간주되도록 그룹 평균에 대한 임계 차이를 지정해야 합니다. 즉, 두 그룹 평균의 차이가 이 값보다 작거나 같으면 동등하다고 간주할 수 있습니다.

이 값을 지정하면 "동등성 검정" 보고서가 나타납니다. 지정한 경계는 보고서 맨 위에 표시됩니다. 이 보고서는 동등성 검정을 제공하는 테이블과 이를 표시하는 산점도로 구성됩니다. 동등성 검정 및 신뢰 구간은 스튜던트 t 임계값을 기반으로 합니다.

참고 : 동등성 검정은 스튜던트 t 방법에만 사용할 수 있습니다.

동등성 TOST 검정

TOST(Two One-Sided Tests: 두 개의 단측 검정) 방법은 평균에 실제적인 차이가 있는지 검정하는데 사용됩니다(Schirmann 1987). 실제 차이가 임계값을 초과한다는 귀무가설에 대해 두 개의 단측 합동 분산 t -검정이 구성됩니다. 두 검정이 모두 기각되면 평균의 차이가 통계적으로 두 임계값 중 어느 하나도 초과하지 않습니다. 따라서 그룹은 실제로 동등하다고 간주됩니다. 검정이 하나만 기각되거나 둘 다 기각되지 않으면 그룹이 실제로 동등하지 않을 수 있습니다.

"동등성 TOST 검정" 보고서에는 각 비교에 대해 다음과 같은 정보가 제공됩니다.

- 차이 - 추정된 평균 차이
- 하한 t 비, 상한 t 비 - 두 개의 단측 합동 분산 유의성 검정에 대한 하한 및 상한 t 비
- 하한 p 값, 상한 p 값 - 하한 및 상한 t 비에 해당하는 p 값
- 최대 p 값 - 하한 및 상한 p 값의 최대값
- 평균 차이에 대한 $1-2\alpha$ 신뢰 구간의 하한 및 상한

참고 : 동등성 TOST 검정은 스튜던트 t 방법에만 사용할 수 있습니다.

동등성 검정 산점도

이 산점도에서는 동등성 검정에 따라 실제로 동등한 평균과 동등하지 않은 평균을 나타내기 위해 색상을 사용합니다. 이 그림을 차이 그림(diffogram) 또는 평균-평균 산점도라고도 합니다.

이 그림에는 대각선 위의 실선 참조선 및 음영 참조 대역이 표시됩니다. 대역 너비는 실제 차이의 두 배입니다. 선분에 있는 점의 좌표는 해당 그룹에 대한 평균입니다. 각 선분이 쌍별 비교를 위한 $1-2\alpha$ 신뢰 구간에 해당하는 세 번째 축이 대각선에 내포되어 있습니다. 이러한 점 중 하나를 커서로 가리키면 비교되는 그룹 및 추정된 차이를 나타내는 툴팁이 표시됩니다. 선분이 대각선 대역 내에 완전히 포함되면 평균이 실제로 동등하다고 할 수 있습니다.

참고 : 동등성 검정 산점도는 스튜던트 t 방법에만 사용할 수 있습니다.

동등성 검정 산점도에는 다음 옵션이 있습니다.

참조선 표시 산점도의 점에 대한 참조선을 표시합니다. 산점도에 점이 많이 있는 경우 권장되지 않습니다. 점이 많은 경우에는 점을 커서로 가리켜 툴팁 라벨을 보는 것이 더 좋습니다.

제거

이 옵션은 "스튜던트 t 전체 쌍별 비교" 보고서에서 "동등성 검정" 보고서를 제거합니다.

Tukey HSD 전체 쌍별 비교의 예

Lipid Data.jmp 샘플 데이터 테이블을 예로 들어보겠습니다. 두 연령 (25 세, 35 세) 과 평균 키를 대상으로 성별 및 비흡연자와 과거 흡연자 (Smoking History 는 각각 no 와 quit) 의 Cholesterol 차이에 관심이 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Lipid Data.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. Cholesterol 을 선택하고 **Y** 를 클릭합니다.
4. Gender, Age, Height 및 Smoking History 를 선택하고 **추가**를 클릭합니다.
5. **실행**을 클릭합니다.
6. "반응 Cholesterol" 옆의 빨간색 삼각형을 클릭하고 **추정값 > 다중 비교**를 선택합니다.
7. "추정값 유형" 목록에서 **사용자 정의 추정값**을 클릭합니다.
8. "Gender 수준 선택" 목록에서 **female**(기본적으로 선택되어 있어야 함)과 **male**을 선택합니다.
9. "Smoking History 수준 선택" 목록에서 **no** 와 **quit** 를 선택합니다.
10. Age 목록에서 처음 두 행에 25 와 35 를 입력합니다.
Height 목록에는 값을 입력하지 마십시오. Height 값이 지정되지 않았으므로 Height 열의 평균 값이 다중 비교 보고서에 사용됩니다.
11. **추정값 추가**를 클릭합니다.
지정한 수준의 가능한 모든 조합이 "비교를 위한 추정값" 보고서에 나타납니다.
12. "초기 비교 선택" 목록에서 **전체 쌍별 비교 - Tukey HSD** 를 선택합니다.
창이 그림 3.31 에 표시된 것과 같이 완료되었는지 확인합니다.

그림 3.31 항목이 채워진 사용자 정의 추정값 창

추정값 유형
☐ 최소 제곱 평균 추정값
☒ 사용자 정의 추정값

Gender 수준 선택
female
male

Smoking History 수준 선택
no
quit
cigarettes
pipes

Age	Height
25	.
35	.
.	.
.	.
.	.
.	.

필요한 경우 요인 설정을 선택하고 [추정값 추가] 버튼을 클릭하여 사용자 정의 추정값을 생성하십시오.

추정값 추가

비교를 위한 추정값

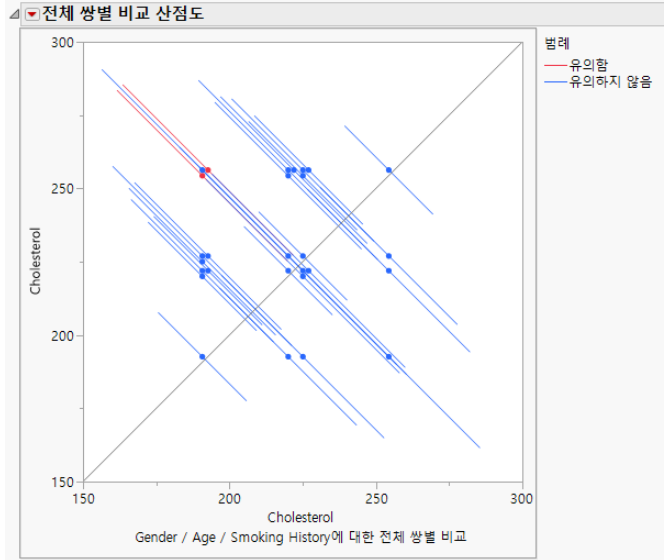
Gender	Age	Height	Smoking History
female	25	69.327632	no
female	25	69.327632	quit
female	35	69.327632	no
female	35	69.327632	quit
male	25	69.327632	no
male	25	69.327632	quit
male	35	69.327632	no
male	35	69.327632	quit

초기 비교 선택
☐ 전체 평균과 비교 - ANOM
☐ 대조군과 비교 - Dunnett
☒ 전체 쌍별 비교 - Tukey HSD
☐ 전체 쌍별 비교 - 스튜던트 t

확인 **취소** **도움말**

13. 확인을 클릭합니다 .

" 전체 쌍별 차이 " 보고서에 따르면 28 개의 쌍별 비교 중 2 개가 유의합니다 . 그림 3.32 에 표시된 " 전체 쌍별 비교 산점도 "에서는 이러한 비교에 대한 신뢰 구간이 빨간색으로 표시됩니다 . 점을 커서로 가리키면 해당 점이 나타내는 쌍별 비교를 확인할 수 있습니다 . 툴팁에는 비교 대상인 두 수준 간의 차이도 포함됩니다 . 그림 3.32 의 빨간색 점 두 개는 여성과 남성 둘 다에 대해 35 세 과거 흡연자와 25 세 비흡연자를 비교하는 점을 나타냅니다 .

그림 3.32 사용자 정의 비교를 위한 전체 쌍별 비교 산점도

기울기 비교

"기울기 비교" 옵션은 명목형 항과 연속형 항이 각각 하나이고 고정 효과에 대한 교호작용 효과가 있을 때 나타납니다. 이 옵션은 ANCOVA(공분산 분석) 모형에서 기울기를 비교할 수 있는 보고서를 생성합니다. 이 보고서에서는 교호작용 효과의 각 수준에 대한 기울기를 전체 기울기와 비교합니다. 비교에는 ANOM(평균 분석)과 전체 평균이 함께 사용됩니다. ANOM(평균 분석) 보고서에 대한 자세한 내용은 "[전체 평균과 비교](#)"(121 페이지)에서 확인하십시오.

전체 평균 기울기는 기울기의 가중 평균이며 이때 가중치는 기울기 추정값의 분산에 반비례합니다. 이러한 분산은 "전체 평균 기울기와의 차이" 테이블에서 "표준 오차" 열의 제곱 값입니다.

결합 요인 검정

"결합 요인 검정" 옵션은 교호작용 효과가 있을 때 나타납니다. JMP는 모형의 각 주효과에 대해 해당 주효과를 포함하는 항의 모든 계수가 0인지 여부를 확인하는 결합 검정을 생성합니다. 이 검정은 모형에 있는 다른 모든 효과를 조건으로 합니다. 구체적으로 말해 결합 검정은 제한된 모형의 일반 선형 가설 검정입니다. 이 모형에서는 지정된 효과에 해당하는 모든 모수와 이 효과를 포함하는 교호작용이 0으로 설정됩니다.

결합 요인 검정 보고서의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. **weight** 를 선택하고 **Y** 를 클릭합니다.

4. **차수** 상자에 2 가 표시되어 있는지 확인합니다 .
5. age, sex 및 height 를 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다 .
6. **실행**을 클릭합니다 .
7. " 반응 weight" 의 빨간색 삼각형을 클릭하고 **추정값 > 결합 요인 검정**을 선택합니다 .

그림 3.33 결합 요인 검정 보고서

결합 요인 검정				
항	DF	제곱합	F 비	Prob > F
age	15	6116.2127	2.8488	0.0139*
sex	7	2113.1080	2.1091	0.0879
height	7	9217.8156	9.2002	<.0001*

age에 대한 검정의 자유도는 15입니다. 이 검정에는 age에 대한 5개 모수, age*sex에 대한 5개 모수, height*age에 대한 5개 모수가 포함됩니다. 이 검정의 귀무가설은 15개 모수가 모두 0이라는 것입니다.

역추정 예측

역추정 예측은 반응 변수의 값이 주어진 경우 통계 모형을 사용하여 설명 변수의 값을 추론할 때 발생합니다. 역추정 예측은 **보정**이라고도 합니다.

"추정값" 메뉴에서 "역추정 예측"을 선택하면 지정된 반응 값에 해당하는 독립 변수 X의 값을 추정할 수 있습니다(그림 3.37). 또한 모형의 다른 설명 변수에 대한 값을 지정할 수 있습니다(그림 3.37). 역추정 예측 계산은 지정된 반응 값에 해당하는 X 값에 대한 신뢰 한계를 제공합니다. 반응 값을 평균 반응 또는 간단히 개별 반응으로 지정할 수 있습니다. 예는 "**역추정 예측의 예**"(132 페이지)에서 확인하십시오.

여러 설명 변수 분석

모형에 여러 설명 변수가 포함된 경우 다른 변수의 지정된 값에 대해 X 값을 예측할 수 있습니다. 심장 박동 수가 60일 때 산소 섭취량이 50이 되는 달리기 시간을 예측하고 싶을 수 있습니다. 남성과 여성 둘 다에 대해 별도의 역추정 예측을 원할 수도 있습니다. 역추정 예측 옵션을 사용하여 이러한 요구 사항을 지정합니다.

역추정 예측 창의 왼쪽에는 설명 변수 목록이 표시됩니다. 예는 그림 3.37에서 확인하십시오. 각 연속형 변수는 처음에 평균으로 설정됩니다. 각 명목형 또는 순서형 변수는 값 순서화에서 가장 낮은 수준으로 설정됩니다. 예측할 변수의 값을 제거하고 결측으로 설정해야 합니다. 또한 역추정 예측을 포함할 다른 변수의 값을 지정해야 합니다(기본 설정과 다른 경우). 창의 오른쪽에 있는 목록에서 관심 있는 하나 이상의 반응 값을 제공할 수 있습니다. 예는 "**여러 모형 효과를 사용하여 단일 X 값 예측의 예**"(134 페이지)에서 확인하십시오.

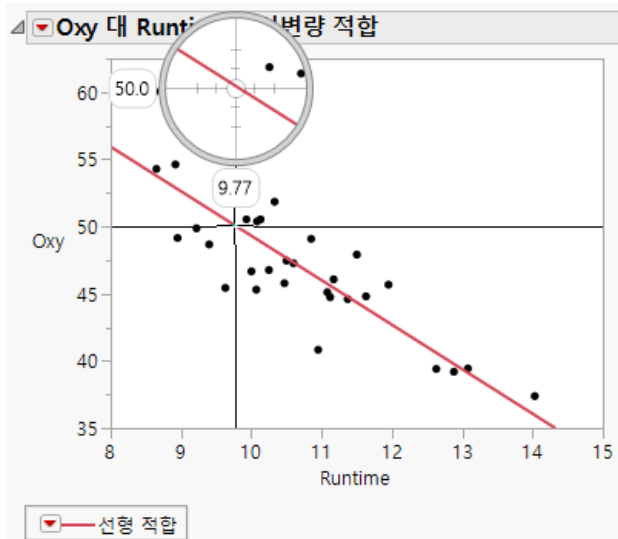
참고 : 역추정 예측의 신뢰 한계는 때때로 단측 구간이 되거나 무한 구간도 될 수 있습니다. 기술 상세 정보는 "통계 상세 정보" 부록의 "**신뢰 한계를 사용한 역추정 예측**"(519 페이지)에서 확인하십시오.

역추정 예측의 예

이 예에서는 Runtime 에서 산소 섭취량을 예측하는 회귀 모형을 적합시킵니다. 그런 다음 지정된 산소 섭취량 값을 도출하는 Runtime 값을 추정합니다. X 변수 (Runtime) 가 하나뿐이므로 X로 Y 적합 플랫폼을 사용하여 역추정 예측값의 시각적 근사를 구하는 것으로 시작합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Fitness.jmp 를 엽니다.
2. **분석 > X 로 Y 적합**을 선택합니다.
3. Oxy 를 선택하고 **Y, 반응**을 클릭합니다.
4. Runtime 을 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.
6. "Oxy 대 Runtime 의 이변량 적합" 의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다.
아래 설명된 대로 십자기호 도구를 사용하여 평균 Oxy 값이 50 이 되는 Runtime 의 근사값을 산출합니다.
7. **도구 > 십자기호**를 선택합니다.
8. 예측선을 클릭한 후 십자기호 도구를 드래그하여 Oxy = 50 에 대한 역추정 예측을 찾습니다.
그림 3.34 에서는 Runtime 이 약 9.77 일 때 Oxy 가 약 50 이 되는 역추정 예측을 제공합니다.

그림 3.34 Fitness.jmp 에 대한 이변량 적합



신뢰 구간과 함께 Runtime 에 대한 정확 예측을 구하려면 모형 적합 시작 창을 사용합니다.

1. Fitness.jmp 샘플 데이터 테이블에서 **분석 > 모형 적합**을 선택합니다.
2. Oxy 를 선택하고 **Y** 를 클릭합니다.
3. Runtime 을 선택하고 **추가**를 클릭합니다.

4. **실행**을 클릭합니다.
5. "반응 Oxy"의 빨간색 삼각형을 클릭하고 **추정값 > 역추정 예측**을 선택합니다.
6. 그림 3.35에 표시된 대로 Oxy 값 4개를 입력합니다.
7. **확인**을 클릭합니다.

그림 3.35 완료된 역추정 예측 규격 창

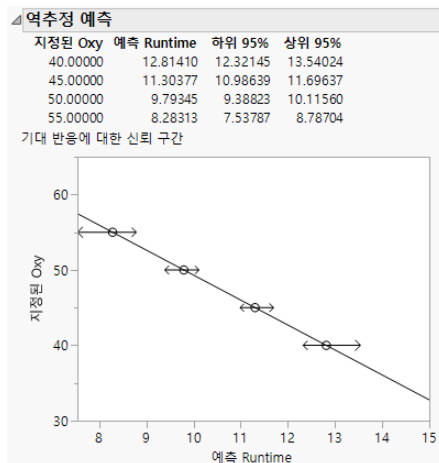
역추정 예측할 하나 이상의 y 값을 지정하십시오.

Runtime (예측할)	신뢰 수준	Oxy
	0.95	40
		45
		50
		55
		.
		.
		.
		.

☐ 기대 반응이 아닌 개별값에 대한 신뢰 구간

"역추정 예측" 보고서에는 지정된 각 Oxy 값에 해당하는 예측 Runtime 값이 제공됩니다. 또한 평균 반응을 얻는 것과 관련하여 Runtime 값에 대한 95% 신뢰 상한 및 하한도 보고서에 표시됩니다.

그림 3.36 역추정 예측 보고서



Oxy 값이 50이 되는 정확 예측 Runtime은 9.7935입니다. 이 값은 그림 3.34에 표시된 이변량 적합 보고서의 Runtime 근사값 9.77에 가깝습니다. "역추정 예측" 보고서에는 Oxy와 Runtime 간의 선형 관계 및 신뢰 구간을 보여 주는 그림도 제공됩니다.

여러 모형 효과를 사용하여 단일 X 값 예측의 예

이 예에서는 RstPulse 가 60 일 때 산소 섭취량이 50 이 되는 Runtime 을 예측합니다. 남성과 여성 둘 다에 대해 Runtime 이 예측됩니다.

1. Fitness.jmp 샘플 데이터 테이블에서 **분석 > 모형 적합**을 선택합니다.
2. Oxy 를 선택하고 **Y** 를 클릭합니다.
3. Sex, Runtime 및 RstPulse 를 선택한 후 **추가**를 선택합니다.
4. **실행**을 클릭합니다.
5. "반응 Oxy" 의 빨간색 삼각형을 클릭하고 **추정값 > 역추정 예측**을 선택합니다.
6. Runtime 값을 예측해야 하므로 해당 값을 삭제합니다.
7. Sex 옆의 **모두**를 선택하여 Sex 의 모든 수준에 대한 Runtime 을 추정합니다.
8. RstPulse 의 평균을 60 으로 바꿉니다.
9. 그림 3.37 에 표시된 대로 Oxy 값 50 을 입력합니다.
10. **확인**을 클릭합니다.

그림 3.37 다중 회귀 모형에 대한 역추정 예측 규칙

역추정 예측할 항목 제외하고 항목 지정하십시오.
예측할 값을 비어 있거나 결측된 상태로 두십시오.
역추정 예측할 하나 이상의 y 값을 지정하십시오.

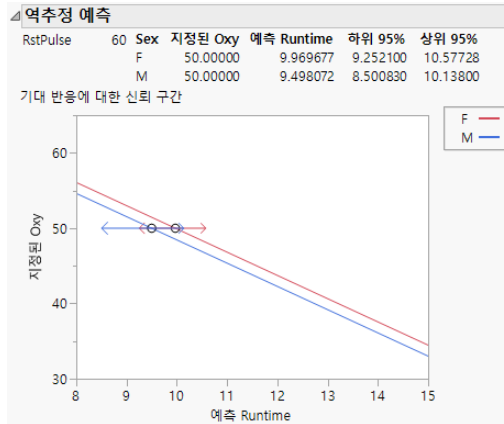
Sex	F ▾	☒ 모두	신뢰 수준	0.95	Oxy	50
Runtime						.
RstPulse	60					.
						.
						.
						.
						.

☐ 기대 반응이 아닌 개별값에 대한 신뢰 구간

확인 **취소** **도움말**

그림 3.38에 표시된 보고서에는 여성과 남성 둘 다에 대한 예측 Runtime 값이 제공됩니다. 또한 평균 반응 50을 제공하는 Runtime 값에 대한 95% 신뢰 구간도 보고서에 포함됩니다.

그림 3.38 다중 회귀 모형에 대한 역추정 예측 보고서



이 그림에서는 RstPulse가 60일 때 여성 및 남성에게 대한 선형 적합을 보여 줍니다. 두 신뢰 구간은 각각 빨간색과 파란색으로 표시됩니다.

Cox 혼합물

참고 : 이 옵션은 혼합물 모형에만 사용할 수 있습니다.

표준 최소 제곱은 Scheffe(1958)에서 제안된 파라미터화를 사용하여 혼합물 모형을 적합시킵니다. 이 모형의 모수를 사용하여 혼합물 성분의 효과를 쉽게 판단할 수 없습니다. Cox 혼합물 모형은 재파라미터화되고 제약된 버전의 Scheffe 모형입니다. 모수 추정값을 사용하여 요인 효과 및 설계 공간의 참조점을 기준으로 한 반응 표면 모양을 도출할 수 있습니다. 자세한 내용은 Cornell 연구 자료(1990)에서 확인하십시오.

"Cox 혼합물" 옵션은 참조 혼합물을 입력하는 창을 엽니다. 참조 혼합물에 대해 합이 1이 되지 않는 성분을 입력하면 합이 1이 되도록 성분의 크기가 비례적으로 조정됩니다. 재조정된 혼합물은 보고서에 "참조 혼합물"로 표시됩니다. 성분 효과도 보고서에 나타납니다. Cox 성분 효과는 요인이 Cox 효과 방향을 따라 최소값에서 최대값으로 이동할 때 예측 반응의 차이입니다.

Cox 혼합물의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Five Factor Mixture.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Y1 을 선택하고 **Y** 를 클릭합니다 .
4. X1 ~ X5 를 선택합니다 .
5. **매크로 > 혼합물 반응 표면**을 선택합니다 .
6. **실행**을 클릭합니다 .
7. " 반응 Y1" 의 빨간색 삼각형을 클릭하고 **추정값 > Cox 혼합물**을 선택합니다 .

8. " 참조 혼합물 " 규격 창에서 다음 값을 입력합니다 .

- X1 옆에 0.15 입력
- X2 옆에 0.25 입력
- X3 옆에 0.4 입력
- X4 옆에 0.1 입력
- X5 옆에 0.1 입력

9. **확인**을 클릭합니다 .

그림 3.39 Cox 혼합물

Cox 참조 혼합 모형				
Cox 모수 추정값				Cox 참조 혼합
모수	추정값	표준 오차	t 비	Prob> t
절편	2.071	1.2541	1.651	0.1035
X1	3.120	7.7538	0.402	0.6887
X2	-5.828	7.0104	-0.831	0.4089
X3	-0.696	3.5822	-0.194	0.8465
X4	3.935	12.6619	0.311	0.7570
X5	8.740	12.5405	0.697	0.4883
X1^2	-15.897	16.3627	-0.972	0.3349
X2^2	17.877	13.5932	1.315	0.1931
X3^2	5.234	8.2498	0.634	0.5280
X4^2	-91.239	159.9347	-0.570	0.5703
X5^2	-145.693	158.3991	-0.920	0.3611
X1*X2	-2.938	24.9012	-0.118	0.9065
X1*X3	-7.617	19.1697	-0.397	0.6924
X2*X3	-16.250	18.7489	-0.867	0.3893
X1*X4	53.690	50.5711	1.062	0.2923
X2*X4	-17.702	49.9082	-0.355	0.7240
X3*X4	-7.808	44.0229	-0.177	0.8598
X1*X5	31.813	50.6186	0.628	0.5319
X2*X5	-2.279	49.3761	-0.046	0.9633
X3*X5	17.983	43.5145	0.413	0.6808
X4*X5	177.431	97.3862	1.822	0.0731
구성 요소 효과				
성분	Cox 효과	구성 요소 범위		
X1	0.80746	0.2200000		
X2	-2.33103	0.3000000		
X3	-0.48747	0.4200000		
X4	0.65585	0.1500000		
X5	1.45663	0.1500000		

보고서에는 Cox 혼합물 모형에 대한 모수 추정값과 표준 오차 및 가설 검정이 표시됩니다. 참조 혼합물이 오른쪽에 나타납니다. 아래쪽에는 성분 효과와 성분 범위가 나타납니다.

모수 검정력

통계적 검정의 검정력은 실제로 차이가 있을 때 검정이 유의할 확률입니다. 검정력은 연구에서 실제 효과가 유의하다고 선언할 가능성을 나타냅니다. "모수 검정력" 옵션은 후향적 검정력 분석을 처리합니다.

참고: 필요한 차이를 감지할 수 있도록 충분한 수의 관측값이 연구에 포함되게 하려면 실험을 설계할 때 검정력에 대한 정보를 사용하십시오. 이 유형의 분석을 **전향적 검정력 분석**이라고 합니다. DOE 플랫폼을 사용하여 연구를 설계하는 것이 좋습니다. DOE > 표본 크기 및 검정력과 DOE > 설계 평가 둘 다 전향적 검정력 분석에 유용합니다. 표준 최소 제곱을 사용한 전향적 검정력 분석의 예는 "**전향적 검정력 분석**"(198 페이지)에서 확인하십시오.

차이를 감지하기 위한 검정력은 다음 요인의 영향을 받습니다.

- 표본 크기
- 알 수 없는 잔차 오차 분산
- 검정의 유의 수준
- 감지할 효과의 크기

이미 연구를 수행하고 데이터를 분석하여 관심 효과가 유의하지 않음을 확인했다고 가정해 보겠습니다. 감지할 수 있는 차이의 크기 또는 수행한 검정의 검정력에 관심이 있을 수 있습니다. 아니면 주어진 크기의 차이를 높은 확률로 감지하기 위해 필요한 관측값 수를 알고 싶을 수도 있습니다.

"모수 검정력" 옵션은 "모수 추정값" 보고서에 후향적 검정력 분석과 관련된 세 개의 값 열을 삽입합니다. 최소 유의값(LSV0.05), 최소 유의 개수(LSN0.05) 및 검정력 계산(조정 검정력0.05)이 제공됩니다.

모수 검정력 계산은 관측된 표본과 변동 프로파일이 동일한 새 표본에 적용됩니다.

주의: LSV0.05, LSN 및 조정 검정력 0.05에서 제공하는 결과를 전향적 검정력 분석에 사용하면 안 됩니다. 이 값은 향후 연구에 내재된 불확실성을 반영하지 않습니다.

- LSV0.05는 최소 유의값입니다. 이 값은 유의 수준 0.05에서 이 검정을 유의하게 만드는 추정값의 가장 작은 절대값입니다. 더 구체적으로 설계에 대한 관측값 수, 평균 제곱 오차, 제곱합 및 교차곱 행렬은 변경되지 않는다고 가정해 보겠습니다. 그런 다음 추정값의 절대값이 LSV0.05보다 작으면 $\text{Prob}>|t|$ 값이 0.05를 초과합니다. 자세한 내용은 "**LSV(최소 유의값)**"(196 페이지)에서 확인하십시오.
- LSN은 최소 유의 개수입니다. 이 값은 유의 수준 0.05에서 이 검정을 유의하게 만드는 관측값 수입니다. 구체적으로 설계에 대한 모수 추정값, 평균 제곱 오차, 제곱합 및 교차곱 행렬은 변경되지 않는다고 가정해 보겠습니다. 그런 다음 관측값 수가 LSN보다 작으면 $\text{Prob}>|t|$ 값이 0.05를 초과합니다. 자세한 내용은 "**LSN(최소 유의 개수)**"(195 페이지)에서 확인하십시오.
- 조정 검정력 0.05는 조정 검정력 값입니다. 이 값은 이 검정이 유의할 확률의 추정값입니다. 현재 연구의 표본 값은 일반적으로 검정력 계산에 사용되는 모수 값으로 대체됩니다. 조정 검정력 계산은 표본 추정값을 비중심성 모수에 대한 계산식에 직접 대체하여 발생하는 편향을 조정합니다 (Wright and O'Brien 1988). 자세한 내용은 "**조정 검정력 및 신뢰 구간**"(196 페이지)에서 확인하십시오.

LSV, LSN 및 조정 검정력은 검정의 민감도를 평가하는 데 유용합니다. 이러한 후향적 계산은 이해를 위한 교육용 도구로도 사용됩니다. 그러나 값을 해석할 때 주의해야 합니다(Hoenig and Heisey 2001).

LSV, LSN 및 조정 검정력에 대한 자세한 내용은 "**검정력 분석**"(193 페이지)에서 확인하십시오. 후향적 분석의 예는 "**후향적 검정력 분석의 예**"(197 페이지)에서 확인하십시오.

추정값 상관계수

"추정값" 메뉴의 "추정값 상관계수" 명령은 모수 추정값에 대한 상관 행렬을 계산합니다. 이러한 상관은 공선성이 있는지 여부를 나타냅니다.

이 행렬의 생성에 대한 이해를 위해 일반적인 최소 제곱 회귀 공식을 고려합니다. 여기서 반응 (Y)은 예측 변수 (x 's) + 오차 (ϵ)의 선형 함수입니다.

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \epsilon$$

데이터 테이블의 각 행에는 반응 값 및 p 예측 변수의 값이 포함됩니다. 각 관측값에 대해 예측 변수 값은 고정 값으로 간주됩니다. 그러나 반응 값은 확률 변수의 실현값으로 간주됩니다.

고정된 예측 변수 값을 고려하면 Y 값 집합에 대해 $\beta_0, \beta_1, \dots, \beta_p$ 계수를 추정할 수 있습니다. 일반적으로 Y 값 집합이 다르면 계수 추정값이 다릅니다. "추정값 상관계수" 옵션은 이러한 모수 추정값의 이론적 상관계수를 계산합니다. 기술 상세 정보는 "**사용자 검정 예에 대한 상세 정보**"(188 페이지)에서 확인하십시오.

모수 추정값의 상관계수는 예측 변수 값과 절편을 나타내는 항에만 의존합니다. 두 모수 추정값 간의 상관계수는 반응 값의 영향을 받지 않습니다.

두 추정값 간에 높은 양의 상관관계가 있으면 해당하는 두 예측 변수 간에 공선성 관계가 존재할 수 있음을 나타냅니다. 그러나 이러한 상관관계를 주의하여 해석해야 합니다(Belsley et al. 1980, p. 185, 92–94). 또한 예측 변수를 재조정하여 평균이 변화되면 모수 추정값과 절편 값의 상관관계가 변경됩니다.

추정값 상관계수의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Socioeconomic.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. Median House Value 를 선택하고 **Y** 를 클릭합니다.
4. Total Population, Median School Years, Total Employment 및 Professional Services 를 선택하고 **추가**를 클릭합니다.
5. "강조" 목록에서 **최소 보고서**를 선택합니다.
6. **실행**을 클릭합니다.
7. "반응 Median House Value" 옆의 빨간색 삼각형을 클릭하고 **추정값 > 추정값 상관계수**를 선택합니다.

그림 3.40 추정값 상관계수 보고서

추정값 상관계수					
상관	절편	Total Population	Median School Years	Total Employment	Professional Services
절편	1.0000	-0.5743	-0.9818	0.4719	0.6903
Total Population	-0.5743	1.0000	0.5871	-0.9746	-0.2398
Median School Years	-0.9818	0.5871	1.0000	-0.5132	-0.7085
Total Employment	0.4719	-0.9746	-0.5132	1.0000	0.1094
Professional Services	0.6903	-0.2398	-0.7085	0.1094	1.0000

보고서에서는 절편과 Median School Years의 모수 추정값 간에 높은 음의 상관관계가 있음을 보여 줍니다(-0.9818). Total Population과 Total Employment 간에도 높은 음의 상관관계가 있습니다(-0.9746).

명목형 효과에 대한 코딩

명목형 모델링 유형의 열을 모형에 입력하면 JMP에서는 이 열을 내부적으로 연속형 표시 변수 집합으로 나타냅니다. 각 변수는 -1, 0, 1 값만 가집니다. 이 코딩은 표시 변수를 사용하여 명목형 변수를 코드화하는 여러 방법 중 하나입니다. 명목형 열에 n 개 수준이 있는 경우 열을 나타내려면 $n-1$ 개의 표시 변수가 필요합니다. $n-1$ 개의 표시 변수가 필요하다는 것은 명목형 열과 관련된 주효과가 $n-1$ 의 자유도를 갖는다는 사실과 직접 관련이 있습니다. 자세한 내용은 "통계 상세 정보" 부록의 "**명목형 요인**"(493 페이지)에서 확인하십시오.

팁: 기본 보고서의 빨간색 삼각형 메뉴에서 열 저장 > 코딩 테이블 저장을 선택하여 코딩을 볼 수 있습니다. 자세한 내용은 "**코딩 테이블 저장**"(167 페이지)에서 확인하십시오.

수준이 4개인 명목형 열이 있다고 가정해 보겠습니다. Cholesterol.jmp 샘플 데이터 테이블의 treatment 열을 예로 사용합니다. treatment 열에는 A, B, Control 및 Placebo라는 4개의 수준이 있습니다. 처음 세 수준은 각각 표시 변수로 나타냅니다. 이러한 표시 변수 이름은 treatment[A], treatment[B] 및 treatment[Control]로 지정됩니다.

지정된 수준의 표시 변수는 해당 수준에 1, 마지막 수준에 -1, 나머지 수준에 0 값을 각각 할당합니다. 표 3.1에서는 이 예에 대한 treatment[A], treatment[B] 및 treatment[Control] 표시 변수의 정의를 보여 줍니다. 표시 변수 treatment[A]를 예로 들어보겠습니다. 표 3.1에 표시된 대로 이 변수는 다음 값을 할당합니다.

- treatment = A 인 행에 값 1 이 할당됩니다.
- treatment = B 또는 Control 인 행에 값 0 이 할당됩니다.
- treatment = Placebo 인 행에 값 -1 이 할당됩니다.

표 3.1 Cholesterol.jmp 의 treatment 에 대한 표시 변수 예시

행에 할당된 treatment	treatment[A]	treatment[B]	treatment[Control]
A	1	0	0
B	0	1	0

표 3.1 Cholesterol.jmp 의 treatment 에 대한 표시 변수 예시 (계속)

행에 할당된 treatment	treatment[A]	treatment[B]	treatment[Control]
Control	0	0	1
Placebo	-1	-1	-1

수준 순서는 "값 순서" 열 특성 (할당된 경우) 또는 JMP에서 할당한 기본 정렬 방식에 의해 결정됩니다. 일반적으로 기본 정렬 방식은 숫자 정렬 순서(숫자)와 영숫자 정렬 순서(문자 데이터)입니다. 그러나 월 이름과 같은 특정 범주형 값은 기본적으로 적절하게 정렬됩니다. 값 순서화에 대한 자세한 내용은 **JMP 사용의** 에서 확인하십시오.

이러한 변수는 모형을 파라미터화하는 데 사용됩니다. 변수는 일반적으로 데이터 테이블에 표시되지 않지만 이러한 변수에 대한 추정 계수가 "모수 추정값" 및 다른 보고서에 제공됩니다. 사용 가능한 다른 코딩도 많지만 이 코딩이 실용적이고 해석 가능한 것으로 입증되었습니다.

순서형 효과의 코딩에 대한 자세한 내용은 "통계 상세 정보" 부록의 "**순서형 요인**"(503페이지)에서 확인하십시오.

효과 선별

선별 설계에서 오차 자유도를 제공하지 않는 경우가 종종 있습니다. 따라서 효과에 대해 전통적 검정을 사용할 수 없습니다. 이러한 경우 효과 선별이 특히 유용합니다.

이러한 설계의 경우 효과 크기에 대한 대부분의 추론은 비절편 모수의 추정값이 상관관계가 없고 분산이 동일하다고 가정합니다. 이 가정은 많은 전통적 실험 설계와 관련된 모형에 적용됩니다. 그러나 이러한 가정이 해당되지 않는 경우가 있습니다. 두 경우 모두 효과 선별 플랫폼을 사용하면 유의한 효과를 확인하는 데 도움이 됩니다.

효과 선별 플랫폼은 **효과 희소성** 원리를 사용합니다(Box and Meyer 1986). 이 원리는 선별 설계에서 연구하는 효과 중 활성 효과는 상대적으로 적다고 주장합니다. 대부분은 비활성 상태이며, 이는 실제 효과가 0이고 추정값이 랜덤 오차임을 의미합니다.

"효과 선별"에서 다음 옵션을 사용할 수 있습니다.

척도화된 추정값 평균이 0 이고 범위가 2 가 되도록 척도화된 요인에 해당하는 모수 추정값을 제공합니다. 자세한 내용은 "**척도화된 추정값 및 연속형 항 코딩**"(141 페이지)에서 확인하십시오.

정규 확률도 정규성에서 벗어난 모수 추정값을 식별하여 활성 효과를 확인하는 데 도움이 됩니다. 자세한 내용은 "**정규 확률도 보고서**"(146 페이지)에서 확인하십시오.

Bayes 그림 Bayes 방법을 사용하여 모든 모형 항에 대한 사후 확률을 계산합니다. 자세한 내용은 "**Bayes 그림 보고서**"(147 페이지)에서 확인하십시오.

파레토도 직교화 및 표준화된 모수 추정값의 절대값을 해당 절대값의 합과 연결하여 그림에 나타냅니다. 자세한 내용은 "**파레토도 보고서**"(149 페이지)에서 확인하십시오.

척도화된 추정값 및 연속형 항 코딩

모수 추정값은 해당 요인의 척도에 따라 크게 좌우됩니다. 요인을 그램(g)에서 킬로그램(kg)으로 변환하면 모수 추정값이 1,000의 배수로 변경됩니다. 동일한 변경을 제곱(2차) 항에 적용하면 모수 추정값이 100만의 배수로 변경됩니다.

효과 크기를 더 잘 이해하고 비교하려면 모수 추정값을 척도에 따라 변하지 않는 방식으로 검토해야 합니다. 모수 추정값의 크기를 반응에 대한 해당 항의 효과 크기와 연관시키는 척도를 사용하는 것이 적절합니다. 이를 수행하는 방법에는 여러 가지가 있습니다.

보고서의 빨간색 삼각형 메뉴에 있는 효과 선별 > 척도화된 추정값 옵션은 척도화된 요인에 해당하는 계수를 제공합니다. 요인은 평균이 0이고 범위가 2가 되도록 척도화됩니다. 그림 3.41에서는 Drug.jmp에 대한 보고서를 보여 줍니다.

요인에 대한 표본 값의 최대값과 최소값이 표본 평균에서 등거리에 있는 경우 척도화된 요인의 범위는 -1 ~ 1입니다. 이 척도화는 실험 설계에 사용된 기존의 코딩에 해당합니다. 단일 요인을 사용한 회귀의 경우 척도화된 모수 추정값은 요인이 전체 범위에서 이동할 때 예측된 반응 변동의 절반입니다.

척도화된 추정값은 데이터에 코드화되지 않은 값이 포함된 경우 모형 항의 영향을 평가하는 데 중요합니다. 직교 설계의 경우 척도화된 추정값은 코드화되지 않은 데이터의 추정값과 동일합니다.

참고 : "코딩" 열 특성은 -1에서 1 사이의 범위로 값이 코드화되도록 요인 값을 선형으로 척도화합니다. 모수 추정값은 이러한 코드화된 값으로 제공되므로 이 경우에는 척도화된 추정값이 필요하지 않습니다. 척도화된 추정값을 구하는 데 사용되는 변환과 달리 코드화된 값은 평균이 0이 아닐 수 있습니다.

척도화된 추정값의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Drug.jmp를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. Y를 선택하고 **Y**를 클릭합니다.
4. Drug와 x를 선택하여 **모형 효과 생성** 목록에 추가합니다.
5. "강조"를 **최소 보고서**로 변경합니다.
6. **실행**을 클릭합니다.
7. "반응 y"의 빨간색 삼각형을 클릭하고 **효과 선별 > 척도화된 추정값**을 선택합니다.

보고서에서는 연속형 요인 x가 평균을 기준으로 중심화되고 범위의 반으로 척도화되었음을 나타냅니다.

그림 3.41 척도화된 추정값 보고서

▲ 척도화된 추정값

영속형 요인이 모든 수준으로 확장되었습니다.
연속형 요인을 평균으로 빼고, 범위의 반으로 척도화되었습니다.

항	척도화된 추정값	표준 오차	t 비	Prob> t
절편	7.9	0.731352	10.80	<.0001*
Drug[a]	-1.185037	1.060822	-1.12	0.2742
Drug[d]	-1.076065	1.041298	-1.03	0.3109
Drug[f]	2.2611017	1.093974	2.07	0.0488*
x	8.8846543	1.480478	6.00	<.0001*

모형은 각 Drug 그룹에 대해 하나씩 세 개의 평행선을 적합시킵니다. x 값의 범위는 3~21입니다. x의 척도화된 추정값 8.8846543은 Drug 그룹 중 하나에 대해 x = 21에 대한 예측값과 x = 3에 대한 예측값 간 차이의 절반입니다. 보고서의 빨간색 삼각형 메뉴에서 **열 저장 > 예측 계산식**을 선택하여 이 사실을 확인할 수 있습니다. 그런 다음 x = 21 및 x = 3에서 각 Drug 그룹에 대한 예측값을 구하기 위해 행을 추가합니다.

이 특정 데이터 집합의 x 값 범위에서 x의 영향은 약 17.8 단위의 범위에서 반응에 변화를 주는 것입니다. 원시 데이터에 기반한 x의 모수 추정값은 0.9871838이므로 반응에 대한 직접적인 해석을 허용하지 않습니다.

그림 옵션

"정규 확률도", "Bayes 그림" 및 "파레토도" 옵션은 "효과 선별" 보고서의 일부로 나타나는 보고서를 제공합니다. 이러한 보고서를 추정값 간의 이분산 및 상관관계를 수정하도록 구성할 수 있습니다.

참고 : "정규 확률도", "Bayes 그림" 및 "파레토도" 옵션을 사용하려면 모형에 4 개 이상의 모수가 포함되어야 합니다. 이러한 모수 중 하나가 절편일 수 있습니다.

변환

그림 옵션 중 하나를 선택하면 "효과 선별" 보고서 제목 바로 아래에 다음 정보가 나타납니다.

- 추정값의 분산이 동일하고 상관관계가 없는 경우 다음과 같은 두 가지 노트가 나타납니다.
 - 모수 추정값들은 등분산을 가집니다.
 - 모수 추정값 간에 상관 관계가 없습니다.
- 추정값의 분산이 동일하지 않거나 상관관계가 있는 경우 옵션 목록이 관련 노트를 대체합니다. 기본적으로 선택된 목록 항목은 JMP에서 추정값을 변환했음을 나타냅니다. 변환 중 하나 또는 둘 다 실행 취소하려면 해당 목록 항목을 선택합니다.

Lenth PSE 값

Lenth PSE(유사 표준 오차) 테이블은 노트 또는 옵션 목록 바로 아래에 나타납니다. PSE에 대한 설명은 "Lenth PSE"(110페이지)에서 확인하십시오. Lenth 테이블에 나타나는 통계량은 모수 추정값의 분산 및 상관관계에 따라 달라집니다.

모수 추정값의 분산이 동일하고 상관관계가 없는 경우 Lenth 테이블에 다음 통계량이 제공됩니다.

Lenth PSE 추정값에 대한 Lenth 유사 표준 오차입니다.

모수 추정값의 분산이 다르거나 상관관계가 있는 경우 또는 둘 다 해당하는 경우 Lenth 테이블에 다음 통계량이 제공됩니다.

t-검정 척도 Lenth PSE 변환된 모수 추정값에 대해 계산된 Lenth 유사 표준 오차를 변환된 척도의 표준 오차로 나눈 값입니다.

코드화된 척도 Lenth PSE 변환된 모수 추정값에 대한 Lenth 유사 표준 오차입니다.

모수 추정값 모집단 보고서

"모수 추정값 모집단" 보고서에는 모수 추정값에 대한 검정이 제공됩니다. 검정은 옵션 목록에 지정된 변환을 기반으로 합니다.

- **등분산을 가지도록 표준화된 추정값 사용** 옵션은 정규화 변환을 적용하여 분산을 표준화합니다. 이 옵션은 분산이 동일하지 않을 때 기본적으로 선택됩니다.
- **상관 관계가 없도록 직교된 추정값 사용** 옵션은 변환을 적용하여 상관관계를 제거합니다. 이 옵션은 추정값이 상관되어 있을 때 기본적으로 선택됩니다. 적용되는 변환은 순차 제곱합을 계산하는 데 사용되는 변환과 동일합니다. 추정값은 모든 이전 변수가 모형에 입력된 후 변수의 추가 기여도를 측정합니다.
- 추정값의 분산이 동일하고 상관관계가 없음을 나타내는 노트가 있으면 변환이 적용되지 않습니다.

테이블에 나타나는 열은 처음에 노트 또는 옵션 목록에 설명된 선택 항목에 따라 다릅니다. 보고서에서는 p 값이 0.20 이하인 추정값에 해당하는 행이 강조 표시됩니다. 보고서의 모든 버전에 "항", "추정값" 및 t 비와 $\text{Prob}>|t|$ 또는 유사 t 비와 유사 p 값 중 하나가 제공됩니다.

항 모수 추정값에 관심이 있는 모형 항을 제공합니다.

추정값 모수에 대한 추정값을 제공합니다. 모형 효과의 척도가 동일한 경우에만 추정값 크기를 비교할 수 있습니다.

t 비 오차의 자유도가 있는 경우 나타납니다. 이 값은 모수 추정값을 표준 오차로 나눈 값입니다.

Prob>|t| 검정의 p 값을 제공합니다. 변환이 적용되면 이 옵션은 변환된 데이터를 사용하여 검정의 p 값을 제공합니다.

유사 t 비 오차의 자유도가 없는 경우 나타납니다. 모수의 상대 표준 오차가 동일한 경우 유사 t 비는 모수 추정값을 Lenth PSE 로 나눈 값입니다. 상대 표준 오차가 다른 경우에는 "유사 t 비"(110 페이지)에 표시된 대로 계산됩니다.

유사 p 값 오차의 자유도가 없는 경우 나타납니다. 이 p 값은 t 분포를 사용하여 파생됩니다. 자유도는 $m/3$ 이며 가장 가까운 정수로 내림합니다. 여기서 m 은 모수의 수입니다.

등분산을 가지도록 표준화된 추정값 사용을 선택하고 모수 추정값이 상관관계가 없다는 노트가 나타나면 보고서에 "표준화 추정값"이라는 열이 표시됩니다. 이 열은 추정값이 등분산을 가지도록 변환한 결과로 생성된 모수 추정값을 제공합니다.

등분산을 가지도록 표준화된 추정값 사용과 상관 관계가 없도록 직교된 추정값 사용을 둘 다 선택하면 보고서에 " 직교 코드화됨 " 이라는 열이 표시됩니다. 다음 정보가 제공됩니다.

직교 코드화됨 추정값을 직교화하는 데 사용되는 변환 결과로 생성된 모수 추정값을 제공합니다.

직교 t 비 오차의 자유도가 있는 경우 나타납니다. 변환된 추정값의 t 비를 제공합니다.

유사 직교 t 비 오차의 자유도가 없는 경우 나타납니다. 이 값은 직교화된 추정값 (직교 코드화됨) 을 코드화된 척도 Lenth PSE 로 나누어 계산된 t 비입니다.

효과 선별 보고서

그림 3.42에서는 Bicycle.jmp 샘플 데이터 테이블의 "Fit Model" 스크립트를 실행하여 생성되는 "효과 선별" 보고서를 보여 줍니다. 이 형식의 보고서를 생성하려면 **효과 선별 > 정규 확률도**를 선택해야 합니다. 보고서 제목 바로 아래의 노트는 변환이 필요하지 않음을 나타냅니다. 따라서 Lenth PSE가 표시됩니다. 오차에 대한 자유도가 없으므로 잔차 오차의 추정값을 생성할 수 없습니다. 따라서 Lenth PSE를 잔차 오차의 추정값으로 사용하여 유사 t 비를 구합니다. 이러한 t 비에 대해 유사 p 값이 제공됩니다. p 값이 0.20 이하인 세 개의 추정값에 해당하는 비절편 항의 행이 강조 표시됩니다.

그림 3.42 분산이 동일하고 상관관계가 없는 추정값에 대한 효과 선별 보고서

효과 선별				
모수 추정값들은 등분산을 가집니다. 모수 추정값 간에 상관 관계가 없습니다.				
Lenth PSE 0.75				
모수 추정값 모집단				
항	추정값	유사 t 비	유사 p 값	
절편	66.5000	88.6667	<.0001*	
HBars[down]	0.5000	0.6667	0.5649	
Dynamo[off]	-6.0000	-8.0000	0.0096*	
Seat[down]	1.7500	2.3333	0.1266	
Tires[hard]	-1.2500	-1.6667	0.2195	
Gear[low]	-11.2500	-15.0000	0.0023*	
Raincoat[off]	0.2500	0.3333	0.7665	
Brkfast[no]	0.5000	0.6667	0.5649	
직교 t-검정에 유사 표준 오차가 사용되었습니다.				
정규 확률도				

분산이 동일하지 않고 상관관계가 있는 추정값에 대한 효과 선별 보고서

Odor.jmp 샘플 데이터 테이블에서 "Model" 스크립트를 실행하고 **실행**을 클릭합니다. 그림 3.43에 표시된 보고서를 생성하려면 "반응 odor"의 빨간색 삼각형을 클릭하고 **효과 선별 > 정규 확률도**를 선택합니다. "반응 odor"의 빨간색 삼각형을 클릭하고 "Bayes 그림" 또는 "파레토도" 옵션을 선택하여 보고서를 생성할 수도 있습니다.

보고서 제목 바로 아래에 있는 노트는 추정값을 표준화 및 직교화하기 위해 변환이 필요했음을 나타냅니다. "추정값 상관" 보고서에는 상관 행렬이 표시됩니다.

이 보고서에는 t-검정 척도 및 코드화된 척도 Lenth PSE가 표시됩니다. 그러나 오차에 대한 자유도가 있으므로 "모수 추정값 모집단" 보고서의 검정에 Lenth PSE가 사용되지 않습니다. p 값이 0.20

이 변환은 모형 항 순서에서 각 회귀변수를 앞에 오는 회귀변수에 대해 직교정규화합니다. 변환을 통해 대각 요소가 동일한 대각 공분산 행렬이 생성됩니다. 코드화된 추정값은 이 반복 과정의 결과입니다.

참고 : 직교 코드화된 추정값은 모형의 항 순서에 따라 다릅니다. 각 효과의 기여도는 해당 효과가 이전에 입력한 효과와 직교된 후에만 추정됩니다. 주효과를 먼저 입력한 후 이원 교호작용, 삼원 교호작용 등의 순서로 입력하는 것이 좋습니다.

정규 확률도 보고서

"정규 확률도" 보고서 제목 아래에서 정규 확률도 또는 1/2 정규 확률도(Daniel 1959)를 선택합니다. 두 그림 모두 효과 회소성 원리, 즉 비교적 적은 수의 효과가 활성화된다는 개념에 근거합니다. 비활성 효과는 랜덤 잡음을 나타냅니다. 해당 추정값은 평균이 0이고 분산이 σ^2 인 정규 분포를 따른다고 가정할 수 있습니다. 여기서 σ^2 은 잔차 오차 분산을 나타냅니다. 따라서 정규 확률도에서 비활성 효과를 나타내는 추정값은 기울기가 σ 인 선에 가깝게 분포됩니다.

정규 확률도

변환이 필요하지 않은 경우 정규 확률도의 수직 좌표는 추정값을 나타내고 수평 좌표는 정규 분위수를 나타냅니다. 비활성 효과를 나타내는 점은 기울기가 σ 인 선을 따라야 합니다. Lenth PSE를 사용하여 σ 를 추정하고 이 기울기의 파란색 선이 그림에 표시됩니다.

직교성 변환이 적용된 경우 세로 축은 정규 추정값을 나타냅니다. 이러한 값은 "모수 추정값 모집단" 보고서에 있는 직교 t 비 값입니다. 직교 t 비 값은 직교 코드화된 추정값을 코드화된 척도 Lenth PSE로 나눈 값입니다.

추정값은 σ 의 추정값에 의해 정규화되므로 비활성 효과에 해당하는 점은 기울기 1의 선을 따라 분포되어야 합니다. 기울기가 1인 빨간색 선과 기울기가 t -검정 척도 Lenth PSE와 동일한 파란색 선이 그림에 표시됩니다.

모든 경우에 "모수 추정값 모집단" 보고서의 p 값을 기반으로 0.20 수준에서 정규성을 벗어난 추정값은 그림에 라벨이 지정됩니다.

1/2 정규 확률도

1/2 정규 확률도에는 효과의 절대값이 표시됩니다. 축 구성 및 표시된 선은 정규 확률도의 양상을 반영합니다.

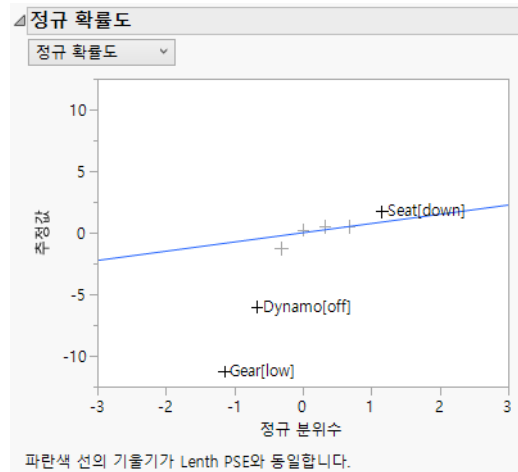
그림 3.44에서는 Bicycle.jmp 샘플 데이터 테이블에 대한 "정규 확률도" 보고서를 보여 줍니다. 이 모형에는 변환이 필요하지 않으므로 정규 분위수에 대한 원시 추정값이 그림에 표시됩니다. 기울기가 Lenth PSE와 동일한 선이 그림에 표시됩니다. 이 그림에서는 Gear, Dynamo 및 Seat가 활성 요인임을 나타냅니다.

정규 확률도의 예

1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Bicycle.jmp를 엽니다.

2. 분석 > 모형 적합을 선택합니다 .
3. Y 를 선택하고 Y 를 클릭합니다 .
4. HBars ~ Brkfast 를 선택하고 추가를 클릭합니다 .
5. 실행을 클릭합니다 .
6. " 반응 Y" 의 빨간색 삼각형을 클릭하고 효과 선별 > 정규 확률도를 선택합니다 .
다음과 같은 정규 확률도가 나타납니다 .

그림 3.44 정규 확률도



Bayes 그림 보고서

Bayes 그림 보고서에는 활성 효과를 확인하는 또 다른 방법이 제공됩니다. 이 보고서는 Bayes 방법을 사용하여 사후 확률을 계산하는 데 도움이 됩니다. Box and Meyer(1986)를 기반으로 한 이 방법은 추정값이 두 분포의 혼합물이라고 가정합니다. 비활성 효과에 해당하는 대부분의 추정값은 분산이 σ^2 인 순수 랜덤 정규 잡음으로 간주됩니다. 활성 효과에 해당하는 나머지 추정값은 분산이 σ^2 보다 K 배 큰 오염 분포에서 오는 것으로 간주됩니다.

항 모수 추정값에 해당하는 모형 항을 제공합니다.

추정값 모수 추정값을 제공합니다. Bayes 그림은 추정된 표준편차가 1인 추정값과 관련하여 생성됩니다. 추정값 간에 상관관계가 없으면 t 비가 사용됩니다. 추정값 간에 상관관계가 있으면 직교 t 비가 사용됩니다.

사전 확률 추정값이 0 이 아닐 확률, 즉 추정값이 오염 분포에 있을 확률을 지정할 수 있습니다. 추정값에 대한 사전 확률은 대개 동일한 값으로 설정됩니다. 처음에는 일반적으로 권장되는 값 0.2 가 표시되지만 값을 변경할 수 있습니다.

K 오염 오차 분산에 대한 오염 분포 분산의 비율을 나타내는 오염 계수 값입니다. 일반적으로 K 는 기본값 10 으로 설정됩니다.

표준 오차 척도 표준 오차 추정값에 대한 자유도가 있는 경우 이 값은 1로 설정됩니다. 보고서에 사용된 추정값은 단위 분산으로 변환 및 척도화되므로 JMP 에서 이 값을 사용합니다. 오차 추정값이 없는 포화 모형의 경우 이 값은 0 으로 설정됩니다. 다른 값을 지정할 경우에는 RMSE 추정값의 배율로 간주하십시오.

DFE 오차 자유도를 제공합니다.

그림 3.45에서는 Bicycle.jmp 샘플 데이터 테이블에 대한 Bayes 그림의 기본 설정이 표시된 규격창을 보여 줍니다. 이 창에서 "시작"을 클릭하면 보고서가 업데이트되어 각 항에 대한 사후 확률 및 막대 차트가 표시됩니다(그림 3.46).

Bayes 그림의 예

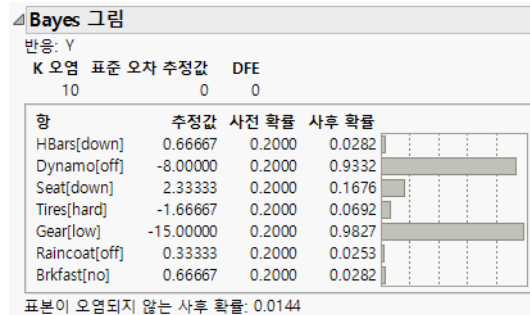
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Bicycle.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. Y 를 선택하고 **Y** 를 클릭합니다.
4. HBars ~ Brkfast 를 선택하고 **추가**를 클릭합니다.
5. **실행**을 클릭합니다.
6. "반응 Y"의 빨간색 삼각형을 클릭하고 **효과 선별 > Bayes 그림**을 선택합니다.

그림 3.45 Bayes 그림 규격

항	추정값	사전 확률	K 오염	표준 오차 척도	DFE
HBars[down]	0.6667	0.2000	10	0	0
Dynamo[off]	-8.0000	0.2000			
Seat[down]	2.3333	0.2000			
Tires[hard]	-1.6667	0.2000			
Gear[low]	-15.0000	0.2000			
Raincoat[off]	0.3333	0.2000			
Brkfast[no]	0.6667	0.2000			

7. **시작**을 클릭하여 사후 확률을 계산합니다.

그림 3.46 Bayes 그림 보고서



"Bayes 그림" 보고서에서 그림 아래의 노트에는 표본이 오염되지 않는 사후 확률이 제공됩니다. 사후 확률은 사전 확률 및 데이터를 기반으로 활성 효과가 전혀 없을 확률입니다. 사후 확률 값 (0.0144)이 작으므로 활성 효과가 존재할 가능성이 있음을 나타냅니다. 사후 확률 열은 적어도 Dynamo와 Gear가 활성 효과임을 나타냅니다.

파레토도 보고서

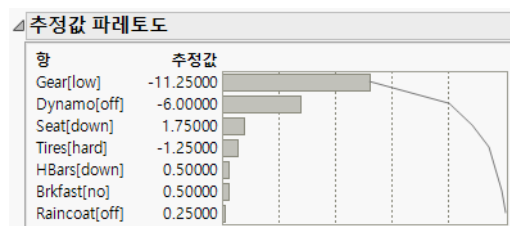
파레토도 보고서에는 추정값의 절대값에 대한 파레토 차트가 제공됩니다. 그림 3.47에서는 Bicycle.jmp 샘플 데이터 테이블에 대한 파레토도 보고서를 보여 줍니다.

- 원래 추정값의 분산이 동일하고 상관관계가 없으면 원래 추정값이 그림에 표시됩니다.
- 원래 추정값의 분산이 동일하지 않고 상관관계가 없으면 t 비가 그림에 표시됩니다.
- 원래 추정값의 분산이 동일하지 않고 상관관계가 있으면 직교 코드화된 추정값이 그림에 표시됩니다.

그림의 누적합 선은 추정값의 절대값을 합합니다. 직교 추정값은 모형에 항을 입력하는 순서에 따라 달라집니다.

참고: 그림에 나타나는 추정값은 동일한 분산을 가지도록 표준화되고 직교하도록 변환됩니다. 이러한 변환을 실행 취소하는 옵션을 사용할 수 있습니다. 자세한 내용은 "[변환](#)"(142 페이지)에서 확인하십시오.

그림 3.47 파레토도



요인 프로파일링

"요인 프로파일링" 메뉴는 추정된 모형을 탐색하고 시각화하는 데 도움이 됩니다. 반응 표면의 모양을 탐색하고, 최적 설정을 찾고, 지정된 잡음 가정을 기반으로 반응 데이터를 시뮬레이션하고, 필요한 경우 반응을 변환할 수 있습니다.

프로파일러, 등고선 프로파일러, 혼합물 프로파일러 및 표면 프로파일러는 표준 최소 제공 모형을 넘어 사용 범위가 확장되는 다목적 도구입니다. 해석 및 사용에 대한 자세한 내용은 **Profilers**에서 확인하십시오.

모형의 반응 열에 한 변수의 가역 변환이 포함되어 있으면 프로파일러 및 교호작용 그림에 변환되지 않은 척도의 반응이 표시됩니다.

"요인 프로파일링"에서 다음 옵션을 사용할 수 있습니다.

참고 : 모형에 표현식 또는 벡터가 효과로 포함된 경우 이러한 옵션 중 대부분을 사용할 수 없습니다.

프로파일러 각 X 변수에 대한 예측 트레이스를 표시합니다. 하나 이상의 반응에 대한 최적 설정을 찾고 시뮬레이션을 사용하여 반응 분포를 탐색할 수 있습니다. 자세한 내용은 "[프로파일러](#)"(151 페이지) 및 **Profilers**의 에서 확인하십시오.

교호작용 그림 모형에 교호작용 효과가 있는 경우 교호작용 그림의 행렬을 표시합니다. 자세한 내용은 "[교호작용 그림](#)"(151 페이지)에서 확인하십시오.

등고선 프로파일러 반응 표면을 그래픽으로 최적화하는 데 유용한 대화식 등고선 프로파일러를 제공합니다. 자세한 내용은 "[등고선 프로파일러](#)"(153 페이지) 및 **Profilers**의 에서 확인하십시오.

혼합물 프로파일러 혼합물 실험 모형의 반응 등고선을 삼원 그림에 표시합니다. 자세한 내용은 "[혼합물 프로파일러](#)"(154 페이지) 및 **Profilers**의 에서 확인하십시오.

참고 : 이 옵션은 "혼합물 효과" 속성을 3 개 이상의 효과에 적용하거나 "혼합물" 특성을 3 개 이상의 열에 적용한 경우에만 나타납니다.

입방체도 입방체 꼭지점에 요인 범위의 극단에 대한 예측값을 표시합니다. 자세한 내용은 "[입방체도](#)"(155 페이지)에서 확인하십시오.

Y의 Box-Cox 변환 분산의 정규성 및 동질성 가정 충족과 관련하여 가장 적합한 반응의 Box-Cox 멱 변환을 찾습니다. 자세한 내용은 "[Y의 Box-Cox 변환](#)"(156 페이지)에서 확인하십시오.

표면 프로파일러 반응 표면의 3 차원 표면 그림을 표시합니다. 자세한 내용은 "[표면 프로파일러](#)"(158 페이지) 및 **Profilers**의 에서 확인하십시오.

프로파일러

참고 : 프로파일러에 대한 자세한 내용은 **Profilers** 의 에서 확인하십시오 .

프로파일러 또는 "예측 프로파일러"에서는 각 X 변수에 대한 예측 트레이스를 보여 줍니다.

그림 3.48에서는 **Reactor.jmp** 샘플 데이터 테이블에 대한 프로파일러의 일부를 보여 줍니다. 각 X 변수에 대한 수직 점선은 **현재 값** 또는 **현재 설정**을 표시합니다. 프로파일러를 사용하면 한 번에 하나의 변수를 변경하여 예측 반응에 대한 영향을 검토할 수 있습니다.

그림 3.48 예측 트레이스 그림

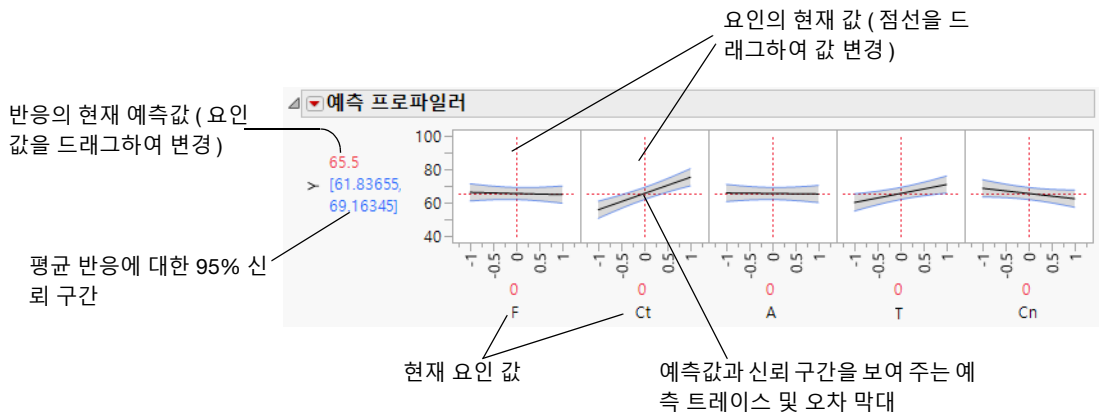


그림 3.48의 F 및 Ct 요인은 연속형입니다. 요인이 명목형이면 가로 축에 수준이 표시됩니다.

각 X 변수에 대해 요인 이름 위의 값이 현재 값입니다. 그래프를 클릭하거나, 새 현재 값을 지정할 점선을 드래그하여 현재 값을 변경합니다.

- 수평 점선은 X 변수의 현재 값에 대한 각 Y 변수의 **현재 예측값**을 나타냅니다.
- 그림 안의 검은색 선은 개별 X 변수의 현재 값을 변경할 때 예측값이 어떻게 변하는지 보여 줍니다. 예측값에 대한 95% 신뢰 구간은 예측 트레이스 (연속형 변수의 경우) 또는 오차 막대 (범주형 변수의 경우) 주위의 점선 곡선으로 표시됩니다.

교호작용 그림

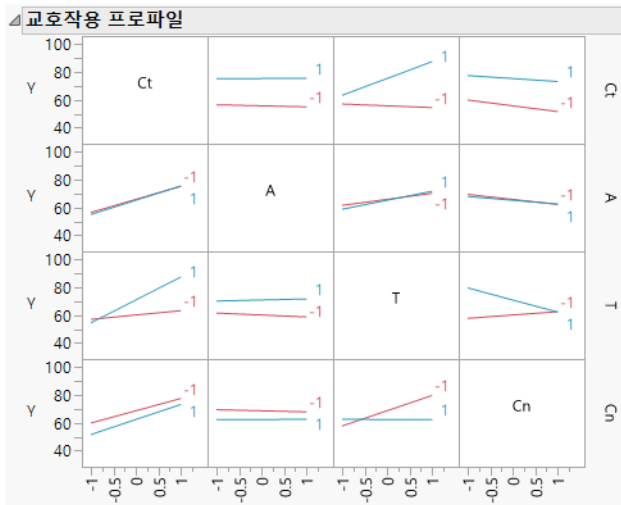
모형에 교호작용 효과가 있는 경우 "교호작용 그림" 옵션은 교호작용 그림의 행렬을 표시합니다. 행렬의 각 셀에는 그림이 나타나는 열에 표시되는 효과에 대해 가로 축이 척도화된 그림이 포함됩니다. 해당하는 행에 표시된 효과와 해당 효과의 교호작용에 대해 선분이 표시됩니다. 따라서 교호작용 그림은 행 효과와 열 효과의 교호작용을 보여 줍니다.

행 효과의 각 수준에 대해 선분이 그림에 표시됩니다. 모형에서 예측한 반응 값이 선분으로 결합됩니다. 평행하지 않는 선분은 교호작용이 있을 수 있다는 시각적 증거를 제공합니다. 그러나 교호작용이 있다는 결론을 내리기 전에 제안된 교호작용의 p 값을 확인해야 합니다. 그림 3.49에서는 Reactor.jmp 샘플 데이터 테이블에 대한 교호작용 그림 행렬을 제공합니다.

교호작용 그림의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Reactor.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. Y 를 선택하고 **Y** 를 클릭합니다.
4. **차수** 상자에 2 가 있는지 확인합니다.
5. Ct, A, T 및 Cn 을 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다.
6. **실행**을 클릭합니다.
7. "반응 Y" 의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 교호작용 그림**을 선택합니다.

그림 3.49 교호작용 그림



T*Cn 교호작용에 해당하는 그림은 그림 맨 아래 행의 세 번째 그림 또는 마지막 그림 열의 세 번째 그림입니다. 두 그림 모두 낮은 수준의 T에서 Y에 대한 Cn의 효과가 상당히 일정하다는 것을 보여 줍니다. 이때 Cn이 설정된 수준이 높은지 아니면 낮은지는 상관이 없습니다. 그러나 높은 수준의 T에서 Y에 대한 Cn의 효과는 해당 수준에 따라 다릅니다. -1의 Cn은 1의 Cn보다 예측 Y 값이 더 높습니다. 이 교호작용은 p 값 < 0.0001 이므로 유의합니다.

특정 설계에서는 이원 교호작용이 다른 이원 교호작용과 별칭 관계를 갖습니다. 별칭 관계가 있으면 "교호작용 프로파일" 그림에서 이러한 교호작용에 해당하는 셀이 흐리게 표시됩니다.

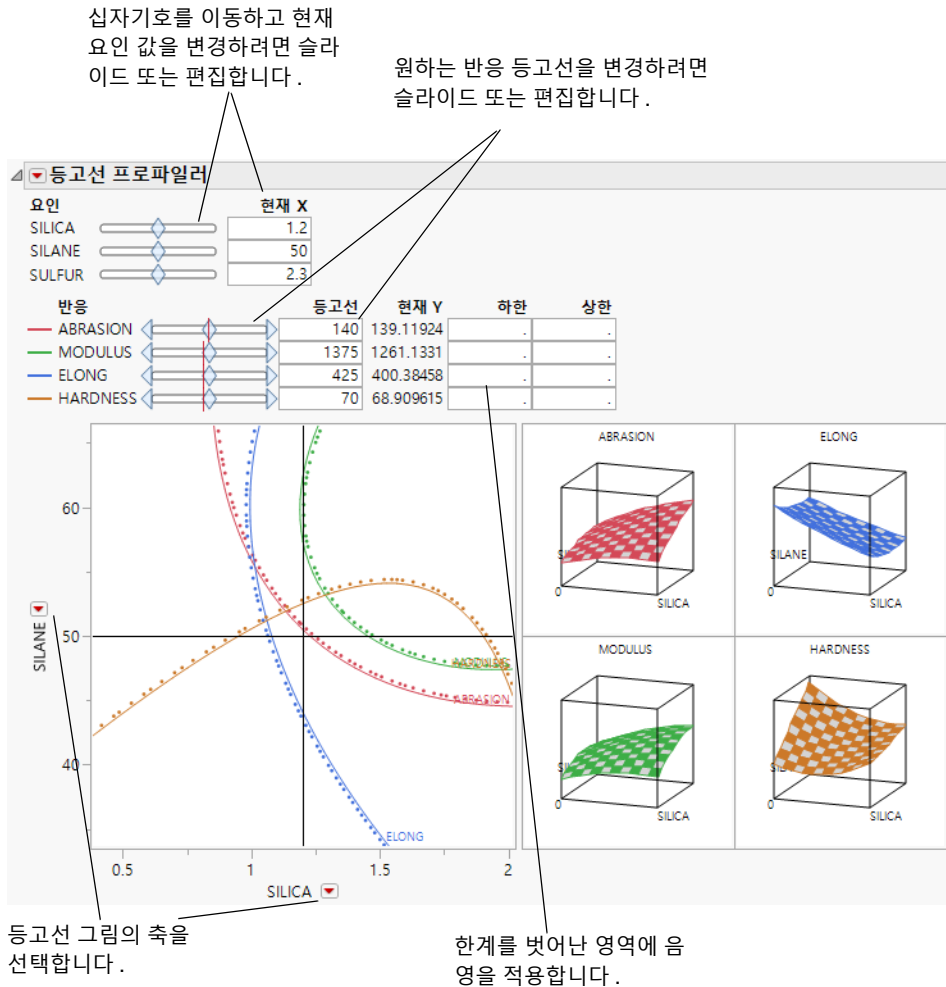
등고선 프로파일러

참고 : 등고선 프로파일러에 대한 자세한 내용은 **Profilers** 의 에서 확인하십시오 .

반응 표면을 그래픽으로 최적화하려면 대화식 등고선 프로파일러를 사용합니다. 등고선 프로파일러에는 적합 모형에 대해 한 번에 두 요인의 등고선이 표시됩니다. 표면 그림도 보고서에 포함됩니다.

그림 3.50에서는 Tiretread.jmp 샘플 데이터 테이블에 대한 등고선 프로파일러 보기를 보여 줍니다. **RSM for 4 Responses** 데이터 테이블 스크립트를 실행하고 "최소 제곱 적합" 보고서 메뉴에서 **프로파일러 > 등고선 프로파일러**를 선택합니다.

그림 3.50 등고선 프로파일러



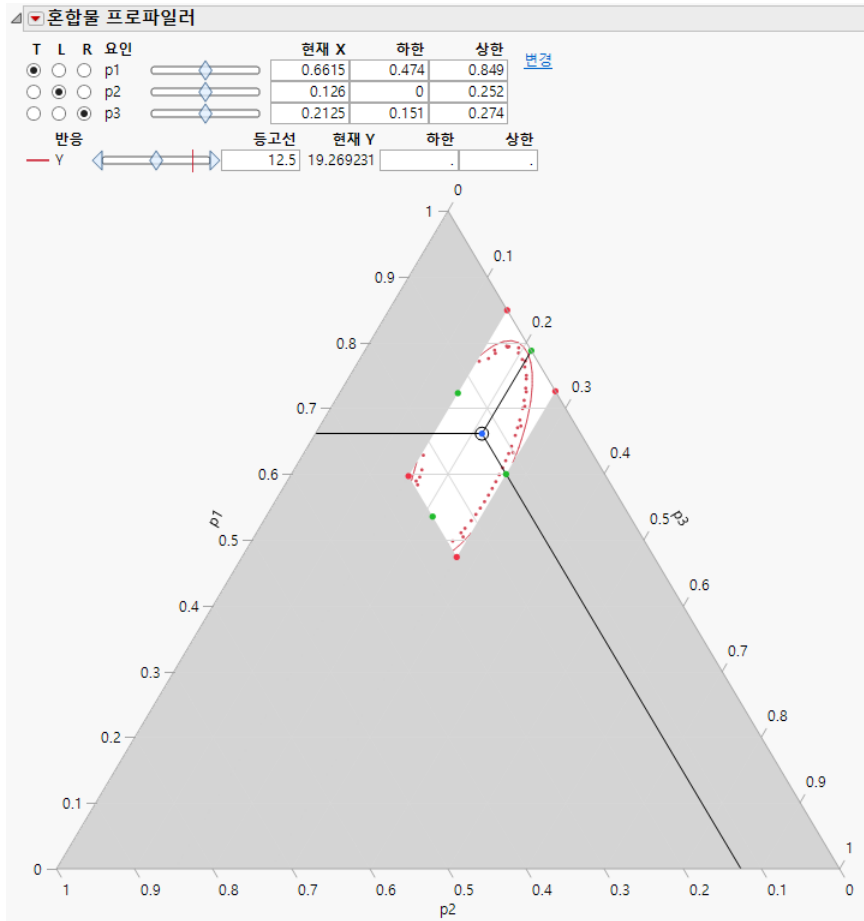
혼합물 프로파일러

참고 : 이 옵션은 효과에 대해 **매크로 > 혼합물 반응 표면** 옵션을 지정한 경우에만 나타납니다. 혼합물 프로파일러에 대한 자세한 내용은 **Profilers** 의 에서 확인하십시오.

혼합물 프로파일러에는 혼합물 실험 모형의 반응 등고선이 삼원 그림에 표시됩니다. 실험에서 세 개 이상의 요인이 혼합물 성분인 경우 혼합물 프로파일러를 사용합니다. 혼합물 프로파일러를 사용하여 실험의 반응 표면을 시각화하고 최적화할 수 있습니다.

그림 3.51에서는 Plasticizer.jmp 샘플 데이터 테이블의 모형에 대한 혼합물 프로파일러를 보여 줍니다. **Model** 데이터 테이블 스크립트를 실행한 후 보고서의 빨간색 삼각형 메뉴에서 **요인 프로파일링 > 혼합물 프로파일러**를 선택합니다. 그림의 왼쪽 위에 있는 여러 라디오 버튼을 선택하여 요인에 대한 그림 축을 수정합니다. 그림의 오른쪽 위에 있는 "하한" 및 "상한" 열을 사용하면 요인과 반응 모두에 대한 제약 조건을 입력할 수 있습니다.

그림 3.51 혼합물 프로파일러



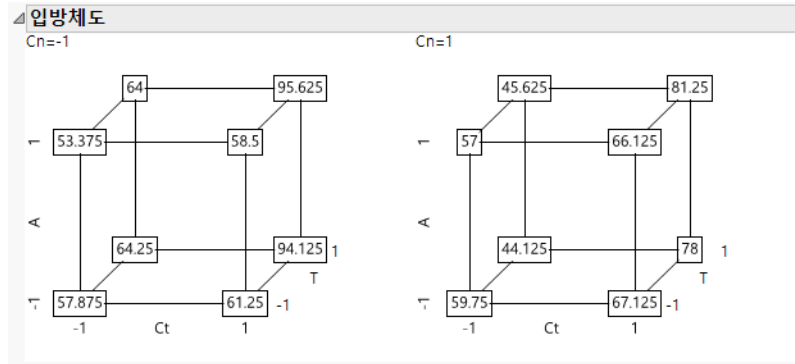
입방체도

"입방체도" 옵션은 요인 범위의 극단에 대한 예측값을 표시합니다. 이러한 값은 입방체의 꼭지점에 나타납니다(그림 3.52). 꼭지점은 요인의 최소 및 최대 관측값으로 정의됩니다. 반응이 여러 개 있으면 여러 반응이 각 꼭지점에 누적되어 표시됩니다.

입방체도의 예

1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Reactor.jmp 를 엽니다.
2. 분석 > 모형 적합을 선택합니다.
3. Y 를 선택하고 Y 를 클릭합니다.
4. 차수 상자에 2 가 있는지 확인합니다.
5. Ct, A, T 및 Cn 을 선택하고 매크로 > 특정 차수까지의 요인을 클릭합니다.
6. 실행을 클릭합니다.
7. "반응 Y" 의 빨간색 삼각형을 클릭하고 요인 프로파일링 > 입방체도를 선택합니다.

그림 3.52 입방체도



Cn = -1에 대한 입방체도와 Cn = 1에 대한 입방체도가 하나씩 있습니다. 요인이 다른 입방체 좌표에 매핑되도록 레이아웃을 변경하려면 첫 번째 입방체에서 요인 이름을 클릭합니다. 그런 다음 원하는 축의 요인 이름을 포함하도록 드래그합니다. 예를 들어 그림 3.52에서 T를 클릭하여 Ct 위로 드래그하면 T와 Ct 및 해당 좌표의 위치가 서로 바뀝니다. 단일 입방체에서 Cn 수준을 보려면 첫 번째 입방체에서 해당 요인 위로 드래그하여 다른 요인과 교환합니다.

Y의 Box-Cox 변환

Box-Cox 척 변환은 분산의 정규성 및 동질성에 대한 일반적인 회귀 가정이 더 상세하게 충족되도록 반응을 변환하는 데 사용됩니다. 그런 다음 회귀 모형을 사용하여 변환된 반응을 적합시킬 수 있습니다. 그러나 다른 이유로 변수를 변환할 때도 Box-Cox 척 변환을 사용할 수 있습니다. 이 변환은 반응 Y가 순양수인 경우에만 적합합니다.

일반적으로 사용되는 변환은 반응을 지정된 횟수만큼 거듭제곱합니다. Box and Cox(1964)에서는 이러한 척 변환 계열을 공식화하고 기술했습니다. 변환 계산식은 모수 λ 로 표현된 연속 정의

를 제공하고 오차 제곱합을 비교할 수 있도록 구성되었습니다. 구체적으로, 다음 방정식을 사용하여 변환 계열을 제공합니다.

$$Y_{\lambda} = \begin{cases} \frac{y^{\lambda}-1}{\lambda y^{\lambda-1}} & \text{단, } \lambda \neq 0 \\ y \ln(y) & \text{단, } \lambda = 0 \end{cases}$$

여기서 $\bar{y}$ 는 기하평균을 나타냅니다.

"Y의 Box-Cox 변환" 옵션은 $\lambda = -2$ 에서 2까지 0.2 단위로 증가하여 변환을 적합시킵니다. λ 값을 선택하기 위해 이러한 각 변환에 대한 가능도 함수가 계산됩니다. 이 함수는 오차가 독립적이며 평균이 0, 분산이 σ^2 인 정규 분포를 따른다고 가정하여 계산됩니다. 가능도를 최대화하는 λ 값이 선택됩니다. 또한 이 값은 λ 값에 대해 SSE를 최소화합니다. SSE를 최소화하는 λ 값은 SSE가 가장 작은 격자 점을 둘러싼 두 증분 격자점 사이의 2차 보간을 사용하여 구합니다. 이 보간법으로 인해 SSE 값이 음수가 되면 SSE를 최소화하는 λ 의 격자 값이 최선의 λ 로 보고됩니다.

"Box-Cox 변환" 보고서에는 λ 값에 대한 SSE(오차 제곱합) 값을 보여 주는 그림이 표시됩니다. 그림의 빨간색 수평선은 λ 에 대한 단측 95% 신뢰 구간을 나타냅니다. 이 신뢰 구간은 Box and Cox(1964, p. 216)에 정의된 신뢰 영역을 기반으로 합니다. 신뢰 영역은 다음 부등식으로 정의됩니다.

$$SSE(\lambda) < SSE(\lambda_{\text{best}}) * \exp(\text{ChiSquareQuantile}(0.95,1) / \text{dfe})$$

다음은 각 요소에 대한 설명입니다.

$SSE(\lambda_{\text{best}})$ 는 보고된 최선의 λ 를 사용하여 계산된 SSE입니다.

$\text{ChiSquareQuantile}(0.95,1)$ 은 자유도가 1인 χ^2 분포의 0.95 번째 분위수입니다.

dfe는 회귀 모형에 대한 분산 분석 테이블의 오차 자유도입니다.

"Box-Cox 변환" 보고서에서는 다음 옵션을 제공합니다.

변환을 사용한 재적합 변환된 Y 변수를 정의하기 위한 람다 값을 지정한 다음, 변환된 변수에 대한 최소 제곱 적합을 제공할 수 있습니다.

변환으로 바꾸기 변환된 Y 변수를 정의하기 위한 람다 값을 지정한 다음, 기존 최소 제곱 적합을 변환된 변수에 대한 적합으로 바꿀 수 있습니다. 반응이 여러 개인 경우 "변환으로 바꾸기"는 변환 대상 반응에 대한 보고서만 바꿉니다.

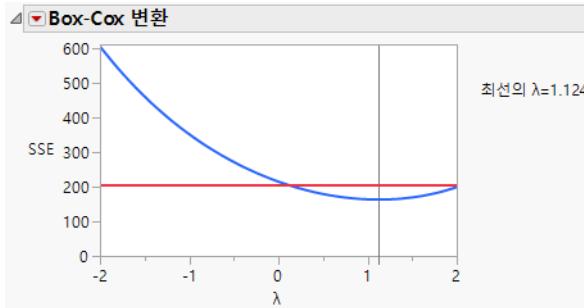
최적 변환 저장 데이터 테이블에 새 열을 생성하고 최적 변환을 위한 계산식을 저장합니다.

특정 변환 저장 람다 값을 지정하고, 지정된 변환을 위한 계산식이 포함된 열을 데이터 테이블에 생성할 수 있습니다.

추정값 테이블 -2에서 2까지 0.2 단위로 증가하는 모든 λ 에 대한 모수 추정값 및 SSE 값을 포함하는 새 데이터 테이블을 생성합니다.

Y의 Box-Cox 변환의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Reactor.jmp**를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. **Y**를 선택하고 **Y**를 클릭합니다.
4. **차수** 상자에 2가 있는지 확인합니다.
5. **F, Ct, A, T** 및 **Cn**을 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다.
6. **실행**을 클릭합니다.
7. "반응 Y"의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > Y의 Box-Cox 변환**을 선택합니다.

그림 3.53 Y의 Box-Cox 변환

이 그림에서는 최선의 λ 값이 0.1과 2.0 사이임을 보여 줍니다. JMP가 λ 값의 0.2 단위 격자에서 최적 값 두 개 사이의 보간을 사용하여 선택한 값은 1.124입니다.

8. (선택 사항) 그래프를 생성하는 데 사용된 SSE 값을 보려면 **추정값 테이블**을 선택합니다.

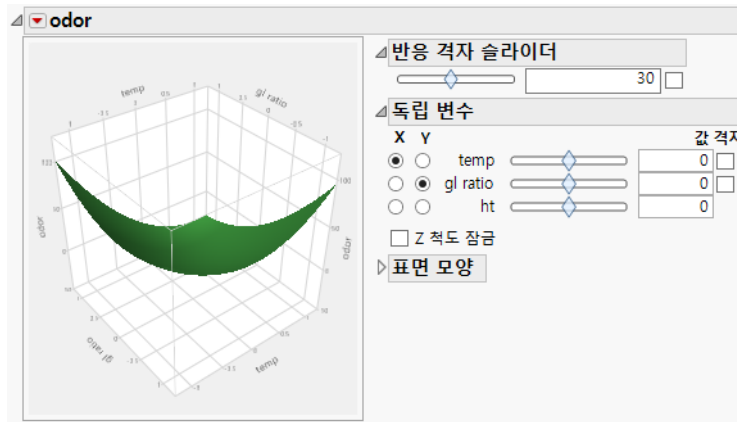
표면 프로파일러

참고 : 표면 프로파일러에 대한 자세한 내용은 **Profilers**의 에서 확인하십시오.

표면 프로파일러에는 반응 표면의 3차원 표면 그림이 표시됩니다.

그림 3.54에서는 **Odor.jmp** 샘플 데이터 테이블의 모형에 대한 표면 프로파일러를 보여 줍니다. **Model** 데이터 테이블 스크립트를 실행한 후 보고서의 빨간색 삼각형 메뉴에서 **요인 프로파일링 > 표면 프로파일러**를 선택합니다. "독립 변수" 아래의 라디오 버튼을 사용하여 축의 변수를 변경할 수 있습니다. 또한 "표면 모양" 아래에서 "실제값"을 클릭하여 그림에 점을 표시할 수도 있습니다.

그림 3.54 표면 그림



행 진단

행 진단 메뉴는 행 또는 관측값과 관련된 문제를 해결합니다.

회귀 그림 범주형 효과의 각 수준에 대한 데이터 산점도와 회귀선을 보여 주는 "회귀 그림" 보고서를 표시합니다.

이 옵션은 모형에 연속형 효과가 정확히 하나만 있고 범주형 효과가 최대 한 개인 경우에만 나타납니다. 이 경우 "회귀 그림" 보고서가 기본적으로 제공됩니다.

실제값 대 예측값 그림 Y의 예측값에 대한 Y의 관측값을 보여 주는 "실제값 대 예측값 그림"을 표시합니다. 자세한 내용은 "[레버리지 그림](#)"(161 페이지)에서 확인하십시오.

참고: "모형 적합" 시작 창의 "강조"에서 "효과 레버리지" 또는 "효과 선별"을 선택하고 R^2 값이 0.999 미만인 경우 "실제값 대 예측값 그림"이 기본적으로 표시됩니다.

효과 레버리지 그림 모형의 각 효과에 대한 "레버리지 그림" 보고서를 표시합니다. 이 그림에서는 관측값이 해당 효과에 대한 검정에 어떤 영향을 미치는지 보여 주고 다중공선성에 대한 정보를 제공합니다. 자세한 내용은 "[레버리지 그림](#)"(161 페이지)에서 확인하십시오.

참고: "모형 적합" 시작 창의 "강조"에서 "효과 레버리지"를 선택하고 R^2 값이 0.999 미만인 경우 효과 레버리지 그림이 기본적으로 표시됩니다. 이 그림은 "전체 모형" 보고서의 오른쪽에 나타납니다. "강조"에서 다른 옵션을 선택하면 "효과 상세 정보" 보고서에 효과 레버리지 그림이 나타납니다. 모든 경우에 효과 레버리지 그림을 표시하려면 회귀 보고서 > 효과 상세 정보 옵션을 선택해야 합니다.

잔차 대 예측값 그림 "잔차 대 예측값 그림" 보고서를 표시합니다. 이 그림에서는 Y의 예측값에 대해 그림에 표시된 잔차를 보여 줍니다. 일반적으로 0 주위에 무작위로 흩어진 잔차 값을 원합니다.

참고 : " 모형 적합 " 시작 창의 " 강조 " 에서 " 효과 레버리지 " 또는 " 효과 선별 " 을 선택하고 R^2 값이 0.999 미만인 경우 " 잔차 대 예측값 그림 " 이 기본적으로 표시됩니다.

잔차 대 행 번호 그림 " 잔차 대 행 번호 그림 " 보고서를 표시합니다. 행 번호에 대한 잔차 값이 그림에 표시됩니다. 이 그림은 관측값의 행 순서로 인한 패턴을 탐지하는 데 도움이 될 수 있습니다.

스튜던트화 잔차 그림 " 스튜던트화 잔차 " 그림을 표시합니다. 그림의 각 점은 현재 관측값이 삭제된 상태에서 구한 표준편차 추정값을 사용하여 계산됩니다. 이러한 잔차를 **RStudent** 또는 외부 스튜던트화 잔차라고도 합니다.

이 그림은 다음과 같은 두 가지 한계 집합을 포함합니다.

- 그림에 빨간색으로 나타나는 외부 한계는 95% Bonferroni 한계입니다. 이러한 한계는 $\pm t_{\text{Quantile}}(0.025/n, n-p-1)$ 에 배치됩니다. 여기서 n 은 관측값 수이고 p 는 예측 변수의 수입니다.
- 그림에 녹색으로 나타나는 내부 한계는 95% 개별 t 분포 한계입니다. 이러한 한계는 $\pm t_{\text{Quantile}}(0.025, n-p-1)$ 에 배치됩니다. 여기서 n 은 관측값 수이고 p 는 예측 변수의 수입니다.

빨간색 한계를 벗어나는 점은 가능성 있는 이상치로 처리해야 합니다. 녹색 한계를 벗어나지만 빨간색 한계 내에 있는 점은 가능하지만 확신도가 적은 이상치로 처리해야 합니다. 이러한 한계에 대한 95% 신뢰 수준은 " 모형 규격 " 창의 " 유의 수준 설정 " 옵션에서 선택한 항목의 영향을 받지 않습니다.

주의 : 열 저장 > 스튜던트화 잔차를 사용하여 저장된 잔차는 외부 스튜던트화 잔차가 아닙니다.

참고 : 모형에 임의 효과가 포함되고 시작 창에서 " 방법 " 에 " REML " 을 지정한 경우, " 스튜던트화 잔차 " 그림에 한계가 포함되지 않으며 그림에 표시된 점은 외부 스튜던트화되지 않습니다.

잔차 대 정규 분위수 그림 (시작 창에서 " 방법 " 에 " REML " 을 지정한 경우에는 사용 불가능) " 잔차 대 정규 분위수 그림 " 을 표시합니다. 정규 분포의 분위수에 대한 잔차 값이 그림에 표시됩니다. 이 그림은 잔차의 정규성 가정을 평가하는 데 도움이 될 수 있습니다.

Press Press 통계량 및 RMSE(제공근 평균 제곱 오차) 를 제공하는 "Press" 보고서를 표시합니다. Press 통계량은 여러 모형을 비교할 때 유용합니다. Press 통계량이 낮은 모형이 선호됩니다. 자세한 내용은 "Press"(164 페이지) 에서 확인하십시오.

Durbin-Watson 검정 (빈도 열을 지정한 경우에는 사용 불가능) 잔차에 1 차 자기상관이 있는지 여부를 검정하기 위한 통계량을 제공하는 "Durbin-Watson" 보고서를 표시합니다. 또한 잔차와 확률 <DW(통계량과 관련된 정확 확률) 의 자기상관도 보고서에 표시됩니다. 이 옵션은 시계열 데이터에만 사용할 수 있으며 관측값이 시간순으로 정렬되어 있다고 가정합니다.

레버리지 그림

X에 대한 효과 레버리지 그림은 다음과 같은 방식으로 유용합니다.

- X에 대한 가설 검정에 영향을 미칠 수 있는 점을 확인할 수 있습니다.
- 일반적이지 않은 패턴과 모형 가정 위반을 발견할 수 있습니다.
- 다중공선성 문제를 발견할 수 있습니다.

생성

효과에 대한 레버리지 그림에서는 다른 효과가 이미 모형에 있는 경우 이 효과를 모형에 추가할 때 미치는 영향을 보여 줍니다. 설명을 위해 단일 연속형 효과 X에 대한 효과 레버리지 그림을 생성하는 경우를 가정해 보겠습니다. 더 일반적인 상황에서 가로 축 척도에 대한 자세한 내용은 "[가로 축 척도화](#)"(162페이지)에서 확인하십시오.

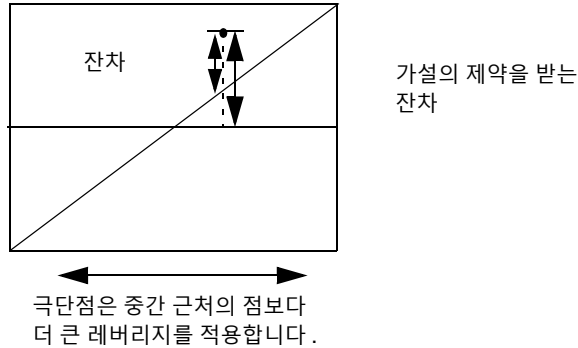
X를 제외한 모든 예측 변수에 대해 반응 Y가 회귀 분석되고 잔차가 계산됩니다. 이러한 잔차를 Y 잔차라고 합니다. 그런 다음 모형의 다른 모든 예측 변수에 대해 X가 회귀 분석되고 잔차가 계산됩니다. 이러한 잔차를 X 잔차라고 합니다. X 잔차에는 모형에서 X를 제외하고 계산된 Y 잔차에 있는 정보 외에 추가 정보가 포함될 수 있습니다.

기본적으로 X에 대한 효과 레버리지 그림은 Y 잔차에 대한 X 잔차의 산점도입니다(그림 3.57). JMP는 그림을 해석하고 다른 그림과 비교하는 데 도움이 되도록 Y의 평균을 Y 잔차에 추가하고 X의 평균을 X 잔차에 추가합니다. 변환된 Y 잔차를 Y 레버리지 잔차라고 하고 변환된 X 잔차를 X 레버리지 값이라고 합니다. 효과 레버리지 그림의 점은 이러한 X 레버리지와 Y 레버리지 잔차 쌍입니다.

JMP는 이러한 점과 평균에 대한 신뢰 대역에 최소 제곱선을 적합시킵니다. 적합선은 빨간색 실선이고 신뢰 대역은 빨간색으로 음영 처리됩니다. 최소 제곱선의 기울기는 Y가 X 및 다른 예측 변수에 대해 회귀 분석되는 모형에서 X에 대한 계수의 추정값입니다. 파란색 수평 파선은 Y 레버리지 잔차의 평균에 설정됩니다. 이 선은 X 잔차가 Y 잔차와 선형 관계가 없는 상황을 설명합니다. 적합선의 기울기가 0이 아닌 경우 모형에 X를 추가하면 변동을 설명하는 데 유용할 수 있습니다.

그림 3.55에서는 잔차가 레버리지 그림에 어떻게 표시되는지 보여 줍니다. 점에서 적합선까지의 거리는 효과를 포함하는 모형의 잔차입니다. 점에서 수평선까지의 거리는 모형에 효과가 없을 때의 잔차 오차입니다. 즉, 레버리지 그림의 평균 선은 모수(효과)의 가설 값이 0으로 제한된 모형을 나타냅니다.

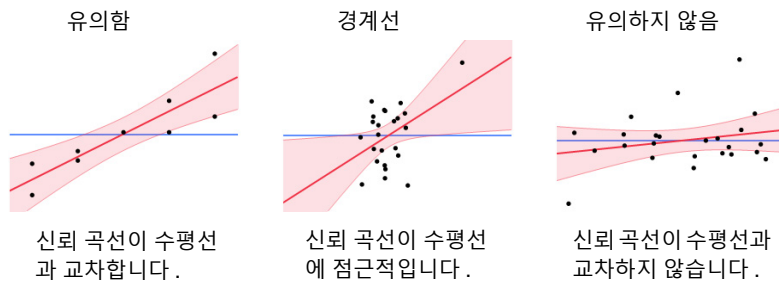
그림 3.55 일반 레버리지 그림의 예



신뢰 곡선

적합선에 대한 신뢰 곡선이 레버리지 그림에 표시됩니다. 이러한 곡선은 관심 검정이 5% 수준(또는 "모형 적합" 시작 창의 "유의 수준 설정"에 지정한 수준)에서 유의한지 여부를 시각적으로 보여 줍니다. 곡선 사이의 신뢰 영역에 가설을 나타내는 수평선이 포함되어 있으면 효과가 유의하지 않습니다. 곡선이 선과 교차하면 효과가 유의합니다. 그림 3.56의 예를 참조하십시오.

그림 3.56 레버리지 그림에 표시된 유의성 비교



가로 축 척도화

예측 변수 X의 모델링 유형이 연속형이면 가로 축이 X 단위로 척도화됩니다. 가로 축 범위는 X 값 범위를 반영합니다. 레버리지 그림에서 적합선 기울기는 X에 대한 모수 추정값입니다. 그림 3.57의 왼쪽 그림을 참조하십시오.

효과가 명목형 또는 순서형이거나, 교호작용과 같은 복합 효과인 경우 가로 축에 효과의 값을 직접 나타낼 수 없습니다. 이 경우 가로 축은 반응 단위로 척도화되고 적합선은 기울기가 1인 대각선입니다. "전체 모형" 레버리지 그림(모든 모수 값이 0이라는 가설이 지정됨)에는 이 척도화가 사용됩니다. 자세한 내용은 "[레버리지 그림 상세 정보](#)"(189페이지)에서 확인하십시오. 이 그림의 경우 가로 축은 그림 3.57의 오른쪽 그림에 표시된 것처럼 전체 모형에 대한 예측 반응 값을 기준으로 척도화됩니다.

단순 회귀 분석에서 선형 효과에 대한 레버리지 그림은 기존의 예측 변수에 대한 실제 반응 값 그림과 동일합니다.

레버리지

이러한 그림은 모형에 효과를 포함하기 위한 검정에서 점의 영향을 시각화하는 데 도움이 되므로 레버리지라는 용어가 사용됩니다. 그림의 중심에서 수평으로 멀리 떨어진 점은 중심에 가까운 점보다 효과 검정에 더 많은 영향을 미칩니다. 효과 검정에는 잔차 제곱합을 해당 효과가 제거된 모형의 잔차 제곱합과 비교하는 과정이 포함됩니다. 극단에서는 가설의 제약을 받기 전과 후의 잔차 차이가 비교적 큰 경향이 있습니다. 따라서 이러한 잔차는 해당 효과의 가설 검정에 대한 제곱합에 더 크게 기여합니다.

다중공선성

다중공선성은 두 개 이상의 예측 변수가 밀접하게 관련되어 있거나, 좀 더 엄밀히 말하면 거의 선형 종속 관계에 관련된 상태를 의미합니다. 다중공선성이 존재하면 표준 오차가 팽창되고 모수 추정값이 불안정해질 수 있습니다. 효과가 다른 예측 변수와 동일 선상에 있으면 점의 가로 값이 그림의 가운데 쪽으로 모이는 경향이 있습니다. 이 상황은 적합선의 기울기가 불안정함을 나타냅니다.

전체 모형 실제값 대 예측값 그림

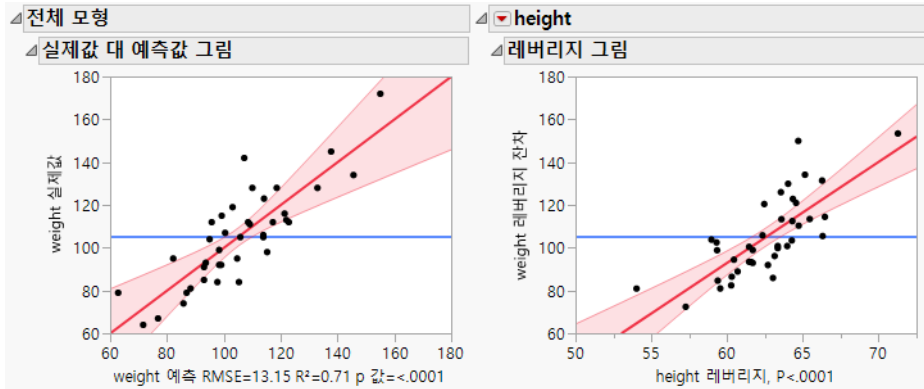
"효과 레버리지 그림" 옵션은 모형의 각 효과에 대한 레버리지 그림을 생성합니다. 실제값 대 예측값 그림도 레버리지 그림으로 간주할 수 있습니다. 이 그림을 사용하면 모형의 모든 모수(절편 제외)가 0이라는 검정을 시각화할 수 있습니다. 동일한 검정이 "분산 분석" 보고서에서 분석적으로 수행됩니다. 이 그림에 대한 자세한 내용은 "[레버리지 그림 상세 정보](#)"(189페이지)에서 확인하십시오.

선형 효과에 대한 레버리지 그림의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. **weight** 를 선택하고 **Y** 를 클릭합니다 .
4. **height, age** 및 **sex** 를 선택하고 **추가**를 클릭합니다 .
5. **실행**을 클릭합니다 .

그림 3.57에서는 전체 모형에 대한 "실제값 대 예측값 그림"과 height 효과에 대한 "레버리지 그림"을 보여 줍니다. 왼쪽의 전체 모형 그림에서는 모든 효과를 검정합니다. 신뢰 곡선이 반응(weight)의 평균에서 수평선과 교차하므로 모형이 유의하다는 것을 추론할 수 있습니다. 또한 오른쪽의 height에 대한 "레버리지 그림"에서는 모형에 age와 sex가 있더라도 height가 유의함을 보여 줍니다. 두 그림 모두 영향력 있는 점 또는 다중공선성과 관련된 문제를 나타내지 않습니다.

그림 3.57 전체 모형 및 효과 레버리지 그림



Press

Press 또는 예측 오차 제곱합 통계량은 leave-one-out 교차 검증을 사용하여 계산된 예측 오차의 추정값입니다. leave-one-out 교차 검증에서는 각 관측값이 차례로 제거됩니다. 특정 관측값을 예로 들겠습니다. 관측값을 주지 않고 모형을 적합시킨 후 이 관측값에 대한 예측값을 계산합니다. 해당 관측값에 대한 잔차가 계산됩니다. 이 절차가 모든 관측값에 적용되며 잔차의 제곱합이 계산되어 Press 값을 제공합니다.

Press 통계량은 다음과 같이 구합니다.

$$\text{Press} = \sum_{i=1}^n (\hat{y}_{(i)} - y_i)^2$$

여기서 n 은 관측값 수, y_i 는 i 번째 관측값의 반응 관측값, $\hat{y}_{(i)}$ 는 i 번째 관측값의 반응 예측값입니다. 이러한 값은 해당 관측값을 포함하지 않은 모형 적합을 기반으로 합니다.

Press RMSE는 $\sqrt{\text{Press}/n}$ 으로 정의됩니다.

Press R^2 은 $1 - \text{Press}/SS_{\text{Total}}$ 로 정의됩니다.

열 저장

각 "열 저장" 옵션은 하나 이상의 새 열을 현재 데이터 테이블에 추가합니다. 적합 모형이 REML 인 경우 추가 "열 저장" 옵션이 나타납니다. 자세한 내용은 ["REML 열 저장 옵션"](#) (180 페이지) 에서 확인하십시오.

다음 사항에 유의하십시오.

- 계산식이 생성되면 "계산식" 열 특성으로 입력됩니다.
- 대부분의 새 열에는 해당 열을 설명하고 최소 제곱 적합으로 열이 생성되었음을 나타내는 "노트" 열 특성이 추가됩니다.
- "예측 계산식" 및 "예측값" 옵션의 경우 "예측" 열 특성이 추가됩니다. 이 특성은 모형 비교 (분석 > 예측 모델링 > 모형 비교) 를 수행할 때 JMP 에서 내부적으로 사용됩니다. 또한 열의 출처를 기록하므로 여러 모형을 적합시키는 경우에도 유용합니다.

"열 저장" 에서 다음 옵션을 사용할 수 있습니다.

예측 계산식 계산식과 예측값이 모두 포함된 예측 계산식 <열 이름> 이라는 새 열을 생성합니다. "예측" 열 특성이 추가되어 예측 소스를 나타냅니다.

참고 : 예측 계산식 <열 이름> 은 <열 이름> 에서 특정한 특성을 상속합니다. 여기에는 반응 한계, 규격 한계 및 관리 한계가 포함됩니다. 예측 계산식 <열 이름> 을 저장한 후에는 <열 이름> 에 대한 이러한 특성을 변경해도 예측 계산식 <열 이름> 에서 업데이트되지 않습니다.

자세한 내용은 ["예측 계산식"](#) (167 페이지) 에서 확인하십시오.

예측값 지정된 모형에서 계산한 예측값이 포함된 예측값 <열 이름> 이라는 새 열을 생성합니다. "노트" 와 "예측" 열 특성이 모두 추가되어 예측 소스를 나타냅니다.

참고 : 예측값 <열 이름> 은 <열 이름> 에서 특정한 특성을 상속합니다. 여기에는 반응 한계, 규격 한계 및 관리 한계가 포함됩니다. 예측값 <열 이름> 을 저장한 후에는 <열 이름> 에 대한 이러한 특성을 변경해도 예측값 <열 이름> 에서 업데이트되지 않습니다.

잔차 반응 관측값에서 예측값을 뺀 값이 포함된 잔차 <열 이름> 이라는 새 열을 생성합니다.

평균 신뢰 구간 평균 <열 이름> 의 95% 하한 및 평균 <열 이름> 의 95% 상한이라는 두 개의 새 열을 생성합니다. 이 열에는 평균 반응에 대한 95% 신뢰 하한과 신뢰 상한이 포함되어 있습니다.

참고 : Shift 키를 누른 상태로 이 옵션을 선택하면 계산에 사용할 α 수준을 입력하는 대화상자가 나타납니다.

개별값 신뢰 구간 개별 <열 이름> 의 95% 하한과 개별 <열 이름> 의 95% 상한이라는 두 개의 새 열을 생성합니다. 이 열에는 개별 반응 값에 대한 95% 신뢰 하한과 신뢰 상한이 포함되어 있습니다.

참고 : Shift 키를 누른 상태로 이 옵션을 선택하면 계산에 사용할 α 수준을 입력하는 대화상자가 나타납니다.

스튜던트화 잔차 잔차를 표준 오차로 나눈 값이 포함된 스튜던트화 잔차 < 열 이름 > 이라는 새 열을 생성합니다.

외부 스튜던트화 잔차 (적합 방법이 REML 인 경우에는 사용 불가능) 잔차를 현재 행을 제외하는 표준 오차 추정값으로 나눈 값이 포함된 외부 스튜던트화 잔차 < 열 이름 > 이라는 새 열을 생성합니다. 자세한 내용은 "**스튜던트화 잔차 그림**"(160 페이지)에서 확인하십시오.

해트 값 h < 열 이름 > 이라는 새 열을 생성합니다. 열 값은 $X(X'X)^{-1}X'$ 행렬의 대각 값이며 hat 값이라고도 합니다.

예측값의 표준 오차 예측 평균 반응의 표준 오차가 포함된 예측 표준 오차 < 열 이름 > 이라는 새 열을 생성합니다.

잔차의 표준 오차 잔차 값의 표준 오차가 포함된 표준 오차 잔차 < 열 이름 > 이라는 새 열을 생성합니다.

개별값의 표준 오차 개별 예측 반응 값의 표준 오차가 포함된 표준 오차 개별 < 열 이름 > 이라는 새 열을 생성합니다.

효과 레버리지 쌍 각 레버리지 그림에 대한 X 레버리지 값과 Y 레버리지 잔차가 포함된 새 열 집합을 생성합니다. 모형의 각 효과에 대해 두 개의 열이 추가됩니다. 반응 열 이름이 R 이고 효과가 X 인 경우 새 열 이름은 다음과 같습니다.

- R 에 대한 X 의 X 레버리지
- R 에 대한 X 의 Y 레버리지

열 패널에서 이러한 열은 레버리지라는 열 그룹으로 구성됩니다.

Cook 거리 영향력 Cook 거리 영향력 통계량의 값이 포함된 Cook 거리 영향력 < 열 이름 > 이라는 새 열을 생성합니다.

예측 표준 오차 계산식 예측값의 표준 오차에 대한 계산식과 값이 모두 포함된 예측값 표준 오차 < 열 이름 > 이라는 새 열을 생성합니다.

참고 : 저장된 계산식이 클 수 있습니다. 계산식이 필요하지 않으면 "예측값의 표준 오차" 옵션을 사용합니다.

평균 신뢰 한계 계산식 평균 < 열 이름 > 의 95% 하한 및 평균 < 열 이름 > 의 95% 상한이라는 두 개의 새 열을 생성합니다. 이 열은 평균 반응에 대한 95% 신뢰 하한 및 상한의 계산식과 값이 모두 포함됩니다.

참고 : Shift 키를 누른 상태로 이 옵션을 선택하면 계산에 사용할 α 수준을 입력하는 대화상자가 나타납니다.

개별값 신뢰 한계 계산식 개별 < 열 이름 > 의 95% 하한과 개별 < 열 이름 > 의 95% 상한이라는 두 개의 새 열을 생성합니다. 이 열은 개별 반응 값에 대한 95% 신뢰 하한 및 상한의 계산식과 값이 모두 포함됩니다.

참고 : Shift 키를 누른 상태로 이 옵션을 선택하면 계산에 사용할 α 수준을 입력하는 대화상자가 나타납니다.

코딩 테이블 저장 첫 번째 열에 모든 모형 모수에 대한 JMP 코딩이 표시되는 새 데이터 테이블을 생성합니다. 마지막 열은 반응 값을 제공합니다. 반응 열을 두 개 이상 입력한 경우 이러한 모든 열이 코딩 테이블의 마지막 열로 나타납니다.

참고 : 코딩 데이터 테이블에는 분석에 사용된 데이터 테이블 이름을 제공하는 "원래 데이터"라는 테이블 변수가 포함되어 있습니다. 기준 변수가 지정된 경우에는 "원래 데이터" 테이블 변수에 기준 변수와 해당 수준이 제공됩니다.

JMP PRO 예측 계산식 게시 예측 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다. "계산식 저장소" 보고서가 열려 있지 않으면 "계산식 저장소" 보고서가 생성됩니다. 자세한 내용은 **Predictive and Specialized Modeling** 의 에서 확인하십시오.

JMP PRO 표준 오차 계산식 게시 표준 오차 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다. "계산식 저장소" 보고서가 열려 있지 않으면 "계산식 저장소" 보고서가 생성됩니다. 자세한 내용은 **Predictive and Specialized Modeling** 의 에서 확인하십시오.

JMP PRO 평균 신뢰 한계 계산식 게시 평균 반응에 대한 신뢰 한계 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다. "계산식 저장소" 보고서가 열려 있지 않으면 "계산식 저장소" 보고서가 생성됩니다. 자세한 내용은 **Predictive and Specialized Modeling** 의 에서 확인하십시오.

JMP PRO 개별 신뢰 한계 계산식 게시 개별 반응 값에 대한 신뢰 한계 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다. "계산식 저장소" 보고서가 열려 있지 않으면 "계산식 저장소" 보고서가 생성됩니다. 자세한 내용은 **Predictive and Specialized Modeling** 의 에서 확인하십시오.

예측 계산식

예측 계산식 <열 이름>은 예측 계산식을 포함하므로 예측값 <열 이름>과 다릅니다. 예측 계산식을 보려면 예측 계산식 <열 이름> 열 머리글을 마우스 오른쪽 버튼으로 클릭하고 **계산식**을 선택합니다. 모형이 크면 예측 계산식에 상당한 공간이 필요할 수 있습니다. 예측값 열에 계산식이 필요하지 않으면 **열 저장 > 예측값** 옵션을 사용합니다. 계산식에 대한 자세한 내용은 **JMP 사용** 의 에서 확인하십시오.

"예측 계산식" 옵션은 새 행의 값을 예측하거나 프로파일러와 함께 사용하는 데 유용합니다. 프로파일러는 "최소 제곱 적합" 보고서(요인 프로파일링)에서 사용할 수 있습니다. 그러나 데이터 테이블에 계산식 열이 포함된 경우 **그래프** 메뉴에 제공된 프로파일러를 사용할 수도 있습니다. 여러 반응을 분석하는 경우 **그래프** 메뉴에서 프로파일러에 액세스하는 것이 유용할 수 있습니다.

참고 : **그래프 > 프로파일러**를 선택하여 프로파일러에 액세스하는 경우 먼저 "예측 계산식" 및 "예측 표준 오차 계산식"을 사용하여 계산식 열을 데이터 테이블에 저장하십시오. 그런 다음 이 두 계산식을 "프로파일러" 창의 "Y, 예측 계산식" 역할에 배치합니다. **확인**을 클릭한 후 "예측값 표준 오차 <열 이름>"을 사용하여 "예측 계산식 <열 이름>"에 대한 신뢰 구간을 생성할지 여부를 지정합니다. 그렇지 않으면 JMP에서 "예측값 표준 오차 <열 이름>"에 대한 별도의 프로파일러 그림을 생성합니다.

효과 요약 보고서

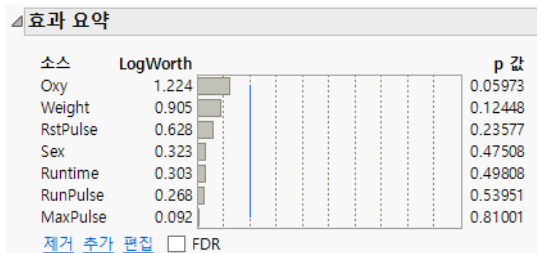
"효과 요약" 옵션은 대화식 보고서를 표시합니다. 이 보고서에는 모형의 효과에 대한 LogWorth 또는 FDR LogWorth 값 그림이 제공됩니다. 모형에서 효과를 추가하거나 제거할 수 있는 컨트롤도 보고서에 제공됩니다. 모형 적합 보고서는 "효과 요약" 보고서의 변경 사항에 따라 자동으로 업데이트됩니다.

"효과 요약" 보고서는 다음 분석법에서 사용할 수 있습니다.

- 표준 최소 제곱
- 명목형 로지스틱
- 순서형 로지스틱
- 비례 위험
- 모수 생존
- 일반화 선형 모형

그림 3.58에서는 Fitness.jmp 데이터 테이블에 대한 "효과 요약" 보고서의 초기 보기를 표시합니다. **FDR** 라벨이 지정된 체크박스는 요약 테이블에 나타나는 열을 제어합니다.

그림 3.58 효과 요약 보고서



효과 요약 테이블 열

"효과 요약" 테이블에는 다음과 같은 열이 포함되어 있습니다.

소스 모형 효과를 p 값이 작은 것부터 순서대로 나열합니다.

LogWorth 각 모형 효과에 대해 $-\log_{10}(p \text{ 값})$ 으로 정의된 LogWorth 를 표시합니다 . 이 변환은 그래프 작성에 적절한 척도를 제공하도록 p 값을 조정합니다 . $-\log_{10}(0.01) = 2$ 이므로 2 를 초과하는 값은 0.01 수준에서 유의합니다 .

FDR LogWorth 각 모형 효과에 대해 $-\log_{10}(\text{FDR } p \text{ 값})$ 으로 정의된 FDR(False Discovery Rate) LogWorth 를 표시합니다 . 이 통계량은 유의성을 그림에 표시하고 평가하는 데 가장 적합합니다 . 그러나 유의성 순서에 따라 크게 좌우되고 , 양의 상관관계가 있는 검정에 대해 보수적이며 , 유의 수준에서 실험별 보호를 제공하지 않습니다 . LogWorth 열을 **FDR LogWorth** 열로 바꾸려면 **FDR** 체크박스를 선택합니다 .

막대 그래프 LogWorth 또는 FDR LogWorth 값의 막대 그래프를 표시합니다 . 이 그래프에는 정수 값에 수직 파선이 있고 2 에 파란색 참조선이 있습니다 .

p 값 각 모형 효과에 대한 p 값을 표시합니다 . 일반적으로 이 값은 모형 보고서의 " 효과 검정 " 테이블 또는 " 효과 가능도비 검정 " 테이블에 표시된 유의성 검정에 해당하는 p 값입니다 .

FDR p 값 각 모형 효과에 대해 Benjamini-Hochberg 기법을 사용하여 계산된 FDR(False Discovery Rate) p 값을 표시합니다 . 이 기법은 여러 검정에 대한 FDR(False Discovery Rate) 을 제어하기 위해 p 값을 조정합니다 . **p 값** 열을 **FDR p 값** 열로 바꾸려면 **FDR** 체크박스를 선택합니다 .

FDR 수정에 대한 자세한 내용은 Benjamini and Hochberg 연구 자료 (1995) 에서 확인하십시오 . FDR(False Discovery Rate) 에 대한 자세한 내용은 Predictive and Specialized Modeling 의 또는 Westfall et al. 연구 자료 (2011) 에서 확인하십시오 .

효과 유전성 열 유의성이 높은 고차 효과의 성분인 저차 효과를 식별합니다 . 저차 효과는 캐럿으로 식별됩니다 . 자세한 내용은 " 효과 유전성 "(171 페이지) 에서 확인하십시오 .

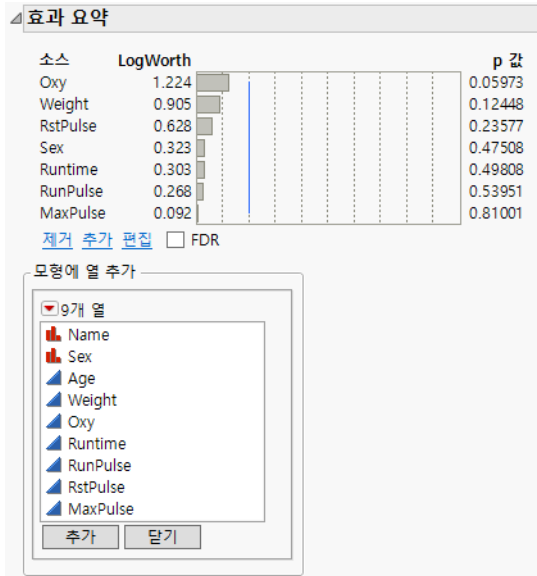
효과 요약 테이블 옵션

요약 테이블 아래의 옵션을 사용하여 효과를 추가 및 제거할 수 있습니다 .

제거 선택한 효과를 모형에서 제거합니다 . 하나 이상의 효과를 제거하려면 대상 효과에 해당하는 행을 선택하고 **제거** 버튼을 클릭합니다 .

추가 데이터 테이블의 모든 열 목록이 포함된 패널을 엽니다 . 모형에 추가할 열을 선택한 후 열 선택 목록 아래의 **추가**를 클릭하면 해당 열이 모형에 추가됩니다 . 패널을 닫으려면 **닫기**를 클릭합니다 . 그림 3.59 에서는 열 추가 패널을 보여 줍니다 .

그림 3.59 효과 요약 열 추가 패널



편집 "열 선택" 목록과 "효과" 지정 패널이 포함된 "모형 편집" 패널을 엽니다. "효과" 패널은 "모형 적합" 시작 창의 "모형 효과 생성" 패널과 비슷합니다. "모형 편집" 패널을 사용하여 개별, 교차, 내포 및 변환 효과를 추가할 수 있습니다. "매크로" 메뉴를 사용하여 여러 효과를 추가할 수도 있습니다. "추가", "교차", "내포", "매크로" 및 "변환"을 사용하여 효과를 생성하는 방법에 대한 자세한 내용은 "[모형 효과 생성](#)"(37페이지)에서 확인하십시오.

"효과" 패널 오른쪽의 다음 옵션을 사용할 수 있습니다.

커밋 업데이트를 모형에 적용합니다.

닫기 모형을 변경하지 않고 패널을 닫습니다.

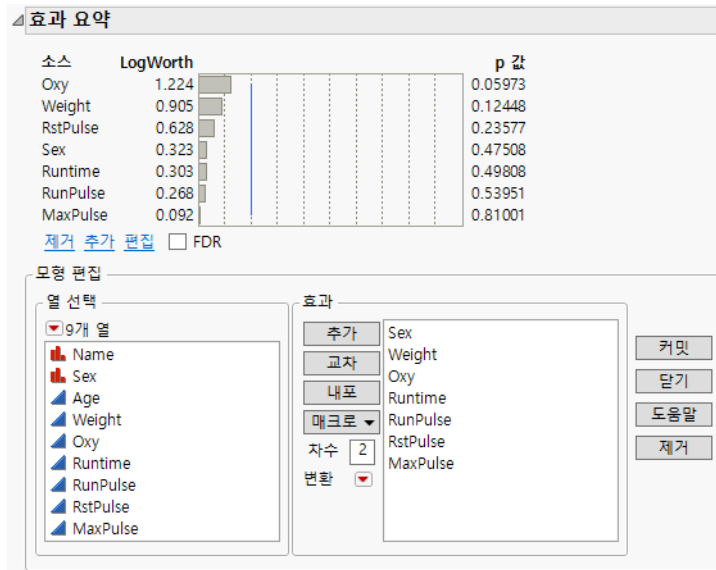
제거 "효과" 목록에서 선택한 하나 이상의 효과를 제거합니다.

그림 3.60에서는 "모형 편집" 패널을 보여 줍니다.

팁 : "편집" 버튼을 사용하면 모형에 대한 업데이트를 최대한 제어할 수 있습니다. "제거" 및 "추가" 버튼의 기능이 포함되고 모형에 추가할 효과를 생성할 수 있습니다.

실행 취소 모형의 효과에 대한 변경 사항을 취소할 수 있습니다.

그림 3.60 효과 요약 모형 편집 패널



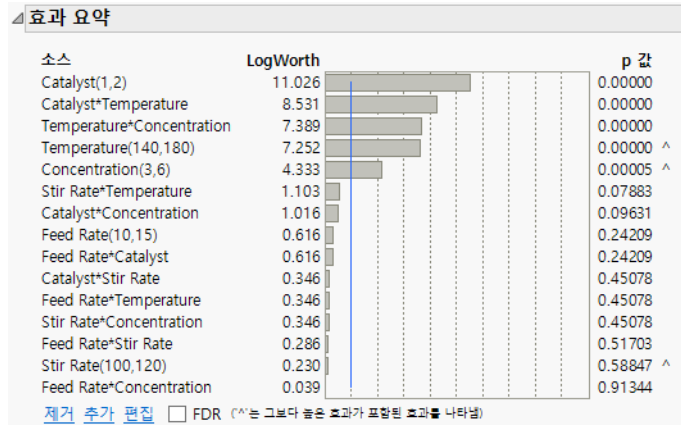
효과 유전성

모형에 유의한 고차 효과가 포함된 경우 일부 또는 모든 저차 성분이 유의하지 않더라도 유지할 수 있습니다. 강한 효과 유전성 원리에 따르면 모형에 고차 효과가 포함되면 모든 저차 성분도 포함되어야 합니다. 약한 효과 유전성 원리는 성분 체인이 포함되어야 함을 나타냅니다.

고차 효과의 저차 성분이 "효과 요약" 테이블에서 고차 효과 아래에 나타나면 맨 오른쪽 열에 캐럿이 표시됩니다. 캐럿은 포함하는 고차 효과가 저차 효과보다 유의하다는 것을 나타냅니다. 저차 효과를 포함하는 모든 고차 효과가 저차 효과보다 유의성이 낮으면 저차 효과 행에 캐럿이 나타나지 않습니다.

캐럿으로 표시된 효과를 제거하는 경우 두 가지 효과 제거 방법 중 하나를 선택할 수 있습니다. 캐럿으로 표시된 효과를 포함하여 선택한 모든 효과를 제거하려면 **선택한 모든 효과 제거**를 선택합니다. 모형에 계속 유지되는 고차 효과를 제외하고 선택한 효과만 제거하려면 **비포함 효과만 제거**를 선택합니다.

그림 3.61에서는 저차 효과를 포함하는 고차 효과 아래에 세 가지 저차 효과가 표시된 "효과 요약" 테이블의 예를 보여 줍니다. 예를 들어 Stir Rate(100,120)가 Stir Rate*Temperature 아래에 나타납니다.

그림 3.61 Reactor 32 Runs.jmp 에 대해 효과 유전성이 포함된 효과 요약 테이블

다중 반응

다중 반응의 경우 각 반응 모형에 각 효과가 나타나지만 "효과 요약" 보고서는 하나만 나타냅니다. 테이블에는 각 효과에 대해 해당 효과의 p 값 중 최소 p 값이 표시됩니다. 효과를 추가하거나 제거하면 모든 반응에 대한 모형에 적용됩니다.

혼합 및 임의 효과 모형 보고서 및 옵션

혼합 및 임의 효과 모형은 모형 적합 시작 창에서 지정할 수 있습니다. 표준 최소 제곱 분석법은 분산 성분 공분산 구조를 적합시킵니다. 이러한 모형을 적합시키기 위해 REML 과 EMS 라는 두 가지 방법이 제공됩니다.

참고 : **JMP PRO** JMP Pro 사용자는 모형 적합 창에서 혼합 모형 분석법을 사용하는 것이 좋습니다. 혼합 모형 분석법은 표준 최소 제곱 분석법에서 제공하는 것보다 광범위한 공분산 구조 집합을 제공합니다.

- " 혼합 모형 및 임의 효과 모형 "
- "REML(제한 최대 가능도) 방법 "
- "EMS(전통적) 모형 적합 보고서 "

혼합 모형 및 임의 효과 모형

임의 효과 모형은 모든 요인이 임의 효과를 나타내는 모형입니다. 자세한 내용은 "임의 효과"(173 페이지)에서 확인하십시오. 이러한 모형을 분산 성분 모형이라고도 합니다. 대개 임의 효과 모형은 계층적 모형입니다. 고정 효과와 임의 효과가 모두 포함된 모형을 혼합 모형이라고 합니다.

반복 측정 및 분할구 모형은 혼합 모형의 특수한 경우입니다. 혼합 모형이라는 용어는 임의 효과 모형을 포함하기 위해 사용되는 경우가 많습니다.

혼합 모형을 적합시키려면 모형 적합 시작 창에서 임의 효과를 지정해야 합니다. 그러나 모든 모형 효과가 임의 효과인 경우에는 계량형/계수형 게이지 차트 플랫폼에서 모형을 적합시킬 수도 있습니다. 이 플랫폼에서는 특정 모형만 적합시킬 수 있습니다. 계량형/계수형 게이지 차트 플랫폼에서 사용되는 적합 방법은 분산 성분 추정값에 음수를 허용하지 않습니다. 계량형/계수형 게이지 차트 플랫폼에서 분산 성분 모형을 적합시키는 방법에 대한 자세한 내용은 **Quality and Process Methods**의 에서 확인하십시오.

임의 효과

임의 효과는 수준이 일부 모집단의 랜덤 표본으로 간주되는 요인입니다. 대개 관심 대상은 임의 효과의 정확한 수준이 아니라 수준에 의해 반영되는 변동(분산 성분)입니다. 그러나 임의 효과의 특정 수준에 대한 반응을 예측하려는 상황도 있습니다. 기술적으로 임의 효과는 평균이 0이고 분산이 0이 아닌 정규 분포를 따르는 것으로 간주됩니다.

두 개의 특정 오븐이 금형 수축에 미치는 영향이 다른지 여부에 관심이 있다고 가정해 보겠습니다. 오븐은 한 번에 하나의 배치(50개 금형)만 처리할 수 있습니다. 한 배치당 50개씩 무작위로 선택한 세 개의 금형 배치를 두 오븐에 각각 연속적으로 넣어 연구를 설계합니다. 배치가 처리되면 각 배치에서 무작위로 선택한 5개 부품에 대해 수축이 측정됩니다.

Batch는 각 배치에 대해 하나씩 6개 수준이 있는 요인입니다. 따라서 모형에는 Oven과 Batch라는 두 가지 요인이 포함됩니다. 특히 각 오븐이 수축에 미치는 영향을 비교하는 데 관심이 있으므로 Oven은 고정 효과입니다. 그러나 특정 6개의 배치가 평균 수축에 미치는 영향에는 관심이 없습니다. 이러한 배치는 이 실험에서 선택할 수 있고 분석 결과를 일반화해야 하는 배치의 전체 모집단을 나타냅니다. Batch는 임의 효과로 간주됩니다. 가능한 모든 배치에서 수축 변동을 연구하는 이 실험에서 Batch 요인은 관심 대상입니다. 요인이 설명하는 수축의 변동량을 추정하는 데 관심이 있습니다. 또한 한 오븐에서 하나의 배치만 처리할 수 있으므로 Batch가 Oven에 내포됩니다.

이제 두 가지 사료 체계가 적용된 암탉의 알 무게에 관심이 있다고 가정해 보겠습니다. 10마리 암탉을 무작위로 선택하여 5마리는 사료 체계 A에 할당하고 5마리는 사료 체계 B에 할당합니다. 그러나 10마리의 암탉은 연구 설계에서 고려되지 않은 유전적 차이가 있습니다. 이 경우 특정 암탉이 낳은 달걀의 예측 무게 및 전체 암탉 간의 달걀 무게에 대한 분산에 관심이 있습니다.

전통적 선형 혼합 모형

JMP에서는 다음과 같이 전통적 선형 혼합 효과 모형을 적합시킵니다.

$$\begin{aligned} Y &= X\beta + Z\gamma + \epsilon \\ \gamma &\sim N(0, G) \\ \epsilon &\sim N(0, \sigma^2 I_n) \end{aligned}$$

다음은 각 요소에 대한 설명입니다.

- Y - 반응의 $n \times 1$ 벡터
- X - 고정 효과에 대한 $n \times p$ 설계 행렬
- β - 설계 행렬 X 를 사용하는 알려지지 않은 고정 효과의 $p \times 1$ 벡터
- Z - 임의 효과에 대한 $n \times s$ 설계 행렬
- γ - 설계 행렬 Z 를 사용하는 알려지지 않은 임의 효과의 $s \times 1$ 벡터
- ε - 알 수 없는 랜덤 오차의 $n \times 1$ 벡터
- G - 임의 효과의 각 수준에 대해 동일한 항목이 있는 $s \times s$ 대각 행렬
- I_n - $n \times n$ 항등 행렬
- γ 와 ε 는 독립적 관계

G 의 대각 요소와 σ^2 을 분산 성분이라고 합니다. 이러한 분산 성분은 고정 효과 벡터(β) 및 임의 효과 벡터(γ)와 함께 추정 대상인 모형 모수입니다.

경우에 따라 이 모형의 공분산 구조를 분산 성분 구조라고도 합니다(SAS Institute Inc. 2020d, ch. 83 이 공분산 구조만 표준 최소 제곱 분석법에서 사용할 수 있습니다.

JMP PRO 혼합 모형 분석법은 잔차, AR(1)(1차 자기회귀), 비정형 및 공간을 포함한 다양한 공분산 구조를 적합시킵니다. 자세한 내용은 "혼합 모형" 장의 "**반복 구조 탭**"(339페이지)에서 확인하십시오.

임의 효과가 있는 모형의 적합 방법 (REML, EMS)

JMP에서는 임의 효과가 있는 모형을 적합시키기 위해 다음 두 가지 방법을 제공합니다.

- REML - 제한 최대 가능도 (항상 권장되는 방법)
- EMS - 기대 평균 제곱 (이전 교재로 교육할 때만 사용)

REML 방법은 기존의 EMS 방법을 대체하는 주요 적합 방법입니다. REML은 적용 가능성 면에서 EMS 방법보다 훨씬 더 일반적입니다. REML 방법은 Patterson and Thompson 연구 자료 (1974)에서 최초로 공개되었습니다. Wolfinger et al. 연구 자료 (1994) 및 Searle et al. 연구 자료 (1992)도 참조하십시오.

EMS 방법은 강력한 성능을 가진 컴퓨터가 나오기 전에 개발되었으며 **적률 방법**이라고도 합니다. 연구원들은 균형 잡힌 상황이 되도록 제한했고, 임의 효과 및 혼합 모형에 대한 추정값을 쉽게 계산할 수 있는 EMS 방법을 사용했습니다. 요즘도 임의 효과가 포함된 모형을 시작할 때 EMS 방법을 사용하는 교재가 많기 때문에 JMP에서 EMS 옵션을 제공합니다. 자세한 내용은 McCulloch et al., 2008, Poduri, 1997, Searle et al., 1992 등에서 확인하십시오.

REML 방법은 고정 효과 모수에 종속되지 않는 제한 가능도 함수의 최대 가능도 추정을 수행합니다. 이렇게 하면 분산 성분의 추정값이 생성되어 고정 효과의 추정값을 구하는 데 사용됩니다. 정밀도 추정값은 모수에 대한 공분산 행렬의 추정값을 기반으로 합니다. 데이터가 불균형한 경우에도 REML은 유용한 추정값, 검정 및 신뢰 구간을 제공합니다.

EMS 방법은 관측 평균 제곱을 기대 평균 제곱에 일치시켜 분산 성분의 추정값을 계산합니다. 균형 설계의 경우 추정값을 구하는 방법을 지정하는 복잡한 규칙 집합을 사용합니다. 이 기법을 불균형 데이터에 적용하면 문제가 있습니다.

균형 데이터의 경우 REML 추정값은 EMS 추정값과 동일합니다. 그러나 EMS와 달리 REML은 불균형 데이터에서도 정상적으로 수행됩니다.

임의 효과 및 적합 방법 지정

임의 효과가 있는 모형은 모형 적합 시작 창에서 지정됩니다. 임의 효과를 지정하려면 "모형 효과 생성" 목록에서 해당 효과를 강조 표시하고 **속성 > 임의 효과**를 선택합니다. 그러면 모형 효과 목록의 효과 이름에 "& 랜덤"이 추가됩니다. 임의 효과의 정의는 "**임의 효과**"(173 페이지)에서 확인하십시오. 임의 효과를 별도의 효과 탭에서 지정할 수도 있습니다. 자세한 내용은 "모형 규칙" 창의 "**모형 효과 생성 탭**"(42 페이지)에서 확인하십시오.

"모형 적합" 시작 창에서 효과에 "& 랜덤" 속성이 추가되면 "REML(권장)" 또는 "EMS(전통적)" 적합 방법을 선택할 수 있습니다.

주의: 교차 및 내포 관계를 명시적으로 선언해야 합니다. 예를 들어 각 개체가 하나의 그룹에만 있는 경우 개체 ID는 해당 개체를 포함하는 그룹을 식별할 수도 있습니다. 이러한 경우 개체 ID를 그룹에 내포된 것으로 선언해야 합니다. 설계 구조를 명시적으로 정의할 수 있도록 주의하십시오.

분산 성분에 대한 무제한 파라미터화

분산 성분을 파라미터화하는 방식에는 두 가지가 있습니다(무제한 및 제한 방식). 고정 효과와 임의 효과의 교호작용 같은 혼합 효과가 모형에 있을 때 이 문제가 발생합니다. 이러한 교호작용 항은 임의 효과로 간주됩니다.

제한된 방식의 경우 임의 효과의 각 수준에 대해 고정 효과의 전체 수준에서 교호작용 효과의 합이 0으로 간주됩니다. 무제한 방식의 경우 혼합 항은 평균 0과 공통 분산을 사용하는 정규 분포의 독립된 랜덤 구현으로 간주됩니다. 이 가정은 일반적으로 잔차 오차에 적용되는 가정과 비슷합니다.

JMP와 SAS에서는 무제한 방식을 사용합니다. 대부분의 통계학 교재에는 제한된 방식이 사용되므로 이를 구분하는 것이 중요합니다. 두 가지 방식 모두 60년 동안 광범위하게 교육에 사용되었습니다. 두 가지 방식에 대한 자세한 내용은 Cobb 연구 자료(1998, Section 13.3)에서 확인하십시오.

임의 분산

분산은 항상 양수이지만 분산의 비편향 추정값이 음수인 상황이 있을 수 있습니다. 효과가 매우 약하거나, 분산 성분에 해당하는 수준이 매우 적으면 실험에서 음수 추정값이 발생할 수 있습니다. 의도하지 않았지만 관측된 데이터에서 음수 추정값이 생성될 수 있습니다.

비유계 분산 성분

JMP에서는 REML 과 EMS 둘 다에 대해 음수 추정값을 생성할 수 있습니다. REML 의 경우 "모형 적합" 시작 창에 "비유계 분산 성분"과 "분산 성분만 추정"이라는 두 가지 옵션이 있습니다. 기본적으로 "비유계 분산 성분" 옵션이 선택되어 있습니다. 이 옵션을 선택 취소하면 분산 성분 추정값이 음수가 아닌 값으로 제한됩니다.

고정 효과에 관심이 있는 경우 "비유계 분산 성분" 옵션을 선택된 상태로 두어야 합니다. 분산 추정값을 음수가 아닌 값으로 제한하면 고정 효과 검정에서 편향이 발생합니다.

분산 성분만 추정

"REML 분산 성분 추정값" 보고서만 표시하려면 이 옵션을 선택합니다. 분산 성분에만 관심이 있는 경우 분산 성분을 음수가 아닌 값으로 제한할 수 있습니다. "비유계 분산 성분" 옵션을 선택 취소하고 "분산 성분만 추정" 옵션을 선택하는 것이 적절할 수 있습니다.

REML(제한 최대 가능도) 방법

선택한 적합 방법에 따라 "최소 제곱 적합" 보고서에 서로 다른 분석 결과가 제공되고 "열 저장" 및 "프로파일러"에 대한 추가 메뉴 옵션이 제공됩니다. 특히 분산과 자유도의 경우 일반적인 방식으로 분할되지 않으므로 분산 분석 보고서가 표시되지 않습니다. "REML 분산 성분 추정값" 보고서에서 잔차 분산 추정값을 얻을 수 있습니다. 자세한 내용은 "[REML 분산 성분 추정값](#)"(178 페이지)에서 확인하십시오. "효과 검정" 보고서는 고정 효과가 검정되는 "고정 효과 검정" 보고서로 대체됩니다. 추가 보고서는 임의 효과에 대한 예측값과 분산 성분에 대한 상세 정보를 제공합니다.

그림 3.62에서는 REML 방법을 사용하여 생성된 Investment Castings.jmp 샘플 데이터 적합 보고서를 보여 줍니다. **Model: REML** 스크립트를 실행한 후 모형을 적합시킵니다. Casting은 임의 효과이며 Temperature에 내포되어 있습니다.

그림 3.62 REML 방법에 대한 최소 제곱 적합 보고서

반응 Shrinkage

모수 추정값

임의 효과 예측

항	BLUP	표준 오차	DFDen	t 비	Prob> t
Temperature[1]:Casting[1]	-0.353959	0.337993	7.939	-1.05	0.3258
Temperature[1]:Casting[3]	0.0524938	0.337993	7.939	0.16	0.8805
Temperature[1]:Casting[5]	0.5139011	0.337993	7.939	1.52	0.1672
Temperature[1]:Casting[7]	-0.212436	0.337993	7.939	-0.63	0.5473
Temperature[2]:Casting[2]	0.4367848	0.337993	7.939	1.29	0.2326
Temperature[2]:Casting[4]	-0.094099	0.337993	7.939	-0.28	0.7878
Temperature[2]:Casting[6]	-0.862231	0.337993	7.939	-2.55	0.0343*
Temperature[2]:Casting[8]	0.519546	0.337993	7.939	1.54	0.1631

REML 분산 성분 추정값

임의 효과	분산 비율	분산 성분	표준 오차	95% 하한	95% 상한	Wald p 값	총계 백분율
Casting[Temperature]	1.511973	0.3205234	0.2161944	-0.10321	0.7442566	0.1382	60.191
잔차		0.2119901	0.0611963	0.1292489	0.4102654		39.809
합계		0.5325135	0.2204825	0.2716915	1.4771758		100.000

-2*로그 가능성도 = 57.247690262

참고: 합계는 양의 분산 성분의 합입니다.

음수 추정값을 포함한 합계 = 0.5325135

분산 성분 추정값의 공분산 행렬

반복

고정 효과 검정

소스	N모수	DF	DFDen	F 비	Prob > F
Temperature	1	1	6	6.5130	0.0434*

임의 효과 예측

이 보고서에는 모형의 각 항에 대해 BLUP(최량 선형 비편향 예측 변수)의 경험적 추정값 및 해당 계수가 0인지 여부에 대한 검정이 제공됩니다.

참고: "임의 효과 예측" 보고서를 표시하려면 회귀 보고서 > 모수 추정값 옵션을 선택해야 합니다.

항 모형에서 임의 효과에 해당하는 항입니다.

BLUP 각 임의 효과에 대한 BLUP(최량 선형 비편향 예측 변수)의 경험적 추정값입니다. 자세한 내용은 "최량 선형 비편향 예측 변수"(178 페이지)에서 확인하십시오.

표준 오차 BLUP의 표준 오차입니다.

DFDen 효과가 0인지에 대한 검정의 분모 자유도입니다. 대부분의 경우 *t*-검정의 자유도는 분수입니다.

t 비 효과가 0인지 검정하기 위한 *t* 비입니다. *t* 비는 BLUP를 표준 오차로 나눈 값입니다.

Prob>|t| 검정의 *p* 값입니다.

95% 하한 BLUP에 대한 95% 신뢰 하한입니다. 이 열은 회귀 보고서 > 모든 신뢰 구간 표시 옵션을 선택했거나, 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 95% 하한을 선택한 경우에만 나타납니다.

95% 상한 BLUP에 대한 95% 신뢰 상한입니다. 이 열은 회귀 보고서 > 모든 신뢰 구간 표시 옵션을 선택했거나, 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 95% 상한을 선택한 경우에만 나타납니다.

최량 선형 비편향 예측 변수

BLUP(최량 선형 비편향 예측 변수)는 임의 효과의 추정량을 나타냅니다. 구체적으로 모든 비편향 추정량 중에서 평균 제곱 예측 오차를 최소화하는 추정량입니다. "임의 효과 예측" 보고서에는 BLUP 또는 경험적 BLUP의 추정값이 제공됩니다. BLUP는 알려지지 않은 분산 성분의 값에 의존하므로 경험적입니다. 분산 성분의 추정값이 BLUP 계산식으로 대체되어 보고서에 표시된 추정값이 생성됩니다.

REML 분산 성분 추정값

"모형 적합" 시작 창에서 "REML"을 적합 방법으로 선택하면 "REML 분산 성분 추정값" 보고서가 제공됩니다. 이 보고서에는 다음 열이 포함됩니다.

임의 효과 모형의 임의 효과입니다.

분산 비율 잔차의 분산 성분에 대한 효과의 분산 성분 비율입니다. 이 값은 효과의 추정 분산을 모형의 추정 오차 분산과 비교합니다.

분산 성분 효과에 대한 추정 분산 성분입니다. "합계"의 분산 성분은 양의 분산 성분만 합한 값입니다. 모든 분산 성분의 합은 테이블 아래에 제공됩니다.

표준 오차 분산 성분 추정값에 대한 표준 오차입니다.

95% 하한 분산 성분에 대한 95% 신뢰 하한입니다. 자세한 내용은 "[분산 성분에 대한 신뢰 구간](#)"(178 페이지)에서 확인하십시오.

95% 상한 분산 성분에 대한 95% 신뢰 상한입니다. 자세한 내용은 "[분산 성분에 대한 신뢰 구간](#)"(178 페이지)에서 확인하십시오.

Wald p 값 공분산 모수가 0이라는 검정에 대한 p 값입니다. 이 열은 "모형 적합" 시작 창에서 "비유계 분산 성분"을 선택한 경우에만 나타납니다.

총계 백분율 합계의 분산 성분에 대한 효과의 분산 성분 비율을 백분율로 나타낸 값입니다.

분산 성분 제곱근 (표준편차) 해당 분산 성분의 제곱근입니다. 이 값은 효과에 대한 표준편차의 추정값입니다. 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 분산 성분 제곱근 (표준편차)을 선택한 경우에만 나타납니다.

CV 분산 성분에 대한 변동 계수입니다. 이 값은 분산 성분의 제곱근에 100을 곱하여 평균 반응으로 나눈 값입니다. 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > CV를 선택한 경우에만 나타납니다.

표준 KHC Kackar-Harville 수정입니다. 자세한 내용은 "[Kackar-Harville 수정](#)"(179 페이지)에서 확인하십시오. 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > 표준 KHC를 선택한 경우에만 나타납니다.

분산 성분에 대한 신뢰 구간

신뢰 한계를 계산하는 데 사용되는 방법은 "모형 적합" 시작 창에서 "비유계 분산 성분"을 선택했는지 여부에 따라 다릅니다. "비유계 분산 성분"은 기본적으로 선택되어 있습니다.

- "비유계 분산 성분" 을 선택한 경우 Wald 기반 신뢰 구간이 계산됩니다 . 이러한 구간은 점근적으로 유효하지만 표본이 작은 경우에는 신뢰할 수 없습니다 .
- "비유계 분산 성분"을 선택하지 않은 경우, 즉 모수의 하한 제약 조건이 0인 경우에는 Satterthwaite 근사가 사용됩니다 (Satterthwaite 1946).

Kackar-Harville 수정

REML 방법에서 고정 효과의 표준 오차는 분산 성분 추정값을 사용하여 추정됩니다. 그러나 이러한 추정값의 변동성을 고려하지 않으면 표준 오차가 과소추정됩니다. 증가한 변동을 설명하기 위해 Kackar-Harville 수정(Kackar and Harville 1984, Kenward and Roger 1997)을 사용하여 고정 효과의 공분산 행렬이 조정됩니다. 고정 효과의 공분산 행렬을 포함하는 모든 계산에 이 수정이 사용됩니다. 여기에는 최소 제곱 평균, 고정 효과 검정, 신뢰 구간 및 예측 분산이 포함됩니다. 통계 상세 정보는 "[Kackar-Harville 수정](#)"(192 페이지)에서 확인하십시오.

표준 KHC는 Kackar-Harville 수정의 Frobenius(행렬) norm입니다. 균형 잡힌 설계의 경우 표준 KHC가 작은 경향이 있습니다.

분산 성분 추정값의 공분산 행렬

이 보고서에는 분산 성분에 대한 점근 공분산 행렬의 추정값이 제공됩니다. 이 값은 관측된 Fisher 정보 행렬의 역입니다.

반복

G의 분산 성분 및 σ^2 의 추정값은 이러한 모수에만 의존하는 잔차 로그 가능도 함수를 최대화하여 구합니다. 반복 절차에서는 잔차 로그 가능도 함수를 최대화하거나, 이와 동등하게 음의 잔차 로그 가능도의 두 배 (-2* 로그 가능도)를 최소화하려고 합니다. "반복" 보고서에는 이 절차에 대한 상세 정보가 제공됩니다.

반복 반복 번호입니다.

-2* 로그 가능도 음의 로그 가능도의 두 배입니다. 이것은 목적 함수입니다. 자세한 내용은 "통계 상세 정보" 부록의 "[가능도, AICc 및 BIC](#)"(514 페이지)에서 확인하십시오.

그라디언트 norm 목적 함수의 그라디언트(1차 도함수) norm입니다.

모수 "모수"라벨이 지정된 열과 나머지 열은 각각 임의 효과에 해당합니다. 열 순서는 "REML 분산 성분 추정값" 보고서에서 임의 효과가 나열된 순서를 따릅니다. 각 반복에서 열의 값은 해당 시점에서 분산 성분의 추정값입니다.

수렴 기준은 기본 공차가 10^{-8} 인 그라디언트를 기반으로 합니다. "모형 적합" 시작 창에서 수렴 설정 > 수렴 한계 옵션을 선택한 후 원하는 공차를 지정하여 기준을 변경할 수 있습니다.

고정 효과 검정

REML 을 사용하는 경우 "효과 검정" 보고서에 고정 효과에 대한 검정이 제공됩니다. 이 보고서에는 다음 열이 포함됩니다.

소스 모형의 고정 효과입니다.

모수 수 효과와 관련된 모수의 수입니다.

DF 효과와 관련된 자유도입니다.

DFDen 분모 자유도입니다. 이 값은 Kenward-Roger 수정을 사용하여 공분산 행렬을 조정할 때 얻은 통계량 분포에 대한 근사를 기반으로 합니다. 자세한 내용은 "[Kackar-Harville 수정](#)"(179 페이지) 및 "[임의 효과](#)"(76 페이지)에서 확인하십시오.

F 비 계산된 F 비입니다.

Prob > F 효과 검정의 p 값입니다.

REML 열 저장 옵션

REML 방법을 사용하는 경우 "열 저장" 메뉴에 6 개의 추가 옵션이 나타납니다. 이러한 옵션 이름은 조건부로 시작합니다. 이 접두사는 해당 열을 계산할 때 기대값 0 대신 임의 효과와 관련된 항의 예측값을 사용한다는 것을 나타냅니다.

조건부 예측 계산식 예측 계산식을 데이터 테이블의 새 열에 저장합니다.

조건부 예측값 예측값을 데이터 테이블의 새 열에 저장합니다.

조건부 잔차 잔차를 데이터 테이블의 새 열에 저장합니다.

조건부 평균 CI 평균에 대한 신뢰 구간을 저장합니다.

조건부 개별값 CI 개별값에 대한 신뢰 구간을 저장합니다.

JMP PRO 조건부 계산식 게시 조건부 예측 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다. "계산식 저장소" 보고서가 열려 있지 않으면 "계산식 저장소" 보고서가 생성됩니다. 자세한 내용은 Predictive and Specialized Modeling 의에서 확인하십시오.

REML 프로파일러 옵션

REML 방법을 사용하는 경우 요인 프로파일링 > 프로파일러를 선택하면 "예측 프로파일러" 열의 빨간색 삼각형 메뉴에 "조건부 예측"이라는 새 옵션이 나타납니다. 조건부 값은 기대값 0 대신 임의 효과에 대한 예측값을 사용합니다.

참고: 프로파일러에는 모든 요인 수준 조합에 대한 조건부 예측값과 조건부 평균 신뢰 구간이 표시됩니다. 일부 조합은 내포로 인해 의미가 없을 수도 있습니다.

EMS(전통적) 모형 적합 보고서

주의 : EMS 는 사용하지 않는 것이 좋습니다 . REML 이 권장되는 방법입니다 .

"EMS"를 적합 방법으로 선택하면 4 개의 새 보고서가 표시됩니다. 고정 효과와 임의 효과 둘 다에 대한 검정은 "임의 효과에 대해 검정" 보고서에서 수행되므로 "효과 검정" 보고서는 표시되지 않습니다.

기대 평균 제곱

모형 효과에 대한 기대 평균 제곱은 잔차 오차 분산을 포함하여 분산 성분과 고정 효과 값의 선형 결합입니다. 이 테이블에는 각 모형 효과의 기대 평균 제곱을 정의하는 계수가 제공됩니다. 행렬의 행은 왼쪽에 나열된 효과에 해당합니다. 열은 상단에 표시된 분산 성분에 해당합니다. 각 기대 평균 제곱에는 계수가 1인 잔차 분산이 포함됩니다. 이 정보는 테이블 아래에 제공됩니다.

그림 3.63에서는 Investment Castings.jmp 샘플 데이터 테이블에 대한 "기대 평균 제곱" 보고서를 보여 줍니다. "Model: EMS" 스크립트를 실행한 후 모형을 실행합니다.

그림 3.63 기대 평균 제곱 보고서

기대 평균 제곱			
행 평균 제곱 대 열 분산 성분			
EMS			
	절편	Temperature	Casting[Temperature]&랜덤
절편	0	0	0
Temperature	0	16	4
Casting[Temperature]&랜덤	0	0	4
+ 1.0 x 잔차 분산			

테이블에서 알 수 있듯이 Casting[Temperature]에 대한 기대 평균 제곱은 다음과 같습니다.

$$4\sigma_{Casting[Temperature]}^2 + \sigma_{Error}^2$$

분산 성분 추정값

분산 성분의 추정값은 기대 평균 제곱을 해당하는 관측 평균 제곱에 일치시킨 후 계산하여 구합니다. "분산 성분 추정값" 보고서에는 추정된 분산 성분이 제공됩니다.

성분 임의 효과를 나열합니다.

분산 성분 추정값 분산 성분의 추정값을 제공합니다.

총계 백분율 분산 성분 합계에 대한 분산 성분의 비율을 제공합니다.

CV 분산 성분에 대한 변동 계수를 제공합니다. 이 값은 분산 성분의 제곱근에 100 을 곱하여 평균 반응으로 나눈 값입니다.

참고 : 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > CV 를 선택한 경우에만 나타납니다.

분모 합성 검정

검정할 각 효과에 대해 F 통계량이 생성됩니다. 이 통계량의 분모는 귀무가설 하에서 분자 평균 제곱의 기대값을 갖는 평균 제곱입니다. 이 분모는 고정 효과와 관련된 값 및 분산 성분에서 생성되거나 합성됩니다.

소스 검정할 효과를 표시합니다.

MS Den F -검정의 분모에 대한 추정 평균 제곱을 제공합니다.

DF Den 합성 분모의 자유도를 제공합니다. 이 값은 Satterthwaite 방법 (Satterthwaite 1946) 을 사용하여 생성됩니다.

MS 합성 (분모) 분모 합성에 사용된 분산 성분을 제공합니다. 잔차 오차 분산은 항상 이 합성의 일부입니다.

임의 효과에 대해 검정

이 보고서에는 고정 효과와 임의 효과에 대한 검정이 제공됩니다.

소스 검정할 효과를 나열합니다. 여기에는 고정 효과와 임의 효과가 포함됩니다.

SS 효과에 대한 제곱합을 제공합니다.

MS Num 분자 평균 제곱을 제공합니다.

DF Num 분자 자유도를 제공합니다.

F 비 검정에 대한 F 비를 제공합니다. 이 값은 분모 평균 제곱에 대한 분자 평균 제곱의 비율입니다. 분모 평균 제곱은 "분모 합성 검정" 보고서에서 얻을 수 있습니다.

Prob > F 효과 검정을 위한 p 값을 제공합니다.

주의: 최소 제곱 평균에 대한 표준 오차와 대비 F -검정에 대한 분모는 합성 분모를 사용합니다. 공통 수준에서 비교되는 교차 효과가 포함된 검정과 같은 특정 상황에서는 이러한 검정이 적절하지 않을 수 있습니다. 사용자 검정은 잔차 오류를 사용하여 수행되고 레버리지 그림은 잔차 오류를 사용하여 생성되므로 역시 적절하지 않을 수 있습니다.

EMS 프로파일러

EMS 방법을 사용하는 경우 요인 프로파일링 > 프로파일러를 선택하면 프로파일러에서 고정 효과 모형을 기반으로 예측 및 조건부 평균 신뢰 구간을 제공합니다. 이러한 값은 임의 효과에 대한 예측값을 기반으로 하지 않습니다.

모형 항 사이에 선형 종속성이 있는 모형

예측 변수 행렬의 열 사이에 선형 종속성이 있으면 몇 가지 표준 최소 제곱 보고서에 영향을 줍니다.

- "특이성 상세 정보"

- "모수 추정값 보고서"
- "효과 검정 보고서"
- "예"

특이성 상세 정보

선형 회귀 모형은 $Y = X\beta + \varepsilon$ 로 표현됩니다. 여기서 X 는 첫 번째 열이 1로 구성되고 나머지 열이 모형의 비절편 항에 대한 값으로 구성된 행렬입니다. 모형이 절편을 포함하여 p 개 항으로 구성된 경우 X 는 $n \times p$ 행렬이며, 여기서 n 은 관측값 수입니다. 벡터 b 로 나타내는 모수 추정값은 일반적으로 다음 계산식에 의해 제공됩니다.

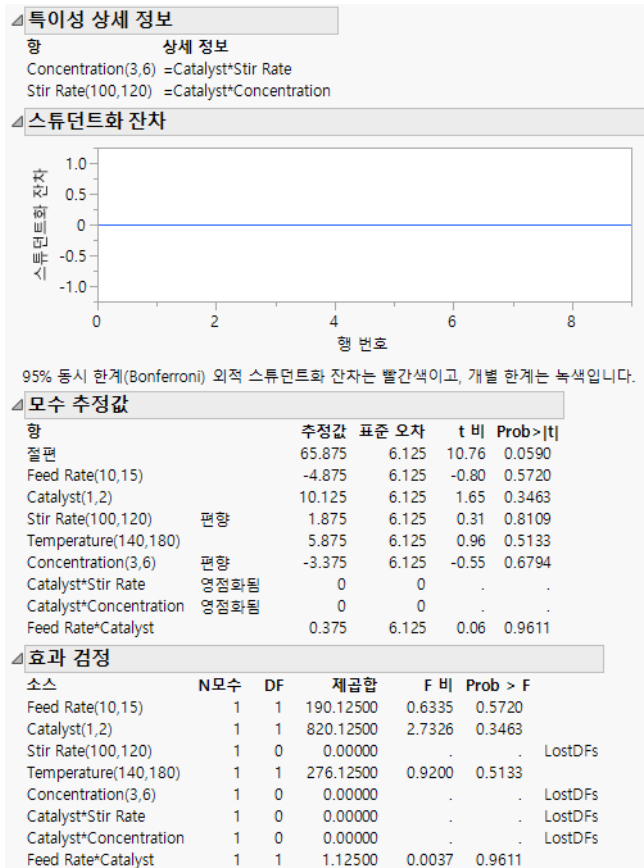
$$b = (X'X)^{-1}X'Y$$

그러나 이 계산식은 $X'X^{-1}$ 이 있다고 가정합니다. 즉, $p \times p$ 행렬 $X'X$ 를 반전할 수 있거나 완전 계수라고 가정합니다. X 열 사이에 선형 종속성이 있으므로 $X'X$ 를 반전할 수 없는 상황이 종종 발생합니다.

이 경우 $X'X$ 행렬은 특이 행렬이고 "최소 제곱 적합" 보고서에 "특이성 상세 정보" 보고서가 포함됩니다. 이 보고서에는 선형 종속성을 설명하는 표현식 테이블이 포함되어 있습니다. 이러한 선형 종속성과 관련된 항은 별칭이 지정됩니다(교락).

그림 3.64에서는 Reactor 8 Runs.jmp 샘플 데이터 테이블에 대한 보고서를 보여 줍니다. 이러한 보고서를 생성하려면 Percent Reacted 를 Y 로 지정하여 모형을 적합시킵니다. Feed Rate, Catalyst, Stir Rate, Temperature, Concentration, Catalyst*Stir Rate, Catalyst*Concentration 및 Feed Rate*Catalyst 를 모형 효과로 입력합니다.

그림 3.64 선형 종속성이 있는 모형에 대한 특이성 및 모수 추정값 보고서



모수 추정값 보고서

$X'X$ 가 특이 행렬인 경우 일반화 역행렬을 사용하여 추정값을 구합니다. 이 방법을 사용하면 선형 종속성과 관련된 모수의 일부(전부는 아님)를 추정할 수 있습니다. 모수는 관련된 항이 모형에 입력되는 순서에 따라 추정되므로 마지막으로 입력한 항은 모수가 추정되지 않는 항입니다. "모수 추정값" 보고서에 추정값이 제공되고 추정할 수 없는 모수의 추정값은 0으로 표시됩니다.

그러나 선형 종속성과 관련된 항의 모수 추정값은 고유하지 않습니다. 관련된 항은 별칭 관계이므로 최소 제곱 기준을 충족하는 추정값 벡터가 무한히 많습니다. 이러한 경우 "모수 추정값" 보고서에서 해당 추정값의 왼쪽에 "편향"이 표시됩니다. 모수를 추정할 수 없는 선형 종속성 관련 항의 경우 "모수 추정값" 보고서에서 추정값 0의 왼쪽에 "영점화됨"이 표시됩니다. 예는 그림 3.64에서 확인하십시오.

오차 추정값에 사용할 수 있는 자유도가 있는 경우 편향 추정값을 사용하여 추정된 모수에 대한 t -검정이 수행됩니다. 그러나 이러한 검정은 추정값이 고유하지 않다는 점을 고려하여 신중하게 해석해야 합니다.

효과 검정 보고서

표준 최소 제곱 적합에서는 존재하는 모형 자유도와 같은 수의 모수만 추정할 수 있습니다. "효과 검정" 보고서의 검정을 수행할 때 각 효과는 모형에 입력된 마지막 효과로 간주됩니다.

- 모든 모형 자유도가 다른 효과에 사용되는 경우 효과의 DF가 0으로 표시됩니다. DF가 0이면 제곱합을 계산할 수 없습니다. 따라서 해당 효과를 검정할 수 없습니다.
- 일부 모형 자유도가 다른 효과에 사용되는 경우 나머지 효과의 DF는 0이 아닙니다. 그러나 DF가 모수의 수 ("모수 수") 보다 작을 수 있으며, 이는 관련 모수 중 일부만 검정할 수 있다는 의미입니다.

오차에 대한 자유도가 있다고 가정하고 효과에 대한 자유도가 0이 아니면 F-검정이 수행됩니다. DF가 모수 수보다 작으면 효과에 해당하는 행의 오른쪽 끝에 LostDFs라는 설명이 표시됩니다(그림 3.64). 이러한 효과는 손실된 자유도를 흡수한 별칭 관계의 효과에 기인하지 않은 모형 제곱합만 설명할 수 있습니다. 따라서 "효과 검정" 보고서에 제공된 제곱합은 효과와 관련된 "실제" 제곱합보다 적게 표시될 가능성이 높습니다. 검정이 유의하면 유의성이 의미가 있습니다. 그러나 유의성 부족은 신중하게 해석해야 합니다.

통계 상세 정보는 SAS Institute Inc. (2020c).

예

Singularity.jmp 샘플 데이터 테이블을 엽니다. 반응 Y, 4개의 예측 변수 X1, X2, X3, A 및 5개의 관측값이 있습니다. 예측 변수는 4개 수준의 명목형 변수 A를 제외하고 모두 연속형입니다. 또한 연속형 효과 사이에 선형 종속성이 있습니다. 즉, $X_3 = X_1 + X_2$ 입니다.

고유하지 않은 추정값

선형 종속성이 있을 때 추정값이 고유하지 않은지 확인하려면 다음을 수행하십시오.

- 도움말 > 샘플 데이터 라이브러리**를 선택하고 Singularity.jmp를 엽니다.
- Model 1** 스크립트를 실행합니다. 그러면 X1, X2, X3 순서로 효과가 입력된 "모형 적합" 시작창이 열립니다.
- 실행**을 클릭하고 보고서 창은 열린 상태로 둡니다.
- Model 2** 스크립트를 실행합니다. 그러면 X1, X3, X2 순서로 효과가 입력된 "모형 적합" 시작창이 열립니다.
- 실행**을 클릭하고 보고서 창은 열린 상태로 둡니다.

두 보고서(그림 3.65)를 비교합니다. 두 보고서의 맨 위에 있는 "특이성 상세 정보" 보고서에는 선형 종속성이 표시됩니다($X_1 = X_3 - X_2$).

이제 두 모형의 "모수 추정값" 보고서를 비교합니다. 예를 들어 X1에 대한 추정값이 Model 1은 -1.25이고 Model 2는 2.75입니다. 두 모형 모두 모형 자유도가 2이므로 효과와 관련된 항 중 두 개만 추정됩니다. "분산 분석" 보고서를 참조하십시오. 추정된 두 항의 추정값은 "편향"이라는 라벨이 지정되며, 나머지 추정값은 0으로 설정되고 "영점화됨"이라는 라벨이 지정됩니다.

"효과 검정" 보고서에서는 검정이 수행되지 않았음을 보여 줍니다. 각 행에 LostDFs 라벨이 지정되어 있습니다. 이런 결과가 발생하는 이유는 이러한 효과 중 하나에 대한 효과 검정을 수행하려면 해당 효과를 모형에 마지막으로 입력해야 하기 때문입니다. 그러나 다른 두 효과가 모형 자유도 2와 관련된 모형 제곱합을 완전히 설명합니다. 따라서 관심 효과에 대한 자유도 또는 관련 제곱합이 남아 있지 않습니다.

그림 3.65 Model 1(왼쪽) 및 Model 2(오른쪽) 에 대한 최소 제곱 적합 보고서

반응 Y						반응 Y					
특이성 상세 정보						특이성 상세 정보					
항 상세 정보						항 상세 정보					
X1 = - X2 + X3						X1 = X3 - X2					
적합 요약						적합 요약					
분산 분석						분산 분석					
소스	DF	제곱합	평균 제곱	F 비		소스	DF	제곱합	평균 제곱	F 비	
모형	2	97.800000	48.9000	97.8000		모형	2	97.800000	48.9000	97.8000	
오차	2	1.000000	0.5000	Prob > F		오차	2	1.000000	0.5000	Prob > F	
수정 합계	4	98.800000		0.0101*		수정 합계	4	98.800000		0.0101*	
모수 추정값						모수 추정값					
항	추정값	표준 오차	t 비	Prob> t		항	추정값	표준 오차	t 비	Prob> t	
절편	11.75	0.433013	27.14	0.0014*		절편	11.75	0.433013	27.14	0.0014*	
X1 편향	-1.25	0.433013	-2.89	0.1020		X1 편향	2.75	0.661438	4.16	0.0533	
X2 편향	-4	0.353553	-11.31	0.0077*		X3 편향	-4	0.353553	-11.31	0.0077*	
X3 영점화됨	0	0	.	.		X2 영점화됨	0	0	.	.	
효과 검정						효과 검정					
효과 상세 정보						효과 상세 정보					

LostDFs

LostDFs 에 대해 더 자세히 알아보려면 아래 단계를 수행하거나 Fit Model Report 데이터 테이블 스크립트를 실행하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Singularity.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 클릭합니다 .
3. Y 를 선택하고 **Y** 를 클릭합니다 .
4. X1 과 A 를 선택하고 **추가**를 클릭합니다 .
5. **강조**를 **최소 보고서**로 설정합니다 .
6. **실행**을 클릭합니다 .

그림 3.66 에 보고서의 일부가 표시되어 있습니다. "특이성 상세 정보" 보고서에서는 X1 및 A 효과와 연결된 3개 항과 관련하여 선형 종속성이 있음을 보여 줍니다. 명목형 효과를 코드화하는 방법에 대한 자세한 내용은 "[사용자 검정 예에 대한 상세 정보](#)"(188페이지)에서 확인하십시오. "분산 분석" 보고서에서는 모형 자유도가 3이라는 것을 보여 줍니다. "모수 추정값" 보고서에서는 3개의 항 X1, A[a], A[b]에 대한 "편향" 추정값과 4번째 항 A[c]에 대한 "영점화됨" 추정값을 보여 줍니다.

"효과 검정" 보고서에서는 X_1 을 검정할 수 없음을 보여 줍니다. A 를 먼저 입력하고 A 가 세 개의 모형 자유도를 설명해야 하기 때문입니다. 그러나 A 는 검정할 수 있지만 자유도가 두 개뿐입니다. X_1 을 먼저 입력하고 모형 자유도 중 하나를 설명해야 합니다. A 에 대한 검정은 부분적이므로 신중하게 해석해야 합니다.

그림 3.66 X_1 및 A 를 사용한 모형의 최소 제곱 적합 보고서

반응 Y

특이성 상세 정보

항 상세 정보

모델 $= 2 * X_1 - A[a] - A[b] + 3 * A[c]$

회귀 그림

적합 요약

분산 분석

소스	DF	제곱합	평균 제곱	F 비
모형	3	98.300000	32.7667	65.5333
오차	1	0.500000	0.5000	Prob > F
수정 합계	4	98.800000		0.0905

모수 추정값

항	추정값	표준 오차	t 비	Prob> t
절편	14.416667	0.399653	36.07	0.0176*
X_1	-2.583333	0.399653	-6.46	0.0977
A[a]	-5.333333	0.471405	-11.31	0.0561
A[b]	3.166667	0.552771	5.73	0.1100
A[c]	영점화됨	0	0	.

효과 검정

소스	N모수	DF	제곱합	F 비	Prob > F
X_1	1	0	0.000000	.	LostDFs
A	3	2	64.500000	64.5000	0.0877

효과 상세 정보

표준 최소 제곱 분석법에 대한 통계 상세 정보

- "강조 규칙"
- "사용자 검정 예에 대한 상세 정보"
- "추정값 상관계수"
- "레버리지 그림 상세 정보"
- "Kackar-Harville 수정"
- "검정력 분석"

강조 규칙

"모형 적합" 시작 창의 기본 강조는 행 수 (n), "모형 효과 생성" 목록에 입력된 효과 수 (k) 및 효과에 적용된 속성을 기반으로 합니다.

- $n > 1000$ 이면 "강조"가 "최소 보고서"로 설정됩니다.
- $n \leq 1000, k \leq 4$ 이면 "강조"가 "효과 레버리지"로 설정됩니다.
- $n \leq 1000, k \geq 10$ 이면 "강조"가 "효과 선별"로 설정됩니다.

- $n \leq 1000$, $4 < k < 10$ 및 $n - k > 20$ 이면 "강조"가 "효과 레버리지"로 설정됩니다.
- 효과에 "임의 효과" 속성이 있으면 "강조"가 "최소 보고서"로 설정됩니다.
- 이러한 조건 중 어느 것도 해당되지 않으면 "강조"가 "효과 선별"로 설정됩니다.

사용자 검정 예에 대한 상세 정보

"사용자 검정의 예"(116 페이지)에서는 Cholesterol.jmp 샘플 데이터 테이블을 사용하여 세 가지 대비를 검정합니다. 구체적으로 다음 내용을 비교하려고 합니다.

- treatment A의 평균 반응 대 treatment B의 평균 반응
- 결합된 treatment A/B의 평균 반응 대 대조군의 평균 반응
- 결합된 treatment A/B의 평균 반응 대 결합된 대조군 / 플라시보 그룹의 평균 반응

"사용자 검정" 열에 입력하는 대비 계수를 유도하기 위해 다음을 수행하십시오. 4개의 treatment 그룹에 대한 이론적 효과를 α_A , α_B , $\alpha_{Control}$ 및 $\alpha_{Placebo}$ 로 나타냅니다. 이것은 treatment 효과가므로 합이 0으로 제한됩니다. 표시 변수와 관련된 모수는 처음 세 효과만 나타내므로 처음 세 효과를 기준으로 대비를 공식화해야 합니다. 자세한 내용은 "사용자 검정 예에 대한 상세 정보"(188 페이지) 및 "통계 상세 정보" 부록의 "모수 해석"(493 페이지)에서 확인하십시오.

검정할 가설은 다음과 같이 모형 효과를 기준으로 작성할 수 있습니다.

- treatment A와 treatment B 비교 :

$$\alpha_A - \alpha_B = 0$$

- treatment A/B와 대조군 비교 :

$$0.5(\alpha_A + \alpha_B) - \alpha_{Control} = 0$$

- treatment A/B와 대조군 / 플라시보 그룹 비교 :

$$0.5(\alpha_A + \alpha_B) - 0.5(\alpha_{Control} + \alpha_{Placebo}) = \alpha_A + \alpha_B = 0$$

이 대비를 위한 대비 계수를 얻으려면 모형 효과를 기준으로 placebo 효과를 작성해야 합니다. 구체적으로 $\alpha_A + \alpha_B + \alpha_{Control} + \alpha_{Placebo} = 0$ 이라는 사실을 사용합니다. 그러면 $\alpha_{Placebo} = -\alpha_A - \alpha_B - \alpha_{Control}$ 입니다. 결론은 다음과 같습니다.

$$\begin{aligned} 0.5(\alpha_A + \alpha_B) - 0.5(\alpha_{Control} + \alpha_{Placebo}) &= 0.5(\alpha_A + \alpha_B) - 0.5(\alpha_{Control} - \alpha_A - \alpha_B - \alpha_{Control}) \\ &= 0.5(\alpha_A + \alpha_B) - 0.5\alpha_{Control} + 0.5(\alpha_A + \alpha_B + \alpha_{Control}) \\ &= \alpha_A + \alpha_B \\ &= 0 \end{aligned}$$

추정값 상관계수

관측값 n 개와 예측 변수 $p - 1$ 개가 있는 데이터 집합을 예로 들겠습니다. 행렬 $\mathbf{X}$ 를 설계 행렬로 정의합니다. 즉, $\mathbf{X}$ 는 첫 번째 열이 1로 구성되고, 나머지 $p - 1$ 개 열이 $p - 1$ 개의 예측 변수 값으로 구성된 $n \times p$ 행렬입니다. 명목형 열은 표시자 예측 변수를 기준으로 코드화됩니다. 각 열은 행렬 $\mathbf{X}$ 의 열입니다.

회귀 계수 벡터의 추정값은 다음과 같습니다.

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

여기서 $\mathbf{Y}$ 는 반응 값의 벡터를 나타냅니다.

일반적인 회귀 가정에서 $\hat{\beta}$ 의 공분산 행렬은 다음과 같습니다.

$$\text{Cov}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

여기서 σ^2 은 반응의 분산을 나타냅니다.

추정값에 대한 상관 행렬은 공분산 행렬의 각 항목을 대각 항목의 제곱근의 곱으로 나누어 얻습니다. 다음과 같이 $\mathbf{V}$ 를 대각 행렬이 되도록 정의합니다. 이 행렬의 각 항목은 공분산 행렬의 대각 항목에 대한 제곱근입니다.

$$\mathbf{V} = \text{Sqrt}(\text{Diag}(\text{Cov}(\hat{\beta})))$$

그러면 모수 추정값에 대한 상관 행렬이 다음과 같이 지정됩니다.

$$\text{Corr}(\hat{\beta}) = \sigma^2\mathbf{V}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{V}^{-1}$$

레버리지 그림 상세 정보

효과 레버리지 그림은 편회귀 잔차 레버리지 그림(Belsley et al. 1980) 또는 추가 변수 그림(Cook and Weisberg 1982)이라고도 합니다. Sall(1990)에서는 이러한 그림을 선형 가설에 적용하도록 일반화했습니다.

JMP에서는 다음과 같은 두 가지 유형의 레버리지 그림을 제공합니다.

- 효과 레버리지 그림에는 다른 모든 효과가 모형에 있을 때 해당 효과가 모형에 없다는 가설을 기준으로 관측값이 표시됩니다.
- "실제값 대 예측값 그림" 보고서에 나오는 전체 모형 레버리지 그림에는 요인 효과가 없다는 가설을 기준으로 관측값이 표시됩니다.

효과 레버리지 그림에서는 하나의 효과만 0으로 가정됩니다. 그러나 전체 모형에 대한 실제값 대 예측값 그림에서는 모든 효과가 0으로 가정됩니다. Sall(1990)에서는 레버리지 그림의 개념을 임의의 선형 가설로 일반화하며, 전체 모형 레버리지 그림이 그 예입니다. 이 섹션에 요약된 해당 연구 자료의 상세 정보는 JMP에 있는 두 가지 유형의 그림에 특화되어 있습니다.

생성

추정 가능한 관심 가설이 다음과 같다고 가정해 보겠습니다.

$$L\beta = 0$$

레버리지 그림은 각 점에서 경사진 회귀선까지의 거리가 무제약 잔차를 표시하도록 점을 그려 이 검정의 특징을 나타냅니다. 0에서 수평선까지의 거리는 적합이 가설의 제약을 받을 때 잔차를 표시합니다. 이 두 잔차 집합의 제곱합 간의 차이는 가설로 인한 제곱합입니다. 이 값은 F -검정의 주성분이 됩니다.

가설의 제약을 받는 모수 추정값을 다음과 같이 작성할 수 있습니다.

$$b_0 = b - (X'X)^{-1}L'\lambda$$

여기서 b 는 최소 제곱 추정값입니다.

$$b = (X'X)^{-1}X'y$$

람다는 가설 제약 조건에 대한 라그랑주 승수이며 다음과 같이 계산됩니다.

$$\lambda = (L(X'X)^{-1}L')^{-1}Lb$$

무제약 잔차와 가설의 제약을 받는 잔차는 각각 다음과 같습니다.

$$r = y - Xb$$

$$r_0 = r + X(X'X)^{-1}L'\lambda$$

각 관측값에 대해 가로 축 값이 v_x 이고 세로 축 값이 v_y 인 점을 가정해 보겠습니다. 다음은 각 요소에 대한 설명입니다.

- v_x 는 제약된 잔차에서 무제약 잔차를 뺀 $r_0 - r$ 이며, 제약 조건이 적용된 후 남은 정보를 반영합니다.
- v_y 는 가로 축 값에 무제약 잔차를 더한 값입니다.

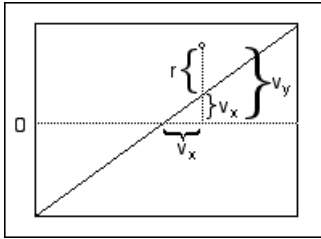
따라서 이 점의 x 및 y 좌표는 다음과 같습니다.

$$v_x = X(X'X)^{-1}L'\lambda \quad \text{및} \quad v_y = r + v_x$$

이러한 점을 기반으로 레버리지 그림이 생성됩니다. 그림 3.67에서는 평균 반응이 0이고 실선 기울기가 1인 레버리지 그림 생성 구조를 보여 줍니다.

JMP의 레버리지 그림에는 반응 평균 $\bar{y}$ 에 수평 점선이 있습니다. 그림에 표시된 점은 $(v_x + \bar{y}, v_y)$ 로 지정됩니다.

그림 3.67 레버리지 그림 생성



레버리지 그림에 검정 중첩

단순 선형 회귀에서 반응 기대값에 대한 신뢰 한계를 예측 변수 $\mathbf{x}$ 의 평활 함수로 그림에 표시할 수 있습니다.

$$\text{Upper}(\mathbf{x}) = \mathbf{x}\mathbf{b} + t_{\alpha/2} s \sqrt{\mathbf{x}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}'}$$

$$\text{Lower}(\mathbf{x}) = \mathbf{x}\mathbf{b} - t_{\alpha/2} s \sqrt{\mathbf{x}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}'}$$

여기서 $\mathbf{x} = [1 \ x]$ 는 예측 변수의 2-벡터입니다.

그림 3.56에 표시된 것처럼 이러한 신뢰 곡선을 통해 해당 가설 검정의 유의성을 시각적으로 평가할 수 있습니다.

- 유의함: 기울기 모수가 0과 유의한 차이가 있으면 신뢰 곡선이 반응 평균에서 수평선과 교차합니다.
- 경계선: 기울기 모수에 대한 t -검정이 유의성 한계에 있으면 신뢰 곡선이 반응 평균에서 수평선에 점근적입니다.
- 유의하지 않음: 기울기 모수가 0과 유의한 차이가 없으면 신뢰 곡선이 반응 평균에서 수평선과 교차하지 않습니다.

레버리지 그림은 신뢰 곡선을 표시하여 이 개념을 반영합니다. 그림이 적절하게 중앙에 오도록 조정됩니다. 가로 축의 점을 z 로 나타냅니다. 함수를 다음과 같이 정의합니다.

$$\text{Upper}(z) = z + \sqrt{s^2 t_{\alpha/2}^2 \bar{h} + (F_{\alpha}/F) z^2}$$

및

$$\text{Lower}(z) = z - \sqrt{s^2 t_{\alpha/2}^2 \bar{h} + (F_{\alpha}/F) z^2}$$

여기서 F 는 가설에 대한 F 통계량이고 F_{α} 는 유의 수준 α 에 대한 참조 값입니다.

또한 $\bar{h} = \bar{\mathbf{x}}(\mathbf{X}'\mathbf{X})^{-1}\bar{\mathbf{x}}'$ 입니다. 여기서 $\bar{\mathbf{x}}$ 는 예측 변수에 대한 적절한 중간 값(예: 평균)으로 구성된 행 벡터입니다.

이러한 함수는 다음과 같이 단순 선형 회귀에 대한 신뢰 곡선과 동일한 방식으로 작동합니다.

- F 통계량이 참조 값보다 크면 신뢰도 함수가 가로 축과 교차합니다.
- F 통계량이 참조 값과 같으면 신뢰도 함수의 가로 축이 점근선으로 표시됩니다.
- F 통계량이 참조 값보다 작으면 신뢰도 함수가 교차하지 않습니다.

또한 $\text{Upper}(z) - \text{Lower}(z)$ 가 z 에서의 예측값에 대한 유효한 신뢰 구간이어야 합니다.

Kackar-Harville 수정

고정 효과의 분산 행렬은 항상 Kackar-Harville 수정을 포함하도록 변경됩니다. BLUP의 분산 행렬 및 BLUP과 고정 효과 간의 공분산은 Kackar-Harville 수정 대상이 아닙니다. 이 방법의 근거는 임의 효과에 수준이 많이 있을 때 BLUP에 대한 수정은 계산 및 메모리 사용량이 많을 수 있다는 것입니다.

SAS에서는 $\text{DDFM}=\text{KENWARDROGER}$ 가 설정된 경우에만 Kackar-Harville 수정이 고정 효과와 BLUP에 대해 모두 수행됩니다.

- 고정 효과 모수만 포함하는 선형 결합의 표준 오차는 $\text{PROC MIXED DDFM}=\text{KENWARDROGER}$ 와 일치합니다. 이 경우 PROC MIXED 와 JMP 에서 사용하는 서로 다른 파라미터화 간의 변환에 주의할 기울여야 합니다.
- BLUP 모수만 포함하는 선형 결합의 표준 오차는 $\text{PROC MIXED DDFM}=\text{SATTERTHWAITE}$ 와 일치합니다.
- 불균형 데이터의 경우 고정 효과와 BLUP를 모두 포함하는 선형 결합의 표준 오차는 DDFM 옵션에 대한 PROC MIXED 와 일치하지 않습니다. 그러나 이러한 표준 오차는 $\text{DDFM}=\text{SATTERTHWAITE}$ 및 $\text{DDFM}=\text{KENWARDROGER}$ 옵션을 사용한 결과 사이에 있습니다. 균형 데이터의 경우 Kackar-Harville 수정이 null 이므로 JMP 는 DDFM 옵션에 관계없이 SAS와 일치합니다.

자유도

고정 효과 모수의 선형 결합만 포함하는 검정의 자유도는 Kenward-Roger 수정을 사용하여 계산됩니다. 따라서 이러한 검정의 JMP 결과는 $\text{DDFM}=\text{KENWARDROGER}$ 옵션을 사용하는 PROC MIXED 와 일치합니다. 선형 결합에 BLUP가 있는 경우 JMP 는 Satterthwaite 근사를 사용하여 자유도를 구합니다. 결과는 이전 단락에서 표준 오차에 대해 설명한 것과 유사한 패턴을 따릅니다.

Kackar-Harville 수정 및 Kenward-Roger DF 방법에 대한 자세한 내용은 Kenward and Roger 연구 자료(1997)에서 확인하십시오. Satterthwaite 방법은 SAS Institute Inc. (2020d, ch. 83).

검정력 분석

검정력 계산과 관련된 옵션은 연속형 반응 모형에만 사용할 수 있습니다. 다음과 같은 컨텍스트에서 검정력 및 관련된 검정 상세 정보를 사용할 수 있습니다.

모수 추정값

각 모수 추정값에 대한 후향적 검정 상세 정보를 얻으려면 보고서의 빨간색 삼각형 메뉴에서 **추정값 > 모수 검정력**을 선택합니다. 이 옵션은 현재 연구 데이터를 기반으로 각 모수에 대한 0.05 유의 수준 검정의 최소 유의값, 최소 유의 개수 및 조정 검정력을 표시합니다.

효과 또는 효과 상세 정보

특정 효과의 F -검정에 대한 후향적 또는 전향적 상세 정보를 얻으려면 효과의 빨간색 삼각형 메뉴에서 **검정력 분석**을 선택합니다. 효과 선별 및 최소 보고서 분석법의 경우 각 효과에 대한 보고서는 "효과 상세 정보" 아래에 있습니다. 효과 레버리지 분석법의 경우 효과에 대한 보고서는 보고서의 첫 번째(전체 모형) 열 오른쪽에 있습니다.

최소 제곱 평균 대비

하나 이상의 대비 검정에 대한 후향적 또는 전향적 상세 정보를 얻으려면 효과의 빨간색 삼각형 메뉴에서 **최소 제곱 평균 대비**를 선택합니다. 원하는 대비를 정의하고 "완료"를 클릭합니다. "대비" 빨간색 삼각형 메뉴에서 **검정력 분석**을 선택합니다.

사용자 검정

사용자 검정에 대한 후향적 또는 전향적 상세 정보를 얻으려면 반응의 빨간색 삼각형 메뉴에서 **추정값 > 사용자 검정**을 선택합니다. 원하는 대비를 정의하고 **완료**를 클릭합니다. "사용자 검정" 빨간색 삼각형 메뉴에서 **검정력 분석**을 선택합니다.

첫 번째를 제외한 모든 경우에 "검정력 분석"을 선택하면 "검정력 상세 정보 창"이 열립니다. 그러면 "검정력 상세 정보 창"에 정보를 입력하여 필요에 따라 계산을 수정할 수 있습니다.

효과 크기

효과 크기(δ)는 귀무가설과 관련 모수의 실제 값 간의 차이를 측정하는 것입니다. 귀무가설은 0으로 설정된 단일 선형 대비 또는 이러한 대비 여러 개로 공식화할 수 있습니다. δ 의 값은 대비의 실제 값과 가설 값 0 간의 차이를 반영합니다.

일반적으로 효과 크기는 다음과 같이 지정됩니다.

$$\delta = \sqrt{SS_{Hyp(Pop)}/n}$$

여기서 $SS_{Hyp(Pop)}$ 는 모집단 모수로 주어진 검정 대상 가설의 제곱합이고 n 은 총 관측값 수입니다. 가능한 관측값이 있는 경우 추정된 효과 크기는 가설에 대해 계산된 제곱합을 δ 의 계산식에 대입하여 계산됩니다.

균형 일원 배치

i 번째 그룹의 평균 반응이 α_i 인 k 개 수준의 균형 일원 배치에 대한 특수한 경우를 예로 들겠습니다.

$$\delta^2 = \frac{\sum (\alpha_i - \bar{\alpha})^2}{k}$$

JMP 모수를 코드화하여 $i=1, 2, \dots, k-1$ 에 대해 다음과 같이 됩니다.

$$\beta_i = (\alpha_i - \bar{\alpha})$$

및

$$\beta_k = - \sum_{m=1}^{k-1} \alpha_m$$

따라서 이러한 모수에 관해 2 수준 균형 배치에 대한 δ 는 다음과 같이 제공됩니다.

$$\delta^2 = \frac{\beta_1^2 + (-\beta_1)^2}{2} = \beta_1^2$$

$$\text{또는 } \delta = |\beta_1|$$

불균형 일원 배치

k 개 수준이 있는 불균형 일원 배치에서 i 번째 그룹의 평균 반응이 α_i , n_i 는 관측값, $n = \sum n_i$ 인 경우:

$$\delta^2 = \sum \frac{n_i}{n} (\alpha_i - \bar{\alpha})^2$$

효과 크기 및 검정력

검정력은 실제 효과 크기가 지정된 값일 때 α 유의 수준에서 가설에 대한 F -검정이 유의할 확률입니다. 실제 효과 크기가 δ 인 경우 검정 통계량은 다음과 같이 계산된 비중심성 모수를 갖는 비중심 F 분포를 따릅니다.

$$\lambda = (n\delta^2)/\sigma^2$$

귀무가설이 참인 경우, 즉 효과 크기가 0이면 비중심성 모수는 0이고 검정 통계량은 중심 F 분포를 따릅니다.

검정력은 λ 에 따라 증가합니다. 특히 검정력은 표본 크기 n 과 효과 크기 δ ,에 따라 증가하고 오차 분산 σ^2 에 따라 감소합니다.

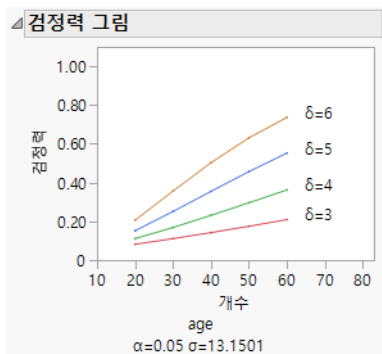
Cohen(1977)과 같은 일부 참고 도서에는 JMP에서 사용하는 원시 효과 크기 대신 표준화된 효과 크기 $\Delta = \delta/\sigma$ 가 사용됩니다. 표준화된 효과 크기의 경우 비중심성 모수는 $\lambda = n\Delta^2$ 입니다.

"검정력 상세 정보 창"에서 δ 는 처음에 $\sqrt{SS_{Hyp}/n}$ 로 설정되어 있습니다. SS_{Hyp} 는 가설에 대한 제곱합이고 n 은 현재 연구의 관측값 수입니다. SS_{Hyp} 는 데이터에서 계산된 δ 의 추정값이지만 이러한 추정값은 편향되어 있습니다(Wright and O'Brien 1988). δ 에 대한 표본 추정값을 사용하여 검정력을 계산하려면 "검정력 계산" 대신 "조정 검정력 및 신뢰 구간" 계산을 사용할 수 있습니다. 조정 검정력 계산은 편향이 부분적으로 수정된 δ 의 추정값을 사용합니다. 자세한 내용은 "통계 상세 정보" 부록의 "조정 검정력 계산"(518 페이지)에서 확인하십시오.

표본 크기별 검정력 그림

표본 크기별 검정력 그림을 보려면 "검정력" 보고서 맨 아래의 빨간색 삼각형 메뉴에서 "검정력 그림" 옵션을 선택합니다. 그러면 "검정력" 테이블의 "검정력" 및 "개수" 열이 그림으로 표시됩니다. 그림 3.68에 표시된 그림은 "후향적 검정력 분석의 예"(197 페이지)에서 얻은 "검정력" 테이블을 그림으로 나타낸 결과입니다.

그림 3.68 표본 크기별 검정력 그림



LSN(최소 유의 개수)

LSN(최소 유의 개수)은 델타, 시그마 및 알파 값이 주어졌을 때 유의한 검정 결과를 생성하는 최소 관측값 수입니다. 델타, 시그마 및 알파는 각각 효과 크기, 오차 표준편차 및 유의 수준을 나타냅니다.

참고 : LSN은 유의 확률을 고려하지 않으므로 권장되는 표본 크기가 아닙니다. 이 값은 지정된 델타 및 시그마 값을 기반으로 계산됩니다.

LSN에는 다음과 같은 특성이 있습니다.

- LSN이 실제 표본 크기 n 보다 작으면 효과가 유의합니다.
- LSN이 n 보다 크면 효과가 유의하지 않습니다. 추가 데이터가 현재 표본과 동일한 구조 결과를 나타낼 것이라고 생각하는 경우 LSN은 유의성을 달성하는 데 필요한 데이터 양을 제안합니다.
- LSN이 n 과 같으면 p 값이 유의 수준 알파와 같습니다. 이 경우 검정은 유의성의 경계에 있습니다.

- $n = \text{LSN}$ 일 때 계산된 효과 크기 검정의 검정력은 항상 0.5 보다 크거나 같습니다. 그러나 검정력은 0.5 에 가까울 수 있으며 이는 계획 목적상 낮은 것으로 간주됩니다.

LSV(최소 유의값)

LSV(최소 유의값)는 단일 자유도 가설 검정에 대해 계산됩니다. 여기에는 좀 더 일반적인 선형 대비는 물론 개별 모형 모수의 유의성에 대한 검정이 포함됩니다. LSV는 지정된 유의 수준에서 유의한 것으로 간주될 최소 효과 크기(절대값)입니다. LSV는 확률 척도가 아니라 모수 척도에 따라 검정의 민감도를 측정합니다.

LSV에는 다음과 같은 특성이 있습니다.

- 모수 추정값 또는 대비의 절대값이 LSV 보다 크거나 같으면 유의성 검정의 p 값이 알파보다 작거나 같습니다.
- 모수 추정값 또는 대비의 절대값은 유의성 검정의 p 값이 알파와 같을 때만 LSV와 같습니다.
- LSV는 모수 또는 모수의 선형 결합에 대한 $1 - \alpha$ 신뢰 구간의 반지름입니다. $1 - \alpha$ 신뢰 구간은 모수 또는 대비의 추정값에서 중심화됩니다.

검정력

검정력은 해당 검정이 유의한 결과를 제공할 확률입니다. 검정력은 효과 크기 δ , 유의 수준 α , 오차 표준편차 σ 및 표본 크기 n 의 함수입니다. 검정력은 지정된 유의 수준에서 지정된 효과 크기를 감지할 확률입니다. 일반적으로 실제적 또는 과학적 중요도의 차이를 감지할 검정력이 높은 연구를 설계하기를 원합니다.

검정력에는 다음과 같은 특성이 있습니다.

- 모수의 실제 값이 가설 값이면 검정력이 검정의 유의 수준과 같습니다. 일반적으로 유의 수준은 0.05 와 같이 작은 값입니다. 가정된 가설이 참일 때 유의한 결과를 얻을 확률이 낮아야 하므로 작은 값이 적절합니다.
- 일반적으로 모수의 실제 값이 가설 값이 아니면 검정력이 가능한 한 커야 합니다.
- 표본 크기가 증가하고, 오차 분산이 감소하고, 실제 모수 값과 가설 값의 차이가 증가할수록 검정력이 증가합니다.

조정 검정력 및 신뢰 구간

후향적 검정력 분석에서는 일반적으로 표본 추정값이 검정력 계산과 관련된 모집단 모수를 대체합니다. 이로 인해 비중심성 모수 추정값이 양의 편향을 갖게 됩니다(Wright and O'Brien 1988). 조정 검정력 계산은 이 편향이 부분적으로 수정된 비중심성 모수의 추정값을 기반으로 합니다.

조정 검정력에 대한 신뢰 구간을 생성할 수도 있습니다. 이 신뢰 구간은 넓은 경향이 있습니다. 자세한 내용은 Wright and O'Brien 연구 자료(1988)에서 확인하십시오.

조정 검정력 및 신뢰 구간 계산은 데이터에서 추정된 δ 값(기본적으로 제공되는 값)에만 관련됩니다. 다른 델타 값의 경우 조정 검정력 및 신뢰 구간이 제공되지 않습니다.

자세한 내용은 "통계 상세 정보" 부록의 "**조정 검정력 계산**"(518 페이지)에서 확인하십시오.

후향적 검정력 분석의 예

이 예에서는 Big Class.jmp 샘플 데이터 테이블을 사용한 후향적 검정력 분석을 보여 줍니다. "검정력 상세 정보 창" (그림 3.69)에서는 α , σ , δ 및 개수 (연구 크기)에 대한 값 범위에서 다양한 통계량을 탐색할 수 있습니다. "완료"를 클릭하면 창이 계산 결과로 바뀝니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. **weight**를 선택하고 **Y**를 클릭합니다.
4. **age**, **sex** 및 **height**를 효과로 추가합니다.
5. **실행**을 클릭합니다.
6. "age"의 빨간색 삼각형을 클릭하고 **검정력 분석**을 선택합니다.

그림 3.69 age에 대한 검정력 상세 정보 창

검정력 상세 정보 창

age

각 열에 1개 또는 2개의 값이나 값 시퀀스를 입력하십시오.

	α	σ	δ	개수
시작:	0.050	13.15009	3	20
끝:	.	.	6	60
증분:	.	.	1	10

☒ 검정력 계산
☒ 최소 유의 개수 계산
☐ 최소 유의값 계산
☐ 조정 검정력 및 신뢰 구간

완료 취소 도움말

모든 시퀀스 조합에 대해 계산이 수행됩니다.

7. 그림 3.69에 표시된 대로 **시작** 상자의 δ 값을 3으로 바꾸고 **끝** 및 **증분** 상자에 각각 6과 1을 입력합니다.
8. 그림 3.69에 표시된 대로 **시작** 상자의 **개수** 값을 20으로 바꾸고 **끝** 및 **증분** 상자에 각각 60과 10을 입력합니다.
9. **검정력 계산** 및 **최소 유의 개수 계산**을 선택합니다.
10. **완료**를 클릭합니다.
11. "검정력 상세 정보 창"이 "검정력 상세 정보" 보고서로 바뀝니다.

그림 3.70 age에 대한 검정력 상세 정보 보고서

검정력 상세 정보				
검정 age				
검정력				
α	σ	δ	개수	검정력
0.0500	13.15009	3	20	0.0828
0.0500	13.15009	3	30	0.1117
0.0500	13.15009	3	40	0.1426
0.0500	13.15009	3	50	0.1755
0.0500	13.15009	3	60	0.2099
0.0500	13.15009	4	20	0.1117
0.0500	13.15009	4	30	0.1694
0.0500	13.15009	4	40	0.2317
0.0500	13.15009	4	50	0.2969
0.0500	13.15009	4	60	0.3630
0.0500	13.15009	5	20	0.1524
0.0500	13.15009	5	30	0.2515
0.0500	13.15009	5	40	0.3554
0.0500	13.15009	5	50	0.4575
0.0500	13.15009	5	60	0.5529
0.0500	13.15009	6	20	0.2063
0.0500	13.15009	6	30	0.3572
0.0500	13.15009	6	40	0.5035
0.0500	13.15009	6	50	0.6320
0.0500	13.15009	6	60	0.7368
최소 유의 개수				
α	σ	δ	개수(LSN)	
0.0500	13.15009	3	216.8578	
0.0500	13.15009	4	123.8892	
0.0500	13.15009	5	80.93391	
0.0500	13.15009	6	57.6814	

이 분석은 계산 시 Big Class.jmp 샘플 데이터 테이블의 구조와 동일한 구조의 연구를 가정하므로 후향적 검정력 분석입니다. 예를 들어 이 예의 검정력 계산은 모형에 입력된 효과와 각 연령(age) 및 성별(sex) 그룹의 참가자 수에 따라 달라집니다. 또한 σ 값은 향후 연구의 대표 값으로 대체할 수 있지만 현재 연구에서 도출되었습니다.

그림 3.70에 표시된 검정력 결과에 대한 자세한 내용은 "검정력"(196 페이지)에서 확인하십시오. LSN(최소 유의 개수)에 대한 자세한 내용은 "LSN(최소 유의 개수)"(195 페이지)에서 확인하십시오.

전향적 검정력 분석

전향적 분석은 "지정된 크기의 차이가 존재하는 경우 제안된 표본 크기, 유의 수준 및 오차 분산 추정값이 주어지면 차이를 감지할 수 있습니까?"라는 질문에 답하는 데 도움이 됩니다. 전향적 검정력 분석에서는 그룹 평균 및 표본 크기의 추정값을 데이터 테이블에 제공해야 합니다. 또한 "검정력 상세 정보 창"에서 오차 표준편차 σ 의 추정값도 제공해야 합니다.

동일한 그룹 크기

독립된 세 개 그룹의 평균을 비교하는 경우를 가정해 보겠습니다. 지정된 검정력을 달성하기 위한 표본 크기를 계산하려면 DOE > 표본 크기 및 검정력을 선택한 후 k 표본 평균을 선택합니다. 그런 다음 "표준편차" 옆에 오차 표준편차의 추정값을 입력합니다. 감지할 최소 차이를 반영하는 평균을 전향적 평균 목록에 입력합니다. 예를 들어 두 평균 사이의 8 단위 차이를 감지하려면 평균의 극단값을 입력합니다(예: 40, 40, 48). 검정력은 총 평균으로부터의 편차를 기반으로 하므로 원하는 차이를 반영하는 값만 입력할 수 있습니다(예: 0, 0, 8).

계속을 클릭하면 검정력 대 표본 크기를 보여 주는 그래프가 표시됩니다. 대신 "표본 크기" 창에서 검정력 또는 표본 크기를 지정하면 다른 통계량이 계산되어 "표본 크기" 창에 표시됩니다. 특히 검정력을 지정할 경우 제공되는 표본 크기는 필요한 총 표본 크기입니다. k 표본 평균 계산은 동일한 그룹 크기를 가정합니다. 세 그룹의 경우 표본 크기를 3으로 나누어 개별 그룹 크기를 구합니다. K 표본 평균에 대한 자세한 내용은 **실험 설계 가이드**의 에서 확인하십시오.

동일하지 않은 그룹 크기

크기가 다른 그룹을 사용하는 연구를 설계한다고 가정해 보겠습니다. 박테리아 수를 줄이는 것으로 알려진 두 가지 처리를 연구하기 위한 실험을 계획해야 합니다. 이러한 처리 효과를 처리가 적용되지 않은 대조군의 결과와 비교하려고 합니다. 또한 둘 중 한 처리 그룹과 대조군의 평균 간에 최소 8 단위의 차이를 감지하려고 합니다. 그러나 대조군은 각 처리 그룹의 두 배여야 합니다. 또한 두 처리 그룹은 크기가 같아야 합니다. 이전 연구에 따르면 오차 표준편차는 약 5 또는 6입니다.

이 상황에 대한 전향적 검정력 분석을 수행하려면 **Bacteria.jmp** 샘플 데이터 테이블에 표시된 몇 가지 기본 정보를 포함하는 데이터 테이블을 생성합니다.

그림 3.71 Bacteria.jmp 데이터 테이블

	Group	Group Means	Relative Sizes
1	Control	40	2
2	Treatment A	40	1
3	Treatment B	48	1

- **Group** 열은 그룹을 식별합니다.
- **Means** 열은 열 간에 감지해야 할 최소 차이를 반영합니다. 여기서 대조군의 평균은 약 40이라고 가정합니다. 둘 중 한 처리 그룹의 평균이 대조군의 평균보다 8 단위 이상 크면 검정이 유의하다는 결론을 내리려고 합니다. 이러한 이유로 두 처리 그룹 중 하나에 평균 48을 할당하고 다른 처리 그룹의 평균은 대조군의 평균과 동일하게 설정합니다. 아니면 대조군과 처리 그룹 중 하나에 평균 0을 할당하고 나머지 처리 그룹에 평균 8을 할당할 수 있습니다. 그룹 평균의 차이는 모집단 값입니다.
- **Relative Sizes** 열에는 처리 그룹의 원하는 상대 크기가 표시됩니다. 이 열은 대조군이 각 처리 그룹의 두 배여야 함을 나타냅니다. 아니면 상대 크기 기준을 준수하는 처리 크기에 대한 초기 추측 값으로 시작할 수 있습니다.

참고 : **Relative Sizes** 열에는 "빈도" 역할이 할당되어야 합니다. "열" 패널에서 열 이름 오른쪽에 있는 기호를 확인하십시오.

그런 다음 모형 적합을 사용하여 일원 분산 분석 모형을 적합시킵니다(그림 3.72). 시작 창에서 **Relative Sizes**가 **빈도**로 선언되어 있습니다. 또한 "최소 보고서" 강조 옵션이 선택되어 있습니다.

그림 3.72 박테리아 연구에 대한 모형 적합 시작 창

모형 규격

열 선택

6개 열

- Group
- Group Means
- Relative Sizes
- Grand Mean
- Contributions to Delta Squared
- Delta

역할 변수 선택

Y

- Group Means (선택적)

가중치

- Relative Sizes (선택적 숫자)

검증

- Relative Sizes (선택적 숫자)

기준

- 선택적

분석법: 표준 최소 제곱

강조: 최소 보고서

도움말 실행

재호출

대화상자 열린 채 유지

제거

모형 효과 생성

추가

교차

내포

마이크로

자수 2

속성

변환

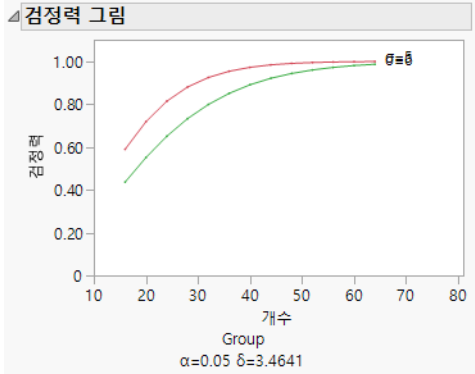
결과 없음

실행을 클릭하여 "최소 제곱 적합" 보고서를 생성합니다. 그룹 내에 오차 변동이 없는 데이터 테이블을 지정했으므로 이 보고서에는 "제공된 평균 제공 오차" 및 오차의 제곱합이 0.0으로 표시됩니다. 검정력 분석을 얻으려면 오차 변동에 대해 제안된 값 범위를 입력해야 합니다. 특히 오차 변동이 약 5에서 최대 6이 될 수 있다는 정보가 있습니다.

1. "효과 상세 정보" 옆의 표시 아이콘을 클릭하여 이 보고서를 엽니다.
2. "Group"의 빨간색 삼각형을 클릭하고 **검정력 분석**을 선택합니다.
3. 의심되는 오차 변동 범위를 탐색하려면 σ 아래의 첫 번째 상자에 5를 입력하고 두 번째 상자에 6을 입력합니다 (그림 3.73).
4. δ 는 3.464102로 입력되어 있습니다. 이 값은 그룹 평균의 지정된 차이에 해당하는 효과 크기입니다. 데이터 테이블에는 효과 크기 계산을 보여 주는 세 개의 숨겨진 열이 포함되어 있습니다. 자세한 내용은 "**불균형 일원 배치**" (194 페이지)에서 확인하십시오.
5. 해당 연구 크기 범위에 대해 검정력을 탐색하려면 **개수** 아래의 첫 번째 상자에 16, 두 번째 상자에 64, 세 번째 상자에 증분 4를 각각 입력합니다 (그림 3.73).
6. **검정력 계산**을 선택합니다.
7. **완료**를 클릭합니다.

"검정력 상세 정보" 보고서의 테이블 맨 아래에 있는 화살표를 클릭하면 그림 3.75에 표시된 것과 같이 σ 의 두 값에 대한 검정력 대 표본 크기 그림이 표시됩니다. 여기서 빨간색 표식은 $\sigma=5$ 에 해당하고 녹색 표식은 $\sigma=6$ 에 해당합니다.

그림 3.75 박테리아 연구에 대한 검정력 그림



4 장

표준 최소 제곱 예 모형의 공통 클래스 분석

이 장에서는 몇 가지 표준 최소 제곱 항목에 대해 지침 자료를 이용한 예를 제공합니다. 여기에는 분산 분석, 공분산 분석, 반응 표면 모형, 분할구 설계, 임의 효과 모수 추정, 매듭이 지정된 스플라인 적합, Bayes 방법을 제공하는 스크립트를 사용한 활성 요인 식별 등이 포함됩니다.

목차

일원 분산 분석 예.....	205
동일한 기울기의 공분산 분석 예	207
서로 다른 기울기의 공분산 분석 예.....	209
반응 표면 모형 예.....	212
전체 반응 표면 모형 적합	212
모형 축소.....	213
반응 표면 보고서 검토.....	214
예측 프로파일러를 사용하여 임계점 찾기	216
등고선 프로파일러를 사용하여 표면 보기	217
분할구 설계 예	219
임의 효과 모수 추정 예	223
매듭이 지정된 스플라인 효과 예	225
활성 요인에 대한 Bayes 그림 예	227

일원 분산 분석 예

일원 분산 분석에서는 명목형 변수에 의해 식별된 서로 다른 각 그룹에 서로 다른 평균이 적합됩니다. JMP 모형을 지정하려면 연속형 반응 열과 명목형 효과 열을 선택합니다. 이 예에서는 Drug.jmp의 데이터를 사용합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Drug.jmp를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. **Y**를 선택하고 **Y**를 클릭합니다.
4. **Drug**를 선택하고 **추가**를 클릭합니다.
5. **실행**을 클릭합니다.

이 예에서 Drug에는 a, d, f라는 세 가지 수준이 있습니다. 표준 최소 제곱 적합 방법은 이 규격을 선형 모형으로 변환합니다. 명목형 변수는 1, 0, -1 값만 가정하는 일련의 표시 변수를 정의합니다. 선형 모형은 다음과 같이 지정됩니다.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

다음은 각 요소에 대한 설명입니다.

- y_i 는 i 번째 관측값에 대한 관측된 반응입니다.
- x_{1i} 는 i 번째 관측값에 대한 첫 번째 표시 변수의 값입니다.
- x_{2i} 는 i 번째 관측값에 대한 두 번째 표시 변수의 값입니다.
- $\beta_0, \beta_1, \beta_2$ 는 각각 절편, 첫 번째 표시 변수, 두 번째 표시 변수에 대한 모수입니다.
- ε_i 는 독립적이며 정규 분포를 따르는 오차 항입니다.

첫 번째 표시 변수 x_1 은 다음과 같이 정의됩니다. 여기서 Drug = a는 값 1, Drug = d는 값 0, Drug = f는 값 -1을 각각 표시 변수에 적용합니다.

$$x_{1i} = \begin{cases} 1, & \text{Drug} = a \\ 0, & \text{Drug} = d \\ -1, & \text{Drug} = f \end{cases}$$

두 번째 표시 변수 x_2 에는 다음 값이 제공됩니다.

$$x_{2i} = \begin{cases} 0, & \text{Drug} = a \\ 1, & \text{Drug} = d \\ -1, & \text{Drug} = f \end{cases}$$

이 파라미터화와 관련하여 세 가지 수준의 평균 추정값은 다음과 같이 정의됩니다.

$$\mu_a = \beta_0 + \beta_1$$

$$\mu_d = \beta_0 + \beta_2$$

$$\mu_f = \beta_0 - \beta_1 - \beta_2$$

β_i 를 계산하면 다음과 같이 됩니다.

$$\beta_0 = \frac{(\mu_a + \mu_d + \mu_f)}{3} = \mu \quad (\text{전체 수준의 평균})$$

$$\beta_1 = \mu_a - \mu$$

$$\beta_2 = \mu_d - \mu$$

따라서 회귀변수가 각 수준의 표시자에서 마지막 수준의 표시자를 뺀 값으로 코드화되면 한 수준의 모수는 해당 수준의 반응과 모든 수준의 평균 반응 간의 차이로 해석됩니다. 명목형 요인의 파라미터화 해석에 대한 자세한 내용은 부록 "통계 상세 정보"(487페이지)에서 확인하십시오.

그림 4.1에서는 Drug 효과에 대한 레버리지 그림과 최소 제곱 평균 테이블을 보여 줍니다. 그림 4.2에서는 약물 데이터 일원 분석에 대한 모수 추정값 보고서와 효과 검정 보고서를 보여 줍니다.

그림 4.1 Drug 에 대한 레버리지 그림 및 최소 제곱 평균 테이블

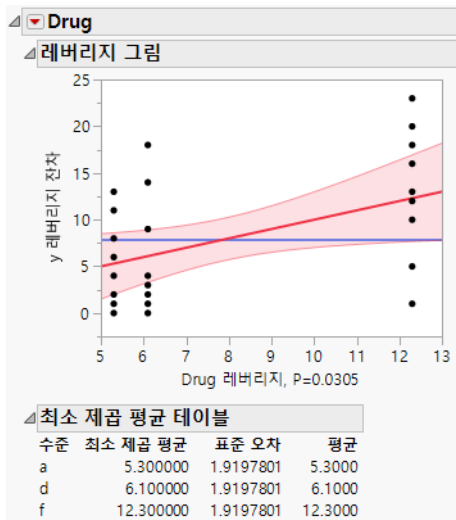


그림 4.2 Drug.jmp 에 대한 모수 추정값 및 효과 검정

▲ 모수 추정값

항	추정값	표준 오차	t 비	Prob> t
절편	7.9	1.108386	7.13	<.0001*
Drug[a]	-2.6	1.567494	-1.66	0.1088
Drug[d]	-1.8	1.567494	-1.15	0.2609

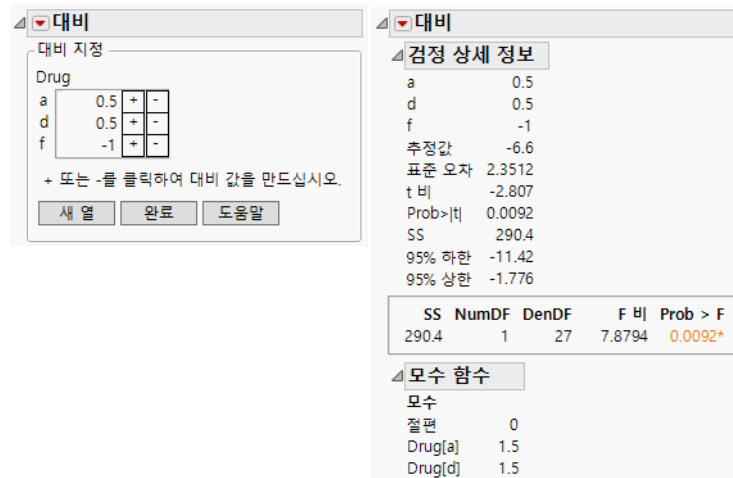
▲ 효과 검정

소스	N모수	DF	제곱합	F 비	Prob > F
Drug	2	2	293.60000	3.9831	0.0305*

다음과 같이 최소 제곱 평균의 대비를 사용하여 Drug 효과를 더 자세히 연구할 수 있습니다.

1. "Drug"의 빨간색 삼각형을 클릭하고 **최소 제곱 평균 대비 ...**를 선택합니다.
2. Drug a 및 d의 + 상자를 클릭하고 Drug f의 - 상자를 클릭하여 약물 a와 d의 평균을 f와 비교하는 대비를 정의합니다 (그림 4.3 참조).
3. **완료**를 클릭합니다.

그림 4.3 Drug 실험에 대한 대비 예



"대비" 보고서에서는 약물 f의 최소 제곱 평균이 다른 두 약물에 대한 최소 제곱 평균의 평균과 유의하게 다르다는 것을 보여 줍니다.

동일한 기울기의 공분산 분석 예

연속형 회귀변수 항이 있는 분산 분석 모형을 공분산 분석이라고 합니다. Drug.jmp 샘플 데이터 테이블에서 x 열은 공변량입니다.

공변량은 모형 방정식에 x_{3i} 항을 추가합니다. 공분산 분석 모형은 다음과 같이 작성됩니다.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i$$

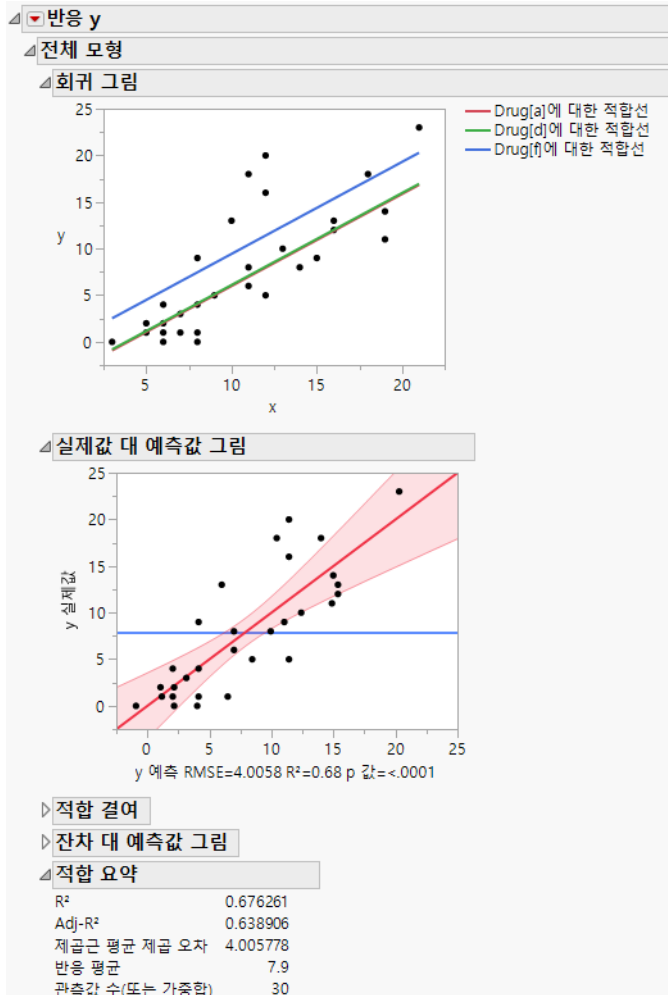
두 가지 모형 효과가 있습니다. 하나는 두 모수를 포함하는 명목형 주효과이고 다른 하나는 한 모수와 관련된 연속형 공변량입니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Drug.jmp를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. Y를 선택하고 **Y**를 클릭합니다.
4. Drug와 x를 모두 선택하고 **추가**를 클릭합니다.

5. 실행을 클릭합니다.

보고서의 "회귀 그림"에서는 기울기가 동일한 모형이 적합되었음을 보여 줍니다(그림 4.4). 주 효과 모형(Drug 효과만)과 비교하여 R^2 은 22.8%에서 67.6%로 증가합니다. 제공된 평균 제곱 오차는 6.07에서 4.0으로 감소합니다. 그림 4.4에서 볼 수 있듯이 전체 모형에 대한 F -검정 유의 확률은 0.03에서 0.0001 미만으로 감소합니다.

그림 4.4 약물 데이터에 대한 공분산 분석



약물 데이터 테이블에는 반복된 관측값이 포함되어 있습니다. 예를 들어 행 1과 9는 둘 다 Drug = a 이고 x = 11 입니다. 모형을 적합시킬 때 반복된 관측값을 사용하여 변동의 순수 오차 추정값을 생성할 수 있습니다. 공변량의 지정되지 않은 함수 형식 또는 명목형 효과의 교호작용에 대한 또 다른 오차 추정값을 생성할 수 있습니다. 이러한 추정값은 적합 결여 검정의 기반이 됩니다. 적합 결여 오차가 유의하면 데이터의 일부 효과가 모형에 의해 설명되지 않음을 나타냅니다.

"적합 결여" 보고서에는 약물 데이터에 대한 이 검정의 결과가 표시됩니다. Prob > F 값 = 0.7507에서 알 수 있듯이 적합 결여 오차는 유의하지 않습니다.

공변량 x 는 이전에 Drug 변수에 의해 설명된 반응 변동의 대부분을 설명합니다. 따라서 모형이 훨씬 적은 오차로 적합되더라도 Drug 효과는 더 이상 유의하지 않습니다. 주효과 모형에서 관측된 Drug의 유의성은 공변량에 의해 설명되는 것으로 보입니다.

공분산 모형의 최소 제곱 평균은 일반 평균과 다릅니다. 이 최소 제곱 평균은 공변량 x 가 반응 y 에 미치는 효과에 대해 조정되기 때문입니다. 최소 제곱 평균은 공변량 x 가 어떤 중립 값으로 고정될 때 Drug의 세 가지 각 수준에 대한 예측값입니다. 중립 값은 공변량의 평균인 10.7333으로 선택됩니다.

최소 제곱 평균은 "모수 추정값" 보고서에 제공된 모수 추정값을 사용하여 계산됩니다.

예측 표현식 $-2.696 - 1.185 \cdot \text{Drug}[a] - 1.0761 \cdot \text{Drug}[d] + 0.98718 \cdot x$

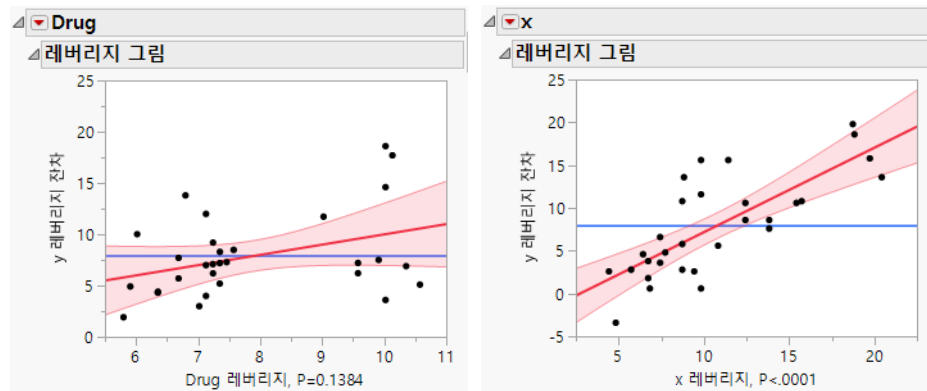
a: $-2.696 - 1.185 \cdot (1) - 1.0761 \cdot (0) + 0.98718 \cdot (10.7333) = 6.71$

d: $-2.696 - 1.185 \cdot (0) - 1.0761 \cdot (1) + 0.98718 \cdot (10.7333) = 6.82$

f: $-2.696 - 1.185 \cdot (-1) - 1.0761 \cdot (-1) + 0.98718 \cdot (10.7333) = 10.16$

그림 4.5에서는 각 효과에 대한 레버리지 그림을 보여 줍니다. 공변량이 유의하므로 Drug의 레버리지 값은 최소 제곱 평균에서 어느 정도 분산되어 있습니다.

그림 4.5 약물 검정 데이터에 대한 레버리지 그림 비교



서로 다른 기울기의 공분산 분석 예

이 예에서는 Drug.jmp 샘플 데이터 테이블을 계속 사용하여 공변량의 기울기가 Drug 수준에 따라 달라지는 모형을 적합시킵니다. 모형을 적합시킨 후 공변량 x 의 특정 값에서 Drug 수준에 대한 최소 제곱 평균을 비교합니다.

1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Drug.jmp를 엽니다.
2. 분석 > 모형 적합을 선택합니다.

- 3. Y 를 선택하고 Y 를 클릭합니다 .
- 4. Drug 와 x 를 모두 선택하고 매크로 > 특정 차수까지의 요인을 클릭합니다 .
이렇게 하면 차수 상자에 지정된 차수까지의 항이 모형에 추가됩니다 . 기본 차수 값은 2 입니
다 . 따라서 주효과 Drug, x 및 교호작용 Drug*x 가 모형 효과 목록에 추가됩니다 .
- 5. 실행을 클릭합니다 .

이 규격은 공변량의 기울기를 Drug 수준에 따라 달라지게 하는 두 개의 열을 선형 모형에 추가합
니다(x_{4i} 및 x_{5i}). Drug에 대한 표시 변수에 공변량 값을 곱하여 새 변수가 생성되고 다음 계산식을
제공합니다.

$$y_i = \beta_0 + \beta_1x_{1i} + \beta_2x_{2i} + \beta_3x_{3i} + \beta_4x_{4i} + \beta_5x_{5i} + \epsilon_i$$

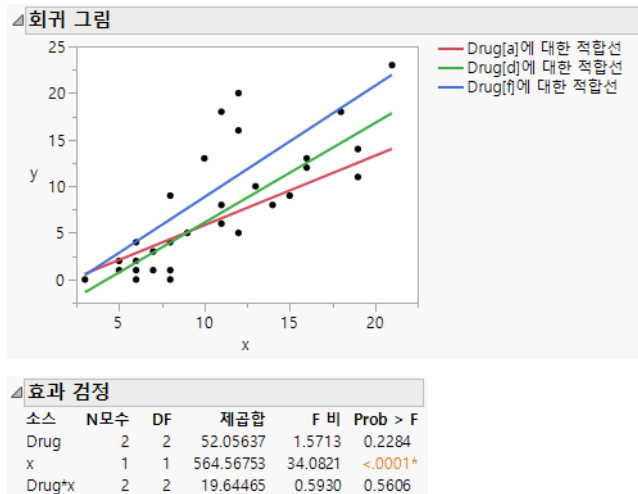
표 4.1에서는 이 모형의 코딩을 보여 줍니다. X의 평균은 10.7333입니다. 이 값은 연속형 항을 중
심화하는 데 사용됩니다.

표 4.1 서로 다른 기울기의 공분산 분석 코딩

회귀변수	효과	값
X_1	Drug[a]	a: +1, d: 0, f: -1
X_2	Drug[d]	a: 0, d: +1, f: -1
X_3	X	X의 값
X_4	Drug[a]*(X - 10.733)	a: X - 10.7333, d: 0, f: -(X - 10.7333)
X_5	Drug[d]*(X - 10.733)	a: 0, d: X - 10.7333, f: -(X - 10.7333)

그림 4.6에 보고서의 일부가 표시되어 있습니다. "회귀 그림"에서는 기울기가 다른 적합선을 보
여 줍니다. "효과 검정" 보고서에서는 교호작용의 p 값 0.56을 제공합니다. 이것은 유의하지 않
으므로 모형에 서로 다른 기울기를 포함할 필요가 없음을 나타냅니다.

그림 4.6 교호작용 그림



스포트라이트 분석 수행

이제 공변량 x 의 특정 값에서 Drug 수준에 대한 최소 제곱 평균을 비교하려고 합니다. 공분산 분석 모형에서 이러한 유형의 비교를 스포트라이트 분석이라고도 합니다. 스포트라이트 분석에 대한 자세한 내용은 Spiller et al. 연구 자료 (2013)에서 확인하십시오.

1. "반응 y "의 빨간색 삼각형을 클릭하고 **추정값 > 다중 비교**를 선택합니다.
2. "다중 비교" 창에서 **사용자 정의 추정값**을 선택합니다.
3. **Drug 수준 선택** 아래의 세 값을 모두 선택합니다.
4. x 아래의 첫 번째 상자에 12.5를 입력합니다.
5. **추정값 추가**를 클릭합니다.

그러면 $x = 12.5$ 에서 Drug의 세 수준 비교가 추가됩니다.

6. **확인**을 클릭합니다.

"사용자 정의 추정값" 보고서에는 공변량 x 가 12.5로 설정된 각 Drug 수준의 최소 제곱 평균 추정값이 표시됩니다. "사용자 정의 추정값에 대한 다중 비교" 옆의 빨간색 삼각형에는 추정값 간의 차이를 검정할 수 있는 옵션이 포함되어 있습니다.

그림 4.7 사용자 정의 추정값 보고서

사용자 정의 추정값에 대한 다중 비교

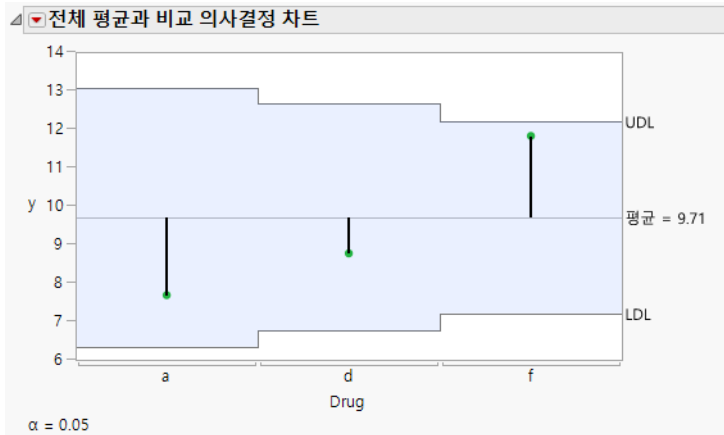
사용자 정의 추정값

$x = 12.5$

Drug	추정값	표준 오차	DF	95% 하한	95% 상한
a	7.684713	1.5772057	24	4.4295208	10.939906
d	8.771371	1.4401229	24	5.7991033	11.743639
f	11.822214	1.2943345	24	9.1508392	14.493590

7. "사용자 정의 추정값에 대한 다중 비교" 옆의 빨간색 삼각형을 클릭하고 **전체 평균과 비교**를 선택합니다.

그림 4.8 전체 평균과 비교 의사결정 차트



"전체 평균과 비교" 옵션은 평균과 세 가지 최소 제곱 평균 간의 차이에 대한 ANOM(평균 분석) 차트를 생성합니다. ANOM 차트를 참조하여 $x = 12.5$ 에서 반응에 대한 Drug 효과는 유의하지 않다고 결론을 내릴 수 있습니다.

반응 표면 모형 예

이 예에서는 반응 표면 모형을 적합시킵니다. 반응을 최소화하는 것이 목적입니다.

- "전체 반응 표면 모형 적합"
- "모형 축소"
- "반응 표면 보고서 검토"
- "예측 프로파일러를 사용하여 임계점 찾기"
- "등고선 프로파일러를 사용하여 표면 보기"

전체 반응 표면 모형 적합

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Custom RSM.jmp를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.

데이터 테이블에 **Model** 스크립트가 포함되어 있어서 "모형 규격" 창이 **Model** 스크립트에 지정된 대로 채워집니다. 다음 사항에 유의하십시오.

- 주효과는 **모형 효과 생성** 목록에 **&RS** 접미사가 추가된 상태로 표시되며 이는 반응 표면 매크로가 적용되었음을 나타냅니다.

- 세 가지 예측 변수 X_1, X_2, X_3 의 전체 반응 표면에 대한 효과가 표시됩니다.
- **&RS** 접미사가 추가된 항이 모형에 포함되어 있으므로 분석 결과에 " 반응 표면 " 보고서가 포함됩니다.

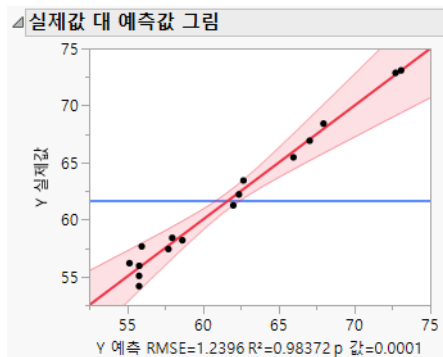
그림 4.9 반응 표면 분석을 위한 모형 적합 시작 창

3. 실행을 클릭합니다.

모형 축소

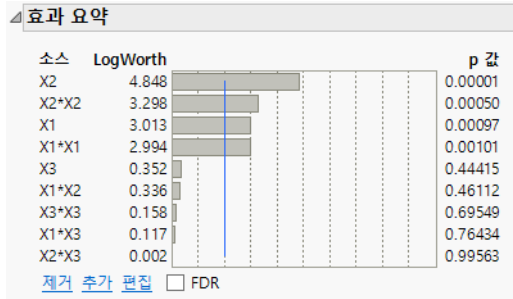
"실제값 대 예측값 그림"에서는 모형이 유의함을 보여 줍니다. 적합 결여를 나타내는 증거가 없습니다.

그림 4.10 전체 모형에 대한 실제값 대 예측값 그림



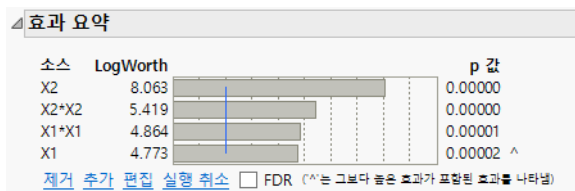
"효과 요약" 보고서에서는 여러 효과가 유의하지 않음을 나타냅니다. 특히 $X2 \times X3$ 은 최소 유의 효과입니다(p 값 = 0.99563). 이제 대화식 "효과 요약" 보고서를 사용하여 모형을 체계적으로 축소합니다.

그림 4.11 효과 요약 보고서



- "효과 요약" 보고서에서 $X2 \times X3$ 을 클릭하고 **제거**를 클릭합니다.
모형이 업데이트됩니다.
"효과 요약" 보고서의 "p 값" 열에 따르면 $X1 \times X3$ 이 유의하지 않습니다.
- $X1 \times X3$ 을 클릭하고 **제거**를 클릭합니다.
 $X3 \times X3$ 의 p 값은 이 효과가 유의하지 않음을 나타냅니다.
- $X3 \times X3$ 을 클릭하고 **제거**를 클릭합니다.
- $X1 \times X2$ 를 클릭하고 **제거**를 클릭합니다.
 $X3$ 이 유의하지 않습니다. 이 효과는 고차 효과에 포함되지 않으므로 효과 유전성 원리를 위반하지 않고 제거할 수 있습니다. 자세한 내용은 "표준 최소 제곱 모형" 장의 "**효과 유전성**"(171 페이지)에서 확인하십시오.
- $X3$ 을 클릭하고 **제거**를 클릭합니다.

그림 4.12 모형 축소 후 효과 요약 보고서



나머지 효과는 모두 유의합니다.

반응 표면 보고서 검토

- "반응 Y"의 빨간색 삼각형을 클릭하고 **추정값 > 예측 표현식 표시**를 선택합니다.

그림 4.13 예측 표현식

예측 표현식	
55.721136364	
+ 2.349 • X1	
+ 5.003 • X2	
+ X1 • (X1 • 4.4365909091)	
+ X2 • (X2 • 5.0765909091)	

팁 : 예측 표현식은 예측 계산식입니다 . **열 저장 > 예측 계산식**을 선택하여 이 계산식을 얻을 수도 있습니다 .

다음 단계에서는 예측 표현식을 참조하여 모형 계수를 확인할 수 있습니다 .

2. " 반응 표면 " 보고서와 " 정준 곡률 " 보고서를 차례로 엽니다 .

그림 4.14 반응 표면 보고서

반응 표면				
계수				
	X1	X2		Y
X1	4.4365909	0		2.349
X2		5.0765909		5.003

해	
변수	임계값
X1	-0.26473
X2	-0.492752
해: 최소값	
해의 예측값 54.177592	

정준 곡률		
	고유값 및 고유 벡터	
고유값	5.0766	4.4366
X1	0.00000	1.00000
X2	1.00000	0.00000

첫 번째 테이블에는 2 차 모형 계수가 행렬 형식으로 제공됩니다 . $X1 \cdot X1$ 의 계수는 4.4365909, $X2 \cdot X2$ 의 계수는 5.0765909, $X1 \cdot X2$ 의 계수는 0 입니다 . 선형 효과 계수 ($X1$ 의 2.349, $X2$ 의 5.003) 는 Y 라벨이 지정된 열에 제공됩니다 .

" 해 " 보고서에는 임계값이 표시됩니다 . 이러한 값은 최대값 , 최소값 또는 안장점이 발생하는 지점을 나타냅니다 . 이 예의 " 해 " 보고서에서는 $X1 = -0.265$, $X2 = -0.493$ 의 임계값에서 반응 표면의 최소값이 54.18 임을 보여 줍니다 .

" 정준 곡률 " 보고서에는 2 차 모수 추정값 행렬의 고유 구조가 표시됩니다 . 고유 구조는 곡률의 형태와 방향을 식별하는 데 유용합니다 . 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 정준 곡률 보고서 "(81 페이지) 에서 확인하십시오 .

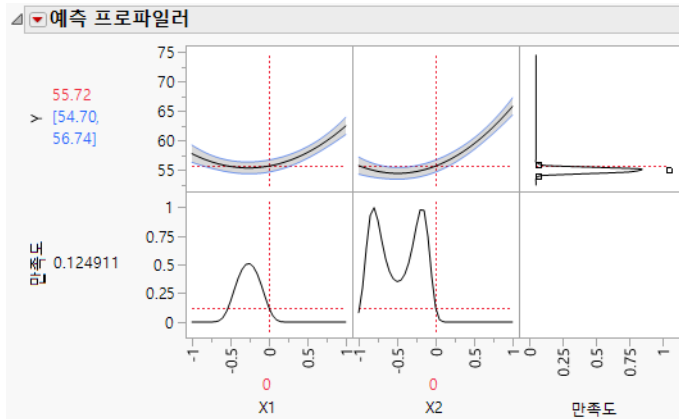
이 예에서는 두 고유값이 모두 양수이므로 표면이 최소값을 얻는다는 것을 나타냅니다 . 가장 큰 곡률의 방향은 최대 고유값 (5.0766) 에 해당합니다 . 방향은 해당하는 고유 벡터 성분에 의

해 정의됩니다. 첫 번째 방향의 경우 고유 벡터 값이 1.00 인 X2 가 방향을 결정합니다. 두 번째 방향은 고유 벡터 값이 1.00 인 X1 에 의해 결정됩니다.

예측 프로파일러를 사용하여 임계점 찾기

"예측 프로파일러" 보고서에는 X1 및 X2에 대한 트레이스를 따라 반응 표면의 2차 동작이 표시됩니다. Y에 "반응 한계" 열 특성이 설정되어 있으므로 프로파일러에 만족도 함수도 표시됩니다.

그림 4.15 목표값 일치가 목표인 예측 프로파일러



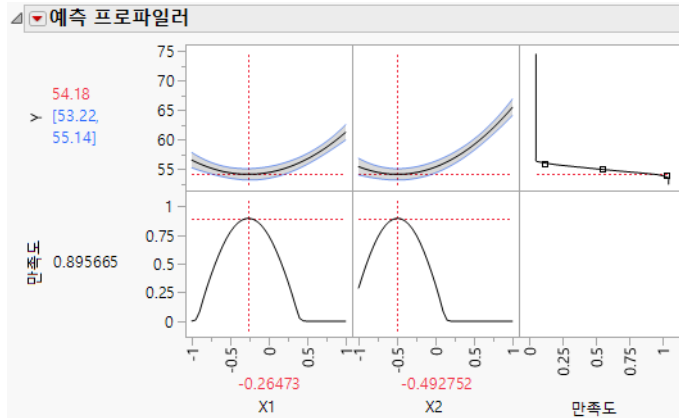
"반응 한계" 열 특성의 목표는 "목표값 일치"로 설정됩니다. 이 예에서는 목표값 일치 대신 Y를 최소화하는 데 관심이 있습니다. 따라서 다음과 같이 설정을 변경해야 합니다.

1. Ctrl 키를 누르고 "예측 프로파일러"의 오른쪽 위 셀을 클릭합니다.
2. "반응 목표" 대화상자의 옵션 목록에서 **최소화**를 선택합니다.
3. **확인**을 클릭합니다.

이제 만족도 함수에 Y 최소화 목표가 반영됩니다.

4. "예측 프로파일러"의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다.

그림 4.16 최소화가 목표이고 만족도가 최대화된 예측 프로파일러



설계 영역 내에서 Y를 최소화하는 설정이 프로파일러 아래에 나타납니다. 이 설정은 "해" 보고서에 제공된 "임계값"과 정확히 일치합니다.

등고선 프로파일러를 사용하여 표면 보기

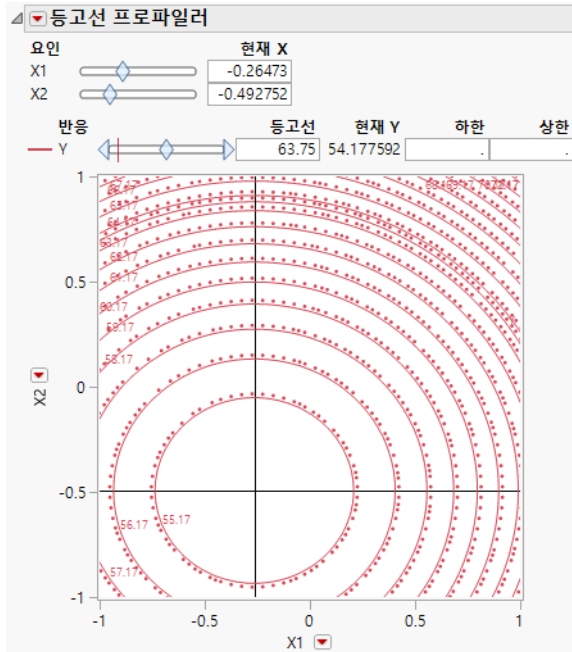
"등고선 프로파일러"에는 반응 표면의 등고선이 표시됩니다. 이는 임계점 영역에서 예측 반응에 대한 대체 프로파일러 시각화를 제공합니다.

1. "반응 Y"의 빨간색 삼각형을 클릭하고 **요인 프로파일링 > 등고선 프로파일러**를 선택합니다.
2. "등고선 프로파일러"의 빨간색 삼각형을 클릭하고 **등고선 격자**를 선택합니다.
3. **증분**에 1을 입력합니다.
4. **확인**을 클릭합니다.

등고선이 1 단위 간격으로 표시됩니다. 자세한 내용은 그림 4.17에서 확인하십시오.

5. "예측 프로파일러"의 빨간색 삼각형을 클릭하고 **요인 설정 > 연결 프로파일러**를 선택합니다.

그림 4.17 임계점에 십자기호가 있는 등고선 프로파일러

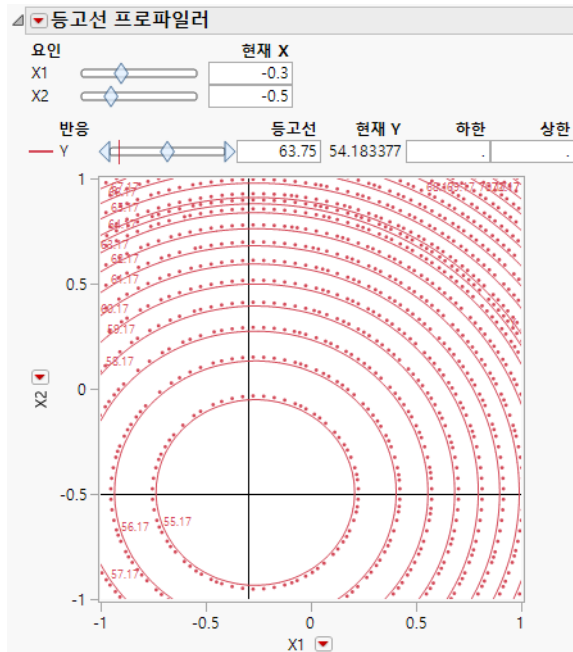


등고선 프로파일러를 예측 프로파일러에 연결하면 등고선 프로파일러의 **현재 X** 값이 예측 프로파일러에 표시된 X 값에 연결됩니다. 예측 프로파일러의 X 값은 Y 가 최소화되는 임계점을 제공합니다. 등고선 프로파일러의 십자기호는 임계점을 표시합니다. **현재 Y** 값은 예측 프로파일러에 따라 예측된 최소값인 54.177592 입니다.

공정 요인의 값을 반응을 최적화하는 값으로 정확하게 설정할 수 없는 경우가 종종 있습니다. 등고선 프로파일러를 사용하면 공정 요인의 대체 설정을 식별하는 데 도움이 됩니다. 다음 단계에서는 공정을 소수점 이하 한 자리만 사용하여 X1 및 X2 값으로 설정할 수 있으며, 공정 설정은 방향에 관계없이 소수점 이하 한 자리씩 변할 수 있다고 가정합니다.

6. "등고선 프로파일러" 보고서의 **현재 X** 아래에서 X1 옆에 -0.3 을 입력하고 X2 옆에 -0.5 를 입력합니다.

그림 4.18 $X_1 = -0.3, X_2 = -0.5$ 를 표시하는 등고선 프로파일러



십자기호는 가장 안쪽 등고선 내에 있으며 현재 Y(현재 X 설정의 예측 Y 값) 는 예측 최소값 54.177592 와 약간 다른 54.183377 입니다 .

7. 등고선 프로파일러에서 십자기호를 클릭하고 드래그하여 $X_1 = -0.3$ 및 $X_2 = -0.5$ 의 0.1 단위 반지름 내에서 현재 X 값을 탐색합니다 .

예측 Y 값은 모두 54.4 미만입니다 . 실제로 설정이 가장 안쪽 등고선 내의 어떤 한 지점까지 이동하면 예측 Y 가 등고선 값 55.17 보다 작습니다 .

분할구 설계 예

임의 효과 수준은 더 많은 수준의 모집단에서 무작위로 선택됩니다 . 추론을 위해 임의 효과의 분포를 평균 0 과 특정 분산 (분산 성분이라고 함) 의 정규 분포로 가정합니다 .

어떤 의미에서는 잔차 오차를 구성하는 임의 효과가 모든 모형에 하나 이상 있다고 할 수 있습니다 . 개별 관측값은 훨씬 더 큰 모집단에서 무작위로 선택된다고 가정하고 오차 항은 평균이 0 이고 분산이 σ^2 인 것으로 가정합니다 .

가장 일반적인 임의 효과 모형은 반복 측정 또는 분할구 모형입니다 . 표 4.2 에는 분할구 모형의 효과 유형이 나열되어 있습니다 . 이러한 모형을 사용하는 경우 실험은 두 가지 계층을 포함합니다 . 일부 효과를 실험의 주구 또는 개체에 적용합니다 . 그런 다음 이러한 주구를 분할하거나 서로 다른 시간에 개체를 측정하고, 나머지 효과를 해당 하위 단위 내에 적용합니다 . 주구 또는 개

체를 설명하는 효과는 주구 효과이고 하위구 또는 반복 측정을 설명하는 효과는 하위구 효과입니다. 일반적으로 하위 단위 효과는 모형에서 생략되고 잔차 오차로 흡수됩니다.

표 4.2 분할구 모형의 효과 유형

분할구 모형	효과 유형	반복 측정 모형
주구 처리	고정 효과	개체 간 처리
주구 ID	임의 효과	개체 ID
하위구 처리	고정 효과	개체 내 처리
하위구 ID	임의 효과	반복 측정 ID

이러한 각 사례는 계층화된 모형으로 처리할 수 있으며 몇 가지 기존 방식을 사용하여 타당하게 모형을 적합시킬 수 있습니다. 다음과 같이 두 가지 다른 실험으로 처리됩니다.

1. 주구 실험은 주구 또는 개체를 실험 단위로 사용하여 오차 항을 구성합니다.
2. 하위구 처리는 개별 측정을 실험 단위로 사용하여 오차 항을 구성합니다(잔차 오차로 유지).

주구를 검정하는 기존 방식은 다음 중 하나를 수행하는 것입니다.

- 전체 측정에서 평균을 구하여 주구 효과에 적합시킵니다.
- 주구 평균 제곱을 주구 ID 평균 제곱으로 나누어 F 비를 계산합니다.
- 분할 또는 반복 측정으로 서로 다른 열이 생성되도록 데이터를 구성합니다. 그런 다음 MANOVA 모형을 적합시키고 단변량 통계량을 사용합니다.

이러한 방식은 단순한 구조이거나 데이터가 완전하고 균형을 이룰 때 사용할 수 있습니다. 그러나 임의 효과의 구조에 작동하는 더 일반적인 모형이 있습니다. 이 일반화 모형은 고정 효과와 임의 효과를 모두 포함하므로 **혼합 모형**이라고 합니다.

계층화된 설계의 가장 일반적인 유형은 균형 분할구로, 대개 시간에 따른 반복 측정 형식입니다. 일부 효과에 대한 하나의 실험 단위가 세분화되고(대개 시간별) 다른 효과가 이 하위 단위에 적용됩니다.

Animals.jmp 샘플 데이터 테이블의 가상 데이터를 예로 들겠습니다. 이 연구에서는 여우와 코요테의 계절별 사냥 습관 차이에 대한 정보를 수집했습니다. 1년 동안 각 계절마다 여우 3마리와 코요테 3마리를 표적으로 삼아 주기적으로 관찰했습니다. 1년 중 다른 계절마다 굴을 돌아다니는 평균 거리(마일)를 기록하여 가장 근사한 값으로 반올림했습니다. 모형은 다음과 같은 특성으로 정의됩니다.

- 연속형 반응 변수 **miles**
- **fox** 또는 **coyote** 값을 갖는 **species** 효과
- **fall, winter, spring** 및 **summer** 값을 갖는 **season** 효과
- 여우와 코요테 둘 다에 대해 명목형 값 1, 2, 3을 갖는 동물 식별 코드 **subject**

모형에는 다음과 같은 두 개의 계층이 있습니다.

1. 상위 계층은 개체 간 계층으로, 개체 변화에 따라 여우 또는 코요테가 되는 효과 (species 효과) 가 검정됩니다.
2. 하위 계층은 개체 내 계층으로, 한 개체 내에서 계절 변화에 따라 4 가지 계절 (season 효과) 에 대한 반복 측정 요인이 검정됩니다. 개체 내 변동은 잔차 오차에 반영됩니다.

season 효과는 F 통계량의 분모에 잔차 오차를 사용할 수 있습니다. 그러나 개체 간 변동은 잔차 오차에 의해 측정되지 않으며 모형에서 species 내의 subject(subject[species]) 효과로 포착되어야 합니다. 개체 간 효과 species에 대한 F 통계량은 잔차 오차 대신 이 내포 효과를 F 비 분모에 사용합니다.

참고 : JMP Pro 사용자는 혼합 모형 분석법을 사용하여 이 모형을 생성할 수 있습니다.

이 데이터에 대한 분할구 모형을 지정하려면 다음 단계를 수행하십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Animals.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. miles 를 선택하고 **Y** 를 클릭합니다.
4. species 와 subject 를 선택하고 **추가**를 클릭합니다.
5. " 열 선택 " 목록에서 species 를 선택합니다.
6. " 모형 효과 생성 " 목록에서 subject 를 선택합니다.
7. **내포**를 클릭합니다.

그러면 species 내의 subject(subject[species]) 효과가 모형에 추가됩니다.

8. 내포 효과 subject[species] 를 선택합니다.
9. **속성 > 임의 효과**를 선택합니다.

이제 이 내포 효과는 species 효과의 오차 항으로 식별되고 "subject[species]& 랜덤 " 으로 표시됩니다.

10. " 열 선택 " 목록에서 season 을 선택하고 **추가**를 클릭합니다.

속성 메뉴를 사용하여 효과를 임의 효과로 정의하면 대화상자의 오른쪽 위에 " 방법 " 옵션 (**REML** 및 **EMS**) 이 나타납니다. **REML** 옵션이 기본적으로 선택되어 있습니다. 완료된 시작 창이 그림 4.19 에 표시되어 있습니다.

그림 4.19 모형 적합 대화상자

11. 실행을 클릭합니다.

그림 4.20 에 보고서가 표시되어 있습니다. 두 고정 효과 (species 와 season) 가 모두 유의합니다. "REML 분산 성분 추정값" 보고서에는 species 에 내포된 subject 의 추정값 및 잔차 분산이 제공됩니다.

그림 4.20 REML 분석 보고서 (일부분)

반응 miles

적합 요약

R ²	0.823497
Adj-R ²	0.786338
제곱근 평균 제곱 오차	1.219062
반응 평균	4.458333
관측값 수(또는 가중합)	24

모수 추정값

항	추정값	표준 오차	DFDen	t 비	Prob> t
절편	4.4583333	0.42287	4	10.54	0.0005*
species[COYOTE]	1.4583333	0.42287	4	3.45	0.0261*
season[fall]	-0.625	0.431003	15	-1.45	0.1676
season[spring]	1.7083333	0.431003	15	3.96	0.0012*
season[summer]	0.875	0.431003	15	2.03	0.0605

임의 효과 예측

REML 분산 성분 추정값

임의 효과	분산 비율	분산 성분	표준 오차	95% 하한	95% 상한	Wald p 값	총계 백분율
subject[species]	0.4719626	0.7013889	0.7707006	-0.809157	2.2119344	0.3628	32.063
잔차		1.4861111	0.5426511	0.8109483	3.5597535		67.937
합계		2.1875	0.8609382	1.1475492	5.700522		100.000

-2*로그 가능성도 = 78.806486054

참고: 합계는 양의 분산 성분의 합입니다.

음수 추정값을 포함한 합계 = 2.1875

분산 성분 추정값의 공분산 행렬

반복

고정 효과 검정

소스	N모수	DF	DFDen	F 비	Prob > F
species	1	1	4	11.8932	0.0261*
season	3	3	15	10.6449	0.0005*

효과 상세 정보

임의 효과 모수 추정 예

임의 효과에는 이중 특성이 있습니다. 임의 효과는 잔차 오차(예: 주구 실험 단위와 관련된 오차)를 나타낸다는 것이 한 가지 특성입니다. 다른 특성은 모수를 임의 효과의 각 수준과 연관시켜 고정 효과처럼 사용된다는 것입니다. 모수는 평균 0 및 분산 성분에 의해 추정된 분산을 사용하는 정규 분포에서 파생된다는 추가 정보가 있습니다. 이 추가 정보에 따라 모수 추정값은 0으로 축소됩니다.

임의 효과와 관련된 모수 추정값을 BLUP(최량 선형 비편향 예측 변수)라고 합니다. 이러한 BLUP는 연구원에 따라 관심 모수로 간주되거나 방법론의 흥미 없는 부산물로 간주되기도 합니다.

BLUP 모수 추정값은 임의 효과 최소 제곱 평균을 추정하는 데 사용되며 이러한 평균은 결국 총 평균으로 축소됩니다. 효과의 분산 및 효과의 수준당 관측값 수에 따라 축소 정도가 다릅니다. 분산 추정값이 크면 축소가 거의 없습니다. 분산 성분이 작으면 더 크게 축소됩니다. 분산 성분이 0이면 효과 수준이 정확히 0으로 축소됩니다. 축소가 반전되는 매우 큰 음의 분산 성분을 얻을 수도 있습니다. 고정 효과를 분산 성분이 매우 큰 특수 임의 효과로 간주할 수 있습니다.

REML 방법은 각 개별 수준에 대한 정보와 수준 간 분산에 대한 정보의 균형을 조정합니다. 수준당 관측값 수가 많으면 추정값이 덜 축소됩니다. 수준당 관측값 수가 매우 적으면 추정값이 더 축소됩니다. 관측값이 무한히 많은 경우에는 축소되지 않고 추정값이 고정 효과와 동일합니다.

여러 야구 선수의 타율 정보가 있다고 가정해 보겠습니다. 선수 간 타율 실적에 대한 분산 성분은 타율에서 선수 간의 일반적인 변동 크기를 설명합니다. 예를 들어 선수가 경기에 참여한 횟수가 몇 번뿐이면 타율이 비정상적으로 작거나 큼니다. 그러면 해당 추정값은 단지 몇 타석에만 근거하므로 신뢰할 수 없는 경우가 많습니다. 그러나 이 추정값을 총 평균과 혼합하면 추정값이 총 평균에 가까이 축소되므로 추정값을 더 신뢰할 수 있습니다. 오랜 타율 기록이 있는 선수의 경우 기록이 부족한 선수보다 훨씬 적게 축소할 수 있습니다.

이를 직접 확인할 수 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Baseball.jmp** 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. **Batting** 을 선택하고 **Y** 를 클릭합니다.
4. **Player** 를 선택하고 **추가**를 클릭합니다.
5. "모형 효과 생성" 상자에서 "Player" 를 선택하고 "속성" 목록에서 **임의 효과**를 선택합니다.
6. **실행**을 클릭합니다.

표 4.3에서는 REML(권장) 적합에 대한 Player 보고서의 최소 제곱 평균을 보여 줍니다. 또한 EMS 방법을 사용하여 계산된 적률 방법 추정값도 표시됩니다. 적률 방법 추정값은 일반적인 Player 평균입니다. 3타석 기록만 있는 Suarez에 대한 REML 추정값은 타석 수가 더 많은 다른 선수에 비해 총 평균에 더 가까워지도록 많이 축소됩니다.

표 4.3 적률 방법과 REML 간 추정값 비교

	적률 방법	REML	N
분산 성분	0.01765	0.019648	
Anderson	0.29500000	0.29640407	6
Jones	0.20227273	0.20389793	11
Mitchell	0.32333333	0.32426295	6
Rodriguez	0.55000000	0.54713393	6
Smith	0.35681818	0.35702094	11
Suarez	0.55000000	0.54436227	3
최소 제곱 평균	일반 평균과 같음	평균에서 축소됨	

매듭이 지정된 스플라인 효과 예

평활 다항식의 분할을 지정된 효과에 적합시키려면 **매듭이 지정된 스플라인 효과** 옵션을 사용합니다. 이 속성을 선택하면 매듭 점 수를 지정할 수 있는 창이 나타납니다. 매듭이 지정된 스플라인은 주효과 연속형 항에 대해서만 구현됩니다.

JMP는 점을 배치할 때 문헌의 권장 사항을 따릅니다. 매듭이 지정된 스플라인은 Stone 스플라인 또는 Stone-Koo 스플라인이라고도 합니다. 자세한 내용은 Stone and Koo 연구 자료(1985)에서 확인하십시오. 점이 100개 이하인 경우 첫 번째 매듭과 마지막 매듭은 각각 최소값과 최대값 내의 5번째 점입니다. 그렇지 않은 경우 첫 번째 매듭과 마지막 매듭은 각각 매듭이 5개 이하일 때는 0.05 분위수와 0.95 분위수에 배치되고, 매듭이 5개보다 많을 때는 0.025 분위수와 0.975 분위수에 배치됩니다. 점이 30개 이하인 경우를 제외하고 기본 매듭 수는 5개입니다. 30개 이하인 경우 기본값은 3개 매듭입니다.

매듭이 지정된 스플라인은 평활 스플라인과 대조적으로 다음과 같은 특성을 나타냅니다.

- 매듭이 지정된 스플라인은 항이 많은 일반 모형 내에서 사용되지만 평활 스플라인은 이변량 회귀에 사용됩니다.
- 회귀 기저가 반응의 함수가 아닙니다.
- 매듭이 지정된 스플라인은 절약성 원리에 따라 k 개 매듭 점의 곡률을 위해 $k-2$ 개의 항만 추가합니다.
- 매듭이 지정된 스플라인은 데이터 범위를 벗어나는 외삽이 다항식이 아니라 직선이라는 점에서 순수 다항식에 비해 보수적입니다.
- 간단한 방법으로 곡률을 검정할 수 있습니다.

매듭이 지정된 스플라인 효과를 사용한 곡률 검정의 예

곡률을 검정하려면 다음을 수행하십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Growth.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. **ratio** 열을 선택하고 **Y** 를 클릭합니다.
4. **age** 열을 선택하고 **추가**를 클릭합니다.
5. **효과** 패널에서 **age** 를 선택하고 **속성 > 매듭이 지정된 스플라인 효과**를 선택합니다.
6. 매듭 수에 5 를 입력하고 **확인**을 클릭합니다.

그림 4.21 모형 적합 시작 창

7. **실행**을 클릭합니다.
8. "반응 ratio"의 빨간색 삼각형을 클릭하고 **추정값 > 사용자 검정**을 선택합니다.
"사용자 검정" 보고서에 열이 하나만 있습니다. 이 예에서는 세 개의 열이 필요합니다.
9. **열 추가** 버튼을 두 번 클릭하여 총 세 개의 열을 생성합니다.
10. 첫 번째 열의 **age& 매듭 지정 @4.5**에 1을 입력합니다.
11. 두 번째 열의 **age& 매듭 지정 @20.25**에 1을 입력합니다.
12. 세 번째 열의 **age& 매듭 지정 @36**에 1을 입력합니다.

그림 4.22 곡률에 대한 사용자 검정의 값

13. **완료**를 클릭합니다.

"사용자 검정" 보고서의 "Prob > F" 값이 작으므로 곡률이 있음을 의미합니다.

그림 4.23 곡률 보고서

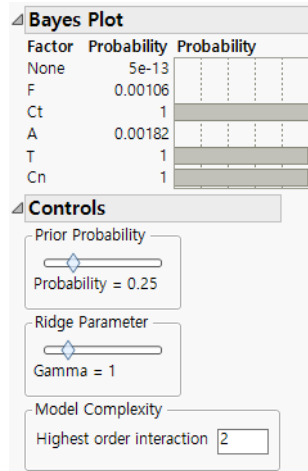
사용자 검정			
모수			
절편	0	0	0
age&대름 지정	0	0	0
age&대름 지정@4.5	1	0	0
age&대름 지정@20.25	0	1	0
age&대름 지정@36	0	0	1
=	0	0	0
값	-1.436874e-5	0.0000365148	-0.000033071
표준 오차	1.9803969e-6	6.0647971e-6	8.0723366e-6
t 비	-7.255483783	6.020782207	-4.096860562
Prob> t	5.275629e-10	8.1625874e-8	0.0001152566
SS	0.0583271224	0.0401646174	0.0185968833
제곱할	0.112524874		
분자 DF	3		
F 비	33.852401355		
Prob > F	1.955866e-13		

활성 요인에 대한 Bayes 그림 예

실험 설계를 수행하고 활성 요인을 판별하려는 경우를 가정해 보겠습니다. JMP 를 사용하여 여러 가지 방법으로 이를 처리할 수 있습니다. 이 예에서는 Bayes 방법으로 활성 요인을 식별하는데 사용할 수 있는 스크립트를 보여 줍니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Reactor.jmp** 를 엽니다.
2. **Samples > Scripts** 폴더에서 **BayesPlotforFactors.jsl** 샘플 스크립트를 엽니다.
3. **편집 > 스크립트 실행**을 선택합니다.
4. Y 를 선택하고 **Y, Response** 를 클릭합니다.
5. F, Ct, A, T 및 Cn 을 선택하고 **X, Factor** 를 클릭합니다.
6. **확인**을 클릭합니다.

그림 4.24 요인 활동에 대한 Bayes 그림



"Model Complexity"에서는 고려할 최고 차수 교호작용이 2임을 나타냅니다. 따라서 최대 2차 교호작용을 포함하는 가능한 모든 모형이 생성됩니다. "Prior Probability"에 할당된 값을 기반으로 가능한 각 모형의 사후 확률이 계산됩니다. 요인에 대한 확률은 해당 요인이 포함된 각 모형에 대한 확률의 합입니다.

이 방법으로 Ct, T 및 Cn이 활성 요인으로 식별되고 A와 F가 비활성 요인으로 식별됩니다.

능형 회귀 모수가 0(허용되지 않음)이면 모든 모형이 최소 제곱으로 적합됩니다. 능형 회귀 모수가 증가하면 모형의 모수 추정값은 0에 가까워지도록 축소됩니다. 능형 회귀 모수에 대한 자세한 내용 및 이 값이 0이 될 수 없는 이유는 Box and Meyer 연구 자료(1993)에서 확인하십시오.

단계별 회귀 모형 변수 선택을 사용하여 모형 찾기

단계별 회귀는 회귀 모형에 대한 효과의 부분집합을 선택하기 위한 방법입니다. 이 방법은 다음과 같은 상황에서 유용하게 사용됩니다.

- 모형에 대한 항 선택을 설명하는 이론이 거의 없는 경우
 - 적합도가 높은 예측 변수를 대화식으로 탐색하려는 경우
 - 불필요한 항을 추정함으로써 발생하는 분산을 줄여 모형의 예측 성능을 향상시키려는 경우
- 범주형 예측 변수에 대해 다음을 수행할 수 있습니다.
- 연관된 항이 모형에 추가되는 방식을 결정하기 위해 여러 규칙 중에서 선택합니다.
 - 효과 유전성을 적용합니다.

단계별 플랫폼을 사용하면 가능한 모든 모형을 탐색하고 모형 평균화를 수행할 수도 있습니다.

목차

단계별 회귀 개요	231
예 : 단계별 회귀 사용	231
단계별 보고서	233
단계별 플랫폼 옵션	233
단계별 회귀 제어 패널	234
현재 추정값 보고서	240
단계 기록 보고서	241
교차, 교호작용 또는 다항식 항이 있는 모형	242
결합 규칙의 예	242
범주형 및 순서형 효과가 있는 모형	244
계층 항 생성	244
명목형 항이 있는 모형의 예	245
계층 항에 대한 제한 규칙의 예	249
이항 및 순서형 로지스틱 단계별 회귀 수행	251
예: 로지스틱 단계별 회귀 사용	252
모든 가능 모형 옵션	253
예: 모든 가능 모형 옵션 사용	254
모형 평균화 옵션	255
예: 모형 평균화 옵션 사용	256
단계별 회귀의 검증 옵션	257
단계별 회귀에서 두 개 또는 세 개의 값이 있는 검증 데이터 집합	257
단계별 회귀의 K 폴드 교차 검증	260

단계별 회귀 개요

JMP에서 단계별 회귀는 모형 적합 플랫폼의 분석법입니다. **단계별** 기능은 다른 최소 제곱 플랫폼에서 제공하는 것과 동일한 추정값을 계산하지만 여러 모형 중에서 검색 및 선택을 쉽게 할 수 있습니다.

이 방식에는 주의해야 할 부작용이 있습니다. 선택된 모형은 고정 모형 내에서 검정된 것이 아니므로 모형에 대한 통계량의 유의 수준이 일반적인 통계적 가정을 위반합니다. 긍정적인 면도 있습니다. 이 방식은 30년 동안 항 수를 줄이는 데 유용하게 사용되고 있습니다. **Subset Selection in Regression**(Miller 1990) 설명서에서 모형 선택 통계량에 통계적 의미를 부여합니다.

이 장에서는 고정 모형에서 계산이 유효함을 나타내기 위해 유의 **확률**이라는 용어를 기계적으로 사용하며, 실제 유의 확률은 보고된 확률과 다르다는 것을 인식하고 있습니다.

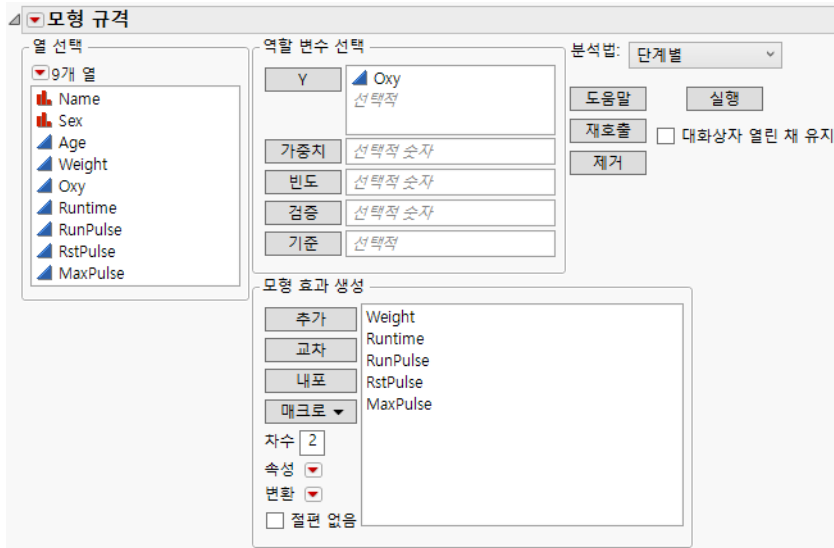
예 : 단계별 회귀 사용

Fitness.jmp 데이터 테이블에는 유산소 피트니스 연구 결과가 포함되어 있습니다. 미리 정해진 거리 동안 러닝머신에서 달리는 사람의 산소 섭취량을 측정하는 특수 검정을 사용하여 유산소 피트니스를 평가할 수 있습니다. 그러나 피트니스를 평가하고 산소 섭취량을 예측하는 간단한 측정을 사용한 계산식을 찾는 것이 더 경제적인 수 있습니다. 이러한 방정식을 식별하기 위해 1.5마일을 달린 31명의 참가자를 대상으로 연령, 체중, 달리기 시간 및 맥박을 측정했습니다.

참고 : 설명을 위해 MaxPulse 및 RunPulse의 특정 값은 Rawlings(1988, p.105)에 보고된 데이터에서 변경되었습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Fitness.jmp를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. Oxy를 선택하고 **Y**를 클릭합니다.
4. Weight, Runtime, RunPulse, RstPulse, MaxPulse를 선택하고 **추가**를 클릭합니다.
5. **분석법**에서 **단계별**을 선택합니다.

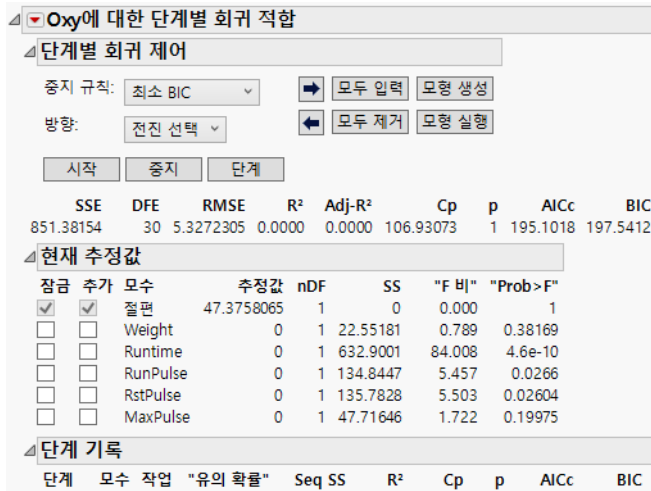
그림 5.1 완료된 모형 적합 시작 창



JMP Pro 검증은 JMP Pro에서만 사용할 수 있습니다.

6. 실행을 클릭합니다.

그림 5.2 단계별 보고서 창



산소 섭취량을 예측하는 적합한 방정식을 찾으려면 서로 다른 여러 회귀 모형을 비교해야 합니다. 단계별 보고서 창의 옵션을 사용하여 효과 조합이 포함된 모형을 검색하고 원하는 모형을 선택합니다.

단계별 보고서

단계별 보고서 창에는 다음과 같은 요소가 포함되어 있습니다.

플랫폼 옵션 "단계별 회귀 적합"의 빨간색 삼각형 메뉴에는 모든 변수에 영향을 주는 옵션이 포함되어 있습니다. 자세한 내용은 "[단계별 플랫폼 옵션](#)"(233 페이지)에서 확인하십시오.

단계별 회귀 제어 회귀변수 효과 확률 제한, 효과 선택 방법 결정, 선택 프로세스 시작 또는 중지, 모형 생성 등을 수행하는 컨트롤이 포함되어 있습니다. 자세한 내용은 "[단계별 회귀 제어 패널](#)"(234 페이지)에서 확인하십시오.

현재 추정값 모형 효과를 추가, 제거 및 잠글 수 있는 보고서입니다. 자세한 내용은 "[현재 추정값 보고서](#)"(240 페이지)에서 확인하십시오.

단계 기록 모형에 항을 추가한 효과를 기록하는 보고서입니다. 자세한 내용은 "[단계 기록 보고서](#)"(241 페이지)에서 확인하십시오.

단계별 플랫폼 옵션

"단계별 회귀 적합"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

K 폴드 교차 검증 (연속형 반응에만 사용 가능) 선택 프로세스에서 K 폴드 교차 검증을 수행합니다. 이 옵션을 선택하면 최대 K 폴드 R^2 중지 규칙을 사용할 수 있습니다 ("[단계별 회귀 제어 패널](#)"(234 페이지)). 검증에 대한 자세한 내용은 "[단계별 회귀의 검증 옵션](#)"(257 페이지)에서 확인하십시오.

모든 가능 모형 (연속형 반응에만 사용 가능) 지정된 한계까지의 가능한 모든 모형을 적합시키고 각 항 수에 대한 최적 모형을 표시합니다. 하나의 모형에 적합시킬 최대 항 수를 입력합니다. 또한 모형의 각 항 수에 대해 표시할 최적 모형 결과의 최대 수를 입력합니다. 범주형 변수는 표시 변수를 사용하여 나타냅니다. 자세한 내용은 "[범주형 및 순서형 효과가 있는 모형](#)"(244 페이지)에서 확인하십시오. 강한 효과 유전성을 충족하는 모형으로 표시되는 모형을 제한할 수 있습니다. 자세한 내용은 "[모든 가능 모형 옵션](#)"(253 페이지)에서 확인하십시오.

모형 평균화 (연속형 반응에만 사용 가능) 단일 최적 모형을 선택하는 대신 여러 모형의 적합을 평균화할 수 있습니다. 자세한 내용은 "[모형 평균화 옵션](#)"(255 페이지)에서 확인하십시오.

기준 기록 그림 AICc 와 BIC 대 모수의 수 그림을 생성합니다. "기준 기록" 그림에는 두 개의 음영 영역이 포함됩니다. 최소 AICc 값을 V^{best} 로 정의합니다. 그러면 녹색 영역은 $[V^{\text{best}}, V^{\text{best}}+4]$ 범위로 정의됩니다. 노란색 영역은 $(V^{\text{best}}+4, V^{\text{best}}+10]$ 범위로 정의됩니다.

R^2 기록 그림 (연속형 반응에만 사용 가능) 훈련 및 검증 R^2 대 모수의 수 그림을 생성합니다.

기록 지우기 단계 기록을 지우고 재설정합니다.

JMP PRO 검증과 함께 모형 내보내기 (단계별 시작 창에서 검증 열을 입력한 경우에만 사용 가능) "모형 생성"을 선택하면 "모형 규격"창에 검증 열을 추가합니다. "모형 실행"을 선택하면 검증 열을 사용하여 모형을 실행합니다.

이 옵션은 기본적으로 선택되어 있습니다.

모형 대화상자 현재 모형에 대해 완료된 "모형 적합" 시작 창을 표시합니다.

단계별 회귀 제어 패널

"단계별 회귀 제어" 패널을 사용하여 회귀변수 효과 확률을 제한하고, 효과 선택 방법을 결정하고, 선택 프로세스를 시작 또는 중지하고, 모형을 실행할 수 있습니다. 제외되거나 결측된 행이 있는지 나타내는 노트가 "시작" 버튼 아래에 표시됩니다.

그림 5.3 단계별 회귀 제어 패널

SSE	DFE	RMSE	R ²	Adj-R ²	Cp	p	AICc	BIC
6940	31	14.962318	0.0000	0.0000	647.07317	1	267.3641	269.8818

중지 규칙

중지 규칙은 선택되는 모형을 결정합니다. "p 값 임계"를 제외한 모든 중지 규칙에 대해 "전진 선택" 및 "후진 제거" 방향만 허용됩니다. 검증을 사용하는 중지 규칙은 "최대 검증 R²" 및 "최대 K 폴드 R²" 뿐입니다. 자세한 내용은 "단계별 회귀의 검증 옵션"(257 페이지)에서 확인하십시오.

p 값 임계 p 값(유의 수준)을 사용하여 모형에서 효과를 입력하고 제거합니다. "p 값 임계"를 선택하면 다음과 같은 두 가지 옵션이 나타납니다.

선택 유의확률 전진 선택 단계 중에 모형에 효과를 입력해야 하는 최대 p 값을 지정합니다.

제거 유의확률 후진 제거 단계 중에 모형에서 효과를 제거해야 하는 최소 p 값을 지정합니다.

최소 AICc 수정된 Akaike 정보 기준 최소값을 사용하여 최적 모형을 선택합니다. 자세한 내용은 "통계 상세 정보" 부록의 "가능도, AICc 및 BIC"(514 페이지)에서 확인하십시오.

최소 BIC Bayes 정보 기준 최소값을 사용하여 최적 모형을 선택합니다. 자세한 내용은 "통계 상세 정보" 부록의 "가능도, AICc 및 BIC"(514 페이지)에서 확인하십시오.

JMP Pro 최대 검증 R² 검증 데이터 집합의 최대 R²을 사용하여 최적 모형을 선택합니다. 이 옵션은 두 개 또는 세 개의 구분되는 값을 포함하는 검증 열을 사용하는 경우에만 선택할 수 있습니다. 검증에 대한 자세한 내용은 "단계별 회귀에서 두 개 또는 세 개의 값이 있는 검증 데이터 집합"(257 페이지)에서 확인하십시오.

최대 K 폴드 R² K 폴드 교차 검증의 최대 R²을 사용하여 최적 모형을 선택합니다. 단계별 회귀의 빨간색 삼각형 메뉴에서 "K 폴드 교차 검증" 옵션을 선택하여 "최대 K 폴드 R²" 중지 규칙에 액세스할 수 있습니다. JMP Pro 사용자는 4 개 이상의 값이 있는 검증 데이터 집합을 사용

하여 옵션에 액세스할 수 있습니다. 이 옵션을 선택하면 폴드 수를 지정해야 합니다. 검증에 대한 자세한 내용은 "[단계별 회귀의 K 폴드 교차 검증](#)"(260 페이지)에서 확인하십시오.

방향

선택한 방향은 모형에서 효과가 입력 및 제거되는 방식을 제어합니다. 다음 옵션 중 하나를 선택합니다.

전진 선택 p 값이 가장 작은 항을 입력합니다. " p 값 임계" 중지 규칙을 선택한 경우 해당 항은 **선택 유의확률**에 지정된 수준에서 유의해야 합니다. 자세한 내용은 "[전진 선택 예](#)"(238 페이지)에서 확인하십시오.

후진 제거 p 값이 가장 큰 항을 제거합니다. " p 값 임계" 중지 규칙을 선택한 경우 해당 항은 **제거 유의확률**에 지정된 수준에서 유의하지 않아야 합니다. 자세한 내용은 "[후진 제거 예](#)"(239 페이지)에서 확인하십시오.

참고: "방향"에서 "후진 제거"를 선택한 경우 "시작" 또는 "단계"를 클릭하기 전에 "모두 입력"을 클릭해야 합니다.

혼합 P 값 중지 규칙을 선택한 경우에만 사용할 수 있습니다. 이 옵션은 전진 선택 단계와 후진 제거 단계를 교대로 실행합니다. **선택 유의확률**을 충족하는 가장 유의한 항을 포함하고 **제거 유의확률**을 충족하는 가장 유의하지 않은 항을 제거합니다. 나머지 항이 유의할 때까지 항을 계속 제거한 후 전진 선택 방향으로 변경됩니다.

시작, 중지, 단계 버튼

"시작", "중지" 및 "단계" 버튼을 사용하면 모형에서 항이 입력 또는 제거되는 방식을 제어할 수 있습니다.

참고: 모든 중지 규칙은 p 값에 따른 입력 (전진 선택 방향) 또는 제거 (후진 제거 방향)로 정의된 모형만 고려합니다. 중지 규칙은 모든 가능 모형을 고려하지 않습니다.

시작 항 입력 (전진 선택 방향) 또는 제거 (후진 제거 방향)를 자동화합니다. 적합 모형 중 선택한 중지 규칙에 따라 최적으로 간주되는 모형이 마지막에 나열됩니다. " p 값 임계" 중지 규칙을 제외하고 중지 규칙 통계량 동작에서 국소적 감소를 무시하는 모형이 최적 모형으로 선택됩니다. 최적 모형 오른쪽에 있는 버튼을 클릭하면 "모형 생성" 및 "모형 실행" 옵션에 사용할 모형이 선택되지만 선택을 변경해도 됩니다.

- " p 값 임계"의 경우 최적 모형은 "선택 유의확률" 및 "제거 유의확률" 기준을 기반으로 합니다. 자세한 내용은 "[p 값 임계](#)"(234 페이지)에서 확인하십시오.
- "최소 AICc" 및 "최소 BIC"의 경우 최적 모형을 찾을 때까지 자동 적합이 계속됩니다. 최적 모형은 최소 AICc 또는 BIC 값을 사용한 모형이며, 각각 AICc 또는 BIC 값이 더 큰 모형이 10 개까지 있을 수 있습니다. 이 모형은 "모수" 열의 "최선" 항목과 "작업" 열의 "특정" 항목으로 지정됩니다.

- "최대 검증 R^2 "(JMP Pro 만 해당) 및 "최대 K 폴드 R^2 "의 경우 최적 모형을 찾을 때까지 자동 적합이 계속됩니다. 최적 모형은 검증 R^2 또는 R^2 K 폴드 값을 사용한 모형이며, 각각 검증 R^2 또는 R^2 K 폴드 값이 더 작은 모형이 10 개까지 있을 수 있습니다. 이 모형은 "모수" 열의 "최선" 행과 "작업" 열의 "특정" 행으로 지정됩니다.

중지 "시작" 버튼을 사용하여 시작된 자동 선택 프로세스를 중지합니다.

단계 항을 하나씩 입력하거나 (전진 선택 방향) 하나씩 제거합니다 (후진 제거 방향). 언제든지 "단계 기록" 보고서의 오른쪽에 있는 버튼을 클릭하여 모형을 선택할 수 있습니다. 선택한 모형 항은 "현재 추정값" 보고서에서 업데이트됩니다. "모형 생성" 또는 "모형 실행"을 클릭하면 이 모형이 사용됩니다.

규칙

참고: 모형에 관련 항이 포함된 경우에만 나타납니다. 명목형 또는 순서형 변수가 있으면 관련 항이 생성되고 "현재 추정값" 테이블에 나타납니다.

모형에 항 계층이 있을 때 적용되는 규칙을 변경하려면 "규칙"을 사용합니다. 계층은 다음과 같은 방식으로 나타낼 수 있습니다.

- 변수가 다른 변수의 성분이면 계층이 생성됩니다. 예를 들어 모형에 A, B 및 $A*B$ 변수가 포함된 경우 계층에서 A 와 B 는 $A*B$ 에 대한 상위 항입니다.
- 명목형 또는 순서형 변수를 포함하는 경우에도 계층이 생성됩니다. 트리 구조에서 다른 항 위에 있는 항이 상위 항입니다. 자세한 내용은 "[계층 항 생성](#)"(244페이지)에서 확인하십시오.

다음 옵션 중 하나를 선택합니다.

결합 상위 항목이 있는 항을 입력할 때 두 개의 개별 검정에 대한 p 값을 계산합니다. 첫 번째 p 값, p_1 은 항을 상위 항과 그룹화한 후 입력에 대한 해당 그룹의 유의 확률을 결합 F-검정으로 검정하여 계산됩니다. 두 번째 p 값, p_2 는 상위 항이 이미 모형에 추가된 후 해당 항의 입력에 대한 유의 확률을 검정한 결과입니다. 상위 항목이 있는 항의 입력에 대한 최종 유의 확률은 $\max(p_1, p_2)$ 입니다.

팁: 결합 규칙을 사용하면 상위 항에 특히 강한 효과가 있을 수 있는 유의하지 않은 교호작용 항이 포함되지 않습니다. 이 시나리오에서는 강한 주효과로 인해 입력에 대한 그룹의 유의 확률 p_1 이 매우 작아질 수 있습니다. 그러나 두 번째 검정에서 교호작용 자체는 유의하지 않습니다. 결과적으로 p_2 가 더 크고 입력에 대한 최종 유의 확률로 사용됩니다.

주의: 상위 항목이 있는 항의 자유도 값은 입력에 대한 두 가지 유의 확률 중 어느 것이 더 큰 지에 따라 다릅니다. 입력에 대한 최종 유의 확률에 사용되는 검정은 "현재 추정값" 테이블의 자유도 nDF 를 결정합니다. 따라서 p_1 이 사용되면 nDF 는 결합 검정을 위한 그룹의 항 수와 같고, p_2 가 사용되면 nDF 는 1 입니다.

결합 옵션은 기본 규칙입니다. 자세한 내용은 "[교차, 교호작용 또는 다항식 항이 있는 모형](#)"(242페이지)에서 확인하십시오.

제한 상위 항목이 추가될 때까지 상위 항목을 포함한 항을 추가할 수 없도록 제한합니다. 자세한 내용은 "범주형 및 순서형 효과가 있는 모형"(244 페이지) 및 "계층 항에 대한 제한 규칙의 예"(249 페이지)에서 확인하십시오.

규칙 없음 선택 루틴으로 인해 계층이 깨지는지 여부에 관계없이 루틴에서 완전히 자유롭게 항을 선택할 수 있습니다.

전체 효과 해당 효과가 포함된 항이 유의한 경우 전체 효과만 입력합니다. 이 규칙은 3 개 이상의 수준이 있는 범주형 변수가 가능한 모형 효과로 입력된 경우에만 적용됩니다. 자세한 내용은 "규칙"(249 페이지)에서 확인하십시오.

버튼

"단계별 회귀 제어" 패널에는 다음 버튼이 포함되어 있습니다.

시작 완료될 때까지 선택 프로세스를 자동화합니다.

중지 선택 프로세스를 중지합니다.

단계 선택 프로세스를 한 번에 한 단계씩 증가시킵니다.

화살표 버튼   선택 프로세스에서 한 단계씩 앞뒤로 이동합니다.

모두 입력 잠기지 않은 모든 항을 모형에 입력합니다.

모두 제거 잠기지 않은 모든 항을 모형에서 제거합니다.

모형 생성 "현재 추정값" 테이블에 표시된 모형에서 "모형 적합" 창에 대한 모형을 생성합니다. 명목형 또는 순서형 항이 있는 경우 **모형 생성**은 모형에 필요한 항을 포함하는 임시 변환 열을 생성합니다.

모형 실행 "현재 추정값" 테이블에 표시된 모형을 실행합니다. 명목형 또는 순서형 항이 있는 경우 **모형 실행**은 모형에 필요한 항을 포함하는 임시 변환 열을 생성합니다.

통계량

"단계별 회귀 제어" 패널 아래에 다음 통계량이 표시됩니다.

SSE 현재 모형에 대한 오차 제곱합입니다.

DFE 현재 모형에 대한 오차 자유도입니다.

RMSE 현재 모형에 대한 제곱근 평균 제곱 오차 (잔차)입니다.

R^2 랜덤 오차가 아니라 모형의 항에 기인할 수 있는 반응의 변동 비율입니다.

Adj- R^2 모수의 개수가 서로 다른 모형들을 비교하기 좋도록 계산에 자유도를 사용하여 R^2 값을 조정합니다. 여러 모형을 검토 중이고 모형의 항 수에 맞게 조정해야 하므로 조정 R^2 은 단계별 절차에서 유용합니다.

Cp 모형 선택을 위한 Mallows C_p 기준입니다. 이 값은 총 제곱 오차의 대체 측도이며 다음과 같이 정의할 수 있습니다.

$$C_p = \left(\frac{SSE_p}{s^2} \right) - (N - 2p)$$

여기서 s^2 은 전체 모형에 대한 MSE 이고, SSE_p 는 절편을 포함하여 p 개 변수가 있는 모형에 대한 오차 제곱합입니다. p 는 x 변수의 수 +1 입니다. C_p 가 p 로 그래프 표시된 경우 Mallows(1973)에서는 C_p 가 p 에 처음 접근하는 모형을 선택할 것을 권장합니다.

p 절편을 포함한 모형의 모수 수입니다.

AICc 수정된 Akaike 정보 기준입니다. 자세한 내용은 "통계 상세 정보" 부록의 "[가능도, AICc 및 BIC](#)"(514 페이지)에서 확인하십시오.

BIC Bayes 정보 기준 자세한 내용은 "통계 상세 정보" 부록의 "[가능도, AICc 및 BIC](#)"(514 페이지)에서 확인하십시오.

전진 선택 예

전진 선택에서는 모형에 항이 입력되고 모든 항이 유의할 때까지 가장 유의한 항이 추가됩니다.

1. "[예: 단계별 회귀 사용](#)"(231 페이지)의 단계를 완료합니다.

방향은 기본적으로 "전진 선택"으로 되어 있습니다.

2. 단계를 클릭합니다.

그림 5.4에서는 한 단계 후 가장 유의한 항인 Runtime 이 모형에 입력되었음을 확인할 수 있습니다.

3. 시작을 클릭합니다.

그림 5.5에서는 RstPulse 와 Weight 를 제외한 모든 항이 추가되었음을 확인할 수 있습니다.

그림 5.4 전진 선택의 한 단계 후 현재 추정값 테이블

현재 추정값							
잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☒	☒	절편	82.4217727	1	0	0.000	1
☐	☐	Weight	0	1	1.323628	0.171	0.68267
☐	☒	Runtime	-3.3105554	1	632.9001	84.008	4.6e-10
☐	☐	RunPulse	0	1	15.36208	2.118	0.15673
☐	☐	RstPulse	0	1	0.130138	0.017	0.89814
☐	☐	MaxPulse	0	1	1.567361	0.202	0.65632

그림 5.5 전진 선택의 세 단계 후 현재 추정값 테이블

현재 추정값							
잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☒	☒	절편	80.9007896	1	0	0.000	1
☐	☐	Weight	0	1	4.989591	0.827	0.37137
☐	☒	Runtime	-2.9701867	1	443.2028	73.971	3.25e-9
☐	☒	RunPulse	-0.3751142	1	55.14175	9.203	0.00529
☐	☐	RstPulse	0	1	0.350744	0.056	0.81399
☐	☒	MaxPulse	0.35421891	1	41.34703	6.901	0.01403

후진 제거 예

후진 제거에서는 모든 항이 모형에 입력된 후 나머지 항이 모두 유의할 때까지 가장 유의하지 않은 항이 제거됩니다.

1. "예: 단계별 회귀 사용"(231 페이지)의 단계를 완료합니다.
2. 모두 입력을 클릭합니다.

그림 5.6 모형에 입력된 모든 효과

현재 추정값							
잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☒	☒	절편	82.3936054	1	0	0.000	1
☐	☒	Weight	-0.0509071	1	4.83788	0.772	0.38784
☐	☒	Runtime	-2.9518165	1	366.3375	58.489	5.29e-8
☐	☒	RunPulse	-0.3970425	1	59.51519	9.502	0.00495
☐	☒	RstPulse	0.01239004	1	0.199033	0.032	0.85995
☐	☒	MaxPulse	0.38479281	1	45.83023	7.317	0.01212

3. 방향에서 "후진 제거"를 선택합니다.
4. 단계를 두 번 클릭합니다.

첫 번째 후진 제거 단계에서 RstPulse 가 제거되고 두 번째 후진 제거 단계에서 Weight 가 제거됩니다.

그림 5.7 항이 제거된 현재 추정값 및 단계 기록 테이블

현재 추정값							
잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☒	☒	절편	80.9007896	1	0	0.000	1
☐	☐	Weight	0	1	4.989591	0.827	0.37137
☐	☒	Runtime	-2.9701867	1	443.2028	73.971	3.25e-9
☐	☒	RunPulse	-0.3751142	1	55.14175	9.203	0.00529
☐	☐	RstPulse	0	1	0.350744	0.056	0.81399
☐	☒	MaxPulse	0.35421891	1	41.34703	6.901	0.01403

단계 기록									
단계	모수	작업	"유의 확률"	Seq SS	R ²	Cp	p	AICc	BIC
1	모두	추가	.	.	0.8161	6	6	157.051	162.22 ○
2	RstPulse	제거	0.8600	0.199033	0.8158	4.0318	5	153.721	158.825 ○
3	Weight	제거	0.3714	4.989591	0.8100	2.8284	4	151.592	156.362 ●

그림 5.7에 표시된 "현재 추정값" 및 "단계 기록" 테이블에는 후진 제거 단계별 선택 프로세스가 요약되어 있습니다. "단계 기록" 테이블에서 세 번째 단계의 BIC 값이 156.362입니다. "단계"를 다시 클릭하여 모형에서 다른 모수를 제거하면 BIC 값이 159.984로 증가합니다. 이러한 이유로 단계 3 모형을 선택합니다. 이 모형은 "시작" 버튼으로 생성하는 모형과 동일합니다.

현재 추정값 보고서

"현재 추정값" 보고서를 사용하여 모형 효과를 입력, 제거 및 잠글 수 있습니다. 절편은 모형에 영구적으로 잠겨 있습니다. 이 보고서에는 다음 열이 포함됩니다.

그림 5.8 현재 추정값 테이블

SSE	DFE	RMSE	R ²	Adj-R ²	Cp	p	AICc	BIC
851.38154	30	5.3272305	0.0000	0.0000	106.93073	1	195.1018	197.5412
△ 현재 추정값								
잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"	
☒	☒	절편	47.3758065	1	0	0.000	1	
☐	☐	Weight	0	1	22.55181	0.789	0.38169	
☐	☐	Runtime	0	1	632.9001	84.008	4.6e-10	
☐	☐	RunPulse	0	1	134.8447	5.457	0.0266	
☐	☐	RstPulse	0	1	135.7828	5.503	0.02604	
☐	☐	MaxPulse	0	1	47.71646	1.722	0.19975	

잠금 모형에서 항을 잠그거나 잠그지 않습니다. 선택된 항은 모형에서 입력하거나 제거할 수 없습니다.

추가 항이 현재 모형에 있는지 여부를 나타냅니다. 항의 체크박스를 클릭하여 효과를 수동으로 모형에 가져오거나 모형에서 내보낼 수 있습니다.

모수 효과 이름입니다.

추정값 현재 모수 추정값이며, 효과가 현재 모형에 없으면 값이 0입니다.

nDF 항의 자유도 수입니다. 항이 모형에 입력되면 다른 항도 모형에 추가되는 경우 항의 자유도는 두 개 이상입니다.

Wald/ 점수 카이제곱 (반응이 범주형인 경우에만 표시됨) 모형의 각 항에 대한 검정 통계량입니다. 모형에 아직 포함되지 않은 항에 대한 검정 통계량은 모형에 항을 포함하는 스코어 검정을 기반으로 합니다. 모형에 이미 포함된 항에 대한 검정 통계량은 모형에서 항을 제거하는 Wald 검정을 기반으로 합니다.

"유의 확률" (반응이 범주형인 경우에만 표시됨) nDF 자유도에 기반한 Wald/ 점수 카이제곱 검정 통계량과 관련된 p 값입니다. "유의 확률"은 모형에 포함할 다음 항을 결정하는 데 사용됩니다.

SS (반응이 연속형인 경우에만 표시됨) 항이 모형에 입력될 경우 오차 (잔차) SS(제곱합)의 감소 또는 항이 모형에서 제거될 경우 오차 SS의 증가입니다. 항이 특정 방식으로 제한되면 SS가 0으로 보고될 수 있습니다.

F 비 (반응이 연속형인 경우에만 표시됨) 항 효과가 0임을 검정하기 위한 기존의 검정 통계량입니다. 이 값은 t 비의 제곱입니다. 모형이 적합한 것으로 선택되었지만 항을 검정하기 위한 F 분포가 없으므로 인용 부호로 묶여 있습니다.

"Prob>F" (반응이 연속형인 경우에만 표시됨) F 통계량과 관련된 p 값입니다. 실제 유의 확률로 신뢰할 수 없으므로 "F 비"와 마찬가지로 인용 부호로 묶여 있습니다.

R 모형의 다른 효과와의 다중 상관입니다. 이 열은 보고서를 마우스 오른쪽 버튼으로 클릭하고 열 > R을 선택한 경우에만 나타납니다.

단계 기록 보고서

각 단계를 수행할 때 모형에 항을 추가한 효과가 "단계 기록" 보고서에 기록됩니다. 예를 들어 Fitness JMP 예의 "단계 기록" 보고서에는 모형에 항이 입력된 순서 및 각 모형에 대한 통계량이 표시됩니다. 자세한 내용은 "[예: 단계별 회귀 사용](#)"(231 페이지)에서 확인하십시오.

오른쪽의 라디오 버튼을 사용하여 모형을 선택합니다.

그림 5.9 단계 기록 보고서

단계 기록									
단계	모수	작업	"유의 확률"	Seq SS	R ²	Cp	p	AICc	BIC
1	Runtime	추가	0.0000	632.9001	0.7434	7.8825	2	155.397	158.81 ○
2	RunPulse	추가	0.1567	15.36208	0.7614	7.4298	3	155.787	159.984 ○
3	MaxPulse	추가	0.0140	41.34703	0.8100	2.8284	4	151.592	156.362 ●

범주형 반응에 대한 단계 기록

범주형 반응이 있는 모형의 "단계 기록" 보고서에는 L-R 카이제곱, "유의 확률", 카이제곱 입력 및 "유의 확률" 입력의 네 가지 추가 열이 포함됩니다.

L-R 카이제곱 및 "유의 확률" 열에는 전체 모형 대 축소 모형의 가능도비 검정 통계량과 p 값이 포함됩니다. 여기서 전체 모형은 지정된 항을 포함하는 모형이고 축소 모형은 지정된 항을 포함하지 않습니다.

카이제곱 입력 및 "유의 확률" 입력 열에는 모형에 포함할 가장 최근 항을 선택하는 데 사용된 Wald/점수 카이제곱 및 "유의 확률" 값이 포함됩니다.

그림 5.10 범주형 반응에 대한 단계 기록 (첫 번째 단계 전)

현재 추정값							
				Wald/			
잠금	추가	모수	추정값	nDF	점수 카이제곱	"유의 확률"	
☒	☒	절편[M]	-0.0645385	1	0	1	
☐	☐	Oxy	0	1	7.275453	0.00699	
☐	☐	Weight	0	1	10.13425	0.00146	
☐	☐	Runtime	0	1	5.764797	0.01635	
☐	☐	RunPulse	0	1	1.865024	0.17205	
☐	☐	RstPulse	0	1	2.693096	0.10078	
☐	☐	MaxPulse	0	1	1.467541	0.22573	

단계 기록										
단계	모수	작업	L-R 카이제곱	"유의 확률"	카이제곱 입력	"유의 확률" 입력	R ²	p	AICc	BIC

그림 5.11 범주형 반응에 대한 단계 기록 (첫 번째 단계 후)

현재 추정값									
				Wald/					
잠금	추가	모수	추정값	nDF	점수	카이제곱	"유의 확률"		
☒	☒	절편[M]	-16.228294	1	0	0	1		
☐	☐	Oxy	0	1	8.750599	0.0031			
☐	☒	Weight	0.20792272	1	6.994297	0.00818			
☐	☐	Runtime	0	1	6.371395	0.0116			
☐	☐	RunPulse	0	1	1.428737	0.23197			
☐	☐	RstPulse	0	1	3.510878	0.06097			
☐	☐	MaxPulse	0	1	0.365726	0.54534			
단계 기록									
단계	모수	작업	L-R	카이제곱	"유의 확률"	카이제곱 입력	"유의 확률" 입력	R ²	p
1	Weight	추가	12.16669	0.0005	10.1342	0.00146	0.2833	2	35.2047
									37.6441

그림 5.11의 "단계 기록" 테이블에서 Weight에 대한 카이제곱 입력 및 "유의 확률" 입력 값은 그림 5.10의 Weight에 대한 Wald/점수 카이제곱 및 "유의 확률" 값과 일치합니다.

교차, 교호작용 또는 다항식 항이 있는 모형

일부 모형, 특히 실험 설계와 관련된 모형은 교호작용 항을 포함합니다. 연속형 요인의 경우 교호작용은 효과를 나타내는 열의 곱입니다. 명목형 및 순서형 요인의 경우 교호작용은 범주형 수준을 나타내는 항의 곱이 포함된 모형 항으로 정의됩니다.

교호작용 항이 있으면 교차 효과의 저차 성분이 모형에 포함되도록 모형 선택 프로세스에 제한을 적용하려는 경우가 종종 있습니다. 이 방법은 효과 유전성 원리에 의해 제안됩니다. 자세한 내용은 실험 설계 가이드의 에서 확인하십시오. 예를 들어 이원 교호작용이 모형에 포함된 경우 성분 주효과(상위)도 포함되어야 합니다.

결합 규칙의 예

1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Reactor.jmp 를 엽니다.
2. 분석 > 모형 적합을 선택합니다.
3. Y 를 선택하고 Y 를 클릭합니다.
4. "차수" 상자에 2 를 입력합니다.
5. F, Ct, A, T 및 Cn 을 선택하고 매크로 > 특정 차수까지의 요인을 클릭합니다.
6. 분석법에서 단계별을 선택합니다.
7. 실행을 클릭합니다.

그림 5.12 결합 규칙을 사용한 초기 현재 추정값 보고서

현재 추정값							
잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☒	☒	절편	65.5	1	0	0.000	1
☐	☐	F	0	1	15.125	0.066	0.79972
☐	☐	Ct	0	1	3042	23.412	3.67e-5
☐	☐	A	0	1	3.125	0.014	0.90823
☐	☐	T	0	1	924.5	4.611	0.03998
☐	☐	Cn	0	1	312.5	1.415	0.24363
☐	☐	F*Ct	0	1	15.125	0.066	0.79972
☐	☐	F*A	0	3	22.75	0.031	0.9926
☐	☐	F*T	0	1	6.125	0.027	0.87178
☐	☐	F*Cn	0	1	0.125	0.001	0.98161
☐	☐	Ct*A	0	1	6.125	0.027	0.87178
☐	☐	Ct*T	0	1	1404.5	7.612	0.00979
☐	☐	Ct*Cn	0	1	32	0.139	0.71193
☐	☐	A*T	0	1	36.125	0.157	0.69476
☐	☐	A*Cn	0	1	6.125	0.027	0.87178
☐	☐	T*Cn	0	1	968	4.863	0.03525

그림 5.12의 모형에는 5개의 연속형 요인에 대한 최대 2요인 교호작용의 모든 항이 포함되어 있습니다. "규칙"(236페이지)에 설명된 결합, 제한 및 전체 효과 규칙을 사용하여 교호작용 항 입력력을 제어할 수 있습니다.

결합 규칙은 두 가지 검정을 기반으로 교호작용 항의 입력을 결정합니다. 자세한 내용은 "결합"(236페이지)에서 확인하십시오. 자유도(nDF)를 기반으로 두 검정 중 p 값에 사용된 검정을 확인할 수 있습니다. 예를 들어 교호작용 항 $F*A$ 의 nDF 값은 3입니다. 이는 $F*A$ 가 상위 항 F 및 A 로 그룹화되고, 3 자유도 결합 F -검정을 기반으로 하여 입력 대상으로 고려된다는 의미입니다. 반면에 교호작용 항 $Ct*T$ 의 nDF 값은 1입니다. 이는 상위 항 Ct 와 T 가 이미 모형에 포함된 후 $Ct*T$ 의 유의성을 검정하는 1 자유도 F -검정을 기반으로 하여 $Ct*T$ 가 입력 대상으로 고려된다는 의미입니다. 단계를 한 번 클릭하여 Ct 가 단독으로 입력되는지 확인합니다.

8. 단계를 다시 클릭하여 $Ct*T$ 가 T 와 함께 입력되는지 확인합니다 (Ct 는 이미 모형에 있음).

그림 5.13 결합 규칙을 사용한 현재 추정값 보고서 (한 단계)

현재 추정값							
잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☒	☒	절편	65.5	1	0	0.000	1
☐	☐	F	0	1	15.125	0.263	0.61236
☐	☒	Ct	9.75	2	4446.5	39.676	6.74e-9
☐	☐	A	0	1	3.125	0.054	0.81819
☐	☒	T	5.375	2	2329	20.781	2.93e-6
☐	☐	Cn	0	1	312.5	6.715	0.01524
☐	☐	F*Ct	0	2	30.25	0.256	0.7764
☐	☐	F*A	0	3	22.75	0.123	0.9459
☐	☐	F*T	0	2	21.25	0.178	0.83755
☐	☐	F*Cn	0	1	0.125	0.002	0.96335
☐	☐	Ct*A	0	2	9.25	0.077	0.92601
☐	☒	Ct*T	6.625	1	1404.5	25.064	2.72e-5
☐	☐	Ct*Cn	0	1	32	0.562	0.45989
☐	☐	A*T	0	2	39.25	0.334	0.7194
☐	☐	A*Cn	0	1	6.125	0.106	0.74747
☐	☐	T*Cn	0	1	968	43.488	4.49e-7

유의한 교호작용 항이 있는 경우 같은 단계에서 여러 항을 입력할 수 있습니다. 단계 버튼을 두 번 클릭하면 $Ct*T$ 는 포함된 두 효과 Ct 및 T 와 함께 입력됩니다. 그러나 두 성분 항을 제거하지 않으

면 교차 항을 제거할 수 없으므로 한 단계 뒤로는 대칭이 아닙니다. 이제 C_t 와 T 의 자유도는 각각 2입니다. 이는 단계별 방법에서 C_t 또는 T 를 제거하는 경우 $C_t * T$ 도 제거해야 하기 때문입니다. "방향"을 **후진 제거**로 변경하고 **단계**를 클릭하면 $C_t * T$ 가 제거되고 C_t 및 T 의 자유도가 1로 변경됩니다.

범주형 및 순서형 효과가 있는 모형

일반적으로 단계별 회귀에서는 모형에 범주형 효과가 있는 상황을 다루지 않았습니다. 다음 사항에 유의하십시오.

- 회귀 모형에 명목형 또는 순서형 효과가 포함된 경우 해당 효과는 표시자 열 집합으로 나타냅니다.
- 범주형 효과에 수준이 두 개뿐인 경우 해당 효과는 단일 열로 나타냅니다.
- 범주형 효과에 k 개 수준이 있는 경우 ($k > 2$) 이 효과는 $k-1$ 개 열로 나타내야 합니다.

JMP에서 사용되는 표준 플랫폼에 대한 규칙은 모수 추정값 평균이 전체 수준에서 0인 항으로 명목형 변수를 나타내는 것입니다.

단계별 플랫폼에서 범주형 변수(명목형 및 순서형)는 계층적 방식으로 코드화됩니다. 이것은 다른 최소 제곱 적합 플랫폼의 코딩과 다릅니다. 계층적 코딩에서 범주형 변수의 수준은 반응의 평균을 가장 많이 분리하는 수준 그룹으로 연속하여 분할됩니다. 분할을 통해 $k-1$ 개 항으로 k 수준 범주형 변수를 나타낼 수 있습니다.

참고: 계층적 코딩에서 생성된 초기 항은 가장 큰 분리를 담당하는 그룹을 나타냅니다. 이 코딩 체계의 장점은 이러한 정보화 항이 모형에 일찍 입력될 가능성이 있다는 것입니다.

계층 항 생성

계층 항은 파티션 분석과 유사한 트리 구조를 사용하여 생성됩니다. 그러나 최대화되는 기준은 그룹 간 제곱합(SSB)입니다.

수준이 k 개인 명목형 변수의 경우 k 개 수준은 SSB가 가장 큰 두 개의 수준 그룹으로 분할됩니다. 이러한 두 수준 그룹을 A1과 A2라고 합니다. 여기서 A1의 평균이 더 작고 A2의 평균이 더 큽니다. A1과 A2의 두 수준 그룹은 A1의 수준 값이 1이고 A2의 수준 값이 -1인 표시 변수를 정의하는 데 사용됩니다. 이 변수는 명목형 변수에 대한 첫 번째 계층 항입니다.

초기 두 그룹 A1과 A2 내의 각 수준에 대해 SSB가 가장 큰 두 개의 수준 그룹으로 분할됩니다. SSB가 가장 큰 수준 그룹이 A1의 수준에 속한다고 가정해 보겠습니다. 이러한 두 그룹을 B1과 B2라고 합니다. 여기서 B1의 평균이 더 작고 B2의 평균이 더 큽니다. B1과 B2의 두 수준 그룹은 B1의 수준 값은 1, B2의 수준 값은 -1, A2의 수준 값은 0인 계층 변수를 정의하는 데 사용됩니다. 다음 변수를 생성하려면 B1, B2 및 A2의 수준 분할을 고려해야 합니다. SSB를 최대화하는 분할이 다음 계층 변수를 정의합니다. 이 프로세스는 $k-1$ 개 계층 항이 생성될 때까지 계속됩니다.

순서형 변수의 경우 분할에서 고려되는 수준 그룹에는 순서가 인접한 수준만 포함됩니다. 이렇게 하면 생성된 항이 수준 순서를 준수합니다.

규칙 및 계층 항

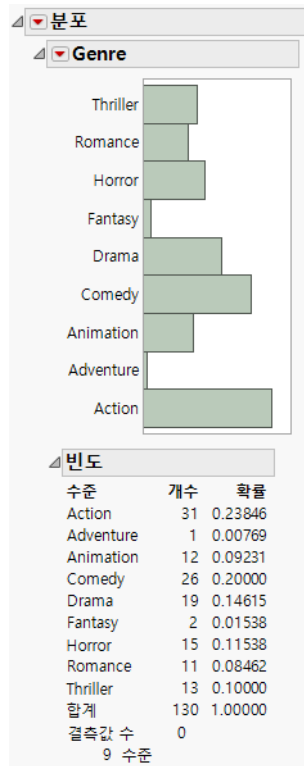
결합 규칙 또는 **제한** 규칙을 사용하는 경우 계층에서 위에 있는 모든 항이 입력되지 않으면 해당 항을 모형에 입력할 수 없습니다. **전체 효과** 규칙을 사용하고 범주형 변수에 대한 항을 입력하면 관련된 모든 항이 입력됩니다. 예는 "[예제의 계층 항 생성](#)"(248 페이지)에서 확인하십시오.

명목형 항이 있는 모형의 예

이 예에서는 2011 년에 개봉한 영화 관련 데이터를 사용합니다 . 특히 총 수입액을 나타내는 World Gross 값에 관심이 있습니다 . 잠재적 예측 변수는 Rotten Tomatoes Score, Audience Score 및 Genre 입니다 . 두 점수 (Score) 변수는 연속형이지만 Genre 는 명목형입니다 . 단계별 회귀를 사용하여 모형을 축소하기 전에 관심 변수를 탐색하려고 합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Hollywood Movies.jsp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. Genre 를 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .

그림 5.14 Genre 분포



Genre 에 9 개 수준이 있으므로 8 개의 모형 향으로 표시됩니다. 결측 데이터 때문에 추가 데이터 탐색 결과 8 개 수준만 단계별 회귀에 고려됩니다.

- 데이터 테이블의 " 열 " 패널에서 관심 열을 선택합니다 (Rotten Tomatoes Score, Audience Score, World Gross).
- 분석 > 선별 > 결측값 탐색을 선택합니다.
- Y, 열을 클릭한 후 확인을 클릭합니다.

그림 5.15 결측 열 보고서

결측 열 보고서 인터페이스는 결측값을 포함하는 열만 표시, 열 선택하기, 열에 색상 적용, 행 제외, 행에 색상 적용 등의 기능을 제공합니다.

열	결측 수
Rotten Tomatoes Score	2
Audience Score	1
World Gross	2

Rotten Tomatoes Score 의 2 개 행 , Audience Score 의 1 개 행 , World Gross 의 2 개 행에 각각 결측값이 있습니다.

8. " 결측 열 " 보고서에서 **열** 아래에 나열된 세 열을 선택합니다 .

9. **행 선택하기**를 클릭합니다 .

데이터 테이블의 " 행 " 패널에서 3 개 행이 선택되어 있음을 확인할 수 있습니다 . 이 세 행에
는 예측 변수 또는 반응에 대한 결측 데이터가 포함되어 있으므로 단계별 분석에서 자동으로
제외됩니다 . Adventure 범주에 포함된 항목은 128 번 행뿐이므로 해당 범주는 분석에서 완
전히 제거됩니다 . 결과적으로 단계별 분석에 사용할 Genre에는 8 개 범주만 있습니다 . 결측
데이터의 효과를 확인했으므로 이제 단계별 분석을 수행합니다 .

10. **분석 > 모형 적합**을 선택합니다 .

11. Rotten Tomatoes Score, Audience Score 및 Genre 를 선택하고 **추가**를 클릭합니다 .

Rotten Tomatoes Score, Audience Score 및 Genre를 예측 변수로 사용하여 표준 최소 제곱 모
형을 World Gross 에 적합시키면 잔차의 이분산성이 매우 커집니다 . 이것은 재무 데이터의
대표적인 예입니다 . 로그 변환을 사용하면 등분산 회귀 가정을 더 잘 충족시킬 수 있습니다 .

12. " 열 선택 " 목록에서 World Gross 를 마우스 오른쪽 버튼으로 클릭하고 **변환 > 로그**를 선택합
니다 .

변환된 변수 로그 [World Gross] 가 " 열 선택 " 목록 아래쪽에 나타납니다 .

13. 로그 [World Gross] 를 선택하고 **Y**를 클릭합니다 .

14. " 분석법 " 목록에서 **단계별**을 선택합니다 .

15. **실행**을 클릭합니다 .

그림 5.16 모형 항 목록을 보여 주는 현재 추정값 테이블

현재 추정값							
잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☒	☒	절편	4.09998703	1	0	0.000	1
☐	☐	Rotten Tomatoes Score	0	1	10.78976	3.303	0.07154
☐	☐	Audience Score	0	1	8.715114	2.655	0.10577
☐	☐	Genre{Drama&Horror&Thriller&Fantasy&Romance&Comedy-Action&Animation}	0	1	49.78148	16.850	7.25e-5
☐	☐	Genre{Drama&Horror&Thriller-Fantasy&Romance&Comedy}	0	1	9.560063	2.918	0.09008
☐	☐	Genre{Drama-Horror&Thriller}	0	1	2.027141	0.608	0.43718
☐	☐	Genre{Horror-Thriller}	0	1	0.013011	0.004	0.95043
☐	☐	Genre{Fantasy&Romance-Comedy}	0	1	1.428612	0.428	0.51439
☐	☐	Genre{Fantasy-Romance}	0	1	0.000376	0.000	0.99157
☐	☐	Genre{Action-Animation}	0	1	1.362765	0.408	0.52426

" 현재 추정값 " 테이블에서 Genre 는 7 개 항으로 표시됩니다 . 이 중 두 개의 항으로 모형을
생성하여 이러한 항이 어떻게 정의되는지 확인합니다 .

16. **추가** 아래에서 Genre 의 처음 두 항 옆에 있는 상자를 선택합니다 .

- Genre{Drama&Horror&Thriller&Fantasy&Romance&Comedy-Action&Animation}
- Genre{Drama&Horror&Thriller-Fantasy&Romance&Comedy}

17. **모형 생성**을 클릭합니다 .

이 두 항이 " 모형 규격 " 창의 모형 효과 목록에 임시 변환 열로 추가됩니다 . 다음 섹션에서
이 열에 대해 설명합니다 .

예제의 계층 항 생성

결측값으로 인해 Genre는 8개 수준의 명목형 변수입니다. "현재 추정값" 테이블에는 Genre가 7개 항목으로 표시됩니다. Genre의 수준이 8개이므로 이것은 적절합니다. 아래에서는 Genre를 나타내는 처음 두 항목에 대해 설명합니다. 후속 항목은 이와 비슷한 방식으로 정의됩니다.

첫 번째 항

처음 나타나는 항은 Genre{Drama&Horror&Thriller&Fantasy&Romance&Comedy-Action&Animation}입니다. 이 변수의 형식은 Genre{A1 - A2}이며, A1과 A2는 빼기 기호(-)로 구분됩니다. 이 표기는 다음 두 수준 집합 사이에 그룹 간 제곱항을 기준으로 최대 분리가 발생한다는 것을 나타냅니다.

- Drama, Horror, Thriller, Fantasy, Romance, Comedy(A1로 표시)
- Action, Animation(A2로 표시)

Genre{Drama&Horror&Thriller&Fantasy&Romance&Comedy-Action&Animation} 항을 모형에 포함하면 해당 항을 나타내는 임시 변환 열이 모형에 사용됩니다. 열에 다음 값이 포함됩니다.

- 1 - Drama, Horror, Thriller, Fantasy, Romance, Comedy
- -1 - Action, Animation

두 번째 항

두 번째로 나타나는 항은 Genre{Drama&Horror&Thriller-Fantasy&Romance&Comedy}입니다. 이 수준 집합은 첫 번째 항의 첫 번째 분할(A1)에 완전히 포함됩니다. 이 표기는 다음 수준을 대조합니다.

- Drama, Horror, Thriller
- Fantasy, Romance, Comedy

알고리즘은 Drama, Horror, Thriller, Fantasy, Romance, Comedy 수준(A1)과 Action, Animation 수준(A2)의 모든 분할 중에서 이 분할이 그룹 간 제곱항이 가장 큰 것으로 결정합니다.

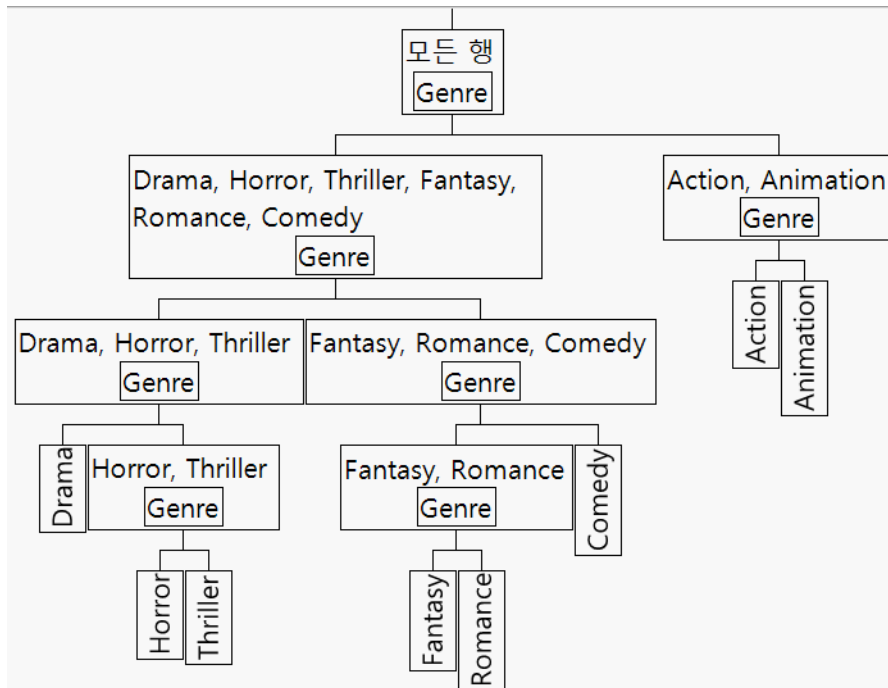
이 항을 모형에 포함하면 해당 항을 나타내는 임시 변환 열이 모형에 사용됩니다. 열에 다음 값이 포함됩니다.

- 1 - Drama, Horror, Thriller
- -1 - Fantasy, Romance, Comedy
- 0 - Action, Animation

항 계층

그룹 간 제곱합 기준에 따라 항 분할이 계속됩니다. 그림 5.17에서는 항 정의를 생성하는 계층을 보여 줍니다.

그림 5.17 계층적 코딩에 사용되는 분할을 보여 주는 트리



규칙

결합 규칙 또는 **제한** 규칙을 사용하는 경우 계층에서 위에 있는 모든 항이 입력되지 않으면 해당 항을 모형에 입력할 수 없습니다. 예를 들어 Genre{Action-Animation}을 입력하면 JMP에서 Genre{Drama&Horror&Thriller&Fantasy&Romance&Comedy-Action&Animation}도 입력합니다.

전체 효과 규칙을 사용하고 Genre 항 중 하나를 입력하면 모든 Genre 항이 입력됩니다.

계층 항에 대한 제한 규칙의 예

명목형 또는 순서형 항이 있는 모형의 경우 모형을 생성하거나 실행하면 모형에 포함된 계층 항을 포함하는 임시 변환 열이 모형 적합에 사용됩니다. 모형 자체는 새 "모형 적합" 창에 나타납니다. 이 예에서는 **단계별** 회귀에서 계층적 효과가 있는 모형을 생성하는 방법을 자세히 보여 줍니다.

핫도그 종류 Type(Meat, Beef, Poultry) 과 핫도그 크기 Size(Jumbo, Regular, Hors d'oeuvre) 의 함수로 핫도그의 온스당 비용 (\$/oz) 을 조사하는 간단한 모형이 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Hot Dogs2.jmp 를 엽니다.

2. 분석 > 모형 적합을 선택합니다.
3. \$/oz 를 선택하고 Y 를 클릭합니다.
4. Type 과 Size 를 선택하고 추가를 클릭합니다.
5. 분석법에서 단계별을 선택합니다.
6. 실행을 클릭합니다.
7. 중지 규칙에서 p 값 임계를 선택합니다.
8. 규칙에서 제한을 선택합니다.

그림 5.18 P 값 임계 및 제한 규칙이 설정된 단계별 회귀 제어 패널

단계별 회귀 제어

중지 규칙: p 값 임계 ➡ 모두 입력 모형 생성
 선택 확률 0.25 ⬅ 모두 제거 모형 실행
 제거 확률 0.1

방향: 전진 선택

규칙: 제한

시작 중지 단계

SSE	DFE	RMSE	R ²	Adj-R ²	Cp	p	AICc	BIC
0.1186093	53	0.0473066	-0.000	-0.0000	40.296085	1	-173.048	-169.306

현재 추정값

잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☒	☒	절편	0.1112963	1	0	0.000	1
☐	☐	Type(Poultry&Meat-Beef)	0	1	0.040492	26.954	3.51e-6
☐	☐	Type(Poultry-Meat)	0	0	0	.	.
☐	☐	Size(Hors d'oeuvre-Regular&Jumbo)	0	1	0.00299	1.345	0.25145
☐	☐	Size(Regular-Jumbo)	0	0	0	.	.

기본 규칙으로 선택된 "결합"을 "제한"으로 변경하면 두 항에 대한 "F 비" 및 "Prob>F" 값이 결측으로 표시됩니다. 해당 항은 Type(Poultry-Meat) 와 Size(Regular-Jumbo) 입니다. 이러한 두 항은 상위 항이 입력될 때까지 모형에 입력될 수 없기 때문입니다.

9. 단계를 클릭합니다.

Type(Poultry&Meat-Beef) 항이 모형에 입력됩니다. 이 항은 "Prob>F" 값이 가장 작으며 해당 값은 "선택 유의확률" 임계값 0.25 미만입니다.

그림 5.19 하나의 항이 추가된 단계별 제어판

현재 추정값

잠금	추가	모수	추정값	nDF	SS	"F 비"	"Prob>F"
☒	☒	절편	0.11864706	1	0	0.000	1
☐	☒	Type(Poultry&Meat-Beef)	-0.0283529	1	0.040492	26.954	3.51e-6
☐	☐	Type(Poultry-Meat)	0	1	0.012426	9.648	0.00309
☐	☐	Size(Hors d'oeuvre-Regular&Jumbo)	0	1	0.001038	0.687	0.4112
☐	☐	Size(Regular-Jumbo)	0	0	0	.	.

Type(Poultry-Meat) 항에 대한 "F 비" 및 "Prob>F" 값이 나타납니다. 상위 항이 모형에 입력 되었으므로 이제 Type(Poultry-Meat) 를 입력할 수 있습니다.

10. 단계를 클릭합니다.

Type{Poultry-Meat} 는 나머지 항 중 "Prob>F" 값이 가장 작고 해당 값이 "선택 유의확률" 임계값 미만이므로 모형에 입력할 다음 항입니다.

11. 단계를 클릭합니다.

"Prob>F" 값이 0.1577 인 Size{Hors d'oeuvre-Regular&Jumbo} 항이 모형에 입력됩니다. 이제 상위 항이 모형에 있으므로 Size{Regular-Jumbo} 항을 모형에 입력할 수 있으며 "Prob>F" 값이 나타납니다.

그러나 Size{Regular-Jumbo} 항의 "Prob>F" 값은 "선택 유의확률" 값 0.25 를 초과하는 0.7566 입니다. 따라서 "단계" 를 다시 클릭하면 모형에 항이 입력되지 않습니다.

그림 5.20 최종 모형에 대한 현재 추정값 보고서

현재 추정값						
잠금	추가	모수	추정값	nDF	SS	"F 비" "Prob>F"
☒	☒	절편	0.10849381	1	0	0.000 1
☐	☒	Type{Poultry&Meat-Beef}	-0.0275278	0	0	.
☐	☒	Type{Poultry-Meat}	-0.0205527	1	0.013985	11.083 0.00164
☐	☒	Size{Hors d'oeuvre-Regular&Jumbo}	-0.0121982	1	0.002596	2.057 0.1577
☐	☐	Size{Regular-Jumbo}	0	1	0.000125	0.097 0.7566

팁 : 전체 단계별 프로세스를 자동으로 실행하려면 "시작" 버튼을 사용하십시오. 이 작업을 확인하려면 모두 제거를 클릭합니다. 그런 다음 시작을 클릭합니다.

12. 모형 생성을 클릭합니다.

"모형 생성" 을 클릭하면 단계별 프로세스에서 선택된 세 가지 모형 효과만 포함하는 모형 적합 시작 창이 나타납니다. 시작 창에서 모형에 입력된 세 가지 계층적 효과를 정의하는 임시 변환 열이 모형에 포함됩니다.

이항 및 순서형 로지스틱 단계별 회귀 수행

모형 적합의 단계별 분석법은 반응이 순서형 또는 명목형일 때 순서형 로지스틱 단계별 회귀를 수행합니다. 명목형 반응은 로지스틱 단계별 회귀 적합 절차에서 순서형 반응으로 처리됩니다. 반응의 수준이 두 개뿐인 경우 순서형 로지스틱 회귀 모형은 명목형 로지스틱 회귀 모형과 동일합니다. 로지스틱 단계별 회귀를 실행하려면 순서형 또는 명목형 반응을 지정하고, 모형에 항을 추가한 후 "분석법" 메뉴에서 "단계별" 을 선택합니다.

로지스틱 모형에 대한 단계별 보고서는 반응이 연속형일 때 제공되는 보고서와 비슷합니다. 다음 요소는 로지스틱 회귀 결과에만 해당됩니다.

- 반응이 범주형이면 모형의 전체 적합이 음의 로그 가능도 (-1* 로그 가능도) 로 제공됩니다. 이 값은 전체 반복 최대 가능도 적합을 기반으로 계산됩니다. 자세한 내용은 "통계 상세 정보" 부록의 "가능도, AICc 및 BIC"(514 페이지) 에서 확인하십시오.
- "현재 추정값" 보고서에는 카이제곱 통계량 ("Wald/ 점수 카이제곱") 및 해당 p 값 ("유의 확률") 이 표시됩니다. 검정 통계량 열에는 현재 모형에 없는 모수에 대한 점수 통계량 및 현재 모형에 있는 모수에 대한 Wald 통계량이 표시됩니다. 회귀 추정값 ("추정값") 은 전체 반복 최대 가능도 적합을 기반으로 합니다.

- " 단계 기록 " 보고서에는 "L-R 카이제곱 " 이 표시됩니다 . 이 값은 모형의 다른 항이 주어졌을 때 해당 회귀 모수가 0 이라는 가설의 가능도비 검정에 대한 검정 통계량입니다 . " 유의 확률 " 은 이 검정의 p 값입니다 .

참고 : 반응이 명목형인 경우 " 모형 생성 " 버튼을 클릭하면 모형 적합의 명목형 로지스틱 분석법을 사용하여 현재 모형을 적합시킬 수 있습니다 . " 모형 적합 " 시작 창이 나타나면 " 실행 " 을 클릭합니다 .

예 : 로지스틱 단계별 회귀 사용

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Fitness.jmp** 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. **Sex** 를 선택하고 **Y** 를 클릭합니다 .
4. **Weight, Runtime, RunPulse, RstPulse** 및 **MaxPulse** 를 선택하고 **추가**를 클릭합니다 .
5. **분석법**에서 **단계별**을 선택합니다 .
6. **실행**을 클릭합니다 .
7. **시작**을 클릭합니다 .

그림 5.21 로지스틱 단계별 보고서

현재 추정값

잠금	추가	모수	추정값	nDF	점수	Wald/카이제곱	"유의 확률"
☒	☒	절편[M]	-35.807846	1		0	1
☐	☒	Weight	0.28229655	1	5.471429	0.01933	
☐	☒	Runtime	1.32680366	1	2.545732	0.11059	
☐	☐	RunPulse		0	0.470408	0.4928	
☐	☐	RstPulse		0	0.007216	0.9323	
☐	☐	MaxPulse		0	0.076497	0.7821	

단계 기록

단계	모수	작업	L-R 카이제곱	"유의 확률"	카이제곱 입력	"유의 확률" 입력	R ²	p	AICc	BIC	
1	Weight	추가	12.16669	0.0005	10.1342	0.00146	0.2833	2	35.2047	37.6441	
2	Runtime	추가	8.017826	0.0046	6.37139	0.0116	0.4700	3	29.6472	33.0603	
3	RunPulse	추가	1.440088	0.2301	1.44684	0.22903	0.5036	4	30.8567	35.0542	
4	MaxPulse	추가	0.071809	0.7887	0.06994	0.79142	0.5052	5	33.6464	38.4164	
5	RstPulse	추가	0.007156	0.9326	0.00722	0.93228	0.5054	6	36.7393	41.8432	
6	최선	특정				0.00722	0.93228	0.4700	3	29.6472	33.0603

중지 규칙에 따라 **Weight** 와 **Runtime** 두 변수가 모형에 입력됩니다 .

8. **모형 생성**을 클릭합니다 .

그림 5.22 축소 모형에 대한 모형 규격 창

모형 규격

열 선택

9개 열

Name
Sex
Age
Weight
Oxy
Runtime
RunPulse
RstPulse
MaxPulse

역할 변수 선택

Y: Sex (선택적)

가중치: 선택적 숫자

변도: 선택적 숫자

검증: 선택적 숫자

기준: 선택적

분석법: 명목형 로지스틱

목표 수준: M

도움말

실행

재호출

대화상자 열린 채 유지

제거

모형 효과 생성

추가: Weight, Runtime

교차

내포

매크로

자수: 2

속성

변환

절편 없음

두 변수가 모형 효과로 포함된 모형 규격 창이 나타납니다. "분석법" 이 "명목형 로지스틱" 으로 선택되어 있습니다. 반응이 순서형인 경우 "분석법" 은 "순서형 로지스틱" 입니다.

모든 가능 모형 옵션

"모든 가능 모형" 옵션을 사용하면 예측 변수를 사용하여 생성할 수 있는 모든 모형을 조사할 수 있습니다. 이 옵션은 단계별 회귀의 빨간색 삼각형에서 액세스합니다.

다음 사항에 유의하십시오.

- 이 옵션은 모형 수가 5 백만 개를 초과하는 큰 문제에는 실용적이지 않습니다.
- 범주형 예측 변수는 표시 변수로 나타납니다. 자세한 내용은 "[범주형 및 순서형 효과가 있는 모형](#)" (244 페이지) 에서 확인하십시오.

다음 옵션은 표시되는 모형 수를 제한합니다.

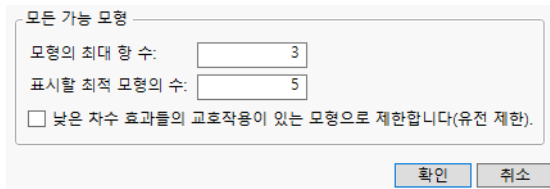
모형의 최대 항 수: 모형의 최대 항 수 값을 입력합니다.

표시할 최적 모형의 수: 표시할 각 크기의 최대 모형 수를 입력합니다. R^2 값에 따라 최적 모형이 나타납니다.

낮은 차수 효과들의 교호작용이 있는 모형으로 제한합니다 (유전 제한). 교차 효과가 포함된 경우 모든 저차 효과를 포함하는 모형만 표시합니다. 이러한 모형은 강한 효과 유전성을 충족합니다. 이 옵션은 예측 변수에 교호작용 또는 다항식 항이 포함된 경우에 유용합니다.

예 : 모든 가능 모형 옵션 사용

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Fitness.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. Oxy 를 선택하고 **Y** 를 클릭합니다 .
4. Runtime, RunPulse, RstPulse 및 MaxPulse 를 선택하고 **추가**를 클릭합니다 .
5. **분석법**에서 **단계별**을 선택합니다 .
6. **실행**을 클릭합니다 .
7. 단계별 회귀의 빨간색 삼각형 메뉴를 클릭하고 **모든 가능 모형**을 선택합니다 .
8. 최대 항 수에 3 을 입력하고 최적 모형의 수에 5 를 입력합니다 .

그림 5.23 모든 가능 모형 팝업 대화상자


모든 가능 모형

모형의 최대 항 수:

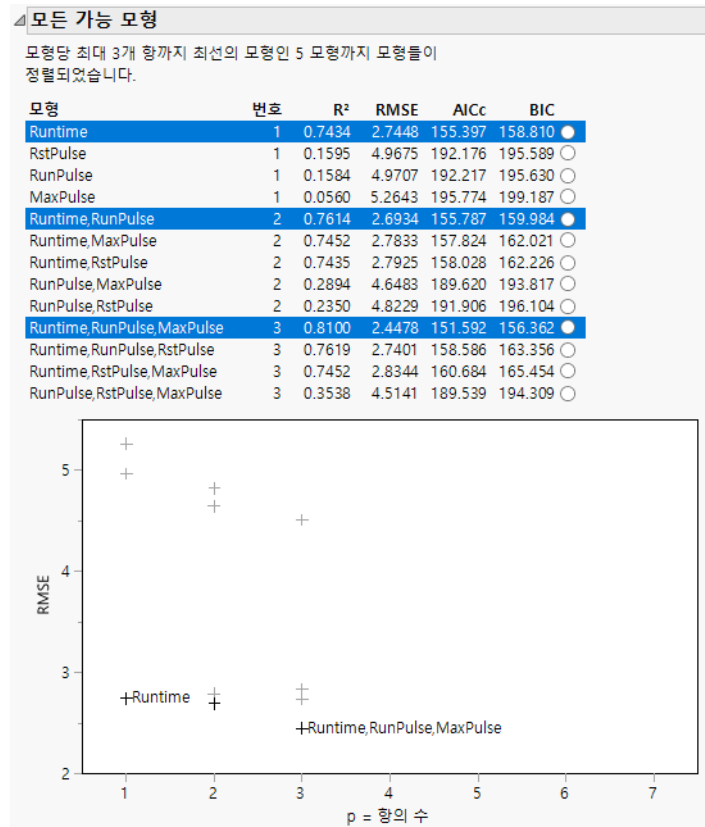
표시할 최적 모형의 수:

☐ 낮은 차수 효과들의 교호작용이 있는 모형으로 제한합니다(유전 제한).

9. **확인**을 클릭합니다 .

가능한 모든 모형(모형에 최대 3개 항 포함)이 적합됩니다.

그림 5.24 모든 가능 모형 보고서



포함된 변수의 수가 작은 것부터 순서대로 모형이 나열됩니다. 각 변수 수에 대해 R^2 이 가장 높은 모형이 강조 표시됩니다. 테이블 오른쪽의 라디오 버튼 열을 사용하면 한 번에 하나의 모형을 선택하여 결과를 확인할 수 있습니다.

참고 : 모형 선택에 권장되는 기준은 BIC 또는 AICc 값이 가장 작은 모형을 선택하는 것입니다. 일부 분석가는 C_p 통계량도 확인하려고 합니다. Mallows C_p 통계량이 계산되지만 처음에는 테이블에 숨겨져 있습니다. 값을 표시하려면 테이블을 마우스 오른쪽 버튼으로 클릭하고 **열 > C_p** 를 선택하십시오.

모형 평균화 옵션

모형 평균화 기법을 사용하면 단일 모형을 선택하는 대신 여러 모형의 적합을 평균화할 수 있습니다. 평균화된 각 단일 모형보다 결과 평균 모형의 예측 성능이 더 우수한 경우가 종종 있습니다. 모형 평균화 기능은 데이터 과대적합 모형을 방지하는 데 유용합니다. 모형에 입력할 항이 많이 선택되면 적합 시 모수 추정값이 부풀려지는 경향이 있습니다. 모형을 평균화하면 더 약한

항에 대한 추정값이 축소되므로 우수한 예측이 생성됩니다. 모형은 다음과 같이 계산되는 각 모형의 AICc 가중치를 고려하여 평균화됩니다.

$$\text{AICcWeight} = \exp[-0.5(\text{AICc} - \text{AICcBest})]$$

AICcBest는 적합된 모형 중에서 가장 작은 AICc 값입니다. AICc 가중치 값은 각 모형에 대해 계산되어 내림차순으로 정렬되며 합이 1이 되도록 척도화됩니다. 합이 1보다 작은 척도화된 AICc 가중치 값에서 지정된 누적 AICc 가중치 경계 값을 뺀 값은 0으로 설정됩니다. 이로 인해 약한 모형은 평균화된 모형에 사용되지 않습니다. 평균화된 모형의 모수는 AICcWeight가 0이 아닌 모형에 대해 모수 추정값의 가중 평균입니다.

예 : 모형 평균화 옵션 사용

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Fitness.jmp** 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. **Oxy** 를 선택하고 **Y** 를 클릭합니다.
4. **Runtime, RunPulse, RstPulse** 및 **MaxPulse** 를 선택하고 **추가**를 클릭합니다.
5. **분석법**에서 **단계별**을 선택합니다.
6. **실행**을 클릭합니다.
7. 단계별 회귀의 빨간색 삼각형 메뉴를 클릭하고 **모형 평균화**를 선택합니다.
8. "모형의 최대 항 수"에 3을 입력합니다.
9. "누적 AICc 가중치 경계"의 0.95를 그대로 유지합니다.

그림 5.25 모형 평균화 창

모형 평균화

모형의 최대 항 수:

누적 AICc 가중치 경계

10. **확인**을 클릭합니다.

그림 5.26 모형 평균화 보고서

0.9707 AICc 가중치 분위수 경계를 사용하여 1 ~ 3개의 항이 있는 평균 모형, 14개의 모형 적합 중 5개가 사용됩니다.

모수	추정값	표준 오차
절편	82.4063	.
Runtime	-3.0422	0.3465967
RunPulse	-0.2832	0.1053199
RstPulse	-0.000286	0.0126522
MaxPulse	0.2603	0.1141399

[예측 계산식 저장](#)

"모형 평균화" 보고서에는 각 모수에 대한 평균 추정값 및 표준 오차가 나타납니다. 표시된 표준 오차는 추정값의 편향을 0에 가깝게 반영합니다.

11. **예측 계산식 저장**을 클릭하여 예측 계산식을 원래 데이터 테이블에 저장합니다.

단계별 회귀의 검증 옵션

JMP에서 단계별 회귀 모형에 대한 교차 검증을 수행하려면 "단계별 회귀 적합"의 빨간색 삼각형을 클릭하고 **K 폴드 교차 검증**을 선택합니다. 자세한 내용은 "단계별 회귀의 K 폴드 교차 검증"(260 페이지)에서 확인하십시오.

JMP PRO JMP Pro에서는 "모형 적합" 창에 검증 열을 지정할 수 있습니다. 검증 열은 숫자 데이터 유형이어야 하고 두 개 이상의 구분되는 값을 포함해야 합니다.

- 열에 두 개의 값이 포함된 경우 더 작은 값이 훈련 데이터 집합을 정의하고 더 큰 값이 검증 데이터 집합을 정의합니다.
- 열에 세 개의 값이 포함된 경우 작은 값부터 순서대로 훈련 데이터 집합, 검증 데이터 집합, 테스트 데이터 집합을 각각 정의합니다.
- 열에 네 개 이상의 구분되는 값이 포함되고 반응이 연속형인 경우 해당 값은 k 폴드 검증을 위한 폴드를 정의합니다.

검증 열 사용에 대한 자세한 내용은 "단계별 회귀에서 두 개 또는 세 개의 값이 있는 검증 데이터 집합"(257 페이지)에서 확인하십시오.

단계별 회귀에서 두 개 또는 세 개의 값이 있는 검증 데이터 집합

JMP PRO 두 개 또는 세 개의 값이 있는 검증 열을 지정하는 경우 단계별 회귀는 훈련 데이터 집합을 기반으로 모형을 적합시킵니다. 검증 데이터 집합 및 테스트 데이터 집합에 대한 모형 적합 통계량이 보고됩니다. 이러한 통계량이 정의되는 방식에 대한 자세한 내용은 "검증 및 테스트 데이터 집합 통계량 정의"(258 페이지)에서 확인하십시오.

반응이 연속형인 경우 다음 통계량이 "단계별 회귀 제어" 패널에 나타납니다.

- 검증 R^2 ("단계 기록" 보고서에도 표시됨)
- RASE 검증
- 테스트 R^2 (테스트 데이터 집합이 있는 경우)
- RASE 테스트 (테스트 데이터 집합이 있는 경우)

반응이 이항 명목형 또는 순서형인 경우 다음 통계량이 "단계별 회귀 제어" 패널에 나타납니다.

- 검증 R^2 ("단계 기록" 보고서에도 표시됨)
- 검증 평균 로그 오차
- 테스트 R^2 (테스트 데이터 집합이 있는 경우)

- 테스트 평균 로그 오차 (테스트 데이터 집합이 있는 경우)

최대 검증 R^2

모형 적합 창에서 두 개 또는 세 개의 값이 있는 검증 열을 지정하는 경우 "중지 규칙"은 기본적으로 "최대 검증 R^2 "으로 설정됩니다. 이 규칙은 검증 데이터 집합에 대해 R^2 통계량을 최대화하는 모형을 찾습니다. "전진 선택" 또는 "후진 제거"로 설정된 방향과 함께 이 규칙을 적용할 수 있습니다.

참고: 최대 검증 R^2 은 p 값에 따른 입력 (전진 선택 방향) 또는 제거 (후진 제거 방향)로 정의된 모형만 고려합니다. 이 규칙은 모든 가능 모형을 고려하지 않습니다.

"단계" 버튼을 사용하여 항을 하나씩 입력하거나(전진 선택 방향) 하나씩 제거할 수 있습니다(후진 제거 방향). 언제든지 "단계 기록" 보고서의 "검증 R^2 " 오른쪽에 있는 버튼을 클릭하여 모형을 선택할 수 있습니다. 선택한 모형 항은 "현재 추정값" 보고서에서 업데이트됩니다. "모형 생성" 또는 "모형 실행"을 클릭하면 이 모형이 사용됩니다.

전진 선택 방향

전진 선택 방향을 사용하는 경우 단계별 회귀는 가장 작은 다음 p 값을 기준으로 항을 추가하여 연속 모형을 생성합니다.

"단계" 대신 "시작"을 클릭하면 항 입력 프로세스가 자동으로 진행됩니다. 적합 모형 중 최적으로 간주되는 모형이 마지막에 나열됩니다. 이 모형은 검증 R^2 의 국소적 감소를 무시하여 얻습니다. 특히 검증 R^2 이 가장 큰 모형이며, 검증 R^2 값이 더 작은 모형이 10개까지 있을 수 있습니다. 이 모형은 "모수" 열의 "최선" 항과 "작업" 열의 "특정" 항으로 지정됩니다. "검증 R^2 " 오른쪽의 버튼을 클릭하면 이 최적 모형이 선택되지만 선택을 변경해도 됩니다.

후진 제거 방향

후진 제거 방향을 사용하는 경우 단계별 회귀는 가장 큰 다음 p 값을 기준으로 항을 제거하여 연속 모형을 생성합니다.

"후진 제거" 방향을 사용하려면 먼저 "모두 입력"을 클릭하여 모든 항을 모형에 입력해야 합니다. 후진 제거 방향은 전진 선택 방향과 비슷한 방식으로 작동합니다. "단계" 대신 "시작"을 클릭하면 항 제거 프로세스가 자동으로 진행됩니다. 최적 모형은 검증 R^2 이 가장 큰 모형이며, 검증 R^2 값이 더 작은 모형이 10개까지 있을 수 있습니다.

검증 및 테스트 데이터 집합 통계량 정의

이 섹션에는 검증 R^2 및 RASE 검증이 정의되어 있습니다. 테스트 R^2 및 RASE 테스트는 테스트 데이터 집합에 대해 완전히 유사한 방식으로 계산됩니다.

연속형 반응

검증 R² 검증 데이터 집합에 대한 R² 측도이며 다음과 같이 계산됩니다.

- 검증 데이터 집합의 각 관측값에 대해 예측 오차를 계산합니다. 이 값은 실제 반응과 훈련 데이터 집합 모형에서 예측한 반응 간의 차이입니다.
- 예측 오차의 제곱합을 계산하여 $SSE_{Validation}$ 을 구합니다.
- 검증 데이터 집합의 실제 반응과 평균 간의 차이에 대한 제곱합을 계산합니다. 이 값이 $SST_{Validation}$ 입니다.
- 검증 R² 은 다음과 같습니다.

$$\text{검증 R}^2 = 1 - \frac{SSE_{Validation}}{SST_{Validation}}$$

참고 : 검증 R² 은 음수일 수 있습니다.

RASE 검증 검증 데이터 집합에 대한 평균 제곱 예측 오차의 제곱근입니다. 이 값은 다음과 같이 계산됩니다.

- 검증 데이터 집합의 각 관측값에 대해 예측 오차를 계산합니다. 이 값은 실제 반응과 훈련 데이터 집합 모형에서 예측한 반응 간의 차이입니다.
- 예측 오차의 제곱합을 계산하여 $SSE_{Validation}$ 을 구합니다.
- 검증 데이터 집합의 관측값 수를 $n_{Validation}$ 으로 나타냅니다.
- RASE 검증은 다음과 같습니다.

$$\text{RASE검증} = \sqrt{\frac{SSE_{Validation}}{n_{Validation}}}$$

이항 명목형 또는 순서형 반응

검증 R² 검증 데이터 집합에 대한 엔트로피 R² 측도 (McFadden R² 이라고도 함) 이며 다음과 같이 계산됩니다.

- 훈련 데이터 집합을 사용하여 모형이 적합됩니다.
- 모든 관측값에 대한 예측 확률을 구합니다.
- 훈련 데이터 집합 모형에 기반한 예측 확률을 사용하여 검증 데이터 집합의 관측값에 대해 모형의 가능도가 계산됩니다. 이 통계량을 $Likelihood_{FullValidation}$ 이라고 합니다.
- 검증 데이터 집합의 데이터를 사용하여 축소 모형 (예측 변수 없음) 의 가능도가 계산됩니다. 이 통계량을 $Likelihood_{ReducedValidation}$ 이라고 합니다.
- 검증 R² 은 다음과 같습니다.

$$\text{검증 } R^2 = 1 - \frac{\log(\text{Likelihood\_Full}_{\text{Validation}})}{\log(\text{Likelihood\_Reduced}_{\text{Validation}})}$$

참고 : 검증 R^2 은 음수일 수 있습니다.

검증 평균 로그 오차 검증 데이터 집합의 평균 로그 오차는 다음과 같이 계산됩니다.

- 검증 데이터 집합의 각 관측값에 대해 훈련 데이터 집합에 기반한 모형에 의해 결정된 예측 확률의 로그를 계산합니다.
- 이 로그의 합을 검증 데이터 집합의 관측값 수로 나눈 후 결과 값의 음수를 구합니다.

팁 : 검증 평균 로그 오차 값이 작을수록 좋습니다.

단계별 회귀의 K 폴드 교차 검증

K 폴드 교차 검증은 데이터를 무작위로 k 개의 부분집합으로 나눕니다. 결과적으로 k 개 집합이 각각 검증 데이터 집합으로 사용되고 나머지 데이터는 모형 적합을 위한 훈련 데이터 집합으로 사용됩니다. 총 k 개 모형이 적합되고 k 개의 검증 통계량이 계산됩니다. 최종 모형에서는 최량 검증 통계량을 제공하는 모형이 선택됩니다. 이 방법은 제한된 양의 데이터를 효율적으로 사용하기 때문에 소규모 데이터 집합에 유용합니다.

참고 : K 폴드 교차 검증은 연속형 반응에만 사용할 수 있습니다.

JMP에서 "단계별 회귀 적합"의 빨간색 삼각형을 클릭하고 **K 폴드 교차 검증**을 선택합니다.

JMP Pro에서는 다음 두 가지 방법으로 k 폴드 교차 검증에 액세스할 수 있습니다.

- "단계별 회귀 적합"의 빨간색 삼각형을 클릭하고 **K 폴드 교차 검증**을 선택합니다.
- 네 개 이상의 구분되는 값이 포함된 검증 열을 지정합니다.

R^2 K 폴드 통계량

K 폴드 교차 검증을 수행하면 "단계별 회귀 제어" 패널의 다른 통계량 오른쪽에 " R^2 K 폴드" 통계량이 나타납니다. R^2 K 폴드는 k 폴드에 대한 검증 R^2 값의 평균입니다.

최대 K 폴드 R^2

k 폴드 교차 검증을 사용하는 경우 "중지 규칙"은 기본적으로 "최대 K 폴드 R^2 "으로 설정됩니다. 이 규칙은 R^2 K 폴드 통계량을 최대화합니다.

참고 : 최대 K 폴드 R^2 은 p 값에 따른 입력 (전진 선택 방향) 또는 제거 (후진 제거 방향) 로 정의된 모형만 고려합니다. 이 규칙은 모든 가능 모형을 고려하지 않습니다.

최대 K 폴드 R^2 중지 규칙은 최대 검증 R^2 중지 규칙과 비슷한 방식으로 작동합니다. 자세한 내용은 "[최대 검증 \$R^2\$](#) "(258 페이지)에서 확인하십시오. " R^2 검증"을 " R^2 K 폴드"로 바꾸어 참조하면 됩니다.

모형 적합 플랫폼의 일반화 회귀 분석법은 JMP Pro에서만 사용할 수 있습니다.

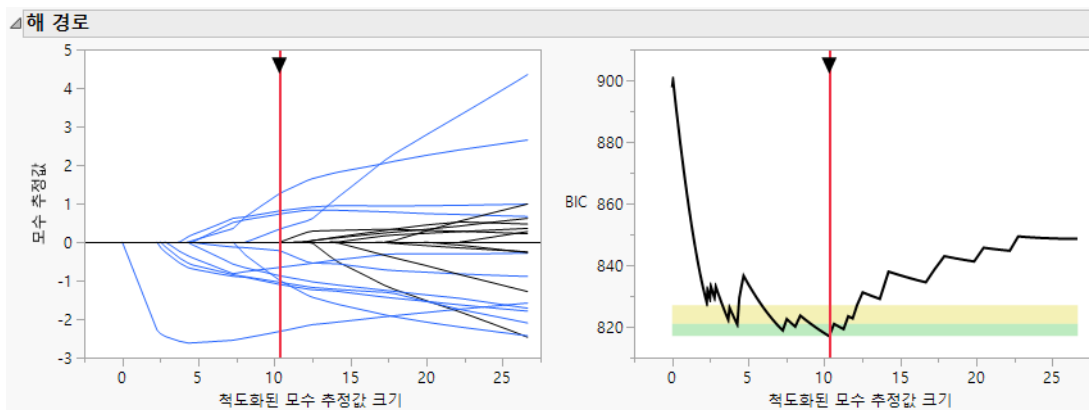
JMP Pro에서 모형 적합 플랫폼의 일반화 회귀 분석법은 축소 방법을 포함한 다양한 변수 선택 기법을 제공하며 특히 상관관계가 있는 고차원 데이터 모델링을 처리합니다. 이 중에서 Lasso 회귀와 Elastic Net이라는 두 가지 기법은 모델링 절차의 일부로 변수 선택을 수행합니다.

많은 변수를 포함하는 대규모 데이터 집합은 대개 다중공선성 문제를 나타냅니다. 최신 데이터 집합에는 관측값보다 변수가 더 많이 포함될 수 있으므로 기존 모델링 기법을 사용하려면 변수 선택이 필요합니다. 다중공선성 및 많은 예측 변수가 있으면 전통적 설계 기법의 단점이 드러납니다.

설계된 실험을 포함하여 상관관계가 거의 없는 소규모 데이터 집합에서도 Lasso 회귀와 Elastic Net이 유용합니다. 예측 모형을 생성하거나, 모형 축소 또는 향후 연구를 위해 변수를 선택할 때 이 기법을 사용할 수 있습니다.

일반화 회귀 분석법은 많은 모델링 상황에 유용합니다. 이 분석법을 사용하면 반응 변수에 대한 다양한 분포를 지정할 수 있습니다. 반응이 연속형, 이항, 개수 또는 영과일일 때 사용합니다. 또한 변수 선택에 관심이 있거나 예측 변수의 공선성이 의심될 때 사용합니다. 보다 일반적으로는 다른 기법을 사용하여 생성한 모형과 비교하는 모형을 적합시킬 때 사용합니다.

그림 6.1 Elastic Net 적함에 대한 해 경로



목차

일반화 회귀 분석법 개요.....	265
일반화 회귀의 예.....	266
일반화 회귀 분석법 시작.....	269
분포.....	270
일반화 회귀 보고서 창.....	277
일반화 회귀 보고서 옵션.....	278
모형 시작 제어판.....	278
반응 분포.....	279
추정 방법 옵션.....	279
고급 제어.....	283
검증 방법 옵션.....	286
조기 중지.....	287
시작.....	287
모형 적합 보고서.....	287
회귀 그림.....	288
모형 요약.....	288
추정 상세 정보.....	291
해 경로.....	291
중심화 및 척도화된 예측 변수에 대한 모수 추정값.....	294
원래 예측 변수에 대한 모수 추정값.....	295
활성 모수 추정값.....	296
효과 검정.....	297
모형 적합 옵션.....	297
일반화 회귀 분석법에 대한 통계 상세 정보.....	305
추정 방법에 대한 통계 상세 정보.....	306
고급 제어에 대한 통계 상세 정보.....	308
분포에 대한 통계 상세 정보.....	309

JMP PRO 일반화 회귀 분석법 개요

일반화 회귀 분석법은 정규화(벌점) 회귀 기법을 특징으로 합니다. 이 기법은 모형 계수를 0에 가까워지도록 축소하여 더 나은 모형을 적합시키려고 합니다. 결과 추정값은 편향되어 있습니다. 편향이 증가하면 예측 분산이 감소되어 벌점 없는 모형에 비해 전체 예측 오차를 낮출 수 있습니다. 이 중에서 Elastic Net과 Lasso 회귀라는 두 가지 기법은 모델링 절차의 일부로 변수 선택을 포함합니다.

Elastic Net 및 Lasso 회귀와 같은 모델링 기법은 일반적으로 공선성이 문제가 되는 대규모 데이터 집합에 특히 유용합니다. 또한 최신 데이터 집합에는 관측값보다 변수가 더 많이 포함될 때가 종종 있습니다. 이 상황을 $p > n$ 문제라고도 하며, 여기서 n 은 관측값 수이고 p 는 예측 변수의 수입니다. 기존 모델링 기법을 사용할 경우 이러한 데이터 집합에 대해 변수 선택이 필요합니다.

설계된 실험을 포함하여 상관관계가 거의 없는 소규모 데이터 집합에도 Elastic Net과 Lasso 회귀를 사용할 수 있습니다. 예측 모형을 생성하거나, 모형 축소 또는 향후 연구를 위해 변수를 선택할 때 이 기법을 사용할 수 있습니다.

이 분석법은 다음과 같은 종류의 모델링 기법을 제공합니다.

- 최대 가능도
- 단계 기반 추정
- 벌점 회귀 방법

Elastic Net과 Lasso 회귀는 비교적 최근의 기법입니다(Tibshirani 1996, Zou and Hastie 2005). 두 기법 모두 모형 계수의 크기에 벌점을 적용하여 지속적 축소가 발생합니다. 축소 크기는 조정 모수에 의해 결정됩니다. 최적의 축소 수준은 여러 검증 방법 중 하나에 의해 결정됩니다. 두 기법 모두 계수를 0으로 축소할 수 있습니다. 이러한 방식으로 변수 선택이 모델링 절차에 기본 제공됩니다. Elastic Net 모형은 Lasso 회귀와 능형 회귀 둘 다 특수 사례로 포함합니다. 자세한 내용은 "[추정 방법에 대한 통계 상세 정보](#)"(306 페이지)에서 확인하십시오.

일반화 회귀 모델링 기법에 대한 상세 정보

- 최대 가능도 방법은 전통적 설계 방식입니다. 이 방법은 다른 기법과 비교할 수 있는 기준을 제공하며 가설 검정과 같은 기존 추론 기법에 대해 가장 적합합니다.
- 전진 선택은 단계별 회귀 방법입니다. 전진 선택에서는 항이 모형에 입력됩니다. 모든 항이 모형에 포함되거나 자유도가 남아 있지 않을 때까지 유의성이 가장 높은 항이 추가됩니다.
- Lasso 회귀에는 두 가지 단점이 있습니다. 상관관계가 높은 여러 변수가 있을 때 해당 그룹에서 하나의 변수만 선택하려는 경향이 있습니다. 변수 수 (p)가 관측값 수 (n)를 초과할 경우 Lasso 회귀는 최대 n 개의 예측 변수를 선택합니다.
- 반면, Elastic Net은 상관관계가 있는 그룹에서 모든 변수를 선택하여 적절한 계수를 적합시키는 경향이 있습니다. 또한 $p > n$ 일 때 n 개를 초과하는 예측 변수를 선택할 수 있습니다.
- 능형 회귀는 첫 번째로 제안된 벌점 회귀 방법입니다(Hoerl 1962, Hoerl and Kennard 1970). 능형 회귀는 계수를 0으로 축소하지 않으므로 변수 선택을 수행하지 않습니다.

- 이중 Lasso 회귀는 초기 Lasso 회귀 모형에서 변수 선택을 수행하여 선택 단계와 축소 단계를 구분하려고 합니다. 그런 다음 초기 모형에서 선택한 변수가 두 번째 Lasso 회귀 모형의 입력 변수로 사용됩니다.
- 2 단계 전진 선택은 두 단계의 단계별 전진 회귀를 수행합니다. 첫 번째 단계에서 주효과에 대한 변수 선택을 수행합니다. 그런 다음 두 번째 단계에서 고차 효과를 모형에 입력할 수 있습니다.

일반화 회귀 분석법은 Lasso 회귀 및 Elastic Net의 적응형 버전에도 적합합니다. 이러한 적응형 버전은 실제 활성 집합에 포함되지 않은 변수보다 실제 활성 집합의 변수에 벌점을 적게 줍니다. 실제 활성 집합은 모형에서 반응에 실제 영향을 미치는 항 집합을 나타냅니다. 적응형 Lasso 회귀 및 Elastic Net은 oracle 특성을 유지하기 위해 개발되었습니다. oracle 특성이 유지됨으로써 추정값은 점진적으로 예측 변수의 실제 활성 집합에 모형을 적합시켰을 때 얻을 수 있는 값이 됩니다. 더 구체적으로 말하면 모형이 계수가 0이어야 하는 예측 변수를 올바르게 식별할 수 있습니다. 추정값은 실제 활성 집합만 포함하여 시작했을 때 얻을 수 있는 값에 수렴합니다. 자세한 내용은 "[적응형 방법](#)"(307 페이지)에서 확인하십시오.

일반화 회귀 분석법을 사용하면 반응 변수에 대한 다양한 분포를 지정할 수 있습니다. 분포 적합에는 정규, Cauchy, 스튜던트 t , 지수, 감마, Weibull, 로그 정규, 베타, 이항, 베타 이항, Poisson, 음이항, 영과잉 이항, 영과잉 베타 이항, 영과잉 Poisson, 영과잉 음이항 및 영과잉 감마가 포함됩니다. 이러한 유연성을 통해 범주형 반응과 개수 반응, 연속형 반응 및 오른쪽으로 치우친 연속형 반응을 적합시킬 수 있습니다. 또한 분위수 회귀 모형과 Cox 비례 위험 모형도 적합시킬 수 있습니다. 일부 분포의 경우 중도절단 데이터에 모형을 적합시킬 수 있습니다. 이 분석법은 모형 선택을 위한 다양한 검증 기준을 제공하고 훈련, 검증 및 테스트 열을 지원합니다. 자세한 내용은 "[분포](#)"(270 페이지)에서 확인하십시오.

JMP 일반화 회귀의 예

Diabetes.jmp 샘플 데이터 테이블의 데이터는 442명의 당뇨병 환자에 대한 측정값으로 구성됩니다. 관심 반응은 Y 이며 기준 측정을 수행하고 1년 후 측정된 질병 진행률입니다. 질병 진행과 관련이 있다고 간주되는 10가지 변수도 기준에서 측정됩니다. 이 예에서는 일반화 회귀 기법을 사용하여 예측 모형을 개발하는 방법을 보여 줍니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Diabetes.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. "열 선택" 목록에서 Y 를 선택하고 Y 를 클릭합니다.
4. Age ~ Glucose 를 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다.
이렇게 하면 차수 2(차수 상자의 기본값)까지의 모든 항이 모형에 추가됩니다.
5. "열 선택" 목록에서 Validation 을 선택하고 **검증**을 클릭합니다.
6. "분석법" 목록에서 **일반화 회귀**를 선택합니다.
7. **실행**을 클릭합니다.

"모형 비교" 보고서, "모형 시작" 제어판 및 "검증 열을 사용한 정규 표준 최소 제곱" 보고서가 포함된 "일반화 회귀" 보고서가 나타납니다.

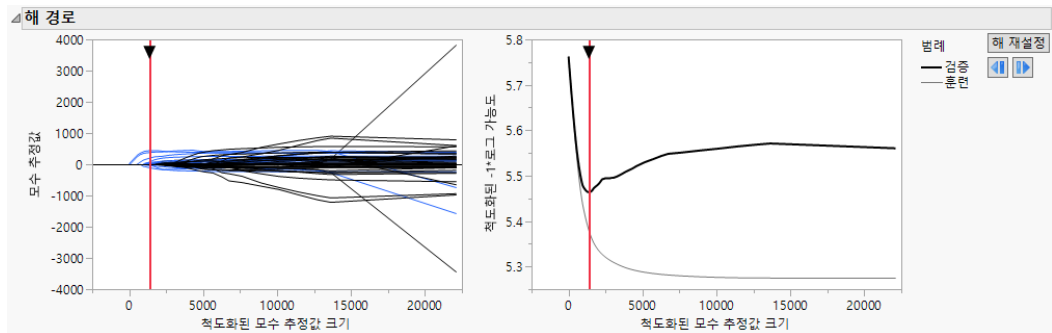
"모형 시작" 제어판에서 다음 사항에 유의하십시오.

- "모형 적합" 시작 창에서 "분포"에 "정규"를 지정했으므로 "반응 분포"가 "정규"로 설정됩니다.
- 기본 추정 방법은 Lasso 회귀입니다.
- "모형 적합" 창에서 검증 열을 지정했으므로 "검증 방법"이 "검증 열"로 설정됩니다.

8. 시작을 클릭합니다.

"검증 열을 사용한 정규 Lasso 회귀" 보고서가 나타납니다. "해 경로" 보고서 (그림 6.2)에서는 모수 추정값 및 척도화된 -1^* 로그 가능도 그림을 보여 줍니다. 척도화된 모수 추정값의 크기가 감소하면 축소량이 증가합니다. 그림 맨 오른쪽의 추정값은 최대 가능도 추정값입니다. 빨간색 수직선은 검증 기준 (이 예에서는 Validation 열에 정의된 홀드백 표본)에 의해 선택된 모수 값을 나타냅니다.

그림 6.2 해 경로 그림



9. "검증 열을 사용한 정규 Lasso 회귀" 옆의 빨간색 삼각형을 클릭하고 0 이 아닌 항 선택을 선택합니다.

이 옵션은 "원래 예측 변수에 대한 모수 추정값" 보고서 (그림 6.3)에서 0 이 아닌 항을 강조 표시하고 해 경로 그림에서 해당 경로를 강조 표시합니다. 데이터 테이블의 해당 열도 선택됩니다. 55 개의 모수 추정값 중 11 개만 0 이 아닙니다. 정규 분포에 대한 척도 모수 (시그마)도 추정되어 "원래 예측 변수에 대한 모수 추정값" 보고서 하단에 있는 별도의 테이블에 표시됩니다. 55 개의 모수 추정값 중 일부만 그림 6.3에 나와 있습니다.

그림 6.3 원래 예측 변수에 대한 모수 추정값 보고서의 일부

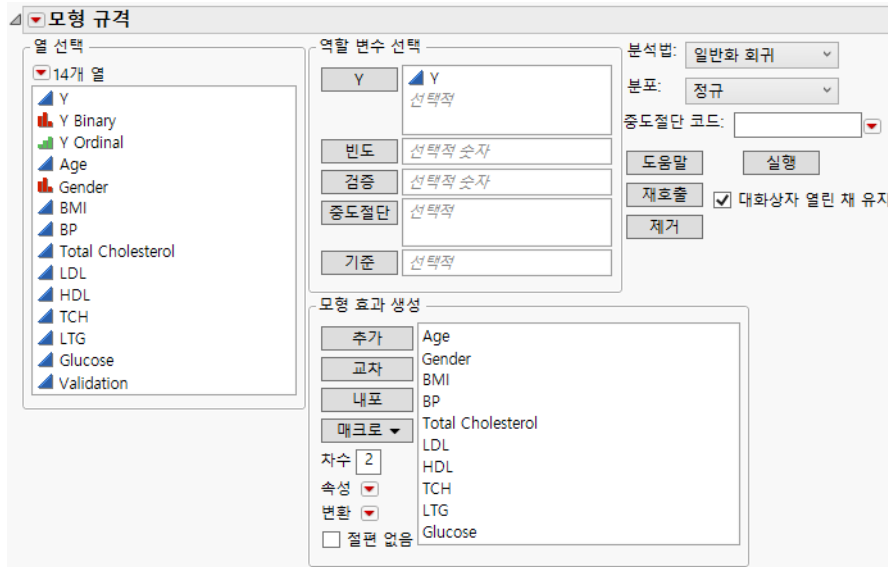
원래 예측 변수에 대한 모수 추정값							
항	추정값	표준 오차	Wald 카이제곱	Prob > ChiSquare	95% 하한	95% 상한	
절편	-247.8422	42.926975	33.334183	<.0001*	-331.9775	-163.7068	
Age	0	0	0	1.0000	0	0	
Gender[1-2]	8.7286488	6.9678967	1.5692446	0.2103	-4.928178	22.385475	
BMI	5.702121	0.9303784	37.562422	<.0001*	3.8786128	7.5256293	
BP	0.8156913	0.2526706	10.421787	0.0012*	0.320466	1.3109166	
Total Cholesterol	0	0	0	1.0000	0	0	
LDL	0	0	0	1.0000	0	0	
HDL	-0.47425	0.3986141	1.4155011	0.2341	-1.25552	0.3070188	
TCH	0	0	0	1.0000	0	0	
LTG	41.286654	7.0844177	33.963392	<.0001*	27.401451	55.171858	
Glucose	0	0	0	1.0000	0	0	
(Age-48.5181)*(Gender[1-2])	-0.082735	0.3621512	0.0521915	0.8193	-0.792538	0.6270682	
(Age-48.5181)*(BMI-26.3758)	0	0	0	1.0000	0	0	
(Age-48.5181)*(BP-94.647)	0.0043544	0.0171735	0.0642885	0.7998	-0.029305	0.0380137	
(Age-48.5181)*(Total Cholesterol-189.14)	0	0	0	1.0000	0	0	
(Age-48.5181)*(LDL-115.439)	0	0	0	1.0000	0	0	
(Age-48.5181)*(HDL-49.7885)	0	0	0	1.0000	0	0	
(Age-48.5181)*(TCH-4.07025)	0	0	0	1.0000	0	0	
(Age-48.5181)*(LTG-4.64141)	0	0	0	1.0000	0	0	
(Age-48.5181)*(Glucose-91.2602)	0.0231484	0.0233144	0.9858084	0.3208	-0.022547	0.0688439	
Gender[1-2]*(BMI-26.3758)	0	0	0	1.0000	0	0	
Gender[1-2]*(BP-94.647)	0	0	0	1.0000	0	0	
Gender[1-2]*(Total Cholesterol-189.14)	0	0	0	1.0000	0	0	
Gender[1-2]*(LDL-115.439)	0	0	0	1.0000	0	0	
Gender[1-2]*(HDL-49.7885)	-0.221009	0.4728172	0.2184913	0.6402	-1.147714	0.7056955	
Gender[1-2]*(TCH-4.07025)	0	0	0	1.0000	0	0	
Gender[1-2]*(LTG-4.64141)	0	0	0	1.0000	0	0	
Gender[1-2]*(Glucose-91.2602)	0	0	0	1.0000	0	0	
(BMI-26.3758)*(BP-94.647)	0.0254032	0.0525392	0.2337819	0.6287	-0.077572	0.1283782	
(BMI-26.3758)*(Total Cholesterol-189.14)	0	0	0	1.0000	0	0	
(BMI-26.3758)*(LDL-115.439)	0	0	0	1.0000	0	0	
(BMI-26.3758)*(HDL-49.7885)	0	0	0	1.0000	0	0	
(BMI-26.3758)*(TCH-4.07025)	0	0	0	1.0000	0	0	
(BMI-26.3758)*(LTG-4.64141)	0	0	0	1.0000	0	0	
(BMI-26.3758)*(Glucose-91.2602)	0.0210214	0.0597655	0.1237153	0.7250	-0.096117	0.1381596	
Prob >							
정규 분포 모수	추정값	표준 오차	Wald 카이제곱	ChiSquare	95% 하한	95% 상한	
적도	52.512586	2.3230199	511.00007	<.0001*	47.959551	57.065621	

예측 계산식을 저장하려면 "검증 열을 사용한 정규 Lasso 회귀" 보고서 옆의 빨간색 삼각형을 클릭하고 **열 저장 > 예측 계산식 저장**을 선택합니다.

JMP PRO 일반화 회귀 분석법 시작

분석 > 모형 적합을 선택하고 Y에 대해 하나 이상의 열을 입력한 후 분석법 메뉴에서 일반화 회귀를 선택하여 일반화 회귀 분석법을 시작합니다.

그림 6.4 일반화 회귀가 선택된 모형 적합 시작 창



"모형 적합" 창에서 모든 분석법에 공통된 요소에 대한 자세한 내용은 "모형 규격" 장(29 페이지)에서 확인하십시오. "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오. 여기에서는 일반화 회귀 분석법에 대한 정보를 제공합니다.

일반화 회귀 분석법에 사용되는 명목형 변수의 파라미터화는 다른 모형 적합 분석법을 사용한 파라미터화와 다릅니다. 일반화 회귀 분석법은 표시 함수 파라미터화를 사용합니다. 이 파라미터화 방법에서 명목형 변수 수준의 표시자에 해당하는 추정값은 같은 수준의 평균 반응과 마지막 수준의 평균 반응 간 차이에 대한 추정값입니다. 마지막 수준은 값 순서 코딩이 가장 높은 수준이며 표시 함수가 모형에 포함되지 않은 수준입니다.

모형 효과에 결측값이 있는 경우 이러한 결측값을 정보 범주로 처리할 수 있습니다. "모형 규격"의 빨간색 삼각형 메뉴에서 "결측값 정보화" 옵션을 선택합니다.

절편 항이 없는 모형을 지정하려면 "모형 적합" 창의 "모형 효과 생성" 패널에서 "절편 없음" 옵션을 선택합니다. 이 옵션을 선택하는 경우 다음 사항에 유의하십시오.

- 예측 변수가 중심화되거나 척도화되지 않습니다.
- 보고서 창에서 승산비, 위험 비율 및 발생률 비율을 사용할 수 없습니다.
- 순서형 로지스틱 분포에서는 "절편 없음" 옵션을 사용할 수 없습니다.

주의 : 결과는 모형 효과의 척도에 민감하므로 Lasso 회귀 또는 Elastic Net 에서 "절편 없음" 옵션을 사용하는 것은 권장되지 않습니다. 대신 이 추정 방법의 적응형 버전을 사용하는 것이 좋습니다.

중도절단

다음 방법 중 하나를 사용하여 반응 변수에 대한 중도절단을 지정할 수 있습니다.

- 구간 중도절단, 오른쪽 중도절단 및 왼쪽 중도절단 반응의 경우 "감지 한계" 열 특성을 포함하는 열을 지정합니다. 열 특성에 지정된 한계는 중도절단되지 않은 것으로 간주되는 반응 값 범위를 정의합니다.
 - 구간 중도절단 반응의 경우 "감지 한계" 열 특성에서 "감지 하한" 과 "감지 상한" 둘 다에 비결측값을 지정합니다.
 - 오른쪽 중도절단 반응의 경우 "감지 한계" 열 특성에서 "감지 하한"에는 결측값을 지정하고 "감지 상한"에는 비결측값을 지정합니다.
 - 왼쪽 중도절단 반응의 경우 "감지 한계" 열 특성에서 "감지 하한"에는 비결측값을 지정하고 "감지 상한"에는 결측값을 지정합니다.
- 오른쪽 중도절단 반응의 경우 시작 창에서 오른쪽 중도절단된 관측값에 대한 표시자가 포함된 열을 "중도절단" 열로 지정합니다. 그런 다음 "중도절단 코드" 목록에서 해당 열의 값 중 오른쪽 중도절단된 관측값을 지정하는 값을 선택합니다.
- 구간 중도절단 및 왼쪽 중도절단 반응의 경우 "Y" 열 역할에 중도절단 구간을 정의하는 두 열을 지정합니다.
 - 구간 중도절단 반응의 경우 첫 번째 Y 변수는 하한을 제공합니다.
 - 왼쪽 중도절단 반응의 경우 첫 번째 Y 변수는 결측값을 포함합니다.
 - 구간 중도절단 반응과 왼쪽 중도절단 반응 모두 두 번째 Y 변수는 각 반응의 상한을 제공합니다.

Y 및 중도절단을 지원하는 분포에 대한 두 열을 지정하면 열이 중도절단을 나타내는지 여부를 묻는 경고가 표시됩니다. "아니요"를 선택할 경우 해당 열은 별도의 반응으로 처리됩니다.

참고 : 일반화 회귀 플랫폼 환경 설정의 "두 개의 반응 열에 대한 처리" 설정을 사용하여 두 반응에 대한 기본 동작을 지정할 수 있습니다.

정규, 지수, 감마, Weibull, 로그 정규 또는 Cox 비례 위험 분포가 지정된 경우 중도절단을 사용할 수 있습니다.

JMP[®] PRO 분포

분석법 메뉴에서 **일반화 회귀**를 선택하면 **분포** 옵션이 나타납니다. 여기서 Y의 분포를 지정할 수 있습니다. 약어 ZI는 **영과잉(Zero-Inflated)**을 의미합니다. 분포는 반응에 따라 연속형, 이산형 및 영과잉의 세 가지 범주로 구분됩니다. 아래에 옵션에 대한 설명이 나옵니다.

참고 : " 모형 규격 " 창에서 Y 변수를 여러 개 지정하는 경우 지정된 모든 Y 변수에 대해 동일한 반응 분포를 사용해야 합니다. 동일한 " 일반화 회귀 " 보고서에서 각 반응 변수마다 개별 분포를 적합시키려면 스크립트를 사용해야 합니다.

JMP PRO 연속형

정규 Y 는 평균이 μ 이고 표준편차가 σ 인 정규 분포를 가집니다. 정규 분포는 대칭적이며, 표준 크기가 충분히 크면 중심 극한 정리에 의해 다양한 다른 분포가 정규 분포와 가까워질 수 있습니다. μ 의 연결 함수는 항등입니다. 즉, Y 의 평균은 선형 모형으로 표현됩니다.

참고 : 정규 분포가 지정된 경우 "표준 최소 제곱"이 "최대 가능성도" 추정 방법을 대체합니다.

정규 분포의 척도 모수는 σ 입니다. 추정 방법에 별점이 없는 경우 척도 모수 σ 의 추정값은 RMSE(제곱근 평균 제곱 오차) 입니다. RMSE 는 σ^2 의 일반적인 비편향 추정량에 대한 제곱근입니다. 표시된 결과는 중도절단된 관측값이 포함된 경우가 아니면 표준 최소 제곱 적합과 동등합니다.

참고 : 일반화 회귀 분석법에 사용되는 명목형 변수의 파라미터화는 표준 최소 제곱 분석법을 사용한 파라미터화와 다릅니다. 이 차이 때문에 명목형 또는 순서형 효과를 포함하는 모형에 대한 모수 추정값이 달라집니다.

자세한 내용은 "분포에 대한 통계 상세 정보"(309 페이지)에서 확인하십시오.

Cauchy Y 는 위치 모수가 μ 이고 척도 모수가 σ 인 Cauchy 분포를 가집니다. Cauchy 분포는 평균과 표준편차가 정의되어 있지 않습니다. 중앙값과 최빈값은 모두 μ 입니다. 대부분의 데이터는 본질적으로 Cauchy 분포를 따르지 않습니다. 그러나 이상치 비율이 큰 (최대 50%) 데이터에 대해 로버스트 회귀를 수행하는 데 유용합니다. μ 의 연결 함수는 항등입니다. 자세한 내용은 "분포에 대한 통계 상세 정보"(309 페이지)에서 확인하십시오.

t(5) Y 는 자유도가 5 이고 위치 모수가 μ , 척도 모수가 σ 인 스튜던트 t 분포를 가집니다. 스튜던트 t 분포는 대칭적이며, 정규 분포와 Cauchy 분포 사이의 범위를 포괄하는 로버스트 옵션입니다. 스튜던트 t 분포의 자유도가 무한대에 가까워지면 정규 분포와 동등합니다. 스튜던트 t 분포의 자유도가 1 이면 Cauchy 분포와 동등합니다. μ 의 연결 함수는 항등입니다. 자세한 내용은 "분포에 대한 통계 상세 정보"(309 페이지)에서 확인하십시오.

지수 Y 는 평균 모수가 μ 인 지수 분포를 가집니다. 지수 분포는 오른쪽으로 치우친 분포로, 대개 수명이나 연속된 이벤트 사이의 시간을 모델링하는 데 사용됩니다. μ 의 연결 함수는 로그입니다. 자세한 내용은 "분포에 대한 통계 상세 정보"(309 페이지)에서 확인하십시오.

감마 Y 는 평균 모수가 μ 이고 산포 모수가 σ 인 감마 분포를 가집니다. 감마는 널리 사용되는 다른 분포 계열을 포함하는 유연한 분포입니다. 예를 들어 지수 분포는 감마 분포의 특수한 경우로, $\sigma = \mu$ 입니다. 카이제곱 분포도 감마 분포에서 파생될 수 있습니다. μ 의 연결 함수는 로그입니다. 자세한 내용은 "분포에 대한 통계 상세 정보"(309페이지)에서 확인하십시오.

Weibull Y 는 평균 모수가 μ 이고 척도 모수가 σ 인 Weibull 분포를 가집니다. Weibull 분포는 유연한 분포로, 대개 수명이나 이벤트까지의 시간을 모델링하는 데 사용됩니다. μ 의 연결 함수

수는 항등입니다. 자세한 내용은 "분포에 대한 통계 상세 정보"(309 페이지)에서 확인하십시오.

로그 정규 Y 는 평균 모수가 μ 이고 척도 모수가 σ 인 로그 정규 분포를 가집니다. 로그 정규 분포는 오른쪽으로 치우친 분포로, 대개 수명이나 이벤트까지의 시간을 모델링하는 데 사용됩니다. μ 의 연결 함수는 항등입니다. 자세한 내용은 "분포에 대한 통계 상세 정보"(309 페이지)에서 확인하십시오.

베타 Y 는 평균 모수가 μ 이고 산포 모수가 σ 인 베타 분포를 가집니다. 베타 분포에 대한 반응은 0에서 1 사이이며 대개 비율을 모델링하는 데 사용됩니다. μ 의 연결 함수는 로짓입니다. 자세한 내용은 "분포에 대한 통계 상세 정보"(309 페이지)에서 확인하십시오.

분위수 회귀 분위수 회귀는 반응에 대한 지정된 조건부 분위수를 모델링합니다. 기본 분포의 형태에 대해서는 가정하지 않습니다. 분위수 회귀를 선택하면 "분포" 메뉴 아래에 "분위수" 상자가 나타납니다. 원하는 분위수를 지정합니다.

"모형 대화상자" 창에서 "분위수"에 0.5(기본값)를 지정할 경우 분위수 회귀는 반응의 조건부 중앙값을 모델링합니다. 분위수 회귀는 회귀 계수로 표현되는 조건부 분위수의 변화 비율이 분위수에 따라 달라지는 경우에 특히 유용합니다. 최소 제곱 회귀와 비교할 때 분위수 회귀의 장점은 이질적 조건부 분포를 사용하여 데이터를 모델링할 수 있는 유연성입니다.

분위수 회귀는 반복 방식을 통해 목적 함수를 최소화하여 적합됩니다. 분위수 회귀에 대한 자세한 내용은 [Koenker and Hallock 연구 자료 \(2001\)](#) 및 [Portnoy and Koenker 연구 자료 \(1997\)](#)에서 확인하십시오.

분위수 회귀를 선택할 경우 "추정 방법"에서 "최대 가능도"만 사용할 수 있으며 "검증 방법"에서는 "없음"만 사용할 수 있습니다.

참고: 분위수 회귀 적합에 시간이 많이 소요되는 경우 진행률 표시줄이 나타납니다. 진행률 표시줄에는 목적 함수의 상대적 변동이 표시됩니다. "현재 추정값 채택"을 클릭하면 계산이 중지되며 보고된 모수 추정값은 해당 지점에서 최적 모형 적합에 해당합니다.

Cox 비례 위험 Cox 비례 위험 모형은 예측 변수가 있는 사건 발생 시간 데이터에 대한 회귀 모형입니다. 이 모형은 예측 변수와 위험 함수 간의 승법 관계를 기반으로 합니다. 예측 변수가 생존 시간에 미치는 영향을 조사하는 데 이 모형을 사용할 수 있습니다. 모형에는 일반적인 위험 함수를 제공하기 위해 예측 변수로 척도화된 임의의 기준 위험 함수가 포함됩니다. 비례 위험 모형은 각 예측 변수에 대한 모수 추정값과 표준 오차를 생성합니다. Cox 비례 위험 모형은 [D. R. Cox\(1972\)](#)에 의해 처음 제안되었습니다. 비례 위험 모형에 대한 자세한 내용은 [Kalbfleisch and Prentice 연구 자료 \(2002\)](#)에서 확인하십시오.

Cox 비례 위험을 선택할 경우 "검증 방법"에서 "BIC"와 "AICc"만 사용할 수 있습니다. 또한 능형 회귀 추정 방법을 사용할 수 없습니다.

참고: 반응에 동일 값이 있으면 Efron 가능도가 사용됩니다. 자세한 내용은 [Efron 연구 자료 \(1981\)](#)에서 확인하십시오. 이 방법은 모형 적합 플랫폼의 비례 위험 분석법 또는 비례 위험 모형 적합 플랫폼에서 사용되는 동일 값 처리 방법과 다릅니다.

JMP PRO 이산형

이항 Y 는 p 와 n 모수가 있는 이항 분포를 가집니다. 반응 Y 는 n 번의 독립 시행 (모든 시행에 대한 확률 p 는 고정됨) 에서 총 성공 횟수를 나타냅니다. 이 분포에서는 표본 크기 열을 사용할 수 있습니다. 나열된 열이 없으면 표본 크기가 1 이라고 가정합니다. p 의 연결 함수는 로짓입니다. 명목형 모델링 유형이 있는 이진 반응 변수를 선택하면 이항 반응 분포만 사용할 수 있습니다. 자세한 내용은 "[분포에 대한 통계 상세 정보](#)"(309페이지)에서 확인하십시오.

"분포"에서 "이항"을 선택하면 다음 방법 중 하나를 사용하여 반응 변수를 지정해야 합니다.

- 요약되지 않음: 데이터가 사건 빈도로 요약되지 않은 경우 단일 이진 열을 반응으로 지정합니다. 이 열의 모델링 유형이 명목형이면 수준 중 하나를 목표 수준으로 지정할 수 있습니다. 기본 목표 수준 값은 수준 순서에 따라 두 수준 중 더 높은 값입니다.
- 빈도 열로 요약됨: 데이터가 성공 및 실패 빈도로 요약되는 경우 단일 이진 열을 반응으로 지정합니다. 이 열의 모델링 유형이 명목형이면 수준 중 하나를 목표 수준으로 지정할 수 있습니다. 기본 목표 수준 값은 수준 순서에 따라 두 수준 중 더 높은 값입니다. 빈도 열을 빈도 역할에 할당합니다.
- 두 번째 Y 로 입력된 표본 크기 열로 요약됨: 데이터가 사건 (성공) 빈도 및 시행 횟수로 요약되는 경우, 성공 횟수 - 시행 횟수의 순으로 두 개의 연속형 열을 Y 로 지정합니다.

참고: 이항 분포가 지정된 경우 "[로지스틱 회귀](#)"가 "[최대 가능도](#)" 추정 방법을 대체합니다.

베타 이항 Y 는 성공 확률 p , 시행 횟수 n 및 과대산포 모수 δ 가 있는 베타 이항 분포를 가집니다. 이 분포는 이항 분포의 과대산포 버전입니다.

산포 모수 δ 를 갖는 베타 이항 분포를 모수 p 및 $n = 20$ 을 갖는 이항 분포와 비교하려면 JMP Samples/Scripts 폴더에 있는 demoBetaBinomial.jsl 을 실행하십시오.

베타 이항 분포를 사용하려면 각 관측값에 대해 표본 크기가 1 보다 커야 합니다. 따라서 사용자는 표본 크기 열을 지정해야 합니다. 표본 크기 열을 삽입하려면 성공 횟수 - 시행 횟수의 순으로 두 개의 연속형 열을 Y 로 지정합니다. p 의 연결 함수는 로짓입니다. 자세한 내용은 "[분포에 대한 통계 상세 정보](#)"(309 페이지)에서 확인하십시오.

다항 Y 는 이산형 수준이 세 개 이상인 다항 분포를 가집니다. 반응 변수의 모델링 유형은 명목형 또는 순서형이어야 합니다. 이 모형은 반응 변수의 각 수준에 대해 별도의 절편 및 효과 모수를 적합시킵니다. 반응 변수의 수준이 k 개이면 모형에 $k - 1$ 개의 절편 및 효과 모수가 포함됩니다. 다항 분포의 연결 함수는 다항 로짓입니다. 자세한 내용은 "[통계 상세 정보](#)" 장의 "[명목형 반응](#)"(489 페이지)에서 확인하십시오.

순서형 로지스틱 Y 는 순서형 수준이 있는 다항 분포를 가집니다. 반응 변수의 모델링 유형은 순서형이어야 합니다. 이 모형은 반응 변수의 각 수준에 대해 절편을 적합시킵니다. 효과 모수는 반응 변수의 모든 수준에서 공통입니다. 순서형 로지스틱 분포의 연결 함수는 순서형 로짓입니다. 자세한 내용은 "[통계 상세 정보](#)" 장의 "[순서형 반응](#)"(491 페이지)에서 확인하십시오.

참고 : 일반화 회귀에서 순서형 로지스틱에 대한 절편 파라미터화는 모형 적합의 순서형 로지스틱 분석법에서와 다릅니다. 일반화 회귀의 첫 번째 절편 항은 순서형 로지스틱 분석법의 첫 번째 절편 항에 해당합니다. 일반화 회귀의 후속 절편 항은 반응 변수의 순서형 수준에 대한 절편 항 간의 연속 차이입니다.

Poisson Y 는 평균이 λ 인 Poisson 분포를 가집니다. 일반적으로 Poisson 분포는 특정 구간 내의 사건 수를 모델링하며 대개 개수 데이터로 표현됩니다. λ 의 연결 함수는 로그입니다. Poisson 회귀는 Y 가 정수가 아닌 값이더라도 허용됩니다. 자세한 내용은 "[분포에 대한 통계 상세 정보](#)"(309 페이지)에서 확인하십시오.

음이항 Y 는 평균이 μ 이고 산포 모수가 σ 인 음이항 분포를 가집니다. 일반적으로 음이항 분포는 지정된 실패 횟수에 도달하기 전의 성공 횟수를 모델링합니다. 또한 음이항 분포는 특정 조건에서 감마 Poisson 분포와 동등합니다. 음이항 분포와 감마 Poisson 분포 간 연결에 대한 자세한 내용은 기본 분석의 에서 확인하십시오.

평균 λ 및 산포 모수 σ 를 갖는 감마 Poisson 분포를 평균 λ 를 갖는 Poisson 분포와 비교하려면 JMP Samples/Scripts 폴더에 있는 demoGammaPoisson.jsl을 실행하십시오.

μ 의 연결 함수는 로그입니다. 음이항 회귀는 Y 가 정수가 아닌 값이더라도 허용됩니다. 자세한 내용은 "[분포에 대한 통계 상세 정보](#)"(309 페이지)에서 확인하십시오.

JMP PRO 영과잉

ZI 이항 Y 는 모수 p, n 및 영과잉 모수 π 가 있는 영과잉 이항 분포를 가집니다. 반응 Y 는 n 번의 독립 시행 (모든 시행에 대한 확률 p 는 고정됨)에서 총 성공 횟수를 나타냅니다. 이 분포에서는 표본 크기 열을 사용할 수 있습니다. 나열된 열이 없으면 표본 크기가 1이라고 가정합니다. p 의 연결 함수는 로짓입니다. 자세한 내용은 "[분포에 대한 통계 상세 정보](#)"(309 페이지)에서 확인하십시오.

ZI 베타 이항 Y 는 성공 확률 p , 시행 횟수 n , 과대산포 모수 δ 및 영과잉 모수 π 가 있는 베타 이항 분포를 가집니다. 이 분포는 ZI 이항 분포의 과대산포 버전입니다. ZI 베타 이항 분포를 사용하려면 각 관측값에 대해 표본 크기가 1보다 커야 합니다. 따라서 사용자는 표본 크기 열을 지정해야 합니다. 표본 크기 열을 삽입하려면 성공 횟수 - 시행 횟수의 순으로 두 개의 연속형 열을 Y 로 지정합니다. p 의 연결 함수는 로짓입니다. 자세한 내용은 "[분포에 대한 통계 상세 정보](#)"(309 페이지)에서 확인하십시오.

ZI Poisson Y 는 평균 모수가 λ 이고 영과잉 모수가 π 인 영과잉 Poisson 분포를 가집니다. λ 모수는 0 펄창 없이 Poisson 분포의 관측값을 기반으로 하는 조건부 평균입니다. λ 의 연결 함수는 로그입니다. ZI Poisson 회귀는 Y 가 정수가 아닌 값이거나 0 값이 관측되지 않더라도 허용됩니다. 자세한 내용은 "[분포에 대한 통계 상세 정보](#)"(309 페이지)에서 확인하십시오.

ZI 음이항 Y 는 위치 모수 μ , 산포 모수 σ 및 영과잉 모수 π 가 있는 영과잉 음이항 분포를 가집니다. μ 모수는 0 펄창 없이 음이항 분포의 관측값을 기반으로 하는 조건부 평균입니다. μ 의 연결 함수는 로그입니다. ZI 음이항 회귀는 Y 가 정수가 아닌 값이거나 0 값이 관측되지 않더라도 허용됩니다. 자세한 내용은 "[분포에 대한 통계 상세 정보](#)"(309 페이지)에서 확인하십시오.

ZI 감마 Y 는 평균 모수가 μ 이고 영과잉 모수가 π 인 영과잉 감마 분포를 가집니다. 많은 경우 0이 아닌 반응이 감마 분포를 따른다고 간주할 수 있습니다. 보험금 청구의 경우 여기에 해당합니다. 청구 값은 감마 분포에 근사하지만 청구가 없는 보험 증권에 대한 데이터에는 0도 있습니다. 영과잉 감마 분포는 데이터를 0 반응과 0이 아닌 반응으로 나누지 않고도 이러한 데이터를 직접 처리할 수 있습니다. μ 모수는 0 팽창 없이 감마 분포의 관측값을 기반으로 하는 조건부 평균입니다. μ 의 연결 함수는 로그입니다. 자세한 내용은 "[분포에 대한 통계 상세 정보](#)"(309 페이지)에서 확인하십시오.

표 6.1에는 다양한 분포가 할당된 Y 변수의 데이터 유형, 모델링 유형 및 기타 요구 사항이 나와 있습니다.

표 6.1 분포에 대한 Y 요구 사항

분포	데이터 유형	모델링 유형	기타
정규	숫자	연속형	
Cauchy	숫자	연속형	
t(5)	숫자	연속형	
지수	숫자	연속형	양수
감마	숫자	연속형	양수
Weibull	숫자	연속형	양수
로그 정규	숫자	연속형	양수
베타	숫자	연속형	0에서 1 사이
분위수 회귀	숫자	연속형	
Cox 비례 위험	숫자	연속형	음수 아님
이항, 요약되지 않음	임의	임의	이진
이항, 빈도 열로 요약됨	임의	임의	이진
이항, 두 번째 Y 로 입력된 개수 열로 요약됨	숫자	연속형	음수 아님
베타 이항	숫자	연속형	음수 아님
다항	임의	명목형 또는 순서형	
순서형 로지스틱	임의	순서형	
Poisson	숫자	임의	음수 아님

표 6.1 분포에 대한 Y 요구 사항 (계속)

분포	데이터 유형	모델링 유형	기타
음이항	숫자	임의	음수 아님
영과잉 이항	숫자	임의	음수 아님
영과잉 베타 이항	숫자	임의	음수 아님
영과잉 Poisson	숫자	임의	음수 아님
영과잉 음이항	숫자	임의	음수 아님
영과잉 감마	숫자	연속형	음수 아님

이러한 분포를 파라미터화하는 방법에 대한 자세한 내용은 "[분포에 대한 통계 상세 정보](#)"(309 페이지)에서 확인하십시오. 표 6.2에는 상세 정보가 요약되어 있습니다.

표 6.2 분포, 모수 및 연결 함수

분포	모수	평균 모형 연결 함수
정규	μ, σ	항등 (μ)
Cauchy	μ, σ	항등 (μ)
t(5)	μ, σ	항등 (μ)
지수	μ	로그 (μ)
감마	μ, σ	로그 (μ)
Weibull	μ, σ	항등 (μ)
로그 정규	μ, σ	항등 (μ)
베타	μ	로짓 (μ)
이항	n, p	로짓 (p)
베타 이항	n, p, δ	로짓 (p)
Poisson	λ	로그 (μ)
음이항	μ, σ	로그 (μ)
영과잉 이항	$n, p, \pi(\text{영과잉})$	로짓 (p)
영과잉 베타 이항	$n, p, \delta, \pi(\text{영과잉})$	로짓 (p)

표 6.2 분포, 모수 및 연결 함수 (계속)

분포	모수	평균 모형 연결 함수
영과잉 Poisson	λ, π (영과잉)	로그 (μ)
영과잉 음이항	μ, σ, π (영과잉)	로그 (μ)
영과잉 감마	μ, σ, π (영과잉)	로그 (μ)

적절한 분포를 선택한 후 **실행**을 클릭합니다. 그러면 "일반화 회귀" 보고서 창이 나타납니다.

일반화 회귀 보고서 창

"모형 적합" 시작 창에서 **실행**을 클릭하면 다음 항목이 포함된 "일반화 회귀" 보고서 창이 나타납니다.

- 모형 비교 보고서 - 보고서의 적합된 모든 모형을 비교할 수 있습니다. "모형 시작" 제어판을 사용하여 새 모형이 적합될 때마다 "모형 비교" 보고서에 추가됩니다. "표시" 열의 체크박스를 사용하여 적합 모형에 대한 보고서를 표시하거나 숨길 수 있습니다. "모형 비교" 보고서의 다른 열에는 적합된 각 모형에 대한 정보 및 적합 통계량이 포함됩니다. "모형 비교" 보고서의 열을 기준으로 모형을 정렬하려면 원하는 열 머리글을 클릭합니다. 처음 클릭하면 오름차순으로 정렬되고 해당 열 머리글을 두 번째 클릭하면 내림차순으로 정렬됩니다.

참고: 하나 이상의 모형이 적합되면 "모형 비교" 보고서가 나타납니다.

- 모형 시작 제어판 - 모형을 적합시킵니다. 자세한 내용은 "[모형 시작 제어판](#)"(278 페이지)에서 확인하십시오. 모형을 적합시키면 적합된 모형의 유형을 설명하는 개요가 제목에 추가됩니다. 자세한 내용은 "[모형 적합 보고서](#)"(287 페이지) 및 "[모형 적합 옵션](#)"(297 페이지)에서 확인하십시오.
 - 모형 항 사이에 선형 종속성이 있는 경우 모형 항이 충족하는 선형 함수를 보여 주는 "특이성 상세 정보" 보고서가 "모형 시작" 제어판에 포함됩니다. 자세한 내용은 "표준 최소 제곱 모형" 장의 "[모형 항 사이에 선형 종속성이 있는 모형](#)"(182 페이지)에서 확인하십시오.
- 최대 가능도 보고서 - 최대 가능도 추정을 사용하여 적합된 모형의 결과를 보여 줍니다. 자세한 내용은 "[최대 가능도](#)"(280 페이지)에서 확인하십시오. "최대 가능도" 보고서는 다음 조건이 충족될 경우에만 나타납니다.
 - 예측 변수 사이에 선형 종속성이 없습니다.
 - 관측값이 예측 변수보다 더 많습니다.
 - 예측 변수가 250 개 이하입니다.

참고 : 정규 분포가 지정된 경우 이 보고서에 "표준 최소 제곱" 라벨이 지정됩니다. 이항 분포가 지정된 경우 이 보고서에 "로지스틱 회귀" 라벨이 지정됩니다.

JMP^{PRO} 일반화 회귀 보고서 옵션

모형 대화상자 현재 분석을 위해 완료된 "모형 적합" 시작 창을 표시합니다.

난수 시드값 설정 K 폴드 및 홀드백 검증에 사용되는 난수화 과정을 위한 시드값을 설정합니다. 분석을 재현하려는 경우에 유용합니다. 시드를 양수 값으로 설정하고 스크립트를 저장하면 시드값이 스크립트에 자동으로 저장됩니다. 스크립트를 실행하면 항상 동일한 교차 검증 분석이 생성됩니다.

코딩 테이블 저장 첫 번째 열에 모든 모형 모수에 대한 JMP 코딩이 표시되는 새 데이터 테이블을 생성합니다. 마지막 열에는 반응 변수의 값이 표시됩니다. 이항 반응에 대한 표본 크기 열을 지정한 경우 반응 열과 표본 크기 열이 모두 테이블에 표시됩니다. 두 개의 반응 열을 사용하여 중도절단을 지정한 경우 두 반응 열이 모두 테이블에 표시됩니다. "감지 한계" 열 특성이 포함된 반응 열을 사용하여 중도절단을 지정한 경우에는 코딩 테이블의 반응 열에 "감지 한계" 열 특성이 포함됩니다.

참고 : 코딩 데이터 테이블에는 분석에 사용된 데이터 테이블 이름을 제공하는 "원래 데이터"라는 테이블 변수가 포함되어 있습니다. 기준 변수가 지정된 경우에는 "원래 데이터" 테이블 변수에 기준 변수와 해당 수준도 제공됩니다.

다음 옵션에 대한 자세한 내용은 **JMP 사용**의 에서 확인하십시오.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

JMP^{PRO} 모형 시작 제어판

"모형 시작" 제어판에는 다음과 같은 옵션이 있습니다.

- "반응 분포"(279 페이지)
- "추정 방법 옵션"(279 페이지)
- "고급 제어"(283 페이지)

- " 검증 방법 옵션 "(286 페이지)
- " 조기 중지 "(287 페이지)
- " 시작 "(287 페이지)

JMP PRO 반응 분포

제어판에서 사용할 수 있는 반응 분포는 "모형 적합" 시작 창에 지정된 반응 변수 속성에 따라 다릅니다. 적절한 반응 분포가 "반응 분포" 옵션에 제공됩니다. 반응 분포에 대한 자세한 내용은 "분포"(270 페이지)에서 확인하십시오.

JMP PRO 추정 방법 옵션

사용 가능한 추정 방법은 선택 및 별점이 없는 기법, 단계별 모형 선택 기법 및 별점 회귀 기법으로 그룹화할 수 있습니다.

최대 가능도, 표준 최소 제곱 및 로지스틱 회귀 방법은 "모형 적합" 시작 창에 지정된 전체 모형을 적합시킵니다. 변수 선택은 수행되지 않습니다. 이러한 모형은 다른 방법과의 비교를 위한 기준으로 사용될 수 있습니다.

참고: 제공된 보고서에 대해 최대 가능도, 표준 최소 제곱 및 로지스틱 회귀 중 하나만 사용할 수 있습니다. 이 추정 방법의 이름은 "모형 적합" 시작 창에 지정된 분포에 따라 다릅니다.

후진 제거, 전진 선택, 가지 치기 전진 선택, 최량 부분집합 및 2단계 전진 선택 방법은 각 단계에서 모형에 추가되거나 제거되는 변수를 기반으로 합니다. 그러나 회귀 계수에 별점을 부과하지 않습니다.

Dantzig 선택기, Lasso 회귀, Elastic Net, 능형 회귀 및 이중 Lasso 회귀 방법은 별점 회귀 기법입니다. 모형의 예측 능력을 향상시키기 위해 회귀 계수 크기를 축소하고 추정값 분산을 줄입니다.

참고: 데이터의 공선성이 높은 경우 Lasso 회귀와 Elastic Net 의 적응형 버전에서 적절한 해를 제공하지 못할 수 있습니다. 적응형 버전은 MLE 가 적절한 추정값을 제공한다고 가정하기 때문입니다. 이러한 경우에는 "적응형" 옵션을 사용하지 않는 것이 좋습니다.

이러한 기법에는 다음과 같은 두 가지 유형의 별점이 사용됩니다.

- l_1 별점 - 회귀 계수의 절대값 합에 별점을 부과합니다.
- l_2 별점 - 회귀 계수의 제곱합에 별점을 부과합니다.

관측 데이터에 대한 기본 추정 방법은 Lasso 회귀입니다. 데이터 테이블에 DOE 스크립트가 포함되어 있고 특이성이 없는 경우 기본 추정 방법은 효과 유전성 옵션이 활성화된 전진 선택입니다. 데이터 테이블에 DOE 스크립트가 포함되어 있고 설계 행렬에 특이성이 있는 경우 기본 추정 방법은 효과 유전성 옵션이 활성화된 2단계 전진 선택입니다.

모형 적합에 다음 방법을 사용할 수 있습니다.

선택 및 별점이 없는 추정 방법

최대 가능도 모형 모수에 대한 MLE(최대 가능도 추정값) 를 계산합니다. 별점이 부과되지 않습니다. 분위수 회귀의 경우 최대 가능도 추정 방법만 사용할 수 있습니다. "모형 적합" 시작 창에서 "검증" 열을 지정한 경우 최대 가능도 모형이 훈련 데이터 집합에 적합됩니다. 다음 조건이 충족되면 최대 가능도 모형 보고서가 기본적으로 나타납니다.

- 예측 변수 사이에 선형 종속성이 없습니다.
- 관측값이 예측 변수보다 더 많습니다.
- 예측 변수가 250 개 이하입니다.

"최대 가능도" 옵션을 사용하면 일반화 회귀 분석법에서 지원하는 반응 분포에 대한 전통적 모형을 생성할 수 있습니다. 또한 최대 가능도 기반 모형을 모형 비교를 위한 기준으로 사용할 수 있습니다.

정규 또는 이항 분포가 지정된 경우 "최대 가능도" 방법을 각각 "표준 최소 제곱" 또는 "로지스틱 회귀" 라고 합니다.

표준 최소 제곱 정규 분포를 지정하면 "최대 가능도" 추정 방법이 "표준 최소 제곱" 추정 방법으로 대체됩니다. 기본 보고서는 일반적인 표준 최소 제곱 결과를 제공하는 "표준 최소 제곱" 보고서입니다.

로지스틱 회귀 이항 분포를 지정하면 "최대 가능도" 추정 방법이 "로지스틱 회귀" 추정 방법으로 대체됩니다. 기본 보고서는 "로지스틱 회귀" 보고서입니다. 로지스틱 결과는 최대 가능도 결과와 동일합니다.

단계별 추정 방법

참고: 다항 분포가 지정된 경우에는 단계별 추정 방법을 사용할 수 없습니다.

후진 제거 후진 제거 회귀를 사용하여 모수 추정값을 계산합니다. 선택한 모형은 선택한 검증 방법과 관련된 최적 해를 제공합니다. 후진 제거는 모형에 모든 모수를 포함하고 모형에 절편만 남을 때까지 각 단계에서 하나의 효과를 제거하는 것으로 시작됩니다. 각 단계마다 각 모수에 대해 Wald 검정을 사용하여 제거할 모수를 결정합니다.

주의: 후진 제거를 위한 해 경로의 가로 축은 다른 추정 방법에서 동일한 축의 반대입니다. 따라서 후진 제거 추정 방법에 대한 해 경로에서 왼쪽에서 오른쪽으로 이동하면 모형에 항이 추가되는 것이 아니라 제거됩니다.

전진 선택 단계별 전진 회귀를 사용하여 모수 추정값을 계산합니다. 각 단계에서 스코어 검정이 가장 유의한 효과가 모형에 추가됩니다. 선택한 모형은 선택한 검증 방법과 관련된 최적 해를 제공합니다.

교호작용이 있고 효과 유전성 옵션이 활성화된 경우 교락 효과는 다음과 같은 방식으로 처리됩니다. 주어진 단계에서 스코어 검정이 가장 유의한 효과가 효과 유전성을 위반하는 경우

교락 효과가 생성됩니다. 교락 효과는 스코어 검정이 가장 유의한 효과 및 효과 유전성을 충족하는 데 필요한 기타 비활성 효과를 포함합니다. 교락 효과에 스코어 검정이 가장 유의한 효과가 있는 경우 교락 효과의 모든 효과가 모형에 추가됩니다.

가지 치기 전진 선택 전진 단계와 후진 단계를 혼합하여 모수 추정값을 계산합니다. 절편만 있는 모형으로 알고리즘이 시작됩니다. 첫 번째 단계에서는 스코어 검정이 가장 유의한 효과가 모형에 추가됩니다. 첫 번째 단계 이후 알고리즘은 각 단계에서 다음과 같은 세 가지 가능성을 고려합니다.

1. 모형에 없는 효과 중 스코어 검정이 가장 유의한 효과를 추가합니다.
2. 모형에 있는 효과 중 Wald 검정이 가장 유의하지 않은 효과를 제거합니다.
3. 한 단계에서 위의 두 작업을 모두 수행합니다.

각 단계에서 수행되는 작업을 선택하기 위해 알고리즘은 지정된 검증 방법을 사용합니다. 예를 들어 BIC 검증 방법이 지정된 경우 알고리즘은 결과 BIC 값이 가장 작은 작업을 선택합니다. 교호작용이 있고 효과 유전성 옵션이 활성화된 경우 교락 효과는 효과 추가에만 고려되고 효과 제거에는 고려되지 않습니다.

모형이 포화 상태가 되면 알고리즘은 후진 단계를 시도하여 모형이 개선되는지 확인합니다. 알고리즘의 최대 단계 수는 모수 수의 5 배입니다. 선택한 모형은 선택한 검증 방법과 관련된 최적 해를 제공합니다.

가지 치기 전진 선택은 단계별 회귀 분석법의 혼합 단계 옵션 대신 사용할 수 있습니다. 그러나 모형에 추가되거나 제거되는 변수를 결정할 때 p 값을 사용하지 않습니다.

팁: 가지 치기 전진 선택 추정 방법에는 조기 중지 옵션을 사용하지 않는 것이 좋습니다.

최량 부분집합 각 단계에서 모형의 활성 효과 수를 늘려 모수 추정값을 계산합니다. 각 단계마다 단계 번호에 지정된 수의 효과를 포함하는 모든 가능 모형 중에서 적절한 모형이 선택됩니다. 해 경로 그림의 가로 축에 있는 값은 모형의 활성 효과 수를 나타냅니다. 단계 0 은 절편만 있는 모형에 해당합니다. 단계 1 은 활성 효과가 하나만 포함된 모형 중 최적 모형에 해당합니다. 단계는 "모형 시작" 보고서의 "고급 제어"에 지정된 "최대 효과 수" 값까지 계속됩니다. 자세한 내용은 "**고급 제어**"(283 페이지)에서 확인하십시오.

팁: 최량 부분집합 추정 방법은 계산량이 많습니다. 따라서 큰 문제에는 권장되지 않습니다.

2 단계 전진 선택 (모형에 2 차 이상 효과가 있는 경우에만 사용 가능) 모수 추정값을 두 단계로 계산합니다. 첫 번째 단계에서는 주효과에 대해 단계별 전진 회귀 모형을 실행하여 모형에 유지할 효과를 결정합니다. 두 번째 단계에서는 첫 번째 단계에서 선택된 주효과로만 구성된 모든 고차 효과에 대해 단계별 전진 회귀 모형을 실행합니다. 이 방법은 강한 효과 유전성을 가정합니다.

첫 번째 단계에서 유지되지 않은 항은 "모수 추정값" 보고서에 0 인 항으로 나타납니다. 그러나 두 번째 단계 모형을 적합시킬 때 무시됩니다. 첫 번째 단계에서 선택된 항은 두 번째 단계에 강제 적용되지 않으며 두 번째 단계에서 선택할 수 있습니다.

벌점 추정 방법

Dantzig 선택기 (정규 분포가 지정되고 "절편 없음" 옵션이 선택되지 않은 경우에만 사용 가능) 선형 프로그래밍 방식으로 l_1 벌점을 적용하여 모수 추정값을 계산합니다. 자세한 내용은 Candes and Tao 연구 자료 (2007) 에서 확인하십시오. Dantzig 선택기는 설계된 실험의 결과를 분석하는 데 유용합니다. 직교 문제의 경우 Dantzig 선택기와 Lasso 회귀에서 동일한 결과를 제공합니다. 자세한 내용은 "**Dantzig 선택기**" (306 페이지) 에서 확인하십시오.

올가미 l_1 벌점을 적용하여 모수 추정값을 계산합니다. l_1 벌점으로 인해 일부 계수가 0으로 추정될 수 있습니다. 따라서 적합 절차의 일부로 변수 선택이 수행됩니다. 일반 Lasso 회귀의 경우 모든 계수에 동일하게 벌점이 적용됩니다.

적응형 Lasso 회귀 회귀 계수 절대값의 가중 합에 벌점을 적용하여 모수 추정값을 계산합니다. l_1 벌점의 가중치는 oracle 특성 (Zou 2006) 을 유지하는 것과 같은 방식으로 데이터에 의해 결정됩니다. 이 옵션은 MLE 를 사용하여 l_1 벌점에 가중치를 적용합니다. 예측 변수의 수가 관측값 수를 초과하거나 예측 변수 사이에 엄격한 선형 종속성이 있으면 MLE 를 계산할 수 없습니다. 회귀 모수의 MLE 를 계산할 수 없는 경우 일반화 역 해 또는 능형 회귀 해가 l_1 벌점 가중치에 사용됩니다. 자세한 내용은 "**적응형 방법**" (307 페이지) 에서 확인하십시오.

일반적으로 "Lasso 회귀" 및 "적응형 Lasso 회귀" 옵션은 예측 변수의 상관관계가 높을 때 간결한 모형을 선택합니다. 이러한 기법은 상관관계가 높은 예측 변수 그룹 중 하나만 선택하는 경향이 있습니다. 고차원 데이터의 예측 변수는 대개 상관관계가 높습니다. 이러한 유형의 데이터에는 Lasso 회귀보다 Elastic Net 을 사용하는 것이 좋습니다. 자세한 내용은 "**Lasso 회귀**" (306 페이지) 에서 확인하십시오.

Elastic Net l_1 벌점과 l_2 벌점을 모두 적용하여 모수 추정값을 계산합니다. l_1 벌점을 적용하면 변수 선택이 수행됩니다. l_2 벌점은 능형 회귀와 같이 계수를 축소하여 예측 능력을 향상시킵니다.

적응형 Elastic Net 적응형 l_1 벌점과 l_2 벌점을 사용하여 모수 추정값을 계산합니다. 이 옵션은 MLE 를 사용하여 l_1 벌점에 가중치를 적용합니다. 예측 변수의 수가 관측값 수를 초과하거나 예측 변수 사이에 엄격한 선형 종속성이 있으면 MLE 를 계산할 수 없습니다. 회귀 모수의 MLE 를 계산할 수 없는 경우 일반화 역 해 또는 능형 회귀 해가 l_1 벌점 가중치에 사용됩니다. "고급 제어" 패널에서 "Elastic Net 알파" 값을 설정할 수 있습니다. 자세한 내용은 "**적응형 방법**" (307 페이지) 에서 확인하십시오.

예측 변수의 상관관계가 높을 때 Elastic Net 의 예측 정확성이 Lasso 회귀보다 높은 경향이 있습니다. 실제로 능형 회귀와 Lasso 회귀 모두 Elastic Net 의 특수한 사례입니다. 예측 능력 면에서는 적응형 Elastic Net 이 Elastic Net 및 적응형 Lasso 회귀를 능가하는 경우가 많습니다. Elastic Net 에는 상관관계가 높은 예측 변수 그룹을 선택하고 관련 예측 변수에 적절한 모수 추정값을 할당하는 기능이 있습니다. 자세한 내용은 "**Elastic Net**" (307 페이지) 에서 확인하십시오.

참고 : "Elastic Net" 적합을 선택하고 "Elastic Net 알파" 를 결측으로 설정하면 알고리즘이 Lasso 회귀, Elastic Net, 능형 회귀 적합을 순서대로 계산합니다. 적합에 시간이 많이 소요되는 경우 진행률 표시줄이 나타납니다. "현재 추정값 채택" 을 클릭하면 계산이 중지되며 보고된 모수 추정값은 해당 지점에서 최적 모형 적합에 해당합니다. 진행률 표시줄은 알고리즘이 각각 Lasso 회귀, Elastic Net 및 능형 회귀를 적합시키고 있음을 나타냅니다. 이 정보를 사용하여 "현재 추정값 채택" 을 클릭할 시기를 결정할 수 있습니다.

능형 회귀 능형 회귀를 사용하여 모수 추정값을 계산합니다. 능형 회귀는 l_2 벌점을 적용하고 모수 추정값이 0 이 되지 않는 편향 회귀 기법입니다. 모형의 모든 예측 변수를 유지하려는 경우에 유용합니다. 자세한 내용은 "**능형 회귀**" (306 페이지) 에서 확인하십시오.

이중 Lasso 회귀 모수 추정값을 두 단계로 계산합니다. 첫 번째 단계에서는 Lasso 회귀 모형을 적합시켜 두 번째 단계에 사용할 항을 결정합니다. 두 번째 단계에서는 첫 번째 단계에서 결정된 항을 사용하여 Lasso 회귀 모형을 적합시킵니다. 표시된 해 경로 결과 및 모수 추정값 보고서는 두 번째 단계 적합에 대한 것입니다. 첫 번째 단계에서 모형에 추가되는 변수가 없으면 두 번째 단계 없이 첫 번째 단계의 결과가 보고서에 나타납니다.

이중 Lasso 회귀는 관측값 수가 예측 변수의 수보다 적을 때 특히 유용합니다. 변수 선택 및 축소 작업을 두 단계로 나누면 두 번째 단계의 Lasso 회귀에서 모형에 포함되어야 할 항에 과도하게 벌점을 부과할 가능성이 적습니다. 이중 Lasso 회귀는 완화 Lasso 회귀와 비슷합니다. 완화 Lasso 회귀는 Hastie et al. 연구 자료 (2009, p. 91) 에 설명되어 있습니다.

적응형 이중 Lasso 회귀 모수 추정값을 두 단계로 계산합니다. 첫 번째 단계에서는 적응형 Lasso 회귀 모형을 적합시켜 두 번째 단계에 사용할 항을 결정합니다. 두 번째 단계에서는 첫 번째 단계에서 결정된 항을 사용하여 적응형 Lasso 회귀 모형을 적합시킵니다. 두 번째 단계에서는 첫 번째 단계 모형에 포함된 항만 고려하고 첫 번째 단계의 모수 추정값을 기반으로 한 가중치를 사용합니다. "고급 제어" 의 "적응형 벌점 가중치" 옵션을 사용하여 가중치 계산 방법을 선택할 수 있습니다. 자세한 내용은 "**고급 제어 옵션**" (284 페이지) 에서 확인하십시오. 표시된 결과는 두 번째 단계 적합에 대한 것입니다. 첫 번째 단계에서 모형에 추가되는 변수가 없으면 두 번째 단계 없이 첫 번째 단계의 결과가 보고서에 나타납니다. 자세한 내용은 "**적응형 방법**" (307 페이지) 에서 확인하십시오.

JMP® 고급 제어

"고급 제어" 옵션을 사용하여 모형 적합 과정의 다양한 요소를 조정할 수 있습니다. 여러 컨트롤이 조정 모수의 격자와 관련이 있습니다.

조정 모수

Lasso 회귀 및 능형 회귀 추정 방법의 해 경로는 단일 조정 모수에 따라 다릅니다. Elastic Net의 해 경로는 가능도 벌점에 대한 조정 모수 및 Elastic Net 알파에 따라 다릅니다. Elastic Net의 가능도에 대한 벌점은 Lasso 회귀 및 능형 회귀 추정 방법과 관련된 벌점의 가중 합입니다. Elastic Net 알파는 이러한 두 벌점의 가중치를 결정합니다. 자세한 내용은 "**추정 방법에 대한 통계 상세 정보**" (306 페이지) 및 "**고급 제어에 대한 통계 상세 정보**" (308 페이지) 에서 확인하십시오.

조정 모수가 0이면 해에 별점이 부과되지 않고 최대 가능도 추정값이 계산됩니다. 조정 모수가 증가하면 별점이 증가합니다.

해는 선택한 검증 방법과 관련하여 별점이 부과된 음의 로그 가능도 함수를 최소화하는 모수 추정값 집합입니다. 현재 해는 해 경로 그림에서 빨간색 수직 실선으로 지정됩니다.

참고 : 해 경로 그림에서 척도화된 모수 추정값 크기가 감소하면 조정 모수 값이 증가합니다. MLE 에 가까운 추정값은 큰 크기와 관련이 있고 별점이 크게 부과된 추정값은 작은 크기와 관련이 있습니다.

다음 사항에 유의하십시오.

- 조정 모수가 너무 작으면 일반적으로 데이터가 과대적합되어 분산이 큰 모형이 생성됩니다.
- 조정 모수가 너무 크면 일반적으로 데이터가 과소적합됩니다.

조정 모수 격자

해를 얻기 위해 미세 격자에서 조정 모수가 증가합니다.

- Lasso 회귀, Elastic Net 알파가 지정된 Elastic Net 및 능형 회귀의 경우 해를 제공하는 조정 모수의 값은 조정 모수 격자에서 최량 적합을 제공하는 값입니다.

참고 : "Elastic Net 알파" 는 기본적으로 0.99 로 설정되어 있습니다.

- "Elastic Net 알파" 에 값을 설정하지 않으면 알파 값도 미세 격자에서 증가합니다. 조정 모수의 값이 고정된 경우 알파는 연속된 10 개의 알파 값이 검증 방법에 의해 결정된 최량 적합을 개선하지 못할 때까지 변경됩니다. 이 과정은 조정 모수 값의 전체 격자에 대해 반복됩니다. 조정 모수 및 알파의 최종 값은 조정 모수 격자에서 최량 적합을 제공하는 값입니다.

조정 모수 값 격자의 범위는 대부분의 경우 0 부터 모든 비절편 항이 0 인 가장 작은 값까지입니다. 모든 비절편 항이 0 인 조정 모수의 가장 작은 값을 상한으로 정의합니다. 조정 모수의 하한은 0.0001 로 설정되는 다음 두 경우를 제외하고 0 입니다.

- 설계 행렬이 특이 행렬인 경우 최대 가능도 추정값을 계산할 수 없습니다. 0.0001 하한을 사용하면 MLE 에 가까운 추정값을 계산할 수 있습니다.
- 이항 또는 다항 분포를 선택한 경우 0.0001 하한을 사용하면 분리를 방지할 수 있습니다.

고급 제어 옵션

효과 유전성 적용 관련 고차 효과보다 저차 효과를 먼저 모형에 입력해야 합니다. 즉, 대부분의 경우 X 가 모형에 없으면 X^2 도 모형에 없다는 의미입니다. 그러나 "전진 선택" 을 제외한 다른 추정 방법의 경우 모형에 X^2 이 추가되고 동일한 단계에서 X 가 제거될 수 있습니다. 이 옵션은 데이터 테이블에 DOE 스크립트가 포함된 경우 활성화되지만 기본적으로 해제되어 있습니다.

Elastic Net 알파 Elastic Net 에 대한 α 모수를 설정합니다. 이 α 모수는 Elastic Net 계수를 추정할 때 l_1 및 l_2 별점 조정 모수의 혼합을 결정합니다. 기본값은 $\alpha = 0.99$ 이며, 이는 l_1 별점의 계

수를 0.99로 설정하고 l_2 벌점의 계수를 0.01로 설정합니다. 이 옵션은 "추정 방법"에 "Elastic Net"을 선택한 경우에만 사용할 수 있습니다. 자세한 내용은 "[추정 방법에 대한 통계 상세 정보](#)"(306 페이지)에서 확인하십시오.

격자 점 수 조정 모수에 대한 하한과 상한 사이의 격자점 수를 지정합니다. 각 격자점 값에서 해당 조정 모수 값에 대한 모수 추정값을 얻습니다. 기본 격자점 수는 150 개입니다.

최소 벌점률 조정 모수의 상한에 대한 하한 비율의 최소값을 나타냅니다. 조정 모수의 하한이 0 이면 해가 MLE 를 제공합니다. MLE 또는 MLE 에 매우 가까운 해를 포함하지 않으려는 경우 "최소 벌점률"을 0 이 아닌 값으로 설정할 수 있습니다. 이중 Lasso 회귀 추정 방법의 경우 이 옵션에 지정된 값은 적합의 첫 번째 단계에서만 사용됩니다. 설계 행렬에 특이성이 있으면 기본값이 0.0001 입니다. 그렇지 않으면 기본값이 0 입니다.

격자 척도 격자 척도의 분포를 선택하는 옵션을 제공합니다. 선형, 제곱근 또는 로그 척도 중에서 선택할 수 있습니다. 선택한 척도에 따라 지정된 격자점 수와 같은 수의 격자점이 조정 모수의 하한과 상한 사이에 분포됩니다. 기본 격자 척도는 제곱근입니다. 자세한 내용은 "[고급 제어에 대한 통계 상세 정보](#)"(308 페이지)에서 확인하십시오.

첫 번째 단계 해 이중 Lasso 회귀 및 2 단계 전진 선택의 첫 번째 단계에서 해를 선택하는 옵션을 제공합니다. 기본적으로 지정된 검증 방법에 따라 최량 적합 해가 선택되고 초기에 표시됩니다(최량 적합). 녹색 또는 노란색 영역에 있는 l_1 노름 값이 더 크거나 더 작은 모형을 초기에 표시하도록 선택할 수 있습니다. 예를 들어 "노란색 영역에서 가장 작음"을 선택할 경우 초기에 표시되는 해는 노란색 영역에서 l_1 노름이 가장 작은 모형입니다. 자세한 내용은 "[비교 가능 모형 영역](#)"(293 페이지)에서 확인하십시오.

최대 효과 수 최량 부분집합 추정 방법에 대해 모형에서 고려할 최대 효과 수를 지정합니다. 이 옵션을 사용하여 모형 적합에 필요한 계산 수를 제한할 수 있습니다. 기본값은 10 입니다.

초기에 표시된 해 "해 경로" 보고서에서 현재 모형으로 초기에 표시되는 해를 선택하는 옵션을 제공합니다. 현재 모형은 수직 실선으로 식별됩니다. 자세한 내용은 "[현재 모형 표시자](#)"(291 페이지)에서 확인하십시오. 최량 적합 해는 수직 점선으로 식별됩니다. 기본적으로 지정된 검증 방법에 따라 최량 적합으로 간주되는 해가 표시됩니다.

녹색 또는 노란색 영역에 있는 l_1 노름 값이 더 크거나 더 작은 모형을 초기에 표시하도록 선택할 수 있습니다. 예를 들어 "노란색 영역에서 가장 작음"을 선택할 경우 초기에 표시되는 해는 노란색 영역에서 l_1 노름이 가장 작은 모형입니다. 자세한 내용은 "[비교 가능 모형 영역](#)"(293 페이지)에서 확인하십시오.

적응형 벌점 가중치 적응형 이중 Lasso 회귀의 두 번째 단계에서 사용되는 벌점 가중치를 계산하는 옵션을 제공합니다. 기본적으로 "역 해" 옵션이 선택되어 있습니다. 이 옵션은 첫 번째 단계 적합의 모수 추정값을 사용하여 벌점 가중치를 계산합니다.

"역 모형 평균" 옵션은 AICc 또는 BIC 모형의 가중 평균인 해의 모수 추정값을 사용하여 벌점 가중치를 계산합니다. AICc 검증 방법을 선택하는 경우 AICc 모형이 사용됩니다. 그렇지 않은 경우 BIC 모형이 사용됩니다. "역 모형 평균" 옵션을 사용하면 최대 가능도 해(있는 경우)가 적응형 이중 Lasso 회귀 모형의 해 경로에서 맨 오른쪽 점으로 나타납니다.

항 강제 적용 모형에 강제로 적용할 항(있는 경우)을 선택할 수 있습니다. 모형에 강제 적용된 항은 벌점에 포함되지 않습니다.

JMP PRO 검증 방법 옵션

다음 방법을 사용하여 모형 적합을 검증할 수 있습니다.

참고 : "분위수 회귀" 분포에는 "없음" 검증 방법 옵션만 허용됩니다. "최대 가능도" 추정 방법에는 "없음" 과 "검증 열" 검증 방법만 허용됩니다. "Cox 비례 위험" 분포에는 "BIC", "AICc" 및 "없음" 검증 방법만 허용됩니다. "Dantzig 선택기" 추정 방법에는 "BIC" 와 "AICc" 검증 방법만 허용됩니다.

K 폴드 조정 모수의 각 값에 대해 다음 단계가 수행됩니다.

- 관측값이 k 개의 부분집합 또는 폴드로 분할됩니다.
- 그러면 각 폴드가 검증 데이터 집합으로 사용됩니다. 모형은 폴드에 없는 관측값에 적합됩니다. 폴드에 있는 관측값에 대해 해당 모형에 기반한 로그 가능도가 계산되어 검증 로그 가능도를 제공합니다.
- k 폴드에 대한 검증 로그 가능도의 평균이 계산됩니다. 이 값은 조정 모수 값에 대한 검증 로그 가능도로 사용됩니다.

최대 검증 로그 가능도를 갖는 조정 모수의 값은 최종 해를 생성하는 데 사용됩니다. 최종 모형을 생성하기 위해 조정 모수의 최적값에 대해 파생된 모든 k 개 모형이 전체 데이터 집합에 적합됩니다. 이 중에서 검증 로그 가능도가 가장 높은 모형이 최종 모형으로 선택됩니다. 이 최종 모형에 사용된 훈련 데이터 집합이 훈련 데이터 집합으로 지정되고 이 모형의 홀드아웃 폴드가 검증 데이터 집합입니다. 이러한 훈련 데이터 집합과 검증 데이터 집합이 최종 해에 대해 보고된 결과 및 그림에 사용됩니다.

홀드백 검증 데이터 집합에 대해 지정된 비율의 데이터를 무작위로 선택하고 데이터의 다른 부분을 사용하여 모형을 적합시킵니다. 최종 해는 검증 데이터 집합에 대한 음의 로그 가능도를 최소화하는 값입니다. 이 방법은 큰 데이터 집합에 유용합니다. 랜덤 선택은 단순 랜덤 표집을 기반으로 할 때보다 더 균형 있는 훈련 데이터 집합 및 검증 데이터 집합을 생성하기 위해 모형 요인에 대한 층화 표집을 기반으로 합니다.

Leave-One-Out leave-one-out 교차 검증을 수행합니다. 이 방법은 폴드 수가 행 수와 같은 K 폴드와 동등합니다. 중간 또는 큰 데이터 집합에는 이 옵션을 사용하면 안 됩니다. 심지어 중간 크기 정도의 관측값 수 일 때도 처리 시간이 오래 걸릴 수 있습니다. 최종 해에 대해 보고된 결과 및 그림에 사용되는 훈련 데이터 집합과 검증 데이터 집합 (한 행) 은 K 폴드 검증에 대해 수행된 대로 결정됩니다.

BIC 해 경로에서 BIC(Bayes 정보 기준)를 최소화합니다. 자세한 내용은 "통계 상세 정보" 부록의 "[가능도, AICc 및 BIC](#)"(514 페이지)에서 확인하십시오.

AICc 해 경로에서 수정 AICc(Akaike 정보 기준)를 최소화합니다. AICc 는 검증 방법의 기본 설정입니다. 자세한 내용은 "통계 상세 정보" 부록의 "[가능도, AICc 및 BIC](#)"(514 페이지)에서 확인하십시오.

참고 : 모수 수가 표본 크기에 근접하거나 초과하는 경우 AICc 가 정의되지 않습니다.

ERIC 해 경로에서 ERIC(확장된 정규화 정보 기준) 를 최소화합니다 . 자세한 내용은 "**모형 적합 상세 정보**"(289 페이지)에서 확인하십시오 . 지수 계열 분포와 Lasso 회귀 및 적응형 Lasso 회귀 추정 방법에만 사용할 수 있습니다 .

없음 검증을 사용하지 않습니다 . 최대 가능도 추정 방법 및 분위수 회귀에만 사용할 수 있습니다 .

검증 열 " 모형 적합 " 창에서 지정한 열을 " 검증 " 역할이 할당된 것으로 사용합니다 . 최종 해는 검증 데이터 집합에 대한 음의 로그 가능도를 최소화하는 값입니다 . Dantzig 선택기 추정 방법이 지정되거나 분위수 회귀 또는 Cox 비례 위험 분포가 지정된 경우에는 이 옵션을 사용할 수 없습니다 .

JMP PRO 초기 중지

초기 중지 옵션은 다음과 같은 초기 중지 규칙을 추가합니다 .

- 전진 선택의 경우 모형에 변수를 추가하여 검증 측도를 개선하지 못하는 단계가 연속으로 10 번이 되면 알고리즘이 종료됩니다 . 해는 10 번의 연속된 단계 앞에 오는 단계의 모형입니다 .
- Lasso 회귀 , Elastic Net 및 능형 회귀의 경우 연속된 10 개의 조정 모수 값이 검증 방법에 의해 결정된 최량 적합을 개선하지 못하면 알고리즘이 종료됩니다 . 해는 10 개의 연속된 값 앞에 오는 조정 모수 값에 해당하는 추정값입니다 .

참고 : AICc 및 BIC 검증 방법의 경우 4 개 이상의 예측 변수가 모형에 입력될 때까지 초기 중지가 발생하지 않습니다 .

JMP PRO 시작

시작을 클릭하면 보고서가 열립니다 . 보고서 제목은 선택한 반응 분포, 추정 방법 및 검증 방법을 지정합니다 . "모형 시작" 제어판으로 돌아가 추가 분석을 수행하고 다른 반응 분포, 추정 방법 및 검증 방법을 선택할 수 있습니다 .

JMP PRO 모형 적합 보고서

" 모형 시작 " 패널에서 지정하는 각 추정 방법 및 검증 방법에 대해 보고서가 생성됩니다 . 보고서는 선택한 분포, 추정 방법 및 검증 방법을 제목에 지정합니다 .

기본적으로 다음 보고서가 제공됩니다 .

- 회귀 그림
- 모형 요약
- 추정 상세 정보 (Lasso 회귀 , Elastic Net 및 능형 회귀에만 표시됨)
- 해 경로 (최대 가능도 추정 방법 및 분위수 회귀를 제외하고 모두 표시됨)
- 원래 예측 변수에 대한 모수 추정값

- 효과 검정

JMP PRO 회귀 그림

참고 : " 회귀 그림 " 보고서는 연속형 예측 변수가 하나이고 범주형 예측 변수가 최대 한 개일 때만 나타납니다. " 분포 " 옵션이 " 다항 ", " 순서형 로지스틱 " 또는 "Cox 비례 위험 " 으로 설정된 경우에는 사용할 수 없습니다. 반응은 연속형이어야 합니다.

"회귀 그림" 보고서는 세로 축에 반응 값이 표시되고 가로 축에 연속형 예측 변수가 표시된 그림을 보여 줍니다. 점 위에 회귀선이 표시됩니다. 모형에 범주형 예측 변수가 있는 경우 범주형 예측 변수의 각 수준에 별도의 회귀선이 있고 그림 옆에 범례가 나타납니다. 반응이 성공 횟수 및 시행 횟수로 지정된 경우에는 성공 횟수를 시행 횟수로 나눈 값이 그림의 세로 축에 표시됩니다.

JMP PRO 모형 요약

" 모형 요약 " 보고서에서는 적합된 모형을 설명하고 적합 자체에 대한 요약 정보를 제공합니다.

JMP PRO 모형 설명 상세 정보

" 모형 요약 " 보고서의 첫 번째 부분에서는 적합된 모형을 설명하는 정보를 제공합니다.

반응 " 모형 적합 " 창에서 "Y" 역할에 할당된 열입니다. 두 개의 열을 사용하여 구간 중도절단을 지정한 경우 두 열 이름이 모두 나열됩니다.

분포 " 모형 적합 " 창에서 선택한 분포입니다. 분위수 회귀의 경우 반응에 대해 지정된 분위수 값도 표시됩니다.

추정 방법 " 모형 시작 " 패널에서 선택한 추정 방법입니다.

검증 방법 " 모형 시작 " 패널에서 선택한 검증 방법입니다.

평균 모형 연결 " 모형 적합 " 창에서 선택한 분포를 기반으로 평균에 대한 모형 연결 함수입니다.

위치 모형 연결 위치 모수에 대한 모형 연결 함수이며 " 모형 적합 " 창에서 "Cauchy" 또는 "t(5)" 분포를 선택한 경우 표시됩니다.

척도 모형 연결 " 모형 적합 " 창에서 선택한 분포를 기반으로 척도 모수에 대한 모형 연결 함수입니다.

확률 모형 연결 " 모형 적합 " 창에서 선택한 분포를 기반으로 확률에 대한 모형 연결 함수입니다.

산포 모형 연결 " 모형 적합 " 창에서 선택한 분포를 기반으로 산포 모수에 대한 모형 연결 함수입니다.

영과잉 모형 연결 " 모형 적합 " 창에서 선택한 분포를 기반으로 영과잉 모수에 대한 모형 연결 함수입니다.

감지 하한 반응 열에 할당된 " 감지 한계 " 열 특성에 지정된 감지 하한입니다.

감지 상한 반응 열에 할당된 "감지 한계" 열 특성에 지정한 감지 상한입니다.

중도절단 열 "모형 적합" 창에서 "중도절단" 역할에 할당된 열입니다.

중도절단 코드 오른쪽 중도절단된 관측값을 지정하는 중도절단 열의 값입니다. 이 값은 "모형 적합" 창의 "중도절단 코드" 목록에 지정된 값입니다.

JMP PRO 모형 적합 상세 정보

"모형 요약" 보고서의 두 번째 부분에서는 모형 적합과 관련된 통계량을 제공합니다. "검증 방법"에서 "홀드백" 또는 "검증 열"을 선택하는 경우 이러한 통계량은 훈련 데이터 집합과 검증 데이터 집합에 대해 별도로 계산됩니다. "검증 방법"에서 "K 폴드" 또는 "Leave-One-Out"을 선택하면 "모형 요약" 보고서의 이 부분을 사용할 수 없습니다.

행 수 행 수입니다.

빈도 합 "모형 적합" 창에서 "빈도" 또는 "가중치" 역할에 할당된 열 값의 합입니다.

참고: -1* 로그 가능도, BIC, AICc 및 ERIC의 경우 값이 작을수록 좋습니다. 자세한 내용은 "통계 상세 정보" 부록의 "[가능도, AICc 및 BIC](#)"(514 페이지)에서 확인하십시오.

-1* 로그 가능도 현재 모형에 대한 가능도 함수의 음의 자연 로그입니다. 자세한 내용은 "통계 상세 정보" 장의 "[가능도, AICc 및 BIC](#)"(514 페이지)에서 확인하십시오.

참고: 분위수 회귀에는 -1* 로그 가능도를 사용할 수 없습니다.

목적 함수 (분위수 회귀에만 사용 가능) 지정된 분위수 회귀 모형을 적합시키기 위해 최소화되는 함수의 값입니다. 최소화되는 함수는 점검-손실 함수입니다.

모수 수 현재 모형에서 0이 아닌 모수의 수입니다.

BIC 다음과 같이 정의되는 Bayes 정보 기준입니다.

$$BIC = -2\text{LogLikelihood} + k\ln(n)$$

자세한 내용은 "통계 상세 정보" 장의 "[가능도, AICc 및 BIC](#)"(514 페이지)에서 확인하십시오.

AICc 다음과 같이 정의되는 수정된 Akaike 정보 기준입니다.

$$AICc = -2\text{LogLikelihood} + 2k + 2k(k+1)/(n-k-1)$$

자세한 내용은 "통계 상세 정보" 장의 "[가능도, AICc 및 BIC](#)"(514 페이지)에서 확인하십시오.

ERIC (지수 계열 분포와 Lasso 회귀 또는 적응형 Lasso 회귀 추정 방법이 지정된 경우에만 사용 가능) 확장된 정규화 정보 기준입니다. 자세한 내용은 Hui et al. 연구 자료 (2015)에서 확인하십시오. ERIC는 다음과 같이 정의됩니다.

$$ERIC = -2\text{LogLikelihood} + (k-2)\ln(n\phi/\lambda)$$

여기서 λ 는 조정 모수의 값이고 ϕ 는 장애 모수입니다.

일반화 R² (분위수 회귀에는 사용 불가능) 일반 회귀 모형에 적용할 수 있는 R² 측도의 확장입니다. 일반화 R²은 적합 모형의 가능도 (L_M)를 절편만 있는 (상수) 모형의 가능도 (L₀)와 비

교합니다. 이 값은 최대값이 1 이 되도록 척도화됩니다. 이항 분포를 제외한 다른 분포의 일반화 R^2 은 다음과 같이 정의됩니다.

$$\text{일반화 } R^2 = 1 - (L_0/L_M)^{2/n}$$

이항 분포가 지정된 경우 일반화 R^2 은 다음과 같이 정의됩니다.

$$\text{일반화 } R^2 = \frac{1 - (L_0/L_M)^{2/n}}{1 - L_0^{2/n}}$$

일반화 R^2 값이 1 이면 완벽한 모형을 나타내고 0 이면 상수 모형과 같음을 나타냅니다. 일반화 R^2 척도는 표준 최소 제곱 설정 시 연속형 정규 반응에 대한 기준 R^2 로 단순화됩니다. 일반화 R^2 를 Nagelkerke 또는 Craig and Uhler R^2 라고도 하는데, 이는 Cox-Snell 유사 R^2 를 정규화한 것입니다. 자세한 내용은 Nagelkerke 연구 자료 (1991) 에서 확인하십시오.

참고 : 정규 분포가 지정되면 일반화 R^2 이 R^2 으로 바뀝니다.

주의 : 서로 다른 반응 분포를 사용하는 모형에 대해 일반화 R^2 값을 비교하면 안 됩니다. 비교는 주어진 반응 분포를 사용하고 절편만 있는 모형을 대상으로 합니다.

R^2 (정규 분포가 지정된 경우에만 사용 가능) 랜덤 오차가 아니라 모형에 기인할 수 있는 반응의 변동 비율을 추정합니다. R^2 값이 1 이면 완벽한 모형을 나타내고 0 이면 상수 모형과 같음을 나타냅니다. R^2 값은 다음과 같이 계산됩니다.

$$1 - \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2}$$

Adj- R^2 (정규 분포가 지정되고 추정 방법에 별점이 포함되지 않은 경우에만 사용 가능) 모형의 모수 수에 대한 R^2 통계량을 조정합니다. Adj- R^2 을 사용하면 모수 수가 다른 모형을 쉽게 비교할 수 있습니다. 계산에 자유도가 사용됩니다. Adj- R^2 값은 다음과 같이 계산됩니다.

$$1 - \frac{\frac{1}{N - (p - 1)} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{\frac{1}{N - 1} \sum_{i=1}^N (Y_i - \bar{Y})^2}$$

여기서 N 은 관측값 수이고 p 는 모수 수입니다.

참고 : 검증 데이터 집합이 있는 경우 조정 R^2 통계량은 훈련 데이터 집합에 대해서만 보고됩니다.

RASE (정규 분포가 지정된 경우에만 사용 가능) RASE(제곱근 평균 제곱 오차)는 현재 모형에서 평균 제곱 예측 오차의 제곱근입니다. 자세한 내용은 "표준 최소 제곱 모형" 장의 "**RASE**"(82 페이지)에서 확인하십시오.

람다 벌점 (Dantzig 선택기, Lasso 회귀, Elastic Net, 능형 회귀 및 이중 Lasso 회귀 추정 방법에만 사용 가능) 현재 모형에 대한 조정 모수 λ 의 값입니다. 자세한 내용은 "**추정 방법에 대한 통계 상세 정보**"(306 페이지)에서 확인하십시오.

JMP PRO 추정 상세 정보

"추정 상세 정보" 보고서에서는 Dantzig 선택기, Lasso 회귀, Elastic Net, 능형 회귀 및 이중 Lasso 회귀 추정 방법에 대한 고급 제어 설정을 보여 줍니다. 이러한 컨트롤에 대한 자세한 내용은 "**고급 제어**"(283 페이지)에서 확인하십시오.

JMP PRO 해 경로

참고 : "해 경로" 보고서는 최대 가능도를 제외한 모든 추정 방법과 분위수 회귀를 제외한 모든 분포에 대해 나타냅니다.

"해 경로" 보고서에서는 다음 두 가지 그림을 보여 줍니다.

- 해 경로 그림 - 추정된 모수의 값을 표시합니다.
- 검증 그림 - 선택한 검증 방법에 해당하는 검증 통계량의 값을 표시합니다.

두 그림에 대한 수평 척도는 척도화된 모수 추정값 크기로 제공됩니다. 이 값은 평균 모형에 대한 척도화된 모수 추정값의 절대값을 합하여 정의되는 l_1 노름입니다. 절편, 산포 모수 및 영과잉 모수에 해당하는 추정값은 l_1 노름 계산에서 제외됩니다. 다음 사항에 유의하십시오.

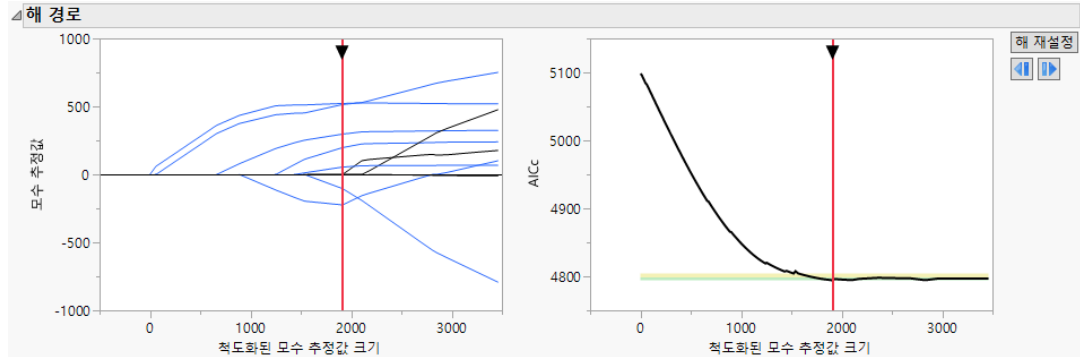
- l_1 노름 값이 큰 추정값은 MLE에 가깝습니다.
- l_1 노름 값이 작은 추정값은 벌점이 크게 부과됩니다.
- l_1 노름이 감소하면 조정 모수 값이 증가합니다.

JMP PRO 현재 모형 표시자

두 그림 모두 "원래 예측 변수에 대한 모수 추정값" 보고서에 표시된 해의 l_1 노름 값에서 빨간색 수직 실선이 추가됩니다. 두 그림 중 하나에서 빨간색 수직선 맨 위의 화살표를 드래그하면 벌점 크기가 변경되어 새 현재 모형을 나타낼 수 있습니다. 검증 그림에서도 그림의 아무 곳이나 클릭하여 모형을 변경할 수 있습니다. 빨간색 수직선을 드래그하여 새 모형을 나타내면 보고서의 결과가 현재 선택된 모형을 반영하도록 업데이트됩니다. 수직 파선은 최량 적합 모형에 유지됩니

다. 검증 그림 옆의 **해 재설정** 버튼을 클릭하면 빨간색 수직선 및 해당 결과가 초기 해로 돌아옵니다. 일부 검증 방법의 경우 비교 가능한 모형을 식별하는 영역이 검증 그림에 제공됩니다. 자세한 내용은 "**비교 가능 모형 영역**"(293 페이지)에서 확인하십시오.

그림 6.5 Diabetes.jmp 에 대한 해 경로 보고서 (AICc 검증을 사용한 Lasso 회귀)



해 경로 그림에 대한 자세한 내용은 "**해 경로 그림**"(292 페이지)에서 확인하십시오. 검증 그림에 대한 자세한 내용은 "**검증 그림**"(293 페이지)에서 확인하십시오.

JMP PRO 해 경로 그림

해 경로 그림에서 경로를 선택하여 "모수 추정값" 보고서에서 해당 항목을 강조 표시할 수 있습니다. 이렇게 하면 데이터 테이블의 해당 열도 선택됩니다. 하나의 보고서에서 행을 선택하면 다른 보고서의 해당 행과 해 경로 그림의 해당 경로가 강조 표시됩니다. 여러 경로 또는 행을 선택하려면 Shift 키를 누르고 클릭합니다.

모수 추정값은 해 경로 그림의 세로 축을 사용하여 그림에 표시됩니다. 이러한 값은 척도화된 모수 추정값입니다. 중심화 및 척도화된 예측 변수로 표현된 모형에 대해 값이 파생됩니다("중심화 및 척도화된 예측 변수에 대한 모수 추정값"(294 페이지) 참조).

예측 변수의 수가 관측값 수보다 적으면 일반적으로 해 경로 그림에 추정값의 전체 범위(0 ~ MLE에 의해 주어진 벌점 없는 적합)가 표시됩니다. 그렇지 않으면 벌점이 부과되지 않은 해에 가까운 크기로 그림이 확장됩니다. 이 현상은 끝에서 두 번째 격자점에서 MLE 해까지의 이동이 너무 커서 끝에서 두 번째 격자점까지의 해에 대한 상세 정보가 분명하지 않을 때 발생합니다. 이 경우 MLE가 최종 해가 아니면 축이 끝에서 두 번째 격자점까지만 확장되도록 해 경로 그림의 척도가 재조정됩니다.

JMP PRO 해 ID

내부적으로 해 경로의 각 해에는 해 ID가 할당됩니다. 초기에 제공된 해가 아닌 다른 해를 선택하기 위해 조정 모수를 변경하면 해당 해 ID가 "스크립트 저장" 옵션으로 생성된 스크립트에 나타납니다. 해 ID는 해 ID 설정(N) 명령의 N 값입니다. 해 ID를 저장하면 스크립트를 실행할 때 선택한 해를 다시 생성할 수 있습니다.

JMP PRO 검증 그림

검증 그림에는 전체 조정 모수 값 또는 척도화된 모수 추정값 크기 값에 대한 모형의 적합 능력을 나타내는 통계량 그림이 표시됩니다. 표시된 통계량은 선택한 검증 방법에 따라 다릅니다. 표 6.3에는 각 검증 방법에 대해 표시되는 통계량이 나열되어 있습니다. 모든 검증 방법에서 값이 작을수록 좋습니다. K 폴드 및 Leave-One-Out 검증 방법과 값이 세 개 이상인 검증 열의 경우 표시되는 통계량은 전체 폴드에서 척도화된 음의 로그 가능도 값의 평균입니다.

표 6.3에서 척도화된 -1*로그 가능도는 음의 로그 가능도를 음의 로그 가능도가 계산되는 집합의 관측값 수로 나눈 값입니다.

표 6.3 검증 방법과 해당 검증 통계량 및 영역

검증 방법	검증 통계량	조정 모수 영역
K 폴드	전체 K 폴드에 대해 척도화된 -1*로그 가능도 값의 평균	두 개
홀드백	척도화된 -1*로그 가능도	없음
Leave-One-Out	모든 폴드에 대해 척도화된 -1*로그 가능도 값의 평균	두 개
BIC	훈련 데이터의 BIC	두 개
AICc	훈련 데이터의 AICc	두 개
ERIC	훈련 데이터의 ERIC	두 개
두 개 또는 세 개의 값이 있는 검증 열	척도화된 -1*로그 가능도	없음
$K > 3$ 개의 값이 있는 검증 열	전체 K 폴드에 대해 척도화된 -1*로그 가능도 값의 평균	두 개

JMP PRO 비교 가능 모형 영역

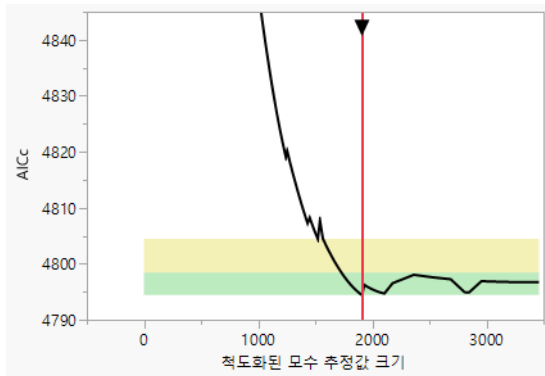
모형이 최적 모형으로 추정되지만 이 선택과 관련하여 불확실성이 존재할 수 있습니다. 경쟁 모형도 거의 적합할 수 있으며 유용한 정보를 포함할 수 있습니다. AICc, BIC, K 폴드 및 Leave-One-Out 검증 방법과 4 개 이상의 값이 있는 검증 열의 경우 고려할 만한 경쟁 모형을 식별하는 영역이 검증 그림에 제공됩니다. 이 영역을 벗어나는 모형은 권장되지 않습니다. 자세한 내용은 Burnham and Anderson 연구 자료(2004) 및 Burnham et al. 연구 자료(2011)에서 확인하십시오.

영역은 검증 통계량 값의 구간입니다. 이러한 영역은 전체 가로 축에 걸쳐 녹색 또는 노란색 직사각형으로 표시됩니다. 검증 통계량 값이 영역에 포함되면 모형이 해당 영역에 포함됩니다. 빨간

색 수직 실선을 드래그하여 영역 내의 해를 탐색할 수 있습니다. 자세한 내용은 "[현재 모형 표시자](#)"(291 페이지)에서 확인하십시오.

그림 6.6에서는 두 영역을 표시하기 위해 세로 축이 확장된 Diabetes.jmp에 대한 검증 그림을 보여 줍니다.

그림 6.6 Diabetes.jmp에 대한 검증 그림 (AICc 검증을 사용한 Lasso 회귀)



BIC, AICc 및 ERIC 검증에 대한 영역

이러한 검증 방법의 경우 두 영역이 그림에 표시됩니다. 최적 해를 위한 검증 BIC, AICc 및 ERIC 값을 V^{best} 로 나타냅니다.

- 녹색 영역은 모형이 최적 모형과 유사하다는 강력한 증거가 있는 모형을 식별합니다. 녹색 영역 구간은 $[V^{best}, V^{best+4}]$ 입니다.
- 노란색 영역은 모형이 최적 모형과 유사하다는 약한 증거가 있는 모형을 식별합니다. 노란색 영역 구간은 $(V^{best+4}, V^{best+10}]$ 입니다.

K 폴드 검증, Leave-One-Out 검증 및 4 개 이상의 값이 있는 검증 열에 대한 영역

이러한 검증 방법의 경우 두 영역이 그림에 표시됩니다. 최적 모형을 위한 해에서 각 검증 데이터 집합에 대해 척도화된 음의 로그 가능도 함수가 실행됩니다. 이러한 값의 표준 오차를 L^{SE} 로 나타냅니다. 최적 해에 대한 척도화된 음의 로그 가능도는 L^{best} 로 나타냅니다.

- 녹색 영역은 모형이 최적 모형과 유사하다는 강력한 증거가 있는 모형을 식별합니다. 녹색 영역 구간은 $[L^{best}, L^{best+L^{SE}}]$ 입니다.
- 노란색 영역은 모형이 최적 모형과 유사하다는 약한 증거가 있는 모형을 식별합니다. 노란색 영역 구간은 $(L^{best+L^{SE}}, L^{best+2.5*L^{SE}}]$ 입니다.

JMP PRO 중심화 및 척도화된 예측 변수에 대한 모수 추정값

"중심화 및 척도화된 예측 변수에 대한 모수 추정값" 보고서에서는 모형의 모든 모수에 대한 추정값 및 기타 결과를 제공합니다. 초기 테이블에는 모형의 예측 변수에 대한 계수가 포함됩니다.

추가 테이블에는 척도, 산포 또는 영과잉 모수와 같은 다른 모형 모수가 포함됩니다. 자세한 내용은 "분포"(270 페이지)에서 확인하십시오. 두 테이블 모두 동일한 결과 열을 포함합니다.

팁: "중심화 및 척도화된 예측 변수에 대한 모수 추정값" 보고서에서 항을 클릭하여 해 경로 그림에서 해당 경로를 강조 표시할 수 있습니다. 데이터 테이블의 해당 열도 선택됩니다. 이 방법은 추가 분석을 실행할 때 유용합니다. 여러 행을 선택하려면 Shift 키를 누르고 항을 클릭합니다.

일반화 회귀 분석법의 모든 적합에 대해 모든 예측 변수는 평균 0 으로 중심화되고 제곱합이 1 과 같도록 척도화됩니다.

- 각 관측값에서 평균을 뺍니다.
- 그런 다음 각 차이를 평균과의 차이 제곱합의 제곱근으로 나눕니다.

이렇게 하면 적용된 벌점과 관련하여 모든 예측 변수가 동등한 관계가 됩니다.

참고: 시작 창에서 "절편 없음" 옵션을 선택하면 예측 변수가 중심화 및 척도화되지 않습니다.

"중심화 및 척도화된 예측 변수에 대한 모수 추정값" 보고서에는 중심화 및 척도화된 예측 변수로 표현된 모형의 모수 추정값이 제공됩니다. 추정값은 지정한 검증 방법에 의해 결정됩니다. 이 추정값은 해 경로 그림에 빨간 수직선으로 표시됩니다.

이 보고서는 다음 정보를 제공합니다.

항 모형 항 목록입니다. "고급 제어" 옵션을 사용하여 모형에 강제로 적용한 항 옆에는 "강제 적용됨" 이 표시됩니다.

추정값 중심화 및 척도화된 모형 항에 해당하는 모수 추정값입니다.

표준 오차 추정값의 표준 오차입니다. 이 값은 M- 추정 및 샌드위치 계산식 (Zou 2006, Huber and Ronchetti 2009) 을 사용하여 구합니다.

Wald 카이제곱 모수가 0 인지 여부에 대한 Wald 검정의 카이제곱 값입니다.

Prob > ChiSquare Wald 검정의 p 값입니다.

95% 하한 모수에 대한 95% 신뢰 구간의 하한입니다. "모형 적합" 창의 "모형 규격" 빨간색 삼각형 메뉴에서 "유의 수준 설정" 을 선택하여 α 수준을 변경할 수 있습니다.

95% 상한 모수에 대한 95% 신뢰 구간의 상한입니다. "모형 적합" 창의 "모형 규격" 빨간색 삼각형 메뉴에서 "유의 수준 설정" 을 선택하여 α 수준을 변경할 수 있습니다.

특이성 상세 정보 (모형 항 사이에 선형 종속성이 있는 경우에만 사용 가능) 모형 항이 충족하는 선형 함수입니다.

JMP PRO 원래 예측 변수에 대한 모수 추정값

"원래 예측 변수에 대한 모수 추정값" 보고서에서는 모형의 모든 모수에 대한 추정값 및 기타 결과를 제공합니다. 초기 테이블에는 모형의 예측 변수에 대한 계수가 포함됩니다. 추가 테이블에

는 척도, 산포 또는 영과잉 모수와 같은 다른 모형 모수가 포함됩니다. 자세한 내용은 "**분포**"(270 페이지)에서 확인하십시오. 두 테이블 모두 동일한 결과 열을 포함합니다.

팁 : "원래 예측 변수에 대한 모수 추정값" 보고서에서 항을 클릭하여 해 경로 그림에서 해당 경로를 강조 표시할 수 있습니다. 데이터 테이블의 해당 열도 선택됩니다. 이 방법은 추가 분석을 실행할 때 유용합니다. 여러 행을 선택하려면 Shift 키를 누르고 항을 클릭합니다.

"원래 예측 변수에 대한 모수 추정값" 보고서에는 원래(중심화 및 척도화되지 않음) 예측 변수로 표현된 모형의 모수 추정값이 제공됩니다.

이 보고서는 다음 정보를 제공합니다.

항 모형 항 목록입니다. "고급 제어" 옵션을 사용하여 모형에 강제로 적용한 항 옆에는 "강제 적용됨"이 표시됩니다.

추정값 원래 측정값으로 표현된 주어진 모형 항에 해당하는 모수 추정값입니다.

표준 오차 추정값의 표준 오차입니다. 이 값은 M-추정 및 샌드위치 계산식 (Zou 2006, Huber and Ronchetti 2009)을 사용하여 구합니다.

Wald 카이제곱 모수가 0 인지 여부에 대한 Wald 검정의 카이제곱 값입니다.

Prob > ChiSquare Wald 검정의 p 값입니다.

95% 하한 모수에 대한 95% 신뢰 구간의 하한입니다. "모형 적합" 창의 "모형 규격" 빨간색 삼각형 메뉴에서 "유의 수준 설정"을 선택하여 α 수준을 변경할 수 있습니다.

95% 상한 모수에 대한 95% 신뢰 구간의 상한입니다. "모형 적합" 창의 "모형 규격" 빨간색 삼각형 메뉴에서 "유의 수준 설정"을 선택하여 α 수준을 변경할 수 있습니다.

특이성 상세 정보 (모형 항 사이에 선형 종속성이 있는 경우에만 사용 가능) 모형 항이 충족하는 선형 함수입니다.

VIF (정규 분포인 경우에만 사용 가능하며, 모수 추정값 테이블에서 마우스 오른쪽 버튼을 클릭하고 열 > VIF를 선택한 경우에만 표시됨) 모형의 각 항에 대한 VIF(분산 팽창 계수)입니다. VIF 값이 높으면 모형의 항 사이에 공선성 문제가 있음을 나타냅니다.

i^{th} 항 x_i 의 VIF는 다음과 같이 계산됩니다.

$$VIF_i = \frac{1}{1 - R_i^2}$$

여기서 R_i^2 은 다른 설명 변수의 함수로 나타낸 x_i 회귀에 대한 R^2 입니다.

JMP® PRO 활성 모수 추정값

"활성 모수 추정값" 보고서에서는 "원래 예측 변수에 대한 모수 추정값" 보고서의 일부를 제공합니다. "활성 모수 추정값" 보고서에는 0이 아닌 모수만 표시됩니다.

JMP PRO 효과 검정

"효과 검정" 보고서에서는 다음 정보를 제공합니다.

소스 모형의 효과 목록입니다.

모수 수 효과와 관련된 모수의 수입니다.

DF Wald 카이제곱 검정의 자유도입니다. 이 값은 모형의 효과와 관련된 0 이 아닌 모수 추정값의 수입니다.

Wald 카이제곱 효과와 관련된 모든 모수가 0인지 여부에 대한 Wald 검정의 카이제곱 값입니다.

Prob > ChiSquare Wald 카이제곱 검정의 p 값입니다.

정규 분포를 사용하고 추정 방법에 별점이 없으면 보고서에 "Wald 카이제곱" 과 "Prob > ChiSquare" 대신 다음 열이 나타납니다.

제공함 효과가 0 이라는 가설을 위한 제공함입니다.

F 비 효과가 0 인지 검정하기 위한 F 통계량입니다. "F 비" 는 효과의 평균 제공을 오차 평균 제공으로 나눈 비율입니다. 오차의 평균 제공은 효과 제공함을 자유도로 나눈 값입니다.

Prob > F 효과 검정의 p 값입니다.

효과 계수가 0으로 추정된 경우 다음 사항에 유의하십시오.

- 효과의 자유도가 1 이면 해당 효과에 대한 행의 맨 오른쪽에 "제거" 라고 표시됩니다.
- 효과의 자유도가 1 을 초과하는 경우 "제거된 수준" 이라고 표시되고 그 뒤에 모수 추정값이 0 인 항에 해당하는 수준 수가 나옵니다.

JMP PRO 모형 적합 옵션

각 모형 적합 보고서에는 다음 옵션이 포함된 빨간색 삼각형 메뉴가 있습니다.

주의 : 시작 창에서 표현식 데이터 유형 또는 벡터 모델링 유형이 있는 열을 지정하면 플랫폼의 많은 옵션을 사용할 수 없습니다.

회귀 보고서 지정된 모형 적합에 대해 표시되는 보고서를 사용자 정의할 수 있습니다. 다음 보고서 중 "중심화 및 척도화된 예측 변수에 대한 모수 추정값" 보고서와 "활성 모수 추정값" 보고서를 제외한 모든 보고서가 기본적으로 표시됩니다.

모형 요약 모형에 대한 규격 및 적합도 통계량 정보를 포함하는 "모형 요약" 보고서를 표시하거나 숨깁니다. 이 옵션은 적용 가능한 모형에 대한 "추정 상세 정보" 보고서도 표시합니다. 자세한 내용은 "모형 요약"(288 페이지) 및 "추정 상세 정보"(291 페이지) 에서 확인하십시오.

해 경로 (최대 가능도 모형에는 사용 불가능) 해 경로 및 검증 경로 그림을 표시하거나 숨깁니다. 자세한 내용은 "해 경로"(291 페이지) 에서 확인하십시오.

중심화 및 척도화된 예측 변수에 대한 모수 추정값 중심화 및 척도화된 모수 추정값 테이블을 표시하거나 숨깁니다. 자세한 내용은 "[중심화 및 척도화된 예측 변수에 대한 모수 추정값](#)"(294 페이지)에서 확인하십시오.

원래 예측 변수에 대한 모수 추정값 데이터의 원래 척도로 표현된 모수 추정값 테이블을 표시하거나 숨깁니다. 자세한 내용은 "[원래 예측 변수에 대한 모수 추정값](#)"(295 페이지)에서 확인하십시오.

활성 모수 추정값 (최대 가능도 또는 능형 회귀 모형에는 사용 불가능) 현재 선택된 모형에 대한 활성 또는 0이 아닌 모수 추정값 테이블을 표시하거나 숨깁니다.

해 경로 요약 표시 (최대 가능도 또는 능형 회귀 모형에는 사용 불가능) 해 경로 그림과 검증 경로 그림에서 활성 집합이 변경되는 점에 대한 적합 통계량 테이블이 포함된 보고서를 표시하거나 숨깁니다. 사용 가능한 통계량은 추정 방법에 따라 다릅니다. BIC 검증을 사용한 정규 Lasso 회귀 모형에 사용할 수 있는 조건부 모형 확률에 대한 자세한 내용은 Hu et al. 연구 자료 (2019)에서 확인하십시오. BIC, AICc 또는 ERIC 검증을 지정하면 BIC 및 AICc 열의 셀에 검증 그림과 동일한 방식으로 색상이 적용됩니다. 자세한 내용은 "[비교 가능 모형 영역](#)"(293 페이지)에서 확인하십시오.

효과 검정 각 효과에 대한 검정을 표시하거나 숨깁니다. 주어진 효과에 대한 효과 검정은 해당 효과와 관련된 모든 모수가 0이라는 귀무가설을 검정합니다. 명목형 또는 순서형 효과에는 수준 수에 따라 관련된 모수가 여러 개 있을 수 있습니다. 이러한 효과에 대한 효과 검정은 관련된 모든 모수가 0인지 여부를 검정합니다. 다항 분포를 지정한 경우에는 전체 반응 수준에 대해 효과가 결합됩니다. 자세한 내용은 "[효과 검정](#)"(297 페이지)에서 확인하십시오.

예측 표현식 표시 추정된 모형에 대한 방정식을 포함하는 "예측 표현식" 보고서를 표시하거나 숨깁니다. 예는 "표준 최소 제곱 모형" 장의 "[예측 표현식 표시](#)"(107 페이지)에서 확인하십시오.

0이 아닌 항 선택 보고서에서 계수가 0이 아닌 항을 강조 표시합니다. 또한 데이터 테이블에서 관련된 모든 열을 선택합니다. "추정 방법"에서 "능형 회귀"를 선택한 경우에는 이 옵션을 사용할 수 없습니다.

0인 항 선택 보고서에서 계수가 0인 항을 강조 표시합니다. 또한 데이터 테이블에서 관련된 모든 열을 선택합니다. "추정 방법"에서 "능형 회귀"를 선택한 경우에는 이 옵션을 사용할 수 없습니다.

활성 집합 다시 시작 (벡터 모델링 유형이 있는 예측 변수가 포함된 모형에는 사용 불가능) 모수 추정값이 0이 아닌 항을 기반으로 한 항 집합이 "모형 효과 생성" 목록에 포함된 "모형 적합" 시작 창을 여는 옵션이 포함되어 있습니다. 이러한 항은 활성 효과입니다. 시작 창의 다른 모든 규격은 원래 분석에 사용된 것과 같습니다.

참고: 기준 변수가 포함된 보고서에서 "활성 집합 다시 시작" 옵션 중 하나를 선택하는 경우 기준 변수는 "모형 적합" 시작 창에 추가되지 않습니다.

활성 효과를 사용하여 다시 시작 활성 효과만 사용하여 "모형 효과 생성" 목록을 채웁니다.

활성 주효과 및 2차 요인 적합 다시 시작 활성 효과를 사용하여 생성된 2차 요인으로 "모형 효과 생성" 목록을 채웁니다.

활성 주효과 및 3차 요인 적합 다시 시작 활성 효과를 사용하여 생성된 3차 요인으로 "모형 효과 생성" 목록을 채웁니다.

활성 주효과 및 완전 요인 적합 다시 시작 활성 효과를 사용하여 생성된 완전 요인으로 "모형 효과 생성" 목록을 채웁니다.

활성 주효과 및 2차 다항식 적합 다시 시작 활성 효과를 사용하여 생성된 2차 다항식으로 "모형 효과 생성" 목록을 채웁니다.

활성 주효과 및 3차 다항식 적합 다시 시작 활성 효과를 사용하여 생성된 3차 다항식으로 "모형 효과 생성" 목록을 채웁니다.

활성 주효과 및 반응 표면 모형 다시 시작 활성 효과를 사용하여 생성된 반응 표면 모형으로 "모형 효과 생성" 목록을 채웁니다.

비활성 경로 숨기기 현재 활성화되지 않은 경로는 흐리게 나타나도록 해 경로 모수 추정값 그림에서 비활성 경로의 투명도를 조정합니다.

승산비 (이항 분포가 지정되고 모형에 절편이 포함된 경우에만 사용 가능하며, 벡터 모델링 유형이 있는 예측 변수가 포함된 모형에는 사용 불가능) 범주형 예측 변수에 대한 승산비와 연속형 예측 변수에 대한 단위 승산비 및 범위 승산비가 포함된 보고서를 표시합니다. 승산비는 두 사건에 대한 승산 비율입니다. 한 사건의 승산은 관심 사건이 발생할 확률 대 발생하지 않을 확률의 비율입니다. 관심 사건은 "모형 적합" 시작 창의 "목표 수준"에 정의됩니다. 각 범주형 예측 변수에 대해 승산비 보고서가 나타납니다. 범주형 모형 항의 모든 수준 조합에 대해 승산비가 표시됩니다.

연속형 예측 변수가 있는 경우 다음과 같은 두 개의 추가 보고서가 나타납니다.

- 단위 승산비 보고서. 단위 승산비는 연속형 모형 항의 1 단위 변화에 대해 계산됩니다.
- 범위 승산비 보고서. 범위 승산비는 연속형 모형 항의 전체 범위에 대해 계산됩니다.

승산비 보고서의 신뢰 구간은 Wald 기반 구간입니다. 모형 항에 대한 승산비는 해당 모형 항이 고차 효과에 포함되지 않은 경우에만 의미가 있습니다.

참고: 모형에 교호작용이 있으면 "다중 비교" 옵션을 사용하여 승산비를 구할 수 있습니다. 자세한 내용은 "[다중 비교](#)"(300 페이지)에서 확인하십시오.

발생률 비율 (Poisson 또는 음이항 분포가 지정되고 모형에 절편이 포함된 경우에만 사용 가능) 범주형 예측 변수에 대한 발생률 비율과 연속형 예측 변수에 대한 단위 발생률 비율 및 범위 발생률 비율이 포함된 보고서를 표시합니다. 발생률 비율은 두 사건에 대한 발생률 비율입니다. 모형 항의 발생률은 주어진 기간 동안 새로 발생하는 사건 수입니다.

각 범주형 예측 변수에 대해 발생률 비율 보고서가 나타납니다. 범주형 모형 항의 모든 수준 조합에 대해 발생률 비율이 표시됩니다.

연속형 예측 변수가 있는 경우 다음과 같은 두 개의 추가 보고서가 나타납니다.

- 단위 발생률 비율 보고서. 단위 발생률 비율은 연속형 모형 항의 1 단위 변화에 대해 계산됩니다.
- 범위 발생률 비율 보고서. 범위 발생률 비율은 연속형 모형 항의 전체 범위에 대해 계산됩니다.

발생률 비율 보고서의 신뢰 구간은 Wald 기반 구간입니다. 모형 항에 대한 발생률 비율은 해당 모형 항이 고차 효과에 포함되지 않은 경우에만 의미가 있습니다.

위험 비율 (Cox 비례 위험 분포가 지정된 경우에만 사용 가능) 범주형 예측 변수에 대한 위험 비율과 연속형 예측 변수에 대한 단위 위험 비율 및 범위 위험 비율이 포함된 보고서를 표시합니다. 위험 비율은 두 사건에 대한 위험 비율입니다. 시간 t 에서 사건의 위험 비율은 사건이 시간 t 까지 생존했다고 가정할 때 추가 시간 동안 생존하지 못할 조건부 확률입니다.

각 범주형 예측 변수에 대해 위험 비율 보고서가 나타납니다. 범주형 모형 항의 모든 수준 조합에 대해 위험 비율이 표시됩니다.

연속형 예측 변수가 있는 경우 다음과 같은 두 개의 추가 보고서가 나타납니다.

- 단위 위험 비율 보고서. 단위 위험 비율은 연속형 모형 항의 1 단위 변화에 대해 계산됩니다.
- 범위 위험 비율 보고서. 범위 위험 비율은 연속형 모형 항의 전체 범위에 대해 계산됩니다.

위험 비율 보고서의 신뢰 구간은 Wald 기반 구간입니다. 모형 항에 대한 위험 비율은 해당 모형 항이 고차 효과에 포함되지 않은 경우에만 의미가 있습니다.

추정값 공분산 모수 추정값의 공분산을 보여 주는 행렬을 표시합니다. 이러한 값은 M- 추정 및 샌드위치 계산식 (Zou 2006, Huber and Ronchetti 2009) 을 사용하여 계산됩니다. 공분산 행렬은 0 인 항을 포함하지 않습니다.

추정값 상관계수 모수 추정값의 상관계수를 보여 주는 행렬을 표시합니다. 이러한 값은 M- 추정 및 샌드위치 계산식 (Zou 2006, Huber and Ronchetti 2009) 을 사용하여 계산됩니다. 상관 행렬은 0 인 항을 포함하지 않습니다.

역추정 예측 (벡터 모델링 유형이 있는 예측 변수가 포함된 모형에는 사용 불가능) Y 및 다른 X 변수에 대한 특정 값이 주어졌을 때 X 값을 예측합니다. 이 옵션은 연속형 변수를 예측하는 데만 사용할 수 있습니다. 역추정 예측에 대한 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 역추정 예측 "(131 페이지) 에서 확인하십시오.

다중 비교 (벡터 모델링 유형이 있는 예측 변수가 포함된 모형 또는 범주형 예측 변수가 포함되지 않은 모형에는 사용 불가능) " 다중 비교 " 시작 창을 표시합니다. " 다중 비교 " 시작 창과 보고서에 대한 자세한 내용은 " 표준 최소 제곱 모형 " 장의 " 다중 비교 "(117 페이지) 에서 확인하십시오. 다중 비교는 선형 예측 변수 척도에서 수행됩니다. 이항 분포가 지정되면 승산비에 대해 다중 비교가 수행됩니다. Poisson 분포가 지정되면 발생률 비율에 대해 다중 비교가 수행됩니다. Cox 비례 위험 분포가 지정되면 위험 비율에 대해 다중 비교가 수행됩니다.

혼동 행렬 (이항, 다항 또는 순서형 로지스틱 분포가 지정된 경우에만 사용 가능) 실제 반응 수준과 예측 반응 수준을 테이블 형식으로 생성하는 행렬을 표시합니다. 적합한 모형은 예측 반응 수준이 실제 반응 수준과 동일해야 합니다. 혼동 행렬을 사용하면 예측 반응이 실제 반응에 잘 맞는지 평가할 수 있습니다. 검증을 사용한 경우 훈련 데이터 집합, 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 혼동 행렬이 표시됩니다.

확률 임계 설정 (이항 분포가 지정된 경우에만 사용 가능) 반응을 분류하기 위한 경계 확률을 지정합니다. 기본적으로 관측값은 예측 확률이 0.5를 초과할 경우 목표 수준으로 분류됩니다. 임계값을 변경하여 목표 수준으로 분류하기 위한 경계를 0.5 이외의 값으로 지정합니다. 혼동 행렬의 예측 비율은 지정된 임계값에 따른 분류를 반영하도록 업데이트됩니다.

반응에 "수의 행렬" 열 특성이 있는 경우 수의 행렬에 의해 확률 임계 초기값이 결정됩니다.

프로파일러 (벡터 모델링 유형이 있는 예측 변수가 포함된 모형에는 사용 불가능) 적합 모형을 탐색할 수 있는 다양한 프로파일러를 제공합니다.

참고: 행 수가 500 보다 작거나 같고 예측 변수의 수가 30 보다 작거나 같은 경우 해 경로 그림에서 현재 모형 표시자를 드래그하면 프로파일러 그림이 계속 업데이트됩니다. 그렇지 않은 경우에는 마우스 버튼을 놓으면 그림이 업데이트됩니다.

프로파일러 예측 프로파일러를 표시합니다. 모수 추정값이 0 이고, 계수가 0 이 아닌 교호작용 항에 포함되지 않은 예측 변수는 프로파일러에 나타나지 않습니다. 예측 프로파일러에 대한 자세한 내용은 **Profilers** 의 에서 확인하십시오.

분포 프로파일러 (이항 또는 분위수 회귀 분포가 지정된 경우에는 사용 불가능) 예측 변수 및 반응의 누적 분포 함수 프로파일러를 표시합니다. 반응은 맨 오른쪽 셀에 표시됩니다.

분위수 프로파일러 (이항 또는 분위수 회귀 분포가 지정된 경우에는 사용 불가능) 예측 반응을 예측 변수의 함수 및 누적 분포 함수의 분위수로 보여 주는 프로파일러를 표시합니다. 분위수는 "확률" 이라고 하며 맨 오른쪽 셀에 표시됩니다.

생존 프로파일러 (정규, 지수, Weibull, 로그 정규 또는 Cox 비례 위험 분포가 지정된 경우에만 사용 가능) 생존 함수를 예측 변수 및 반응의 함수로 보여 주는 프로파일러를 표시합니다. 반응은 맨 오른쪽 셀에 표시됩니다.

위험 프로파일러 (정규, 지수, Weibull, 로그 정규 또는 Cox 비례 위험 분포가 지정된 경우에만 사용 가능) 위험 비율을 예측 변수 및 반응의 함수로 보여 주는 프로파일러를 표시합니다. 반응은 맨 오른쪽 셀에 표시됩니다.

사용자 검정 사용자 가설을 검정할 수 있는 "사용자 검정" 보고서를 표시합니다. 모형에 해 경로가 있는 경우 해를 업데이트하면 사용자 검정 결과가 업데이트됩니다. 사용자 검정에 대한 자세한 내용은 "표준 최소 제곱 모형" 장의 "**사용자 검정**"(115 페이지) 에서 확인하십시오. "사용자 검정" 의 빨간색 삼각형 메뉴에는 사용자 검정 보고서를 제거하는 옵션이 포함되어 있습니다.

진단 그림 현재 모형이 얼마나 적합한지 평가하는 데 도움이 되는 다양한 그림을 제공합니다. 검증 열을 지정하거나 "검증 방법" 에서 "K 폴드", "홀드백" 또는 "Leave-One-Out" 을 선택한 경우 아래 옵션을 사용하여 훈련 데이터 집합, 검증 데이터 집합 및 테스트 데이터 집합 (해당하는 경우) 을 보거나, 이러한 데이터 집합에 대해 별도의 그림을 생성할 수 있습니다. "K 폴드" 또는 "Leave-One-Out" 을 선택한 경우 그림은 예측 오차를 최적화하는 검증 데이터 집합 및 이에 대한 훈련 데이터 집합에 해당합니다. 자세한 내용은 "**K 폴드**"(286 페이지) 에서 확인하십시오.

참고 : 해 경로 그림에서 현재 모형 표시자를 드래그하면 모든 진단 그림이 계속 업데이트됩니다.

진단 번들 (이항, 다항, 순서형 로지스틱 또는 Cox 비례 위험 분포가 지정된 경우에는 사용 불가능) 잔차 대 예측값 그림, 잔차 대 행 번호 그림, 잔차 히스토그램 및 관측된 반응보다 더 큰 반응이 관측될 확률 히스토그램을 포함하는 네 개의 그래프를 표시합니다.

모든 관측값을 사용하여 그래프가 생성됩니다. 검증 열을 사용하거나 "검증 방법"에서 "K 폴드", "홀드백" 또는 "Leave-One-Out"을 선택한 경우 체크박스를 사용하여 훈련 데이터 집합, 검증 데이터 집합 및 테스트 데이터 집합(해당하는 경우)을 선택할 수 있습니다. 데이터 테이블에서 이러한 데이터 집합에 해당하는 행이 선택되고 그래프에서 해당 점과 영역이 강조 표시됩니다. 이 옵션을 사용하여 모형 적합이 전체 데이터 집합에서 유사한지 여부를 판별할 수 있습니다.

"더 큰 반응이 관측될 적합 확률" 히스토그램은 모형의 적합도를 평가하는 데 도움이 됩니다. 다음과 같이 분포에 따라 다른 기준이 적용됩니다.

- 영과잉 분포와 분위수 회귀를 제외한 다른 분포의 경우 "올바른" 모형은 거의 균등한 값 분포를 표시해야 합니다.
- 영과잉 분포의 경우 히스토그램은 0에 점질량을 표시하고 다른 곳에서는 거의 균등한 분포를 표시해야 합니다.
- 분위수 회귀의 경우 히스토그램은 지정된 분위수 왼쪽에는 거의 균등한 값 분포를 표시하고 지정된 분위수 오른쪽에는 약간 더 높은 값을 거의 균등한 분포로 표시해야 합니다.

기준 생존 및 위험 그림 (Cox 비례 위험 분포가 지정된 경우에만 사용 가능) 반응 변수에 대한 기준 비례 위험 함수의 생존 및 위험 함수를 보여 주는 "기준 생존 및 위험" 그림을 표시합니다. 아래에는 그림에 표시된 값이 포함된 테이블이 있습니다.

참고 : Cox 비례 위험 분포가 지정된 경우 "진단 그림"에 "기준 생존 및 위험 그림" 옵션만 있습니다.

ROC 곡선 (이항, 다항 또는 순서형 로지스틱 분포가 지정된 경우에만 사용 가능) ROC(Receiver Operating Characteristic) 곡선을 표시합니다. 검증을 사용한 경우 훈련 데이터 집합, 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 ROC 곡선이 표시됩니다.

ROC 곡선은 반응 수준을 올바르게 분류하기 위한 적합 확률 능력을 측정합니다. 주대각과 곡선의 거리가 멀수록 적합도가 높습니다. ROC 곡선에 대한 소개는 기본 분석의 에서 확인하십시오.

반응 수준이 세 개 이상이면 ROC 곡선 그림에 각 반응 수준에 대한 ROC 곡선이 표시됩니다. 주어진 반응 수준에 대해 이 곡선은 해당 수준으로 올바르게 분류하기 위한 ROC 곡선입니다. ROC 곡선에 대한 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오.

Lift 곡선 (이항, 다항 또는 순서형 로지스틱 분포가 지정된 경우에만 사용 가능) 모형에 대한 Lift 곡선을 표시합니다. 검증을 사용한 경우 훈련 데이터 집합, 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 Lift 곡선이 표시됩니다.

Lift 곡선은 적합 확률이 감소할 때 반응 수준이 얼마나 효과적으로 분류되는지 보여 줍니다. 적합 확률은 가로 축을 따라 내림차순으로 표시됩니다. 적합 확률의 수직 좌표는 해당 확률 이상에 대한 올바른 분류 비율을 전체 올바른 분류 비율로 나눈 값입니다. 적합 확률이 임계값을 초과하는 관측값만 선택한 경우 Lift 곡선을 사용하여 대부분의 관측값을 올바르게 분류할 수 있는지 확인할 수 있습니다.

반응 수준이 세 개 이상이면 Lift 곡선 그림에 각 반응 수준에 대한 Lift 곡선이 표시됩니다. 주어진 반응 수준에 대해 이 곡선은 해당 수준으로 올바르게 분류하기 위한 Lift 곡선입니다. Lift 곡선에 대한 자세한 내용은 **Predictive and Specialized Modeling** 의 에서 확인하십시오.

실제값 대 예측값 그림 (이항, 다항, 순서형 로지스틱 또는 Cox 비례 위험 분포가 지정된 경우에는 사용 불가능) 세로 축에 실제 Y 값을 표시하고 가로 축에 예측 Y 값을 표시합니다. 검증을 사용한 경우 훈련 데이터 집합, 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 그림이 표시됩니다.

잔차 대 예측값 그림 (이항, 다항, 순서형 로지스틱 또는 Cox 비례 위험 분포가 지정된 경우에는 사용 불가능) 세로 축에 잔차를 표시하고 가로 축에 예측 Y 값을 표시합니다. 검증을 사용한 경우 훈련 데이터 집합, 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 그림이 표시됩니다.

예측 변수별 잔차 도표화 (이항, 다항, 순서형 로지스틱 또는 Cox 비례 위험 분포가 지정되거나, 벡터 모델링 유형이 있는 예측 변수가 포함된 모형에는 사용 불가능) 모형의 각 예측 변수에 대해 세로 축에 잔차를 표시하고 가로 축에 예측 변수 값을 표시합니다. 모형의 예측 변수에 대해 각각 하나씩 그림이 있습니다. 검증을 사용한 경우 훈련 데이터 집합, 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 그림이 표시됩니다.

정규 분위수 그림 (정규 분포가 지정되고 중도절단이 없는 경우에만 사용 가능) 세로 축에 정규 분위수를 표시하고 가로 축에 표준화 잔차를 표시합니다. 검증을 사용한 경우 훈련 데이터 집합, 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 그림이 표시됩니다.

열 저장 적합 모형을 기반으로 한 열을 데이터 테이블에 저장할 수 있습니다. "분포"에서 "Cox 비례 위험"을 선택한 경우 사용할 수 있는 옵션에 대한 내용은 **"Cox 비례 위험 모형의 열 저장 옵션"**(305 페이지)에서 확인하십시오. 다른 모든 분포의 경우 다음 열을 데이터 테이블에 저장할 수 있습니다.

함수 예측 계산식 저장 (반응 열에 "FDE FPC 수" 열 특성이 포함된 경우에만 사용 가능) 각 함수 주성분에 대한 예측 계산식이 포함된 각 FDE 주성분 반응 열을 데이터 테이블에 저장합니다. 이 옵션은 함수 데이터 탐색기 플랫폼의 고유 함수 열과 예측 계산식의 선형 조합인 모형 예측 계산식도 추가합니다. 모형 예측 계산식을 사용하여 원래 반응을 프로파일링할 수 있는 스크립트가 데이터 테이블에 추가됩니다. 반응은 FDE 주성분 반응 열의 "FDE 출력" 열 특성에 지정됩니다. 함수 주성분에 대한 자세한 내용은 **Predictive and Specialized Modeling** 의 에서 확인하십시오.

참고 : " 함수 예측 계산식 저장 " 옵션은 보고서 창에서 모든 FDE 주성분 반응에 대한 계산식 열을 저장합니다. 단일 반응에 여러 모형이 적합된 경우 각 반응의 최종 모형을 사용하여 해당 반응에 대한 예측 계산식을 생성합니다.

예측 계산식 저장 관측된 (표준화되지 않음) 데이터 값이 주어졌을 때 예측 계산식이 포함된 열을 데이터 테이블에 저장합니다. 예측 계산식은 0 인 항을 포함하지 않습니다. 평균 계산식에 대한 자세한 내용은 "[분포에 대한 통계 상세 정보](#)"(309페이지)에서 확인하십시오.

반응 열이 범주형인 경우 이 옵션은 각 반응 수준에 대한 확률 열과 가장 확률이 높은 반응이 포함된 열을 생성합니다. 가장 확률이 높은 반응 열에는 모형에 따른 확률이 가장 높은 수준이 포함됩니다. 확률 임계값이 0.5 가 아닌 경우 이 옵션은 확률 임계값을 기반으로 가장 확률이 높은 반응을 포함하는 추가 열을 생성합니다.

평균 신뢰 구간 평균 반응의 95% 신뢰 하한 및 상한이 포함된 두 열을 데이터 테이블에 저장합니다.

참고 : " 모형 적합 " 창의 " 모형 규격 " 빨간색 삼각형 메뉴에서 " 유의 수준 설정 " 을 선택하여 신뢰 구간의 α 수준을 변경할 수 있습니다.

예측값의 표준 오차 예측 평균 반응의 표준 오차가 포함된 열을 데이터 테이블에 저장합니다.

잔차 계산식 저장 Y에서 예측 계산식을 뺀 형식으로 제공되는 잔차의 계산식이 포함된 열을 데이터 테이블에 저장합니다. 잔차 계산식은 0 인 항을 포함하지 않습니다. " 분포 " 에서 " 이항 " 을 선택한 경우에는 사용할 수 없습니다.

분산 계산식 저장 예측 분산에 대한 계산식이 포함된 열을 데이터 테이블에 저장합니다. 예측 분산은 선택한 분포의 분산 계산식을 사용하여 계산됩니다. 연결 함수에 관련된 모수의 값은 추정된 선형 성분에 역 연결 함수를 적용하여 추정됩니다. 다른 모수는 모수 추정값으로 대체됩니다. 분산 계산식에 대한 자세한 내용은 "[분포에 대한 통계 상세 정보](#)"(309페이지)에서 확인하십시오. " 분포 " 에서 " 이항 " 을 선택한 경우에는 사용할 수 없습니다.

선형 예측 변수 저장 설계 행렬과 모수 추정값 벡터의 곱을 구하는 계산식이 포함된 열을 데이터 테이블에 저장합니다. 일반적으로 이것을 $X\beta$ 라고 합니다. 이 계산식은 0 인 항을 포함하지 않습니다.

검증 열 저장 (K 폴드, 홀드백 또는 Leave-One-Out 검증 방법이 지정된 경우에만 사용 가능) 폴드에 행 할당을 설명하는 열을 저장합니다. K 폴드의 경우 행이 할당된 폴드가 열에 나열됩니다. 홀드백의 경우 각 행은 훈련 데이터 집합 또는 검증 데이터 집합에 속하는 것으로 식별됩니다. Leave-One-Out 의 경우 행 값은 해당 행이 제외되는 순서를 나타냅니다.

참고 : 시작 창에서 검증 열을 선택한 경우 " 검증 열 저장 " 옵션이 나타나지 않습니다.

분포 계산식 저장 (이항 또는 분위수 회귀 분포가 지정된 경우에는 사용 불가능) 누적 분포 함수에 대한 계산식이 포함된 열을 데이터 테이블에 저장합니다.

생존 계산식 저장 (연속형 분포가 지정된 경우에만 사용 가능) 관측된 시간에서의 생존 확률에 대한 계산식이 포함된 열을 데이터 테이블에 저장합니다. 생존 함수는 1 에서 누적 분포 함수를 뺀 값과 같습니다.

시뮬레이션 계산식 저장 적합 모형에 대한 추정 모수를 사용하여 시뮬레이션 값을 생성하는 계산식이 포함된 열을 데이터 테이블에 저장합니다. 이 열은 "시뮬레이션" 유틸리티에서 "스위치 인할 열"로 사용할 수 있습니다. 자세한 내용은 기본 분석의 에서 확인하십시오.

Cook 거리 영향력 (정규 분포 및 표준 최소 제곱 추정 방법이 지정된 경우에만 사용 가능) Cook 거리 영향력 통계량의 값이 포함된 열을 데이터 테이블에 저장합니다.

해트 값 (정규 분포 및 표준 최소 제곱 추정 방법이 지정된 경우에만 사용 가능) $X(X'X)^{-1}X'$ 의 주대각 요소가 포함된 열을 데이터 테이블에 저장합니다. 이러한 값을 **hat 값**이라고도 합니다.

예측 계산식 게시 예측 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다. "계산식 저장소" 보고서가 열려 있지 않으면 "계산식 저장소" 보고서가 생성됩니다. 자세한 내용은 Predictive and Specialized Modeling 의 에서 확인하십시오.

Cox 비례 위험 모형의 열 저장 옵션

생존 계산식 저장 관측된 시간에서의 생존 확률에 대한 계산식이 포함된 열을 데이터 테이블에 저장합니다.

Cox Snell 잔차 계산식 저장 Cox-Snell 잔차에 대한 계산식이 포함된 열을 데이터 테이블에 저장합니다. Cox-Snell 잔차는 양수입니다. Cox-Snell 잔차에 대한 자세한 내용은 Meeker and Escobar 연구 자료 (1998, sec. 17.6.1) 에서 확인하십시오.

Martingale 잔차 계산식 저장 Martingale 잔차에 대한 계산식이 포함된 열을 데이터 테이블에 저장합니다. Martingale 잔차는 개인에 대해 관측된 사건 수와 조건부 기대 사건 수 간의 차이로 정의됩니다. Martingale 잔차의 평균은 0 이고 범위는 음의 무한대와 1 사이입니다. 자세한 내용은 Fleming and Harrington 연구 자료 (1991) 에서 확인하십시오.

선형 예측 변수 저장 설계 행렬과 모수 추정값 벡터의 곱을 구하는 계산식이 포함된 열을 데이터 테이블에 저장합니다. 일반적으로 이것을 $X\beta$ 라고 합니다. 이 계산식은 0 인 항을 포함하지 않습니다.

적합 제거 적합 보고서를 제거합니다.

일반화 회귀 분석법에 대한 통계 상세 정보

- "추정 방법에 대한 통계 상세 정보"
- "고급 제어에 대한 통계 상세 정보"
- "분포에 대한 통계 상세 정보"

JMP[®] PRO 추정 방법에 대한 통계 상세 정보

별점 회귀 방법은 별점을 적용하여 회귀 계수에 편향을 도입합니다.

JMP[®] PRO 능형 회귀

능형 회귀에서 회귀 계수에 l_2 별점이 적용됩니다. 능형 회귀 계수 추정값은 다음과 같이 제공됩니다.

$$\hat{\beta}^{ridge} = \operatorname{argmin}_{\beta} \left\{ \sum_{i=1}^N -\operatorname{LogLikelihood}(\beta; y_i) + \frac{\lambda}{2} \sum_{j=1}^p \beta_j^2 \right\},$$

여기서 $\sum_{j=1}^p \beta_j^2$ 은 l_2 별점, λ 는 조정 모수, N 은 행 수, p 는 변수 수입니다.

JMP[®] PRO Dantzig 선택기

Dantzig 선택기에서 회귀 계수에 l_{∞} 별점이 적용됩니다. Dantzig 선택기에 대한 계수 추정값은 다음 기준을 충족합니다.

$$\min_{\beta} \|X^T(y - X\beta)\|_{\infty} \quad \text{subject to } \|\beta\|_1 \leq t$$

여기서 $\|v\|_{\infty}$ 는 벡터 v 성분의 최대 절대값인 l_{∞} 노름을 나타냅니다.

JMP[®] PRO Lasso 회귀

Lasso 회귀에서 회귀 계수에 l_1 별점이 적용됩니다. Lasso 회귀의 계수 추정값은 다음과 같이 제공됩니다.

$$\hat{\beta}^{lasso} = \operatorname{argmin}_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^N -\operatorname{LogLikelihood}(\beta; y_i) + \lambda \sum_{j=1}^p |\beta_j| \right\},$$

여기서 $\sum_{j=1}^p |\beta_j|$ 는 l_1 벌점, λ 는 조정 모수, N 은 행 수, p 는 변수 수입니다.

JMP PRO Elastic Net

Elastic Net 은 l_1 벌점과 l_2 벌점을 결합합니다. Elastic Net 의 계수 추정값은 다음과 같이 제공됩니다.

$$\hat{\beta}^{enet} = \operatorname{argmin}_{\beta} \left\{ \sum_{i=1}^N -\operatorname{LogLikelihood}(\beta; y_i) + \lambda \sum_{j=1}^p \left(\alpha |\beta_j| + \frac{(1-\alpha)}{2} \beta_j^2 \right) \right\},$$

다음은 방정식에 사용되는 표기입니다.

- $\sum_{j=1}^p |\beta_j|$ - l_1 벌점
- $\sum_{j=1}^p \beta_j^2$ - l_2 벌점
- λ - 조정 모수
- α - l_1 벌점과 l_2 벌점의 결합을 결정하는 모수
- N - 행 수
- p - 변수 수

팁 : Elastic Net 에서 단일 예측 변수에 대한 가변 α 및 λ 의 축소 효과를 보여 주는 두 개의 샘플 스크립트가 있습니다. **도움말 > 샘플 데이터**를 선택하고 **샘플 스크립트 디렉터리 열기**를 클릭한 후 `demoElasticNetAlphaLambda.jsl` 또는 `demoElasticNetAlphaLambda2.jsl` 을 선택합니다. 각 스크립트에는 스크립트 사용 방법 및 목적에 대한 설명이 포함되어 있습니다.

JMP PRO 적응형 방법

적응형 Lasso 회귀 방법은 가중 벌점을 사용하여 일관된 계수 추정값을 제공합니다. l_1 벌점의 가중 형식은 다음과 같습니다.

$$\sum_{j=1}^p \frac{|\beta_j|}{\tilde{\beta}_j},$$

여기서 $\tilde{\beta}_j$ 는 MLE 입니다 (MLE 가 있는 경우). MLE 가 없고 반응이 정규 분포인 경우 최소 제곱을 사용하여 추정이 수행되고 $\tilde{\beta}_j$ 는 일반화 역행렬을 사용하여 구한 해입니다. 반응이 정규 분포가 아닌 경우 $\tilde{\beta}_j$ 는 능형 회귀 해입니다.

적응형 Lasso 회귀의 경우 이 가중 l_1 벌점을 사용하여 $\hat{\beta}^{lasso}$ 계수를 결정합니다.

적응형 Elastic Net 은 이 가중 l_1 벌점을 사용하고 가중 l_2 벌점도 부과합니다. 적응형 Elastic Net 의 가중 l_2 벌점은 다음과 같습니다.

$$\sum_{j=1}^p \left(\frac{\beta_j}{\tilde{\beta}_j} \right)^2,$$

여기서 $\tilde{\beta}_j$ 는 MLE 입니다 (MLE 가 있는 경우). MLE 가 없고 반응이 정규 분포인 경우 최소 제곱을 사용하여 추정이 수행되고 $\tilde{\beta}_j$ 는 일반화 역행렬을 사용하여 구한 해입니다. 반응이 정규 분포가 아닌 경우 $\tilde{\beta}_j$ 는 능형 회귀 해입니다.

JMP^{PRO} 고급 제어에 대한 통계 상세 정보

JMP^{PRO} 격자

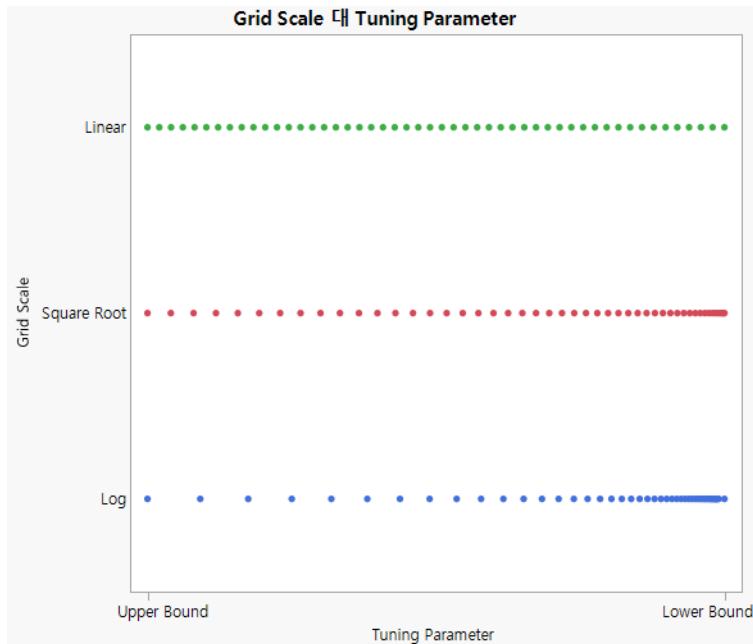
능형 회귀 및 Lasso 회귀에서 벌점 가능도를 가장 잘 최소화하는 조정 모수는 조정 모수 값의 격자를 검색하여 찾습니다. 이 값 격자는 조정 모수의 하한과 상한 사이에 있습니다. "고급 제어" 에서 격자점 수를 지정할 수 있습니다.

하한은 0.0001로 설정되는 특수한 경우 외에는 0입니다. 자세한 내용은 "[조정 모수](#)"(283 페이지) 에서 확인하십시오. 조정 모수의 하한이 0이면 해에 벌점이 부과되지 않고 계수가 MLE입니다. 상한은 모든 비절편 항이 0인 가장 작은 값입니다.

하한과 상한 사이의 값 격자를 반복 검색하여 최적의 조정 모수 값을 결정합니다. 가능한 조정 모수 격자를 설정할 때 선형, 로그 및 제곱근이라는 세 가지 척도를 사용할 수 있습니다. 기본 격자 척도는 제곱근입니다.

벌점이 부과되지 않은 추정값과 이전 단계 사이의 차이가 큰 경우가 있습니다. 이러한 큰 차이로 인해 해 경로가 왜곡될 수 있습니다. 로그 척도는 큰 값이 거의 없는 작은 조정 모수 값을 검색하는데 중점을 두는 반면, 선형 척도는 최소값에서 최대값까지 검색을 균등하게 분산시킵니다. 제곱근 척도는 다른 두 척도를 절충한 방법입니다. 그림 6.7에서는 다양한 격자 척도를 보여 줍니다.

그림 6.7 조정 모수 격자 척도 옵션



JMP^{PRO} 분포에 대한 통계 상세 정보

아래에서는 일반화 회귀 분석법에 의한 분포 적합을 모형 적합에 사용된 모수를 기준으로 구분하여 보여 줍니다. 설명을 통해 구체적으로 언급되어 있지는 않지만 일반화 회귀 분석법을 사용하면 이산형 분포에 대해 정수가 아닌 값을 지정할 수 있습니다.

JMP^{PRO} 연속형 분포

정규 분포

$$f(y|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right], -\infty < y < \infty$$

$$E(Y) = \mu$$

$$Var(Y) = \sigma^2$$

Cauchy 분포

$$f(y|\mu, \sigma) = \left\{ \pi\sigma \left[1 + \left(\frac{y-\mu}{\sigma} \right)^2 \right] \right\}^{-1}, -\infty < y < \infty$$

$$E(Y) = \text{정의되지 않음}$$

$$Var(Y) = \text{정의되지 않음}$$

t(5) 분포

$$f(y|\mu, \sigma) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} \frac{1}{\sqrt{v\pi\sigma^2}} \left[1 + \frac{(x-\mu)^2}{v\sigma^2}\right]^{-\left(\frac{v+1}{2}\right)}, -\infty < y < \infty, v = 5$$

$$E(Y) = \mu$$

$$Var(Y) = \sigma^2 \frac{v}{v-2}$$

지수 분포

$$f(y|\theta) = \frac{1}{\mu} \exp\left[-\frac{y}{\mu}\right], y > 0$$

$$E(Y) = \mu$$

$$Var(Y) = \mu^2$$

감마 분포

$$f(y|\mu, \sigma) = \frac{y^{(\mu/\sigma)-1} \exp[-y/\sigma]}{\Gamma[\mu/\sigma] \sigma^{\mu/\sigma}}, y > 0$$

$$E(Y) = \mu$$

$$Var(Y) = \mu\sigma$$

Weibull 분포

$$f(y|\mu, \sigma) = \frac{1}{y\sigma} \exp\left[\frac{\log(y)-\mu}{\sigma}\right] \exp\left\{-\exp\left[\frac{\log(y)-\mu}{\sigma}\right]\right\}, y > 0$$

$$E(Y) = \exp(\mu)\Gamma[1+\sigma]$$

$$Var(Y) = \exp(2\mu)\{\Gamma[1+2\sigma] - (\Gamma[1+\sigma])^2\}$$

로그 정규 분포

$$f(y|\mu, \sigma) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}[\log(y) - \mu]^2\right\}, y > 0$$

$$E(Y) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

$$Var(Y) = [\exp(\sigma^2) - 1] \exp(2\mu + \sigma^2)$$

베타 분포

$$f(y|\mu, \sigma) = \frac{\Gamma[1/\sigma]}{\Gamma[\mu/\sigma]\Gamma[(1-\mu)/\sigma]} y^{\mu/\sigma-1} (1-y)^{\frac{(1-\mu)}{\sigma}-1}, y \in (0, 1)$$

$$E(Y) = \mu$$

$$Var(Y) = \mu(1-\mu) \frac{\sigma}{\sigma+1}$$

JMP PRO 이산형 분포

이항 분포

$$f(y|n, p) = \binom{n}{y} p^y (1-p)^{n-y}, y = 0, 1, 2, \dots, n$$

$$E(Y) = np$$

$$Var(Y) = np(1-p)$$

베타 이항 분포

$$f(y|n, p, \delta) = \binom{n}{y} \frac{\Gamma\left[\frac{1}{\delta}-1\right] \Gamma\left[y+p\left(\frac{1}{\delta}-1\right)\right] \Gamma\left[n-y+(1-p)\left(\frac{1}{\delta}-1\right)\right]}{\Gamma\left[p\left(\frac{1}{\delta}-1\right)\right] \Gamma\left[(1-p)\left(\frac{1}{\delta}-1\right)\right] \Gamma\left[n-1+\frac{1}{\delta}\right]}, y = 0, 1, 2, \dots, n$$

$$E(Y) = np$$

$$Var(Y) = np(1-p)[1 + (n-1)\delta]$$

Poisson 분포

$$f(y|\lambda) = \frac{\lambda^y}{y!} \exp(-\lambda), y = 0, 1, 2, \dots$$

$$E(Y) = \lambda$$

$$Var(Y) = \lambda$$

음이항 분포

$$f(y|\mu, \sigma) = \frac{\Gamma[y + (1/\sigma)]}{\Gamma[y + 1]\Gamma[1/\sigma]} \left[\frac{(\mu\sigma)^y}{(1 + \mu\sigma)^{y + (1/\sigma)}} \right], y = 0, 1, 2, \dots$$

$$E(Y) = \mu$$

$$Var(Y) = \mu + \sigma\mu^2$$

JMP PRO **영과잉 분포****영과잉 이항 분포**

$$f(y|n, p, \pi) = \begin{cases} \pi + (1 - \pi)(1 - p)^n, & \text{for } y = 0 \\ (1 - \pi) \binom{n}{y} p^y (1 - p)^{n-y}, & \text{for } y = 1, 2, \dots, n \end{cases}$$

$$E(Y) = np(1 - \pi)$$

$$Var(Y) = (1 - \pi)np(1 - p) + n^2p^2(1 - \pi)^2$$

영과잉 베타 이항 분포

$$f(y|n, p, \delta, \pi) = \begin{cases} \pi + (1 - \pi) \frac{\Gamma[\frac{1}{\delta} - 1] \Gamma[p(\frac{1}{\delta} - 1)] \Gamma[n + (1 - p)(\frac{1}{\delta} - 1)]}{\Gamma[p(\frac{1}{\delta} - 1)] \Gamma[(1 - p)(\frac{1}{\delta} - 1)] \Gamma[n - 1 + \frac{1}{\delta}]}, & \text{for } y = 0 \\ (1 - \pi) \binom{n}{y} \frac{\Gamma[\frac{1}{\delta} - 1] \Gamma[y + p(\frac{1}{\delta} - 1)] \Gamma[n - y + (1 - p)(\frac{1}{\delta} - 1)]}{\Gamma[p(\frac{1}{\delta} - 1)] \Gamma[(1 - p)(\frac{1}{\delta} - 1)] \Gamma[n - 1 + \frac{1}{\delta}]}, & y = 1, 2, \dots, n \end{cases}$$

$$E(Y) = np(1 - \pi)$$

$$Var(Y) = (1 - \pi)np\{1 + (1 - p)[1 + (n - 1)\delta]\} - [np(1 - \pi)]^2$$

영과잉 Poisson 분포

$$f(y|\lambda, \pi) = \begin{cases} \pi + (1 - \pi)\exp[-\lambda] \quad \text{단, } y = 0 \\ (1 - \pi)\frac{\lambda^y}{y!}\exp[-\lambda], \text{ for } y = 1, 2, \dots \end{cases}$$

$$E(Y) = (1 - \pi)\lambda$$

$$Var(Y) = \lambda(1 - \pi)(1 + \lambda\pi)$$

영과잉 음이항 분포

$$f(y|\mu, \sigma, \pi) = \begin{cases} \pi + (1 - \pi)(1 + \mu\sigma)^{-(1/\sigma)}, \text{ for } y = 0 \\ (1 - \pi)\frac{\Gamma[y + (1/\sigma)]}{\Gamma[y + 1]\Gamma[1/\sigma]}\left[\frac{(\mu\sigma)^y}{(1 + \mu\sigma)^{y + (1/\sigma)}}\right], \text{ for } y = 1, 2, \dots \end{cases}$$

$$E(Y) = (1 - \pi)\mu$$

$$Var(Y) = \mu(1 - \pi)[1 + \mu(\sigma + \pi)]$$

영과잉 감마 분포

$$f(y|\mu, \sigma, \pi) = \begin{cases} \pi, \text{ for } y = 0 \\ (1 - \pi)\frac{\exp(-y/\sigma)}{\Gamma[\mu/\sigma]\sigma^{\mu/\sigma}y^{1 - \mu/\sigma}}, \text{ for } y > 0 \end{cases}$$

$$E(Y) = \mu(1 - \pi)$$

$$Var(Y) = \mu(1 - \pi)(\sigma + \mu) - (1 - \pi)^2\mu^2$$

7 장

일반화 회귀 예 정규화 기법을 사용하여 모형 생성

이 장에서는 일반화 회귀 플랫폼을 사용하는 몇 가지 모형에 대해 지침 자료를 이용한 예를 제공합니다.

목차

Poisson 일반화 회귀 예.....	317
이항 일반화 회귀 예.....	319
영과잉 Poisson 회귀 예.....	321
가로형 데이터에 대한 일반화 회귀 예.....	323

Poisson 일반화 회귀 예

Liver Cancer.jmp 샘플 데이터 테이블에는 136명의 환자에 대한 간암 Node Count(결절 수) 값이 포함되어 있습니다. 또한 BMI, Age(연령), Time(시간), Markers(표지자), Hepatitis(간염) 및 Jaundice(황달)라는 6가지 잠재적 관련 변수에 대한 측정값도 포함됩니다. 이러한 변수는 데이터 테이블의 각 열에서 "노트" 열 특성에 설명되어 있습니다. "노트" 열 특성을 보려면 열 이름을 마우스 오른쪽 버튼으로 클릭하고 "열 정보"를 선택한 후 "열 특성" 아래의 "노트"를 선택합니다.

이 예에서는 6개의 예측 변수를 사용하여 Node Count에 대한 예측 모형을 개발합니다. Node Count는 Poisson 분포를 사용하여 모델링됩니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Liver Cancer.jmp를 엽니다.

2. **분석 > 모형 적합**을 선택합니다.

3. "열 선택" 목록에서 Node Count를 선택하고 **Y**를 클릭합니다.

4. BMI ~ Jaundice를 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다.

이렇게 하면 차수 2(차수 상자의 기본값)까지의 모든 항이 모형에 추가됩니다.

5. "열 선택" 목록에서 Validation을 선택하고 **검증**을 클릭합니다.

6. "분석법" 목록에서 **일반화 회귀**를 선택합니다.

7. "분포" 목록에서 **Poisson**을 선택합니다.

8. **실행**을 클릭합니다.

"모형 비교" 보고서, "모형 시작" 제어판 및 "검증 열을 사용한 Poisson 최대 가능도" 보고서가 포함된 "일반화 회귀" 보고서가 나타납니다. 기본 추정 방법은 "Lasso 회귀"입니다.

9. **적응형** 상자를 선택합니다.

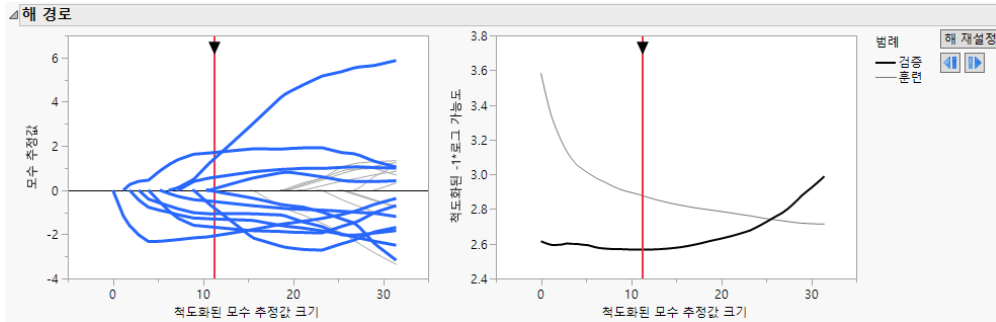
10. **시작**을 클릭합니다.

11. "검증 열을 사용한 Poisson 적응형 Lasso 회귀" 옆의 빨간색 삼각형을 클릭하고 **0이 아닌 항 선택**을 선택합니다.

그림 7.1에는 해 경로가 나와 있습니다. 계수가 0이 아닌 항의 경로가 강조 표시됩니다. 해가 MLE에서 멀어질수록 해 경로가 그림의 오른쪽에서 왼쪽으로 움직인다고 생각할 수 있습니다. 여러 항에 상당히 일찍 0으로 축소하는 경로가 있습니다.

해 경로 그림의 세로 축은 표준화된 예측 변수의 모수 추정값을 나타냅니다. 빨간색 수직선은 교차 검증에 의해 결정된 최적 축소에서의 값을 나타냅니다. 이 점에서 11개의 항의 계수가 0이 아닙니다. 빨간색 수직선은 검증 데이터 집합의 최소 척도화된 $-1 \times$ 로그 가능도 값을 나타냅니다.

그림 7.1 0 이 아닌 항이 강조 표시된 적응형 Lasso 회귀 적합의 해 경로



"원래 예측 변수에 대한 모수 추정값" 보고서 (그림 7.2)에서는 비중심화 및 비척도화 데이터에 대한 모수 추정값을 보여 줍니다. 모수 추정값이 0이 아닌 11개 항이 강조 표시됩니다. 여기에 교호작용 효과가 포함되어 있습니다. 모든 예측 변수 열이 계수가 0이 아닌 항에 나타나므로 데이터 테이블에서 6개의 모든 예측 변수 열이 선택됩니다.

"효과 검정" 보고서에서 계수 추정값이 0인 10개 효과가 "제거"로 지정됩니다. "효과 검정" 보고서는 0.05 수준에서 하나의 효과 (Age*Markers 교호작용) 만 유의함을 나타냅니다.

그림 7.2 0 이 아닌 항이 강조 표시된 모수 추정값 보고서

원래 예측 변수에 대한 모수 추정값							
항	추정값	표준 오차	Wald 카이제곱	Prob > ChiSquare	95% 하한	95% 상한	
절편	1.9658064	1.5268652	1.6576014	0.1979	-1.026794	4.9584071	
BMI	-0.014557	0.0655906	0.0492589	0.8244	-0.143113	0.1139977	
Age	-0.000344	0.0052439	0.0043119	0.9476	-0.010622	0.0099335	
Time	0	0	0	1.0000	0	0	
Markers[0-1]	0.3853241	0.3347727	1.3248061	0.2497	-0.270818	1.0414665	
Hepatitis[0-1]	0	0	0	1.0000	0	0	
Jaundice[0-1]	-0.220764	0.1645347	1.8002793	0.1797	-0.543246	0.1017185	
(BMI-23.1737)*(Age-56.3994)	0.0007336	0.0014188	0.267356	0.6051	-0.002047	0.0035144	
(BMI-23.1737)*(Time-9.00945)	0	0	0	1.0000	0	0	
(BMI-23.1737)*Markers[0-1]	-0.066574	0.0511526	1.6938495	0.1931	-0.166831	0.0336832	
(BMI-23.1737)*Hepatitis[0-1]	0.0269878	0.0827496	0.1063662	0.7443	-0.135198	0.1891741	
(BMI-23.1737)*Jaundice[0-1]	0	0	0	1.0000	0	0	
(Age-56.3994)*(Time-9.00945)	0	0	0	1.0000	0	0	
(Age-56.3994)*Markers[0-1]	-0.023498	0.0083023	8.0106599	0.0047*	-0.03977	-0.007226	
(Age-56.3994)*Hepatitis[0-1]	0	0	0	1.0000	0	0	
(Age-56.3994)*Jaundice[0-1]	0	0	0	1.0000	0	0	
(Time-9.00945)*Markers[0-1]	-0.007529	0.0104633	0.5178189	0.4718	-0.028037	0.0129784	
(Time-9.00945)*Hepatitis[0-1]	0	0	0	1.0000	0	0	
(Time-9.00945)*Jaundice[0-1]	0.0009619	0.0093476	0.0105895	0.9180	-0.017359	0.0192828	
Markers[0-1]*Hepatitis[0-1]	-0.422619	0.397399	1.1309549	0.2876	-1.201507	0.3562684	
Markers[0-1]*Jaundice[0-1]	0	0	0	1.0000	0	0	
Hepatitis[0-1]*Jaundice[0-1]	0	0	0	1.0000	0	0	

12. "원래 예측 변수에 대한 모수 추정값" 보고서에서 (Age - 56.3994)*Markers[0-1] 행을 클릭합니다.

이렇게 하면 해 경로 그림에서 해당 효과의 경로가 강조 표시되고 데이터 테이블에서 Age 및 Markers 열이 선택됩니다.

13. "검증 열을 사용한 Poisson 적응형 Lasso 회귀" 옆의 빨간색 삼각형을 클릭하고 열 저장 > 예측 계산식 저장 및 열 저장 > 분산 계산식 저장을 선택합니다.

Node Count 예측 계산식과 Node Count 분산이라는 두 열이 데이터 테이블에 추가됩니다.

14. 데이터 테이블에서 열 머리글 중 하나를 마우스 오른쪽 버튼으로 클릭하고 **계산식**을 선택하여 계산식을 확인합니다. 아니면 "열" 패널에서 열 이름 오른쪽의 더하기 기호를 클릭합니다.

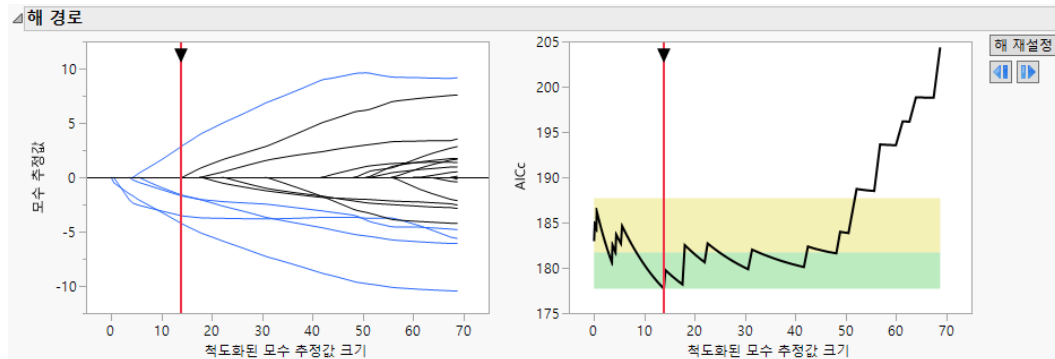
예측 계산식 저장 열의 예측 계산식은 모형의 추정된 선형 부분에 지수 함수를 적용합니다. Poisson 분포의 분산이 평균과 같으므로 Node Count 분산의 예측 분산 계산식은 동일한 계산식으로 제공됩니다.

이항 일반화 회귀 예

이 예에서는 Liver Cancer.jmp 샘플 데이터 테이블의 이항 반응 Severity(심각도)에 대한 예측 모형을 개발하는 방법을 보여 줍니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Liver Cancer.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. " 열 선택 " 목록에서 **Severity** 를 선택하고 **Y** 를 클릭합니다.
4. **BMI ~ Jaundice** 를 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다.
차수 2(**차수** 상자의 기본값)까지의 모든 항이 모형에 추가됩니다.
5. " 분석법 " 목록에서 **일반화 회귀**를 선택합니다.
" 분포 " 목록에 " 이항 " 분포가 자동으로 표시됩니다. Y가 이항 변수이고 명목형 모델링 유형일 때는 이 분포만 사용할 수 있습니다.
6. **실행**을 클릭합니다.
" 모형 비교 " 보고서, " 모형 시작 " 제어판 및 " 로지스틱 회귀 " 보고서가 포함된 " 일반화 회귀 " 보고서가 나타납니다. 기본 추정 방법은 "Lasso 회귀" 입니다.
7. " 추정 방법 "에서 **Elastic Net** 을 선택합니다.
8. **적응형** 상자를 선택합니다.
9. **시작**을 클릭합니다.
"AICc 검증을 사용한 이항 적응형 Elastic Net" 보고서가 나타납니다. 그림 7.3 에는 해 경로가 나와 있습니다.

그림 7.3 해 경로 그림



계수가 0 이 아닌 항의 경로가 파란색으로 표시됩니다. 최적 모수 값은 MLE 에서 상당히 축소됩니다. 오른쪽의 검증 그림은 몇 가지 모형을 최적 모형으로 고려할 수 있음을 나타냅니다. 해당 모형을 보려면 녹색 부분에 검은색 선이 있는 영역에서 빨간색 세로 막대를 움직입니다.

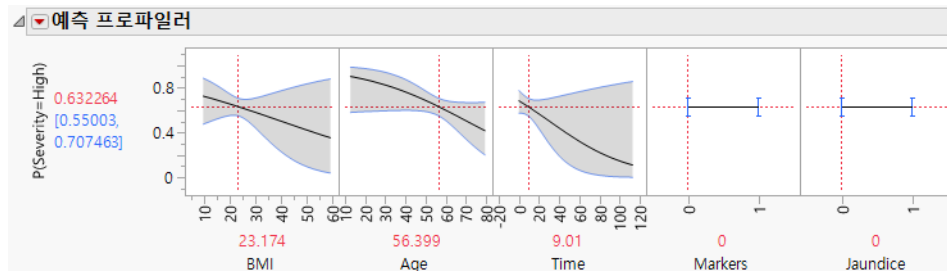
10. "AICc 검증을 사용한 이항 적응형 Elastic Net" 옆의 빨간색 삼각형을 클릭하고 **0 인 항 선택** 옵션을 선택합니다.

"원래 예측 변수에 대한 모수 추정값" 보고서에서 계수 추정값이 0 인 16 개 항이 강조 표시됩니다. 이 항은 "효과 검정" 보고서에 "제거"로 지정됩니다.

또한 "효과 검정" 보고서는 0.05 수준에서 유의한 항이 없음을 보여 줍니다. 그러나 Time*Markers 교호작용의 p 값 (0.0626) 이 작고 Time 효과의 p 값 (0.1458) 도 작습니다.

11. "AICc 검증을 사용한 이항 적응형 Elastic Net" 옆의 빨간색 삼각형을 클릭하고 **프로파일러 > 프로파일러**를 선택합니다.

그림 7.4 Severity = High 일 때 Time 이 낮을 확률에 대한 프로파일러

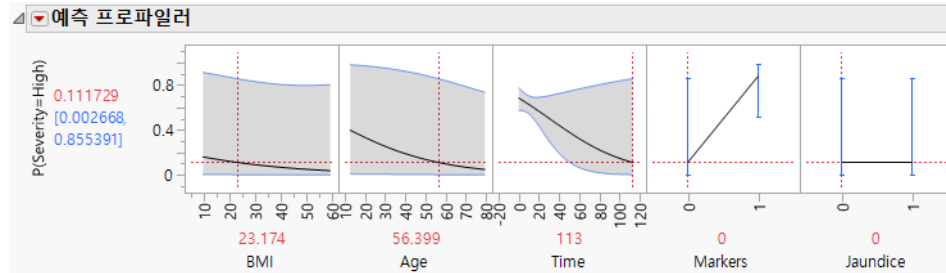


예측 프로파일러를 검토하여 Time 및 Time*Markers 교호작용이 Severity에 미치는 영향을 확인합니다.

참고 : Hepatitis 예측 변수는 활성 (0 이 아님) 항에 나타나지 않으므로 프로파일러에 표시되지 않습니다. Markers 와 Jaundice 는 활성 교호작용 항에 나타나므로 주효과가 활성이 아니더라도 프로파일러에 표시됩니다.

12. Time 의 빨간색 파선을 왼쪽에서 오른쪽으로 움직여 **Markers** 와의 교호작용을 확인합니다 (그림 7.4 및 그림 7.5). 진단 이후 Time 값이 작은 환자의 경우 **Markers** 는 **Severity** 에 거의 영향을 주지 않습니다. 그러나 장기간 진단을 받고 연구에 참여한 환자의 경우 **Markers** 가 중요합니다. 이러한 환자에게 정상 표지자는 **Severity** 가 높을 확률이 낮음을 나타냅니다.

그림 7.5 Severity = High 일 때 Time 이 높을 확률에 대한 프로파일러



영과잉 Poisson 회귀 예

Fishing.jmp 샘플 데이터 테이블에는 공원을 방문하는 그룹이 잡은 물고기 수에 영향을 미치는 다양한 요인 연구를 위한 가상의 데이터가 포함되어 있습니다. 여행 중인 가족 또는 그룹에 대한 250개의 반응이 데이터 테이블에 포함되어 있습니다. 이 예에서는 **Fish Caught**(잡힌 물고기 수)를 **Live Bait**(산 미끼), **Fishing Poles**(낚싯대), **Camper**(캠퍼), **People**(사람) 및 **Children**(어린이)의 함수로 모델링합니다. 이러한 변수는 데이터 테이블의 각 열에서 "노트" 열 특성에 설명되어 있습니다. "노트" 열 특성을 보려면 열 이름을 마우스 오른쪽 버튼으로 클릭하고 "열 정보"를 선택한 후 "열 특성" 아래의 "노트"를 선택합니다.

데이터 테이블에는 **Fished**(낚시 여부)라는 숨겨진 열이 있습니다. 데이터 수집 중 그룹의 누군가가 실제로 낚시를 했는지 여부는 확인되지 않았습니다. 그러나 두 가지 경우(그룹에서 낚시를 한 사람이 아무도 없거나 그룹에서 낚시를 한 모든 사람이 운이 없었음) 중 하나를 통해 물고기를 잡지 못하는 상황이 발생할 수 있다는 점을 강조하기 위해 **Fished** 열이 테이블에 포함되었습니다.

따라서 두 가지 원인을 통해 반응이 0일 수 있습니다. 이 문제를 해결하기 위해 영과잉 분포를 적합시킬 수 있습니다. Poisson 분포는 낚시를 한 사람에 대한 개수 데이터에 적합하므로 영과잉 Poisson 분포를 적합시킵니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Fishing.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. "열 선택" 목록에서 **Fish Caught** 를 선택하고 **Y** 를 클릭합니다.
4. **Live Bait ~ Children** 을 선택하고 **매크로 > 특정 차수까지의 요인**을 클릭합니다.
차수 2(차수 상자의 기본값)까지의 항이 모형에 추가됩니다.
5. "열 선택" 목록에서 "Validation" 을 선택하고 **검증**을 클릭합니다.
6. "분석법" 목록에서 **일반화 회귀**를 선택합니다.

7. " 분포 " 목록에서 **ZI Poisson** 을 선택합니다.

8. **실행**을 클릭합니다.

" 모형 비교 " 보고서, " 모형 시작 " 제어판 및 " 검증 열을 사용한 ZI Poisson 최대 가능도 " 보고서가 포함된 " 일반화 회귀 " 보고서가 나타납니다. 기본 추정 방법은 "Lasso 회귀" 입니다.

9. " 추정 방법 " 목록에서 **Elastic Net** 을 선택합니다.

10. **시작**을 클릭합니다.

" 검증 열을 사용한 ZI Poisson Elastic Net" 보고서가 나타납니다. " 해 경로 ", " 원래 예측 변수에 대한 모수 추정값 " 보고서 및 " 효과 검정 " 보고서에서 하나의 항이 0 임을 나타냅니다. 두 개의 모수 추정값 보고서에서 마지막 줄에 추정값이 표시된 영과잉 모수는 매우 유의합니다. 이것은 반응 (Fish Caught) 에서 나타난 일부 변동은 일부 그룹이 낚시를 하지 않았기 때문일 수 있음을 나타냅니다.

그림 7.6 원래 예측 변수에 대한 모수 추정값 보고서

원래 예측 변수에 대한 모수 추정값							
항	추정값	표준 오차	Wald	카이제곱	Prob > ChiSquare	95% 하한	95% 상한
절편	1.3599776	0.3237335	17.647701		<.0001*	0.7254716	1.9944837
Live Bait[0-1]	-0.379555	0.1919937	3.9081951		0.0481*	-0.755856	-0.003255
Fishing Poles	0.3034685	0.0465239	42.547553		<.0001*	0.2122833	0.3946537
Camper[0-1]	-0.363094	0.1899392	3.654349		0.0559	-0.735368	0.0091795
People	0.0776017	0.121377	0.4087617		0.5226	-0.160293	0.3154962
Children	-0.192656	0.1028943	3.5057658		0.0612	-0.394325	0.009013
Live Bait[0-1]*(Fishing Poles-0.856)	-0.026766	0.0668764	0.160184		0.6890	-0.157841	0.1043095
Live Bait[0-1]*Camper[0-1]	-0.336382	0.48759	0.4759452		0.4903	-1.292041	0.6192764
Live Bait[0-1]*(People-2.516)	-0.164164	0.1994702	0.6773258		0.4105	-0.555118	0.2267908
Live Bait[0-1]*(Children-0.936)	0.1025463	0.2362161	0.1884607		0.6642	-0.360429	0.5655213
(Fishing Poles-0.856)*Camper[0-1]	-0.446199	0.1120088	15.869111		<.0001*	-0.665732	-0.226666
(Fishing Poles-0.856)*(People-2.516)	-0.00611	0.0853344	0.0051267		0.9429	-0.173362	0.1611423
(Fishing Poles-0.856)*(Children-0.936)	-0.17587	0.080167	4.8127458		0.0282*	-0.332995	-0.018746
Camper[0-1]*(People-2.516)	-0.195261	0.3035369	0.4138148		0.5200	-0.790182	0.3996608
Camper[0-1]*(Children-0.936)	0.1899805	0.2998548	0.4014171		0.5264	-0.397724	0.7776852
(People-2.516)*(Children-0.936)	0	0	0		1.0000	0	0
ZI Poisson							
분포 모수	추정값	표준 오차	Wald	카이제곱	Prob > ChiSquare	95% 하한	95% 상한
영과잉	0.7815222	0.0354433	486.19905		<.0001*	0.7120545	0.8509898

" 효과 검정 " 보고서는 0.05 수준에서 네 개의 항, 즉 Live Bait, Fishing Poles, Fishing Poles* Camper 및 Fishing Poles*Children 이 유의함을 나타냅니다.

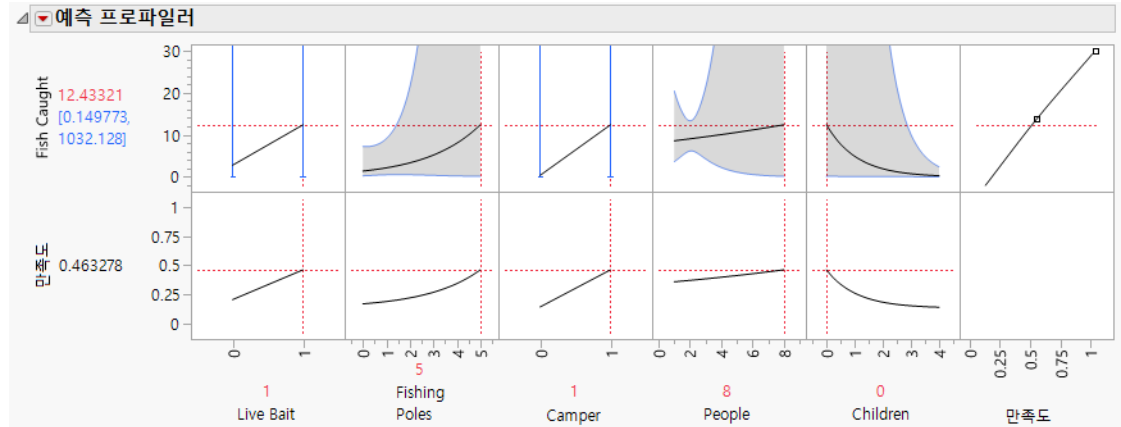
11. " 검증 열을 사용한 ZI Poisson Elastic Net" 옆의 빨간색 삼각형을 클릭하고 **프로파일러 > 프로파일러**를 선택합니다.

12. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 함수**를 선택합니다.

반응에 함수가 적용되어 Fish Caught 수를 최대화하는 것이 만족도가 높음을 나타냅니다. 만족도 함수에 대한 자세한 내용은 Profilers 의 에서 확인하십시오.

13. " 예측 프로파일러 " 의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다.

그림 7.7 Fish Caught 가 최대화된 예측 프로파일러



예측 변수의 설정을 변경하여 유의한 효과, 즉 Live Bait, Fishing Poles, Fishing Poles*Camper 및 Fishing Poles*Children의 영향을 확인할 수 있습니다. 예를 들어 Live Bait는 더 많은 물고기와 관련이 있고, Camper는 캠핑을 하지 않는 사람보다 낚싯대를 더 많이 가져와 더 많은 물고기를 잡는 경향이 있습니다.

14. "검증 열을 사용한 ZI Poisson Elastic Net" 열의 빨간색 삼각형을 클릭하고 열 저장 > 예측 계산식 저장 및 열 저장 > 분산 계산식 저장을 선택합니다.

Fish Caught 예측 계산식과 Fish Caught 분산이라는 두 열이 데이터 테이블에 추가됩니다.

15. 데이터 테이블에서 열 머리글 중 하나를 마우스 오른쪽 버튼으로 클릭하고 계산식을 선택하여 계산식을 확인합니다. 아니면 "열" 패널에서 열 이름 오른쪽의 더하기 기호를 클릭합니다. 두 계산식에서 모두 추정된 영과잉 모수가 0.781522155로 나타납니다.

가로형 데이터에 대한 일반화 회귀 예

예측 변수가 관측값 수보다 많은 경우 기존 회귀 방법은 실용적이지 않습니다. 이러한 경우 변수 선택을 포함하는 회귀 방법을 사용하여 회귀 모형을 적합시킬 수 있습니다. 이 예에서는 변수 선택 수준이 서로 다른 세 모형을 비교합니다. 세 모형이 모두 검증 그림의 녹색 영역 내에 있으므로 이러한 모형 모두 최적 모형과 유사하다는 강력한 증거가 있습니다.

Prostate Cancer.jmp 샘플 데이터 테이블에는 165 명의 남성으로부터 수집한 혈청 표본의 결과가 포함되어 있습니다. 이 중 약 절반이 전립선암을 앓고 있습니다. 혈청 표본에서 측정된 667 개의 단백질이 있습니다.

1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Prostate Cancer.jmp를 엽니다.
2. 분석 > 모형 적합을 선택합니다.
3. "열 선택" 목록에서 Status를 선택하고 Y를 클릭합니다.

이 열은 명목형 반응 열이므로 "분석법"이 "명목형 로지스틱"으로 변경되고 "목표 수준" 옵션이 나타납니다. 이 옵션의 기본값은 데이터 테이블의 "목표 수준" 열 특성에 지정된 값인 "CCD"입니다.

4. "분석법" 목록에서 **일반화 회귀**를 선택합니다.

"분포" 목록에 "이항" 분포가 자동으로 표시됩니다. Y가 이항 변수이고 명목형 모델링 유형일 때는 이 분포만 사용할 수 있습니다.

5. "열 선택" 목록에서 **Proteins** 열 그룹을 선택하고 **추가**를 클릭합니다.

그러면 열 그룹에 포함된 667개 열이 모두 모형에 추가됩니다.

6. **실행**을 클릭합니다.

"모형 시작" 제어판이 포함된 "일반화 회귀" 보고서가 나타납니다. 예측 변수의 수가 관측 값 수보다 많으므로 초기 로지스틱 회귀 모형 적합이 없습니다.

7. "추정 방법"에서 **Elastic Net**을 선택합니다.

8. "고급 제어" 옆의 회색 표시 아이콘을 클릭합니다.

그림 7.8 고급 제어

모형 시작

특이성 상세 정보

반응 분포

이항

추정 방법

Elastic Net

☐ 적응형

고급 제어

Elastic Net 알파 0.99

격자 점 수 150

최소 불점 분율 0.001

격자 척도 제공근

초기에 표시된 해 최량 적합

항 강제 적용

검증 방법

AICc

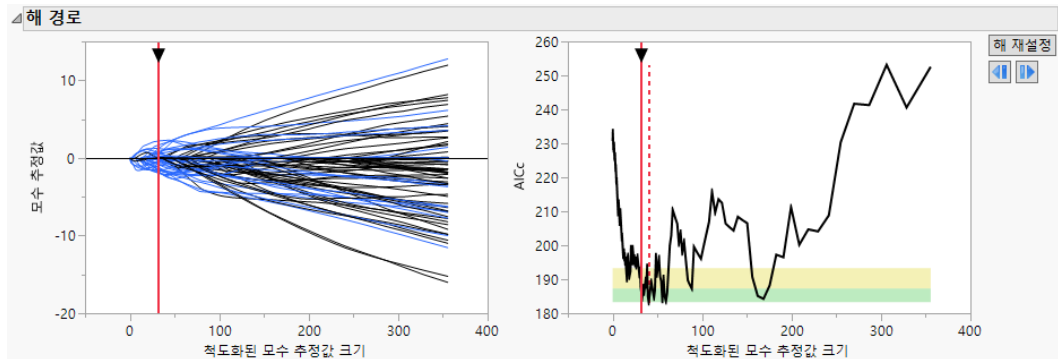
☐ 조기 중지

시작

9. "초기에 표시된 해"에서 **녹색 영역에서 가장 작음**을 선택합니다.

10. **시작**을 클릭합니다.

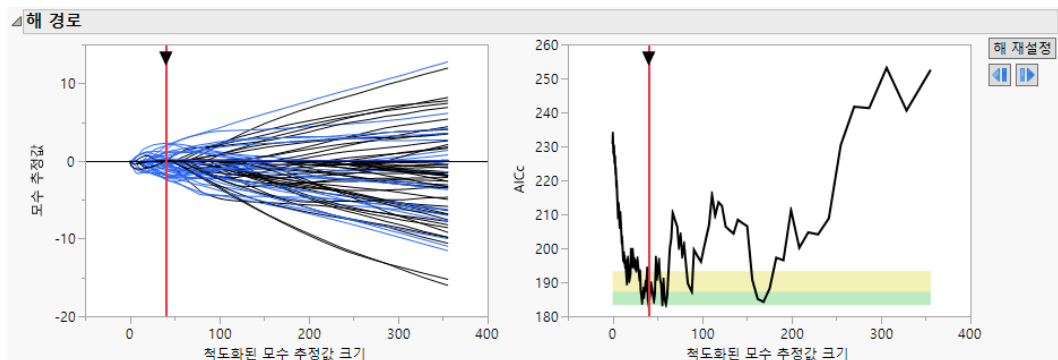
그림 7.9 녹색 영역에서 가장 작은 모형



"해 경로"에서는 최소 AICc 모형과 유사한 것으로 간주되는 가장 작은 모형을 보여 줍니다.
여기서 가장 작은 모형은 모수의 수가 가장 적은 모형을 의미합니다.

11. "AICc 검증을 사용한 이항 Elastic Net" 옆의 회색 표시 아이콘을 클릭합니다.
12. "모형 시작" 옆의 회색 표시 아이콘을 클릭합니다.
13. "초기에 표시된 해"에서 **최량 적합**을 선택합니다.
14. **시작**을 클릭합니다.

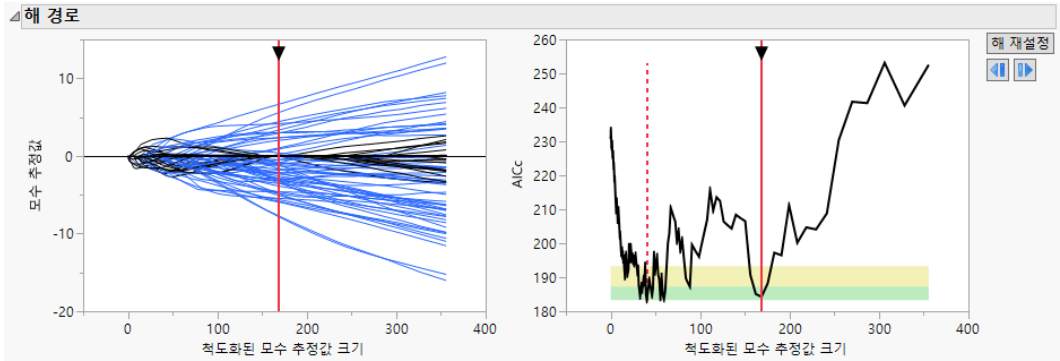
그림 7.10 최량 적합 모형



"해 경로"에서는 최량 적합 모형을 보여 줍니다. 여기서 최량 적합 모형은 최소 AICc 값을 가진 모형을 의미합니다.

15. "AICc 검증을 사용한 이항 Elastic Net" 옆의 회색 표시 아이콘을 클릭합니다.
16. "모형 시작" 옆의 회색 표시 아이콘을 클릭합니다.
17. "초기에 표시된 해"에서 **녹색 영역에서 가장 큼**을 선택합니다.
18. **시작**을 클릭합니다.

그림 7.11 녹색 영역에서 가장 큰 모형



"해 경로"에서는 최소 AICc 모형과 유사한 것으로 간주되는 가장 큰 모형을 보여 줍니다. 여기서 가장 큰 모형은 모수의 수가 가장 많은 모형을 의미합니다.

19. "AICc 검증을 사용한 이항 Elastic Net" 옆의 회색 표시 아이콘을 클릭합니다.

그림 7.12 모형 비교 테이블

모형 비교							
표시	반응 분포	추정 방법	검증 방법	0이 아닌 모수	AICc	BIC	일반화 R ²
☒	이항	Elastic Net	AICc	36	186.07102	277.07256	0.7466199
☒	이항	Elastic Net	AICc	40	183.29544	281.08164	0.8021925
☒	이항	Elastic Net	AICc	57	184.25399	299.49849	0.9824565

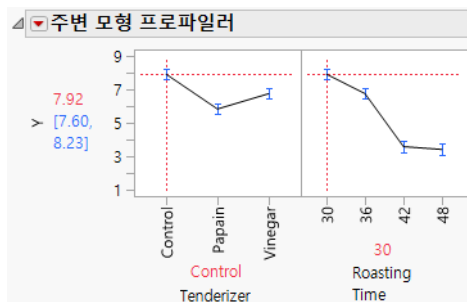
"모형 비교" 보고서에 세 개의 모형이 표시됩니다. "0이 아닌 모수" 열을 사용하여 각 모형의 크기를 식별할 수 있습니다. 모형의 모수 수가 증가하면 "일반화 R²" 값이 증가합니다. 이러한 모형이 모두 녹색 영역에 있으므로 모든 모형이 최적 모형과 유사하다는 강력한 증거가 있습니다.

모형 적합 플랫폼의 혼합 모형 분석법은 JMP Pro 에서만 사용할 수 있습니다.

JMP Pro에서 모형 적합 플랫폼의 혼합 모형 분석법은 복소 공분산 구조의 연속형 반응에 대해 다양한 선형 모형을 적합시킵니다. 이러한 모형에는 랜덤 계수, 반복 측정, 공간 데이터 및 상관된 여러 반응이 있는 데이터가 포함됩니다. 혼합 모형 분석법을 사용하여 직관적인 인터페이스로 선형 혼합 모형 및 모형의 공분산 구조를 간편하게 지정하고, 최대 가능도 방법으로 이러한 모형을 적합시킬 수 있습니다.

모델링 결과는 프로파일러, 표면 그림 및 등고선 그림과 같은 대화식 시각화 도구에 의해 지원됩니다. 이러한 도구를 사용하여 모형에 대한 이해를 보완할 수 있습니다.

그림 8.1 분할구 실험에 대한 주변 모형 프로파일러



목차

혼합 모형 분석법 개요.....	330
예 : 혼합 모형 사용.....	330
혼합 모형 분석법 시작.....	335
모형 적합 시작 창.....	335
데이터 형식.....	340
혼합 적합 보고서.....	341
적합 통계량.....	343
임의 효과 공분산 모수 추정값.....	345
고정 효과 모수 추정값.....	345
반복 효과 공분산 모수 추정값.....	346
랜덤 계수.....	347
임의 효과 예측.....	347
고정 효과 검정.....	347
순차 검정.....	348
다중 비교.....	348
기울기 비교.....	349
주변 모형 추론.....	349
실제값 대 예측값 그림.....	349
잔차 그림.....	349
Profilers.....	350
조건부 모형 추론.....	351
실제값 대 조건부 예측값 그림.....	351
조건부 잔차 그림.....	351
조건부 프로파일러.....	352
변동도.....	352
공분산 및 상관 행렬.....	353
열 저장.....	354
혼합 모형 분석법의 추가 예.....	354
반복 측정 예.....	355
분할구 예.....	370
공간 예: 균일도 시험.....	376
상관관계가 있는 반응의 예.....	384
혼합 모형 분석법에 대한 통계 상세 정보.....	390
수렴 스코어 검정.....	391
랜덤 계수 모형.....	392
반복 측정.....	393
반복 공분산 구조.....	393

공간 및 시간 변동	398
Kackar-Harville 수정	401

JMP PRO 혼합 모형 분석법 개요

JMP Pro의 혼합 모형 분석법을 사용하면 복소 공분산 구조를 가진 모형을 분석할 수 있습니다. 다음과 같은 경우 이 방법으로 분석할 수 있습니다.

- 분할구 실험
- 랜덤 계수 모형
- 반복 측정 설계
- 공간 데이터
- 상관관계가 있는 반응 데이터

분할구 실험은 실험 단위의 수준 또는 크기가 두 개 이상이고 결과적으로 여러 오차 항을 생성하는 실험입니다. 보통 일부 요인은 변경하기 쉽고 다른 요인은 변경하기 어려울 때 이러한 설계가 필요합니다. 자세한 내용은 **실험 설계 가이드**의 에서 확인하십시오.

랜덤 계수 모형은 계층적 또는 다수준 모형이라고도 합니다(Singer 1998, Sullivan et al. 1999). 배치 또는 개체의 절편 및 기울기가 무작위로 달라진다고 추정될 때 이러한 모형이 사용됩니다. 제약 산업의 약물 안정성 시험과 교육 리서치의 개별 성장 연구에서 랜덤 계수 모형이 필요한 경우가 많습니다.

반복 측정 설계, 공간 데이터 및 상관된 반응 데이터는 관측값이 독립적이지 않다는 특성을 공유하므로 상관 구조를 모델링해야 합니다.

- 반복 측정 설계 (개체 내 설계라고도 함)는 오차의 상관관계를 허용하면서 시간 또는 공간에 따른 반응의 변화를 모델링합니다.
- 공간 데이터는 일반적으로 위도와 경도 같은 둘 이상의 차원에서 측정되는 값입니다. 공간 측정값은 주로 공간 근접성의 함수로 상관관계를 보입니다.
- 상관된 반응 데이터는 동일한 실험 단위에 대해 여러 가지 측정을 수행한 결과입니다. 예를 들어 의학 연구에서 개인의 키, 체중 및 혈압 측정값 또는 제조품에 대해 측정한 경도, 강도 및 탄성은 상관관계가 있을 가능성이 큼니다. 이러한 측정값을 개별적으로 연구할 수 있지만 상관된 반응으로 처리하면 유용한 정보를 끌어낼 수 있습니다.

관측값 간의 상관관계를 설명하지 못하면 처리 효과에 대한 잘못된 결론을 내릴 수 있습니다. 그러나 공분산 구조 모수를 추정할 때 데이터의 정보가 사용됩니다. 추정되는 모수의 수는 검정력과 제1종 오류 비율에 영향을 줍니다. 따라서 공분산 모형을 신중하게 선택해야 합니다. 자세한 내용은 "**반복 측정 예**"(355페이지)에서 확인하십시오.

JMP PRO 예 : 혼합 모형 사용

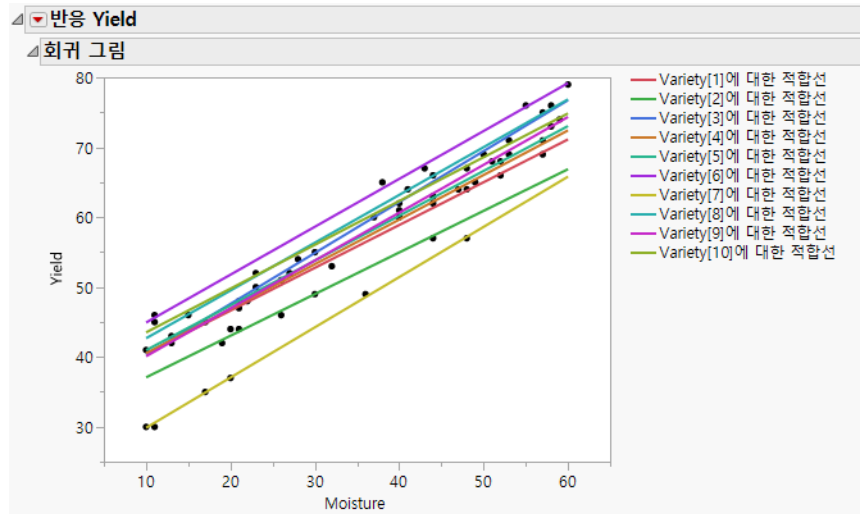
밀 수확량 연구에서 건조 기후 조건에 맞게 적응된 경질 붉은 겨울밀 품종 모집단에서 10종의 밀을 무작위로 선택합니다. 그런 다음 6개의 1에이커 대지에 무작위로 할당합니다. 예정지의 수분 함량은 발아율에 영향을 줄 수 있으며 결과적으로 최종 수확량이 달라질 수 있습니다. 따라서 각

대지에 대해 상위 36인치 토양의 예정지 수분 양이 결정됩니다. 이 연구에서 수분 함량이 수확량에 영향을 주는지 여부를 확인하는 데 관심이 있습니다.

품종이 무작위로 선택되므로 각 품종에 대한 회귀 모형은 품종 모형 모집단에서 선택한 랜덤 모형입니다. 절편과 기울기는 각 품종에 대해 랜덤하며 상관관계가 있을 수 있습니다. 랜덤 계수는 고정 효과에서 중심화됩니다. 고정 효과는 모집단 절편과 기울기이며, 이는 품종의 절편과 기울기 모집단에 대한 기대값입니다. 이 예는 Littell et al. 연구 자료 (2006, p. 320)에서 가져왔습니다.

표준 최소 제곱 분석법에서 REML을 사용하여 모형을 적합시키면 절편과 기울기의 변동을 볼 수 있습니다(그림 8.2). 기울기의 변동성은 크지 않지만 절편은 상당한 차이가 있습니다. 절편과 기울기는 음의 상관관계가 있을 수 있으며 절편이 낮은 품종의 기울기가 더 높은 것으로 보입니다.

그림 8.2 표준 최소 제곱 회귀



절편과 기울기 간의 상관관계를 모델링하기 위해 혼합 모형 분석법을 사용합니다. 이를 통해 모 집단 회귀 방정식과 품종별 방정식을 결정하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Wheat.jmp** 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. **Yield** 를 선택하고 **Y** 를 클릭합니다.
이 열을 **Y** 로 추가하면 적합 분석법이 "표준 최소 제곱" 으로 설정됩니다.
4. "분석법" 목록에서 **혼합 모형**을 선택합니다. 아니면 "혼합 모형" 분석법을 먼저 선택한 후 **Y** 를 클릭하여 **Yield** 를 추가할 수도 있습니다.
5. **Moisture** 를 선택하고 "고정 효과" 탭에서 **추가**를 클릭합니다.

그림 8.3 고정 효과를 보여 주는 완료된 모형 적합 시작 창

모형 규격

열 선택

- 5개 열
- Variety
- Yield
- Moisture
- Pred Formula Yield
- Cond Pred Formula Yield

역할 변수 선택

Y Yield
선택적

기준 선택적

분석법: 혼합 모형

☒ 경계가 없는 분산 성분

도움말 실행

재호출 ☐ 대화상자 열린 채 유지

제거

모형 효과 생성

고정 효과 임의 효과 반복 구조

추가 Moisture

교차

내포

매크로

차수 2

속성

☐ 절편 없음

6. 임의 효과 탭을 선택합니다.

7. Moisture 를 선택하고 추가를 클릭합니다.

8. " 열 선택 " 목록에서 **Variety** 를 선택하고 " 임의 효과 " 탭에서 Moisture 를 선택한 후 **내포 랜덤 계수**를 클릭합니다.

그림 8.4 임의 효과 탭을 보여 주는 완료된 모형 적합 시작 창

모형 규격

열 선택

- 5개 열
- Variety
- Yield
- Moisture
- Pred Formula Yield
- Cond Pred Formula Yield

역할 변수 선택

Y Yield
선택적

기준 선택적

분석법: 혼합 모형

☒ 경계가 없는 분산 성분

도움말 실행

재호출 ☐ 대화상자 열린 채 유지

제거

모형 효과 생성

고정 효과 임의 효과 반복 구조

추가 절편[Variety] & 랜덤 계수(1)

교차 Moisture[Variety] & 랜덤 계수(1)

내포

내포 랜덤 계수

매크로

차수 2

임의 효과는 품종별로 그룹화되고 절편은 랜덤 성분으로 포함됩니다.

9. (선택 사항 .) " 모형 규격 " 의 빨간색 삼각형을 클릭하고 **중심화 다항식** 옵션의 설정을 확인합니다 .

Moisture 효과는 임의 효과로 포함되어 있으므로 " 중심화 다항식 " 옵션이 선택되어 있더라도 평균으로 중심화되지 않습니다 .

10. **실행**을 클릭합니다 .

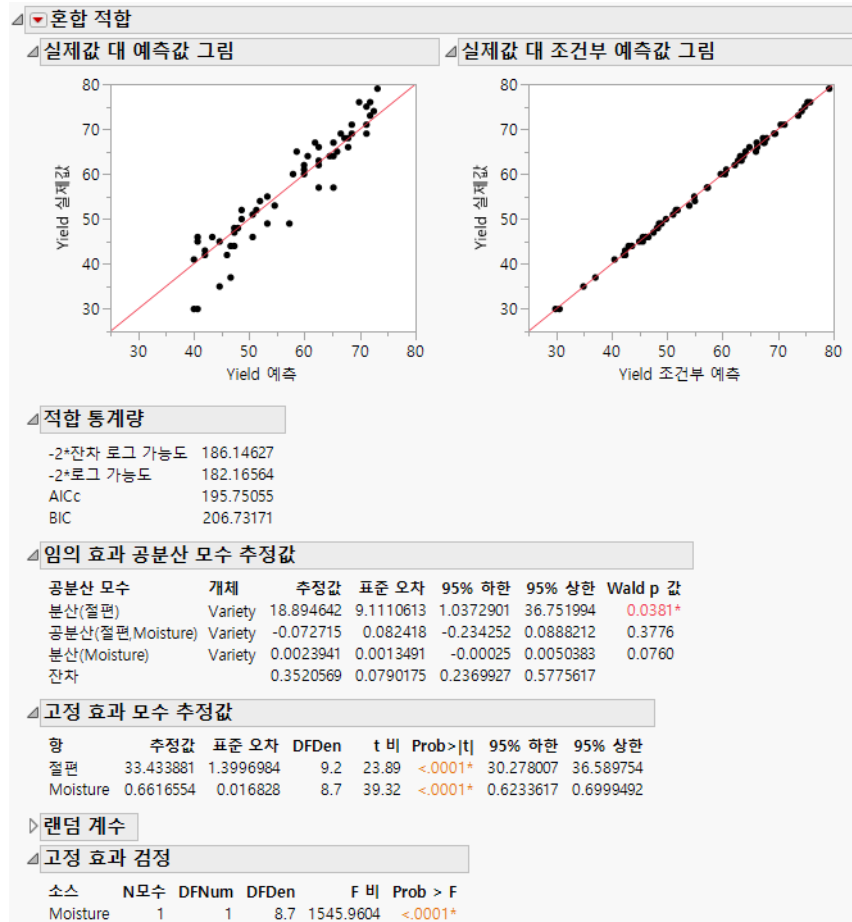
그림 8.5 에 " 혼합 적합 " 보고서가 표시되어 있습니다 . " 실제값 대 예측값 그림 " 은 모형 적합 및 기본 가정에 대해 불일치를 나타내지 않습니다 .

모형 적합에 명백한 문제가 없으므로 통계 검정을 해석하고 회귀 방정식을 구할 수 있습니다 . " 고정 효과 검정 " 보고서에서 볼 수 있듯이 수확량에 대한 수분의 효과는 유의합니다 . " 고정 효과 모수 추정값 " 에 제공된 추정값은 다음 방정식이 추정된 모집단 회귀 방정식임을 나타냅니다 .

$$Yield = 33.43 + 0.66 * Moisture$$

" 임의 효과 공분산 모수 추정값 " 보고서에는 품종의 절편 분산을 나타내는 " 분산 (절편)", 기울기를 나타내는 " 분산 (Moisture)" 및 공분산을 나타내는 " 공분산 (절편 , Moisture)"의 추정값이 제공됩니다 . 이 경우 절편과 기울기는 추정값의 신뢰 구간에 0 이 포함되므로 유의한 상관관계가 없습니다 . 이 보고서에는 잔차 분산의 추정값도 제공됩니다 .

그림 8.5 혼합 적합 보고서



- 모집단 회귀 방정식의 추정값이 있지만 품종 (Variety) 2의 추정 수확량에도 관심이 있습니다.
11. "랜덤 계수" 보고서를 열어 "절편" 및 "Moisture"에 대한 품종 효과의 추정값을 확인합니다. 이러한 계수는 각 품종이 모집단과 어떻게 다른지 추정합니다.

그림 8.6 랜덤 계수 보고서

랜덤 계수		
Variety		
Variety	절편	Moisture
1	0.9577924	-0.049211
2	-2.284289	-0.066697
3	-0.408109	0.0672225
4	0.6960205	-0.023306
5	1.115904	-0.019904
6	4.639151	0.0238887
7	-10.73004	0.0564233
8	2.4011709	0.0224336
9	-0.176207	0.0233566
10	3.7886051	-0.034207
공분산 행렬		
임의 효과	절편	Moisture
절편	18.89464	-0.07272
Moisture	-0.07272	0.002394

" 고정 효과 모수 추정값 " 및 " 랜덤 계수 " 보고서에서 Variety 2 에 대해 다음과 같은 예측 방정식을 구합니다 .

$$Yield = 33.433 + 0.662 * Moisture - 2.284 - 0.067 * Moisture$$

$$Yield = 31.149 + 0.595 * Moisture$$

Variety 2 는 모집단 평균보다 낮은 수확량으로 시작하여 모집단 평균보다 낮은 비율로 Moisture 와 함께 증가합니다 .

JMP PRO 혼합 모형 분석법 시작

혼합 모형은 모형 적합 시작 창에서 선택할 수 있는 여러 분석법 중 하나입니다. 이 섹션에서는 모형 적합 시작 창에서 혼합 모형을 적합 방법으로 선택하는 방법을 설명합니다. 이 선택 항목과 관련된 옵션에 대한 내용도 다룹니다.

JMP PRO 모형 적합 시작 창

고정 효과, 임의 효과, 반복 구조 또는 이들의 조합을 사용하여 모형을 지정할 수 있습니다. 옵션은 사용자가 지정한 모형의 특성에 따라 다릅니다.

혼합 모형 분석법을 사용하여 모형을 적합시키려면 **분석 > 모형 적합**을 선택한 후 "분석법" 목록에서 **혼합 모형**을 선택합니다. 분석법을 선택하기 전에 Y 목록에 연속형 변수를 입력하면 분석법이 기본적으로 "표준 최소 제곱"으로 설정됩니다. "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

혼합 모형 분석법을 사용하여 모형을 적합시킬 때 비유계 분산 성분을 허용할 수 있습니다. 이렇게 하면 추정값이 음수인 분산 성분이 0으로 보고되지 않습니다. 이 옵션은 기본적으로 선택되어 있습니다. 분산 추정값을 0으로 제한하면 고정 효과 검정에서 편향이 발생할 수 있으므로 고

정 효과에 관심이 있다면 이 옵션이 선택된 상태여야 합니다. "비유계 분산 성분" 옵션에 대한 자세한 내용은 "표준 최소 제곱 모형" 장의 "**임의 분산**"(175 페이지)에서 확인하십시오.

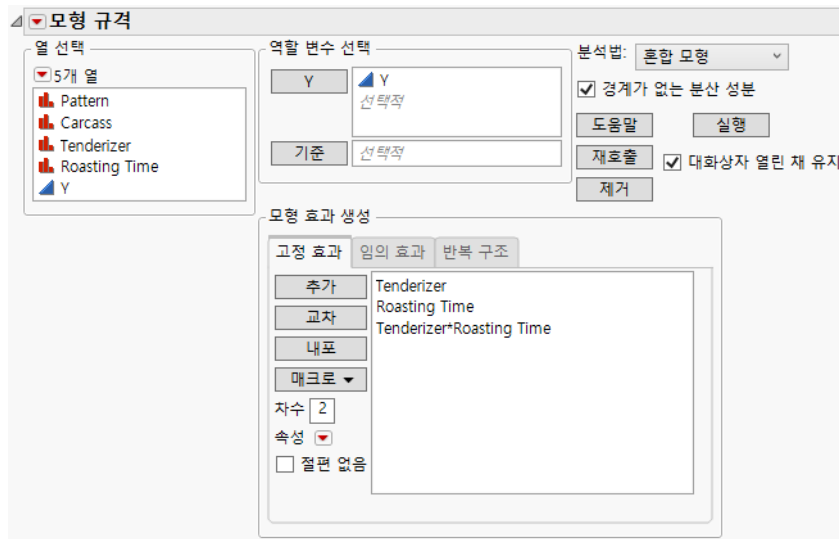
JMP PRO 고정 효과 탭

"고정 효과" 탭에서 모든 고정 효과를 추가합니다. 필요에 따라 "추가", "교차", "내포", "매크로" 및 "속성" 옵션을 사용합니다. 이러한 옵션에 대한 자세한 내용은 "**모형 규격**" 장(29 페이지)에서 확인하십시오.

참고 : 연속형 열이 임의 효과에 포함된 경우에는 "모형 규격"의 빨간색 삼각형 메뉴에서 "중심화 다항식" 옵션이 선택되어 있더라도 해당 열이 중심화되지 않습니다.

그림 8.7에는 Split Plot.jmp 샘플 데이터 테이블 분석의 고정 효과가 나와 있습니다. 모형에 고정 효과가 없을 수도 있습니다. 예는 "**공간 예 : 균일도 시험**"(376 페이지)에서 확인하십시오.

그림 8.7 완료된 고정 효과를 보여 주는 모형 적합 시작 창



JMP PRO 임의 효과 탭

"임의 효과" 탭을 사용하여 기존의 분산 성분 모형 및 랜덤 계수 모형을 지정합니다.

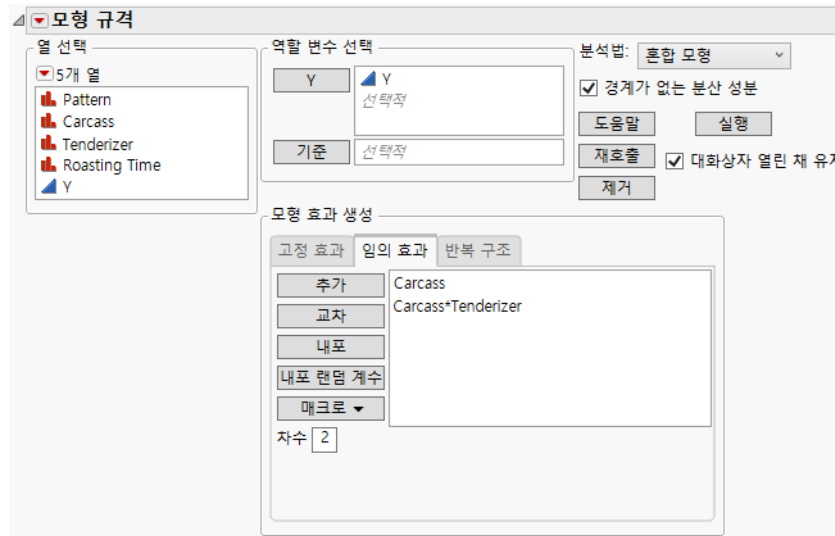
참고 : 연속형 열이 임의 효과에 포함된 경우에는 "모형 규격"의 빨간색 삼각형 메뉴에서 "중심화 다항식" 옵션이 선택되어 있더라도 해당 열이 중심화되지 않습니다.

분산 성분

기존 분산 성분 모형의 경우 "추가", "교차" 또는 "내포" 옵션을 사용하여 랜덤 블록, 주구 오차 항 및 하위구 오차 항과 같은 항을 지정합니다. 이러한 옵션에 대한 자세한 내용은 "모형 규격" 장(29페이지)에서 확인하십시오.

그림 8.8에서는 Carcass가 랜덤 블록인 Split Plot.jmp 샘플 데이터의 임의 효과 지정을 보여 줍니다. "분할구 예"(370 페이지)에서 이 예를 자세히 설명합니다.

그림 8.8 완료된 임의 효과 탭을 보여 주는 모형 적합 시작 창



랜덤 계수

랜덤 계수 모형을 생성하려면 "내포 랜덤 계수" 버튼을 사용하여 랜덤 계수 그룹을 생성합니다.

1. "열 선택" 목록에서 연속형 예측 변수 열을 선택합니다.
2. 임의 효과 탭을 선택한 후 추가를 선택합니다.
3. "임의 효과" 탭에서 이러한 효과를 선택합니다. 수준이 개별 회귀 모형을 정의하는 임의 효과가 포함된 열도 선택합니다. 이 열은 기본적으로 SAS PROC MIXED에서 random 문의 개체입니다.
4. 내포 랜덤 계수 버튼을 클릭합니다.

이 마지막 단계에서는 임의 효과의 수준 내에서 상관되는 랜덤 절편 및 랜덤 기울기 효과를 생성합니다. 개체는 개체 간의 변동성으로 인해 다른 개체에 내포됩니다. 절편이 모든 그룹에 대해 고정되어 있다고 간주하는 경우 절편[그룹]&랜덤 계수(1)을 선택한 후 제거를 클릭합니다.

계층적 선형 모형에서와 마찬가지로 이러한 방식으로 여러 랜덤 계수 그룹을 정의할 수 있습니다. 기울기 및 절편 계수에 대해 처리 효과별 랜덤 배치와 랜덤 배치 효과가 모두 있을 때 이 작업

이 필요할 수 있습니다. 또한 성취도 점수에 대해 랜덤 학생 효과와 랜덤 학교 효과가 있고 학생이 학교에 내포된 경우와 같은 계층적 선형 모형에서도 필요할 수 있습니다.

랜덤 계수는 비정형 공분산 구조를 사용하여 모델링됩니다. 그림 8.9에서는 Wheat.jmp 샘플 데이터의 랜덤 계수 지정을 보여 줍니다. 자세한 내용은 "예: 혼합 모형 사용"(330 페이지)에서 확인하십시오.

그림 8.9 랜덤 계수를 보여 주는 완료된 모형 적합 시작 창

모형 규격

열 선택: 5개 열
 Variety
 Yield
 Moisture
 Pred Formula Yield
 Cond Pred Formula Yield

역할 변수 선택:
 Y: Yield (선택적)
 기준: (선택적)

분석법: 혼합 모형
☒ 경계가 없는 분산 성분
 도움말 실행
 재호출 ☐ 대화상자 열린 채 유지
 제거

모형 효과 생성
 고정 효과 임의 효과 반복 구조
 추가 교차 내포 내포 랜덤 계수 매크로
 차수 2

질문[Variety] & 랜덤 계수(1)
 Moisture[Variety] & 랜덤 계수(1)

JMP PRO 반복 구조 탭

"반복 구조" 탭을 사용하여 모형의 반복 효과에 대한 공분산 구조를 선택합니다.

그림 8.10 반복 구조 탭을 보여 주는 완료된 모형 적합 시작 창

The screenshot shows the '모형 규격' (Model Specification) window in JMP Pro. The '반복 구조' (Repetition Structure) tab is active. In the '역할 변수 선택' (Role Variable Selection) section, 'Yield' is selected as the response variable. In the '모형 효과 생성' (Model Effect Generation) section, the '반복' (Repetition) structure is set to '공간' (Space) and '구형' (Spherical). The '개체' (Individual) section is set to '선택적' (Optional). The '반복' (Repetition) section also shows 'Row' and 'Column' as factors.

구조

반복 구조는 기본적으로 "잔차"로 설정됩니다. 잔차 구조는 관측값 간에 공분산이 없음을 지정합니다. 즉, 오차가 독립적이라는 의미입니다. 잔차 구조와 이분산 구조 외에 다른 모든 공분산 구조도 관측값 간의 공분산을 모델링합니다. 구조에 대한 자세한 내용은 통계 상세 정보 섹션의 "반복 측정"(393 페이지) 및 "공간 및 시간 변동"(398 페이지)에서 확인하십시오.

표 8.1에는 사용 가능한 공분산 구조, 각 구조를 사용하기 위한 요구 사항 및 주어진 구조에 대한 공분산 모수의 수가 나열되어 있습니다. 관측 시간 수는 J 로 표시됩니다.

표 8.1 반복 공분산 구조 요구 사항

구조	반복 열 유형	필요한 반복 열 수	개체	모수 수
잔차	해당 없음	0	해당 없음	0
이분산	범주형	1	선택적	J
비정형	범주형	1	필수	$J(J+1)/2$
AR(1)	연속형	1	선택적	2
복합 대칭	범주형	1	필수	2

표 8.1 반복 공분산 구조 요구 사항 (계속)

구조	반복 열 유형	필요한 반복 열 수	개체	모수 수
Antedependent 등분산	범주형		필수	J
Toeplitz	범주형	1	필수	J
복합 대칭 이분산	범주형	1	필수	$J+1$
Antedependent	범주형		필수	$2J-1$
Toeplitz 이분산	범주형	1	필수	$2J-1$
공간	연속형	2+	선택적	
공간 이방성	연속형	2+	선택적	
공간(너겟 사용)	연속형	2+	선택적	
공간 이방성(너 겟 사용)	연속형	2+	선택적	

잔차 구조를 사용하여 "반복" 또는 "개체" 열을 입력하면 해당 열이 무시됩니다. 이 경우 "잔차 반복 공분산 구조를 선택한 경우에는 반복 열 및 개체 열이 무시됩니다"라는 경고가 표시됩니다.

유형

공간 공분산 구조 중 하나를 선택하면 공간 구조 유형을 선택할 수 있는 "유형" 목록이 나타납니다. 여기서 "거듭제곱", "지수", "가우스" 및 "구형"의 네 가지 유형을 사용할 수 있습니다. 그림 8.10에서는 Uniformity Trial.jmp 샘플 데이터에 대해 공간 구조 및 구형 유형이 선택된 것을 보여줍니다.

반복

반복 측정 구조를 정의하는 열을 입력합니다. 반복 열의 모델링 유형은 공분산 구조에 따라 다릅니다. 각 반복 측정 공분산 구조의 요구 사항에 대한 자세한 내용은 표 8.1에서 확인하십시오.

개체

개체를 정의하는 하나 이상의 열을 입력합니다. 개체 열은 범주형이어야 합니다.

JMP
PRO 데이터 형식

모형 적합 플랫폼에서 혼합 모형 분석법을 사용하려면 모든 반응 측정값이 하나의 반응 열에 포함되어야 합니다. 반복 측정 데이터가 여러 열에 기록되는 경우도 있습니다. 이때 각 행은 개체이고 반복 측정값은 별도의 반응 열에 기록됩니다. 혼합 모형 분석법을 실행하기 전에 이 형식의

데이터를 쌓아야 합니다. **Cholesterol.jmp** 및 **Cholesterol Stacked.jmp** 샘플 데이터 테이블에서는 각각 가로형 형식과 쌓인 형식을 보여 줍니다. 가로형 테이블의 각 행은 쌓인 테이블에서 **Patient**의 한 수준에 해당합니다.

JMP PRO 혼합 적합 보고서

"혼합 적합"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

모형 보고서 혼합 모형 적합과 관련된 보고서를 생성합니다. 이러한 보고서에는 적합 통계량 및 모형 모수에 대한 추정값과 검정이 제공됩니다. 자세한 내용은 "**모형 보고서**"(342 페이지)에서 확인하십시오.

분산 성분의 선형 결합 분산 성분의 선형 결합에 대한 신뢰 구간을 계산할 수 있는 보고서를 표시합니다. 처음에는 이 보고서에 편집 가능한 텍스트 상자 및 모형의 분산 성분 테이블이 포함됩니다. 텍스트 상자를 사용하여 선형 결합에 라벨을 지정할 수 있습니다. 테이블에서 오른쪽 열의 셀에 값을 입력하여 신뢰 구간에 대한 선형 함수를 지정할 수 있습니다. 모수의 선형 결합을 지정하고 "완료"를 클릭하면 지정된 선형 결합에 대한 신뢰 구간이 포함된 테이블이 나타납니다.

이 테이블에는 추정값, 표준 오차 및 두 가지 유형의 신뢰 구간(Satterthwaite, Wald)과 Wald p 값이 포함되어 있습니다. Wald p 값은 추정값이 0과 다르다는 가설 검정에 해당합니다.

팁: Satterthwaite 신뢰 구간은 양수 값으로 제한되므로 지정된 계수가 음수인 경우에는 권장되지 않습니다. 추정값이 음수이면 Satterthwaite 신뢰 구간을 생성할 수 없으며 결측값으로 보고됩니다.

다중 비교 하나 이상의 효과 및 초기 비교를 선택할 수 있는 "다중 비교" 대화상자 창을 엽니다. 이 보고서는 범주형 고정 효과에 사용할 수 있습니다. 자세한 내용은 "**다중 비교**"(348 페이지)에서 확인하십시오.

기울기 비교 (명목형 항과 연속형 항이 각각 하나이고 고정 효과에 대한 교호작용 효과가 있는 경우에만 사용 가능) ANCOVA(공분산 분석) 모형에서 교호작용 효과의 각 수준에 대한 기울기를 비교할 수 있는 보고서를 생성합니다. 자세한 내용은 "**기울기 비교**"(349 페이지)에서 확인하십시오.

역추정 예측 하나 이상의 반응 값에 대해 설명 변수 값을 예측합니다. 자세한 내용은 "표준 최소 제곱 모형" 창의 "**역추정 예측**"(131 페이지)에서 확인하십시오.

주변 모형 추론 주변 예측값과 주변 잔차를 기반으로 그림을 생성합니다. 이러한 그림에는 임의 효과로 인한 변동이 표시됩니다. 자세한 내용은 "**주변 모형 추론**"(349 페이지)에서 확인하십시오.

조건부 모형 추론 조건부 예측값과 조건부 잔차를 기반으로 그림을 생성합니다. 이러한 그림에는 임의 효과가 고려된 후 남아 있는 변동이 표시됩니다. 자세한 내용은 "**조건부 모형 추론**"(351 페이지)에서 확인하십시오.

열 저장 다양한 모형 결과를 데이터 테이블의 열로 저장하는 옵션이 포함되어 있습니다. 자세한 내용은 "[열 저장](#)"(354 페이지)에서 확인하십시오.

모형 대화상자 현재 분석을 위해 완료된 "모형 적합" 시작 창을 엽니다. 자세한 내용은 "[모형 적합 시작 창](#)"(335 페이지)에서 확인하십시오.

다음 옵션에 대한 자세한 내용은 **JMP 사용**의에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

모형 보고서

"모형 보고서"에서 사용할 수 있는 보고서는 사용자가 수행하는 분석 유형에 따라 결정됩니다. 이러한 보고서 중 일부는 기본적으로 표시됩니다.

적합 통계량 모형 적합 통계량 보고서를 표시합니다. 자세한 내용은 "[적합 통계량](#)"(343 페이지)에서 확인하십시오.

임의 효과 공분산 모수 추정값 임의 효과 공분산 모수 추정값 보고서를 표시합니다. 이 보고서는 시작 창에서 임의 효과를 지정한 경우 나타납니다. 자세한 내용은 "[임의 효과 공분산 모수 추정값](#)"(345 페이지)에서 확인하십시오.

고정 효과 모수 추정값 고정 효과 모수 추정값 보고서를 표시합니다. 자세한 내용은 "[고정 효과 모수 추정값](#)"(345 페이지)에서 확인하십시오.

반복 효과 공분산 모수 추정값 반복 효과 공분산 모수 추정값 보고서를 표시합니다. 이 보고서는 시작 창에서 반복 효과를 지정한 경우 나타납니다. 자세한 내용은 "[반복 효과 공분산 모수 추정값](#)"(346 페이지)에서 확인하십시오.

표시자 파라미터화 추정값 (고정 효과 중 명목형 열이 있는 경우에만 사용 가능) "표시 함수 파라미터화" 보고서를 표시합니다. 이 보고서는 명목형 고정 효과 열이 표시자 (SAS GLM) 파라미터화를 사용하여 코드화되고 연속형으로 처리되는 모형을 기반으로 고정 효과에 대한 모수 추정값을 제공합니다. 순서형 열은 일반적인 JMP 코딩 체계를 사용하여 코드화됩니다. SAS GLM 및 JMP 코딩 체계에 대한 설명은 "통계 상세 정보" 부록의 "[요인 모형](#)"(492 페이지)에서 확인하십시오.

주의 : "표시 함수 파라미터화" 보고서에 제공된 표준 오차, t비 및 기타 결과는 "모수 추정값" 보고서의 값과 다릅니다. 이는 추정하는 모수가 서로 다르기 때문입니다.

랜덤 계수 랜덤 계수 보고서를 표시합니다 . 이 보고서는 시작 창에서 임의 효과를 지정한 경우 나타납니다 . 자세한 내용은 "**랜덤 계수**"(347 페이지) 에서 확인하십시오 .

임의 효과 예측 임의 효과 예측 보고서를 표시합니다 . 이 보고서는 시작 창에서 임의 효과를 지정한 경우 나타납니다 . 자세한 내용은 "**임의 효과 예측**"(347 페이지) 에서 확인하십시오 .

고정 효과 검정 (고정 효과를 하나 이상 포함하는 모형에만 사용 가능) 고정 효과 검정을 표시합니다 . 이 보고서는 시작 창에서 고정 효과를 지정한 경우 나타납니다 . 자세한 내용은 "**고정 효과 검정**"(347 페이지) 에서 확인하십시오 .

순차 검정 (고정 효과를 하나 이상 포함하는 모형에만 사용 가능) 모형에 효과가 순차적으로 추가될 때 제곱합이 포함된 " 순차 (유형 1) 검정 " 보고서를 표시합니다 . 순차 제곱합을 기반으로 F- 검정을 수행합니다 . 자세한 내용은 "**순차 검정**"(348 페이지) 에서 확인하십시오 .

JMP PRO 적합 통계량

"적합 통계량" 보고서에는 모형 비교에 사용되는 통계량이 제공됩니다 . 모든 적합 통계량은 값이 작을수록 좋습니다 . 한 모형이 다른 모형에 포함된 경우 두 모형 간의 가능도비 검정을 수행할 수 있습니다 . 그렇지 않은 경우 가능도를 비교하면 유용할 수 있습니다 . 예는 "**공간 구조 모형 적합**"(376 페이지)에서 확인하십시오 .

적합 통계량 보고서 설명에서는 다음과 같은 표기를 사용합니다 .

- 혼합 모형 지정

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

여기서 $\mathbf{y}$ 는 관측값의 $n \times 1$ 벡터 , $\boldsymbol{\beta}$ 는 고정 효과 모수의 벡터 , $\boldsymbol{\gamma}$ 는 임의 효과 모수의 벡터 , $\boldsymbol{\varepsilon}$ 는 오차의 벡터입니다 .

- $\boldsymbol{\gamma}$ 및 $\boldsymbol{\varepsilon}$ 벡터는 다음과 같은 다변량 정규 분포를 따른다고 가정합니다 .

$$E \begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

및

$$\text{Var} \begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}$$

- 이러한 가정을 통해 $\mathbf{y}$ 의 분산은 다음과 같이 계산됩니다 .

$$\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$$

적합 통계량 보고서 설명

-2* 잔차 로그 가능도 음의 잔차 로그 가능도 (목적 함수) 의 두 배인 최종 계산 결과입니다 .

$$-2\log\text{likelihood}_{\mathbf{R}}(\mathbf{G}, \mathbf{R}) = \log|\mathbf{V}| + \log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| + \mathbf{r}'\mathbf{V}^{-1}\mathbf{r} + (n-p)\log(2\pi)$$

다음은 각 요소에 대한 설명입니다.

$$\mathbf{r} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y})$$

p 는 X 의 계수입니다. 모형의 고정 효과 부분이 동일한 모형을 비교하는 경우에만 잔차 가능도를 사용합니다. 자세한 내용은 "통계 상세 정보" 부록의 "[가능도, AICc 및 BIC](#)"(514 페이지)에서 확인하십시오.

-2* 로그 가능도 음의 로그 가능도 함수의 두 배인 계산 결과입니다. 자세한 내용은 "통계 상세 정보" 부록의 "[가능도, AICc 및 BIC](#)"(514 페이지)에서 확인하십시오.

모형에서 고정 효과, 임의 효과 및 반복 효과가 서로 다른 모형을 비교하는 경우 로그 가능도를 사용합니다.

AICc 수정된 Akaike 정보 기준입니다. 자세한 내용은 "통계 상세 정보" 부록의 "[가능도, AICc 및 BIC](#)"(514 페이지)에서 확인하십시오.

BIC Bayes 정보 기준 자세한 내용은 "통계 상세 정보" 부록의 "[가능도, AICc 및 BIC](#)"(514 페이지)에서 확인하십시오.

JMP PRO 수렴 스코어 검정

모형 수렴에 문제가 있는 경우 적합 통계량 아래에 경고 메시지가 표시됩니다. 그림 8.11에서는 수렴 문제의 원인과 가능한 해결 방법을 제안하는 경고를 보여 줍니다. 또한 최종 반복에서 상대 그래디언트 검정도 포함됩니다. 이 검정이 유의하지 않으면 모형이 올바르지만 수렴 기준에 완전히 도달하지 못할 수 있습니다. 이 경우 모형과 결과를 신중하게 사용해야 합니다. 자세한 내용은 "[수렴 스코어 검정](#)"(391 페이지)에서 확인하십시오.

그림 8.11 수렴 스코어 검정

적합 통계량			
-2*잔차 로그 가능도	213.70579		
-2*로그 가능도	198.9947		
AICc	223.48449		
BIC	239.93814		
경고: 모수 추정 알고리즘이 수렴되지 않았습니다.			
상대 그래디언트 기준: 0.048			
수렴 점수 검정			
카이제곱	DF	Prob>ChiSq	
0.048	9	1.0000	
가능한 원인:			
- 모형이 적절하지 않습니다. 고정 효과, 임의 효과 또는 반복 성분이 올바르지 않을 수 있습니다. - 데이터 집합 크기에 비해 모형이 너무 복잡합니다. - 데이터에 문제가 되는 이상치가 있을 수 있습니다. - 데이터의 불균형이 심할 수 있습니다. - 입력 열 사이에 선형 또는 선형에 근사한 종속성이 있습니다. - 반응이 편중되었거나, 이산형이거나, 정규 분포에 근사하지 않습니다.			
가능한 해:			
- 수렴 점수 검정 p 값이 .05보다 클 경우 모형 및 결과를 주의해서 사용하십시오. - 더 단순한 대체 모형을 고려해 보십시오. - 이상치 및 범주형 요인의 드문 수준을 제거하십시오. - 거의 정규 분포에 가깝게 되도록 반응을 변환하십시오. - 최대 반복 수를 늘리십시오.			

JMP PRO 임의 효과 공분산 모수 추정값

"임의 효과 공분산 모수 추정값" 보고서에는 모형에 지정한 임의 효과의 공분산 모수에 대한 상세 정보가 제공됩니다.

공분산 모수 모형에 지정한 임의 효과의 모든 공분산 모수를 나열합니다.

참고 : 모형에 분산 성분만 포함된 경우에는 이 열에 "분산 성분" 라벨이 지정됩니다.

개체 공분산 행렬의 대각 블록을 구성한 개체를 나열합니다.

추정값 효과에 대한 추정 분산 또는 공분산 성분을 제공합니다.

참고 : 모형이 REML 모형과 동등한 경우 총 공분산 모수에 대한 행이 테이블에 추가됩니다. 총 공분산 성분의 추정값은 양의 분산 성분만 합한 값입니다.

표준 오차 공분산 성분 추정값에 대한 표준 오차를 제공합니다.

95% 하한 공분산 성분에 대한 95% 신뢰 하한을 제공합니다. 자세한 내용은 "[분산 성분에 대한 신뢰 구간](#)"(345 페이지)에서 확인하십시오.

95% 상한 공분산 성분에 대한 95% 신뢰 상한을 제공합니다. 자세한 내용은 "[분산 성분에 대한 신뢰 구간](#)"(345 페이지)에서 확인하십시오.

Wald p 값 공분산 모수가 0 이라는 검정에 대한 p 값을 제공합니다. 이 열은 "모형 적합" 시작 창에서 "비유계 분산 성분"을 선택한 경우에만 나타납니다.

JMP PRO 분산 성분에 대한 신뢰 구간

신뢰 한계를 계산하는 데 사용되는 방법은 "모형 적합" 시작 창에서 "비유계 분산 성분"을 선택했는지 여부에 따라 다릅니다. "비유계 분산 성분"은 기본적으로 선택되어 있습니다.

- "비유계 분산 성분"을 선택한 경우 Wald 기반 신뢰 구간이 계산됩니다. 이러한 구간은 점근적으로 유효하지만 표본이 작은 경우에는 신뢰할 수 없습니다. 구간이 더 넓기 때문에 추정값이 0 과 유의한 차이가 없다고 잘못 생각할 수 있습니다.
- "비유계 분산 성분"을 선택하지 않은 경우, 즉 모수의 하한 제약 조건이 0 인 경우에는 Satterthwaite 근사가 사용됩니다(Satterthwaite 1946). 신뢰 구간의 하한도 0으로 제한됩니다.

JMP PRO 고정 효과 모수 추정값

"고정 효과 모수 추정값" 보고서에는 모형에 지정한 고정 효과 모수에 대한 상세 정보가 제공됩니다. 각 모수에 대해 다음과 같은 상세 정보가 제공됩니다.

- 추정값
- 표준 오차
- 추정값이 0 이라는 가설에 대한 t -검정

- 추정값에 대한 95% 신뢰 구간

"고정 효과 모수 추정값" 보고서에는 다음 열이 포함됩니다.

항 추정된 모수에 해당하는 모형 항을 제공합니다. "모형 적합" 시작 창에서 "절편 없음" 옵션을 선택한 경우를 제외하고 첫 번째 항은 항상 절편입니다. 고차항의 일부인 연속형 열은 기본적으로 중심화됩니다. 명목형 또는 순서형 효과는 대괄호 안에 수준 값이 포함된 상태로 나타납니다. 명목형 및 순서형 항의 코딩에 대한 자세한 내용은 "[요인 모형](#)"(492 페이지)에서 확인하십시오.

참고 : 연속형 열이 임의 효과에 포함된 경우에는 "모형 규격"의 빨간색 삼각형 메뉴에서 "중심화 다항식" 옵션이 선택되어 있더라도 해당 열이 중심화되지 않습니다.

추정값 각 항에 대한 모수 추정값을 제공합니다. 이 값은 모형에 있는 항 계수의 추정값입니다.

표준 오차 모수 추정값에 대한 표준 오차 추정값을 제공합니다.

DFDen 효과 검정에 대한 분모 자유도, 즉 오차 자유도를 제공합니다. DFDen은 Kenward-Roger 1 차 근사를 사용하여 계산됩니다. 자세한 내용은 "[Kackar-Harville 수정](#)"(401 페이지)에서 확인하십시오.

t 비 모수의 실제 값이 0 인지 여부를 검정합니다. t 비는 표준 오차에 대한 추정값의 비율입니다. 모형에 대한 일반적인 가정이 주어지면 t 비는 귀무가설 하에서 스튜던트 t 분포를 따릅니다.

Prob>|t| t 비의 양측 검정에 대한 p 값을 나열합니다.

95% 하한 모수에 대한 95% 신뢰 하한을 표시합니다.

95% 상한 모수에 대한 95% 신뢰 상한을 표시합니다.

JMP PRO 반복 효과 공분산 모수 추정값

"반복 효과 공분산 모수 추정값" 보고서에는 모형에 지정한 반복 효과의 공분산 모수에 대한 상세 정보가 제공됩니다. 여기에는 각 모수에 대한 추정값, 표준 오차 및 95% 신뢰경계가 포함됩니다. 등방성 공간 모형의 경우 범위, 너겟 및 문턱(sill) 값을 사용하여 공분산 모수 추정값을 해석합니다. 자세한 내용은 "[변동도](#)"(399 페이지)에서 확인하십시오.

참고 : 분산은 효과 자체의 공분산입니다.

"반복 효과 공분산 모수 추정값" 보고서에는 다음 열이 포함됩니다.

공분산 모수 모형의 반복 효과에 대한 모든 공분산 모수를 나열합니다.

추정값 효과에 대한 추정 분산 또는 공분산 성분을 제공합니다.

표준 오차 분산 또는 공분산 성분 추정값에 대한 표준 오차를 제공합니다.

95% 하한 공분산 성분에 대한 95% 신뢰 하한을 제공합니다. 자세한 내용은 "[분산 성분에 대한 신뢰 구간](#)"(345 페이지)에서 확인하십시오.

95% 상한 공분산 성분에 대한 95% 신뢰 상한을 제공합니다. 자세한 내용은 "**분산 성분에 대한 신뢰 구간**"(345 페이지)에서 확인하십시오.

JMP PRO 랜덤 계수

혼합 모형 분석법을 사용하면 각 임의 효과에 대해 추정된 계수를 보여 주는 보고서와 공분산 추정값 행렬을 보여 주는 보고서가 제공됩니다. 계수 보고서의 각 행은 임의 효과의 한 수준에 해당합니다. 임의 효과의 수준과 관련된 모든 계수 추정값이 행에 표시됩니다. 랜덤 계수 추정값은 고정 효과 추정값과 함께 사용되어 임의 효과의 특정 수준에 대한 예측을 생성합니다.

JMP PRO 임의 효과 예측

이 보고서에는 모형의 각 임의 효과에 대해 BLUP(최량 선형 비편향 예측 변수)라는 추정값, 표준 오차, 자유도 및 Satterthwaite 기반 신뢰 구간이 제공됩니다.

표준 오차를 추정하려면 BLUP 공분산 행렬을 계산해야 하며 이 작업은 시간이 많이 걸릴 수 있습니다. 계산 시간이 길어지면 진행률 표시줄이 나타납니다.

JMP PRO 고정 효과 검정

이 보고서에는 모형의 각 고정 효과에 대한 유의성 검정이 표시됩니다. 주어진 효과에 대한 검정은 해당 효과와 관련된 모든 모수가 0이라는 귀무가설을 검정합니다. 단일 연속형 설명 변수의 경우 효과에 하나의 모수만 있을 수 있습니다. 이 경우 검정은 "고정 효과 모수 추정값" 보고서의 해당 행에 대한 t -검정과 동등합니다. 명목형 또는 순서형 효과에는 수준 수에 따라 관련된 모수가 여러 개 있을 수 있습니다. 이러한 효과에 대한 효과 검정은 관련된 모든 모수가 0인지 여부를 검정합니다.

"고정 효과 검정" 보고서에는 다음 열이 포함됩니다.

소스 모형의 고정 효과를 나열합니다.

모수 수 효과와 관련된 모수의 수를 표시합니다. 연속형 효과에는 하나의 모수가 있습니다. 명목형 또는 순서형 효과에 대한 모수의 수는 수준 수에서 1을 뺀 값입니다. 교차 효과에 대한 모수의 수는 각 개별 효과에 대한 모수의 수를 곱한 값입니다.

DFNum 효과 검정에 대한 분자 자유도를 표시합니다.

DFDen 효과 검정에 대한 분모 자유도, 즉 오차 자유도를 표시합니다. DFDen은 Kenward-Roger 1 차 근사를 사용하여 계산됩니다. 자세한 내용은 "**Kacker-Harville 수정**"(401 페이지)에서 확인하십시오.

F 비 효과가 0인지 검정하기 위한 계산된 F 비를 제공합니다.

Prob > F 효과 검정을 위한 p 값을 제공합니다.

JMP PRO 순차 검정

이 보고서에는 고정 효과에 대한 순차(유형 1) 검정이 표시됩니다. 모형에 효과가 순차적으로 추가될 때 제곱합이 보고서에 포함됩니다. 입력 순서는 "모형 적합" 시작 창의 "모형 효과 생성" 목록에 나타나는 효과의 순서로 정의됩니다.

순차 검정의 기반이 되는 제곱합을 **제1종 제곱합**이라고도 합니다. 이 값은 지정된 효과 입력 순서에 따라 모형을 적합시켜 계산됩니다. 특정 효과를 예로 들겠습니다. 해당 효과 전에 입력한 모든 효과를 포함하는 모형에 대한 모형 제곱합을 계산합니다. 그런 다음 이러한 효과와 지정된 효과를 모두 포함하는 모형에 대한 모형 제곱합을 계산합니다. 지정된 효과에 대한 순차 제곱합은 모형 제곱합의 증가입니다.

다음과 같은 경우 순차 검정이 적절합니다.

- 적절한 순서로 지정된 분산 모형의 균형 분석 (즉, 효과 목록에서 주효과 다음에 이원 교호작용이 나오는 순서)
- 적절한 순서로 지정된 순수 내포 모형
- 적절한 순서로 지정된 다항 회귀 모형

" 순차 (유형 1) 검정 " 보고서에는 다음 열이 포함됩니다.

소스 모형의 고정 효과를 나열합니다.

모수 수 효과와 관련된 모수의 수를 표시합니다. 연속형 효과에는 하나의 모수가 있습니다. 명목형 또는 순서형 효과에 대한 모수의 수는 수준 수에서 1 을 뺀 값입니다. 교차 효과에 대한 모수의 수는 각 개별 효과에 대한 모수의 수를 곱한 값입니다.

DFNum 효과 검정에 대한 분자 자유도를 표시합니다.

DFDen 효과 검정에 대한 분모 자유도, 즉 오차 자유도를 표시합니다. DFDen은 Kenward-Roger 1 차 근사를 사용하여 계산됩니다. 자세한 내용은 "[Kackar-Harville 수정](#)"(401 페이지) 에서 확인하십시오.

F 비 효과가 0 인지 검정하기 위한 계산된 F 비를 제공합니다.

Prob > F 효과 검정을 위한 p 값을 제공합니다.

JMP PRO 다중 비교

"다중 비교" 옵션은 주효과와 교호작용 효과의 최소 제곱 평균을 비교하는 다양한 방법을 제공합니다. 다중 비교 옵션에 대한 자세한 내용은 "표준 최소 제곱 모형" 창의 "[다중 비교](#)"(117페이지)에서 확인하십시오. 혼합 모형 예는 "[6월 \(June\) 의 모든 치료 \(Treatment\) 비교](#)"(367페이지) 및 "[분할구 예](#)"(370페이지)에서 확인하십시오.

모형의 고정 효과 부분만 다중 비교에 사용됩니다. "다중 비교" 보고서에는 최소 제곱 평균 추정값, 표준 오차, 효과 없음에 대한 t -검정 및 95% 신뢰 구간이 표시됩니다. 이 보고서 다음에는 사용자가 선택한 여러 비교 검정이 나옵니다. "전체 쌍별 비교" 보고서에는 동등성 검정이 제공됩니다.

JMP PRO 기울기 비교

"기울기 비교" 옵션은 명목형 항과 연속형 항이 각각 하나이고 고정 효과에 대한 교호작용 효과가 있을 때 나타납니다. 이 옵션은 ANCOVA(공분산 분석) 모형에서 기울기를 비교할 수 있는 보고서를 생성합니다. 이 보고서에서는 교호작용 효과의 각 수준에 대한 기울기를 전체 기울기와 비교합니다. 비교에는 ANOM(평균 분석)과 전체 평균이 함께 사용됩니다. ANOM(평균 분석) 보고서에 대한 자세한 내용은 "표준 최소 제곱 모형" 장의 "**전체 평균과 비교**"(121 페이지)에서 확인하십시오.

전체 평균 기울기는 기울기의 가중 평균이며 이때 가중치는 기울기 추정값의 분산에 반비례합니다. 이러한 분산은 "전체 평균 기울기와의 차이" 테이블에서 "표준 오차" 열의 제곱 값입니다.

JMP PRO 주변 모형 추론

주변 모형 그림은 주변 예측값과 주변 잔차를 기반으로 합니다.

실제값 대 예측값 그림 임의의 효과를 고려하지 않고 실제값 대 모형에서 예측한 값을 그림에 표시합니다. "실제값 대 예측값 그림"은 기본적으로 나타납니다. 자세한 내용은 "**실제값 대 예측값 그림**"(349 페이지)에서 확인하십시오.

잔차 그림 임의의 효과를 고려하지 않고 모형 적합을 평가하는 잔차 그림을 제공합니다. 자세한 내용은 "**잔차 그림**"(349 페이지)에서 확인하십시오.

프로파일러, 등고선 프로파일러, 혼합물 프로파일러, 표면 프로파일러 임의의 효과를 고려하지 않고 반응과 모형 항 간의 관계를 검토할 수 있는 프로파일러를 제공합니다. 자세한 내용은 "**Profilers**"(350 페이지)에서 확인하십시오.

JMP PRO 실제값 대 예측값 그림

"실제값 대 예측값 그림"은 기본적으로 나타납니다. 이 그림을 통해 임의의 효과로 인한 변동을 반영하는 모형 적합을 시각적으로 평가할 수 있습니다. 여기에는 Y 의 주변 예측값에 대해 Y 의 관측값이 표시됩니다. 이러한 값은 열 저장 > 예측 계산식을 선택하여 구한 예측값입니다.

선형 혼합 모형은 $E[Y|\gamma] = X\beta + Z\gamma$ 로 나타냅니다. 여기서 β 는 고정 효과 계수의 벡터이고 γ 는 임의의 효과 계수의 벡터입니다. 주변 예측은 $X\beta$ 로 지정된 예측 모형의 고정 효과 부분에서 구한 예측값입니다.

JMP PRO 잔차 그림

주변 잔차는 고정 효과의 적합만 기반으로 하는 예측 오차를 반영합니다. 주변 잔차는 실제값과 열 저장 > 예측 계산식을 선택하여 구한 예측값 간의 차이입니다.

선형 혼합 모형은 $E[Y|\gamma] = X\beta + Z\gamma$ 로 나타냅니다. 여기서 β 는 고정 효과 계수의 벡터이고 γ 는 임의 효과 계수의 벡터입니다. 주변 잔차는 예측 모형의 고정 효과 부분에서 구한 잔차입니다.

$$r = Y - X\hat{\beta}$$

"잔차 그림" 옵션은 모형 적합을 평가할 수 있는 세 가지 시각적 방법을 제공합니다.

잔차 대 예측값 그림 Y 의 예측값에 대해 그림에 표시된 잔차를 보여 줍니다. 일반적으로 0 주위에 무작위로 흩어진 잔차 값을 원합니다.

잔차 분위수 그림 표준 정규 분포의 분위수에 대해 그림에 표시된 잔차의 분위수를 보여 줍니다. 잔차의 막대 차트도 표시됩니다. 잔차가 정규 분포를 따르는 경우 정규 분위수 그림의 점은 거의 빨간색 주대각선을 따라 분포되어야 합니다. 이 유형의 그림을 분위수 - 분위수 그림 또는 Q-Q 그림이라고도 합니다. 정규 분위수 그림에는 Lilliefors 신뢰경계도 표시됩니다 (Conover 1999).

잔차 대 행 번호 그림 행 번호에 대해 그림에 표시된 잔차를 보여 줍니다. 이 그림은 관측값의 행 순서로 인한 패턴을 탐지하는 데 도움이 될 수 있습니다.

JMP[®] Profilers

"주변 모형 프로파일러"의 그림은 주변 예측값을 기반으로 합니다. 이러한 값은 열 저장 > 예측 계산식을 선택하여 구한 예측값입니다.

선형 혼합 모형은 $E[Y|\gamma] = X\beta + Z\gamma$ 로 나타냅니다. 여기서 β 는 고정 효과 계수의 벡터이고 γ 는 임의 효과 계수의 벡터입니다. 주변 예측은 $X\hat{\beta}$ 로 지정된 예측 모형의 고정 효과 부분에서 구한 예측값입니다.

참고: 주변 모형 프로파일러 그림에서는 고정 효과의 개별 설정에 대한 예측만 보여 줍니다. "프로파일러"에는 고정 효과에 해당하는 셀만 표시됩니다. 임의 효과를 표시할 수 있는 다른 프로파일러에서는 고정 효과의 설정만으로 예측값을 결정합니다.

다음과 같은 네 가지 유형의 프로파일러가 제공됩니다.

- 프로파일러
- 등고선 프로파일러
- 혼합물 프로파일러
- 표면 프로파일러

적합 중인 모형에 적절한 옵션이 활성화됩니다. 프로파일러의 예는 그림 8.21에서 확인하십시오. 표면 프로파일러의 예는 그림 8.42에서 확인하십시오. 표면 프로파일러에 대한 자세한 내용은 Profilers의 에서 확인하십시오.

JMP PRO 조건부 모형 추론

조건부 모형 진단 그림은 조건부 잔차를 기반으로 합니다. 조건부 잔차는 고정 효과와 임의 효과
가 모두 적합된 후의 예측 오차를 반영합니다.

실제값 대 조건부 예측값 그림 임의 효과를 고려하여 실제값 대 모형에서 예측한 값을 그림에 표
시합니다. 임의 효과가 있는 경우 "실제값 대 조건부 예측값 그림"이 기본적으로 나타납니
다. 자세한 내용은 "실제값 대 조건부 예측값 그림"(351 페이지)에서 확인하십시오.

조건부 잔차 그림 임의 효과를 고려하여 모형 적합을 평가하는 잔차 그림을 제공합니다. 자세한
내용은 "조건부 잔차 그림"(351 페이지)에서 확인하십시오.

조건부 프로파일러, 조건부 등고선 프로파일러, 조건부 혼합물 프로파일러, 조건부 표면 프로파일러
임의 효과를 고려하여 반응과 모형 항 간의 관계를 검토할 수 있는 프로파일러를 제공합니다.
자세한 내용은 "조건부 프로파일러"(352 페이지)에서 확인하십시오.

변동도 관측값 간의 거리가 증가할 때 공분산의 변화를 보여 주는 변동도 그림을 제공합니다. "잔
차" 구조를 선택하면 시간 또는 공간 좌표로 사용할 열을 선택할 수 있습니다. 자세한 내용
은 "변동도"(352 페이지)에서 확인하십시오.

JMP PRO 실제값 대 조건부 예측값 그림

"실제값 대 조건부 예측값 그림"은 기본적으로 나타납니다. 이 그림을 통해 임의 효과로 인한 변
동을 설명하는 모형 적합을 시각적으로 평가할 수 있습니다. 여기에는 Y 의 조건부 예측값에 대
해 Y 의 관측값이 표시됩니다. 이러한 값은 열 저장 > 조건부 예측 계산식을 선택하여 구한 예측
값입니다.

선형 혼합 모형은 $E[Y|\gamma] = X\beta + Z\gamma$ 로 나타냅니다. 여기서 β 는 고정 효과 계수의 벡터이고 γ 는 임
의 효과 계수의 벡터입니다. 조건부 예측은 $X\hat{\beta} + Z\hat{\gamma}$ 로 지정된 모형에서 구한 예측입니다.

JMP PRO 조건부 잔차 그림

조건부 잔차는 고정 효과와 임의 효과 둘 다의 적합을 기반으로 하는 예측 오차를 반영합니다. 조
건부 잔차는 실제값과 열 저장 > 조건부 예측 계산식을 선택하여 구한 조건부 예측값 간의 차이
입니다.

선형 혼합 모형은 $E[Y|\gamma] = X\beta + Z\gamma$ 로 나타냅니다. 여기서 β 는 고정 효과 계수의 벡터이고 γ 는
임의 효과 계수의 벡터입니다. 조건부 잔차는 다음과 같이 계산됩니다.

$$r = Y - (X\hat{\beta} + Z\hat{\gamma})$$

"조건부 잔차 그림" 옵션은 모형 적합을 평가할 수 있는 세 가지 시각적 방법을 제공합니다.

조건부 잔차 대 예측값 그림 Y 의 조건부 예측값에 대해 그림에 표시된 조건부 잔차를 보여 줍니
다. 일반적으로 0 주위에 무작위로 흩어진 조건부 잔차를 원합니다.

조건부 잔차 분위수 그림 표준 정규 분포의 분위수에 대해 그림에 표시된 조건부 잔차의 분위수를 보여 줍니다. 조건부 잔차의 막대 차트도 표시됩니다. 조건부 잔차가 정규 분포를 따르는 경우 정규 분위수 그림의 점은 거의 빨간색 주대각선을 따라 분포되어야 합니다. 이 유형의 그림을 분위수-분위수 그림 또는 Q-Q 그림이라고도 합니다. 정규 분위수 그림에는 Lilliefors 신뢰경계도 표시됩니다 (Conover 1999).

조건부 잔차 대 행 번호 그림 행 번호에 대해 그림에 표시된 조건부 잔차를 보여 줍니다. 이 그림은 관측값의 행 순서로 인한 패턴을 탐지하는 데 도움이 될 수 있습니다.

JMP^{PRO} 조건부 프로파일러

조건부 모형 프로파일러 그림은 조건부 예측값을 기반으로 합니다. 이러한 값은 열 저장 > 조건부 예측 계산식을 선택하여 구한 예측값입니다.

선형 혼합 모형은 $E[Y|\gamma] = X\beta + Z\gamma$ 로 나타냅니다. 여기서 β 는 고정 효과 계수의 벡터이고 γ 는 임의 효과 계수의 벡터입니다. 조건부 예측은 $X\hat{\beta} + Z\hat{\gamma}$ 로 지정된 모형에서 구한 예측입니다.

다음과 같은 네 가지 유형의 프로파일러가 제공됩니다.

- 조건부 프로파일러
- 조건부 등고선 프로파일러
- 조건부 혼합물 프로파일러
- 조건부 표면 프로파일러

적합 중인 모형에 적절한 옵션이 활성화됩니다. 프로파일러의 예는 그림 8.21에서 확인하십시오. 표면 프로파일러의 예는 그림 8.42에서 확인하십시오. 표면 프로파일러에 대한 자세한 내용은 Profilers의 에서 확인하십시오.

JMP^{PRO} 변동도

변동도 그림은 관측값의 거리를 기반으로 관측값의 공간 또는 시간 상관관계를 설명합니다. 이 그림에서는 반분산이 거리 또는 시간의 함수로 표시됩니다. 이론적으로 반분산은 지정된 거리만큼 떨어진 위치에서 반응 값 간의 차이에 대한 분산의 2분의 1입니다. 지정된 거리에서 반분산과 상관관계는 반비례 관계를 나타냅니다. 지정된 거리에서 값 간의 상관관계가 작으면 해당 거리에서 관측값 간의 반분산이 큼니다.

모형 적합 창에서 등방성 반복 구조(AR(1), 공간 또는 공간(너갯 사용))를 지정하면 변동도 그림이 기본적으로 표시됩니다. 잔차 구조를 지정하는 경우 빨간색 삼각형 메뉴에서 "변동도" 옵션을 선택하면 변동도 계산에 사용할 연속형 열을 선택할 수 있습니다. 데이터의 공간 또는 시간 구조를 설명하는 열을 개수 제한 없이 포함할 수 있습니다.

초기 변동도 보고서에는 거리에 대한 경험적 반분산 그림이 표시됩니다. 배경 정보를 포함한 자세한 내용은 "Antedependent 공분산 구조"(397페이지)에서 확인하십시오.

등방성 구조에 대한 반분산 곡선

반분산 곡선은 AR(1), 거듭제곱, 지수, 가우스 및 구형과 같은 등방성 공분산 구조에 대해 제공됩니다. 공간 구조의 경우 너겟을 사용한 모형과 사용하지 않은 모형에 대해 곡선이 제공됩니다.

이론적 모형에 대한 곡선은 공분산 모수 추정값을 사용하여 적합됩니다. 기본 계산식은 Chilès and Delfiner 연구 자료(2012) 및 Cressie 연구 자료(1993)에서 확인하십시오.

이론적 모형을 사용하여 데이터가 선택된 등방성 구조에 맞는지 확인할 수 있습니다. 잔차 구조를 선택한 경우 경험적 변동도를 사용하여 데이터가 시간 또는 공간 구조를 나타내는지 확인할 수 있습니다. 점이 수평선을 따르면 거리에 따라 상관관계가 변하지 않으며 잔차 구조가 적절하다는 것을 의미합니다. 점이 패턴을 표시하는 경우 다양한 등방성 모형을 적합시키면 모형을 재적합하는 데 적절한 반복 구조가 제안될 수 있습니다.

JMP^{PRO} 너겟

너겟은 변동도 원점의 값 0에서 매우 작은 분리 거리의 반분산 값으로 수직 이동을 나타냅니다. 너겟을 사용한 변동도 모형은 원점에서 불연속성을 가집니다. 0 바로 위의 거리에 대한 이론적 곡선의 값이 너겟입니다.

JMP^{PRO} 변동도 옵션

AR(1) AR(1) 공분산 구조에 대한 변동도를 그림에 표시합니다.

공간 지수, 가우스, 거듭제곱 또는 구형 공분산 구조에 대한 변동도를 그림에 표시합니다.

공간 (너겟 사용) 너겟을 사용하여 지수, 가우스, 거듭제곱 또는 구형 공분산 구조에 대한 변동도를 그림에 표시합니다.

JMP^{PRO} 공분산 및 상관 행렬

"공분산 및 상관 행렬" 메뉴에는 모형과 관련된 공분산 행렬을 볼 수 있는 옵션이 포함되어 있습니다.

고정 효과의 공분산 모형의 고정 효과에 대한 공분산 행렬을 표시하거나 숨깁니다.

공분산 모수의 공분산 모형의 임의 효과에 대한 공분산 행렬을 표시하거나 숨깁니다. 행렬의 효과는 G 측 임의 효과, R 측 임의 효과, 잔차 효과의 순서로 정렬됩니다.

모든 모수의 공분산 모형의 모든 효과에 대한 공분산 행렬을 표시하거나 숨깁니다. 행렬의 효과는 고정 효과, G 측 임의 효과, R 측 임의 효과, 잔차 효과의 순서로 정렬됩니다.

고정 효과의 상관관계 모형의 고정 효과에 대한 상관 행렬을 표시하거나 숨깁니다.

JMP PRO 열 저장

"열 저장" 메뉴의 각 옵션은 하나 이상의 새 열을 데이터 테이블에 추가합니다.

예측 계산식 계산식과 주변 평균 예측값이 모두 포함된 예측 계산식 < 열 이름 > 이라는 새 열을 생성합니다. "예측" 열 특성이 추가되어 예측 소스를 나타냅니다. 자세한 내용은 "[주변 모형 추론](#)"(349 페이지)에서 확인하십시오.

예측값의 표준 오차 주변 예측 평균 반응에 대한 표준 오차가 포함된 예측 표준 오차 < 열 이름 > 이라는 새 열을 생성합니다.

평균 신뢰 구간 평균 < 열 이름 > 의 95% 하한 및 평균 < 열 이름 > 의 95% 상한이라는 두 개의 새 열을 생성합니다. 이 열에는 평균 반응에 대한 95% 신뢰 하한과 신뢰 상한이 포함되어 있습니다. 이 구간에는 추정의 변동이 포함되고 반응의 변동은 포함되지 않습니다.

개별값 신뢰 구간 (G 측 효과만 포함된 모형에 사용 가능) 개별값 < 열 이름 > 의 95% 하한과 개별값 < 열 이름 > 의 95% 상한이라는 두 개의 새 열을 생성합니다. 이 열에는 개별 반응 값에 대한 95% 신뢰 하한과 신뢰 상한이 포함되어 있습니다. 이 구간에는 반응과 추정의 변동이 모두 포함됩니다.

잔차 반응 관측값에서 주변 평균 예측값을 뺀 값이 포함된 잔차 < 열 이름 > 이라는 새 열을 생성합니다. 자세한 내용은 "[주변 모형 추론](#)"(349 페이지)에서 확인하십시오.

조건부 예측 계산식 계산식과 조건부 평균 예측값이 모두 포함된 조건부 예측 계산식 < 열 이름 > 이라는 새 열을 생성합니다. "예측" 열 특성이 추가되어 예측 소스를 나타냅니다. 자세한 내용은 "[조건부 프로파일러](#)"(352 페이지)에서 확인하십시오.

조건부 예측 표준 오차 조건부 예측 평균 반응에 대한 표준 오차가 포함된 조건부 예측 표준 오차 < 열 이름 > 이라는 새 열을 생성합니다.

조건부 평균 CI (G 측 효과가 포함된 모형에 사용 가능) 조건부 평균 < 열 이름 > 의 95% 하한 및 조건부 평균 < 열 이름 > 의 95% 상한이라는 두 개의 새 열을 생성합니다. 이 열에는 조건부 예측의 기대값에 대한 95% 신뢰 하한과 신뢰 상한이 포함되어 있습니다. 이 신뢰 구간에는 임의 효과가 있는 모형의 임의 효과 추정값이 포함됩니다. 자세한 내용은 "[조건부 모형 추론](#)"(351 페이지)에서 확인하십시오.

조건부 잔차 반응 관측값에서 조건부 평균 예측값을 뺀 값이 포함된 조건부 잔차 < 열 이름 > 이라는 새 열을 생성합니다. 자세한 내용은 "[조건부 모형 추론](#)"(351 페이지)에서 확인하십시오.

시뮬레이션 계산식 저장 (분산 성분 및 랜덤 계수 모형에만 사용 가능) 적합 모형에 대한 추정 모수를 사용하여 시뮬레이션 값을 생성하는 계산식이 포함된 열을 데이터 테이블에 저장합니다. 이 열은 "시뮬레이션" 유틸리티에서 "스위치 인할 열"로 사용할 수 있습니다. 자세한 내용은 기본 분석의에서 확인하십시오.

JMP PRO 혼합 모형 분석법의 추가 예

- "[반복 측정 예](#)"

- " 분할구 예 "
- " 공간 예 : 균일도 시험 "
- " 상관관계가 있는 반응의 예 "

JMP PRO 반복 측정 예

Cholesterol Stacked.jmp 샘플 데이터 테이블을 예로 들겠습니다. 대조 약물에 대해 두 가지 새 콜레스테롤 약물을 테스트하기 위한 연구가 수행되었습니다. 콜레스테롤 수치가 높은 20명의 환자를 각각 네 가지 치료(두 가지 실험 약물, 대조 약물, 플라시보)에 무작위로 할당했습니다. 연구 기간 동안 각 환자의 총 콜레스테롤 수치를 6번 측정했습니다(4월 1일, 5월 1일, 6월 1일 아침과 오후 한 번씩). 이 연구를 통해 신약이 콜레스테롤 수치를 낮추는 데 효과가 있는지, 그리고 시간과 치료 간에 상관관계가 있는지 여부에 관심이 있습니다.

JMP PRO 배경

일반적으로 이러한 설계를 분석하기 위해 다음과 같은 두 가지 방법이 사용되었습니다.

- MANOVA(다변량 분산 분석)
- 수정된 Huynh-Feldt(1976) 또는 Greenhouse-Geisser(1959) 연구 자료를 사용한 시간 단변량 ANOVA(분산 분석) 의 분할구

두 옵션 모두 모형 적합의 MANOVA 분석법을 통해 사용할 수 있습니다. 이러한 양극단 두 옵션은 공분산 구조를 모델링하기 위한 것입니다. MANOVA 분석에서는 모든 변수와 공분산이 개별적으로 추정되는 비정형 공분산 구조를 가정합니다. 시간 분석의 독립 분할구에서는 모든 오차가 독립적이라고 가정합니다. 이는 가우스 데이터에 대해 복합 대칭 공분산 구조를 가정하는 것에 해당합니다.

이 두 모형은 처리 효과에 대해 상당히 다른 결론을 내릴 수 있습니다. 복소 공분산 구조를 가정하면 데이터의 정보를 사용하여 공분산 모수를 추정합니다. 너무 많은 공분산 모수를 적합시키면 모형 과대적합이 발생할 위험이 있습니다. 반복 측정 데이터를 모델링하는 경우 이러한 문제를 상쇄하는 공분산 구조를 찾아야 합니다.

- 모형이 과대적합되면 덜 복잡한 공분산 구조를 가정할 때보다 차이를 감지하는 검정력이 더 작습니다.
- 모형이 과소적합되면 제 1 종 오류 제어가 손실됩니다. 경우에 따라 기각률이 부풀려질 수도 있습니다. 그렇지 않은 경우 분산 팽창으로 인해 기각률이 감소합니다.

JMP PRO 공분산 구조

혼합 모형 분석법은 다양한 공분산 구조를 적합시킵니다. 시간에 따른 반복 측정의 경우 대개 Toeplitz 공분산 구조와 AR(1)(1차 자기회귀) 공분산 구조 모두 적절한 상관 구조를 제공합니다. 이러한 구조는 모형을 과대적합시키지 않고도 상관관계가 있는 관측값을 허용합니다. AR(1)은 공통 분산 모수를 가정하는 반면, 이분산을 사용한 Toeplitz 공분산 행렬은 반복 측정 변수의 각

단위에 대해 고유한 분산을 추정합니다. 자세한 내용은 "[반복 공분산 구조 요구 사항](#)"(339 페이지)에서 확인하십시오.

이 예에서는 네 가지 공분산 구조를 적합시킵니다. 관측 시간 수 J 는 6입니다.

- "[공분산 구조: 비정형](#)"(356 페이지). 비정형 모형은 모든 공분산 모수 $(J(J+1)/2)$ 를 합하여 적합시킵니다. 이 예에서 모형은 21 개의 분산을 적합시킵니다.
- "[공분산 구조: 잔차](#)"(359 페이지). 잔차 모형은 일반적인 분산 성분 구조와 동등합니다. 이 예에서 모형은 두 개의 분산을 적합시킵니다.
- "[공분산 구조: Toeplitz](#)"(361 페이지). Toeplitz 모형은 $2J-1$ 개의 공분산 모수를 적합시킵니다. 이 예에서 모형은 11 개의 분산을 적합시킵니다.
- "[공분산 구조: AR\(1\)](#)"(363 페이지). 이 모형은 두 개의 공분산 모수를 적합시킵니다. 한 모수는 분산을 결정하고 다른 모수는 시간에 따라 공분산이 어떻게 변하는지 결정합니다.

AICc를 사용하여 모형 적합을 평가합니다. BIC 기준도 사용할 수 있습니다. 이 경우 두 기준에 의해 선택된 모형이 동일합니다. 최적의 공분산 구조를 선택한 후 추가 분석을 계속합니다.

- "[AR\(1\) 구조를 사용한 추가 분석](#)"(365 페이지)
- "[AR\(1\)에 대한 회귀 모형의 예](#)"(368 페이지)

팁: 이 예를 진행하는 동안 모형 적합 시작 창을 열어 두십시오.

JMP^{PRO} 데이터 구조

Cholesterol.jmp 데이터 테이블은 반복 측정 데이터를 기록하는 데 일반적으로 사용되는 형식으로 되어 있습니다. 혼합 모형 분석법을 사용하여 이러한 데이터를 분석하려면 각 콜레스테롤 측정값이 Cholesterol Stacked.jmp와 같이 고유의 행에 포함되어야 합니다. Cholesterol Stacked.jmp를 구성하기 위해 테이블 > 쌓기를 사용하여 Cholesterol.jmp의 데이터를 쌓았습니다.

쌓인 테이블의 Days 열은 계산식을 사용하여 생성되었습니다. Days 열은 콜레스테롤 수치를 측정한 시간(일 수)을 연구에 제공합니다. 모델링 유형은 연속형입니다. AR(1) 공분산 구조에는 반복 효과가 연속형이어야 하므로 이 유형이 필요합니다.

JMP^{PRO} 공분산 구조: 비정형

먼저 비정형 공분산 구조를 사용하여 모형을 적합시킵니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cholesterol Stacked.jmp를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. 다음 예에서 시작 창으로 돌아갈 수 있도록 **대화상자 열린 채 유지**를 선택합니다.
4. **Y**를 선택하고 **Y**를 클릭합니다.
5. "분석법" 목록에서 **혼합 모형**을 선택합니다.

6. Treatment, Month, 및 AM/PM 을 선택한 후 **매크로 > 완전 요인**을 선택합니다 .

그림 8.12 완료된 고정 효과 탭을 보여 주는 모형 적합 시작 창

7. **반복 구조** 탭을 선택합니다 .
8. " 구조 " 목록에서 **비정형**을 선택합니다 .
9. Time 을 선택하고 **반복**을 클릭합니다 . **반복** 열린 개체 내에서 반복 측정을 정의합니다 .
10. Patient 를 선택하고 **개체**를 클릭합니다 .

참고 : 비정형 공분산 모형에서는 반복 구조 변수가 중복 값을 가질 수 없습니다 . 이 예에서는 개체가 치료 (treatment) 에 내포되고 , 각 치료 내에서 1, 2, 3, 4, 5 값을 사용하여 환자 (patient) 에 번호가 매겨져 있다고 가정합니다 . 이 분석을 실행하면 경고가 표시됩니다 . 반복 변수의 각 값에 서로 다른 식별자를 지정하기 위해 환자 번호를 다시 매겨야 합니다 . 또는 데이터 테이블에 치료 내의 내포를 나타내는 열을 생성하고 이 효과를 개체로 입력해야 합니다 .

그림 8.13 완료된 반복 구조 탭을 보여 주는 모형 적합 시작 창

11. 실행을 클릭합니다.

그림 8.14 에 " 혼합 적합 " 보고서가 표시되어 있습니다 . AICc 또는 BIC 를 사용하여 세 가지 모형을 비교하려고 하므로 " 적합 통계량 " 보고서를 확인해야 합니다 . 비정형 모형의 AICc 는 703.84 입니다 .

" 반복 효과 공분산 모수 추정값 " 보고서에서는 21 개의 모든 공분산 모수에 대한 추정값을 보여 줍니다 . 예상대로 더 가까운 시간의 관측값이 더 멀리 떨어진 관측값보다 공분산이 큼니다 . 또한 시간에 따라 분산이 증가합니다 .

그림 8.14 혼합 적합 보고서 - 비정형 공분산 구조

혼합 적합					
실제값 대 예측값 그림					
적합 통계량					
-2*잔차 로그 가능도	554.98818				
-2*로그 가능도	557.89101				
AICc	703.83696				
BIC	773.32814				
반복 효과 공분산 모수 추정값					
반복 효과: Time					
개체: Patient					
공분산 모수	추정값	표준 오차	95% 하한	95% 상한	
본산(April AM)	18.725	6.6202872	5.7494754	31.700525	
공분산(April PM, April AM)	18.354884	6.6035621	5.4121399	31.297628	
본산(April PM)	19.268932	6.8125963	5.9164888	32.621376	
공분산(May AM, April AM)	9.2756709	8.4629182	-7.311344	25.862686	
공분산(May AM, April PM)	5.5074058	8.3704001	-10.89828	21.913089	
본산(May AM)	56.603347	20.012305	17.379949	95.826744	
공분산(May PM, April AM)	9.4147226	8.5038584	-7.252534	26.081979	
공분산(May PM, April PM)	6.6230805	8.4532303	-9.944946	23.191108	
공분산(May PM, May AM)	55.365523	19.83529	16.489068	94.241978	
본산(May PM)	57.058089	20.173081	17.519577	96.596602	
공분산(June AM, April AM)	1.1945478	8.6266098	-15.7133	18.102392	
공분산(June AM, April PM)	0.3183447	8.7461245	-16.82374	17.460434	
공분산(June AM, May AM)	1.106725	14.992149	-28.27735	30.490796	
공분산(June AM, May PM)	0.6455101	15.050552	-28.85303	30.14405	
본산(June AM)	63.512277	22.454981	19.501323	107.52323	
공분산(June PM, April AM)	1.5810194	8.7687993	-15.60551	18.76755	
공분산(June PM, April PM)	0.7647113	8.8882627	-16.65596	18.185386	
공분산(June PM, May AM)	0.9262595	15.232066	-28.92804	30.780561	
공분산(June PM, May PM)	0.6543895	15.292238	-29.31785	30.626625	
공분산(June PM, June AM)	63.878686	22.700344	19.386829	108.37054	
본산(June PM)	65.568482	23.181959	20.132677	111.00429	
고정 효과 모수 추정값					
고정 효과 검정					
소스	N모수	DFNum	DFDen	F 비	Prob > F
Treatment	3	3	16.0	274.96713	<.0001*
Month	2	2	15.0	340.48166	<.0001*
Treatment*Month	6	6	18.3	123.47461	<.0001*
AM/PM	1	1	16.0	360.93593	<.0001*
Treatment*AM/PM	3	3	16.0	0.6339843	0.6038
Month*AM/PM	2	2	15.0	1.1988247	0.3289
Treatment*Month*AM/PM	6	6	18.3	1.1642781	0.3671

JMP® 공분산 구조 : 잔차

잔차 공분산 구조는 분할구 모형을 적합시킬 때 적절합니다.

1. "공분산 구조: 비정형"(356 페이지)에 나오는 1 단계 ~ 7 단계를 완료합니다.
2. 반복 구조 탭의 "구조" 목록에서 잔차를 선택합니다.
3. 이전 예에서 계속하는 경우 Time 과 Patient 를 제거합니다.

그렇지 않으면 "잔차 반복 공분산 구조를 선택한 경우에는 반복 열 및 개체 열이 무시됩니다"라는 경고가 표시됩니다. 확인을 클릭하여 분석을 계속할 수도 있습니다.

4. 임의 효과 탭을 선택합니다.

5. Patient 를 선택하고 **추가**를 클릭합니다 .
6. " 임의 효과 " 영역에서 Patient 를 선택하고 Treatment 열을 선택한 후 **내포**를 클릭합니다 .

그림 8.15 완료된 임의 효과 탭을 보여 주는 모형 적합 시작 창

7. **실행**을 클릭합니다 .

그림 8.16 에 " 혼합 적합 " 보고서가 표시되어 있습니다. " 적합 통계량 " 보고서에 따르면 잔차 모형의 AICc 는 832.55 이며 이는 비정형 모형의 703.84 와 비교됩니다 .

" 임의 효과 공분산 모수 추정값 " 보고서에는 두 공분산 모수의 추정값이 표시됩니다 . 이러한 값은 Treatment 에 내포된 Patient 의 분산 및 잔차 분산의 추정값입니다 .

그림 8.17 완료된 반복 구조 탭을 보여 주는 모형 적합 시작 창

모형 규격

열 선택
7개 열
Patient
Y
Treatment
Time
Month
AM/PM
Days

역할 변수 선택
Y 선택적
기준 선택적

분석법: 혼합 모형
☒ 경계가 없는 분산 성분
도움말 실행
재호출 ☒ 대화상자 열린 채 유지
제거

모형 효과 생성
고정 효과 임의 효과 반복 구조
반복 공분산 구조
구조 테플리츠 이분산
반복 Time
개체 Patient 선택적

7. 실행을 클릭합니다.

그림 8.18 혼합 적합 보고서 - Toeplitz 이분산 구조

혼합 적합					
실제값 대 예측값 그림					
적합 통계량					
-2*잔차 로그 가능도	659.09915				
-2*로그 가능도	688.02972				
AICc	788.02972				
BIC	855.59193				
반복 효과 공분산 모수 추정값					
개체: Patient					
공분산 모수	추정값	표준 오차	95% 하한	95% 상한	
테플 상관(1)	0.6230225	0.0856377	0.4551756	0.7908693	
테플 상관(2)	0.2555783	0.1463264	-0.031216	0.5423728	
테플 상관(3)	0.0962006	0.1887883	-0.273818	0.466219	
테플 상관(4)	-0.1112	0.2248683	-0.551934	0.3295338	
테플 상관(5)	0.2427218	0.1955609	-0.140571	0.6260141	
본산(April AM)	17.552396	6.9782256	3.8753253	31.229467	
본산(April PM)	24.652754	8.3215651	8.3427857	40.962721	
본산(May AM)	73.950898	25.930509	23.128035	124.77376	
본산(May PM)	63.668022	21.766899	21.005684	106.33036	
본산(June AM)	69.45679	22.578117	25.204495	113.70909	
본산(June PM)	52.484909	19.171565	14.909332	90.060486	
고정 효과 모수 추정값					
고정 효과 검정					
소스	N모수	DFNum	DFDen	F 비	Prob > F
Treatment	3	3	14.1	230.25344	<.0001*
Month	2	2	23.7	363.8599	<.0001*
Treatment*Month	6	6	26.2	134.01838	<.0001*
AM/PM	1	1	5.2	106.73459	0.0001*
Treatment*AM/PM	3	3	5.2	0.1874794	0.9007
Month*AM/PM	2	2	26.9	0.036706	0.9640
Treatment*Month*AM/PM	6	6	34.6	0.0347301	0.9998

참고 : JMP 의 혼합 모형 분석법은 상관관계를 보고하는 반면 SAS 의 PROC MIXED 는 공분산을 보고합니다 .

"적합 통계량" 보고서에 따르면 이분산을 사용한 Toeplitz 모형의 AICc는 788.03입니다. 이 값을 잔차 모형의 832.55 및 비정형 모형의 703.84와 비교합니다.

Toeplitz 이분산 구조를 사용하는 경우 11개의 공분산 모수를 추정해야 합니다. 이러한 추정값은 "반복 효과 공분산 모수 추정값" 보고서에 표시됩니다. Toeplitz 상관 추정값 다음에 각 시점에 대한 분산 추정값이 나옵니다. 이 행렬이 파라미터화되는 방법에 대한 자세한 내용은 "[반복 추정](#)"(393 페이지)에서 확인하십시오.

JMP PRO 공분산 구조 : AR(1)

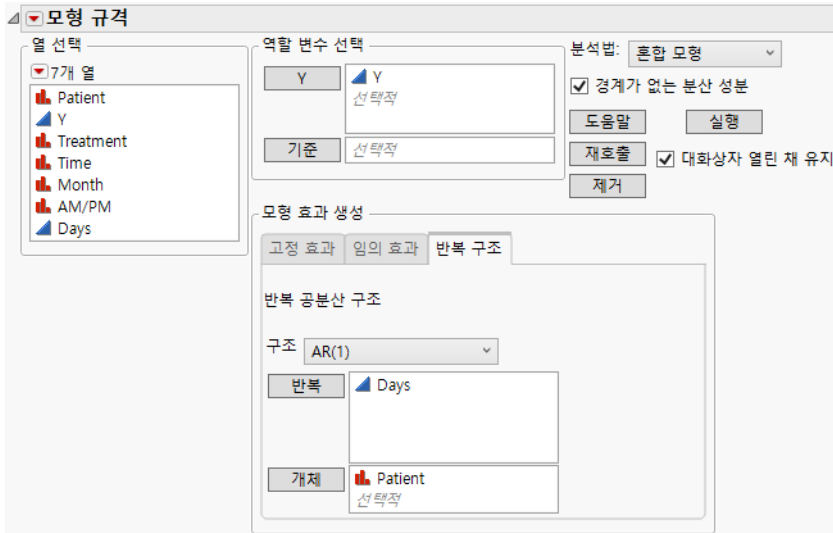
마지막으로 AR(1) 구조를 적합시킵니다 .

1. "[공분산 구조: 비정형](#)"(356 페이지) 에 나오는 1 단계 ~ 6 단계를 완료합니다 .
2. 이전 예에서 계속하는 경우 **반복** 상자의 Time 을 선택한 후 **제거**를 클릭합니다 .
AR(1) 을 사용하는 경우 반복 값에 대한 변수가 연속형이어야 합니다 .

3. " 구조 " 목록에서 **AR(1)** 을 선택합니다 .

4. Days 를 선택하고 **반복** 을 클릭합니다 .

그림 8.19 완료된 반복 구조 탭을 보여 주는 모형 적합 시작 창



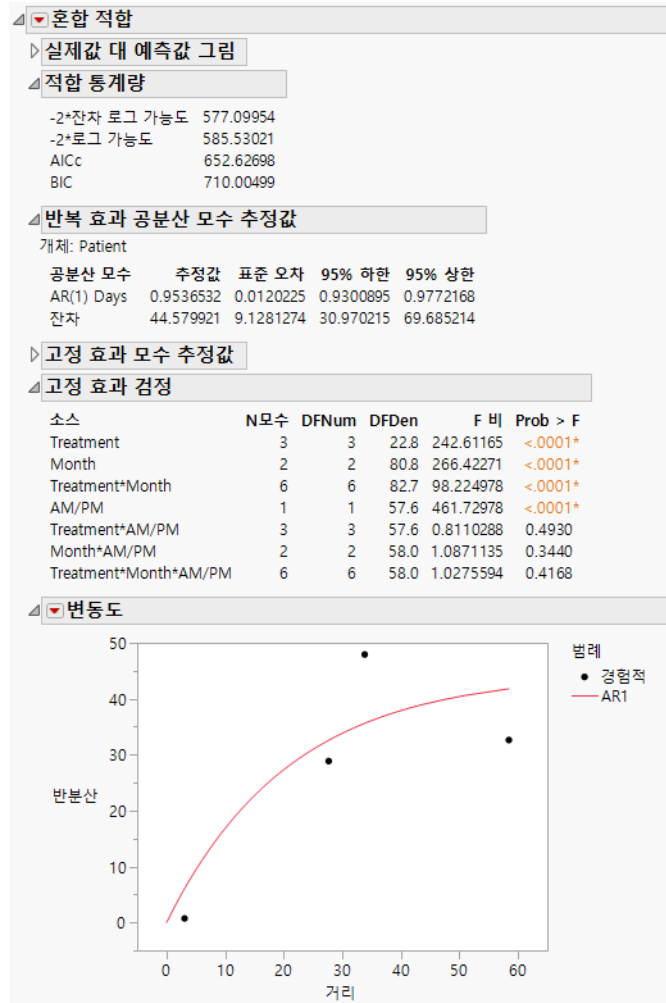
5. **실행** 을 클릭합니다 .

그림 8.20 에 " 혼합 적합 " 보고서가 표시되어 있습니다 . " 적합 통계량 " 보고서에 따르면 AR(1) 모형의 AICc 는 652.63 입니다 . 이 값을 잔차 모형의 832.55, 비정형 모형의 703.84 및 Toeplitz 이분산 모형의 788.03 과 비교합니다 . AICc 기준에 따르면 AR(1) 모형이 네 개 모형 중 가장 적합합니다 .

AR(1) 구조를 사용하는 경우 두 개의 공분산 모수를 추정해야 합니다 . 이러한 추정값은 " 반복 효과 공분산 모수 추정값 " 보고서에 표시됩니다 . AR(1) Days 모수 추정값은 AR(1) 구조의 상관 모수인 ρ 의 추정값입니다 .

변동도 그림에는 AR(1) 모형에 대한 경험적 반분산과 곡선이 표시됩니다 . Days 에는 0 이 아닌 값이 5 개뿐이므로 4 개의 거리 클래스만 가능하고 4 개의 점만 표시됩니다 . AR(1) 구조는 적절해 보입니다 . 다른 구조를 탐색하려면 " 변동도 " 의 빨간색 삼각형 메뉴에서 옵션을 선택합니다 . " 변동도 " 옵션에 대한 자세한 내용은 " 변동도 "(352 페이지) 에서 확인하십시오 .

그림 8.20 혼합 적합 보고서 - AR(1) 공분산 구조



JMP PRO AR(1) 구조를 사용한 추가 분석

AR(1) 모형이 최량 적합을 제공하므로 이 모형을 채택하여 분석을 계속합니다. "고정 효과 검정" 보고서에서는 AM/PM의 주효과뿐만 아니라 Treatment와 Month 간에 유의한 교호작용이 있음을 나타냅니다. 이제 이러한 유의 효과에 대해 살펴보겠습니다.

1. "혼합 적합"의 빨간색 삼각형을 클릭하고 **주변 모형 추론 > 프로파일러**를 선택합니다.
"주변 모형 프로파일러" 보고서 (그림 8.21)를 사용하면 Treatment, Month 및 AM/PM의 다양한 설정이 콜레스테롤 수준(Y)에 미치는 영향을 확인할 수 있습니다.
2. Month 그림에서 빨간색 수직 점선을 April에서 May로 드래그한 후 June으로 드래그합니다.

3 개월 동안 Y 에 대한 예측 AM 측정값이 277.4(April 평균) 에서 177.7(June 평균) 로 감소합니다 .

3. Treatment 그림에서 빨간색 수직 점선을 A 에서 B 로 드래그합니다 .

Month 그림의 선을 April에서 June으로 드래그하면 Treatment B의 경우 Y에 대한 예측 AM 평균이 276.8(April) 에서 191.2(June) 로 감소하는 것을 볼 수 있습니다 .

4. Treatment 그림에서 수직 점선을 Control 로 드래그한 후 Placebo 로 드래그합니다 .

Treatment를 Control 또는 Placebo로 설정하면 3개월 동안 거의 변화가 없습니다(그림 8.22). 다음으로 AM/PM 효과를 살펴봅니다 .

5. 빨간색 수직선을 드래그하여 Treatment 와 Month를 12 개의 모든 수준 조합으로 설정합니다 .

12 개의 모든 조합에 대해 예측 콜레스테롤 수준은 일관되게 아침보다 오후에 더 높아 주효과를 설명합니다 .

Treatment A는 Treatment B보다 5월(May)에 콜레스테롤 수치가 더 낮은 것으로 보입니다 . 이 효과가 유의하면 Treatment A가 B보다 빨리 반응한다고 할 수 있습니다 . 다음 섹션 "[6월 \(June\)의 모든 치료\(Treatment\) 비교](#)"(367 페이지)에서는 치료(Treatment)를 평가하는 방법을 보여 줍니다 .

그림 8.21 Treatment A 에 대한 주변 프로파일러 그림

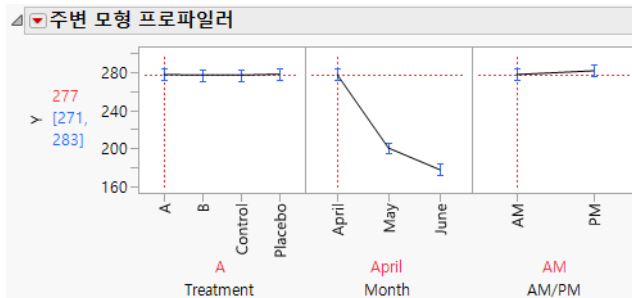
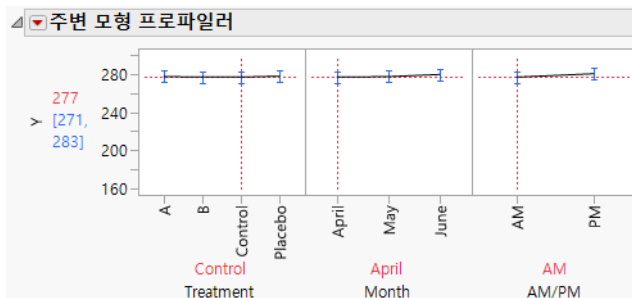


그림 8.22 Control 에 대한 주변 프로파일러 그림



6 월 (June) 의 모든 치료 (Treatment) 비교

연구는 4 월 (April), 5 월 (May), 6 월 (June) 에 걸쳐 수행됩니다 . 6 월 (June) 의 PM 측정에서 어떤 치료법이 차이가 있는지 알아보려고 합니다 .

1. " 혼합 적합 " 의 빨간색 삼각형을 클릭하고 **다중 비교**를 선택합니다 .
2. " 추정값 유형 " 에서 **사용자 정의 추정값**을 선택합니다 .
3. "Treatment 수준 선택 " 패널에서 네 가지 치료 유형을 모두 선택합니다 .
4. "Month 수준 선택 " 패널에서 "June" 을 선택합니다 .
5. "AM/PM 수준 선택 " 패널에서 "PM" 을 선택합니다 .
6. 추정값 추가를 클릭합니다 .
7. " 초기 비교 선택 " 목록에서 **전체 쌍별 비교 - Tukey HSD** 를 선택합니다 .

그림 8.23 완료된 다중 비교 창

추정값 유형

☐ 최소 제곱 평균 추정값
☒ 사용자 정의 추정값

Treatment 수준 선택

A
B
Control
Placebo

Month 수준 선택

April
May
June

AM/PM 수준 선택

AM
PM

필요한 경우 요인 설정을 선택하고 [추정값 추가] 버튼을 클릭하여 사용자 정의 추정값을 생성하십시오.

추정값 추가

비교를 위한 추정값

Treatment	Month	AM/PM
A	June	PM
B	June	PM
Control	June	PM
Placebo	June	PM

초기 비교 선택

☐ 전체 평균과 비교 - ANOM
☐ 대조군과 비교 - Dunnett
☒ 전체 쌍별 비교 - Tukey HSD
☐ 전체 쌍별 비교 - 스튜던트 t

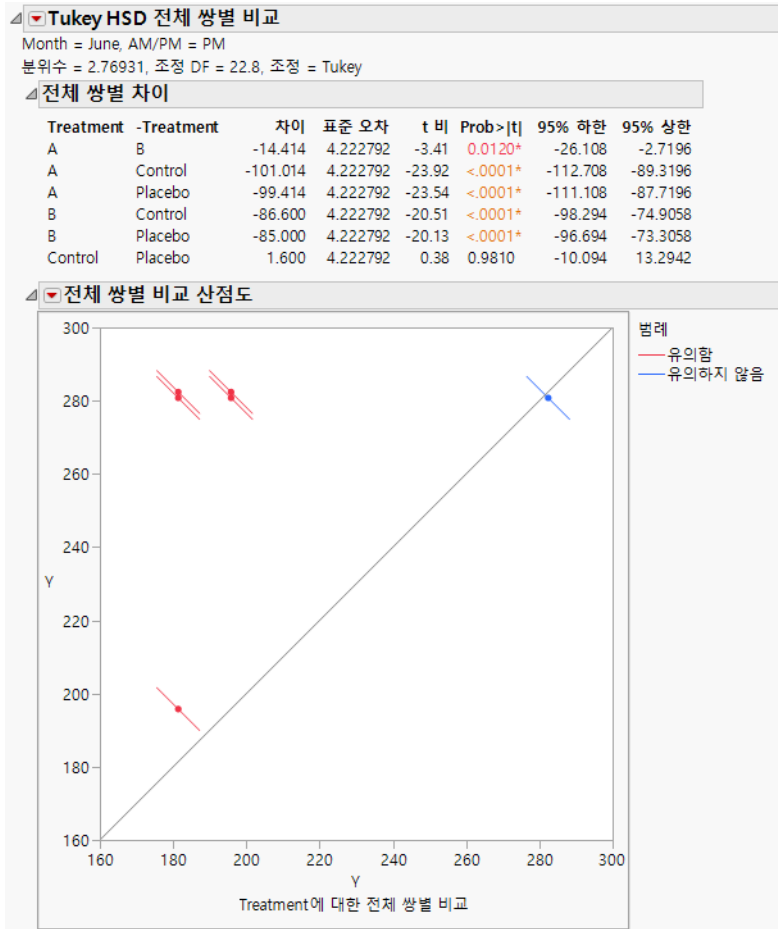
확인

취소

도움말

8. **확인**을 클릭합니다 .

그림 8.24 June PM 의 모든 Treatment 에 대한 Tukey HSD 전체 쌍별 비교 보고서



"Tukey HSD 전체 쌍별 비교" 보고서에는 "전체 쌍별 차이" 보고서와 "전체 쌍별 비교 산점도"가 표시됩니다. June PM 측정에 대해 Control과 Placebo를 제외한 모든 치료가 유의하게 다릅니다.

치료 A와 B 간의 차이를 검토합니다. 평균 차이는 -14.414 이고 신뢰 구간 범위는 -26.108 ~ -2.7196입니다. 치료 A에 의한 콜레스테롤 측정값 감소가 치료 B에 의한 감소보다 2.7 ~ 26.1점 정도 높다는 결론을 내릴 수 있습니다. 치료 A와 B 둘 다 Control과 Placebo에 비해 매우 효과적입니다.

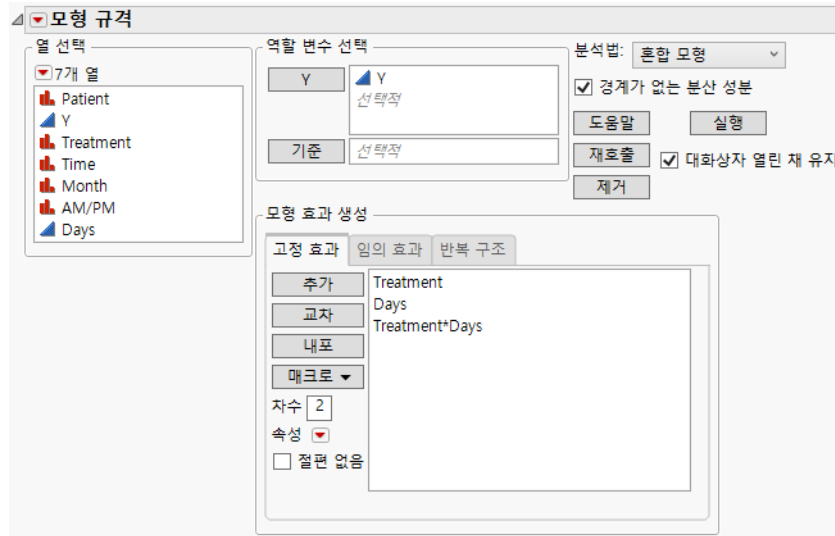
JMP^{PRO} AR(1)에 대한 회귀 모형의 예

Month 및 AM/PM 범주형 효과를 사용하여 콜레스테롤 데이터에 대한 네 가지 공분산 구조를 비교했습니다. 비정형 적합을 위해 범주형 효과가 필요했습니다. 결과적으로 AR(1) 공분산 구조를 사용하기로 결정했습니다.

이제 범주형 효과 대신 연속형 효과 Days 를 사용하여 치료 효과를 모델링한다고 가정해 보겠습니다 . 그런 다음 치료 기간 동안 임의의 시간에서 콜레스테롤 수준을 예측할 수 있습니다 .

1. " 공분산 구조 : AR(1)"(363 페이지) 에서 1 단계 ~ 4 단계를 수행한 후 모형 적합 시작 창으로 돌아갑니다 .
2. " 고정 효과 " 탭에서 기존 고정 효과를 선택하고 **제거**를 클릭합니다 .
3. Treatment 와 Days 를 선택한 후 **매크로 > 완전 요인**을 선택합니다 .

그림 8.25 고정 효과 탭을 보여 주는 모형 적합 시작 창



4. " 모형 규격 "의 빨간색 삼각형을 클릭하고 **중심화 다항식**을 선택 취소합니다 .

참고 : 기본 설정에서는 교호작용 항에 사용되는 연속형 효과가 중심화됩니다 . "중심화 다항식" 옵션을 해제하면 교호작용 항의 연속형 효과가 중심화되지 않습니다 .

5. **실행**을 클릭합니다 .

그림 8.26 에 " 혼합 적합 " 보고서가 표시되어 있습니다 . Treatment 와 Days 의 교호작용이 매우 유의하며 약물에 대해 다양한 회귀를 나타낸다는 것을 확인할 수 있습니다 .

참고 : 다른 날짜의 약물에 대한 결과를 예측하려면 프로파일러를 사용합니다 . 자세한 내용은 Profilers 의 에서 확인하십시오 .

3. Y 를 선택하고 Y 를 클릭합니다 .
4. " 분석법 " 목록에서 **혼합 모형**을 선택합니다 .
5. Tenderizer 와 Roasting Time 을 선택한 후 **매크로 > 완전 요인**을 선택합니다 .

그림 8.27 완료된 고정 효과 탭을 보여 주는 모형 적합 시작 창

모형 규격

열 선택: 5개 열
☒ Pattern
☒ Carcass
☒ Tenderizer
☒ Roasting Time
☒ Y

역할 변수 선택:
Y (선택적)
기준 (선택적)

분석법: 혼합 모형
☒ 경계가 없는 분산 성분
도움말 실행
재호출
☒ 대화상자 열린 채 유지
제거

모형 효과 생성
고정 효과 임의 효과 반복 구조
추가 교차 내포 매크로
차수 2
속성 없음
☐ 절편 없음

결과:
Tenderizer
Roasting Time
Tenderizer*Roasting Time

6. **임의 효과** 탭을 선택합니다 .
7. Carcass 를 선택하고 **추가**를 클릭하여 임의의 carcass 효과를 생성합니다 .
8. Carcass 와 Tenderizer 를 선택하고 **교차**를 클릭합니다 .
Carcass*Tenderizer 교호작용은 주구 요인 Tenderizer 의 오차 항입니다 . 이것은 표준 최소 제곱의 Carcass*Tenderizer& 랜덤 항과 동등합니다 .

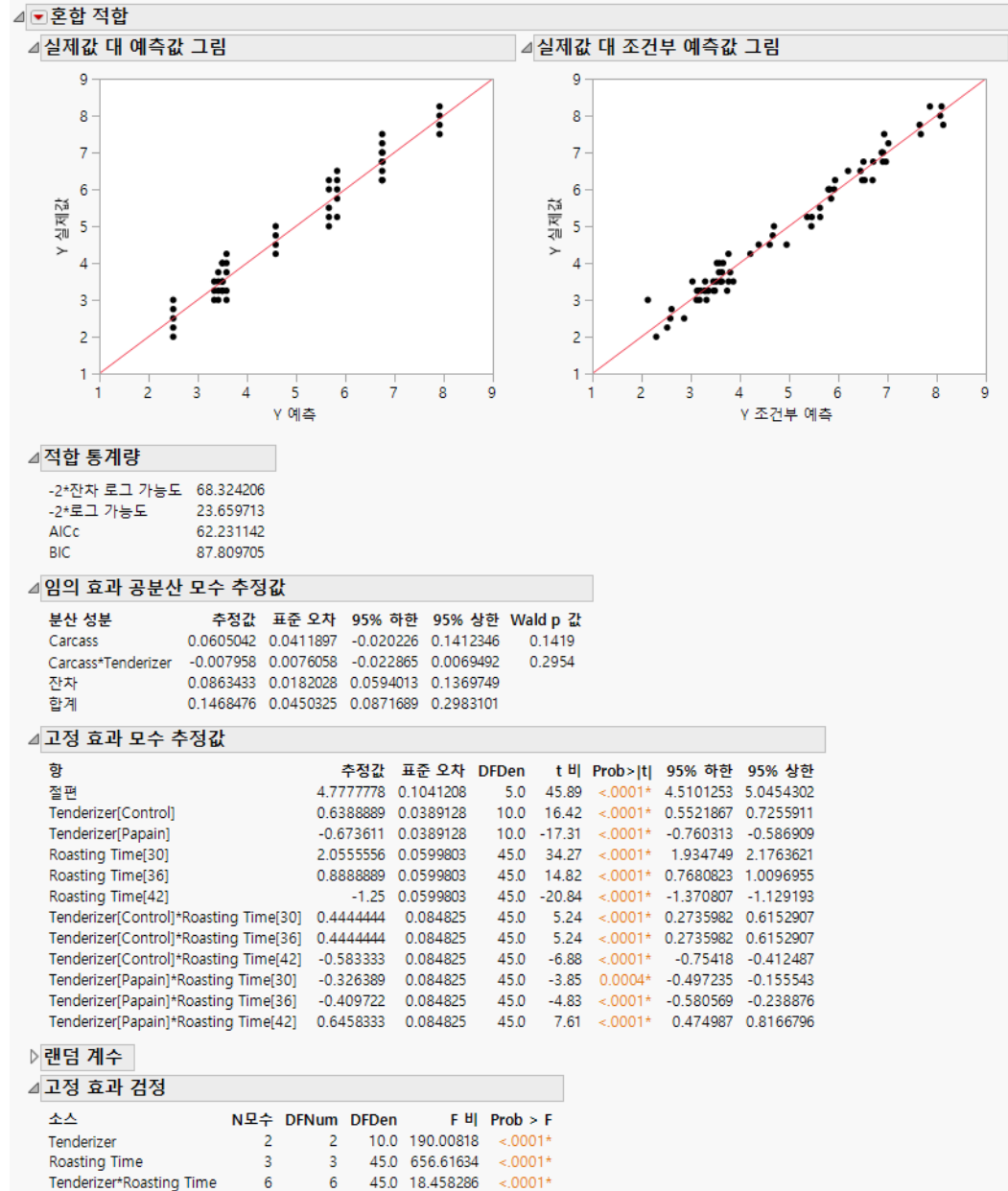
그림 8.28 완료된 임의 효과 탭을 보여 주는 모형 적합 시작 창

9. 실행을 클릭합니다.

그림 8.29 에 " 혼합 적합 " 보고서가 표시되어 있습니다.

" 실제값 대 예측값 그림 " 및 " 실제값 대 조건부 예측값 그림 " 에서 모형 적합 문제가 보이지 않으므로 결과를 해석할 수 있습니다. " 고정 효과 검정 " 보고서에서는 연화제와 로스팅 시간 사이에 유의한 교호작용이 있음을 나타냅니다.

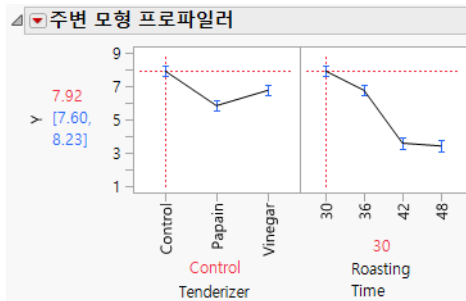
그림 8.29 혼합 적합 보고서



연화제와 로스팅 시간 사이의 교호작용 탐색

1. " 혼합 적합 " 의 빨간색 삼각형을 클릭하고 주변 모형 추론 > 프로파일러를 선택합니다 .

그림 8.30 Roasting Time 이 30 분으로 설정된 주변 모형 프로파일러



2. "Roasting Time" 패널에서 빨간색 수직 파선을 36, 42, 48 로 움직입니다 .

그림 8.30 에서 로스팅 시간이 30 분 또는 36 분일 때 파파인 (Papain) 과 식초 (Vinegar) 연화제 모두 대조군 (Control) 보다 연질 점수가 유의하게 낮습니다 . 그러나 42 분에서는 유의한 차이가 없습니다 . 48 분일 때는 파파인 (Papain) 값이 대조군 (Control) 보다 낮지만 식초 (Vinegar) 값은 그렇지 않습니다 . 42 분을 제외한 모든 시간에서 파파인 (Papain) 의 연질 점수가 식초 (Vinegar) 보다 낮습니다 .

3. " 혼합 적합 " 의 빨간색 삼각형을 클릭하고 **다중 비교**를 선택합니다 .
4. **Tenderizer*Roasting Time** 을 선택합니다 .
5. **전체 쌍별 비교 - Tukey HSD** 를 선택한 후 **확인**을 클릭합니다 .

그림 8.31 에서는 쌍별 비교 목록의 일부를 보여 줍니다 . 프로파일러에서 관측한 파파인 (Papain)과 식초(Vinegar)의 차이는 대부분 통계적으로 유의합니다. 따라서 파파인(Papain)이 더 나은 연화제입니다 .

그림 8.31 다중 비교 (일부분)

▶ 혼합 적합									
▶ Tenderizer*RoastingTime에 대한 다중 비교									
▶ 최소 제곱 평균 추정값									
	Roasting								
Tenderizer	Time	추정값	표준 오차	DF	95% 하한	95% 상한			
Control	30	7.916667	0.15214556	20.355	7.5996515	8.2336819			
Control	36	6.7500000	0.15214556	20.355	6.4329848	7.0670152			
Control	42	3.5833333	0.15214556	20.355	3.2663181	3.9003485			
Control	48	3.4166667	0.15214556	20.355	3.0996515	3.7336819			
Papain	30	5.8333333	0.15214556	20.355	5.5163181	6.1503485			
Papain	36	4.5833333	0.15214556	20.355	4.2663181	4.9003485			
Papain	42	3.5000000	0.15214556	20.355	3.1829848	3.8170152			
Papain	48	2.5000000	0.15214556	20.355	2.1829848	2.8170152			
Vinegar	30	6.7500000	0.15214556	20.355	6.4329848	7.0670152			
Vinegar	36	5.6666667	0.15214556	20.355	5.3496515	5.9836819			
Vinegar	42	3.5000000	0.15214556	20.355	3.1829848	3.8170152			
Vinegar	48	3.3333333	0.15214556	20.355	3.0163181	3.6503485			
▶ TukeyHSD 전체 쌍별 비교									
분위수 = 3.44501, 조정 DF = 45.0, 조정 = Tukey-Kramer									
▶ 전체 쌍별 차이									
	Roasting								
Tenderizer	Time	-Tenderizer	-Roasting	Time	차이	표준 오차	t 비	Prob> t	95% 하한 95% 상한
Control	30	Control	36		1.16667	0.1696500	6.88	<.0001*	0.58222 1.75111
Control	30	Control	42		4.33333	0.1696500	25.54	<.0001*	3.74889 4.91778
Control	30	Control	48		4.50000	0.1696500	26.53	<.0001*	3.91555 5.08445
Control	30	Papain	30		2.08333	0.1616430	12.89	<.0001*	1.52647 2.64019
Control	30	Papain	36		3.33333	0.1616430	20.62	<.0001*	2.77647 3.89019
Control	30	Papain	42		4.41667	0.1616430	27.32	<.0001*	3.85981 4.97353
Control	30	Papain	48		5.41667	0.1616430	33.51	<.0001*	4.85981 5.97353
Control	30	Vinegar	30		1.16667	0.1616430	7.22	<.0001*	0.60981 1.72353
Control	30	Vinegar	36		2.25000	0.1616430	13.92	<.0001*	1.69314 2.80686
Control	30	Vinegar	42		4.41667	0.1616430	27.32	<.0001*	3.85981 4.97353
Control	30	Vinegar	48		4.58333	0.1616430	28.35	<.0001*	4.02647 5.14019
Control	36	Control	42		3.16667	0.1696500	18.67	<.0001*	2.58222 3.75111
Control	36	Control	48		3.33333	0.1696500	19.65	<.0001*	2.74889 3.91778
Control	36	Papain	30		0.91667	0.1616430	5.67	<.0001*	0.35981 1.47353
Control	36	Papain	36		2.16667	0.1616430	13.40	<.0001*	1.60981 2.72353
Control	36	Papain	42		3.25000	0.1616430	20.11	<.0001*	2.69314 3.80686
Control	36	Papain	48		4.25000	0.1616430	26.29	<.0001*	3.69314 4.80686
Control	36	Vinegar	30		8.882e-16	0.1616430	0.00	1.0000	-0.55686 0.55686
Control	36	Vinegar	36		1.08333	0.1616430	6.70	<.0001*	0.52647 1.64019
Control	36	Vinegar	42		3.25000	0.1616430	20.11	<.0001*	2.69314 3.80686
Control	36	Vinegar	48		3.41667	0.1616430	21.14	<.0001*	2.85981 3.97353
Control	42	Control	48		0.16667	0.1696500	0.98	0.9974	-0.41778 0.75111
Control	42	Papain	30		-2.25000	0.1616430	-13.92	<.0001*	-2.80686 -1.69314
Control	42	Papain	36		-1.00000	0.1616430	-6.19	<.0001*	-1.55686 -0.44314
Control	42	Papain	42		0.08333	0.1616430	0.52	1.0000	-0.47353 0.64019
Control	42	Papain	48		1.08333	0.1616430	6.70	<.0001*	0.52647 1.64019
Control	42	Vinegar	30		-3.16667	0.1616430	-19.59	<.0001*	-3.72353 -2.60981
Control	42	Vinegar	36		-2.08333	0.1616430	-12.89	<.0001*	-2.64019 -1.52647
Control	42	Vinegar	42		0.08333	0.1616430	0.52	1.0000	-0.47353 0.64019
Control	42	Vinegar	48		0.25000	0.1616430	1.55	0.9187	-0.30686 0.80686
Control	48	Papain	30		-2.41667	0.1616430	-14.95	<.0001*	-2.97353 -1.85981
Control	48	Papain	36		-1.16667	0.1616430	-7.22	<.0001*	-1.72353 -0.60981
Control	48	Papain	42		-0.08333	0.1616430	-0.52	1.0000	-0.64019 0.47353
Control	48	Papain	48		0.91667	0.1616430	5.67	<.0001*	0.35981 1.47353
Control	48	Vinegar	30		-3.33333	0.1616430	-20.62	<.0001*	-3.89019 -2.77647
Control	48	Vinegar	36		-2.25000	0.1616430	-13.92	<.0001*	-2.80686 -1.69314
Control	48	Vinegar	42		-0.08333	0.1616430	-0.52	1.0000	-0.64019 0.47353
Control	48	Vinegar	48		0.08333	0.1616430	0.52	1.0000	-0.47353 0.64019

JMP PRO 공간 예 : 균일도 시험

Uniformity Trial.jmp 샘플 데이터 테이블을 예로 들겠습니다. 8x8 격자의 대지에 대해 작물 균일도 시험이 수행되었습니다. 균일도 시험에서는 실험 처리가 적용되지 않은 구획에서 테스트 작물을 재배합니다. 일반적으로 측정되는 반응 변수는 수확량입니다. 구획에서 수행할 설계된 실험을 계획하기 위한 배경으로 해당 구획의 변동을 특성화하는 것이 이 연구의 목적입니다. 자세한 내용은 Littell et al. 연구 자료(2006, p. 447)에서 확인하십시오.

목적은 이러한 데이터의 정보를 사용하여 16 개의 처리로 수확량 시험을 설계하는 것입니다. 특히 해당 구획에서 다음과 같은 향후 실험을 수행할지 여부를 결정하려고 합니다.

- 4 개의 블록이 있는 완비 블록 설계 (데이터에 Quarter 로 표시됨)
- 16 개의 블록이 있는 불완비 블록 설계 (데이터에 Subquarter 로 표시됨)
- 공간 상관 오차가 있는 완전히 랜덤화된 설계

이 목적을 위해 공간 데이터를 둘 이상의 차원이 반복 효과로 포함된 반복 측정으로 처리할 수 있습니다. 따라서 "적합 통계량" 보고서의 값을 사용하여 적절한 모형을 비교하고 선택할 수 있습니다. 유의한 공간 변동이 있는지 확인하는 것으로 시작한 후 너겟 효과가 있는지 여부를 확인합니다.

너겟 효과가 있는지 여부를 확인한 후에는 가장 적합한 공간 공분산 구조를 결정합니다. 마지막으로 블록 모형을 적합시키고 이러한 모형을 최적 공간 구조와 비교합니다. 이 예에서는 AICc와 BIC를 모두 사용하여 최적 모형을 선택합니다. "[공간 상관 구조](#)"(399 페이지)에서는 너겟 효과 및 기타 공간 용어에 대한 자세한 내용을 제공합니다.

팁 : 이 섹션에서는 모형을 적합하는 것부터 최적 공분산 구조를 결정하는 것까지 공간 데이터 적합에 대한 다양한 내용을 설명합니다. 각 예를 진행하는 동안 모형 적합 시작 창을 열어 두십시오.

JMP PRO 공간 구조 모형 적합

유의한 공간 변동이 있는지 확인하기 위해 공간 변동을 설명하는 모형을 적합시킬 수 있습니다. 그런 다음 이 공간 모형의 가능도를 공간 변동이 고려되지 않은 모형의 가능도와 비교할 수 있습니다. 독립 오차 모형이 공간 모형 계열에 내포되어 있으므로 이를 수행할 수 있습니다. 독립 오차 모형은 공간 상관(ρ)이 0인 공간 모형입니다. 이는 두 모형에 대한 공식 가능도비 검정을 수행할 수 있다는 의미입니다.

먼저 공간 구조를 설명하는 모형을 적합시킵니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Uniformity Trial.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. 다음 예에서 시작 창으로 돌아갈 수 있도록 **대화상자 열린 채 유지**를 선택합니다.
4. Yield 를 선택하고 **Y** 를 클릭합니다.
5. "분석법" 목록에서 **혼합 모형**을 선택합니다.

6. **반복 구조** 탭을 선택합니다 .
7. " 구조 " 옆의 목록에서 **공간**을 선택합니다 .
8. " 유형 " 옆의 목록에서 **구형**을 선택합니다 .
9. Row 와 Column 을 선택하고 **반복**을 클릭합니다 .

그림 8.32 반복 구조 탭을 보여 주는 완료된 모형 적합 시작 창

모형 규격

역할 변수 선택

Y: Yield (선택적)
기준: 선택적

분석법: 혼합 모형

☒ 경계가 없는 분산 성분

도움말 실행

재호출 ☒ 대화상자 열린 채 유지

제거

모형 효과 생성

고정 효과 임의 효과 **반복 구조**

반복 공분산 구조

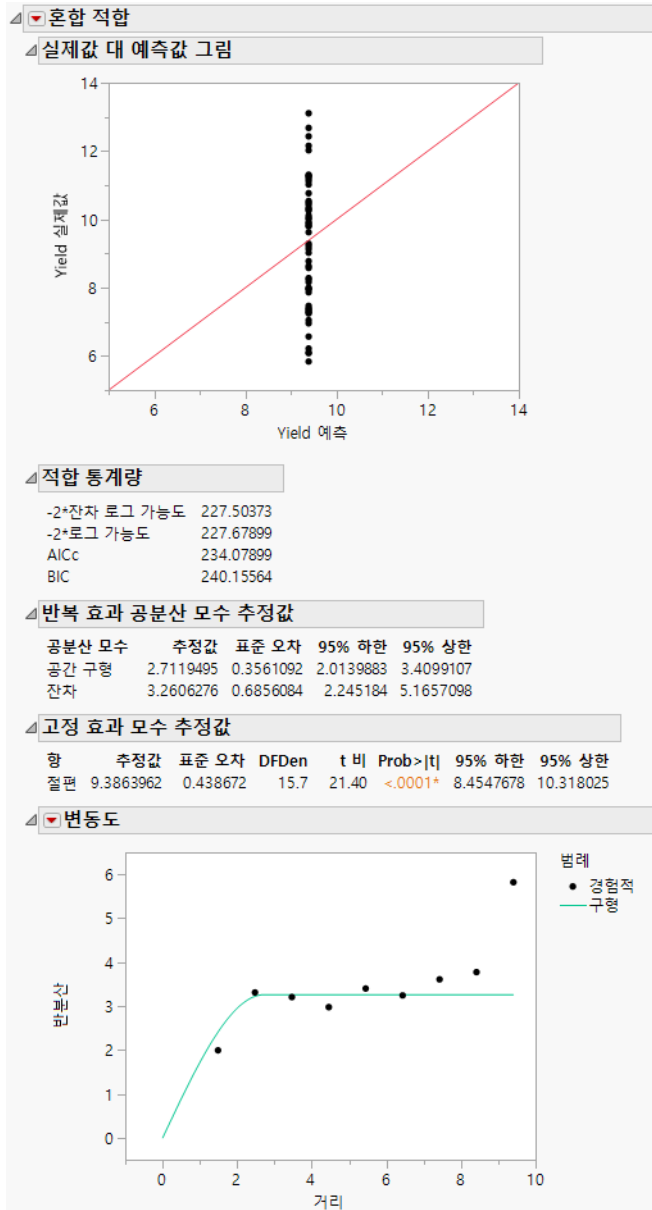
구조: 공간
유형: 구형

반복: Row, Column (선택적)

개제: 선택적

10. **실행**을 클릭합니다 .

그림 8.33 혼합 적합 보고서 - 공간 구형



"혼합 적합" 보고서의 "실제값 대 예측값 그림"에서는 예측 수확량(Yield)이 단일 값을 나타냅니다. 이는 공간 공분산만 적합되었기 때문입니다. "적합 통계량" 보고서에 따르면 -2*로그 가능도는 227.68이고 AICc는 234.08입니다.

등방성 공간 구조가 적합되었으므로 변동도 그림이 표시됩니다. 이 시험은 8 x 8 격자로 설계되므로 매우 먼 거리보다 짧은 거리에 더 많은 점 쌍이 있습니다. 레이아웃은 그림 8.37에서 확인하

십시오. "변동도"에서는 최대 약 8.4 거리에 대해 구형 공간 구조가 가장 적합하다는 것을 보여줍니다. 최종 거리에 대한 거리 클래스는 두 개의 주대각 점 쌍으로만 구성됩니다.

"반복 효과 공분산 모수 추정값" 보고서에는 범위 (공간 구형 = 2.71) 및 문턱(sill)(잔차 = 3.26)의 추정값이 제공됩니다. 자세한 내용은 "변동도"(399페이지)에서 확인하십시오.

JMP PRO 독립 오차 모형 적합

다음으로 독립 오차 모형을 적합시킵니다.

1. "모형 적합" 시작 창으로 돌아갑니다.
2. **반복 구조** 탭을 선택합니다.
3. "구조" 목록에서 **잔차**를 선택합니다.
4. "반복" 효과 목록에서 **Row**와 **Column**을 제거합니다.

그렇지 않으면 "잔차 반복 공분산 구조를 선택한 경우에는 반복 열 및 개체 열이 무시됩니다"라는 경고가 표시됩니다.

5. **실행**을 클릭합니다.

독립 오차 모형에 대한 적합 통계량은 $-2 \times \text{로그 가능도} = 254.22$ 이고 $\text{AICc} = 258.41$ 입니다. 이 값은 각각 공간 상관 모형의 해당 값 ($-2 \times \text{로그 가능도} = 227.68$, $\text{AICc} = 234.08$)을 초과합니다. 이러한 통계량의 값이 작을수록 더 적합하다는 것을 나타내므로 공간 모형의 적합도가 더 좋다고 할 수 있습니다.

JMP PRO 가능도비 검정 수행 (선택 사항)

공식 가능도비 검정에서는 공간 상관 모형이 유의한 변동을 설명하는지 여부를 보여줍니다. 유효한 가능도비 검정을 생성하려면 한 모형이 다른 모형에 내포되어야 합니다.

일반적으로 공간 모형은 공식 가능도비 검정이 아니라 AICc 또는 BIC 를 사용하여 비교됩니다. AICc 또는 BIC 를 평가하는 것이 더 빠르고, 많은 공간 모형이 내포되지 않습니다.

독립 오차 모형이 공간 모형 계열에 내포되어 있으므로 이 예에서 가능도비 검정을 수행할 수 있습니다. 독립 오차 모형은 공간 상관(ρ)이 0인 공간 모형입니다. 이는 두 모형에 대한 공식 가능도비 검정을 수행할 수 있다는 의미입니다.

이 예에서 가능도비 검정 통계량은 $254.22 - 227.68 = 26.54$ 입니다. 이 값을 자유도 1에 대한 카이 제곱 분포와 비교하면 공간 상관이 없다는 귀무가설이 p 값 < 0.0001 로 기각됩니다. 즉, 이러한 데이터에 유의한 공간 변동이 포함되어 있다는 결론을 내릴 수 있습니다.

JMP PRO 공간 공분산 유형 선택

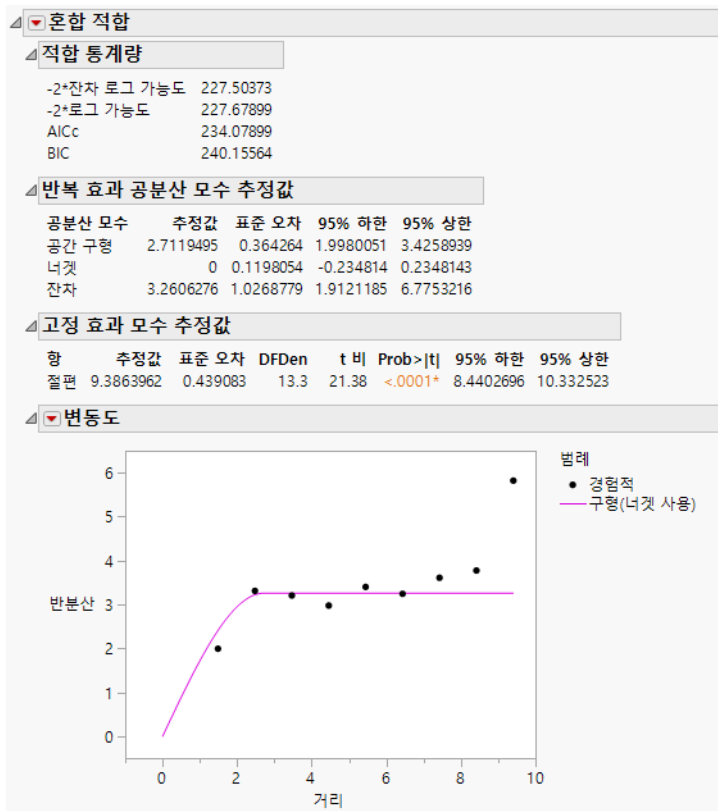
다음으로 데이터에 가장 적합한 공간 공분산 구조를 결정합니다.

- 너겟 효과 (상대적으로 짧은 거리에 대한 변동) 사용 또는 사용 안 함
- 등방성 (공간 상관이 모든 방향으로 같음) 또는 이방성 (공간 상관이 두 방향으로 다름)

- 구조 유형 (구형, 가우스, 지수 또는 거듭제곱)
1. "모형 적합" 시작 창으로 돌아갑니다.
 2. 반복 구조 탭을 선택합니다.
 3. Row 와 Column 을 선택하고 반복을 클릭합니다.
 4. "구조" 목록에서 공간 (너겟 사용) 을 선택합니다.
 5. "유형" 목록에서 구형을 선택합니다.
 6. 실행을 클릭합니다.

그림 8.34 에 "혼합 적합" 보고서가 표시되어 있습니다. 로그 가능도는 기본적으로 너겟이 없는 구형 모형과 같고 AICc 는 약간 높습니다 (236.36 대 234.08). "반복 효과 공분산 모수 추정값" 보고서에서는 너겟 공분산 모수의 추정값이 0 임을 보여 줍니다. 즉, 너겟 효과에 대한 증거가 없습니다.

그림 8.34 혼합 적합 보고서 - 공간 구형 (너겟 사용)



7. "변동도"의 빨간색 삼각형을 클릭하고 공간 > 구형을 선택합니다.

JMP[®] PRO 공간 구조 유형 결정

너겟이 없는 등방성 공간 구조가 적절합니다. 사용 가능한 여러 유형을 비교하여 공간 구조 유형을 결정할 수 있습니다.

1. "모형 적합" 시작 창으로 돌아갑니다.
2. **반복 구조** 탭을 선택합니다.
3. Row 와 Column 을 선택하고 **반복**을 클릭합니다.
4. "구조" 목록에서 **공간**을 선택합니다.
5. "유형" 목록에서 **거듭제곱**을 선택합니다.
6. **실행**을 클릭합니다.

모형 적합에 대한 AICc 및 BIC 값에 유의하십시오.

7. "모형 적합" 시작 창으로 돌아갑니다.
 8. "유형" 목록에서 **지수**를 선택합니다.
 9. **실행**을 클릭합니다.
- 모형 적합에 대한 AICc 및 BIC 값에 유의하십시오.
10. "모형 적합" 시작 창으로 돌아갑니다.
 11. "유형" 목록에서 **가우스**를 선택합니다.
 12. **실행**을 클릭합니다.

모형 적합에 대한 AICc 및 BIC 값에 유의하십시오.

표 8.2에는 이러한 유형 및 수행한 다른 적합에 대해 관측된 AICc 값이 요약되어 있습니다.

표 8.2 공간 모형 적합에 대한 적합 통계량

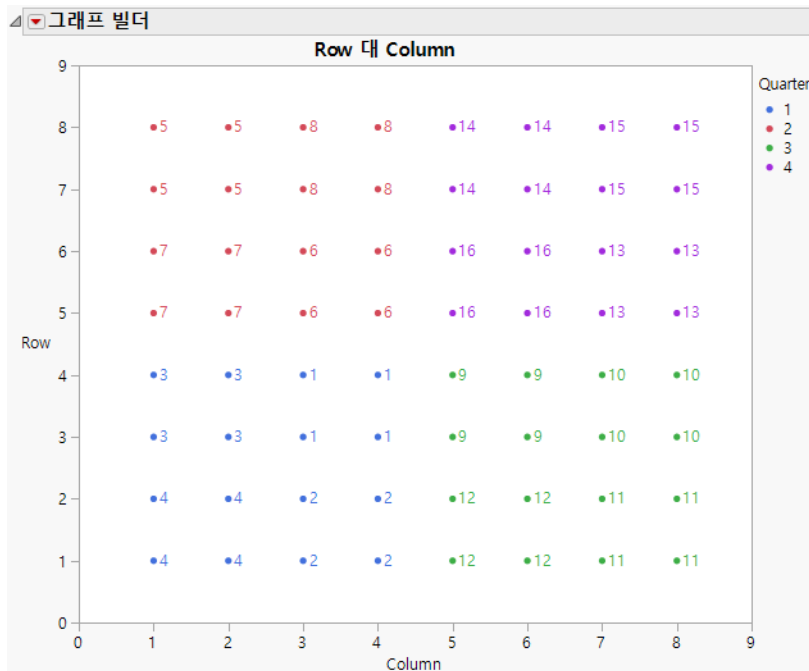
구조	유형	AICc	BIC
공간	구형	234.08	240.16
잔차		258.41	262.53
공간(너겟 사용)	구형	236.36	244.31
공간 이방성	구형	240.54	248.50
공간	거듭제곱	240.24	246.32
공간	지수	240.24	246.32
공간	가우스	238.37	244.44

AICc 값이 가장 작은 최량 적합 모형은 구형 공분산 구조를 사용한 공간 구조 모형입니다. 이제 이 모형을 완비 블록 모형 및 불완비 블록 모형과 비교하여 균일도 시험의 목표를 완료합니다.

JMP PRO 모형을 블록 설계와 비교

1. Uniformity Trial 데이터 테이블로 돌아갑니다.
2. 테이블 패널에서 "Graph Builder" 스크립트를 실행합니다.

그림 8.37 제안된 완비 / 불완비 블록 설계에 대한 그래프 빌더



이 그림에서는 해당 구획에 대해 제안된 완비 블록 설계 및 불완비 블록 설계를 보여 줍니다. 색상은 완비 블록으로 사용되는 4 등분 구획을 나타냅니다. 번호가 매겨진 점은 불완비 블록으로 사용되는 16 등분 구획을 나타냅니다.

완비 블록 모형을 적합시키려면 다음 단계를 수행하십시오.

1. "모형 적합" 시작 창으로 돌아갑니다.
2. **반복 구조** 탭을 선택합니다.
3. "구조" 목록에서 **잔차**를 선택합니다.
4. 효과에서 **Row** 와 **Column** 을 제거합니다.

그렇지 않으면 "잔차 반복 공분산 구조를 선택한 경우에는 반복 열 및 개체 열이 무시됩니다"라는 팝업 창이 표시됩니다.

5. **임의 효과** 탭을 선택합니다.

6. Quarter 를 선택하고 **추가**를 클릭합니다 .

7. **실행**을 클릭합니다 .

불완비 블록 모형을 적합시키려면 다음 단계를 수행하십시오 .

1. " 모형 적합 " 시작 창으로 돌아갑니다 .
2. **임의 효과** 탭을 선택합니다 .
3. Quarter 를 선택하고 **제거**를 클릭합니다 .
4. Subquarter 를 선택하고 **추가**를 클릭합니다 .
5. **실행**을 클릭합니다 .

다음 목록에서는 경쟁 모형에 대한 AICc 및 BIC 값을 모두 보여 줍니다 . 구형 공분산 구조를 사용할 때 최량 모형 적합 결과가 생성됩니다 . 이는 향후에 이 구획을 사용하여 연구를 수행하는 경우 공간 상관 오차가 있는 완전히 랜덤화된 설계가 선호된다는 것을 나타냅니다 .

- 구형 모형 (" [공간 구조 유형 결정](#) "(382 페이지) 참조)
 - AICc: 234.08
 - BIC: 240.16
- 완비 블록 (RCBD) 모형
 - AICc: 259.90
 - BIC: 265.97
- 불완비 블록 모형
 - AICc: 248.77
 - BIC: 254.85

JMP PRO 상관계수가 있는 반응의 예

이 예에서는 관심 있는 특성에 대해 웨이퍼 생산을 처리하는 두 레이아웃의 효과를 연구합니다 . 50개의 각 웨이퍼는 4개의 사분면으로 분할되고 각 사분면에 대해 특성이 측정됩니다 . 일반적으로 이 유형의 데이터는 각 행에 관심 단위 중 하나에 대한 모든 반복 측정값이 포함되는 형식으로 제공됩니다 . 이 유형의 데이터는 각 반응에 대해 별도의 모형을 사용하여 분석되는 경우가 많습니다 . 그러나 단일 단위에 대해 반복 측정을 수행하는 경우 단위 내 상관관계가 있을 가능성이 큼니다 . 이 상관관계를 설명하지 못하면 잘못된 결정 및 예측이 생성될 수 있습니다 . 혼합 모형 분석법을 사용하여 가능한 상관관계를 설명하고 모델링할 수 있습니다 .

반복 측정 데이터에 대한 혼합 모형 분석의 경우 각 반복 측정값이 고유의 행에 포함되어야 합니다 . 모든 반복 측정값이 동일한 행에 포함되는 일반적인 형식의 데이터가 제공되면 테이블 > 쌓기를 사용하여 혼합 모형 분석에 적합한 데이터 테이블을 생성할 수 있습니다 . 자세한 내용은 JMP 사용의 에서 확인하십시오 .

이 예에서는 먼저 일반적인 데이터 테이블 형식을 사용하여 단변량 모형을 적합시킵니다. 그런 다음 혼합 모형 분석법을 사용하여 네 개의 반응에 대한 모형을 적합시키는 동시에 반응 간의 가능한 상관관계를 설명합니다.

JMP PRO 단변량 모형 적합

표준 최소 제곱을 사용하여 네 개의 각 사분면에 대해 단변량 모형을 적합시킵니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Wafer Quadrants.jmp** 를 엽니다.

이 데이터 테이블은 혼합 모형 분석에 사용할 수 있도록 각 **Quadrant** 에서 측정한 각 **Y** 값이 하나의 행에 포함된 구조입니다. 단변량 분석을 수행하기 위해 저장된 스크립트를 사용하여 테이블을 분할합니다.

2. **Wafer Quadrants.jmp** 데이터 테이블에서 **Split Y by Quadrant** 스크립트 옆의 녹색 삼각형을 클릭합니다.

새 데이터 테이블은 반복 측정 데이터를 기록하는 데 일반적으로 사용되는 형식으로 되어 있습니다. **Wafer ID** 의 각 값은 행을 정의하고 해당 웨이퍼에 대한 네 개의 측정값이 단일 행에 제공됩니다.

3. **분석 > 모형 적합**을 선택합니다.

데이터 테이블에 "Model" 스크립트가 있으므로 " 모형 규격 " 창이 이미 채워져 있습니다. **High, High ~ Low, Low** 열은 **Y** 로 입력되고 **Layout** 은 단일 모형 효과입니다.

4. **실행**을 클릭합니다.

그림 8.38 네 가지 단변량 모형

최소 제곱 적합					
효과 요약					
반응 High, High					
적합 요약					
분산 분석					
모수 추정값					
항	추정값	표준 오차	t 비	Prob> t	
절편	97.536737	0.638073	152.86	<.0001*	
Layout[Layout A]	7.7721843	0.638073	12.18	<.0001*	
효과 검정					
효과 상세 정보					
반응 High, Low					
적합 요약					
분산 분석					
모수 추정값					
항	추정값	표준 오차	t 비	Prob> t	
절편	100.85909	0.571089	176.61	<.0001*	
Layout[Layout A]	0.678605	0.571089	1.19	0.2406	
효과 검정					
효과 상세 정보					
반응 Low, High					
적합 요약					
분산 분석					
모수 추정값					
항	추정값	표준 오차	t 비	Prob> t	
절편	81.602404	1.329545	61.38	<.0001*	
Layout[Layout A]	5.344424	1.329545	4.02	0.0002*	
효과 검정					
효과 상세 정보					
반응 Low, Low					
적합 요약					
분산 분석					
모수 추정값					
항	추정값	표준 오차	t 비	Prob> t	
절편	75.080525	1.672511	44.89	<.0001*	
Layout[Layout A]	5.2598871	1.672511	3.14	0.0028*	
효과 검정					
효과 상세 정보					

보고서에 따르면 Layout 은 High, Low 사분면을 제외한 모든 사분면에 대해 통계적으로 유의한 효과입니다.

JMP PRO 혼합 모형 분석 수행

혼합 모형 분석을 사용하여 더 많은 정보를 얻을 수 있습니다.

1. Wafer Quadrants.jmp 로 돌아갑니다. 데이터 테이블을 닫았으면 **도움말 > 샘플 데이터 라이브러리**를 선택하고 Wafer Quadrants.jmp 를 엽니다.

2. **분석 > 모형 적합**을 선택합니다 .
3. **Y**를 선택하고 **Y**를 클릭합니다 .
4. **분석법** 목록에서 **혼합 모형**을 선택합니다 .
5. **열 선택** 목록에서 **Quadrant**와 **Layout**을 선택하고 **매크로 > 완전 요인**을 클릭합니다 .

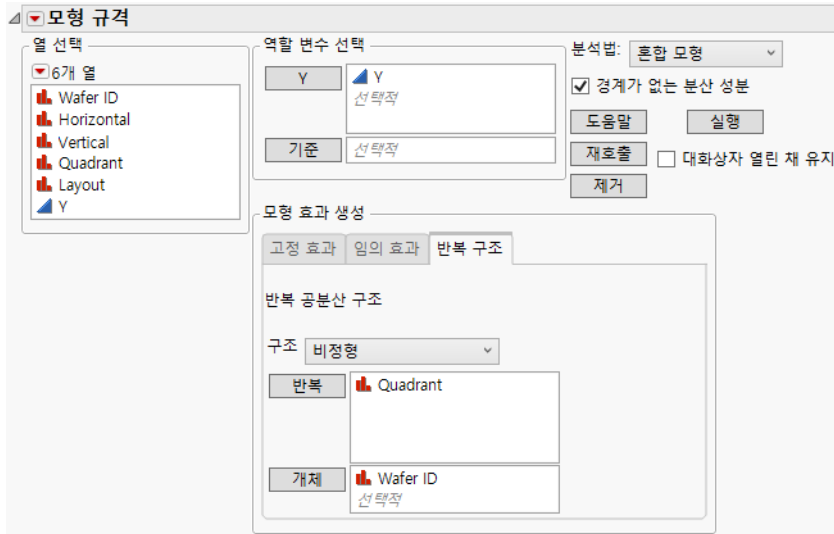
이 모형 규격을 사용하면 반복 측정값에 대한 **Layout**의 효과 및 **Layout**과 **Quadrant**의 가능한 교호작용을 탐색할 수 있습니다 .

그림 8.39 완료된 고정 효과 탭을 보여 주는 모형 적합 시작 창

The screenshot shows the '모형 규격' (Model Specification) dialog box. On the left, under '열 선택' (Column Selection), a list of variables includes 'Wafer ID', 'Horizontal', 'Vertical', 'Quadrant', 'Layout', and 'Y'. The '역할 변수 선택' (Role Variable Selection) section has 'Y' selected as the response variable and 'Quadrant' and 'Layout' as fixed effects. The '분석법' (Analysis Method) is set to '혼합 모형' (Mixed Model). The '고정 효과' (Fixed Effects) tab is active, showing a list of selected effects: 'Quadrant', 'Layout', and 'Quadrant*Layout'. The '반복 구조' (Repeated Structure) tab is also visible.

6. **반복 구조** 탭을 선택합니다 .
7. " 구조 " 목록에서 **비정형**을 선택합니다 .
8. **Quadrant**를 선택하고 **반복**을 클릭합니다 .
9. **Wafer ID**를 선택하고 **개체**를 클릭합니다 .

그림 8.40 반복 구조 탭을 보여 주는 모형 적합 시작 창



10. 실행을 클릭합니다 .

그림 8.41 고정 효과 모수 추정값 보고서가 닫혀 있는 혼합 적합 보고서

혼합 적합					
반복 효과 공분산 모수 추정값					
반복 효과: Quadrant					
개제: Wafer ID					
공분산 모수	추정값	표준 오차	95% 하한	95% 상한	
분산(High, High)	19.835702	4.0489457	11.899914	27.77149	
공분산(High, Low, High, High)	0.0111321	2.5624804	-5.011237	5.0335013	
분산(High, Low)	15.889659	3.2434631	9.5325884	22.24673	
공분산(Low, High, High, High)	15.893138	6.3915311	3.3659676	28.420309	
공분산(Low, High, High, Low)	1.0055411	5.3413847	-9.463381	11.474463	
분산(Low, High)	86.121899	17.579559	51.666597	120.5772	
공분산(Low, Low, High, High)	14.723606	7.7996739	-0.563474	30.010686	
공분산(Low, Low, High, Low)	8.0424897	6.8163252	-5.317262	21.402242	
공분산(Low, Low, Low, High)	2.1512894	15.640276	-28.50309	32.805667	
분산(Low, Low)	136.28412	27.81888	81.76012	190.80813	
고정 효과 검정					
소스	N모수	DFNum	DFDen	F 비	Prob > F
Quadrant	3	3	46.0	149.21352	<.0001*
Layout	1	1	48.0	51.757797	<.0001*
Quadrant*Layout	3	3	46.0	22.379307	<.0001*

" 반복 효과 공분산 모수 추정값 " 보고서에는 네 가지 반응에 대한 추정 분산 및 공분산이 제공됩니다. Low, High, High, High의 공분산에 대한 신뢰 구간에 0이 포함되지 않습니다. 이는 두 사분면의 측정값 간에 양의 공분산이 있음을 나타냅니다. 이 정보는 혼합 모형 분석에 사용되며 반응이 독립적으로 모델링되는 경우 사용할 수 없습니다.

" 고정 효과 검정 " 보고서에서는 유의한 레이아웃(Layout)* 특성(Characteristic) 교호작용이 있음을 나타냅니다.

JMP PRO 프로파일러를 사용하여 레이아웃 * 특성 교호작용 탐색

1. " 혼합 적합 "의 빨간색 삼각형을 클릭하고 **주변 모형 추론 > 프로파일러**를 선택합니다 .
2. 프로파일러 그림에서 먼저 Layout A 에 빨간색 수직 점선을 설정한 후 Layout B 에 설정하여 전체 사분면에서 Y 의 예측값을 비교합니다 .

그림 8.42 레이아웃 A 의 사분면에 대한 프로파일러

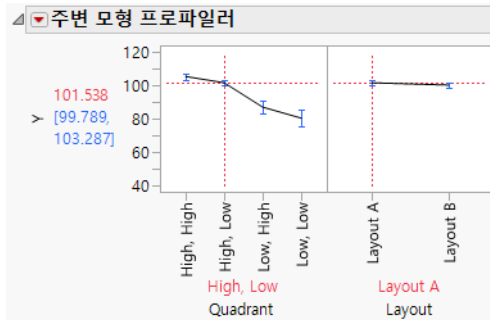
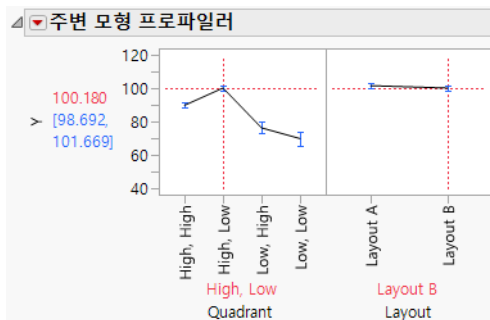


그림 8.43 레이아웃 B 의 사분면에 대한 프로파일러



각 Layout 설정에서 프로파일의 차이는 유의한 교호작용에 대한 정보를 제공합니다 . 교호작용은 부분적으로 High, High 사분면의 차이로 인해 발생하는 것으로 보입니다 .

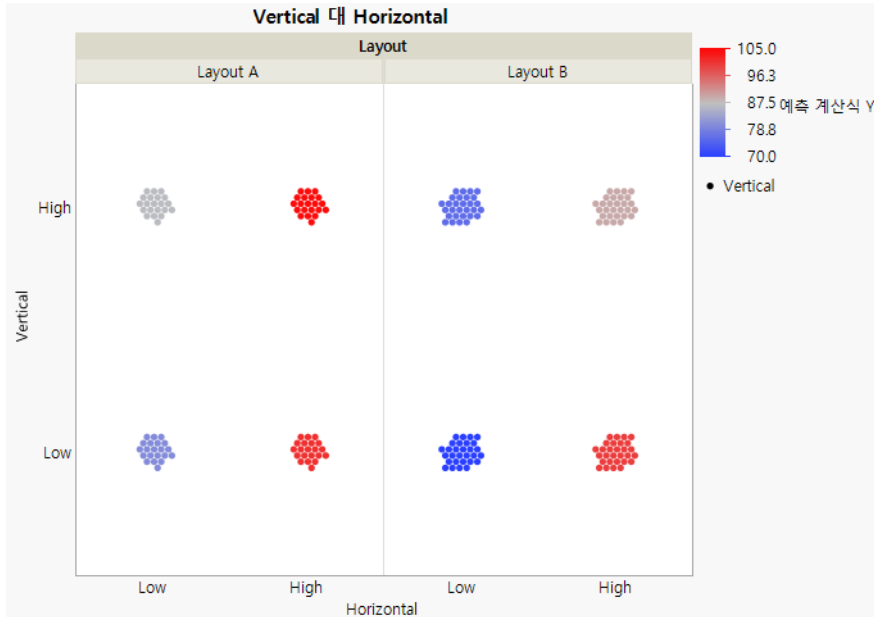
JMP PRO 레이아웃 및 사분면별 Y 그림

그래프 빌더를 사용하여 교호작용의 특성을 탐색합니다 .

1. " 혼합 적합 "의 빨간색 삼각형을 클릭하고 **열 저장 > 예측 계산식**을 선택합니다 .
예측 계산식이 데이터 테이블의 예측 계산식 Y 열에 저장됩니다 .
2. **그래프 > 그래프 빌더**를 선택합니다 .
3. Horizontal 을 X 영역으로 드래그합니다 .
4. Vertical 을 Y 영역으로 드래그합니다 .

5. 예측 계산식 Y 를 **색상** 영역으로 드래그합니다 .
6. Layout 을 **펼침** 영역으로 드래그합니다 .
7. **완료**를 클릭하여 제어판을 숨깁니다 .

그림 8.44 완료된 그래프 빌더 그림



참고 : 그래프 빌더의 점은 무작위로 지터링되므로 그림 8.44 의 그림과 정확히 일치하지 않을 수 있습니다 .

이 그림에서는 색상 강도 척도를 사용하여 8 개의 Layout 및 Quadrant 조합에 대한 예측된 차이를 보여 줍니다 . High, Low 사분면에 대한 예측값은 오른쪽 하단에 있습니다 . 색상 그래프 디엔트는 이러한 예측값에 대해 상대적으로 작은 차이를 보여 줍니다 . 다른 차이는 명확하게 표시됩니다 .

JMP PRO 혼합 모형 분석법에 대한 통계 상세 정보

- "수렴 스코어 검정 "
- " 랜덤 계수 모형 "
- " 반복 측정 "
- " 반복 공분산 구조 "
- " 공간 및 시간 변동 "
- "Kackar-Harville 수정 "

JMP PRO 수렴 스코어 검정

수렴 실패 경고는 알 수 없는 MLE(최대 가능도 추정값)가 모형 적합 알고리즘의 최종 반복에서 제공된 모수와 일치한다는 가설에 대한 스코어 검정을 보여 줍니다. 이 가설 검정은 상대 그래디언트 기준이 스코어 검정 통계량과 대수적으로 동등하므로 가능합니다. 또한 스코어 검정에는 실제 MLE에 대한 정보가 필요 없습니다.

JMP PRO 스코어 검정

먼저 단일 모수 θ 를 가정해 보겠습니다. l 을 θ 의 로그 가능도 함수로 하고 $\mathbf{x}$ 를 데이터로 합니다. 점수는 다음과 같이 θ 에 대한 로그 가능도 함수의 도함수입니다.

$$U(\theta) = \frac{\partial l(\theta|\mathbf{x})}{\partial \theta}$$

관측된 정보 :

$$I(\theta) = -\frac{\partial^2}{\partial \theta^2} l(\theta|\mathbf{x})$$

$H_0: \theta = \theta_0$ 의 스코어 검정에 대한 통계량 :

$$S(\theta_0) = \frac{U(\theta_0)^2}{I(\theta_0)}$$

이 통계량은 귀무가설 하에서 자유도가 1 인 점근적 카이제곱 분포를 따릅니다.

스코어 검정을 여러 모수로 일반화할 수 있습니다. 모수 θ 의 벡터를 가정해 보겠습니다. $H_0: \theta = \theta_0$ 의 스코어 검정에 대한 통계량은 다음과 같습니다.

$$S(\theta_0) = \mathbf{U}'(\theta_0) \mathbf{I}^{-1}(\theta_0) \mathbf{U}(\theta_0)$$

다음은 각 요소에 대한 설명입니다.

$$\mathbf{U}(\theta) = \frac{\partial l(\theta|\mathbf{x})}{\partial \theta}$$

및

$$\mathbf{I}(\theta) = -\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \theta (\partial \theta')}$$

$\mathbf{U}'$ 는 $\mathbf{U}$ 행렬의 전치를 나타냅니다.

검정 통계량은 자유도가 k 인 점근적 카이제곱 분포입니다. 여기서 k 는 비유계 모수의 수입니다.

JMP PRO 상대 그래디언트

혼합 모형 적합 절차의 수렴 기준은 상대 그래디언트 $g'H^{-1}g$ 를 기반으로 합니다. 여기서 $g(\theta) = \mathbf{U}(\theta)$ 는 로그 가능도 함수의 그래디언트이고, $\mathbf{H}(\theta) = -\mathbf{I}(\theta)$ 는 헤시안 행렬입니다.

θ_0 을 알고리즘이 종료되는 θ 의 값이라고 하겠습니다. θ_0 에서 평가된 상대 그래디언트가 스코어 검정 통계량입니다. p 값은 자유도가 k 인 카이제곱 분포를 사용하여 계산됩니다. 이 p 값을 알 수 없는 MLE 값이 θ_0 와 일치하는지 여부를 나타냅니다. "임의 효과 공분산 모수 추정값" 보고서에 나열된 비유계 모수의 수는 k 입니다.

JMP PRO 랜덤 계수 모형

표준 랜덤 계수 모형은 각 개체에 대한 랜덤 절편과 기울기를 지정합니다. y_{ij} 가 i 번째 개체에 대한 j 번째 관측값의 측정값을 나타낸다고 하겠습니다. 그러면 랜덤 계수 모형을 다음과 같이 지정할 수 있습니다.

$$y_{ij} = a_i + b_i x_{ij} + e_{ij}$$

다음은 각 요소에 대한 설명입니다.

$$i = 1, 2, \dots, t$$

$$j = 1, 2, \dots, n_i$$

$$\begin{bmatrix} a_i \\ b_i \end{bmatrix} \sim iid N\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \mathbf{G}\right)$$

$$\mathbf{G} = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix}$$

및

$$e_{ij} \sim iid N(0, \sigma^2)$$

추정되는 고정 성분 및 랜덤 성분을 반영하도록 모형을 재구성할 수 있습니다.

$$y_{ij} = (\alpha + a_i^*) + (\beta + b_i^*)x_{ij} + e_{ij}$$

다음은 각 요소에 대한 설명입니다.

$$a_i^* = \alpha_i - \alpha$$

$$b_i^* = \beta_i - \beta$$

및

$$\begin{bmatrix} a_i^* \\ b_i^* \end{bmatrix} \sim iid N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{G} \right)$$

$\mathbf{G}$ 및 e_{ij} 는 위와 같이 정의됩니다.

JMP PRO 반복 측정

반복 측정 모형의 형식은 $y_{ijk} = \alpha_{ij} + s_{ik} + e_{ijk}$ 이고, 다음은 각 요소에 대한 설명입니다.

α_{ij} 를 처리 및 시간 요인으로 작성할 수 있습니다.

s_{ik} 는 i 번째 처리에 할당된 k 번째 개체의 임의 효과입니다.

$j = 1, \dots, m$ 은 시간에 따른 반복 측정을 나타냅니다.

s_{ik} 를 독립적이며 동일한 분포를 따르는 $N(0, \sigma_s^2)$ 변수로 가정합니다. 처리 요인의 수는 t 로 나타내고 개체 수는 s 로 나타냅니다. 그러면 e_{ijk} 의 분포는 $N(0, \Sigma)$ 이고, 다음과 같습니다.

$$\Sigma = \mathbf{I}_{ts} \otimes \text{Var}(\mathbf{y}_{ik} | s_{ik})$$

및

$$\mathbf{y}'_{ik} | s_{ik} = \left(\begin{bmatrix} y_{i1k} & y_{i2k} & \dots & y_{imk} \end{bmatrix} \middle| s_{ik} \right)$$

처리 내에서 ik 번째 개체에 해당하는 공분산 행렬 Σ 의 대각 성분 블록을 Σ_{ik} 로 나타냅니다. 즉, $\Sigma_{ik} = \text{Var}(\mathbf{y}_{ik} | s_{ik})$ 입니다. 일반적으로 개체 내에서 시간에 따른 관측값은 독립적이지 않으므로 $y_{ijk} | s_{ik}$ 의 분산을 추정해야 합니다. 상관관계를 설명하지 못하면 추론이 왜곡될 수 있습니다.

Σ_{ik} 에 사용할 수 있는 공분산 구조에 대한 자세한 내용은 "[반복 공분산 구조](#)"(393페이지) 및 "[공간 및 시간 변동](#)"(398페이지)에서 확인하십시오.

JMP PRO 반복 공분산 구조

이 섹션에서는 다음 공분산 구조에 대한 파라미터화를 제공합니다.

- "[이분산 공분산 구조](#)"(394 페이지)
- "[비정형 공분산 구조](#)"(394 페이지)
- "[복합 대칭 공분산 구조](#)"(395 페이지)
- "[AR\(1\) 공분산 구조](#)"(396 페이지)
- "[Toeplitz 공분산 구조](#)"(397 페이지)
- "[Antedependent 공분산 구조](#)"(397 페이지)

JMP[®] PRO 이분산 공분산 구조

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ & \sigma_2^2 & 0 & \dots & 0 \\ & & \sigma_3^2 & \dots & 0 \\ & & & \ddots & \vdots \\ & & & & \sigma_m^2 \end{bmatrix}$$

시간 j 에서 관측값 간의 분산은 다음과 같습니다.

$$\sigma_j^2 = \text{Var}(y_{ijk} | s_{ik})$$

분산은 반복 열의 수준마다 다를 수 있습니다. 서로 다른 수준에서 관측값 간의 공분산은 0입니다.

JMP[®] PRO 비정형 공분산 구조

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1[m-1]} & \sigma_{1m} \\ & \sigma_2^2 & \sigma_{23} & \dots & \sigma_{2[m-1]} & \sigma_{2m} \\ & & \sigma_3^2 & \dots & \sigma_{3[m-1]} & \sigma_{3m} \\ & & & \ddots & \vdots & \vdots \\ & & & & \sigma_{m-1}^2 & \sigma_{[m-1]m} \\ & & & & & \sigma_m^2 \end{bmatrix}$$

시간 j 에서 관측값 간의 분산은 다음과 같습니다.

$$\sigma_j^2 = \text{Var}(y_{ijk} | s_{ik})$$

시간 j 및 j' 에서 관측값 간의 공분산은 다음과 같습니다.

$$\sigma_{jj'} = \text{Cov}(y_{ijk}, y_{ij'k} | s_{ik})$$

분산은 반복 열의 수준마다 다를 수 있습니다. 서로 다른 수준에서 관측값 간의 공분산은 고유합니다.

JMP^{PRO} 복합 대칭 공분산 구조

JMP에서 복합 대칭 공분산 구조는 독립 오차 방법의 혼합 모형을 사용하여 구현됩니다. 임의 효과는 두 가지 범주(G 측 또는 R 측)로 분류됩니다. 자세한 내용은 Searle et al. 연구 자료(1992)에서 확인하십시오.

G 측 임의 효과는 임의 효과에 대한 설계 행렬과 관련됩니다. R 측 임의 효과는 잔차 오차와 관련됩니다. 개체 내 분산은 설계 구조의 일부이며 G 측에서 모델링됩니다. 개체 간 분산은 잔차 구조에 속하며 R 측에서 모델링됩니다. 독립 구조의 분산은 다음과 같은 방식으로 모델링됩니다.

- 임의 효과 G 측 분산은 $s_{ik} \sim \text{iid } N(0, \sigma_s^2)$ 으로 모델링됩니다.
- R 측 분산은 $e_{ijk} \sim \text{iid } N(0, \sigma^2)$ 으로 모델링됩니다.

그러면 공분산 행렬이 다음과 같이 정의됩니다.

$$\Sigma_{ik} = \sigma_s^2 \mathbf{J} + \sigma^2 \mathbf{I} = \begin{bmatrix} \sigma_s^2 + \sigma^2 & \sigma_s^2 & \dots & \sigma_s^2 \\ & \sigma_s^2 + \sigma^2 & \dots & \sigma_s^2 \\ & & \ddots & \vdots \\ & & & \sigma_s^2 + \sigma^2 \end{bmatrix}$$

여기서 $\mathbf{J}$ 는 1로 구성된 행렬이고 $\mathbf{I}$ 는 항등 행렬입니다.

모든 분산을 R 측에서 모델링할 수도 있습니다. 가우스 가정을 고려할 때 이 복합 대칭 공분산 구조는 독립 모형 (SAS의 Type=CS) 과 동등합니다. 이 구조는 JMP에서 반복 구조 탭의 복합 대칭 구조를 통해 사용할 수 있습니다. 여기서 반복 관측값 쌍 간의 상관관계는 관측값 간의 시간 차이에 관계없이 동일합니다. 따라서 상관 행렬을 다음과 같이 지정할 수 있습니다.

$$\mathbf{C} = \begin{bmatrix} 1 & \rho & \dots & \rho \\ & 1 & \dots & \rho \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}$$

또한 JMP에서 복합 대칭 구조를 사용하면 모든 시점에서 관측값 간에 공통 분산(σ_e^2)을 가정할 수 있습니다. 그러면 공분산 구조는 $\Sigma = \sigma_e^2 \mathbf{C}$ 이고, 다음과 같습니다.

$$\sigma_e^2 = \sigma_s^2 + \sigma^2$$

및

$$\rho = \frac{\sigma_s^2}{\sigma_s^2 + \sigma^2}.$$

여기서 ρ 은 클래스 간 상관계수이고 σ_e^2 는 잔차 분산입니다. JMP 에서 복합 대칭 이분산 구조를 사용할 수도 있습니다. 이 방법은 시점에 따라 분산이 다를 수 있습니다. 그러면 다음과 같은 공분산 행렬이 생성됩니다.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 & \dots & \rho\sigma_1\sigma_{m-1} & \rho\sigma_1\sigma_m \\ & \sigma_2^2 & \rho\sigma_2\sigma_3 & \dots & \rho\sigma_2\sigma_{m-1} & \rho\sigma_2\sigma_m \\ & & \sigma_3^2 & \dots & \rho\sigma_3\sigma_{m-1} & \rho\sigma_3\sigma_m \\ & & & \ddots & \vdots & \vdots \\ & & & & \sigma_{m-1}^2 & \rho\sigma_{m-1}\sigma_m \\ & & & & & \sigma_m^2 \end{bmatrix}$$

JMP[®] PRO AR(1) 공분산 구조

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho^{|t_1-t_2|}\sigma^2 & \rho^{|t_1-t_3|}\sigma^2 & \dots & \rho^{|t_1-t_{m-1}|}\sigma^2 & \rho^{|t_1-t_m|}\sigma^2 \\ & \sigma^2 & \rho^{|t_2-t_3|}\sigma^2 & \dots & \rho^{|t_2-t_{m-1}|}\sigma^2 & \rho^{|t_2-t_m|}\sigma^2 \\ & & \sigma^2 & \dots & \rho^{|t_3-t_{m-1}|}\sigma^2 & \rho^{|t_3-t_m|}\sigma^2 \\ & & & \ddots & \vdots & \vdots \\ & & & & \sigma^2 & \rho^{|t_{m-1}-t_m|}\sigma^2 \\ & & & & & \sigma^2 \end{bmatrix}$$

여기서 t_j 는 관측값 j 의 시간입니다. 이 구조에서 지정된 시간의 관측값은 동일한 분산(σ^2)을 가집니다. ρ ($-1 < \rho < 1$) 모수는 한 단위 시간 간격을 갖는 두 관측값 간의 상관입니다. 관측값 간의 시간 차이가 증가하면 ρ 모수가 더 크게 거듭제곱되므로 공분산이 감소합니다. 대부분 AR(1)은 제1종 오류 제어를 유지하면서 더 높은 검정력으로 개체 내 상관을 위한 적절한 모형을 제공합니다.

JMP PRO Toeplitz 공분산 구조

Toeplitz 구조에서 고정된 수의 시간 단위로 구분된 관측값은 동일한 상관을 가집니다. AR(1) 상관 구조와 달리 고정된 시간 차이에서 Toeplitz 상관은 임의적입니다. d 단위 간격의 관측값에 대한 상관을 ρ_d 로 나타냅니다. 상관 행렬은 다음과 같이 정의됩니다.

$$C = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{m-1} & \rho_m \\ & 1 & \rho_1 & \cdots & \rho_{m-2} & \rho_{m-1} \\ & & 1 & \cdots & \rho_{m-3} & \rho_{m-2} \\ & & & \ddots & \vdots & \vdots \\ & & & & 1 & \rho_1 \\ & & & & & 1 \end{bmatrix}$$

JMP의 다음 두 옵션은 이 상관 행렬을 사용합니다.

- Toeplitz 구조는 모든 시점의 관측값에 대해 공통 분산 (σ^2)을 가정합니다. 공분산 구조는 $\Sigma = \sigma^2 C$ 입니다.
- Toeplitz 이분산 구조를 사용하면 시점에 따라 분산이 다를 수 있습니다.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_1 \sigma_1 \sigma_2 & \rho_2 \sigma_1 \sigma_3 & \cdots & \rho_{m-1} \sigma_1 \sigma_{m-1} & \rho_m \sigma_1 \sigma_m \\ & \sigma_2^2 & \rho_1 \sigma_2 \sigma_3 & \cdots & \rho_{m-2} \sigma_2 \sigma_{m-1} & \rho_{m-1} \sigma_2 \sigma_m \\ & & \sigma_3^2 & \cdots & \rho_{m-3} \sigma_3 \sigma_{m-1} & \rho_{m-2} \sigma_3 \sigma_m \\ & & & \ddots & \vdots & \vdots \\ & & & & \sigma_{m-1}^2 & \rho_1 \sigma_{m-1} \sigma_m \\ & & & & & \sigma_m^2 \end{bmatrix}$$

JMP PRO Antedependent 공분산 구조

Antedependence 모형은 유연하고 시간에 따라 상관 구조가 달라질 수 있는 일반 모형입니다. 이 모형에서 인접한 시점 $j-1$ 과 j 에서 두 관측값 간의 상관은 고유하며 $\rho_{j|j-1}$ 로 나타냅니다.

인접하지 않은 시점 j 와 j' 에서 관측값 쌍 간의 상관은 그 사이에 있는 모든 인접 상관의 곱입니다.

$$\text{Corr}(y_{ijk}, y_{ij'k} | s_{ik}) = \prod_{k=j}^{j'-1} \rho_{k|k-1}$$

예를 들어 시점 $j=2$ 와 $j'=6$ 에서 관측값 쌍 간의 상관은 $\rho_{21}\rho_{32}\rho_{43}\rho_{54}$ 가 됩니다.

상관 행렬은 다음과 같이 정의됩니다.

$$C = \begin{bmatrix} 1 & \rho_{10} & \rho_{10}\rho_{21} & \cdots & \rho_{10}\cdots\rho_{[m-1][m-2]} & \rho_{10}\cdots\rho_{m[m-1]} \\ & 1 & \rho_{21} & \cdots & \rho_{21}\cdots\rho_{[m-1][m-2]} & \rho_{21}\cdots\rho_{m[m-1]} \\ & & 1 & \cdots & \rho_{32}\cdots\rho_{[m-1][m-2]} & \rho_{32}\cdots\rho_{m[m-1]} \\ & & & \ddots & \vdots & \vdots \\ & & & & 1 & \rho_{m[m-1]} \\ & & & & & 1 \end{bmatrix}$$

JMP의 다음 두 옵션은 이 상관 행렬을 사용합니다.

- Antedependent 등분산 구조는 상관관이 달라질 수 있다는 것은 허용하지만 전체 관측 시간에서 동일한 분산을 가정합니다. 관측값 간의 분산은 항상 σ^2 이고 공분산 행렬은 $\Sigma = \sigma^2 C$ 입니다.
- Antedependent 구조를 사용하면 지정된 시간에서 관측값 간의 분산이 달라질 수 있습니다. 시간 j 에서 관측값 간의 분산은 σ_j^2 로 나타냅니다. 그러면 공분산 행렬이 다음과 같이 정의됩니다.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{10}\sigma_1\sigma_2 & \rho_{10}\rho_{21}\sigma_1\sigma_3 & \cdots & \rho_{10}\cdots\rho_{[m-1][m-2]}\sigma_1\sigma_{m-1} & \rho_{10}\cdots\rho_{m[m-1]}\sigma_1\sigma_m \\ & \sigma_2^2 & \rho_{21}\sigma_2\sigma_3 & \cdots & \rho_{21}\cdots\rho_{[m-1][m-2]}\sigma_2\sigma_{m-1} & \rho_{21}\cdots\rho_{m[m-1]}\sigma_2\sigma_m \\ & & \sigma_3^2 & \cdots & \rho_{32}\cdots\rho_{[m-1][m-2]}\sigma_3\sigma_{m-1} & \rho_{32}\cdots\rho_{m[m-1]}\sigma_3\sigma_m \\ & & & \ddots & \vdots & \vdots \\ & & & & \sigma_{m-1}^2 & \rho_{m[m-1]}\sigma_{m-1}\sigma_m \\ & & & & & \sigma_m^2 \end{bmatrix}$$

JMP PRO 공간 및 시간 변동

단순 모형 $y_i = \mu + e_i$ 를 가정해 보겠습니다. 공간 또는 시간 구조는 오차 항 e_i 를 통해 모델링됩니다. 일반적으로 공간 상관 모형은 $\text{Var}(e_i) = \sigma_i^2$ 및 $\text{Cov}(e_i, e_j) = \sigma_{ij}$ 로 정의할 수 있습니다.

$\mathbf{s}_i$ 가 y_i 의 위치를 나타낸다고 하겠습니다. 이때 $\mathbf{s}_i$ 는 공간 또는 시간을 반영하는 좌표로 지정됩니다. 일반적으로 공간 또는 시간 구조는 공분산이 $\mathbf{s}_i$ 와 $\mathbf{s}_j$ 사이의 유클리드 거리 d_{ij} 에 대한 함수라고 가정하여 제한됩니다. 공분산은 $\text{Cov}(e_i, e_j) = \sigma^2[f(d_{ij})]$ 로 작성할 수 있으며, 여기서 $f(d_{ij})$ 는 관측값 y_i 와 y_j 간의 상관을 나타냅니다.

둘 이상의 위치 좌표가 있는 경우 $f(d_{ij})$ 가 방향에 종속되지 않으면 공분산 구조가 등방성입니다. 그렇지 않으면 구조가 이방성입니다.

JMP PRO 공간 상관 구조

JMP에서 사용할 수 있는 공간 모형의 상관 구조는 아래와 같습니다. 이러한 구조는 제약된 경우 외에는 양수인 ρ 에 의해 파라미터화됩니다.

- 구형

$$f(d_{ij}) = [1 - 1.5(d_{ij}/\rho) + 0.5(d_{ij}/\rho)^3] \times 1_{\{d_{ij} < \rho\}}$$

$$\text{여기서, } 1_{\{d_{ij} < \rho\}} = \begin{cases} 1 \text{ 단, } d_{ij} < \rho \\ 0, \text{ if } d_{ij} \geq \rho \end{cases}$$

- 지수

$$f(d_{ij}) = \exp(-d_{ij}/\rho)$$

- 가우스

$$f(d_{ij}) = \exp(-d_{ij}^2/\rho^2)$$

- 거듭제곱

$$f(d_{ij}) = \rho^{d_{ij}}$$

이방성 모형의 경우 상관 함수는 각 방향에 대해 ρ_k 모수를 포함합니다.

JMP PRO 변동도

공간 과정이 2차 정상 상태인 경우 "**공간 상관 구조**"(399 페이지)에 나열된 구조는 **변동도**를 정의합니다. 지리 통계학에서 차용된 변동도는 공간 변동을 설명하고 추정하는 표준 도구입니다. 여기에서는 반분산을 사용하여 관측값 간의 거리 d_{ij} 에 대한 함수로 공간 변동을 측정합니다.

$Z(\mathbf{s})$ 가 위치 $\mathbf{s}$ 에서 반응의 값을 나타낸다고 하겠습니다. $\mathbf{s}_i$ 와 $\mathbf{s}_j$ 에서 관측값 간의 반분산은 다음과 같이 정의됩니다.

$$\gamma(\mathbf{s}_i, \mathbf{s}_j) = (\text{Var}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)))/2$$

반응에 상수 평균이 있으면 표현식을 다음과 같이 단순화할 수 있습니다.

$$\gamma(\mathbf{s}_i, \mathbf{s}_j) = E[(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2]/2$$

등방성 과정의 경우 반분산은 점 사이의 거리 (h)에만 종속되며 함수를 다음과 같이 지정할 수 있습니다.

$$\gamma(h) = E[(Z(\mathbf{s}_i) - Z(\mathbf{s}_i + h))^2]/2$$

다음은 변동도와 관련된 용어입니다.

너겟 절편으로 정의됩니다. 너겟은 $h=0$ 에서 불연속 이동을 나타냅니다.

문턱 (Sill) 더 먼 거리에 대해 도달한 평탄 시작 구간에서의 반변동도 값으로 정의됩니다. 이 값은 관측값의 분산에 해당합니다. 너겟 효과가 없는 모형에서 문턱 (sill) 은 σ^2 입니다. 너겟 효과가 있는 모형에서 문턱 (sill) 은 $\sigma^2 + c_1$ 이며, 여기서 c_1 은 너겟을 나타냅니다. 부분 문턱 (sill) 은 σ^2 으로 정의됩니다.

범위 반변동도가 문턱 (sill) 에 도달하는 거리로 정의됩니다. 범위 미만의 거리에서는 관측값이 공간적으로 상관됩니다. 범위를 초과하거나 같은 거리의 경우 공간 상관은 사실상 0 입니다. 구형 모형에서 ρ 는 범위입니다. 지수 모형에서는 3ρ 가 실제적 범위입니다. 가우스 모형에서는 $\rho\sqrt{3}$ 이 실제적 범위입니다. 실제적 범위는 공분산이 문턱 (sill) 의 95% 로 축소되는 거리로 정의됩니다.

그림 8.34(380 페이지) 에서 반복 효과 공분산 모수 추정값은 다양한 반변동도 특징을 나타냅니다.

공간 구형 범위 (ρ) 의 추정값입니다.

너겟 c_1 의 척도화된 추정값입니다. 잔차 x 너겟이 c_1 입니다.

잔차 너겟이 없는 모형의 문턱 (sill) 또는 부분 문턱 (sill) 입니다.

JMP^{PRO} 변동도 추정값

지정된 등방성 공간 구조에 대해 추정된 변동도는 "**공간 상관 구조**"(399 페이지)에 나오는 적절한 함수에 대한 관측 데이터의 비선형 최소 제곱 적합을 사용하여 구할 수 있습니다.

JMP^{PRO} 경험적 반분산

경험적 반분산을 계산하려면 변동도 공분산을 위해 선택된 변수에 대한 모든 점 쌍 사이의 거리가 계산됩니다. 거리 범위는 10개의 동일 구간으로 구분됩니다. 데이터에 10개의 구간이 허용되지 않으면 가능한 한 많은 구간이 생성됩니다.

점 쌍으로 구성된 거리 클래스가 생성됩니다. h 번째 거리 클래스는 거리가 h 번째 구간에 속하는 모든 점 쌍으로 구성됩니다.

다음 표기를 고려하십시오.

n 점 쌍의 총 수

C_h 거리가 h 번째로 큰 구간에 속하는 점으로 구성된 거리 클래스

$Z(\mathbf{x})$ $\mathbf{x}$ 에서 반응의 값 ($\mathbf{x}$ = 시간 또는 공간 좌표의 벡터)

$\gamma(h)$ 거리 클래스 C_h 에 대한 반분산

반분산 함수(γ)는 다음과 같이 정의됩니다.

$$\gamma(h) = \begin{cases} \frac{1}{2n} \left\{ \sum_{(x,y) \in C_h} [Z(x) - Z(y)]^2 \right\} & \text{의 경우 } h > 0 \\ \hat{c}_1 & \text{for } h = 0 \end{cases}$$

여기서 $\hat{c}_1$ 은 너겟 효과의 추정값입니다.

JMP PRO Kackar-Harville 수정

고정 효과의 분산 행렬은 항상 Kackar-Harville 수정을 포함하도록 변경됩니다. BLUP의 분산 행렬 및 BLUP과 고정 효과 간의 공분산은 Kackar-Harville 수정 대상이 아닙니다. 이 방법의 근거는 임의 효과에 수준이 많이 있을 때 BLUP에 대한 수정은 계산 및 메모리 사용량이 많을 수 있다는 것입니다. SAS에서는 DDFM=KENWARDROGER가 설정된 경우에만 Kackar-Harville 수정이 고정 효과와 BLUP에 대해 모두 수행됩니다.

공분산 모수와 관련하여 0이 아닌 2차 도함수가 있는 공분산 구조의 경우 Kenward-Roger 공분산 행렬 조정에 2차 항이 포함됩니다. 이 항으로 인해 표준 오차 축소가 발생할 수 있습니다. 또한 조정된 결과 공분산 행렬은 부정부호 행렬이 될 수 있으며 재파라미터화에서 불변하지 않습니다. 1차 Kenward-Roger 공분산 행렬 조정은 계산에서 2차 도함수를 제거합니다. 모든 공간 구조 및 AR(1) 구조는 일반적으로 0이 아닌 2차 도함수를 유도하는 공분산 구조입니다.

JMP에서는 Kenward-Roger 1차 조정을 구현하므로 다음 사항에 유의하십시오.

- 고정 효과 모수만 포함하는 선형 결합의 표준 오차는 PROC MIXED DDFM=KENWARDROGER (FIRSTORDER)와 일치합니다. 이때 PROC MIXED와 JMP에서 사용하는 서로 다른 파라미터화 간의 변환에 주의를 기울여야 합니다.
- BLUP 모수만 포함하는 선형 결합의 표준 오차는 PROC MIXED DDFM=SATTERTHWAITE와 일치합니다.
- 불균형 데이터의 경우 고정 효과와 BLUP를 모두 포함하는 선형 결합의 표준 오차는 DDFM 옵션에 대한 PROC MIXED와 일치하지 않습니다. 그러나 이러한 표준 오차는 DDFM=SATTERTHWAITE 옵션과 DDFM=KENWARDROGER 옵션을 사용하여 구한 값 사이에 있습니다. 균형 데이터의 경우 Kackar-Harville 수정이 null이므로 JMP는 DDFM 옵션에 관계없이 SAS와 일치합니다.

JMP PRO 자유도

고정 효과 모수의 선형 결합만 포함하는 검정의 자유도는 1차 Kenward-Roger 수정을 사용하여 계산됩니다. 따라서 이러한 검정의 JMP 결과는 DDFM=KENWARDROGER(FIRSTORDER) 옵션을 사용하여 PROC MIXED와 일치합니다. 선형 결합에 BLUP가 있는 경우 JMP는 Satterthwaite 근사를 사용하여 자유도를 구합니다. 결과는 이전 단락에서 표준 오차에 대해 설명한 것과 유사한 패턴을 따릅니다.

Kackar-Harville 수정 및 Kenward-Roger DF 방법에 대한 자세한 내용은 Kenward and Roger 연구 자료 (1997)에서 확인하십시오. Satterthwaite 방법은 SAS Institute Inc. (2020d, ch. 83)

다변량 반응 모형 MANOVA 를 사용하여 관계 적합

다변량 모형은 여러 반응(Y 변수)을 효과 집합에 적합시킵니다. 적절한 반응 설계를 사용하여 전체 Y 변수에 대한 함수를 검정할 수 있습니다.

표준 MANOVA(다변량 분산 분석) 모형을 생성하는 것 외에 다음 기법을 사용할 수도 있습니다.

- 반복 측정 분석 - 각 개체에 대해 반복 측정을 수행하고, 전체 측정에서 개체 사이 효과와 개체 내 효과를 모두 분석하려는 경우에 사용할 수 있습니다. 이 다변량 접근 방식은 전체 측정의 상관 구조가 임의적일 때 특히 중요합니다.
- 정준 상관 - 상관관계가 가장 높은 X 및 Y 변수의 선형 결합을 찾을 수 있습니다.
- 판별 분석 - 여러 그룹의 점과 다변량 평균 사이의 거리 계산식을 찾아 점을 해당 점이 속할 가능성이 가장 높은 그룹으로 분류할 수 있습니다. 판별 분석을 더 완벽하게 구현하려면 **판별 플랫폼**을 사용합니다.

다변량 적합은 모수 추정값과 최소 제곱 평균을 보여 주는 가장 기본적인 예비 분석으로 시작됩니다. 그런 다음 Y 변수에 대한 반응 설계를 지정하고 다변량 검정을 수행할 수 있습니다.

목차

다중 반응 모형의 예.....	405
MANOVA 보고서	407
MANOVA 적합 옵션.....	407
반응 지정.....	408
반응 선택 옵션	408
사용자 검정 옵션	409
모형 효과 옵션	409
다변량 검정.....	413
확장된 다변량 보고서	414
다변량 검정 비교	415
단변량 검정 및 구형성 검정	415
반복 측정을 사용하는 다변량 모형	416
반복 측정 예	417
복합 다변량 모형의 예.....	418
판별 분석.....	421
판별 저장 옵션의 예.....	421
MANOVA 분석법에 대한 통계 상세 정보	422
다변량 검정.....	422
근사 F-검정.....	423
정준 상세 정보	424

다중 반응 모형의 예

이 예에서는 Golf Balls.jmp 샘플 데이터 테이블을 사용합니다 (McClave and Dietrich 1988). 세 가지 브랜드의 골프공에 대해 이동 거리 및 내구성 측정값을 비교한 데이터가 포함되어 있습니다. 랜덤으로 표집한 각 브랜드의 골프공 10 개를 로봇 골퍼가 무작위 순서로 쳤습니다. 세 가지 골프공 브랜드에 대해 거리 및 내구성이 동일하다는 가설을 검정하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Golf Balls.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. Distance 와 Durability 를 선택하고 **Y** 를 클릭합니다.
4. Brand 를 선택하고 **추가**를 클릭합니다.
5. **분석법**에서 **MANOVA** 를 선택합니다.

그림 9.1 MANOVA 설정

The screenshot shows the 'Model Specifier' dialog box in JMP. The 'Y' variable is selected as 'Distance' and 'Durability'. The 'X' variable is 'Brand'. The 'Analysis Method' is set to 'MANOVA'. The 'Model Effects' section shows 'Brand' as the selected effect. The 'Number of Rows' is set to 2, and the 'Number of Columns' is set to 2. The 'Number of Rows' and 'Number of Columns' are both set to 2.

6. **실행**을 클릭합니다.

그림 9.2 MANOVA 보고서 창



반응 설계를 아직 지정하지 않았으므로 초기 결과 자체는 흥미롭지 않을 수 있습니다. 반응 설계를 지정하면 다변량 플랫폼에 다변량 추정값 및 검정 테이블이 표시됩니다. 반응 설계를 지정하는 방법에 대한 자세한 내용은 "반응 지정"(408 페이지)에서 확인하십시오.

MANOVA 보고서

MANOVA 보고서 창에는 다음과 같은 요소가 포함되어 있습니다.

MANOVA 적합 빨간색 삼각형 메뉴 저장 옵션을 포함합니다. 자세한 내용은 "**MANOVA 적합 옵션**"(407 페이지)에서 확인하십시오.

반응 지정 다양한 검정에 대한 반응 설계를 지정할 수 있습니다. 자세한 내용은 "**반응 지정**"(408 페이지)에서 확인하십시오.

모수 추정값 각 반응 변수에 대한 모수 추정값을 포함합니다 (표준 오차 또는 t -검정과 같은 상세 정보 제외). 각 반응 변수마다 하나의 열이 있습니다.

최소 제공 평균 모든 반응 열의 전체 최소 제공 평균, 각 명목형 수준의 최소 제공 평균, 평균의 최소 제공 평균 그림을 보고합니다.

부분 상관 X 효과에 대해 조정된 초기 적합에서 구한 잔차의 공분산 행렬 및 부분 상관 행렬을 표시합니다.

전체 E&H 행렬 E 행렬 및 H 행렬을 표시합니다.

- E 행렬의 요소는 잔차의 교차곱입니다.
- H 행렬은 가설 제공합 및 교차곱에 해당합니다.

전체 모형과 모형의 각 효과에 대한 H 행렬이 있습니다. E 및 H 행렬의 주대각 요소는 단변량 F 검정에 대한 가설 (분자) 및 오차 (분모) 제공합에 해당합니다. 주어진 반응 설계에 대한 새 E 행렬과 H 행렬이 이 초기 행렬에서 구성되고, 새 행렬에서 다변량 검정 통계량이 계산됩니다.

MANOVA 적합 옵션

다음과 같은 MANOVA 적합 옵션을 사용할 수 있습니다.

판별 저장 판별 분석을 수행하고 결과를 데이터 테이블에 저장합니다. 자세한 내용은 "**판별 분석**"(421 페이지)에서 확인하십시오.

예측값 저장 예측 반응을 데이터 테이블에 저장합니다.

잔차 저장 잔차를 데이터 테이블에 저장합니다.

모형 대화상자 현재 분석을 위해 완료된 시작 창을 표시합니다.

다음 옵션에 대한 자세한 내용은 **JMP 사용**의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

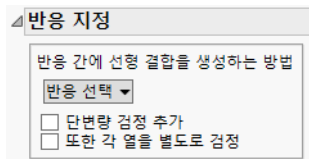
스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

반응 지정

"반응 지정" 패널을 사용하여 다양한 검정에 대한 반응 설계를 지정합니다.

그림 9.3 반응 지정 패널



반응 선택 M 행렬에 대한 선택 옵션을 제공합니다. 자세한 내용은 "[반응 선택 옵션](#)"(408 페이지)에서 확인하십시오.

단변량 검정 추가 조정 및 조정되지 않은 단변량 반복 측정 검정과 다변량 검정을 수행합니다. 반복 측정 모형에 사용됩니다.

또한 각 열을 별도로 검정 각 반응에 대한 단변량 ANOVA 검정과 다변량 검정을 수행합니다.

다음 버튼은 반응 옵션을 선택한 후에만 사용할 수 있습니다.

실행 분석을 수행하고 다변량 추정값 및 검정을 표시합니다. 자세한 내용은 "[다변량 검정](#)"(413 페이지)에서 확인하십시오.

도움말 "반응 지정" 패널에 대한 도움말을 표시합니다.

직교화 행렬을 직교정규화합니다. 직교정규화는 "합"을 제외한 모든 반응 유형에 대한 열 대비 (합이 0) 후에 수행됩니다.

마지막 열 삭제 변환 차원을 축소합니다.

반응 선택 옵션

반응 설계는 M 행렬을 구성합니다. M 행렬의 열은 다변량 분석을 위한 변환 변수 집합을 정의합니다. **반응 선택** 버튼에는 M 행렬을 위한 다음 옵션이 포함되어 있습니다.

반복 측정 합 반응과 대비 반응을 모두 구성하고 실행합니다.

합 단일 값을 제공하는 반응 합입니다.

항등 각각의 개별 반응을 사용합니다 (항등 행렬).

대비 각 반응과 첫 번째 반응을 비교합니다.

다항식 직교 다항식 행렬을 구성합니다.

Helmert 각 반응을 그 아래에 나열된 결합 반응과 비교합니다.

프로파일 각 반응을 그 다음 반응과 비교합니다.

평균 각 반응을 나머지 반응의 평균과 비교합니다.

복합 반응이 두 가지 효과로 구성된 경우 적절한 여러 반응 함수를 생성하고 실행합니다.

사용자 정의 입력한 사용자 M 행렬을 사용합니다.

다변량 회귀에 대해 가장 일반적인 반응 설계는 **반복 측정** 및 **항등**입니다. **대비**, **Helmert**, **프로파일** 및 **평균** 옵션은 동일한 공간을 생성하므로 제공된 검정에 거의 차이가 없습니다. 그러나 "최소 제곱 평균" 및 "모수 추정값" 테이블의 검정과 상세 정보는 각 옵션에 따라 서로 다른 결과를 강조합니다.

반복 측정 및 **복합** 옵션은 반응 효과 이름을 지정하는 대화상자를 표시합니다. 그런 다음 추가 사용자 입력을 기다리지 않고 여러 반응 함수를 적합시킵니다. 그렇지 않은 경우 선택에 따라 제어 판이 확장되어 지정을 구체화할 수 있습니다.

사용자 검정 옵션

사용자 검정 옵션을 사용하여 효과 수준의 사용자 검정을 설정합니다.

참고 : 사용자 검정을 생성하는 방법에 대한 자세한 내용은 "표준 최소 제곱 모형" 장의 "[사용자 검정](#)"(115 페이지)에서 확인하십시오.

모형 효과 옵션

각 효과 이름 옆의 메뉴 아이콘은 다변량 적합에 대한 추가 정보를 요청하는 다음과 같은 옵션을 제공합니다.

검정 상세 정보 다변량 검정 통계량을 생성하는 데 사용되는 $E^{-1}H$ 행렬의 고유값 및 고유 벡터를 표시합니다. 자세한 내용은 "[검정 상세 정보](#)"(410 페이지)에서 확인하십시오.

중심 그림 검정 공간에서 구성된 처음 두 정준 변수에 대한 중심 (다변량 최소 제곱 평균) 을 그림에 표시합니다. 자세한 내용은 "[중심 그림](#)"(411 페이지)에서 확인하십시오.

정준 점수 저장 Canon[1], Canon[2] 등의 변수를 현재 데이터 테이블의 열로 저장합니다. 이 열에는 값과 계산식이 둘 다 있습니다. 예는 "[정준 점수 저장](#)"(411 페이지)에서 확인하십시오. 기술 상세 정보는 "[정준 상세 정보](#)"(424 페이지)에서 확인하십시오.

대비 대비 대화상자에서 지정한 처리 수준의 통계적 대비를 수행합니다.

참고 : 대비 명령은 단일 반응의 회귀와 동일합니다. **최소 제곱 평균 대비** 명령에 대한 설명과 예는 "표준 최소 제곱 모형" 장의 "[최소 제곱 평균 대비 ...](#)"(95 페이지)에서 확인하십시오.

검정 상세 정보

"검정 상세 정보" 보고서에는 전체 모형 또는 지정된 효과의 검정에 대한 정준 상세 정보가 제공됩니다.

고유값 다변량 검정 통계량을 계산하는 데 사용되는 $E^{-1}H$ 행렬의 고유값입니다.

정준 상관 각 고유값과 관련된 정준 상관입니다. 이것은 모형의 다른 모든 효과에 대해 수정된 효과와 변환된 반응의 정준 상관관계입니다.

고유 벡터 $E^{-1}H$ 행렬 또는 이와 동등한 $(E + H)^{-1}H$ 의 고유 벡터입니다.

검정 상세 정보의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Iris.jmp 를 엽니다.

Iris 데이터 (Mardia et al. 1979) 의 Species에는 Virginica, Setosa 및 Versicolor 라는 세 가지 수준이 있습니다. 또한 각 표본에 대해 측정된 네 가지 값 (Petal length, Petal width, Sepal length, Sepal width) 이 있습니다.

2. **분석 > 모형 적합**을 선택합니다.
3. Petal length, Petal width, Sepal length 및 Sepal width 를 선택하고 **Y** 를 클릭합니다.
4. Species 를 선택하고 **추가**를 클릭합니다.
5. **분석법**에서 **MANOVA** 를 선택합니다.
6. **실행**을 클릭합니다.
7. **반응 선택** 버튼을 클릭하고 **항등**을 선택합니다.
8. **실행**을 클릭합니다.
9. "Species" 의 빨간색 삼각형을 클릭하고 **검정 상세 정보**를 선택합니다.
고유값, 고유 벡터 및 정준 상관이 나타납니다.

그림 9.4 검정 상세 정보

Species					
검정	값	근사 F	NumDF	DenDF	Prob>F
Wilks 람다	0.0234386	199.1453	8	288	<.0001*
Pillai 대각합	1.1918988	53.4665	8	290	<.0001*
Hotelling-Lawley	32.47732	582.1970	8	203.4	<.0001*
Roy 최대근	32.191929	1166.9574	4	145	<.0001*
고유값 정준 상관					
32.1919292	0.98482089				
0.28539104	0.47119702				
1.235e-15	0				
-6.174e-16	0				
고유 벡터					
Sepal length	-0.0684059	0.00198791	-0.2350196	0.1176771	
Sepal width	-0.1265612	0.1785267	0.21657608	0.04510419	
Petal length	0.18155288	-0.0768636	0.23964446	0.06563465	
Petal width	0.23180286	0.23417227	-0.2865277	-0.2438953	

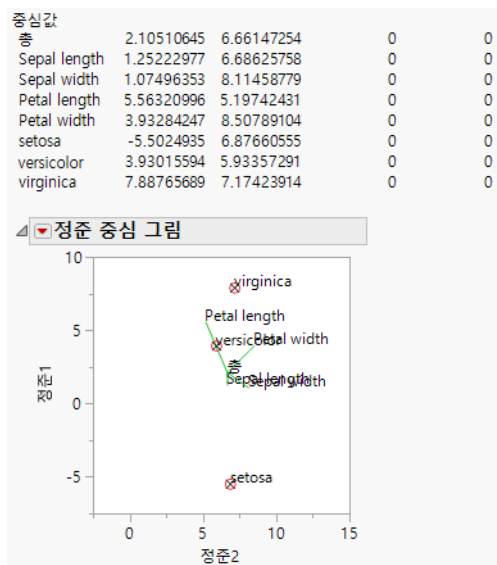
중심 그림

중심 그림 명령은 "Species"의 빨간색 삼각형에서 액세스할 수 있으며, 검정 공간에서 구성된 처음 두 정준 변수에 대한 중심(다변량 최소 제곱 평균)을 그림에 표시합니다(그림 9.5 참조). 첫 번째 정준 축은 세로 축이므로 검정 공간이 1차원일 경우 중심이 세로 축에 정렬됩니다. 중심점은 95% 신뢰 영역에 해당하는 원과 함께 나타납니다(Mardia et al. 1979). 중심 그림이 효과 검정에서 생성되면 검정 중인 효과에 해당하는 원이 빨간색으로 나타납니다. 다른 원은 파란색입니다. 행렬도 반지름은 검정 공간에서 원래 반응 변수의 방향을 보여 줍니다. 자세한 내용은 "[중심 그림 옵션에 대한 상세 정보](#)"(424 페이지)에서 확인하십시오.

중심값 표시 아이콘을 클릭하면 그림 9.5에 표시된 추가 정보가 나타납니다.

고유값이 있는 첫 번째 정준 축은 두 번째 축보다 훨씬 더 많은 분리를 설명합니다. 평균이 잘 분리(구분)되어 있으며, 첫 번째 그룹이 다른 두 그룹보다 더 멀리 떨어져 있습니다. 첫 번째 정준 변수에서는 **petal width** 변수에 비해 **petal length** 변수의 기여도가 높은 것으로 보입니다. 행렬도 반지름 및 관련 고유 벡터를 사용하여 변수 그룹 간의 관계를 확인할 수 있습니다.

그림 9.5 중심 그림 및 중심값



정준 점수 저장

Canon[i] 열을 데이터 테이블에 저장합니다. 여기서 i 는 Y 변수에 대한 i 번째 정준 점수를 나타냅니다. 정준 점수는 다변량 검정 통계량을 생성하는 데 사용되는 $E^{-1}H$ 행렬을 기반으로 계산됩니다. 0이 아닌 고유값에 해당하는 고유 벡터에 대해 정준 점수가 저장됩니다.

정준 상관

정준 상관 분석은 특정 명령은 아니지만 다변량 적합 플랫폼에서 일련의 명령을 사용하여 수행할 수 있습니다.

1. "검정 상세 정보의 예"(410 페이지)에서 1 단계 ~ 8 단계를 수행합니다.
2. "전체 모형"의 빨간색 삼각형을 클릭하고 **검정 상세 정보**를 선택합니다.
3. "전체 모형"의 빨간색 삼각형을 클릭하고 **정준 점수 저장**을 선택합니다.

상세 정보에서 고유값 옆에 정준 상관 값(정준 상관)이 나열됩니다. 저장된 변수는 Canon[1], Canon[2] 등입니다. 이 열에는 값과 계산식이 둘 다 있습니다.

X 측에 대한 정준 변수를 구하려면 X 변수와 Y 변수를 교환하여 동일한 단계를 반복합니다. Canon[n] 열이 데이터 테이블에 이미 추가된 경우에는 Canon[n] 2(또는 다른 숫자)라는 고유한 이름의 새 열이 저장됩니다.

다음은 정준 상관 분석의 추가 예입니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Exercise.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. chins, situps 및 jumps 를 선택하고 Y 를 클릭합니다.
4. weight, waist 및 pulse 를 선택하고 **추가**를 클릭합니다.
5. **분석법**에서 **MANOVA** 를 선택합니다.
6. **실행**을 클릭합니다.
7. **반응 선택** 버튼을 클릭하고 **항등**을 선택합니다.
8. **실행**을 클릭합니다.
9. "전체 모형"의 빨간색 삼각형을 클릭하고 **검정 상세 정보**를 선택합니다.
10. "전체 모형"의 빨간색 삼각형을 클릭하고 **정준 점수 저장**을 선택합니다.

그림 9.6 정준 상관

전체 모형						
검정	값	근사 F	NumDF	DenDF	Prob>F	
Wilks 람다	0.3503905	2.0482	9	34.223	0.0635	
Pillai 대각합	0.6784815	1.5587	9	48	0.1551	
Hotelling-Lawley	1.7719415	2.6397	9	19.053	0.0357*	
Roy 최대근	1.7247387	9.1986	3	16	0.0009*	
고유값 정준 상관						
1.72473874	0.79560815					
0.0419084	0.20055604					
0.00529433	0.07257029					
고유 벡터						
chins	0.02503681	-0.016636	0.05641878			
situps	0.00637953	0.0004622	-0.004547			
jumps	-0.0052909	0.0048507	0.0018787			

출력 정준 변수는 Y 변수의 선형 결합으로 표시된 고유 벡터를 사용합니다. 예를 들어 Canon[1]은 다음과 같이 계산됩니다.

$$0.02503681 * \text{chins} + 0.00637953 * \text{situps} + -0.0052909 * \text{jumps}$$

이 정준 분석은 평균이 0이고 표준편차가 1인 표준화 변수를 생성하지는 않지만 이러한 기능이 있는 계산기를 사용하여 새 표준화 변수를 쉽게 정의할 수 있습니다.

다변량 검정

반응 선택의 예

1. "다중 반응 모형의 예"(405 페이지)에 나오는 1 단계 ~ 6 단계를 완료합니다.
2. 반응 선택 버튼을 클릭하고 **항등**을 선택합니다.
3. 실행을 클릭합니다.

그림 9.7 다변량 검정 보고서

항등						
M 행렬						
M-변환 모수 추정값						
전체 모형						
검정	값	근사 F	NumDF	DenDF	Prob>F	
Wilks 랏다	0.2117852	15.2485	4	52	<.0001*	
Pillai 대각합	0.936487	11.8876	4	54	<.0001*	
Hotelling-Lawley	3.0216575	19.4162	4	30.19	<.0001*	
Roy 최대근	2.7688021	37.3788	2	27	<.0001*	
절편						
검정	값	정확한 F	NumDF	DenDF	Prob>F	
F-검정	3226.6839	41946.890	2	26	<.0001*	
Brand						
검정	값	근사 F	NumDF	DenDF	Prob>F	
Wilks 랏다	0.2117852	15.2485	4	52	<.0001*	
Pillai 대각합	0.936487	11.8876	4	54	<.0001*	
Hotelling-Lawley	3.0216575	19.4162	4	30.19	<.0001*	
Roy 최대근	2.7688021	37.3788	2	27	<.0001*	

M 행렬 보고서에는 사용자가 지정한 반응 설계가 제공됩니다. **M-변환 모수 추정값** 보고서에는 원래 모수 추정값 행렬에 M 행렬의 전치를 곱한 값이 제공됩니다.

참고 : 이 장의 처음에 나왔듯이 행렬 이름 **E**와 **H**는 오차 및 가설 교차곱을 나타냅니다. 반응 설계가 지정된 후 **E**와 **H**는 반응 설계에 의해 변환된 행렬을 나타내며 실제로는 **M'EM** 및 **M'HM**입니다.

확장된 다변량 보고서

다변량 적합에서 가설과 오차로 인한 제곱합은 단일 숫자가 아니라 제곱과 교차곱의 행렬입니다. 그리고 가설 제곱합 및 교차곱 행렬 (**H** 또는 **SSCP**) 이 잔차 행렬 (**E**) 과 비교되는 값의 크기를 측정하는 방법이 많이 있습니다. JMP는 문헌에 일반적으로 설명된 네 가지 다변량 검정을 보고합니다. 정확 유의 수준으로 검정을 원하는 경우 참고 도서에서 표를 검색해야 할 수도 있습니다. 다행히 네 가지 검정 모두 근사 F 검정으로 변환할 수 있습니다. 반응 설계에서 단일 값이 산출되거나 가설이 단일 자유도인 경우 다변량 검정이 동등하며 정확 F 검정을 동일하게 수행합니다. JMP는 검정에 정확 F 라는 라벨을 지정하고, 그렇지 않은 경우에는 근사 F 라벨을 지정합니다.

골프공 예에서는 효과가 하나뿐이므로 전체 모형 검정과 Brand 검정이 동일하며, 근사 F 검정으로 네 가지 다변량 검정을 보여 줍니다. DF가 두 개 (각 반응에 대해 하나씩) 인 단일 절편만 있으므로 이에 대한 F 검정이 정확하며 정확 F 라벨이 지정됩니다.

"전체 모형", "절편" 및 "Brand" 보고서의 빨간색 삼각형 메뉴에는 고유값, 정준 상관, 중심값 목록, 중심 그림 등의 추가 정보를 생성하는 옵션과 정준 변량을 저장할 수 있는 **저장** 명령이 포함되어 있습니다.

효과 (이 예에서는 Brand) 팝업 메뉴에는 대비를 지정하는 옵션도 포함되어 있습니다.

사용자 검정 및 대비 기능은 단일 반응이 있는 회귀의 기능과 동일합니다. 자세한 내용은 "[표준 최소 제곱 모형](#)" 장 (69 페이지) 에서 확인하십시오.

MANOVA 테이블 검정에 대한 계산식은 "[다변량 검정](#)" (422 페이지) 에서 확인하십시오.

확장된 다변량 보고서에는 다음 열이 포함됩니다.

검정 테이블의 각 통계 검정에 라벨을 지정합니다. 반응 함수 값 (**M** 행렬에 지정된 열) 의 수가 1 이거나 효과에 반응 함수당 자유도가 하나뿐인 경우 정확 F 검정이 제공됩니다. 그렇지 않은 경우 Wilks 람다 (Λ), Pillai 대각합, Hotelling-Lawley 대각합 및 Roy 최대근이라는 네 가지 표준 다변량 검정 통계량에 근사 F 검정이 제공됩니다.

값 보고서의 각 다변량 통계 검정에 대한 값입니다.

근사 F 또는 정확 F 다변량 검정에 해당하는 F 값입니다. 반응 설계에서 단일 값이 산출되거나 검정 자유도가 1 인 경우 정확 F 검정입니다.

NumDF 분자 자유도입니다.

DenDF 분모 자유도입니다.

Prob>F F 값에 해당하는 유의 확률입니다.

참고: 구형성 검정 테이블에 대한 자세한 내용은 "[단변량 검정 및 구형성 검정](#)" (415 페이지) 에서 확인하십시오.

다변량 검정 비교

네 가지 표준 다변량 검정은 대개 비슷한 결과를 제공하지만 결과가 다른 경우도 있고 특정 검정이 다른 검정보다 이점이 있을 수도 있습니다. 확실하게 어떤 검정이 가장 우수하다고 할 수는 없습니다. 일반적으로 검정력 측면에서 선호하는 순서는 다음과 같습니다.

1. Pillai 대각합
2. Wilks 람다
3. Hotelling-Lawley 대각합
4. Roy 최대근

귀무가설에서 큰 편차가 있고 고유값이 크게 다를 경우 선호 순서는 반대입니다(Seber 1984).

단변량 검정 및 구형성 검정

반복 측정 모형과 같이 다변량 문제를 단변량 문제로 변환할 수 있는 경우가 있습니다 (Huynh and Feldt 1970). 다음과 같은 경우 다변량 컨텍스트에서 단변량 검정을 사용할 수 있습니다.

- 반응 설계 행렬 $\mathbf{M}$ 이 직교정규 ($\mathbf{M}'\mathbf{M} = \text{항등}$) 인 경우
- $\mathbf{M}$ 이 각 변환의 계수 합이 0 인 둘 이상의 반응을 생성하는 경우
- 구형성 조건이 충족되는 경우. 구형성 조건은 $\mathbf{M}$ -변환 반응 간에 상관관계가 없고 분산이 동일하다는 의미입니다. $\mathbf{M}'\Sigma\mathbf{M}$ 은 항등 행렬에 비례하며, 여기서 Σ 는 $\mathbf{Y}$ 변수의 공분산입니다.

이러한 조건이 유지되면 $\mathbf{E}$ 및 $\mathbf{H}$ 검정 행렬의 주대각 요소가 합산되어 F 검정의 분모와 분자에 대한 단변량 제곱합을 생성합니다. 위의 조건이 유지되지 않으면 오류 메시지가 나타납니다. Golf Balls.jmp의 경우 항등 행렬이 $\mathbf{M}$ 행렬로 지정됩니다. 열 벡터 중심화 및 직교정규화 후에는 항등 행렬을 완전 계수 행렬로 변환할 수 없습니다. 따라서 단변량 요청이 무시됩니다.

단변량 및 구형성 검정의 예

1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Dogs.jmp 를 엽니다.
2. 분석 > 모형 적합을 선택합니다.
3. LogHist0, LogHist1, LogHist3 및 LogHist5 를 선택하고 $\mathbf{Y}$ 를 클릭합니다.
4. drug 와 dep1 을 선택하고 추가를 클릭합니다.
5. "모형 효과 생성" 패널에서 drug 를 선택합니다. "열 선택" 패널에서 dep1 을 선택합니다. 교차를 클릭합니다.
6. "분석법" 에서 MANOVA 를 선택합니다.
7. 실행을 클릭합니다.
8. 단변량 검정 추가 옆의 체크박스를 선택합니다.
9. 반응 선택 메뉴에서 반복 측정을 선택합니다.
"Y 이름" 에 "시간" 을 입력하고 단변량 검정 추가를 선택해야 합니다.

10. **확인**을 클릭합니다.

그림 9.8 구형성 검정

구형성 검정	
Mauchly 기준	0.1752641
카이제곱	16.930873
DF	5
Prob >Chisq	0.0046328

구형성 검정은 구형성 가정을 검정하기 위해 Mauchly 기준을 사용하여 개체 내 효과에 대한 비조정 단변량 F 검정의 적합성을 확인합니다(Anderson 1958). 구형성 검정과 단변량 검정은 항상 직교정규화된 $\mathbf{M}$ 행렬을 사용하여 수행됩니다. 다음 지침을 사용하여 구형성 검정을 해석하십시오.

- 실제 공분산 구조가 구형인 경우 비조정 단변량 F 검정을 사용할 수 있습니다.
- 구형성 검정이 유의하면 실제 공분산 구조가 구형이 아님을 나타냅니다. 따라서 다변량 검정 또는 조정된 단변량 검정을 사용할 수 있습니다.

단변량 F 통계량은 구형성이 없어도 근사 F 분포를 갖지만 분자와 분모의 자유도가 일부 부분 입실론(ϵ) 만큼 감소합니다. Box 연구 자료(1954), Greenhouse and Geisser 연구 자료(1959) 및 Huynh-Feldt 연구 자료(1976)에서는 입실론 자유도 조정을 추정하는 기법을 제공합니다. Muller and Barton 연구 자료(1989)에서는 검정력 연구를 기반으로 Greenhouse-Geisser 버전을 권장합니다.

다변량 보고서의 입실론 조정 검정에는 G-G(Greenhouse-Geisser) 또는 H-F(Huynh-Feldt) 라벨이 지정됩니다. 입실론 조정은 값 열에 표시됩니다.

반복 측정을 사용하는 다변량 모형

다변량 적합의 일반적인 용도 중 하나는 종단 데이터라고도 하는 반복 측정을 사용한 데이터 분석입니다. 시간에 따라 반복해서 개체를 측정하고, 각 시간 측정값이 변수를 구성하도록 데이터가 배열됩니다. 이 상관관계가 구형성이라는 패턴을 만드는 경우 외에는 측정값 간의 상관관계 때문에 데이터를 단일 열에 쌓아 단변량 모형으로 분석하면 안 됩니다. 이 항목에 대한 자세한 내용은 이전에 나온 "**단변량 검정 및 구형성 검정**"(415페이지) 섹션에서 확인하십시오.

반복 측정을 사용하면 분석이 다음과 같은 두 계층으로 구분됩니다.

- 개체 사이 (또는 개체 전체) 효과는 반복 측정 열의 합을 모형 효과에 적합시켜 모델링됩니다. 이 방법은 **합** 반응 함수를 사용하는 것과 같습니다. 이 반응 함수는 1로 구성된 단일 벡터인 $\mathbf{M}$ 행렬입니다.
- 개체 내 효과 (반복 효과 또는 시간 효과)는 반복 측정 열의 차이를 적합시키는 반응 함수로 모델링됩니다. 이 분석은 **대비** 반응 함수 또는 기타 유사한 차분 함수 (**다항식**, **Helmert**, **프로파일** 또는 **평균**)를 사용하여 수행할 수 있습니다. 반복 측정 간의 차이를 모델링하는 경우 차이를 새로운 개체 내 효과 (대개 시간)로 간주하십시오. 모형의 효과를 적합시키는 경우 효과를 개체 내 효과와의 교호작용으로 해석할 수 있습니다. 예를 들어 절편의 효과가 시간

(개체 내) 효과가 되어 반복 측정 간의 전체적인 차이를 보여 줍니다. A 효과가 있는 경우 개체 내 검정은 A* 시간 교호작용에 대한 검정으로 해석되어, 반복 측정 간의 차이가 A 효과에서 어떻게 달라지는지 모델링합니다.

표 9.1(418 페이지)에서는 반응 함수와 모형 효과 사이의 관계를 단변량 모형 지정을 사용할 때와 비교하여 보여 줍니다. **합**(개체 사이) 모형과 **대비**(개체 내) 모형을 모두 사용하여 반응을 단일 열에 쌓고 표준 단변량 적합을 구하여 얻은 검정을 재생성할 수 있어야 합니다.

다음과 같이 직접 / 간접 방법을 사용하여 반복 측정 분석을 수행할 수 있습니다.

- 직접적인 방법은 팝업 메뉴의 "반복 측정" 항목을 사용하는 것입니다. 이 항목을 선택하면 반복 측정에 대해 개체 내 효과를 나타내는 효과의 이름을 지정하도록 요청됩니다. 그런 다음 **대비**와 **합** 반응 함수를 모두 적합시킵니다. 이 방법을 사용하면 개체 내 효과 이름을 사용하여 효과에 적절한 라벨이 지정된다는 이점이 있습니다.
- 간접적인 방법은 두 반응 함수를 개별적으로 지정하는 것입니다. 먼저 **합** 반응 함수를 완료하고 두 번째로 **대비** 또는 차이를 모델링하는 다른 함수 중 하나를 완료합니다. 대비 적합에서 개체 내 효과를 모형 효과와 연결해야 합니다.

반복 측정 예

Cole and Grizzle(1966) 연구를 예로 들어보겠습니다. 결과는 샘플 데이터 폴더의 Dogs.jmp 테이블에 있습니다. 16 마리의 개가 각각 두 개의 수준을 갖는 drug 및 dep1 변수로 정의된 4 개 그룹에 할당됩니다. 종속 변수는 약물을 주입하고 0, 1, 3 및 5 분 후 히스타민의 혈중 농도입니다. 농도 로그는 데이터의 평균과 분산 간의 상관관계를 최소화하는 데 사용됩니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Dogs.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. LogHist0, LogHist1, LogHist3 및 LogHist5 를 선택하고 **Y** 를 클릭합니다.
4. drug 와 dep1 을 선택한 후 **매크로** 메뉴에서 **완전 요인**을 선택합니다.
5. "분석법"에서 **MANOVA** 를 선택합니다.
6. **실행**을 클릭합니다.
7. **반응 선택** 메뉴에서 **반복 측정**을 선택합니다.

"Y 이름"에 "시간"을 입력해야 합니다. **단변량 검정 추가** 체크박스를 선택하면 반응이 단일 열에 쌓인 것처럼 계산되는 단변량 검정이 보고서에 포함됩니다.

8. **확인**을 클릭합니다.

그림 9.9 반복 측정 창

Y 변수를 횡단하는 효과를 나타내려면 항의 이름을 입력하십시오.

Y 이름

시간

☐ 단변량 검정 추가

이 모형은 대비 및 합 반응 설계를 모두 사용하고 지정한 이름을 대비-반응의 효과에 추가하는 것과 동등한 결과를 가져옵니다.

확인

취소

도움말

표 9.1에서는 LogHist0, LogHist1, LogHist3 및 LogHist5 열의 데이터가 단일 Y 열에 쌓인 경우, **합** 및 **대비** 반응 설계에 대한 다변량 검정이 단변량 검정의 라벨 지정 방법에 어떻게 대응하는지 보여 줍니다. 새 행은 명목형 그룹화 변수인 시간으로 식별됩니다.

표 9.1 대응하는 다변량 검정과 단변량 검정

합 M 행렬 개체 사이		대비 M 행렬 개체 내	
다변량 검정	단변량 검정	다변량 검정	단변량 검정
절편	절편	절편	시간
drug	drug	drug	시간*drug
depl	depl	depl	시간*depl

개체 사이 분석이 먼저 생성됩니다. 이 분석은 팝업 메뉴에서 **합**을 선택한 경우와 동일합니다(제목 제외).

다음으로 개체 내 분석이 생성됩니다. 이 분석은 팝업 메뉴에서 **대비**를 선택한 경우와 동일(제목 제외)하지만 보고서의 효과 이름에 개체 내 효과 이름(시간)이 추가되었습니다. 절편 항은 반복 측정값의 전체적인 차이를 추정하므로 이전의 절편 위치는 시간입니다.

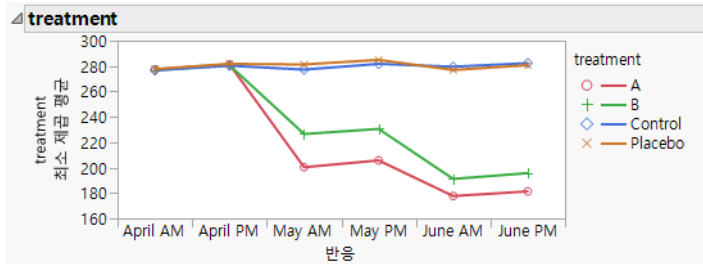
복합 다변량 모형의 예

JMP 는 반복 측정 계층으로 데이터를 처리할 수 있습니다 . Cholesterol.jmp 데이터 테이블을 예로 들어보겠습니다 . 5 명의 참가자로 구성된 그룹은 A, B, Control(대조), Placebo(플라시보)라는 4 개의 treatment(치료) 그룹 중 하나에 속합니다 . 3 개월 동안 한 달에 한 번 아침에 콜레스테롤 수치를 측정하고 오후에 다시 측정했습니다 (가상 데이터). 이 예에서 반응 열은 월별 시간에 따라 순서대로 배열되어 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cholesterol.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .

3. April AM, April PM, May AM, May PM, June AM 및 June PM 을 선택하고 **Y** 를 클릭합니다 .
4. **treatment** 를 선택하고 **추가**를 클릭합니다 .
5. " 분석법 " 옆에서 **MANOVA** 를 선택합니다 .
6. **실행**을 클릭합니다 .

그림 9.10 treatment 그래프



treatment 그래프에서 4 개의 치료 그룹이 매우 유사한 평균 콜레스테롤 값으로 연구를 시작했음을 알 수 있습니다. A 및 B 치료 그룹은 실험 기간이 끝날 때 콜레스테롤 수치가 더 낮은 것으로 보입니다. 대조 및 플라시보 그룹은 변경되지 않았습니다.

7. **반응 선택** 메뉴를 클릭하고 **복합**을 선택합니다 .

데이터 테이블에 반응이 배열되는 방식과 각 반응의 수준 수를 지정하여 이 창을 완료합니다. 콜레스테롤 예에서 시간 열은 월 내에 배열됩니다. 따라서 시간을 하나의 요인으로 지정하고 월 효과를 다른 요인으로 지정합니다. 교호작용 효과 검정은 선택 사항입니다.

8. 그림 9.11 의 옵션을 사용하여 창을 완료합니다 .

그림 9.11 복합 창

반응이 두 요인의 교차로 배열됩니다.
두 개의 요인 이름을 입력하고 가로 수준 수를 설정하십시오.

Time: 2 수준

Month: April AM, April PM, May AM, May PM, June AM, June PM

☒ 교호작용 효과도 함께 생성
☐ 단변량 검정 추가

확인 취소

9. **확인**을 클릭합니다 .

각 효과에 대한 검정이 나타납니다. 그림 9.12 에 보고서의 일부가 표시되어 있습니다. 다음 사항에 유의하십시오 .

- " 시간 " 보고서에는 시간과 treatment 간의 교호작용에 대한 p 값이 0.6038 로 표시되어 교호작용이 유의하지 않음을 나타냅니다. 즉, 오전 (AM) 과 오후 (PM) 사이에 treatment 차

판별 분석

판별 분석은 알려진 반응 값을 기반으로 일원 분류의 일부 수준을 예측하는 방법입니다. 이 기법은 측정 변수가 예측되는 수준의 다변량 평균에 얼마나 가까운지를 기반으로 합니다. 판별 분석은 판별 플랫폼을 사용하여 더 완전하게 구현됩니다. 자세한 내용은 **Multivariate Methods**의 에서 확인하십시오.

JMP에서 측정 변수를 Y 효과로 지정하고 분류 변수를 단일 X 효과로 지정합니다. 다변량 적합 플랫폼은 각 그룹에 대해 공분산이 동일하다는 가정하에 데이터의 평균 및 공분산 행렬 추정값을 제공합니다. MANOVA 플랫폼 이름 옆의 팝업 메뉴에서 **판별 저장** 옵션을 사용하여 판별 정보를 얻을 수 있습니다. 이 명령은 초기 E 및 H 행렬을 사용하여 거리와 확률을 현재 데이터 테이블의 열로 저장합니다.

수준이 k 개인 분류 변수의 경우 JMP는 k 개의 거리 열, k 개의 분류 확률 열, 예측 분류 열 및 두 개의 기타 계산 정보 열을 현재 데이터 테이블에 추가합니다.

판별 저장 옵션의 예

Mardia et al. 연구 자료 (1979) 에 나오는 Fisher 의 Iris 데이터를 검토합니다. 각 표본에 대해 $k = 3$ 개 수준의 종 (species) 과 4 개의 측정값이 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Iris.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. Sepal length, Sepal width, Petal length 및 Petal width 를 선택하고 Y 를 클릭합니다.
4. Species 를 선택하고 **추가**를 클릭합니다.
5. " 분석법 " 옆에서 **MANOVA** 를 선택합니다.
6. **실행**을 클릭합니다.
7. "MANOVA 적합" 의 빨간색 삼각형을 클릭하고 **판별 저장**을 선택합니다.

다음 열이 Iris.jmp 샘플 데이터 테이블에 추가됩니다.

SqDist0 Mahalanobis 거리 계산에 필요한 2 차 형식입니다.

SqDist[setosa] setosa 중심에서 관측값까지의 Mahalanobis 거리입니다.

SqDist[versicolor] versicolor 중심에서 관측값까지의 Mahalanobis 거리입니다.

SqDist[virginica] virginica 중심에서 관측값까지의 Mahalanobis 거리입니다.

Prob0 아래에 사용된 Mahalanobis 거리의 음의 지수 합입니다.

Prob[setosa] setosa 범주에 속할 확률입니다.

Prob[versicolor] versicolor 범주에 속할 확률입니다.

Prob[virginica] virginica 범주에 속할 확률입니다.

Pred Species 확률에서 가장 가능성이 높은 종 (Species) 입니다.

이제 다른 JMP 플랫폼에서 데이터 테이블의 새 열을 사용하여 보고서와 그래프로 판별 분석을 요약할 수 있습니다. 예를 들면 다음과 같습니다.

1. 새 열을 포함하는 업데이트된 Iris.jmp 데이터 테이블에서 **분석 > X로 Y 적합**을 선택합니다.
2. Species를 선택하고 **Y, 반응**을 클릭합니다.
3. Pred Species를 선택하고 **X, 요인**을 클릭합니다.
4. **확인**을 클릭합니다.

"분할표"에 판별 분류가 요약되어 있습니다. 세 개의 오분류가 확인되었습니다.

그림 9.13 예측 및 실제 Species의 분할표

분할표		Species			
개수	총 %	setosa	versicol or	virginic a	합계
Pred Species setosa	행 %				
		50	0	0	50
		33.33	0.00	0.00	33.33
		100.00	0.00	0.00	
versicolor	행 %				
		0	48	1	49
		0.00	32.00	0.67	32.67
		0.00	96.00	2.00	
virginica	행 %				
		0	2	49	51
		0.00	1.33	32.67	34.00
		0.00	4.00	96.00	
합계	행 %				
		50	50	50	150
		33.33	33.33	33.33	

MANOVA 분석법에 대한 통계 상세 정보

- "다변량 검정"
- "근사 F-검정"
- "정준 상세 정보"

다변량 검정

아래에서 **E**는 잔차 교차곱 행렬이고, **H**는 모형 교차곱 행렬입니다. **E**의 주대각 요소는 각 변수의 잔차 제곱합입니다. **H**의 주대각 요소는 각 변수에 대한 모형의 제곱합입니다. 판별 분석 관련 문헌에서는 종종 **E**를 **W**라고 합니다(**W**는 **within**을 나타냄).

다변량 결과 표의 검정 통계량은 $E^{-1}H$ 의 고유값 λ 를 사용해서 구합니다. 다음 목록에서는 각 검정 통계량의 계산 방법을 설명합니다.

참고: 반응 설계가 지정된 후 초기 **E** 및 **H** 행렬의 앞에 **M'**를 곱하고 뒤에 **M**을 곱합니다.

- Wilks 람다

$$\Lambda = \frac{\det(\mathbf{E})}{\det(\mathbf{H} + \mathbf{E})} = \prod_{i=1}^n \left(\frac{1}{1 + \lambda_i} \right)$$

- Pillai 대각합

$$V = \text{Trace}[\mathbf{H}(\mathbf{H} + \mathbf{E})^{-1}] = \sum_{i=1}^n \frac{\lambda_i}{1 + \lambda_i}$$

- Hotelling-Lawley 대각합

$$U = \text{Trace}(\mathbf{E}^{-1}\mathbf{H}) = \sum_{i=1}^n \lambda_i$$

- Roy 최대근

$$\Theta = \lambda_1, \mathbf{E}^{-1}\mathbf{H} \text{의 최대 고유값}$$

$\mathbf{E}$ 와 $\mathbf{H}$ 는 다음과 같이 정의됩니다.

$$\mathbf{E} = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'(\mathbf{X}'\mathbf{X})\mathbf{b}$$

$$\mathbf{H} = (\mathbf{Lb})'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}')^{-1}(\mathbf{Lb})$$

여기서 $\mathbf{b}$ 는 모형 계수의 추정 벡터이며, $\mathbf{A}$ 는 행렬 $\mathbf{A}$ 의 일반화 역행렬을 나타냅니다.

전체 모형 $\mathbf{L}$ 은 0(절편 값)으로 구성된 열로, 행 및 열 개수가 모형의 모수 개수와 같은 단위 행렬과 연결되어 있습니다. 효과에 대한 $\mathbf{L}$ 행렬은 전체 모형 $\mathbf{L}$ 행렬의 일부 행으로 구성된 부분집합입니다.

근사 F-검정

F 값과 자유도를 계산하려면 p 의 순위가 $\mathbf{H} + \mathbf{E}$ 가 되도록 하고 q 의 순위는 $\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}'$ 가 되도록 합니다. 여기서 $\mathbf{L}$ 행렬은 검정 중인 효과와 관련된 $\mathbf{X}'\mathbf{X}$ 의 요소를 식별합니다. v 는 오차 자유도가 되도록 하고 s 는 p 및 q 의 최소값이 되도록 합니다. 또한 $m = 0.5(|p - q| - 1)$ 및 $n = 0.5(v - p - 1)$ 이 되도록 합니다.

표 9.2(423 페이지)에서는 해당 검정 통계량에서 구한 각 근사 F의 계산을 보여 줍니다.

표 9.2 근사 F 통계량

검정	근사 F	분자 DF	분모 DF
Wilks 람다	$F = \left(\frac{1 - \Lambda^{1/t}}{\Lambda^{1/t}} \right) \left(\frac{rt - 2u}{pq} \right)$	pq	$rt - 2u$
Pillai 대각합	$F = \left(\frac{V}{s - V} \right) \left(\frac{2n + s + 1}{2m + s + 1} \right)$	$s(2m + s + 1)$	$s(2n + s + 1)$
Hotelling-Lawley 대각합	$F = \frac{2(sn + 1)U}{s^2(2m + s + 1)}$	$s(2m + s + 1)$	$2(sn + 1)$
Roy 최대근	$F = \frac{\Theta(v - \max(p, q) + q)}{\max(p, q)}$	$\max(p, q)$	$v - \max(p, q) + q$

정준 상세 정보

검정 상세 정보 옵션에 대한 상세 정보

지정된 검정에 대해 "검정 상세 정보" 옵션을 선택하면 고유값, 정준 상관 및 고유 벡터가 보고서에 표시됩니다.

"검정 상세 정보" 옵션에 의해 생성된 정준 상관은 다음과 같이 계산됩니다.

$$\rho_i = \sqrt{\frac{\lambda_i}{1 + \lambda_i}}$$

여기서 λ_i 는 다변량 검정 통계량을 계산하는 데 사용되는 $\mathbf{E}^{-1}\mathbf{H}$ 행렬의 i 번째 고유값입니다. 고유 벡터 라벨이 표시된 행렬은 지정된 검정에 대한 $\mathbf{E}^{-1}\mathbf{H}$ 의 고유 벡터 행렬인 $\mathbf{V}$ 행렬입니다.

참고 : 지정된 검정의 $\mathbf{E}$ 및 $\mathbf{H}$ 행렬은 원래 $\mathbf{E}$ 및 $\mathbf{H}$ 행렬에 대해 $\mathbf{M}'\mathbf{E}\mathbf{M}$ 과 $\mathbf{M}'\mathbf{H}\mathbf{M}$ 을 나타냅니다. $\mathbf{M}$ 행렬은 반응 설계에 의해 정의됩니다. 이 섹션에 사용된 $\mathbf{E}$ 및 $\mathbf{H}$ 는 "다변량 검정"(422 페이지)에 정의되어 있습니다.

중심 그림 옵션에 대한 상세 정보

총 표본 중심 및 효과의 중심값은 다음과 같이 계산됩니다.

$$\text{총} = (c'_1\bar{y}, c'_2\bar{y}, ..., c'_g\bar{y})$$

$$\text{효과}_j = (c'_1\bar{x}_j, c'_2\bar{x}_j, ..., c'_g\bar{x}_j)$$

다음은 각 요소에 대한 설명입니다 .

$$c_i = \left(v_i' \left(\frac{E}{N-r} \right) v_i \right)^{-1/2} v_i$$

N - 관측값 수

v_i - 지정된 검정에 대한 $E^{-1}H$ 의 고유 벡터 행렬인 V 의 i 번째 열

$\bar{x}_j$ - j 번째 효과의 다변량 최소 제곱 평균

$\bar{y}$ - 반응의 전체 평균

g - 0 보다 큰 $E^{-1}H$ 의 고유값 수

r - X 행렬의 계수

참고 : 지정된 검정의 E 및 H 행렬은 원래 E 및 H 행렬에 대해 $M'EM$ 과 $M'HM$ 을 나타냅니다 . M 행렬은 반응 설계에 의해 정의됩니다 . 이 섹션에 사용된 E 및 H 는 "[다변량 검정](#)"(422 페이지)에 정의되어 있습니다 .

효과에 대한 중심 반지름은 다음과 같이 계산됩니다 .

$$d = \sqrt{\frac{\chi_g^2(0.95)}{L(X'X)^{-1}L'}}$$

여기서 g 는 0 보다 큰 $E^{-1}H$ 의 고유값 수이고 , 분모의 L 행렬은 다변량 최소 제곱 평균 계산에서 가져옵니다 .

정준 점수 저장 옵션에 대한 상세 정보

정준 Y 값은 다음과 같이 계산됩니다 .

$$\tilde{Y} = YM'V$$

다음은 각 요소에 대한 설명입니다 .

Y - 반응 변수의 행렬

M' - 반응 설계 행렬의 전치

V - 지정된 검정에 대한 $E^{-1}H$ 의 고유 벡터 행렬

참고 : 지정된 검정의 E 및 H 행렬은 원래 E 및 H 행렬에 대해 $M'EM$ 과 $M'HM$ 을 나타냅니다 . M 행렬은 반응 설계에 의해 정의됩니다 . 이 섹션에 사용된 E 및 H 는 "[다변량 검정](#)"(422 페이지)에 정의되어 있습니다 .

0보다 큰 고유값에 해당하는 고유 벡터에 대해 정준 Y 값이 저장됩니다 .

10 장

로그 선형 분산 모형 반응의 분산 및 평균 모델링

모형 적합 플랫폼의 로그 선형 분산 분석법을 사용하면 회귀 모형으로 반응의 기대값과 분산을 모두 모델링할 수 있습니다. 분산의 로그는 하나의 선형 모형에 적합되고 기대 반응은 여러 선형 모형에 동시에 적합됩니다.

참고 : 값을 추정하려면 잘 설계된 적합 데이터가 많이 있어야 합니다. 평균보다 분산을 적합시키기 위한 데이터가 더 많이 필요합니다.

많은 엔지니어의 실험 목표는 반응 자체를 최대화 또는 최소화하는 것이 아니라 목표 반응에 중점을 두고 변동성을 최소화하는 것입니다. 로그 선형 분산 모형은 분산을 모델링하는 매우 일반적이고 효과적인 방법을 제공하며 반복되지 않은 데이터와 반복이 있는 데이터에 모두 사용할 수 있습니다.

목차

로그 선형 분산 모형 개요	429
산포 효과	429
모형 규격	429
참고	430
로그 선형 분산 분석법 시작	430
예 : 로그 선형 분산 사용	431
로그 선형 보고서	433
로그 선형 플랫폼 옵션	434
열 저장	435
행 진단	435
잔차 검토	435
적합 모형 프로파일링	436
적합 모형 프로파일링의 예	436

로그 선형 분산 모형 개요

로그 선형 분산 모형은 선형 모형을 통해 분산을 모델링하는 방법을 제공합니다. 자세한 내용은 Harvey 연구 자료 (1976), Cook and Weisberg 연구 자료 (1983), Aitken 연구 자료 (1987) 및 Carroll and Ruppert 연구 자료 (1988) 에서 확인하십시오. 선형 모형에는 다음과 같이 평균 반응을 모델링하기 위한 회귀변수 항 외에도 분산의 로그를 모델링하기 위한 회귀변수 항이 있습니다.

평균 모형 : $E(y) = X\beta$

분산 모형 : $\log(\text{Variance}(y)) = Z\lambda$

또는 다음과 같이 지정될 수도 있습니다.

$\text{Variance}(y) = \exp(Z\lambda)$

여기서 X 의 열은 반응 평균에 대한 회귀변수이고 Z 의 열은 반응 분산에 대한 회귀변수입니다. 정규 선형 모형 모수는 β 로 표시되고 λ 는 분산 모형의 모수를 나타냅니다.

로그 선형 분산 모형은 REML을 사용하여 추정됩니다.

산포 또는 로그 분산 효과는 반응 분산의 변화를 모델링할 수 있습니다. 이는 모형 적합 플랫폼에서 로그 선형 분산이라는 적합 분석법에 의해 구현됩니다.

산포 효과

산포 효과 모델링은 Taguchi 프레임워크를 제외하고는 교재에서 아주 자세하게 다루지 않습니다. Taguchi 스타일 실험에서는 외측 배열의 전체 설정에 대해 여러 측정을 수행하고, 이 외측 배열에서 목표를 벗어난 변동성을 측정하는 새 반응을 생성한 후 모형을 적합시켜 변동성을 최소화하는 요인을 찾는 방식으로 처리됩니다. 이 유형의 모델링에는 두 설계의 완전한 Cartesian 곱인 특수 설계가 필요합니다. 이 장에서 설명하는 방법은 더 유연한 모형 기반 방식으로 분산을 모델링합니다. Taguchi가 변동성 모델링을 위해 권장하는 특정 성능 통계량은 $STD = -\log(s)$ 입니다. JMP에 사용되는 방법론에서는 $\log(s^2)$ 이 모델링된 후 평균이 있는 모형과 결합됩니다. $\log(s^2) = 2\log(s)$ 이므로 두 값은 기본적으로 동등합니다.

모형 규격

로그 선형 분산 효과는 "모형 적합" 대화상자에서 효과를 강조 표시한 후 **속성** 드롭다운 메뉴에서 **로그 분산 효과**를 선택하여 지정됩니다. 그러면 해당 효과의 끝에 **& 로그 분산**이 표시됩니다. 이 속성을 사용하면 맨 위의 적합 **분석법**도 **로그 선형 분산**으로 변경됩니다. 반응의 평균과 분산에 모두 효과를 사용하려면 효과를 두 번 지정해야 하며 한 번은 **로그 분산** 옵션을 사용합니다.

로그 분산 속성을 사용하여 지정하는 효과는 모형에서 Z s를 생성하는 효과가 되고, 다른 효과는 모형에서 X s가 됩니다.

참고

평균 모형에 대해 다른 모수가 추정될 때마다 하나 이상의 관측값이 필요하며 더 많으면 좋습니다. 그러나 분산 모수의 경우 적절한 추정값을 얻으려면 각 분산 모수에 대해 몇몇 관측값이 더 필요합니다. 평균보다 분산을 추정하는 데 데이터가 더 많이 필요합니다.

로그 선형 분산 모형은 산포 효과를 적합시키는 매우 유연한 방법이며, 지금까지 문헌에서 받은 것보다 훨씬 더 많은 주의를 기울여야 합니다.

로그 선형 분산 분석법 시작

분석 > 모형 적합을 선택하고 **Y**에 대해 하나 이상의 열을 입력한 후 **분석법** 메뉴에서 **로그 선형 분산**을 선택하여 로그 선형 분산 분석법을 시작합니다.

그림 10.1 로그 선형 분산이 선택된 모형 적합 시작 창

"모형 적합" 창에서 모든 분석법에 공통된 요소에 대한 자세한 내용은 "**모형 규격**" 장(29페이지)에서 확인하십시오. "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오. 여기에서는 로그 선형 분산 분석법에 대한 정보를 제공합니다.

모형 효과에 결측값이 있는 경우 이러한 결측값을 정보 범주로 처리할 수 있습니다. "모형 규격"의 빨간색 삼각형 메뉴에서 "결측값 정보화" 옵션을 선택합니다.

예 : 로그 선형 분산 사용

InjectionMolding.jmp 데이터 테이블에는 네 개의 중앙점이 추가된 7요인 2^{7-3} 부분 요인 설계의 실험 결과가 포함되어 있습니다(Montgomery 1991). 예비 조사에 따르면 평균 반응은 처음 두 요인, 즉 Mold Temperature와 Screw Speed에 따라서만 달라지고, 분산은 Holding Time의 영향을 받는 것 같다고 확인되었습니다.

그림 10.2 사출 성형 데이터

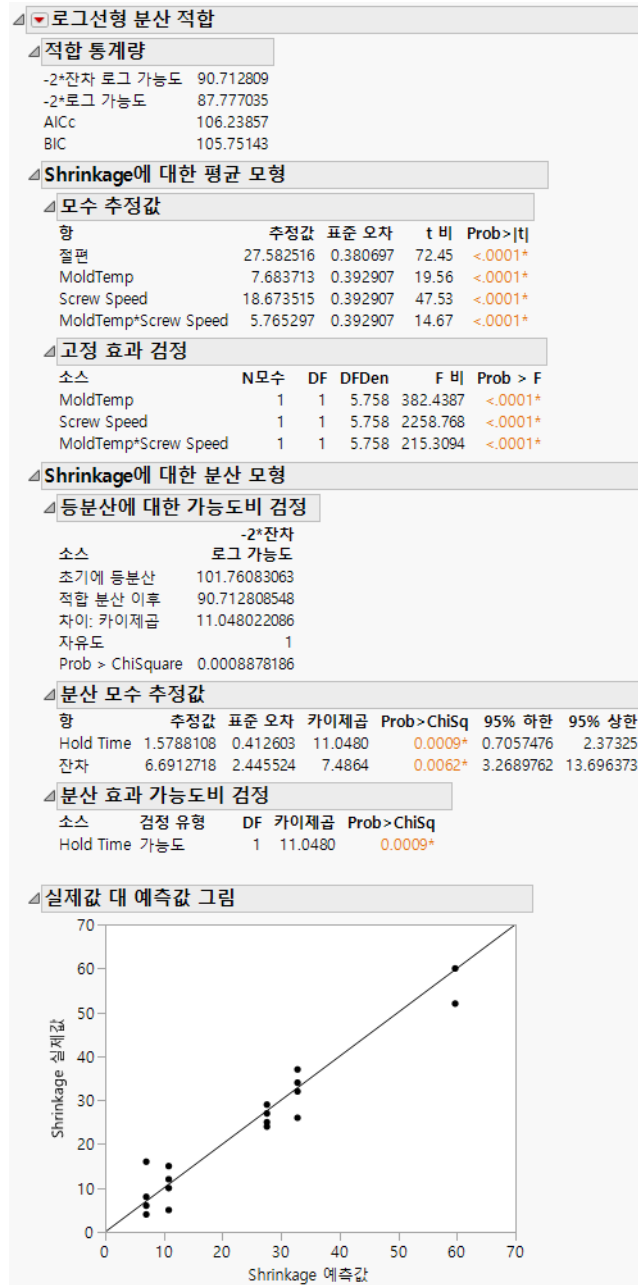


	Pattern	MoldTemp	Screw Speed	Hold Time	Gate Size	Cycle Time	Moisture	Pressure	Shrinkage
1	-----	-1	-1	-1	-1	-1	-1	-1	6
2	++++++	1	-1	-1	-1	1	-1	1	10
3	-----	-1	1	-1	-1	1	1	-1	32
4	++++++	1	1	-1	-1	-1	1	1	60
5	-----	-1	-1	1	-1	1	1	1	4
6	++++++	1	-1	1	-1	-1	1	-1	15
7	-----	-1	1	1	-1	-1	-1	1	26
8	++++++	1	1	1	-1	1	-1	-1	60
9	-----	-1	-1	-1	1	-1	1	1	8
10	++++++	1	-1	-1	1	1	1	-1	12
11	-----	-1	1	-1	1	1	-1	1	34
12	++++++	1	1	-1	1	-1	-1	-1	60
13	-----	-1	-1	1	1	1	-1	-1	16
14	++++++	1	-1	1	1	-1	-1	1	5
15	-----	-1	1	1	1	-1	1	-1	37
16	++++++	1	1	1	1	1	1	1	52
17	0000000	0	0	0	0	0	0	0	25
18	0000000	0	0	0	0	0	0	0	29
19	0000000	0	0	0	0	0	0	0	24
20	0000000	0	0	0	0	0	0	0	27

1. 도움말 > 샘플 데이터 라이브러리를 선택하고 InjectionMolding.jmp 를 엽니다.
2. 분석 > 모형 적합을 선택합니다.

데이터 테이블의 변수에 사전 선택된 역할이 할당되어 있으므로 분석이 자동으로 실행됩니다.

그림 10.3 로그 선형 분산 보고서 창



"Shrinkage에 대한 평균 모형" 보고서는 평균 모형의 모수를 제공하고 "Shrinkage에 대한 분산 모형" 보고서는 분산 모형의 모수를 제공합니다.

로그 선형 보고서

결과 보고서의 위쪽 부분에서는 기대 반응의 적합을 보여 줍니다. "[적합 통계량 보고서 설명](#)"(343 페이지)에 적합 통계량 계산 방법이 설명되어 있습니다. "모수 추정값" 및 "고정 효과 검정"은 REML(제한 최대 가능도)에서 도출되는 점을 제외하고 표준 최소 제곱 분석법의 보고서와 비슷합니다.

그림 10.4 평균 모형 출력

로그선형 분산 적합					
적합 통계량					
-2*잔차 로그 가능도	90.712809				
-2*로그 가능도	87.777035				
AICc	106.23857				
BIC	105.75143				
Shrinkage에 대한 평균 모형					
모수 추정값					
항	추정값	표준 오차	t 비	Prob> t	
절편	27.582516	0.380697	72.45	<.0001*	
MoldTemp	7.683713	0.392907	19.56	<.0001*	
Screw Speed	18.673515	0.392907	47.53	<.0001*	
MoldTemp*Screw Speed	5.765297	0.392907	14.67	<.0001*	
고정 효과 검정					
소스	N모수	DF	DFDen	F 비	Prob > F
MoldTemp	1	1	5.758	382.4387	<.0001*
Screw Speed	1	1	5.758	2258.768	<.0001*
MoldTemp*Screw Speed	1	1	5.758	215.3094	<.0001*

그림 10.5 분산 모형 출력

Shrinkage에 대한 분산 모형							
등분산에 대한 가능도비 검정							
-2*잔차 로그 가능도							
소스	로그 가능도						
초기에 등분산	101.76083063						
적합 분산 이후	90.712808548						
차이: 카이제곱	11.048022086						
자유도	1						
Prob > ChiSquare	0.0008878186						
분산 모수 추정값							
항	추정값	표준 오차	카이제곱	Prob>ChiSq	95% 하한	95% 상한	
Hold Time	1.5788108	0.412603	11.0480	0.0009*	0.7057476	2.37325	
잔차	6.6912718	2.445524	7.4864	0.0062*	3.2689762	13.696373	
분산 효과 가능도비 검정							
소스	검정 유형	DF	카이제곱	Prob>ChiSq			
Hold Time 가능도		1	11.0480	0.0009*			

보고서의 두 번째 부분에서는 분산 모형의 적합을 보여 줍니다. **분산 모수 추정값** 보고서에 추정값 및 관련 통계량이 표시됩니다. 다음과 같은 두 개의 숨겨진 열이 제공됩니다.

- 숨겨진 열 **exp(추정값)** 은 추정값의 지수입니다. 따라서 요인이 +1 및 -1 값을 갖도록 코드화되는 경우 요인의 +1 수준은 분산에 **exp(추정값)** 값을 곱합니다. 마찬가지로 -1 수준은 분산

에 이 열의 역수를 곱합니다. 숨겨진 열을 표시하려면 보고서를 마우스 오른쪽 버튼으로 클릭한 후 표시된 **열** 메뉴에서 해당 열의 이름을 선택합니다.

- **exp(2 추정값 I)** 라벨이 지정된 숨겨진 열은 회귀변수의 범위가 -1 ~ +1 일 때 하위 분산 대비 상위 분산의 비율입니다.

보고서에는 각 추정값의 표준 오차, 카이제곱, p 값 및 프로파일 가능도 신뢰 한계도 표시됩니다. 잔차 모수는 다른 모든 회귀변수가 0일 때 분산의 전체 추정값입니다.

분산 모형의 적합도가 원래 모형보다 유의하게 나은지 확인해 봅니다. 이를 위한 가능도비 검정은 적합 모형을 절편을 제외한 모든 모수가 0인 모형, 즉 등분산 모형과 비교합니다. 이 경우 p 값은 매우 유의합니다. **Hold Time** 변경에 따라 분산이 변경됩니다.

분산 효과 가능도비 검정은 각 항을 차례로 제외하고 모형을 다시 적합시켜 가능도비 검정을 생성합니다. 일반적으로 이러한 검정을 Wald 검정보다 더 신뢰할 수 있습니다.

로그 선형 플랫폼 옵션

"로그선형 분산 적합" 메뉴에는 다음 옵션이 포함되어 있습니다.

열 저장 데이터 테이블에 하나 이상의 열을 생성합니다. 자세한 내용은 "**열 저장**"(435 페이지)에서 확인하십시오.

행 진단 행 진단 그림을 표시합니다. 자세한 내용은 "**행 진단**"(435 페이지)에서 확인하십시오.

프로파일러 프로파일러, 등고선 프로파일러 또는 표면 프로파일러를 엽니다. 자세한 내용은 "표준 최소 제곱 모형" 장의 "**요인 프로파일링**"(150 페이지)에서 확인하십시오.

모형 대화상자 현재 분석을 위해 완료된 시작 창을 표시합니다.

다음 옵션에 대한 자세한 내용은 **JMP 사용**의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

열 저장

"열 저장"에서 다음 옵션을 사용할 수 있습니다.

예측 계산식 평균이라는 새 열을 생성합니다. 새 열에는 지정된 모형에 의해 계산된 평균의 예측값이 포함됩니다.

분산 계산식 분산이라는 새 열을 생성합니다. 새 열에는 지정된 모형에 의해 계산된 분산의 예측값이 포함됩니다.

표준편차 계산식 표준편차라는 새 열을 생성합니다. 새 열에는 지정된 모형에 의해 계산된 표준편차의 예측값이 포함됩니다.

잔차 잔차라는 새 열을 생성합니다. 이 열에는 반응 관측값에서 예측값을 뺀 잔차가 포함됩니다. 자세한 내용은 "[잔차 검토](#)"(435 페이지)에서 확인하십시오.

스튜던트화 잔차 스튜던트화 잔차라는 새 열을 생성합니다. 새 열 값은 잔차를 표준 오차로 나눈 값입니다.

예측값의 표준 오차 예측값의 표준 오차라는 새 열을 생성합니다. 새 열에는 예측값의 표준 오차가 포함됩니다.

개별값의 표준 오차 개별값의 표준 오차라는 새 열을 생성합니다. 새 열에는 개별 예측값의 표준 오차가 포함됩니다.

평균 신뢰 구간 평균의 95% 하한과 평균의 95% 상한이라는 두 개의 새 열을 생성합니다. 새 열에는 예측 평균에 대한 신뢰 구간의 한계가 포함됩니다.

개별값 신뢰 구간 개별값의 95% 하한과 개별값의 95% 상한이라는 두 개의 새 열을 생성합니다. 새 열에는 개별 반응 값에 대한 신뢰 한계가 포함됩니다.

행 진단

"행 진단"에서 다음 옵션을 사용할 수 있습니다.

실제값 대 예측값 그림 Y 에 대한 관측값 대 예측값 그림을 표시합니다. 이 그림은 전체 모형에 대한 레버리지 그림입니다.

스튜던트화 잔차 대 예측값 그림 Y 에 대한 스튜던트화 잔차 대 예측값 그림을 표시합니다.

스튜던트화 잔차 대 행 그림 스튜던트화 잔차 대 행 그림을 표시합니다.

잔차 검토

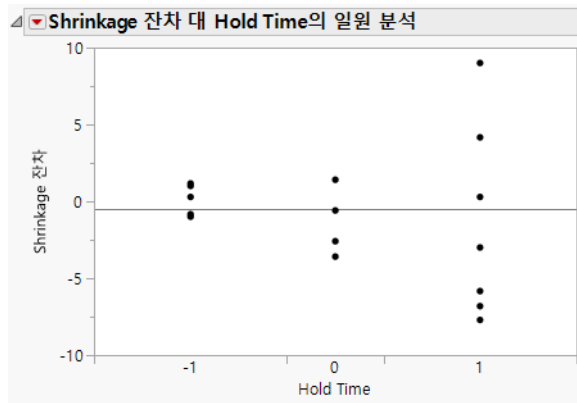
산포 효과를 확인하려면 다음 단계를 수행하십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 `InjectionMolding.jmp`를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.

데이터 테이블의 변수에 사전 선택된 역할이 할당되어 있으므로 분석이 자동으로 실행됩니다.

3. "로그선형 분산 적합"의 빨간색 삼각형을 클릭하고 **열 저장 > 잔차**를 선택합니다.
4. InjectionMolding.jmp 샘플 데이터 테이블의 "열" 패널에서 **Hold Time** 옆에 있는 연속형 아이콘을 마우스 오른쪽 버튼으로 클릭하고 **명목형**을 선택합니다.
5. **분석 > X로 Y 적합**을 선택합니다.
6. Shrinkage 잔차를 선택하고 **Y, 반응**을 클릭합니다.
7. Hold Time 을 선택하고 **X, 요인**을 클릭합니다.
8. **확인**을 클릭합니다.

그림 10.6 잔차 대 산포 효과



이 그림에서 Hold Time이 증가할수록 분산이 위로 올라가는 것을 쉽게 확인할 수 있습니다. 이는 Hold Time을 명목형 요인으로 처리하여 수행됩니다.

적합 모형 프로파일링

프로파일러, 등고선 프로파일러, 또는 표면 프로파일러를 사용하면 적합 모형에 대한 더 많은 정보를 얻을 수 있습니다. 프로파일러 옵션을 선택하려면 "로그선형 분산 적합"의 빨간색 삼각형을 클릭한 후 프로파일러 메뉴에 있는 옵션 중 하나를 선택합니다.

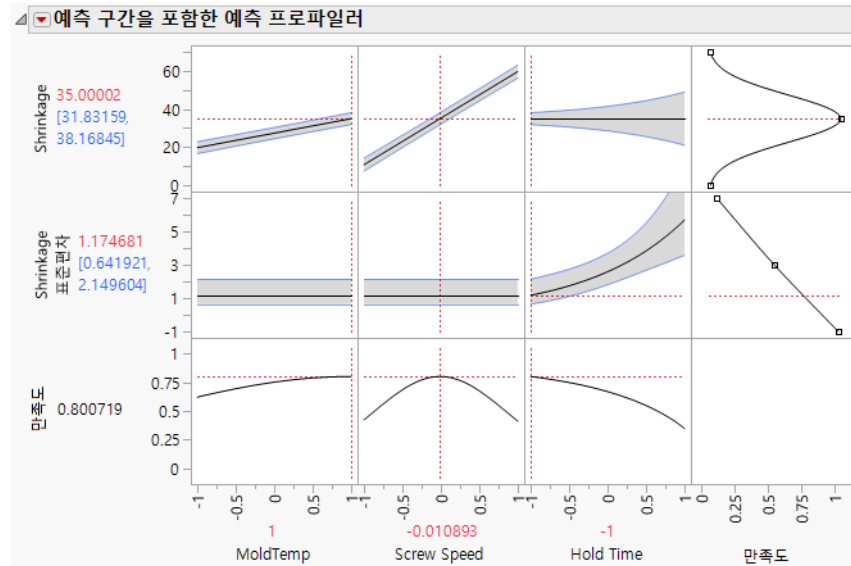
적합 모형 프로파일링의 예

예를 들어 가장 작은 분산에서 반응에 대한 목표 (35)를 달성하는 요인 설정을 찾는 것이 목적이 라고 가정해 보겠습니다. 모형을 적합시키고 보고서 메뉴에서 "프로파일러"를 선택합니다. 예를 들어 그림 10.7에서는 평균의 목표값을 일치시키고 분산을 최소화하도록 설정된 프로파일러를 보여 줍니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 InjectionMolding.jmp를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.

- 데이터 테이블의 변수에 사전 선택된 역할이 할당되어 있으므로 분석이 자동으로 실행됩니다.
3. "로그선형 분산 적합"의 빨간색 삼각형을 클릭하고 **프로파일러 > 프로파일러**를 선택합니다.
 4. "예측 프로파일러"의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 설정**을 선택합니다.
 5. "반응 목표" 창이 나타나면 "최대화"를 "목표값 일치"로 변경합니다.
 6. **확인**을 클릭합니다.
 7. 두 번째 "반응 목표" 창에서 **확인**을 클릭합니다.
 8. "예측 프로파일러"의 빨간색 삼각형을 클릭하고 **최적화 및 만족도 > 만족도 최대화**를 선택합니다.
 9. "예측 프로파일러"의 빨간색 삼각형을 클릭하고 **예측 구간**을 선택합니다.

그림 10.7 예측 구간을 포함한 목표값 일치 및 분산 최소화 프로파일러



평균과 분산(둘 다 로그 분산 분석법으로 모델링됨)의 관계를 확인하는 가장 좋은 방법 중 하나는 평균에 대한 개별 예측 신뢰 구간을 검토하는 것입니다. 정규 신뢰 구간(프로파일러에 기본적으로 표시된 구간)은 개별 예측 신뢰 구간만큼 분산 모형에 대한 정보를 충분히 나타내지 않습니다. 예측 구간은 평균 및 분산 모형을 모두 한 그래프에 표시합니다.

Y 가 모델링된 반응이고 x_n 에서 새 관측값에 대한 예측 구간이 필요한 경우:

$$s^2|_{x_n} = s_Y^2|_{x_n} + s_Y^2|_{x_n}$$

다음은 각 요소에 대한 설명입니다 .

$s^2|x_n$ - x_n 에서 개별 예측에 대한 분산

$s_Y^2|x_n$ - x_n 에서 Y 분포의 분산

$s_{\hat{Y}}^2|x_n$ - $\hat{Y}$ 의 표본 분포 분산이며 평균의 분산

개별 예측의 분산은 Y 분포의 분산을 포함하므로 Y 에 대한 분산이 변화할 때의 효과를 확인할 수 있습니다. 개별 예측 구간은 더 넓을 뿐만 아니라 분산 효과의 변화에 따라 형태를 변경할 수도 있습니다.

11 장

로지스틱 회귀 모형 명목형 또는 순서형 반응에 대한 회귀 모형 적합

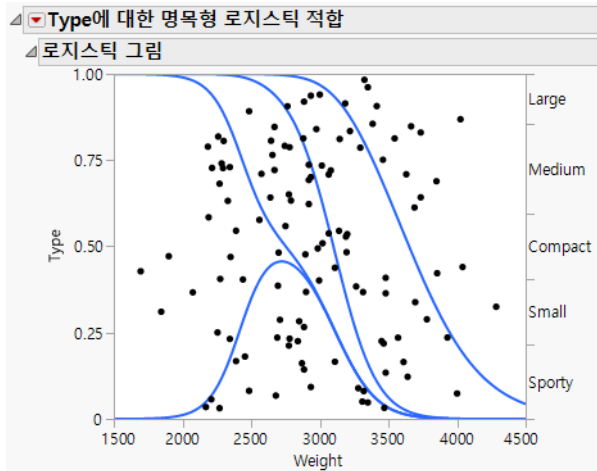
반응 변수에 이산형 값이 있는 경우 모형 적합 플랫폼을 사용하여 로지스틱 회귀 모형을 적합시킬 수 있습니다. 모형 적합 플랫폼은 로지스틱 회귀 모형 적합을 위한 두 가지 분석법을 제공합니다. 사용하는 분석법은 반응 열의 모델링 유형(명목형 또는 순서형)에 따라 달라집니다.

명목형 반응 변수의 경우 명목형 로지스틱 분석법으로 선형 모형을 다중 수준 로지스틱 반응 함수에 적합시킵니다.

순서형 반응 변수의 경우 순서형 로지스틱 분석법으로 누적 반응 확률을 선형 모형의 로지스틱 분포 함수에 적합시킵니다.

두 분석법 모두 모형의 가능도비 검정, 혼동 행렬, ROC 곡선 및 Lift 곡선을 제공합니다. 이항 반응의 경우 명목형 로지스틱 분석법은 승산비(해당 신뢰 구간 포함)도 제공합니다.

그림 11.1 명목형 로지스틱 회귀 모형에 대한 로지스틱 그림



목차

명목형 및 순서형 로지스틱 분석법 개요.....	441
명목형 로지스틱 회귀	441
순서형 로지스틱 회귀	441
로지스틱 회귀 모형을 적합시키는 기타 JMP 플랫폼	442
로지스틱 회귀의 예	442
명목형 로지스틱 회귀의 예	442
순서형 로지스틱 회귀의 예	444
명목형 및 순서형 로지스틱 분석법 시작.....	446
로지스틱 회귀 모형의 검증.....	447
로지스틱 적합 보고서	448
전체 모형 검정	449
적합 상세 정보	450
적합 결여 검정	450
로지스틱 적합 플랫폼 옵션.....	451
명목형 및 순서형 적합 옵션	451
명목형 적합 옵션	453
순서형 적합 옵션	454
로지스틱 회귀의 추가 예.....	455
역추정 예측의 예	455
명목형 로지스틱 모형에 대한 효과 요약의 사용 예.....	457
2차 순서형 로지스틱 모형의 예	459
예: 여러 열의 개수 쌓기.....	462
명목형 및 순서형 로지스틱 분석법에 대한 통계 상세 정보.....	463
로지스틱 회귀 모형	463
승산비	464
통계적 검정의 관계	465

명목형 및 순서형 로지스틱 분석법 개요

로지스틱 회귀는 범주형 Y 반응 변수 수준의 확률을 하나 이상의 X 효과에 대한 함수로 모델링합니다. 모형 적합 플랫폼은 로지스틱 회귀 모형 적합을 위한 두 가지 분석법을 제공합니다. 사용하는 분석법은 반응 열의 모델링 유형(명목형 또는 순서형)에 따라 달라집니다.

로지스틱 회귀 모형 적합에 대한 자세한 내용은 Walker and Duncan 연구 자료(1967), Nelson 연구 자료(1976), Harrell 연구 자료(1986) 및 McCullagh and Nelder 연구 자료(1989)에서 확인하십시오.

로지스틱 회귀 모형의 파라미터화에 대한 자세한 내용은 "로지스틱 회귀 모형"(463 페이지)에서 확인하십시오.

명목형 로지스틱 회귀

반응 변수의 모델링 유형이 명목형이면 플랫폼에서 최대 가능도를 사용하여 선형 모형을 다중 수준 로지스틱 반응 함수에 적합시킵니다. 따라서 하나를 제외한 모든 반응 수준은 X 효과의 값이 주어졌을 때 반응 수준의 확률을 나타내는 로지스틱 곡선으로 모델링됩니다. 최종 반응 수준의 확률은 1에서 다른 적합 확률의 합을 뺀 값입니다. 결과적으로 X 효과의 모든 값에서 반응 수준에 대한 적합 확률의 합은 1입니다.

반응 변수가 이항인 경우 모형 적합 창에서 목표 수준을 설정하여 확률을 모델링할 수준을 지정할 수 있습니다. 기본적으로 모형은 반응 변수의 상위 수준 확률을 추정합니다.

명목형 반응 변수의 모형 적합에 대한 자세한 내용은 "통계 상세 정보" 장의 "명목형 반응"(489 페이지)에서 확인하십시오.

순서형 로지스틱 회귀

반응 변수의 모델링 유형이 순서형이면 플랫폼에서 최대 가능도를 사용하여 누적 반응 확률을 선형 모형의 로지스틱 함수에 적합시킵니다. 각 반응 수준과 같은 수준 또는 그 아래 수준의 반응이 나타날 누적 확률은 곡선으로 모델링됩니다. 이 곡선들은 오른쪽 또는 왼쪽으로 이동된다는 점만 제외하고는 각 수준마다 동일합니다.

팁 : 반응 수준이 많은 경우 순서형 모형이 명목형 모형보다 적합 속도가 훨씬 빠르고 메모리를 적게 사용합니다.

순서형 반응 변수가 있는 모형의 적합에 대한 자세한 내용은 "통계 상세 정보" 장의 "순서형 반응"(491 페이지)에서 확인하십시오.

로지스틱 회귀 모형을 적합시키는 기타 JMP 플랫폼

JMP에서는 다음과 같은 다양한 방법을 사용하여 로지스틱 회귀 모형을 적합시킬 수 있습니다.

- 연속형 주효과 하나만 있는 로지스틱 회귀 모형을 적합시키려면 X로 Y 적합 플랫폼을 사용하여 각 효과에 대한 누적 로지스틱 확률도를 표시할 수 있습니다. 자세한 내용은 [기본 분석](#)의 예에서 확인하십시오.
- 로지스틱 회귀 모형에서 변수 선택을 수행하려면 모형 적합 플랫폼의 단계별 분석법을 사용할 수 있습니다. 자세한 내용은 ["단계별 회귀 모형"](#) 장 (229 페이지) 에서 확인하십시오.
- 로짓 연결 이외의 연결 함수를 사용하는 로지스틱 회귀 모형을 적합시키려면 모형 적합 플랫폼의 일반화 선형 모형 분석법을 사용할 수 있습니다. 자세한 내용은 ["일반화 선형 모형"](#) 장 (467 페이지) 에서 확인하십시오.
-  로지스틱 회귀 모형에서 변수 선택을 수행하고 별점 로지스틱 회귀 모형을 적합시키려면 모형 적합 플랫폼의 일반화 회귀 분석법을 사용할 수 있습니다. 자세한 내용은 ["일반화 회귀 모형"](#) 장 (263 페이지) 에서 확인하십시오.

로지스틱 회귀의 예

이 섹션에는 모형 적합 플랫폼의 각 로지스틱 회귀 분석법 (명목형 및 순서형)에 대한 두 가지 예가 포함되어 있습니다.

- ["명목형 로지스틱 회귀의 예"](#) (442 페이지)
- ["순서형 로지스틱 회귀의 예"](#) (444 페이지)

명목형 로지스틱 회귀의 예

다양한 가열 시간 및 유지 시간으로 준비된 금속 주괴에 대한 실험을 수행한 후 시작 준비 상태를 테스트했습니다. 자세한 내용은 Cox and Snell 연구 자료 (1989)에서 확인하십시오. 데이터는 Ingots.jmp 샘플 데이터 테이블에 포함되어 있습니다. 이 예에서 모형 적합 플랫폼은 heat 및 soak 회귀변수가 있는 로지스틱 회귀 모형을 사용하여 Ready 반응의 확률을 적합시킵니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Ingots.jmp를 엽니다.
범주형 변수 ready의 값 Ready와 Not Ready는 주괴를 사용할 준비가 되었는지 여부를 나타냅니다.
2. **분석 > 모형 적합**을 선택합니다.
3. ready를 선택하고 **Y**를 클릭합니다.
명목형 모델링 유형의 열을 선택했으므로 모형 적합 분석법이 "명목형 로지스틱"으로 업데이트됩니다.
ready는 두 개의 수준만 있는 명목형 열이므로 "목표 수준" 옵션이 나타납니다. 이 옵션을 사용하면 확률을 모델링할 반응 수준을 지정할 수 있습니다.

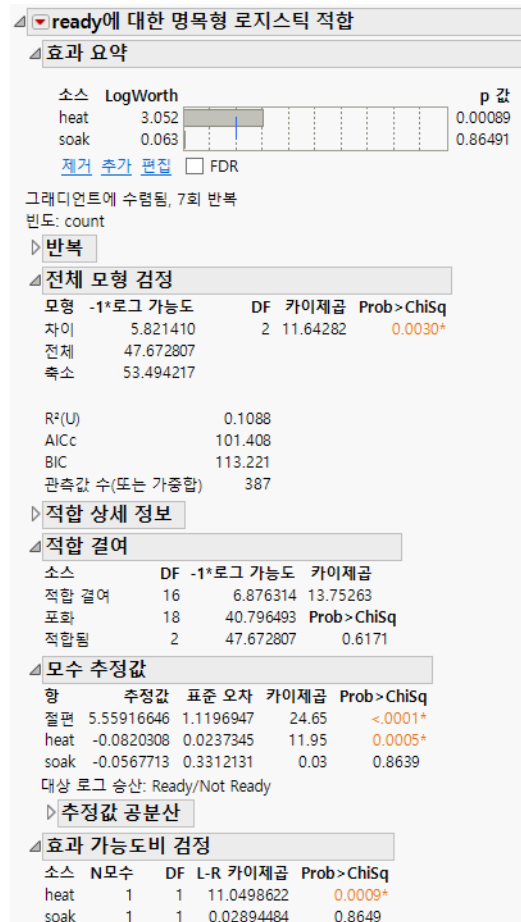
4. " 목표 수준 " 목록에서 "Ready" 를 선택합니다 .

이 모형에서는 목표 수준이 Ready 이므로 Ready 반응의 확률을 모델링합니다 .

5. heat 와 soak 를 선택하고 **추가**를 클릭합니다 .
6. Count 를 선택하고 **빈도**를 클릭합니다 .
7. **실행**을 클릭합니다 .

적합 프로세스가 수렴되면 명목형 회귀 보고서가 나타납니다 .

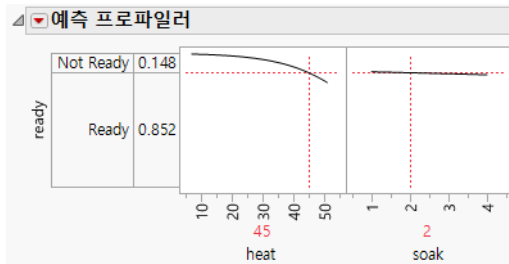
그림 11.2 명목형 로지스틱 적합 보고서



"전체 모형 검정" 보고서에서 카이제곱 통계량(11.64)의 p 값(0.0030)이 작으며 이는 전체 모형이 유의함을 나타냅니다. 그러나 **soak**에 대한 모수 추정값의 p 값은 0.8639이며 이는 유지 시간이 통계적으로 유의하지 않음을 나타냅니다.

8. "ready 에 대한 명목형 로지스틱 적합 " 옆의 빨간색 삼각형을 클릭하고 **프로파일러**를 선택합니다 .

그림 11.3 예측 프로파일러



heat가 45로 설정되고 soak가 2로 설정되면 ready의 확률이 0.85입니다.

이 시점에서 시작 준비 확률이 주어지고 특정 유지 시간에 대한 가열 시간을 찾기 위해 역추정 예측을 사용하는 데 관심이 있을 수 있습니다. 이에 대한 예는 "[역추정 예측의 예](#)"(455페이지)에서 계속됩니다.

순서형 로지스틱 회귀의 예

다양한 치즈 첨가물 (A ~ D) 이 치즈 맛에 영향을 주는지 여부를 테스트하기 위한 실험이 수행되었습니다. 시식 패널이 맛을 측정하여 1(매우 싫음) ~ 9(최고의 맛) 의 순서형 척도를 기록했습니다. 자세한 내용은 McCullagh and Nelder 연구 자료 (1989) 에서 확인하십시오. 데이터는 Cheese.jmp 샘플 데이터 테이블에 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cheese.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. 응답을 선택하고 **Y** 를 클릭합니다.

순서형 모델링 유형의 열을 선택했으므로 모형 적합 분석법이 "순서형 로지스틱" 으로 업데이트됩니다.

4. Cheese 를 선택하고 **추가**를 클릭합니다.
5. Count 를 선택하고 **빈도**를 클릭합니다.
6. **실행**을 클릭합니다.

그림 11.4 순서형 로지스틱 적합 보고서

Response에 대한 순서형 로지스틱 적합				
빈도: Count				
전체 모형 검정				
모형	-1*로그 가능도	DF	카이제곱	Prob>ChiSq
자이	74.22695	3	148.4539	<.0001*
전체	355.67395			
축소	429.90090			
R ² (U)	0.1727			
AICc	734.695			
BIC	770.061			
관측값 수(또는 가중합)	208			
적합 상세 정보				
적합 결여				
소스	DF	-1*로그 가능도	카이제곱	
적합 결여	21	10.15410	20.30819	
포화	24	345.51986	Prob>ChiSq	
적합됨	3	355.67395	0.5018	
모수 추정값				
항	추정값	표준 오차	카이제곱	Prob>ChiSq
절편[1]	-4.6051335	0.4267239	116.46	<.0001*
절편[2]	-3.5499488	0.3073868	133.37	<.0001*
절편[3]	-2.4503882	0.2398702	104.36	<.0001*
절편[4]	-1.381774	0.2001699	47.65	<.0001*
절편[5]	-0.0455233	0.1803299	0.06	0.8007
절편[6]	0.90648953	0.1886609	23.09	<.0001*
절편[7]	2.408151	0.2380517	102.34	<.0001*
절편[8]	3.96800057	0.3466551	131.02	<.0001*
Cheese[A]	-0.8622328	0.2289236	14.19	0.0002*
Cheese[B]	2.48960592	0.2703151	84.82	<.0001*
Cheese[C]	0.8476542	0.2280359	13.82	0.0002*
효과 가능도비 검정				
소스	N모수	DF	L-R 카이제곱	Prob>ChiSq
Cheese	3	3	148.453899	<.0001*

이 예의 모형 적합을 통해 절편만 있는 모형에 대한 -1*로그 가능도 429.9가 전체 모형에 대해 355.67로 감소합니다. 이 감소로 인해 자유도가 3인 전체 모형의 가능도비 카이제곱 통계량이 148.45로 산출됩니다. 따라서 인지된 치즈 맛의 차이는 매우 유의합니다.

가장 선호하는 치즈 첨가물은 음의 모수 추정값이 가장 큰 첨가물입니다. Cheese[D]는 설계 행렬의 자체 열이 없으므로 모수 추정값 보고서에 표시되지 않습니다. 그러나 Cheese D의 효과는 다른 효과의 음의 합계로 계산할 수 있으며 표 11.1에 표시됩니다.

표 11.1 Cheese.jmp의 치즈 첨가물 선호도

Cheese	추정값	환경 설정
A	-1	두 번째 호감도
B	2.4896	최저 호감도
C	0.8477	세 번째 호감도
D	-1	최고 호감도

명목형 로지스틱 모형과의 비교

"적합 결여" 보고서에서는 모형이 데이터를 잘 적합시키는지에 대한 검정을 보여 줍니다.

순서형 문제로 볼 때, 처음 8개의 반응 수준에는 각각 절편이 있지만 네 개의 Cheese 수준에는 세 개의 모수만 있습니다. 따라서 순서형 모형의 자유도는 3입니다. 순서형 모형은 적합 결여 검정에서 적합 모형입니다.

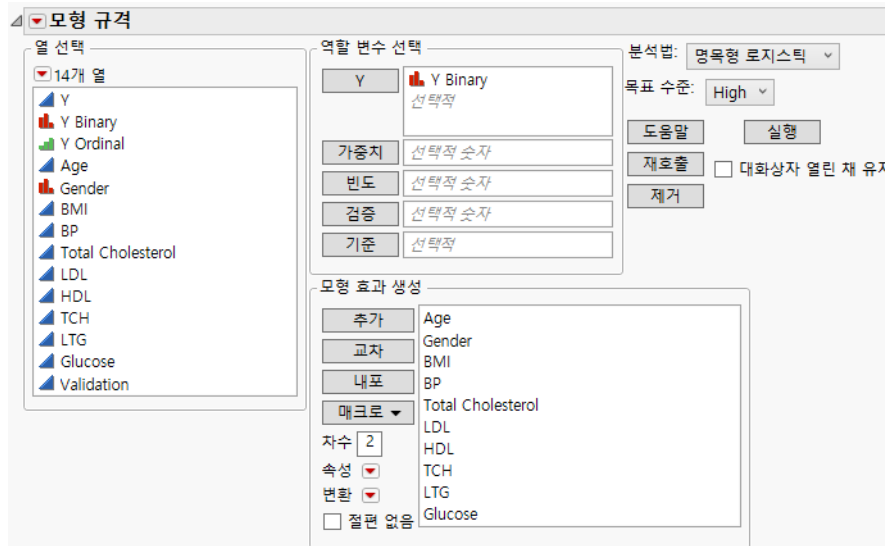
명목형 문제로 볼 때, 처음 8개의 반응 수준에는 각각 절편 및 네 개의 Cheese 수준에 대한 세 개의 모수가 있습니다. 따라서 명목형 모형의 자유도는 $8 \times 3 = 24$ 입니다. 결과적으로 명목형 모형은 적합 결여 검정에서 포화 모형입니다.

이 예에서 순서형 모형의 적합 결여 검정은 명목형 모형을 비교 대상으로 순서형 반응 모형을 검정합니다. 적합 결여가 유의하지 않으면 순서형 모형이 적절하다고 간주합니다.

명목형 및 순서형 로지스틱 분석법 시작

분석 > 모형 적합을 선택하고 **Y**에 대해 하나 이상의 비연속형 열을 입력하여 명목형 로지스틱 및 순서형 로지스틱 분석법을 시작합니다. **Y**에 여러 열이 입력되면 각 반응에 대한 모형이 적합됩니다.

그림 11.5 명목형 로지스틱이 선택된 모형 적합 시작 창



"모형 적합" 창에서 모든 분석법에 공통된 요소에 대한 자세한 내용은 **"모형 규격"** 장(29 페이지)에서 확인하십시오. "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오. 여기에서는 명목형 로지스틱 및 순서형 로지스틱 분석법에 대한 정보를 제공합니다.

모형 효과에 결측값이 있는 경우 이러한 결측값을 정보 범주로 처리할 수 있습니다. "모형 규격"의 빨간색 삼각형 메뉴에서 "결측값 정보화" 옵션을 선택합니다.

절편 항이 없는 모형을 지정하려면 "모형 적합" 창의 "모형 효과 생성" 패널에서 "절편 없음" 옵션을 선택합니다. 순서형 로지스틱 분석법에서는 "절편 없음" 옵션을 사용할 수 없습니다.

로지스틱 회귀 모형의 관심 사건은 "모형 적합" 창의 "목표 수준" 옵션에 의해 정의됩니다. 이 옵션은 명목형 로지스틱 분석법에서 이항 반응 열을 지정하는 경우에만 사용할 수 있습니다.

참고 : 모형 적합 플랫폼의 로지스틱 분석법을 사용하려면 모든 반응이 하나의 열에 있도록 데이터가 쌓인 형식이어야 합니다. 데이터가 여러 열에 포함되도록 형식이 지정되는 경우도 있습니다. 여러 열의 반응을 단일 열 반응으로 변환하는 예는 "[예: 여러 열의 개수 쌓기](#)"(462 페이지)에서 확인하십시오.

JMP PRO 로지스틱 회귀 모형의 검증

검증은 데이터 집합의 일부를 사용하여 모형 모수를 추정하고 다른 부분을 사용하여 모형의 예측 능력을 평가하는 프로세스입니다.

- 훈련 데이터 집합은 모형 모수를 추정하는 데 사용됩니다.
- 검증 데이터 집합은 모형 적합에서 모형의 예측 능력을 평가하거나 검증하는 데 사용됩니다.
- 테스트 데이터 집합은 모형의 예측 능력에 대한 독립된 최종 평가입니다. 테스트 데이터 집합은 검증 열을 사용하는 경우에만 사용할 수 있습니다.

훈련, 검증 및 테스트 데이터 집합은 원래 데이터의 부분집합으로 생성됩니다. 이는 모형 적합 시작 창에서 검증 열을 사용하여 수행됩니다.

검증 열의 값은 데이터가 분할되는 방법과 검증에 사용되는 방법을 결정합니다.

- 열에 두 개의 구분되는 값이 있으면 훈련 및 검증 데이터 집합이 생성됩니다.
- 열에 세 개의 구분되는 값이 있으면 훈련, 검증 및 테스트 데이터 집합이 생성됩니다.
- 열에 네 개 이상의 구분되는 값이 있거나 값이 하나뿐이면 검증이 수행되지 않습니다.

검증 열을 사용하면 "적합 상세 정보" 보고서에 훈련, 검증 및 테스트 데이터 집합에 대한 모형 적합 통계량이 제공됩니다. 또한 각 훈련, 검증 및 테스트 데이터 집합에 대해 별도의 ROC 곡선, Lift 곡선 및 혼동 행렬이 있습니다.

JMP 모델링 플랫폼에서 검증 열이 사용되는 방법에 대한 자세한 내용은 **Predictive and Specialized Modeling**의 에서 확인하십시오.

로지스틱 적합 보고서

명목형 로지스틱 또는 순서형 로지스틱 분석법을 사용하여 모형을 적합시키면 각각 명목형 로지스틱 적합 보고서 또는 순서형 로지스틱 적합 보고서가 생성됩니다. 기본적으로 이러한 보고서에는 다음 보고서가 포함됩니다.

효과 요약 모형에서 효과를 추가하거나 제거할 수 있는 대화식 보고서입니다. 자세한 내용은 "표준 최소 제곱 모형" 장의 "**효과 요약 보고서**"(168 페이지)에서 확인하십시오.

로지스틱 그림 (모형이 단일 연속형 효과로 구성된 경우에만 사용 가능) 로지스틱 확률도에서는 로지스틱 모형의 적합 상태를 보여 줍니다. 가로 축의 각 값에서 세로 방향의 확률 척도가 각 반응 범주의 확률로 분할됩니다. 이 확률은 곡선 사이의 세로 거리로 측정되며 모든 반응 범주 확률의 합계는 1 이 됩니다.

로지스틱 그림에서 점은 데이터 테이블의 관측값을 나타냅니다. 각 점의 가로 위치는 연속형 요인의 값에 따라 결정됩니다. 각 점의 세로 위치는 반응 범주의 값에 해당하는 곡선 사이에 오도록 무작위로 선택됩니다. 고정된 난수 시드값이 사용되므로 세로 위치는 동일한 모형을 여러 번 적합시켜도 달라지지 않습니다.

반복 (명목형 로지스틱 분석법에서만 사용 가능) 모형 적합을 시작한 후 반복 추정 프로세스가 시작되고 반복별로 보고됩니다. 적합 프로세스가 완료되면 반복 기록 보고서를 열어 반복 단계를 확인할 수 있습니다.

전체 모형 검정 절편 모수를 제외한 모든 회귀변수 효과가 생략된 모형과 전체 모형 적합을 비교하는 검정을 표시합니다. 이 검정은 연속형 반응에 대한 분산 분석 테이블과 유사합니다. 전체 모형 검정 보고서에 대한 자세한 내용은 "**전체 모형 검정**"(449페이지)에서 확인하십시오.

적합 상세 정보 모형에 대한 다양한 적합 측도를 표시합니다. 자세한 내용은 "**적합 상세 정보**"(450 페이지)에서 확인하십시오.

적합 결여 (X 효과와 관련한 반복된 점이 있고 모형이 포화되지 않은 경우에만 사용 가능) 모형에 항이 더 필요한지 여부를 확인하는 적합 결여 검정 (적합도 검정이라고도 함) 을 표시합니다. 자세한 내용은 "**적합 결여 검정**"(450 페이지)에서 확인하십시오.

모수 추정값 모수 추정값, 표준 오차 및 관련된 가설 검정을 표시합니다. "추정값 공분산" 보고서에는 모수 추정값의 분산 및 공분산이 포함됩니다.

참고 : "추정값 공분산" 보고서는 순서형 반응 변수가 아니라 명목형 반응 변수에 대해서만 표시됩니다.

특이성 상세 정보 (모형 항 사이에 선형 종속성이 있는 경우에만 표시됨) 모형 항이 충족하는 선형 함수가 포함된 보고서를 표시합니다.

효과 가능도비 검정 가능도비 카이제곱 검정은 전체 모형과 검정할 가설의 제약을 받는 모형 간의 로그 가능도 차이의 두 배로 계산됩니다. 제약이 있는 모형은 효과를 포함하지 않는 모형입니다. 이러한 검정은 각 검정마다 별도의 반복 집합이 필요하므로 시간이 걸릴 수 있습니다.

참고 : 가능도비 검정은 완료하는 데 20 초 미만이 걸릴 것으로 예상되는 경우 플랫폼 기본값입니다. 그렇지 않은 경우 기본 효과 검정은 Wald 검정입니다.

전체 모형 검정

"전체 모형 검정" 테이블에는 절편 모수를 제외한 모든 회귀 모수가 생략된 모형과 전체 모형 적합을 비교하는 검정이 표시됩니다. 이 검정은 연속형 반응에 대한 분산 분석 테이블과 유사합니다. 음의 로그 가능도는 제곱합에 해당하고 카이제곱 검정은 F 검정에 해당합니다.

"전체 모형 검정" 테이블에는 다음과 같은 통계량이 표시됩니다.

모형 모형 라벨을 나열합니다.

차이 전체 모형과 축소 모형 간의 차이입니다. 이 모형은 적합에 대한 전체 회귀변수의 유의성을 측정하는 데 사용됩니다.

전체 절편 및 모든 효과를 포함하는 완전 모형입니다.

축소 절편 모수만 포함하는 모형입니다.

-1* 로그 가능도 각 모형에 대한 음의 로그 가능도입니다. 자세한 내용은 "통계 상세 정보" 부록의 "**가능도, AICc 및 BIC**"(514 페이지)에서 확인하십시오.

DF 전체 모형과 축소 모형 간의 차이에 대한 DF(자유도)입니다.

카이제곱 모든 회귀 모수가 0이라는 가설에 대한 가능도비 카이제곱 검정 통계량입니다. 이 검정 통계량은 적합 모형과 절편만 있는 축소 모형 간의 음의 로그 가능도 차이의 두 배로 계산됩니다.

Prob>ChiSq 지정된 모형이 절편만 포함된 모형보다 더 잘 적합되지 않을 경우 카이제곱 값이 더 클 확률입니다.

$R^2(U)$ 모형 적합에 기인한 전체 불확실성의 비율로, 축소된 음의 로그 가능도 값에 대한 차이의 비율로 정의됩니다. R^2 값은 0(적합이 개선되지 않는 경우)에서 1(완벽한 적합)까지입니다. $R^2(U)$ 값이 1 이면 발생하는 사건의 예측 확률이 1 과 같음을 나타냅니다. 즉, 예측 확률에 불확실성이 없는 것입니다. 로짓 모형의 경우 예측 확률은 거의 확실하지 않으므로 $R^2(U)$ 가 작은 경향이 있습니다.

$R^2(U)$ 는 불확도 계수인 U 또는 McFadden 유사 R^2 이라고도 합니다.

AICc 수정된 Akaike 정보 기준입니다. 자세한 내용은 "통계 상세 정보" 부록의 "**가능도, AICc 및 BIC**"(514 페이지)에서 확인하십시오.

BIC Bayes 정보 기준 자세한 내용은 "통계 상세 정보" 부록의 "**가능도, AICc 및 BIC**"(514 페이지)에서 확인하십시오.

관측값 수 (또는 가중합) 표본의 총 관측값 수입니다. "모형 적합" 창에서 "빈도" 또는 "가중치" 열이 지정되는 경우 이 값은 빈도 또는 가중치 역할에 할당된 열 값의 합입니다.

적합 상세 정보

"적합 상세 정보" 보고서에는 다음과 같은 통계량이 포함됩니다.

엔트로피 R^2 $R^2(U)$ 와 동등합니다. 자세한 내용은 "[전체 모형 검정](#)"(449 페이지) 에서 확인하십시오.

일반화 R^2 일반 회귀 모형에 적용할 수 있는 측도입니다. 이 값은 가능도 함수 L 을 기반으로 하며 최대값이 1 이 되도록 척도화됩니다. 일반화 R^2 측도는 표준 최소 제곱 설정 시 연속형 정규 반응에 대한 기존 R^2 로 단순화됩니다. 일반화 R^2 를 Nagelkerke 또는 Craig and Uhler R^2 라고도 하는데, 이는 Cox-Snell 유사 R^2 를 정규화한 것입니다. 자세한 내용은 Nagelkerke 연구 자료 (1991) 에서 확인하십시오.

평균 -Log p $-\log(p)$ 의 평균입니다. 여기서 p 는 발생한 사건과 연관된 적합 확률입니다.

RASE 제공된 평균 제공 오차입니다. 여기서 차이는 반응과 p (실제로 발생한 사건의 적합 확률) 사이의 차이입니다.

평균 절대 편차 반응과 p (실제로 발생한 사건의 적합 확률) 사이의 차이에 대한 절대값을 평균한 것입니다.

오분류 비율 적합 확률이 가장 높은 반응 범주가 관측된 범주가 아닌 비율입니다.

N 관측값 수입니다.

엔트로피 R^2 및 일반화 R^2 의 경우 값이 1 에 가까울수록 더 나은 적합을 나타냅니다. 평균 -Log p, RASE, 평균 절대 편차 및 오분류 비율의 경우 값이 작을수록 더 나은 적합을 나타냅니다.

효과가 전체적으로 유의한지 검정하기 위해 (전체 모형 검정) 카이제곱 통계량은 적합 모형과 절편만 있는 축소 모형 간의 음의 로그 가능도 차이의 두 배로 계산됩니다.

검증 열을 지정한 경우 각 훈련, 검증 및 테스트 데이터 집합에 대한 열이 "적합 상세 정보" 보고서에 포함됩니다.

적합 결여 검정

적합 결여 검정은 현재 모형에 충분한 정보가 있는지 아니면 복잡 항이 더 필요한지를 확인합니다. 이 검정을 적합도 검정이라고도 합니다. 적합 결여 검정은 데이터에서 모형 효과 값의 모든 조합에 대한 범주를 구성하여 순수 오차 음의 로그 가능도를 계산합니다. "적합 결여" 테이블의 포화 행은 이 로그 가능도를 포함합니다. "적합 결여" 보고서에는 포화 로그 가능도가 적합된 모형보다 유의하게 나은지에 대한 검정도 포함되어 있습니다.

포화 자유도는 $m-1$ 이며, 여기서 m 은 고유 모집단의 수입니다. 적합된 자유도는 절편을 제외한 모수의 수입니다.

"적합 결여" 테이블에는 적합 결여로 인한 오차, 포화 모형의 오차 (순수 오차) 및 적합 모형의 총 오차에 대한 음의 로그 가능도가 포함됩니다. 카이제곱 통계량은 적합 결여를 검정합니다.

로지스틱 적합 플랫폼 옵션

- "명목형 및 순서형 적합 옵션"
- "명목형 적합 옵션"
- "순서형 적합 옵션"

명목형 및 순서형 적합 옵션

"명목형 로지스틱 적합"과 "순서형 로지스틱 적합"의 빨간색 삼각형 메뉴에서 모두 다음 옵션을 사용할 수 있습니다.

로지스틱 그림 (모형이 단일 연속형 효과로 구성된 경우에만 사용 가능) "로지스틱 그림" 보고서를 표시하거나 숨깁니다. 자세한 내용은 "[로지스틱 적합 보고서](#)"(448 페이지)에서 확인하십시오.

가능도비 검정 "효과 가능도비 검정" 보고서를 표시하거나 숨깁니다. 가능도비 카이제곱 검정은 전체 모형과 검정할 가설의 제약을 받는 모형 간의 로그 가능도 차이의 두 배로 계산됩니다. 제약이 있는 모형은 효과를 포함하지 않는 모형입니다. 이러한 검정은 각 검정마다 별도의 반복 집합이 필요하므로 시간이 걸릴 수 있습니다. 따라서 큰 문제의 경우 시간이 오래 걸릴 수 있습니다.

참고: 가능도비 검정은 완료하는 데 20 초 미만이 걸릴 것으로 예상되는 경우 플랫폼 기본값입니다. 이 기본 옵션을 사용하는 것이 좋습니다.

Wald 검정 "효과 Wald 검정" 보고서를 표시하거나 숨깁니다. Wald 카이제곱은 가능도비 검정에 대한 2 차 근사이며 계산의 부산물입니다. Wald 검정은 신뢰도가 낮다고 간주되지만 효과 선별을 위해 적절한 유의성 지표를 제공합니다. 각 모수 추정값 및 효과가 Wald 검정과 함께 표시됩니다. 이 검정은 가능도비 검정을 완료하는 데 20 초를 초과할 것으로 예상되는 경우 기본 검정입니다.

신뢰 구간 모형 모수에 대한 프로파일 - 가능도 신뢰 구간을 표시하거나 숨깁니다. "모형 적합" 창의 "모형 규격" 빨간색 삼각형 메뉴에서 "유의 수준 설정"을 선택하여 신뢰 수준을 변경할 수 있습니다. 각 신뢰 한계에는 모형 적합에 일련의 반복이 필요하며 계산하는 데 시간이 오래 걸릴 수 있습니다. 그뿐만 아니라 한계를 찾는 데 실패할 수도 있습니다.

ROC 곡선 모형에 대한 ROC 곡선을 표시하거나 숨깁니다. ROC(Receiver Operating Characteristic) 곡선은 반응 수준을 정렬하기 위해 모형의 적합 확률의 정렬 효율성을 측정합니다. ROC 곡선은 진단 검정에서 기준점을 설정하는 데도 도움이 될 수 있습니다. 주대각에서 곡선과의 거리가 멀어질수록 적합도가 높습니다. ROC 곡선에 대한 소개는 기본 분석의에서 확인하십시오.

로지스틱 적합에 세 개 이상의 반응 수준이 있는 경우 일반화 ROC 곡선 (파티션 플랫폼의 ROC 곡선과 동일)이 생성됩니다. 이 그림에는 각 반응 수준에 대한 곡선, 즉 다른 모든 수

준에 대한 해당 수준의 ROC 곡선이 있습니다. 자세한 내용은 **Predictive and Specialized Modeling** 의 에서 확인하십시오.

검증 열을 지정한 경우 훈련 데이터 집합, 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 ROC 곡선이 표시됩니다.

Lift 곡선 모형에 대한 Lift 곡선을 표시하거나 숨깁니다. Lift 곡선은 ROC 곡선과 동일한 정보를 보여 주되, 시작점에서 정렬의 풍부도를 극화하는 방법을 사용합니다. 세로 축은 모집단 중 해당 부분이 해당 반응 수준 전체의 비율에 비해 선택된 반응 수준에 얼마나 풍부하게 포함되었는지를 보여 줍니다. 검증 열을 지정한 경우 훈련 데이터 집합, 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 Lift 곡선이 표시됩니다. Lift 곡선에 대한 자세한 내용은 **Predictive and Specialized Modeling** 의 에서 확인하십시오.

혼동 행렬 " 혼동 행렬 " 보고서 및 " 혼동 비율 " 보고서가 포함된 혼동 통계량 보고서를 표시하거나 숨깁니다. 두 보고서 모두 실제 반응 수준과 예측 반응 수준의 이원 분류입니다. 예측 반응 수준은 시작 창에 지정된 목표 수준입니다. 혼동 비율 보고서는 혼동 행렬 보고서와 동일하며 개수를 행 합계로 나눕니다.

적합한 모형은 예측 반응 수준이 실제 반응 수준과 동일해야 합니다. 혼동 행렬 보고서를 통해 예측 반응이 실제 반응에 잘 맞는지 평가할 수 있습니다. 검증 열을 지정한 경우 훈련 데이터 집합, 검증 데이터 집합 및 테스트 데이터 집합에 대해 각각 혼동 행렬이 표시됩니다.

반응이 명목형이고 " 수익 행렬 " 열 특성이 있는 경우 이 옵션을 선택하면 결정 행렬 보고서도 나타납니다. 결정 행렬 보고서에 대한 자세한 내용은 **Predictive and Specialized Modeling** 의 에서 확인하십시오.

프로파일러 모형의 요인 값이 변경될 때 지정된 반응 확률에 대한 적합된 값을 보여 주는 예측 프로파일러를 표시하거나 숨깁니다. 이 기능은 명목형 반응과 순서형 반응에 모두 사용할 수 있습니다. 빨간색 삼각형 메뉴의 옵션에 대한 자세한 내용은 **Profilers** 의 에서 확인하십시오.

등고선 프로파일러 (모형에 두 개 이상의 연속형 요인이 포함된 경우에만 사용 가능) 등고선 프로파일러를 표시하거나 숨깁니다. 빨간색 삼각형 메뉴의 옵션에 대한 자세한 내용은 **Profilers** 의 에서 확인하십시오.

모형 대화상자 현재 분석을 위해 완료된 시작 창을 표시합니다.

효과 요약 모형의 효과를 대화식으로 업데이트할 수 있는 " 효과 요약 " 보고서를 표시하거나 숨깁니다. 자세한 내용은 " 로지스틱 적합 보고서 "(448 페이지) 에서 확인하십시오.

다음 옵션에 대한 자세한 내용은 **JMP 사용** 의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

명목형 적합 옵션

"명목형 로지스틱 적합"의 빨간색 삼각형 메뉴에서 다음 옵션을 사용할 수 있습니다.

그림 옵션 (모형이 단일 연속형 효과로 구성된 경우에만 사용 가능) "그림 옵션" 메뉴에는 다음 옵션이 포함되어 있습니다.

점 표시 로지스틱 그림의 점을 표시하거나 숨깁니다.

비율 곡선 표시 효과 변수의 다양한 값에서 추정된 비율을 적합된 로지스틱 곡선과 비교할 수 있습니다. 이 옵션은 효과의 각 값에 대해 여러 점이 있는 경우에만 유용합니다. 이 경우 각 값의 적절한 비율 추정값을 구하고 이 비율을 적합된 로지스틱 곡선과 비교할 수 있습니다. 퇴화 점 (일반적으로 0 또는 1 일 때) 이 너무 많아지지 않도록 JMP에서는 x 값에 점이 세 개 이상 있는 경우에 비율 값만 표시합니다.

선 색상 그림의 곡선 색상을 지정합니다.

승산비 단위 승산비와 범위 승산비가 포함된 승산비 보고서를 표시하거나 숨깁니다.

그림 11.6 승산비

승산비

ready의 Ready 대 Not Ready 승산

단위 승산 비율

회귀변수의 단위 변화당

항	승산 비율	하위 95%	상위 95%	역수
heat	0.921244	0.87937	0.965111	1.0854892
soak	0.94481	0.493646	1.808313	1.0584137

범위 승산 비율

전제 범위에 대한 회귀변수의 변화당

항	승산 비율	하위 95%	상위 95%	역수
heat	0.027069	0.003496	0.209605	36.942229
soak	0.8434	0.120295	5.913179	1.185677

승산비에 대한 검정 및 신뢰 구간이 Wald 방법에 기반합니다.

역추정 예측 (2 수준 명목형 반응에만 사용 가능) 지정된 확률이 되는 x 값을 찾습니다. 역추정 예측에 대한 자세한 내용은 부록 "통계 상세 정보"(487 페이지)에서 확인하십시오.

확률 계산식 저장 현재 데이터 테이블에 반응 수준의 선형 결합 계산식, 각 반응 수준에 대한 예측 계산식, 가장 확률이 높은 반응을 제공하는 예측 계산식을 포함하는 열을 생성합니다.

r 개 수준이 있는 명목형 반응 모형의 경우 JMP에서 다음 열이 생성됩니다.

- 반응 수준 $j = 1, 2, \dots, r-1$ 에 대한 회귀변수의 선형 결합을 포함하는 선형 j 열
- 마지막 수준 r 에 대한 적합 계산식을 포함하는 확률 r 열
- $j < r$ 일 때 j 수준에 대한 적합 계산식을 포함하는 확률 j 열

- 계산된 확률을 기반으로 각 행의 가장 확률이 높은 수준을 선택하는 가장 확률이 높은 '반응 이름' 열

JMP PRO 확률 계산식 게시 확률 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다. "계산식 저장소" 보고서가 열려 있지 않으면 "계산식 저장소" 보고서가 생성됩니다. 자세한 내용은 **Predictive and Specialized Modeling** 의 에서 확인하십시오.

표시자 파라미터화 추정값 (모형 효과 중 명목형 열이 있는 경우에만 사용 가능) 명목형 열이 표시자 (SAS GLM) 파라미터화를 사용하여 코드화되고 연속형으로 처리되는 모형에 대한 모수 추정값을 제공하는 "표시 함수 파라미터화" 보고서를 표시하거나 숨깁니다. 순서형 열은 일반적인 JMP 코딩 체계를 사용하여 코드화됩니다. SAS GLM 및 JMP 코딩 체계에 대한 설명은 "통계 상세 정보" 부록의 "**요인 모형**"(492 페이지) 에서 확인하십시오.

주의 : "표시 함수 파라미터화" 보고서에 제공된 표준 오차 및 카이제곱 값은 "모수 추정값" 보고서의 값과 다릅니다. 이는 추정하는 모형 모수가 서로 다르기 때문입니다.

순서형 적합 옵션

"순서형 로지스틱 적합"의 빨간색 삼각형 메뉴에서 다음 옵션을 사용할 수 있습니다.

저장 다음 저장 옵션을 포함합니다.

확률 계산식 저장 현재 데이터 테이블에 반응 수준의 선형 결합 계산식, 각 반응 수준에 대한 예측 계산식, 가장 확률이 높은 반응을 제공하는 예측 계산식을 포함하는 열을 생성합니다.

r 개 수준이 있는 순서형 반응 모형의 경우 JMP 에서 다음 열이 생성됩니다.

- 절편 항이 없는 회귀변수의 선형 결합에 대한 계산식을 포함하는 선형 열
- $j = 1, 2, \dots, r-1$ 수준에 대해 각각 반응이 j 수준보다 작거나 같을 누적 확률에 대한 계산식을 포함하는 누적 j 열. 모든 행에 대해 1 인 누적 $j = 1, 2, \dots, r-1$ 은 없습니다.
- $1 < j < r$ 일 때 각각 반응이 j 수준일 확률에 대한 계산식을 포함하는 확률 $j = 1, 2, \dots, r-1$ 열. 확률 j 는 누적 j 와 누적 $j-1$ 의 차이입니다. 확률 1 은 누적 1 이고 확률 r 은 $1 -$ 누적 $[r-1]$ 입니다.
- 계산된 확률을 기반으로 각 행의 가장 확률이 높은 수준을 선택하는 가장 확률이 높은 '반응 이름' 열

JMP PRO 확률 계산식 게시 확률 계산식을 생성하여 계산식 저장소 플랫폼에 계산식 열 스크립트로 저장합니다. "계산식 저장소" 보고서가 열려 있지 않으면 "계산식 저장소" 보고서가 생성됩니다. 자세한 내용은 **Predictive and Specialized Modeling** 의 에서 확인하십시오.

분위수 저장 (반응이 수치형이고 순서형 모델링 유형인 경우에만 사용 가능) 현재 데이터 테이블에 세 가지 확률에 대한 분위수를 적합시키는 OrdQ.05, OrdQ.50 및 OrdQ.95 열을 생성합니다.

기대값 저장 (반응이 수치형이고 순서형 모델링 유형인 경우에만 사용 가능) 현재 데이터 테이블에 기대 순서 열을 생성합니다. 이 열에는 각 행에 대한 적합 반응 확률과 반응 값의 선형 결합이 포함되며 기대값을 제공합니다.

로지스틱 회귀의 추가 예

이 섹션에는 모형 적합 플랫폼에서 로지스틱 회귀를 사용하는 추가 예가 포함되어 있습니다.

- "역추정 예측의 예"
- "명목형 로지스틱 모형에 대한 효과 요약의 사용 예"
- "2 차 순서형 로지스틱 모형의 예"
- "예: 여러 열의 개수 쌓기"

역추정 예측의 예

역추정 예측을 사용하면 지정된 반응 확률에서 하나의 독립 변수에 대한 값을 예측할 수 있습니다. 독립 변수가 두 개 이상인 경우 예측 대상 변수를 제외한 모든 변수의 값을 지정해야 합니다.

금속 주괴에 대한 실험이 수행되었습니다. 다양한 가열 시간 및 유지 시간으로 주괴를 준비한 후 시작 준비 상태를 테스트했습니다. 자세한 내용은 Cox and Snell 연구 자료(1989)에서 확인하십시오. 데이터는 JMP의 `Ingots.jmp` 샘플 데이터 테이블에 포함되어 있습니다.

시작 준비 확률 0.8 과 0.9 에 대해 유지 시간 2 에서 가열 시간을 예측하는 데 관심이 있습니다.

1. 분석을 위해 "명목형 로지스틱 회귀의 예"(442 페이지)의 단계를 수행합니다.
2. "ready 에 대한 명목형 로지스틱 적합" 옆의 빨간색 삼각형을 클릭하고 **역추정 예측**을 선택합니다.
3. `heat`의 값을 삭제합니다.
`heat`의 예측값을 찾아야 하므로 `heat` 값을 비워 둡니다.
4. `soak`에 2 를 입력합니다.
`soak`가 2 일 때 가열 시간을 예측하려고 합니다.
5. "확률" 아래의 처음 두 행에 0.9 와 0.8 을 입력합니다.

그림 11.7 역추정 예측 규격 창

예측할 값을 비어 있거나 결측된 상태로 두십시오.
역추정 예측할 하나 이상의 확률 값을 지정하십시오.

heat

soak

신뢰 수준

양측

확률(ready=Ready)

0.9
0.8
.
.
.
.
.

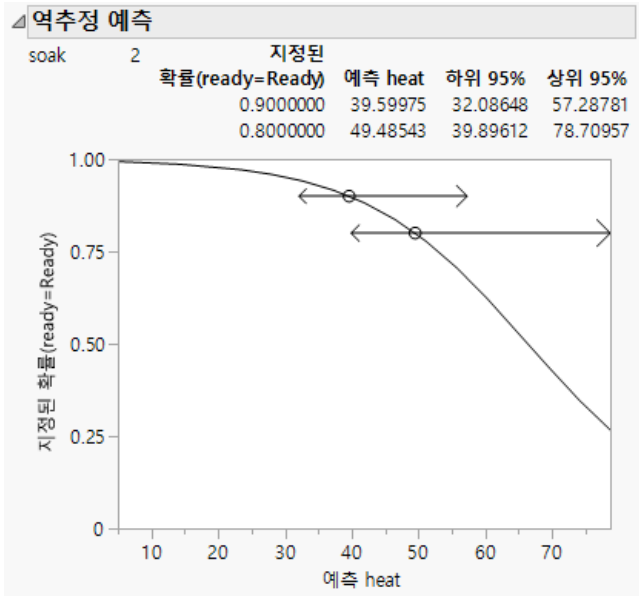
확인

취소

도움말

6. **확인**을 클릭합니다.

그림 11.8 역추정 예측 보고서



soak가 2일 때 주괴 시작 준비가 완료될 확률이 90%인 heat의 예측값은 39.60이며, 32.09 ~ 57.29의 95% 신뢰 구간을 같이 가집니다. soak가 2일 때 주괴 시작 준비가 완료될 확률이 80%인 heat의 예측값은 49.49이며, 39.90 ~ 78.71의 95% 신뢰 구간을 같이 가집니다.

명목형 로지스틱 모형에 대한 효과 요약의 사용 예

세제 브랜드 선호도를 평가하기 위한 시장 조사 연구가 수행되었습니다 . 자세한 내용은 Ries and Smith 연구 자료 (1963) 에서 확인하십시오 . 연구 결과는 Detergent.jmp 샘플 데이터 테이블에 있습니다 . 모형은 다음과 같이 정의됩니다 .

- brand 반응 변수 - m, x 값 포함
- softness(수질 연성) 효과 - soft, medium, hard 값 포함
- previous use 효과 - yes, no 값 포함
- temperature 효과 - high, low 값 포함
- count 개수 변수 - 효과 범주의 각 조합에 대한 빈도 수 제공

3 요인 완전 요인 모형을 지정하여 연구를 시작합니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Detergent.jmp 를 엽니다 .

2. **분석 > 모형 적합**을 선택합니다 .

3. " 열 선택 " 목록에서 brand 를 선택하고 **Y** 를 클릭합니다 .

명목형 반응 변수를 지정했으므로 " 분석법 " 이 " 명목형 로지스틱 " 으로 변경됩니다 .

brand 는 두 개의 수준만 있는 명목형 열이므로 " 목표 수준 " 옵션이 나타납니다 . 이 옵션을 사용하면 확률을 모델링할 반응 수준을 지정할 수 있습니다 .

4. " 목표 수준 " 목록에서 "m" 을 선택합니다 .

5. Count 를 선택하고 **빈도**를 클릭합니다 .

6. softness ~ temperature 를 선택하고 **매크로 > 완전 요인**을 클릭합니다 .

7. **실행**을 클릭합니다 .

그림 11.9 3 요인 요인 모형에 대한 명목형 로지스틱 적합

brand에 대한 명목형 로지스틱 적합

▷ 효과 요약
그래디언트에 수렴됨, 3회 반복
빈도: count

▷ 반복

△ 전체 모형 검정

모형	-1*로그 가능도	DF	카이제곱	Prob>ChiSq
차이	16.41281	11	32.82562	0.0006*
전체	682.24780			
축소	698.66061			

R²(U) 0.0235
AICc 1388.81
BIC 1447.48
관측값 수(또는 가중합) 1008

▷ 적합 상세 정보

▷ 모수 추정값

△ 효과 가능도비 검정

소스	N모수	DF	L-R 카이제곱	Prob>ChiSq
softness	2	2	0.09804239	0.9522
previous use	1	1	22.1316677	<.0001*
softness*previous use	2	2	3.78609668	0.1506
temperature	1	1	3.63914017	0.0564
softness*temperature	2	2	0.19617686	0.9066
previous use*temperature	1	1	2.26089203	0.1327
softness*previous use*temperature	2	2	0.73731691	0.6917

"전체 모형 검정" 보고서에서는 3 요인 완전 요인 모형이 전체적으로 유의하다는 것을 보여줍니다 (Prob>ChiSq = 0.0006).

"효과 가능도비 검정" 보고서에서는 **softness** 를 포함하는 효과가 모형 적합에 유의하게 기여하지 않음을 보여 줍니다. 따라서 **softness** 를 모형에서 제거할 수도 있습니다. "효과 요약" 보고서에서 이를 수행할 수 있습니다.

- "효과 요약" 보고서의 "소스" 아래에서 **softness*previous use ~ softness** 를 선택하고 **제거** 를 클릭합니다.

보고서가 업데이트되어 2 요인 요인 모형이 표시됩니다 (그림 11.10). "전체 모형 검정" 보고서를 보면 2 요인 모형도 전체적으로 유의함을 알 수 있습니다.

그림 11.10 2 요인 요인 모형에 대한 명목형 로지스틱 적합

brand에 대한 명목형 로지스틱 적합

그래디언트에 수렴됨, 3회 반복
빈도: count

반복

전체 모형 검정

모형	-1*로그 가능도	DF	카이제곱	Prob>ChiSq
차이	13.58479	3	27.16958	<.0001*
전체	685.07582			
축소	698.66061			

R²(U) 0.0194
AICc 1378.19
BIC 1397.81
관측값 수(또는 가중합) 1008

적합 상세 정보

모수 추정값

항	추정값	표준 오차	카이제곱	Prob>ChiSq
절편	0.03043265	0.0666827	0.21	0.6481
previous use[no]	-0.3148498	0.0666827	22.29	<.0001*
temperature[low]	-0.1331196	0.0666827	3.99	0.0459*
previous use[no]*temperature[low]	0.11106673	0.0666827	2.77	0.0958

대상 로그 승산: m/x

추정값 공분산

효과 가능도비 검정

소스	N모수	DF	L-R	카이제곱	Prob>ChiSq
previous use	1	1	22.6511053	<.0001*	
temperature	1	1	3.99625478	0.0456*	
previous use*temperature	1	1	2.78794934	0.0950	

세제 브랜드의 이전 사용 여부와 수온이 세제 선호도에 영향을 미친다는 결론을 내릴 수 있습니다. 또한 온도와 이전 사용 여부의 교호작용은 통계적으로 유의하지 않으므로 온도가 이전 사용에 따라 달라진다는 증거가 없습니다.

2 차 순서형 로지스틱 모형의 예

순서형 로지스틱 분석법은 순서형 반응의 높은 수준 또는 낮은 수준의 확률을 최적화하기 위해 2차 표면을 적합시킬 수 있습니다.

이 예에서 전자레인지 팝콘 제조업체는 소비자가 좋아하는 팝콘의 소금 양이 어느 정도인지 확인하려고 합니다. 이를 위해 제조업체는 팝콘 패키지에 추가되는 소금의 양에 따라 양호한 반응의 최대 확률을 찾습니다. 이 실험에서 소금의 양은 0, 1, 2 및 3 티스푼으로 조절됩니다. 응답자는 1(낮음)~5(높음)의 척도로 맛을 평가합니다. 더 양호한 반응의 확률을 최대화하는 양이 최적의 소금 양입니다. 표 11.2에서는 각 소금 수준에 대한 10개의 관측값을 보여 줍니다.

표 11.2 팝콘의 소금

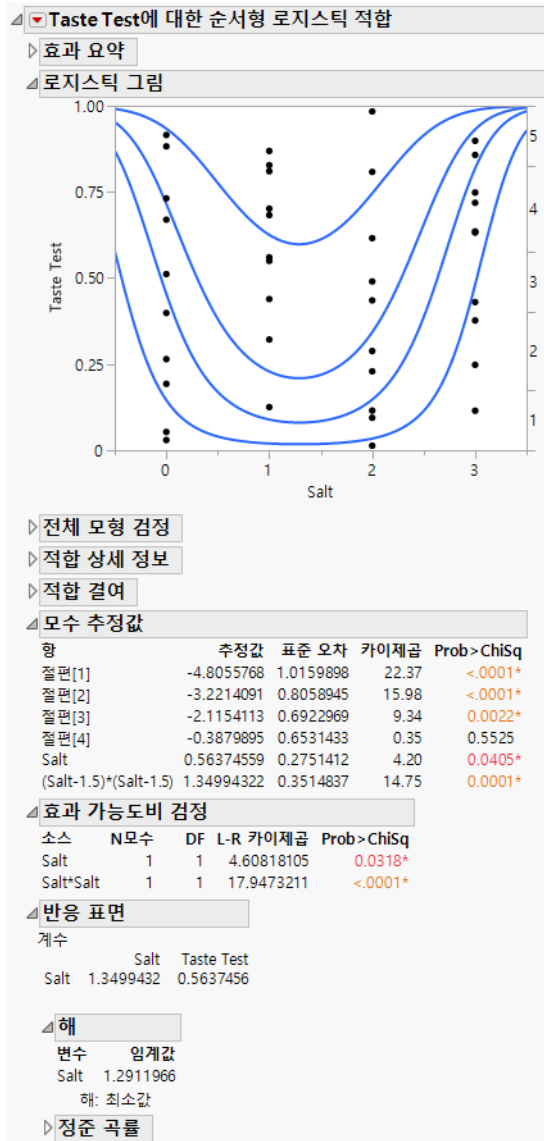
소금 양	소금 평가 반응
소금 없음	1, 3, 2, 4, 2, 2, 1, 4, 3, 4

표 11.2 팝콘의 소금 (계속)

소금 양	소금 평가 반응
1 티스푼	4, 5, 3, 4, 5, 4, 5, 5, 4, 5
2 티스푼	4, 3, 5, 1, 4, 2, 5, 4, 3, 2
3 티스푼	3, 1, 2, 3, 1, 2, 1, 2, 1, 2

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Salt in Popcorn.jmp 를 엽니다 .
2. **분석 > 모형 적합**을 선택합니다 .
3. " 열 선택 " 목록에서 **Taste Test** 를 선택하고 **Y** 를 클릭합니다 .
4. " 열 선택 " 목록에서 **Salt** 를 선택하고 **매크로 > 반응 표면**을 클릭합니다 .
5. **실행**을 클릭합니다 .
6. " 반응 표면 " 옆의 표시 아이콘을 클릭하여 보고서를 엽니다 .

그림 11.11 Salt in Popcorn.jmp 에 대한 순서형 로지스틱 적합



이 보고서에서는 2차 모형이 반응 확률을 어떻게 적합시키는지 보여 줍니다. 곡선은 변하는 로지스틱 곡선이 아니라 동일한 점에서 최소값을 이루는 각 곡선이 쌓여 U자형이 됩니다. 임계값은 $\text{Mean}(X) - 0.5 * (b1/b2)$ 입니다. 여기서 $b1$ 은 선형 계수이고 $b2$ 는 2차 계수입니다. 이 계산식은 중심화된 X 를 대상으로 합니다. "모수 추정값" 테이블에서 최적의 소금 양은 $1.5 - 0.5 * (0.5637/1.3499) = 1.29$ 티스푼임을 계산할 수 있습니다.

특정 소금 양에서 각 곡선 사이의 세로 거리는 해당 소금 양에 대한 다섯 가지 반응 수준의 각 확률을 측정합니다. 가장 높은 반응 수준에 대한 확률은 상단 곡선에서 그림 직사각형 상단까지의

거리입니다. 이 거리는 소금 양이 약 1.3 티스푼일 때 최대값에 도달합니다. 모든 곡선은 동일한 임계점을 공유합니다.

Salt 및 Salt*Salt에 대한 모수 추정값은 임계값을 찾는 데 사용되는 계수입니다. 임계값은 로지스틱 그림에 최소값으로 표시되지만 가장 높은 반응의 확률을 최대화한다는 의미에서 최대값이 됩니다. 보고서의 "해" 부분은 그림 11.11의 "반응 표면" 아래에 표시되어 있으며, 여기서 임계값 아래에 1.29가 표시됩니다.

예 : 여러 열의 개수 쌓기

빈도(개수) 데이터가 데이터 테이블의 여러 열에 나열된 경우 데이터를 로지스틱 회귀에 필요한 형식으로 변환해야 합니다. 예를 들어 Ingots2.jmp 데이터 테이블(그림 11.12)에는 Nready 및 Nnotready 열이 있습니다. 이러한 열은 Heat 및 Soak 값의 각 조합에 대해 시작 준비가 된 주괴와 준비가 되지 않은 주괴의 수를 제공합니다.

그림 11.12 Ingots2.jmp 샘플 데이터 테이블

	Heat	Soak	Nready	Nnotready	Ntotal	P	Loss
1	7	1	10	0	10	0.5000	6.9315
2	7	1.7	17	0	17	0.5000	11.7835
3	7	2.2	7	0	7	0.5000	4.8520
4	7	2.8	12	0	12	0.5000	8.3178
5	7	4	9	0	9	0.5000	6.2383
6	14	1	31	0	31	0.5000	21.4876
7	14	1.7	43	0	43	0.5000	29.8053
8	14	2.2	31	2	33	0.5000	22.8739
9	14	2.8	31	0	31	0.5000	21.4876
10	14	4	19	0	19	0.5000	13.1698
11	27	1	55	1	56	0.5000	38.8162
12	27	1.7	40	4	44	0.5000	30.4985
13	27	2.2	21	0	21	0.5000	14.5561
14	27	2.8	21	1	22	0.5000	15.2492
15	27	4	15	1	16	0.5000	11.0904

로지스틱 회귀 모형을 적합시키기 전에 다음 단계에 따라 Nready 및 Nnotready 열을 단일 열에 쌓으십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Ingots2.jmp 를 엽니다.
2. **테이블 > 쌓기**를 선택합니다.
3. "열 선택" 목록에서 Nready 와 NNotReady 를 선택하고 **열 쌓기**를 클릭합니다.
4. **확인**을 클릭합니다.

이렇게 하면 그림 11.13의 새 테이블이 생성됩니다. 라벨은 반응(Y) 열이고 데이터는 빈도 열입니다.

이 쌓인 데이터 테이블은 "명목형 로지스틱 회귀의 예"(442 페이지)에 사용된 Ingots.jmp 샘플 데이터 테이블과 동등합니다.

그림 11.13 쌓인 데이터 테이블

	Heat	Soak	Ntotal	P	Loss	Label	Data
1	7	1	10	0.5000	6.9315	Nready	10
2	7	1	10	0.5000	6.9315	Nnotready	0
3	7	1.7	17	0.5000	11.7835	Nready	17
4	7	1.7	17	0.5000	11.7835	Nnotready	0
5	7	2.2	7	0.5000	4.8520	Nready	7
6	7	2.2	7	0.5000	4.8520	Nnotready	0
7	7	2.8	12	0.5000	8.3178	Nready	12
8	7	2.8	12	0.5000	8.3178	Nnotready	0
9	7	4	9	0.5000	6.2383	Nready	9
10	7	4	9	0.5000	6.2383	Nnotready	0
11	14	1	31	0.5000	21.4876	Nready	31
12	14	1	31	0.5000	21.4876	Nnotready	0
13	14	1.7	43	0.5000	29.8053	Nready	43
14	14	1.7	43	0.5000	29.8053	Nnotready	0
15	14	2.2	33	0.5000	22.8739	Nready	31
16	14	2.2	33	0.5000	22.8739	Nnotready	2
17	14	2.8	31	0.5000	21.4876	Nready	31
18	14	2.8	31	0.5000	21.4876	Nnotready	0

명목형 및 순서형 로지스틱 분석법에 대한 통계 상세 정보

- "로지스틱 회귀 모형"
- "승산비"
- "통계적 검정의 관계"

로지스틱 회귀 모형

로지스틱 회귀는 명목형 Y 반응을 X 항의 선형 모형에 적합시킵니다. 더 정확히 말하면 로지스틱 함수를 사용하여 반응 수준에 대한 확률을 적합시킵니다. 두 개의 반응 수준에 대한 함수는 다음과 같습니다.

$$P(Y = r_1) = (1 + e^{-Xb})^{-1} \text{ 여기서, } r_1 \text{ 은 첫 번째 반응 수준입니다.}$$

또는 다음과 같이 지정될 수도 있습니다.

$$\log\left(\frac{P(Y = r_1)}{P(Y = r_2)}\right) = Xb \text{ 여기서, } r_1 \text{ 과 } r_2 \text{ 는 각각 두 반응 수준입니다.}$$

참고 : Y 가 이항 변수이고 명목형 모델링 유형인 경우 모형 적합 창에서 목표 수준을 설정하여 확률을 모델링할 수준을 지정할 수 있습니다. 이 섹션에서는 목표 수준이 r_1 로 지정됩니다.

$r(r > 2)$ 개의 명목형 반응 수준에 대해 모형은 다음과 같은 형식의 $r-1$ 개 선형 모형 모수로 정의됩니다.

$$\log\left(\frac{P(Y=j)}{P(Y=r)}\right) = Xb_j$$

최대 가능도 적합은 모형에서 발생한 반응에 기인한 결합 확률을 최대화하기 위해 β 가 선택됨을 의미합니다. 이 적합은 음의 로그 가능도 ($-\text{LogLikelihood}$) 를 최소화하는 것과 동등합니다.

$$\text{Loss} = -\log\text{Likelihood} = \sum_{i=1}^n -\log(\text{Prob}(i\text{번째 행에 } y_j\text{번째 반응이 있음}))$$

승산비

두 개의 반응 수준에 대한 로지스틱 회귀 모형은 다음과 같이 지정됩니다.

$$\log\left(\frac{\text{Prob}(Y=r_1)}{\text{Prob}(Y=r_2)}\right) = Xb \quad \text{여기서, } r_1 \text{ 과 } r_2 \text{ 는 두 반응 수준입니다.}$$

따라서 승산은 다음과 같이 정의됩니다.

$$\frac{\text{Prob}(Y=r_1)}{\text{Prob}(Y=r_2)} = \exp(X\beta) = \exp(\beta_0) \cdot \exp(\beta_1 X_1) \cdots \exp(\beta_i X_i)$$

$\exp(\beta_i(X_i+1)) = \exp(\beta_i X_i) \exp(\beta_i)$ 입니다. 즉, X_i 가 단위 양만큼 변경되면 승산에 $\exp(\beta_i)$ 를 곱하고 이를 단위 승산비라고 합니다. X_i 가 전체 범위에 대해 변경되면 승산에 $\exp((X_{\text{high}} - X_{\text{low}})\beta_i)$ 를 곱하고 이를 범위 승산비라고 합니다. 이항 반응의 경우 반전된 반응 수준의 로그 승산비는 모수의 부호만 변경됩니다. 따라서 보고된 값의 역수를 첫 번째 반응 수준 대신 마지막 수준에 초점을 맞출 수 있습니다.

2수준 명목형 효과는 첫 번째 수준과 두 번째 수준에 대해 각각 1과 -1로 코드화되므로 범위 승산비 또는 역수가 중요합니다.

승산비 옵션을 선택할 때 모수 추정값의 신뢰 구간이 계산되지 않은 경우 승산비 옵션을 선택하면 해당 승산비에 대해 Wald 기반 신뢰 구간이 생성됩니다.

다음 조건이 모두 충족될 경우 승산비 옵션을 선택하면 해당 승산비에 대해 프로파일 가능도 기반 신뢰 구간이 생성됩니다.

- 가능도비 검정 옵션이 선택되었습니다.
- 신뢰 구간 옵션이 선택되었습니다.
- 각 반응의 모수 수가 8 개 미만입니다.

- 행 수가 1000 개 미만입니다 .

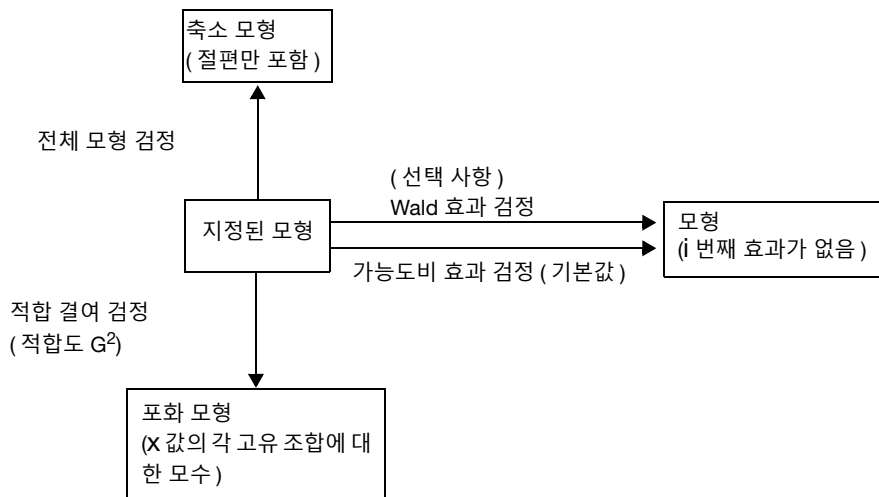
승산비에 대한 신뢰 구간을 계산하는 데 사용된 방법이 승산비 보고서 하단에 표시됩니다.

통계적 검정의 관계

로지스틱 회귀 보고서의 모든 통계적 검정은 그림 11.14 에 설명된 대로 지정된 모형의 적합을 부분집합 또는 상위집합 모형과 비교합니다 . 검정이 유의성을 나타내면 고차 모형이 타당성을 갖습니다 .

- 전체 모형 검정 : 지정된 모형이 절편을 제외한 다른 효과가 없는 축소 모형보다 유의하게 나은 경우
- 적합 결여 검정 : 포화 모형이 지정된 모형보다 유의하게 나은 경우
- 효과 검정 : 지정된 모형이 효과가 제공되지 않은 모형보다 유의하게 나은 경우

그림 11.14 통계적 검정의 관계
모형 적합 비교 검정



12 장

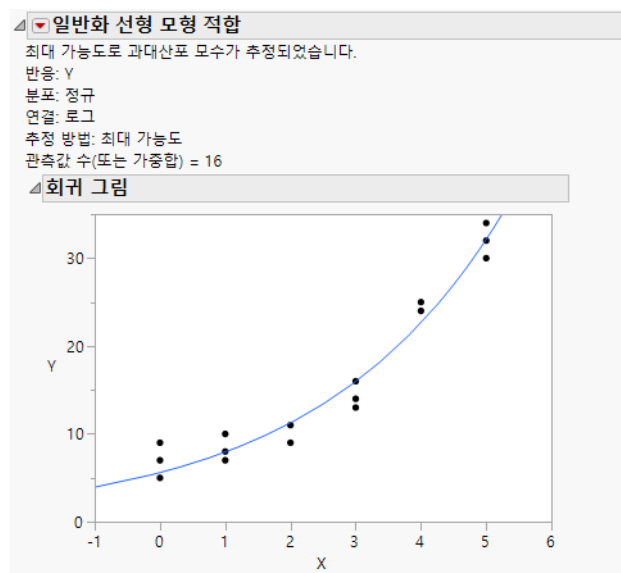
일반화 선형 모형

비정규 반응 분포에 대한 모형 적합

일반화 선형 모형은 기존 선형 모형의 일반적인 요구 사항에 맞지 않는 반응을 적합시키는 통합된 방법을 제공합니다. 예를 들어 빈도 수는 일반화 선형 모형을 사용하여 Poisson 분포 및 적합을 수행하는 경우가 많습니다.

모형 적합 플랫폼의 일반화 선형 모형 분석법을 사용하면 이항, 정규, Poisson 또는 지수 분포를 사용하여 반응에 대한 일반화 선형 모형을 적합시킬 수 있습니다. 이 플랫폼은 기존 선형 모형에 제공되는 것과 유사한 보고서를 제공합니다. 또한 Firth 수정을 사용하여 로지스틱 회귀 모형의 분리도 수용합니다.

그림 12.1 일반화 선형 모형 적합의 예



목차

일반화 선형 모형 분석법 개요	469
일반화 선형 모형의 예	469
일반화 선형 모형 분석법 시작	472
일반화 선형 모형 적합 보고서	473
전체 모형 검정	474
일반화 선형 모형 적합 보고서 옵션	475
일반화 선형 모형 분석법의 추가 예	477
대비를 사용하여 변수 수준의 차이 비교	477
오프셋을 사용하는 Poisson 회귀	478
로그 연결을 사용하는 정규 회귀	480
일반화 선형 모형 분석법에 대한 통계 상세 정보	482
모형 선택 및 이탈도	484

일반화 선형 모형 분석법 개요

기존 선형 모형은 통계 데이터 분석에서 광범위하게 사용됩니다. 그러나 기존 선형 모형의 가정을 위반하는 상황이 있습니다. 이러한 경우 기존 선형 모형은 적절하지 않습니다. 기존 선형 모형은 반응이 연속형이고 모든 관측값에 대해 일정한 분산으로 정규 분포를 따른다고 가정합니다. 이러한 가정이 적절하지 않을 수 있습니다. 예를 들어 개수를 모델링하려는 경우 또는 반응이 증가할수록 관측된 반응의 분산이 증가하는 경우 이 가정은 적절하지 않습니다. 반응 평균이 0과 1 사이의 비율과 같은 특정 범위로 제한되는 경우도 기존 선형 모형의 가정을 위반한다고 볼 수 있습니다.

이와 같이 더 넓은 범위의 데이터 분석 문제에 해당하는 경우 일반화 선형 모형을 적용할 수 있습니다. 일반화 선형 모형은 기존 선형 모형의 확장입니다. 일반화 선형 모형은 선형 성분, 연결 함수 및 분산 함수로 구성됩니다. 연결 함수 $g(\mu_i) = x_i' \beta$ 는 Y_i 의 기대값이 선형 예측 변수와 어떤 관계가 있는지 설명하는 단조 미분 가능 함수입니다. 일반화 선형 회귀의 한 예로 Poisson 회귀가 있으며 여기서 $\log(\mu_i)$ 는 연결 함수입니다. 모형 적합 플랫폼의 일반화 선형 모형 분석법을 사용하여 가능한 일반화 선형 회귀 모형의 전체 목록은 "[일반화 선형 모형 분석법에 대한 통계 상세 정보](#)"(482 페이지)에서 확인하십시오.

적합된 일반화 선형 모형은 기존 선형 모형과 동일한 통계량을 사용하여 요약하고 평가할 수 있습니다. 모형 적합 플랫폼은 일반화 선형 모형에 대한 모수 추정값, 표준 오차, 적합도 통계량, 신뢰 구간 및 가설 검정을 제공합니다. 정확 분포 이론이 일반화 선형 모형에 대해 제공되지 않거나 실현 가능하지 않은 경우도 있다는 점에 유의하십시오. 따라서 일부 추론 절차는 점근적 결과를 기반으로 합니다.

일반화 선형 모형 적합의 중요한 특징은 모형의 설명 변수를 선택하는 것입니다. 적합도 통계량의 변화는 대개 설명 변수의 부분집합이 특정 모형에 미치는 기여도를 평가하는 데 사용됩니다. 이탈도는 도달 가능한 최대 로그 가능도 함수 값과 회귀 모수의 최대 가능도 추정값에서의 로그 가능도 함수 값 차이의 두 배로 정의됩니다. 이탈도는 주로 적합도의 측도로 사용됩니다. 도달 가능한 최대 로그 가능도는 모든 관측값에 대한 모수가 포함된 모형을 사용하여 얻을 수 있습니다.

 일반화 선형 모델링의 변수 선택 및 벌점 방법의 경우 JMP Pro에서 모형 적합 플랫폼의 일반화 회귀 분석법을 사용할 수 있습니다. 자세한 내용은 "[일반화 회귀 모형](#)" 장(263 페이지)에서 확인하십시오.

일반화 선형 모형의 예

이 예에서는 Poisson 회귀를 사용하여 투구게 등지 연구의 개수 데이터를 모델링합니다. 각 암게의 등지에 수계가 있습니다. 이 연구에서는 위성 수계라고 하는 다른 수계가 근처에 있는지 여부를 조사했습니다. CrabSatellites.jmp 데이터 테이블에는 위성 수계의 수를 나열하는 반응 변수와 암게의 색상, 가지 상태, 무게 및 등딱지 폭 길이를 설명하는 변수가 포함되어 있습니다. 위성 수계의 수와 암게를 설명하는 변수 사이의 관계에 관심이 있습니다.

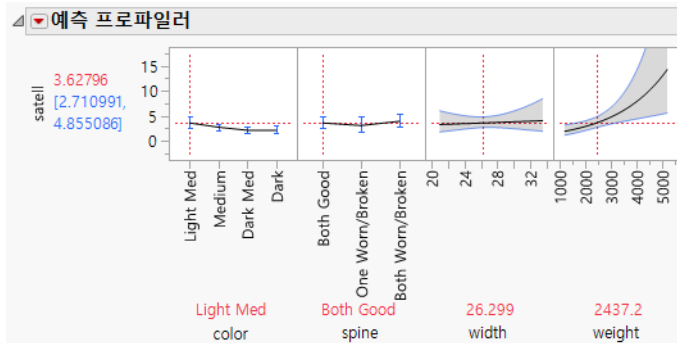
"효과 검정" 보고서에서는 **weight** 가 모형에서 유의한 요인임을 보여 줍니다. **weight** 는 연속형 변수이므로 **weight** 의 p 값 0.0026 이 "모수 추정값" 테이블에서도 동일합니다.

"효과 검정" 보고서를 통해 범주형 변수 **color** 도 유의하다는 것을 알 수 있습니다. **color** 에는 Light Med, Medium, Dark Med 및 Dark 의 네 가지 수준이 있습니다. 처음 세 수준의 모수 추정값이 "모수 추정값" 테이블에 나옵니다. **Dark** 의 모수 추정값은 처음 세 수준에 대한 모수 추정값 합계의 음수입니다.

예측 프로파일러를 사용하여 모형을 자세히 탐색할 수 있습니다.

8. "일반화 선형 모형 적합" 옆의 빨간색 삼각형을 클릭하고 **프로파일러 > 프로파일러**를 선택합니다.

그림 12.3 satell 에 대한 예측 프로파일러



weight 의 신뢰 대역은 **weight** 값이 클 때보다 **weight** 값이 작을 때 모형의 변동성이 적음을 나타냅니다. 프로파일러를 사용하면 범주형 변수의 수준을 쉽게 탐색할 수 있습니다. 프로파일러를 통해 계의 색상이 연한 색에서 진한 색으로 변환에 따라 예측된 위성 계의 수가 줄어드는 것을 확인할 수 있습니다.

일반화 선형 모형 분석법 시작

분석 > 모형 적합을 선택하고 **Y**에 대해 하나 이상의 열을 입력한 후 **분석법** 메뉴에서 **일반화 선형 모형**을 선택하여 일반화 선형 모형 분석법을 시작합니다.

그림 12.4 일반화 선형 모형이 선택된 모형 적합 시작 창

"모형 적합" 창에서 모든 분석법에 공통된 요소에 대한 자세한 내용은 "**모형 규격**" 장(29페이지)에서 확인하십시오. "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오. 여기에서는 일반화 선형 모형 분석법에 대한 정보를 제공합니다.

모형 효과에 결측값이 있는 경우 이러한 결측값을 정보 범주로 처리할 수 있습니다. "모형 규격"의 빨간색 삼각형 메뉴에서 "결측값 정보화" 옵션을 선택합니다.

팁: 모형 적합 플랫폼의 일반화 선형 모형 분석법에서는 "절편 없음" 옵션을 사용할 수 없습니다.

"분석법"에서 "일반화 선형 모형"을 선택하면 모형 적합 시작 창이 변경되어 추가 옵션을 포함합니다. 일반화 선형 모형 분석법에서는 다음 추가 옵션을 사용할 수 있습니다.

분포 반응 변수에 대한 확률 분포를 지정합니다.

연결 함수 선형 모형을 반응 변수와 관련짓는 연결 함수를 지정합니다.

과대산포 검정 및 구간 과대산포 모수가 모형에 포함되도록 지정합니다. 과대산포는 반응 분산이 반응 분포의 이론적 분산에 의해 예상되는 것보다 클 때 발생합니다. Poisson 및 이항 반응 모형에서 과대산포가 발생할 수 있습니다. McCullagh and Nelder 연구 자료 (1989)에 따르면 과대산포는 실제로 흔히 일어나는 일입니다.

참고 : 이 옵션은 "전체 모형 검정" 보고서의 "적합도 통계량" 테이블에 "과대산포" 열을 추가합니다.

Firth 편향 조정 추정값 Firth 편향 조정 방법을 사용하여 모형을 적합시키도록 지정합니다. 이 최대 가능도 기반 방법은 편향 수정을 사용하지 않는 최대 가능도 기반 모형보다 더 나은 추정값과 검정을 생성하는 것으로 나타났습니다. 또한 편향 수정된 MLE 는 로지스틱 유형의 모형에서 발생하기 쉬운 분리 문제도 개선합니다. 로지스틱 회귀의 분리 문제에 대한 자세한 내용은 Firth 연구 자료 (1993) 및 Heinze and Schemper 연구 자료 (2002) 에서 확인하십시오.

오프셋 (역할 변수 선택 버튼으로 표시됨) 오프셋 변수를 지정합니다. 오프셋 변수는 모수가 1.0 으로 고정된 회귀 공변량처럼 처리되는 변수입니다. 오프셋 변수는 로그 연결을 사용하는 Poisson 회귀에서 평균의 모델링을 척도화하는 데 가장 자주 사용됩니다.

이항 반응 분포의 반응 지정

"분포" 에서 "이항" 을 선택하면 다음 방법 중 하나를 사용하여 반응 변수를 지정해야 합니다.

- 데이터가 사건 빈도로 요약되지 않은 경우 단일 이진 열을 반응으로 지정합니다. 반응 열은 명목형이어야 합니다.
- 데이터가 사건 빈도로 요약된 경우 단일 이진 열을 반응으로 지정하고 "빈도" 역할에 빈도 변수를 지정합니다. 반응 열은 명목형이어야 하고 빈도 변수는 각 반응 수준의 수를 포함합니다.
- 데이터가 사건 빈도 및 시행 횟수로 요약된 경우 성공 횟수 - 시행 횟수의 순으로 두 개의 연속형 열을 지정합니다. 성공 횟수 대신 실패 횟수를 지정할 수도 있습니다.

일반화 선형 모형 적합 보고서

기본적으로 "일반화 선형 모형 적합" 보고서에는 모형 규격에 대한 상세 정보와 다음 보고서가 포함됩니다.

특이성 상세 정보 (모형 항 사이에 선형 종속성이 있는 경우에만 표시됨) 모형 항이 충족하는 선형 함수가 포함된 보고서를 표시합니다.

회귀 그림 (연속형 예측 변수가 하나이고 범주형 예측 변수가 최대 한 개인 경우에만 표시됨) 세로 축에 반응을 나타내고 가로 축에 연속형 예측 변수를 나타내는 그림을 표시합니다. 점 위에 회귀선이 표시됩니다. 모형에 범주형 예측 변수가 있는 경우 범주형 예측 변수의 각 수준에 별도의 회귀선이 있고 그림 옆에 범례가 나타납니다.

전체 모형 검정 절편 모수를 제외한 모든 효과가 생략된 모형과 전체 모형 적합을 비교하는 검정을 표시합니다. 이 보고서에는 적합도 통계량 및 수정된 AICc(Akaike 정보 기준) 값도 포함됩니다. 자세한 내용은 "전체 모형 검정"(474 페이지) 에서 확인하십시오.

효과 요약 모형에서 효과를 추가하거나 제거할 수 있는 대화식 보고서입니다. 자세한 내용은 "표준 최소 제곱 모형" 장의 "효과 요약 보고서"(168 페이지) 에서 확인하십시오.

효과 검정 효과 검정은 개별 효과에 대한 모든 모수가 0 이라는 결합 검정입니다. 연속형 효과와 마찬가지로 효과에 모수가 하나만 있는 경우 검정은 모수 추정값 테이블의 검정과 동일합니다.

참고: Firth 조정이 사용되는 경우에도 효과 검정은 별점 없는 가능도 함수를 기반으로 합니다.

모수 추정값 모수 추정값, 표준 오차, 관련된 가설 검정 및 신뢰 한계를 표시합니다. 단순 연속형 효과에는 모수가 하나만 있습니다. 복합 분류 효과가 포함된 모형에는 각 예상 자유도에 대한 모수가 있습니다.

참고: 관측값 수가 1,000 을 초과하면 Wald 기반 신뢰 구간이 표시됩니다. 그렇지 않으면 프로파일 - 가능도 신뢰 구간이 표시됩니다.

스튜던트화 이탈도 잔차 대 예측값 세로 축에 스튜던트화 이탈도 잔차를 나타내고 가로 축에 예측 반응 값을 나타내는 그림을 표시합니다.

전체 모형 검정

"전체 모형 검정" 테이블에는 절편 모수를 제외한 모든 회귀 모수가 생략된 모형과 전체 모형 적합을 비교하는 검정이 표시됩니다. 또한 모형 적합성을 평가하기 위한 두 가지 적합도 통계량과 AICc 값도 포함되어 있습니다.

"전체 모형 검정" 테이블에는 다음과 같은 통계량이 표시됩니다.

모형 모형 라벨을 나열합니다.

차이 전체 모형과 축소 모형 간의 차이입니다. 이 모형은 적합에 대한 전체 회귀변수의 유의성을 측정하는 데 사용됩니다.

전체 절편 및 모든 효과를 포함하는 완전 모형입니다.

축소 절편 모수만 포함하는 모형입니다.

-1* 로그 가능도 각 모형에 대한 음의 로그 가능도입니다. 자세한 내용은 "통계 상세 정보" 부록의 "**가능도, AICc 및 BIC**"(514 페이지)에서 확인하십시오.

L-R 카이제곱 모든 회귀 모수가 0 이라는 가설에 대한 가능도비 카이제곱 검정 통계량입니다. 이 검정 통계량은 적합 모형과 절편만 있는 축소 모형 간의 음의 로그 가능도 차이의 두 배로 계산됩니다.

DF 전체 모형과 축소 모형 간의 차이에 대한 DF(자유도)입니다.

Prob>ChiSq 지정된 모형이 절편만 포함된 모형보다 더 잘 적합되지 않을 경우 카이제곱 값이 더 클 확률입니다.

적합도 통계량 두 가지 적합도 통계량인 Pearson 과 이탈도를 나열합니다.

카이제곱 각 적합도 통계량에 대한 카이제곱 검정 통계량입니다.

DF 각 적합도 통계량에 대한 자유도입니다.

Prob>ChiSq 각 적합도 통계량에 대한 p 값입니다.

과대산포 (시작 창에서 "과대산포 검정 및 구간" 옵션을 선택한 경우에만 표시됨) 과대산포 모형의 추정값입니다. 자세한 내용은 "일반화 선형 모형 분석법에 대한 통계 상세 정보"(482 페이지)에서 확인하십시오.

AICc 수정된 Akaike 정보 기준입니다. 자세한 내용은 "통계 상세 정보" 부록의 "가능도, AICc 및 BIC"(514 페이지)에서 확인하십시오.

일반화 선형 모형 적합 보고서 옵션

"일반화 선형 모형 적합"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

사용자 검정 사용자 가설을 검정할 수 있습니다. 사용자 검정에 대한 자세한 내용은 "표준 최소 제곱 모형" 창의 "**사용자 검정**"(115 페이지)에서 확인하십시오.

대비 변수 내의 수준에 대한 차이를 검정할 수 있습니다. 대비에 공변량이 포함된 경우 대비를 검정할 공변량 값을 지정할 수 있습니다. "대비" 옵션의 예는 "**대비를 사용하여 변수 수준의 차이 비교**"(477 페이지)에서 확인하십시오.

역추정 예측 (연속형 X 변수에만 사용 가능) Y 및 다른 X 변수에 대한 특정 값이 주어졌을 때 X 값을 예측할 수 있습니다. "역추정 예측" 옵션에 대한 자세한 내용은 "표준 최소 제곱 모형" 창의 "**역추정 예측**"(131 페이지)에서 확인하십시오.

추정값 공분산 모형의 모든 효과에 대한 공분산 행렬을 표시하거나 숨깁니다. 모수 추정량의 추정된 공분산 행렬은 다음과 같이 정의됩니다.

$$\Sigma = -\mathbf{H}^{-1}$$

여기서 $\mathbf{H}$ 는 마지막 반복의 모수 추정값을 사용하여 평가된 헤시안 (또는 2 차 도함수) 행렬입니다. 산포 모수는 추정되었는지 아니면 지정되었는지에 관계없이 $\mathbf{H}$ 에 통합되고, 별칭이 지정된 모수에 해당하는 행과 열은 Σ 에 포함되지 않습니다.

추정값 상관계수 모형의 모든 효과에 대한 상관 행렬을 표시하거나 숨깁니다. 상관 행렬은 정규화된 공분산 행렬입니다. Σ 의 각 σ_{ij} 요소에 대해 상관 행렬의 해당 요소는 $\sigma_{ij}/\sigma_i\sigma_j$ 이며, 여기서 $\sigma_i = \sqrt{\sigma_{ii}}$ 입니다.

프로파일러 다음과 같은 프로파일러 하위 메뉴를 표시합니다.

프로파일러 각 X 변수에 대한 예측 트레이스를 검토할 수 있는 예측 프로파일러를 표시하거나 숨깁니다. 예측 프로파일러에 대한 자세한 내용은 "표준 최소 제곱 모형" 창의 "**프로파일러**"(151 페이지)에서 확인하십시오.

등고선 프로파일러 대화식 등고선 프로파일러를 표시하거나 숨깁니다. 등고선 프로파일러에 대한 자세한 내용은 **Profilers** 의 에서 확인하십시오.

표면 프로파일러 3D 표면 프로파일러를 표시하거나 숨깁니다. 표면 프로파일러에 대한 자세한 내용은 **Profilers** 의 에서 확인하십시오.

진단 그림 잔차, 예측값 및 실제값 그림을 위한 하위 메뉴를 표시합니다. 이러한 그림을 통해 이상치를 검색하고 모형의 적합성을 판단할 수 있습니다. 이탈도에 대한 자세한 내용은 "**모형 선택 및 이탈도**"(484 페이지)에서 확인하십시오. 다음과 같은 그림을 사용할 수 있습니다.

스튜던트화 이탈도 잔차 대 예측값 세로 축에 스튜던트화 이탈도 잔차를 나타내고 가로 축에 예측 반응 값을 나타내는 그림을 표시하거나 숨깁니다.

스튜던트화 Pearson 잔차 대 예측값 세로 축에 스튜던트화 Pearson 잔차를 나타내고 가로 축에 예측 반응 값을 나타내는 그림을 표시하거나 숨깁니다.

이탈도 잔차 대 예측값 세로 축에 이탈도 잔차를 나타내고 가로 축에 예측 반응 값을 나타내는 그림을 표시하거나 숨깁니다.

Pearson 잔차 대 예측값 세로 축에 Pearson 잔차를 나타내고 가로 축에 예측 반응 값을 나타내는 그림을 표시하거나 숨깁니다.

회귀 그림 (연속형 예측 변수가 하나이고 범주형 예측 변수가 최대 한 개인 경우에만 사용 가능) 세로 축에 반응을 나타내고 가로 축에 연속형 예측 변수를 나타내는 그림을 표시하거나 숨깁니다. 점 위에 회귀선이 표시됩니다. 모형에 범주형 예측 변수가 있는 경우 범주형 예측 변수의 각 수준에 별도의 회귀선이 있고 그림 옆에 범례가 나타납니다.

선형 예측 변수 그림 (연속형 예측 변수가 하나이고 범주형 예측 변수가 최대 한 개인 경우에만 사용 가능) 세로 축에 역 연결 함수에 의해 변환된 반응을 나타내고 가로 축에 연속형 예측 변수를 나타내는 그림을 표시하거나 숨깁니다. 점 위에 변환된 회귀선이 표시됩니다. 모형에 범주형 예측 변수가 있는 경우 범주형 예측 변수의 각 수준에 별도의 변환된 회귀선이 있고 그림 옆에 범례가 나타납니다.

열 저장 특정 통계량을 현재 데이터 테이블의 새 열로 저장할 수 있는 옵션이 포함된 하위 메뉴를 표시합니다. 잔차 계산식에 대한 자세한 내용은 "[잔차 계산식](#)"(485 페이지)에서 확인하십시오.

예측 계산식 현재 데이터 테이블에 모형을 예측하는 계산식 열을 생성합니다.

예측값 현재 데이터 테이블에 모형에서 예측한 값을 포함하는 열을 생성합니다.

평균 신뢰 구간 현재 데이터 테이블에 모형에 대한 예측 방정식의 95% 신뢰 한계를 포함하는 열을 생성합니다. 이러한 신뢰 한계는 모수 추정값의 변동을 반영합니다.

참고 : "모형 적합" 창의 "모형 규격" 빨간색 삼각형 메뉴에서 "유의 수준 설정"을 선택하여 신뢰 한계의 α 수준을 변경할 수 있습니다.

개별 신뢰 한계 저장 현재 데이터 테이블에 모형에 대한 지정된 개별 값의 95% 신뢰 한계를 포함하는 열을 생성합니다. 이러한 신뢰 한계는 오차의 변동과 모수 추정값의 변동을 반영합니다.

참고 : "모형 적합" 창의 "모형 규격" 빨간색 삼각형 메뉴에서 "유의 수준 설정"을 선택하여 신뢰 한계의 α 수준을 변경할 수 있습니다.

이탈도 잔차 현재 데이터 테이블에 이탈도 잔차를 포함하는 열을 생성합니다.

Pearson 잔차 현재 데이터 테이블에 Pearson 잔차를 포함하는 열을 생성합니다.

스튜던트화 이탈도 잔차 현재 데이터 테이블에 스튜던트화 이탈도 잔차를 포함하는 열을 생성합니다.

스튜던트화 Pearson 잔차 현재 데이터 테이블에 스튜던트화 Pearson 잔차를 포함하는 열을 생성합니다.

모형 대화상자 현재 분석을 위해 완료된 시작 창을 표시합니다.

효과 요약 모형의 효과를 대화식으로 업데이트할 수 있는 "효과 요약" 보고서를 표시하거나 숨깁니다. 자세한 내용은 "표준 최소 제곱 모형" 장의 "[효과 요약 보고서](#)"(168 페이지)에서 확인하십시오.

다음 옵션에 대한 자세한 내용은 **JMP 사용**의 에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

일반화 선형 모형 분석법의 추가 예

- "[대비를 사용하여 변수 수준의 차이 비교](#)"
- "[오프셋을 사용하는 Poisson 회귀](#)"
- "[로그 연결을 사용하는 정규 회귀](#)"

대비를 사용하여 변수 수준의 차이 비교

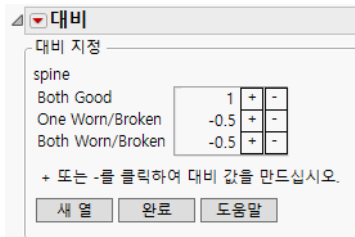
이 예에서는 "[일반화 선형 모형의 예](#)"(469 페이지)의 위성 계 예를 계속합니다. 가시 상태가 좋은 암계가 유인하는 수계(위성)의 수가 가시가 닳거나 부러진 암계와 다른지 여부를 검정하려고 한다고 가정해 보겠습니다.

1. "[일반화 선형 모형의 예](#)"(469 페이지)에 나오는 1 단계 ~ 7 단계를 완료합니다.
2. "일반화 선형 모형 적합" 옆의 빨간색 삼각형을 클릭하고 **대비**를 선택합니다. 그러면 "대비 효과를 선택하십시오" 창이 나타납니다.
3. **spine**(관심 변수)을 선택하고 **확인**을 클릭합니다.

4. 가시 상태가 좋은 계를 가시가 닳거나 부러진 계와 비교하려면 Both Good 옆의 + 버튼을 클릭하고 One Worn/Broken 과 Both Worn/Broken 옆의 - 버튼을 각각 클릭합니다.

이렇게 하면 가시 상태가 좋은 암계를 가시가 닳거나 부러진 암계와 비교하는 대비 지정이 생성됩니다.

그림 12.5 대비 지정 창



5. 완료를 클릭합니다.

그림 12.6 대비 보고서



Prob>Chisq 값 0.8242는 0.05보다 훨씬 크기 때문에 가시 상태에 따라 위성 계 유인에 차이가 있다는 결론을 내릴 수 없습니다.

오프셋을 사용하는 Poisson 회귀

Ship Damage.JMP 샘플 데이터 테이블은 McCullagh and Nelder 연구 자료(1989)의 데이터에서 채택되었습니다. 이 데이터 테이블에는 파도로 인해 선체 앞쪽 부분에 발생한 특정 유형의 손상 정보가 있습니다. 선체 건조 엔지니어는 선체 유형 (Type), 선체 건조 연도 (Yr Made) 및 선체 이용 연도 (Yr Used)의 세 가지 변수와 관련된 손상 위험에 관심이 있습니다.

이 분석에서는 총 이용 개월 수의 로그인 Service 변수를 오프셋 변수로 사용합니다. 오프셋 변수는 모수가 1.0으로 고정된 회귀 공변량처럼 처리되는 변수입니다. 오프셋 변수는 로그 연결을 사

용하는 Poisson 회귀에서 평균의 모델링을 척도화하는 데 가장 자주 사용됩니다. 이 예에서는 수리 횟수가 이용 개월 수에 비례할 것으로 예상되므로 $\log(\text{이용 개월 수})$ 를 사용합니다.

오프셋 변수가 어떻게 사용되는지 확인하기 위해 GLM의 선형 성분을 η 라고 가정합니다. 그런 다음 로그 연결 함수를 사용하여 오프셋이 포함된 평균 모형을 다음과 같이 지정합니다.

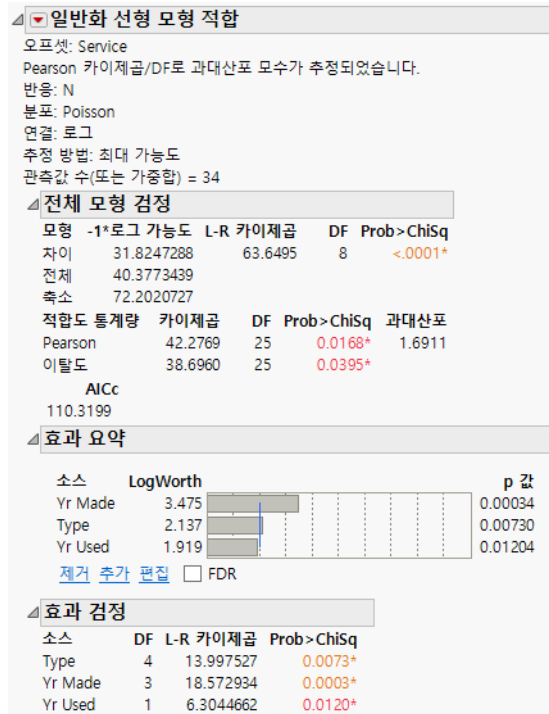
$$\exp[\text{Log}(\text{이용 개월 수}) + \eta] = [(\text{이용 개월 수}) * \exp(\eta)].$$

이 예를 실행하려면 다음 단계를 수행하십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Ship Damage.jmp를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. "분석법" 목록에서 **일반화 선형 모형**을 선택합니다.
4. "분포" 목록에서 **Poisson**을 선택합니다.
"연결 함수" 목록에서 **로그**가 자동으로 선택되어야 합니다.
5. **N**을 선택하고 **Y**를 클릭합니다.
6. **Service**를 선택하고 **오프셋**을 클릭합니다.
7. **Type, Yr Made, Yr Used**를 선택하고 **추가**를 클릭합니다.
8. **과대산포 검정 및 구간** 체크박스를 클릭합니다.
9. **실행**을 클릭합니다.

보고서에서 세 가지 효과 (Type, Yr Made, Yr Used) 가 모두 유의하다는 것을 알 수 있습니다.

그림 12.7 오프셋을 사용하는 Poisson 모형에 대한 보고서 (일부분)

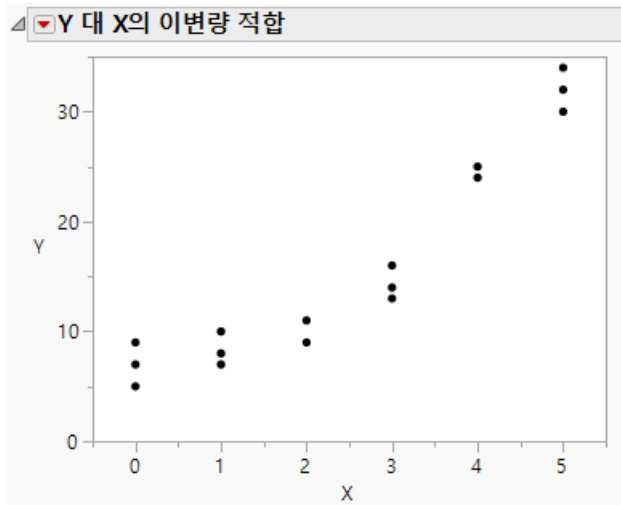


로그 연결을 사용하는 정규 회귀

이 예에서는 Nor.jmp 데이터 테이블을 사용합니다. 여기서 X는 설명 변수이고 Y는 반응 변수입니다. 먼저 X와 Y 사이의 관계를 탐색하여 모형 적합 플랫폼의 일반화 선형 모형 분석법에 사용할 적절한 연결 함수를 결정합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Nor.jmp를 엽니다.
2. **분석 > X로 Y 적합**을 선택합니다.
3. Y를 선택하고 **Y, 반응**을 클릭합니다.
4. X를 선택하고 **X, 요인**을 클릭한 후 **확인**을 클릭합니다.

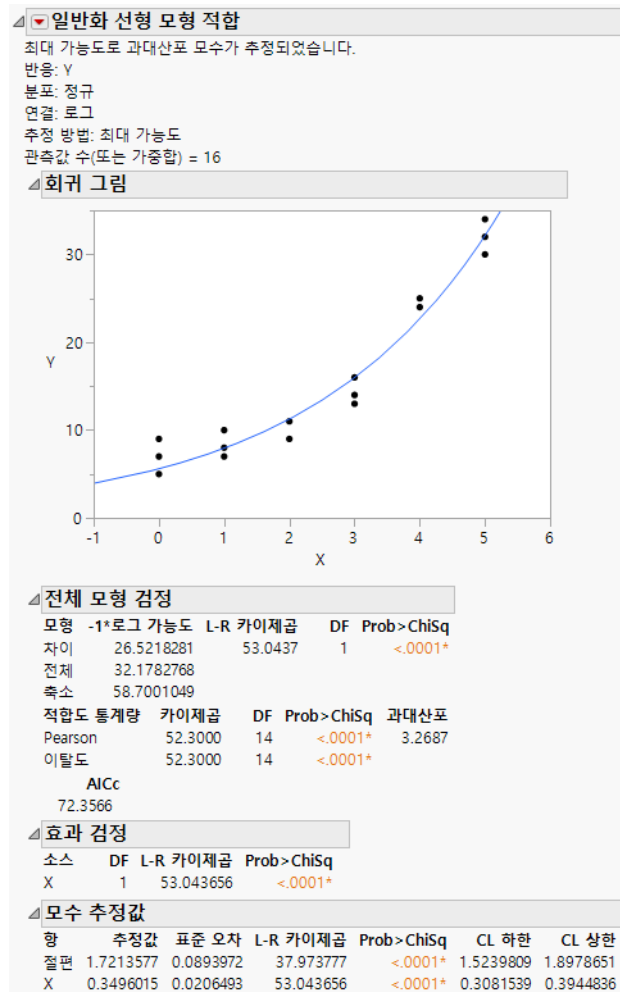
그림 12.8 Nor.jmp 에 대한 X 로 Y 적합 결과



Y 가 X 에 따라 비선형 방식으로 변하고 분산이 거의 일정하다는 것을 확인할 수 있습니다 .
따라서 이러한 데이터를 모델링하려면 로그 연결 함수를 사용하는 정규 분포가 적합합니다 .
즉, $\log(\mu_i) = \mathbf{x}_i' \boldsymbol{\beta}$ 이고 따라서 $\mu_i = \exp(\mathbf{x}_i' \boldsymbol{\beta})$ 입니다 .

5. 분석 > 모형 적합을 선택합니다 .
6. " 분석법 " 목록에서 일반화 선형 모형을 선택합니다 .
7. " 분포 " 목록에서 정규를 선택합니다 .
8. " 연결 함수 " 목록에서 로그를 선택합니다 .
9. Y 를 선택하고 Y 를 클릭합니다 .
10. x 를 선택하고 추가를 클릭합니다 .
11. 실행을 클릭합니다 .

그림 12.9 Nor 결과



일반화 선형 모형 분석법에 대한 통계 상세 정보

일반화 선형 모형을 생성하려면 데이터에 대한 반응 변수와 설명 변수를 선택해야 합니다. 그런 다음 적절한 연결 함수와 반응 확률 분포를 선택해야 합니다. 설명 변수는 연속형 변수, 분류 변

수 및 교호작용의 조합으로 지정할 수 있습니다. 표 12.1 에는 일반화 선형 모형의 몇 가지 일반적인 예가 나열되어 있습니다.

표 12.1 일반화 선형 모형의 예

모형	반응 변수	분포	기본 연결 함수
기존 선형 모형	연속형	정규	항등, $g(\mu) = \mu$
로지스틱 회귀	개수 또는 이진 확률 변수	이항	로짓, $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$
로그 선형 모형의 Poisson 회귀	개수	Poisson	로그, $g(\mu) = \log(\mu)$
지수 회귀	연속형 양수	지수	$\frac{1}{\mu}$

플랫폼에서 모수 벡터의 최대 가능도 추정을 통해 일반화 선형 모형을 데이터에 적합시킵니다. 일반적으로 모수의 최대 가능도 추정에 대한 닫힌 형식의 해는 없습니다. 따라서 이 플랫폼에서는 Nelder and Wedderburn 연구 자료(1972)에서 최초로 공개된 기법을 사용하여 반복 적합 프로세스를 통해 모형의 모수를 수치적으로 추정합니다. 과대산포 모수 ϕ 는 Pearson 적합도 통계량을 자유도로 나누어 추정합니다. 공분산, 표준 오차 및 신뢰 한계는 최대 가능도 추정량의 점근적 정규성을 기반으로 추정된 모수에 대해 계산됩니다.

모형 적합 플랫폼의 일반화 선형 모형 분석법에서 다양한 연결 함수 및 확률 분포를 사용할 수 있습니다. 표 12.2에는 기본 제공 연결 함수가 나열되어 있습니다.

표 12.2 기본 제공 연결 함수

연결 함수 이름	연결 함수 계산식
항등	$g(\mu) = \mu$
로짓	$g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$
프로빗	$g(\mu) = \Phi^{-1}(\mu)$. Φ = 표준 정규 누적 분포 함수
로그	$g(\mu) = \log(\mu)$
역수	$g(\mu) = \frac{1}{\mu}$

표 12.2 기본 제공 연결 함수 (계속)

연결 함수 이름	연결 함수 계산식
검정력	$g(\mu) = \begin{cases} \mu^\lambda & \text{단, } (\lambda \neq 0) \\ \log(\mu) & \text{단, } \lambda = 0 \end{cases}$
보 로그-로그	$g(m) = \log(-\log(1 - \mu))$

검정력 연결 함수를 선택하면 원하는 검정력을 입력할 수 있는 숫자 상자가 나타납니다.
표 12.3에는 반응 변수에 사용할 수 있는 분포와 관련된 분산 함수가 나열되어 있습니다.

표 12.3 반응 분포에 대한 분산 함수

분포	분산 함수
정규	$V(\mu) = 1$
이항	$V(\mu) = \mu(1 - \mu)$
Poisson	$V(\mu) = \mu$
지수	$V(\mu) = \mu^2$

모형 선택 및 이탈도

일반화 선형 모델링의 중요한 특징은 모형의 설명 변수를 선택하는 것입니다. 적합도 통계량의 변화는 대개 설명 변수의 부분집합이 특정 모형에 미치는 기여도를 평가하는 데 사용됩니다. 이탈도는 도달 가능한 최대 로그 가능도와 회귀 모수의 최대 가능도 추정값에서의 로그 가능도 차이의 두 배로 정의됩니다. 이탈도는 주로 적합도의 측도로 사용됩니다. 도달 가능한 최대 로그 가능도는 모든 관측값에 대한 모수가 포함된 모형을 사용하여 얻을 수 있습니다. 표 12.4에는 반응 변수에 사용할 수 있는 각 분포에 대한 이탈도 계산식이 나열되어 있습니다.

표 12.4 반응 분포에 대한 이탈도 계산식

분포	이탈도 계산식
정규	$\sum_i w_i (y_i - \mu_i)^2$
이항	$2 \sum_i w_i m_i \left[y_i \log \left(\frac{y_i}{\mu_i} \right) + (1 - y_i) \log \left(\frac{1 - y_i}{1 - \mu_i} \right) \right]$

표 12.4 반응 분포에 대한 이탈도 계산식 (계속)

분포	이탈도 계산식
Poisson	$2 \sum_i w_i \left[y_i \log \left(\frac{y_i}{\mu_i} \right) - (y_i - \mu_i) \right]$
지수	$2 \sum_i w_i \left[-\log \left(\frac{y_i}{\mu_i} \right) + \left(\frac{y_i - \mu_i}{\mu_i} \right) \right]$

Pearson 카이제곱 통계량은 다음과 같이 정의됩니다 .

$$X^2 = \sum_i \frac{w_i (y_i - \mu_i)^2}{V(\mu_i)}$$

다음은 각 요소에 대한 설명입니다.

$y_i = i$ 번째 반응

$\mu_i =$ 해당하는 예측 평균

$V(\mu_i) =$ 분산 함수

$w_i = i$ 번째 관측값에 대한 알려진 가중치

참고 : 가중치를 지정하지 않으면 모든 관측값에 대해 $w_i = 1$ 입니다 .

변수를 선택하는 한 가지 전략은 모형 시퀀스를 적합시키는 것입니다. 절편 항만 포함된 단순 모형으로 시작하여 각 연속 모형에 추가 설명 변수를 하나씩 포함합니다. 연속 모형 간의 이탈도 값이나 적합 로그 가능도 값의 차이로 추가 설명 변수의 중요도를 측정할 수 있습니다. 점근적 검정을 사용하면 추가 항의 통계적 유의성을 평가할 수 있습니다.

비정규 분포의 경우 역추정 예측에서 t 분포 임계값 대신 정규 임계값이 사용됩니다.

잔차 계산식

이탈도

$$r_{Di} = \sqrt{d_i} (\text{sign}(y_i - \mu_i))$$

스튜던트화 이탈도

$$r_{Di} = \frac{\text{sign}(y_i - \mu_i) \sqrt{d_i}}{\sqrt{\phi(1 - h_i)}}$$

Pearson

$$r_{Pi} = \frac{(y_i - \mu_i)}{\sqrt{V(\mu_i)}}$$

스튜던트화 Pearson

$$r_{Pi} = \frac{y_i - \mu_i}{\sqrt{V(\mu_i)(1 - h_i)}}$$

다음은 각 요소에 대한 설명입니다.

$(y_i - \mu_i)$ = 원시 잔차

$\text{sign}(y_i - \mu_i) = (y_i - \mu_i)$ 가 양수인 경우 1, $(y_i - \mu_i)$ 가 음수인 경우 -1

d_i = 관측값 i 의 총 이탈도에 대한 기여도

ϕ = 산포 모수

$V(\mu_i)$ = 분산 함수

$h_i = W_e^{(1/2)} X (X' W_e X)^{-1} X' W_e^{(1/2)}$ 행렬의 i 번째 주대각 요소. 여기서 W_e 는 기대 정보 행렬을 계산하는 데 사용되는 가중치 행렬입니다.

잔차 및 일반화 선형 모형에 대한 자세한 내용은 SAS Institute Inc. (2020a).

부록 A

통계 상세 정보 선형 모형 적합

이 부록에서는 다양한 유형의 반응 모형, 요인, 설계 코딩 및 파라미터화에 대해 설명합니다. 또한 본문에 설명된 방법에 대한 기타 여러 상세 정보가 포함되어 있습니다.

JMP 시스템은 선형 모형을 연속형, 순서형 및 명목형이라는 라벨이 지정된 세 가지 유형의 반응 모형에 적합시킵니다. 요인에 대한 많은 상세 정보는 다른 반응 모형에서도 동일하지만 JMP는 연속형 반응에 대해서만 그래픽 및 주변 프로파일을 지원합니다. 순서형 및 명목형 반응에는 이러한 기능이 지원되지 않습니다.

효과 적합 및 가설 검정 생성을 위해 컴퓨터 프로그램마다 서로 다른 설계 행렬 코딩을 사용하고 결과적으로 파라미터화에 영향을 줍니다. JMP는 SAS 시스템의 GLM 프로시저와 다른 코딩을 사용하지만 대부분의 경우 JMP와 SAS GLM 프로시저를 통해 생성되는 결과는 동일합니다. 다음 섹션에서는 JMP 코딩 상세 정보를 설명하고, SAS GLM 프로시저의 코딩과 다른 경우 해당 사례를 강조합니다.

목차

반응 모형.....	489
연속형 반응.....	489
명목형 반응.....	489
순서형 반응.....	491
요인 모형.....	492
연속형 요인.....	492
명목형 요인.....	493
순서형 요인.....	503
빈도.....	508
일반적인 가정.....	509
가정된 모형.....	509
상대적 유의성.....	509
다중 추론.....	509
타당도 평가.....	509
대체 방법.....	510
핵심 통계적 개념.....	510
불확도, 통합 개념.....	511
두 개의 기본 적합 기계.....	511
가능도, AICc 및 BIC.....	514
검정력 계산.....	515
LSN 계산.....	516
LSV 계산.....	516
검정력 계산.....	517
조정 검정력 계산.....	518
신뢰 한계를 사용한 역추정 예측.....	519

반응 모형

JMP는 선형 모형을 연속형, 명목형 및 순서형이라는 세 가지 유형의 반응에 적합시킵니다. JMP에서 사용할 수 있는 모형과 방법은 실용적이고 널리 사용되며 통계 소프트웨어 도구에서 일반적인 접근 방식의 요구 사항에 잘 맞습니다. 모든 통계 소프트웨어에서 그렇듯이 사용하기로 선택한 모형의 가정 및 가정이 충족되지 않을 경우 발생하는 결과에 대해 학습해야 합니다. 자세한 내용은 이 장의 "[일반적인 가정](#)"(509 페이지)에서 확인하십시오.

- "[연속형 반응](#)"
- "[명목형 반응](#)"
- "[순서형 반응](#)"

연속형 반응

반응 열 (Y 역할이 할당된 열)이 연속형인 경우 JMP에서 반응 값을 직접 적합시킵니다. 각 관측값에 대한 기본 모형은 다음과 같습니다.

$$Y = (X \text{ 및 모수의 함수}) + \text{오차}$$

통계 검정은 모형의 오차 항이 정규 분포를 따른다는 가정을 기반으로 합니다.

연속형 반응의 적합 원칙

이 적합 원칙을 **최소 제곱**이라고 합니다. 최소 제곱 방법은 오차 제곱합을 최소화하도록 모형의 모수를 추정합니다. 적합 모형의 오차(잔차)는 각 관측값의 실제값과 적합 모형에서 예측한 값 사이의 차이입니다.

오차가 정규 분포를 따르는 경우 최소 제곱 방법은 최대 가능도 추정 방법과 동등합니다. 이는 분석을 통해 가장 가능성이 높은 잔차를 제공하는 모형을 추정한다는 의미입니다. 로그 가능도는 정규 분포에 대한 오차 제곱합의 척도 배수입니다.

연속형 반응의 기준 모형

연속형 측정에 대한 가장 단순한 모형은 하나의 값만 적합시켜 모든 반응 값을 예측합니다. 이 값은 **평균**의 추정값입니다. 평균은 반응 값의 산술평균입니다. 다른 모든 모형이 이 기준 모형과 비교됩니다.

명목형 반응

명목형 반응은 로짓 모형의 간단한 확장을 통해 분석됩니다. 이항 (2 수준) 반응의 경우 로짓 반응 모형은 다음과 같이 지정됩니다.

$$\log\left(\frac{P(y = 1)}{P(y = 2)}\right) = X\beta$$

위의 모형을 다음과 같이 작성할 수도 있습니다.

$$P(y = 1) = F(X\beta)$$

여기서 $F(x)$ 는 표준 로지스틱 분포의 누적 분포 함수입니다.

$$F(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x}$$

반응 수준이 r 개인 경우 JMP 는 반응이 데이터 값에 의해 주어진 r 개의 여러 반응 수준 중 하나 일 확률을 적합시킵니다. 확률 추정값은 모두 양수여야 합니다. 주어진 X 구성에 대해 확률 추정값의 합은 전체 반응 수준에서 1 이어야 합니다. JMP 에서 확률을 예측하는 데 사용되는 함수는 선형 모형과 다중 반응 로지스틱 함수의 합성입니다. 확률 비율의 로그가 선형 모형이므로 로그 선형 모형이라고도 합니다. JMP 는 각 반응 확률을 r 번째 확률과 관련짓고, 별도의 설계 모 수 집합을 이러한 $r - 1$ 개 모형에 적합시킵니다.

$$\log\left(\frac{P(y = j)}{P(y = r)}\right) = X\beta_{(j)} \quad j = 1, \dots, r - 1 \text{의 경우}$$

명목형 반응의 적합 원칙

이 적합 원칙을 최대 가능도라고 합니다. 이 방법은 데이터에 의해 제공된 모든 반응의 결합 확률이 모형에서 얻을 수 있는 최대값이 되도록 모수를 추정합니다. 결합 확률(가능도)을 직접 보고하는 것보다 가능도의 음의 로그 합계를 보고하는 것이 관리하기가 더 쉽습니다.

불확도 (음의 로그 가능도) 는 모형이 표본 데이터에서 실제로 발생한 반응에 기인한 확률의 음의 로그 합입니다. 이 값은 크기가 n 인 표본에 대해 H 로 표시되고 다음과 같이 작성됩니다.

$$H = \sum_{i=1}^n -\log(P(y = y_i))$$

확률 1을 발생한 각 사건에 기인한 것으로 볼 경우 완벽한 적합을 위한 음의 로그 합은 0입니다.

명목형 모형, 특히 반응 수준이 많은 모형을 적합시키려면 시간 및 메모리가 많이 소요될 수 있습니다. JMP 는 음의 로그 가능도 값이 추정값에 수렴할수록 작아지는 것을 보여 주는 반복 기록을 사용하여 계산 진행 정도를 추적합니다.

명목형 반응의 기준 모형

명목형 반응에 대한 가장 단순한 모형은 전체 데이터 테이블에서 각 반응 수준에 대한 발생률로 적합된 일련의 상수 반응 확률입니다. 즉, y 가 반응 수준 j 일 확률은 총 표본 수 n 을 각 반응 수준 n_j 의 총계로 나누어 추정됩니다. 이 확률은 다음과 같이 지정됩니다.

$$p_j = \frac{n_j}{n}$$

다른 모든 모형이 이 기준 모형과 비교됩니다. 명목형 반응에 대한 기준 모형은 연속형 모형에 대한 표본 평균과 동일하게 사용됩니다.

R^2 통계량은 다음과 같이 모형에 의해 설명되는 불확도 비율을 측정합니다.

$$1 - \frac{H(\text{전체 모형})}{H(\text{기준 모형})}$$

그러나 실제로 범주형 모형에 대해 1에 가까운 R^2 을 얻기는 어렵습니다.

순서형 반응

순서형 반응(Y)을 사용할 경우 JMP는 명목형 반응과 마찬가지로 반응이 데이터에 의해 주어진 r 개의 여러 반응 수준 중 하나일 확률을 적합시킵니다.

순서형 데이터에는 연속형 데이터와 같이 순서가 있습니다. 순서는 분석에 사용되지만 정렬된 순서 사이의 간격이나 거리는 사용되지 않습니다. 수치형 반응이 있지만 모형에서 값 간격을 무시하려면 해당 반응 열에 순서형 수준을 할당하면 됩니다. 분류 변수가 있고 수준이 낮음, 중간, 높음과 같이 자연 순서인 경우 순서형 모델링 유형을 사용할 수 있습니다.

순서형 반응은 일련의 병렬 로지스틱 곡선을 누적 확률에 적합시켜 모델링됩니다. 각 곡선은 설계 모수가 동일하지만 절편이 다르며 다음과 같이 지정됩니다.

$$P(y \leq j) = F(\alpha_j + X\beta) \quad j = 1, \dots, r-1 \text{의 경우}$$

여기서 r 개의 반응 수준이 있고 $F(x)$ 는 표준 로지스틱 누적 분포 함수입니다.

$$F(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x}$$

다음과 같이 여러 임계값을 통과할 때 순서형 반응이 변경되는 관측되지 않은 연속형 변수 z 의 관점으로 작성할 수도 있습니다.

$$y = \begin{cases} r & \alpha_{r-1} \leq z \\ j & \alpha_{j-1} \leq z < \alpha_j \\ 1 & z \leq \alpha_1 \end{cases}$$

여기서 z 는 관측 불가능한 선형 모형과 오차의 함수입니다.

$$z = X\beta + \varepsilon$$

ε 는 로지스틱 분포를 따릅니다.

이러한 모형은 반응의 순서 특성을 인식하고 명목형 모형에 비해 필요한 모수 수가 훨씬 적으며 계산이 빠르기 때문에 유용합니다.

순서형 모형을 구성하는 방법은 승산을 모델링하는 대신 누적 분포를 모델링하는 명목형 모형의 경우와 다르지만 수학적으로 동등합니다. 마지막 수준을 제외한 모든 수준에 대해 함수를 적합시키는 대신 하나의 함수만 적합시킨 후 각 누적 반응 확률을 적합시킵니다.

순서형 반응의 적합 원칙

순서형 반응 모형에 대한 최대 가능도 적합 원칙은 명목형 반응 모형의 경우와 동일합니다. 이 방법은 발생하는 모든 반응의 결합 확률이 모형에서 얻을 수 있는 최대값이 되도록 모수를 추정합니다. 명목형 적합에 비해 더 빠르고 메모리를 절약하는 반복 방법을 사용합니다.

기준 모형

순서형 반응에 대한 가장 단순한 모형은 명목형 반응과 마찬가지로 전체 데이터 테이블에서 반응에 대한 발생률로 적합된 일련의 반응 확률입니다.

요인 모형

x 변수 (요인) 를 모델링하여 기대값 또는 확률을 예측하는 방법은 모형 요인의 핵심입니다.

요인은 x 값과 추정할 모수의 선형 결합으로 예측 방정식에 참여합니다. i 가 관측값을 인덱싱하고 j 가 모수를 인덱싱하는 연속형 반응 모형의 경우 일반적인 관측값의 가정된 모형 y_i 는 다음과 같이 작성됩니다.

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i$$

다음은 각 요소에 대한 설명입니다.

y_i - 반응

x_{ij} - 데이터의 함수

ε_i - 관측 불가능한 랜덤 오차 구현

b_j - 추정할 미지의 모수

요인 항에서 선형 모형의 x 가 구성되는 방식은 각 모델링 유형에 따라 다릅니다. 선형 모형 x 는 교호작용 또는 내포 효과와 같은 복합 효과일 수도 있습니다. 복합 효과는 이후에 자세히 논의됩니다.

연속형 요인

연속형 요인은 설계 행렬에 직접 회귀변수로 배치됩니다. 열이 다른 열의 선형 함수인 경우 이 열의 모수는 **영점화됨** 또는 **추정할 수 없음**으로 표시됩니다. 연속형 요인은 다른 요인과 교차할 때 (교호작용 또는 다항식 항) 평균으로 중심화됩니다. 요인의 열 특성이 **혼합물** 또는 **코딩**이거나, 모형을 지정할 때 중심화 다항식 옵션이 해제된 경우에는 중심화가 억제됩니다. 코딩 열 특성이 있으면 적합 전에 요인이 코드화됩니다.

명목형 요인

명목형 요인은 설계 행렬에 대한 표시 변수로 변환됩니다. SAS GLM은 각 명목형 수준에 대한 표시자 열을 생성합니다. JMP는 마지막 수준을 제외한 각 명목형 수준에 대해 동일한 표시자 열을 생성합니다. 마지막 명목형 수준이 나오면 요인의 다른 모든 열에서 하나가 빠집니다. 아래와 같이 GLM 및 JMP에 대해 각각 세 가지 수준이 코드화된 명목형 요인 A를 예로 들어보겠습니다.

표 A.1 명목형 요인 A

	GLM			JMP	
A	A1	A2	A3	A13	A23
A1	1	0	0	1	0
A2	0	1	0	0	1
A3	0	0	1	-1	-1

GLM의 선형 모형 설계 행렬에는 열 사이에 선형 종속성이 있으며 최소 제곱 해는 일반화 역행렬을 사용합니다. 선택된 해는 A3 모수가 0으로 설정되게 합니다.

JMP의 선형 모형 설계 행렬은 결측 셀 또는 기타 부수적 공선성이 있는 경우를 제외하고 완전 계수(full rank)가 되도록 코드화됩니다. 마지막 수준에 대한 A 효과의 모수는 다른 수준의 음의 합이므로 모든 효과 수준에서 모수의 합이 0이 됩니다.

모수 해석

참고 : 명목형 수준의 모수는 해당 수준에 대한 예측 반응과 모든 수준에 대한 평균 예측 반응 간의 차이로 해석됩니다.

요인 수준에 대한 설계 열은 해당 요인 수준의 0-1 표시자에서 마지막 수준의 표시자를 뺀 값으로 생성됩니다. 이 코딩으로 인해 위의 모수 해석이 가능합니다.

표 A.2 모수 해석

JMP 모수 보고서	해석 방법	설계 열 코딩
절편	모든 수준에 대한 평균	$1'$
A[1]	$\alpha_1 - 1/3(\alpha_1 + \alpha_2 + \alpha_3)$	$(A=1) - (A=3)$
A[2]	$\alpha_2 - 1/3(\alpha_1 + \alpha_2 + \alpha_3)$	$(A=2) - (A=3)$

교호작용 및 교차효과

GLM 과 JMP 에서는 교차되는 요인의 설계 열 행에 직접곱을 적용하여 교호작용 효과를 생성합니다. 예를 들어 다음 GLM 코드는

```
PROC GLM;  
  CLASS A B;  
  MODEL A B A*B;
```

아래와 같은 설계 행렬을 생성합니다.

표 A.3 설계 행렬

		A			B			AB								
A	B	1	2	3	1	2	3	11	12	13	21	22	23	31	32	33
A1	B1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0
A1	B2	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0
A1	B3	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0
A2	B1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0
A2	B2	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0
A2	B3	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0
A3	B1	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0
A3	B2	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0
A3	B3	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1

JMP 모형 적합 명령을 사용하고 A 및 B 열에 대한 요인 모형을 요청하면 다음과 같은 설계 행렬이 생성됩니다. 이 행렬의 A13은 이전 행렬의 A1–A3입니다. 그러나 A13B13은 현재 행렬에서 A13*B13입니다.

표 A.4 현재 행렬

		A		B					
A	B	13	23	13	23	A1 B1	A1 B1	A1 B1	A1 B1
A1	B1	1	0	1	0	1	0	0	0
A1	B2	1	0	0	1	0	1	0	0
A1	B3	1	0	-1	-1	-1	-1	0	0

표 A.4 현재 행렬 (계속)

A2	B1	0	1	1	0	0	0	1	0
A2	B2	0	1	0	1	0	0	0	1
A2	B3	0	1	-1	-1	0	0	-1	-1
A3	B1	-1	-1	1	0	-1	0	-1	0
A3	B2	-1	-1	0	1	0	-1	0	-1
A3	B3	-1	-1	-1	-1	1	1	1	1

JMP 코딩을 사용하면 수준 수가 적은 요인의 교호작용 문제에 대해 메모리 및 계산 시간을 절약할 수 있습니다.

3x3 교차 모형에 대해 모수에 관한 셀의 기대값은 다음과 같습니다.

표 A.5 3x3 교차 모형

	B1	B2	B3
A1	$\mu + \alpha_1 + \beta_1 + \alpha\beta_{11}$	$\mu + \alpha_1 + \beta_2 + \alpha\beta_{12}$	$\mu + \alpha_1 - \beta_1 - \beta_2 - \alpha\beta_{11} - \alpha\beta_{12}$
A2	$\mu + \alpha_2 + \beta_1 + \alpha\beta_{21}$	$\mu + \alpha_2 + \beta_2 + \alpha\beta_{22}$	$\mu + \alpha_2 - \beta_1 - \beta_2 - \alpha\beta_{21} - \alpha\beta_{22}$
A3	$\mu - \alpha_1 - \alpha_2 + \beta_1 - \alpha\beta_{11} - \alpha\beta_{21}$	$\mu - \alpha_1 - \alpha_2 + \beta_2 - \alpha\beta_{12} - \alpha\beta_{22}$	$\mu - \alpha_1 - \alpha_2 - \beta_1 - \beta_2 + \alpha\beta_{11} + \alpha\beta_{12} + \alpha\beta_{21} + \alpha\beta_{22}$

내포 효과

GLM은 모형에 없는 것을 기준으로 우측 검정을 판별하므로 GLM의 내포 효과는 교호작용 효과와 동일하게 코드화됩니다. 모형에 포함되지 않은 효과는 포함하는 교호작용(내포와 동등) 효과에 의해 흡수될 수 있습니다.

JMP의 내포 효과는 다르게 코드화됩니다. JMP는 괄호로 묶은 항을 각 그룹에 대한 그룹화 항으로 사용합니다. 내포하는 항의 각 수준 조합에 대해 JMP는 괄호 밖의 효과를 생성합니다. 바깥쪽 항의 수준은 내포하는 항의 전체 수준에서 일치시킬 필요가 없습니다. 각 내포 수준은 설계 열 및 모수 생성과 관련하여 별도로 간주됩니다.

표 A.6 내포 효과

				B(A)					
				A1	A1	A2	A2	A3	A3
A	B	A13	A23	B13	B23	B13	B23	B13	B23
A1	B1	1	0	1	0	0	0	0	0
A1	B2	1	0	0	1	0	0	0	0
A1	B3	1	0	-1	-1	0	0	0	0
A2	B1	0	1	0	0	1	0	0	0
A2	B2	0	1	0	0	0	1	0	0
A2	B3	0	1	0	0	-1	-1	0	0
A3	B1	-1	-1	0	0	0	0	1	0
A3	B2	-1	-1	0	0	0	0	0	1
A3	B3	-1	-1	0	0	0	0	-1	-1

명목형 요인에 대한 최소 제곱 평균

최소 제곱 평균은 다른 모든 요인을 중립 값으로 설정한 후 일부 수준 조합에 해당하는 예측값입니다. 직접 배치된 연속형 회귀변수에 대한 중립 값은 표본 평균으로 정의됩니다. 연관이 없는 명목형 요인이 포함된 효과의 중립 값은 전체 수준에 적용된 평균 효과로 정의됩니다(코딩에서 모두 0이 됨). 순서형 요인은 다른 중립 값을 사용합니다("순서형 최소 제곱 평균"(506페이지) 참조). 최소 제곱 평균을 추정하지 못할 수 있으며, 이에 해당하는 경우 "추정할 수 없음"으로 표시됩니다. JMP의 최소 제곱 평균은 가중치가 사용되는 경우를 제외한 모든 경우에서 SAS PROC GLM의 최소 제곱 평균과 일치합니다(Goodnight and Harvey 1978). 가중치 변수가 사용되면 중립 값에 대해 JMP는 가중 평균을 사용하고 SAS PROC GLM은 비가중 평균을 사용합니다.

유효 가설 검정

일반적으로 JMP에서 생성하는 가설 검정은 SAS PROC GLM(Type III 및 IV 가설) 같은 대부분의 다른 신뢰할 수 있는 프로그램의 가설 검정과 일치합니다. 다음 두 섹션에서는 차이가 있는 경우에 대해 설명합니다.

SAS PROC GLM에서 Type III 및 IV에 대한 가설 검정은 일반 형식의 추정 가능한 함수를 사용하고, 관심 효과 및 관심 효과에 있는 효과만 포함하는 함수를 검색하여 생성됩니다(Goodnight 1978).

동일한 검정이 JMP에서 생성됩니다. 그러나 파라미터화가 다르므로 해당 효과의 모든 모수에 대한 결합 검정을 수행하여 효과를 검정할 수 있습니다(일단 완전 계수 가정). 포함하는 교호작

용 모수는 검정과 관련이 없습니다. 코딩으로 인해 이러한 모수는 포함된 효과에 대한 검정과 관련이 없기 때문입니다.

결측 셀 또는 기타 특이성이 있는 경우 JMP 검정은 GLM 검정과 다릅니다. 아래에서 두 검정의 차이에 대해 설명합니다.

- JMP 검정은 최소한 주효과에 대해 최소 제곱 평균이 다른지를 검정하는 것과 동등합니다. 최소 제곱 평균을 추정할 수 없는 경우에는 검정에 일부 비교를 포함할 수 없으므로 자유도가 손실됩니다. 교호작용의 경우 JMP는 최소 제곱 평균의 차이가 모형의 포함하는 효과로 설명되는 주변 패턴보다 큰지를 검정합니다.
- JMP는 해당 효과가 있는 모형의 SSE를 해당 효과가 없는 모형의 SSE와 비교하여 효과를 검정합니다. JMP는 이 방법이 타당하도록 모형을 파라미터화합니다.
- JMP는 셀 평균 파라미터화가 아니라 구조적 파라미터화를 사용하지만 Hocking(1985, pp. 80–89, 163–166)에서 설명하는 유효 가설 검정을 구현합니다. 유효 가설 검정은 효과에 대해 원하는 가설로 시작하여 "가능한 한 많은" 가설을 포함합니다. 물론 포함하는 효과에 결측 셀이 있으면 완전 가설을 추정할 수 없으므로 이 검정에서 가설의 일부를 제거해야 합니다. 유효 가설은 완전 가설을 최대한 적게 제거합니다.
- JMP와 GLM 및 기타 프로그램에서 결측 셀의 존재와 관련된 가설 검정 간의 차이는 흥미로운 검정으로 간주되지 않습니다. 교호작용이 유의하면 포함된 주효과에 대한 검정은 관심 대상이 아닙니다. 교호작용이 유의하지 않으면 모형에서 해당 교호작용을 항상 제거할 수 있습니다. 일부 검정은 고유하지도 않습니다. 결측 셀 설계의 수준에 라벨을 다시 지정하면 GLM Type IV 검정이 변경될 수 있습니다.

다음 섹션에서는 이 항목에 대한 내용을 더 자세히 설명합니다.

명목형 효과의 특이성 및 결측 셀

설계 열 사이의 선형 종속성을 가정해 보겠습니다. JMP 코딩을 사용하면 데이터가 부족하여 추정이 필요한 조합을 작성할 수 없거나 효과의 교락 또는 공선성 유형이 있는 경우 외에는 선형 종속성 문제가 발생하지 않습니다.

선형 종속성이 있으면 모수에 대한 최소 제곱 해가 고유하지 않을 수 있으며 일부 가설 검정을 수행할 수 없습니다. 설계 열이 모형의 이전 효과에 선형 종속되므로 JMP는 모수 추정값을 순서대로 0으로 설정하는 전략을 선택합니다. 영점화되는 모수 추정값과 추정 불가능한 모수 추정값을 보여주는 특수 열이 보고서에 포함됩니다. 별도의 특이성 보고서에는 선형 종속성이 표시됩니다.

특이성이 있는 경우 JMP에서 검정하는 가설이 GLM에서 선택하는 가설과 다를 수 있습니다. 일반적으로 JMP는 검정을 더 높은 주변화 수준으로 유지하므로 GLM보다 검정 자유도가 적습니다. 즉, JMP 검정은 해당 효과에 대한 전체 최소 제곱 평균의 검정에 해당하지만 GLM 검정은 이 특성이 항상 적용되지는 않습니다.

예를 들어 교호작용이 있는 이원 모형과 하나의 결측 셀을 가정해 봅시다. 여기서 A의 수준은 세 개, B의 수준은 두 개, A3B2는 결측 셀입니다.

표 A.7 교호작용이 있는 이원 모형

A B	A1	A2	B1	A1B1	A2B1
A1 B1	1	0	1	1	0
A2 B1	0	1	1	0	1
A3 B1	-1	-1	1	-1	-1
A1 B2	1	0	-1	-1	0
A2 B2	0	1	-1	0	-1
A3 B2	-1	-1	-1	1	1

이 교호작용이 없는 경우를
가정해 봅니다.

각 셀의 기대값은 다음과 같습니다.

표 A.8 기대값

	B1	B2
A1	$\mu + \alpha_1 + \beta_1 + \alpha\beta_{11}$	$\mu + \alpha_1 - \beta_1 - \alpha\beta_{11}$
A2	$\mu + \alpha_2 + \beta_1 + \alpha\beta_{21}$	$\mu + \alpha_2 - \beta_1 - \alpha\beta_{21}$
A3	$\mu - \alpha_1 - \alpha_2 + \beta_1 - \alpha\beta_{11} - \alpha\beta_{21}$	$\mu - \alpha_1 - \alpha_2 - \beta_1 + \alpha\beta_{11} + \alpha\beta_{21}$

데이터를 포함하는 셀에 추정 가능한 기대값이 있는 것은 분명합니다. 결측 셀에는 추정할 수 없는 기대값이 있습니다. 사실상 기대값은 특이성 보고서에 있는 설계 열의 선형 결합입니다.

$$\mu - \alpha_1 - \alpha_2 - \beta_1 + \alpha\beta_{11} + \alpha\beta_{21}$$

B1 과 B2 의 최소 제곱 평균을 비교하는 검정을 생성하려는 경우를 가정해 보겠습니다. 이 예에서 위의 표에 있는 행 평균은 이러한 최소 제곱 평균을 제공합니다.

$$\begin{aligned} \text{LSM(B1)} &= (1/3)(\mu + \alpha_1 + \beta_1 + \alpha\beta_{11} + \\ &\mu + \alpha_2 + \beta_1 + \alpha\beta_{21} + \\ &\mu - \alpha_1 - \alpha_2 + \beta_1 - \alpha\beta_{11} - \alpha\beta_{21}) \\ &= \mu + \beta_1 \end{aligned}$$

$$\begin{aligned} \text{LSM(B2)} &= (1/3)(\mu + \alpha_1 - \beta_1 - \alpha\beta_{11} + \\ &\mu + \alpha_2 - \beta_1 - \alpha\beta_{21} + \\ &\mu - \alpha_1 - \alpha_2 - \beta_1 + \alpha\beta_{11} + \alpha\beta_{21}) \\ &= \mu - \beta_1 \end{aligned}$$

$$\text{LSM}(B1) - \text{LSM}(B2) = 2\beta_1$$

이것은 β_1 모수에 대한 검정이 최소 제곱 평균이 동일한지 검정하는 것과 동등하다는 것을 보여줍니다. 그러나 β_1 은 추정할 수 없으므로 검정이 불가능하며, 이는 검정에 대한 자유도가 없다는 의미입니다.

이제 A 수준에 대한 최소 제곱 평균 검정을 생성합니다.

$$\begin{aligned}\text{LSM}(A1) &= (1/2)(\mu + \alpha_1 + \beta_1 + \alpha\beta_{11} + \mu + \alpha_1 - \beta_1 - \alpha\beta_{11}) \\ &= \mu + \alpha_1\end{aligned}$$

$$\begin{aligned}\text{LSM}(A2) &= (1/2)(\mu + \alpha_2 + \beta_1 + \alpha\beta_{21} + \mu + \alpha_2 - \beta_1 - \alpha\beta_{21}) \\ &= \mu + \alpha_2\end{aligned}$$

$$\begin{aligned}\text{LSM}(A3) &= (1/2)(\mu - \alpha_1 - \alpha_2 + \beta_1 - \alpha\beta_{11} - \alpha\beta_{21} + \\ &\quad \mu - \alpha_1 - \alpha_2 - \beta_1 + \alpha\beta_{11} + \alpha\beta_{21}) \\ &= \mu - \alpha_1 - \alpha_2\end{aligned}$$

$$\text{LSM}(A1) - \text{LSM}(A3) = 2\alpha_1 + \alpha_2$$

$$\text{LSM}(A2) - \text{LSM}(A3) = 2\alpha_2 + \alpha_1$$

이 중에서 어느 것도 추정할 수 없지만 추정 가능한 다른 비교가 있습니다. 결국 셀이 없는 A 열 두 개를 비교할 수 있습니다.

$$\text{LSM}(A1) - \text{LSM}(A2) = \alpha_1 - \alpha_2$$

이 조합은 효과에 두 개의 모수가 있지만 자유도가 1인 검정을 사용하여 JMP에서 검정됩니다.

내곱과 특이성 조합을 사용하고 이 값이 0 인지 검토하여 추정 가능성을 확인할 수 있습니다.

표 A.9 확인

	특이성	조합
모수	조합	검정
m	1	0
a ₁	-1	1
a ₂	-1	-1
b ₁	-1	0
ab ₁₁	1	0
ab ₂₁	1	0

교호작용에 대한 결측 셀의 설계 열은 항상 주효과(명목형 요인의 경우)의 자유도를 제거하는 것으로 나타났습니다. 따라서 최소 제곱 평균의 추정 불가능과 이러한 최소 제곱 평균에 해당하는 효과를 검정하기 위한 자유도의 손실 사이에는 직접적인 관계가 있습니다.

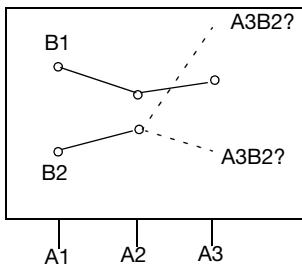
이 내용을 GLM의 방식과 비교해 볼 수 있습니다. 결측 셀이 없을 때는 GLM과 JMP에서 수행하는 검정이 동일합니다. 즉, 두 방법 모두 최소 제곱 평균이 동일한지 효과적으로 검정합니다. 그러나 GLM은 특이성이 발견되면 Type III인지 아니면 Type IV인지에 따라 이러한 셀을 다른 방식으로 처리합니다. Type IV의 경우 찾을 수 있는 추정 가능한 조합을 검색합니다. 이는 고유하지 않을 수 있으며 수준을 제정렬하면 다른 결과를 얻을 수 있습니다. Type III의 경우 고유한 검정을 얻기 위해 추정 가능한 함수의 직교화를 수행합니다. 그러나 셀 평균에 관해 검정을 해석하기가 매우 어려울 수 있습니다.

JMP 방식은 처음에는 GLM을 사용할 때보다 자유도 손실이 더 많아 곤란할 수 있지만 다음과 같은 몇 가지 장점이 있습니다.

1. 검정이 이론적으로 LSM에 연결됩니다.
2. 재파라미터화된 모형의 제곱합 감소를 사용하여 검정을 쉽게 계산할 수 있습니다.
3. 검정이 Hocking의 "유효 가설 검정 (Effective Hypothesis Tests)"과 일치합니다.
4. 검정은 전체 주변 검정입니다. 즉, 검정은 교호작용의 다른 효과를 완벽하게 가로질러 수행하게 됩니다.

마지막으로 몇 가지 세부 사항을 설명합니다. A3B2가 결측 셀인 이전 예에서 셀 평균의 기대값 그래프를 생각해 봅니다.

그림 A.1 셀 평균의 기대값



이 그래프에서는 결측 셀이 있는 셀 평균의 기대값을 보여 줍니다. A1 및 A2 셀의 평균은 B 수준에 대해 프로파일링됩니다. JMP 방식을 사용할 경우 결측 셀의 평균은 교호작용 항에서 허용하는 무엇이든 될 수 있으므로 A3B2 결측 셀이 있으면 B 주효과를 검정할 수 없습니다. 결측 셀의 평균이 표시된 값보다 높으면 B 효과 검정이 유의할 가능성이 큼니다. 값이 더 낮으면 검정이 유의하지 않을 가능성이 큼니다. 요점은 사용자가 모른다는 것입니다. 최소 제곱 평균이 추정 불가능한 것으로 선언되는 것이 이 경우에 해당됩니다. 효과에 대한 가설도 마찬가지로 사용자가 알지 못한다는 것을 가정해야 합니다.

셀 부분집합에 대한 차이와 관련된 가설을 검정하려는 경우 GLM Type IV가 유용합니다. JMP에서는 많은 계산을 통해 효과를 분할하거나 대비를 사용하여 이러한 검정을 직접 생성해야 합니다.

JMP 및 GLM 가설

GLM은 JMP와 다르게 작동하며 결측 셀이 있을 때는 다른 가설 검정을 생성합니다. 특히 GLM은 JMP와 달리 효과에서 내포와 교차의 차이를 인식하지 못합니다. A, B(A), C(A B)라는 3계층의 내포 효과가 있고 지분 설계의 단계에서 수준 수가 다른 경우를 가정해 보겠습니다.

그림 A.10에서는 GLM 모수에 관한 주효과 A의 검정을 보여 줍니다. 첫 번째 열 집합은 JMP에서 수행한 검정입니다. 두 번째 열 집합은 GLM Type IV에서 수행한 검정입니다. 세 번째 열 집합은 JMP의 검정과 동등한 검정이며 처음 두 열에 다음 행렬을 곱한 결과입니다.

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

이는 GLM 검정과 유사합니다. 마지막 열 집합은 GLM Type III 검정입니다. 검정에서 포함하는 효과를 분배하는 방식에 차이가 있습니다. JMP에서는 하향식 계층 구조입니다. GLM Type IV에서는 검정이 상향식입니다. 실제로는 검정 통계량이 유사한 경우가 많습니다.

표 A.10 GLM 및 JMP 가설 비교

모수	A에 대한 JMP 검정		A에 대한 GLM-IV 검정		JMP 회전 검정		A에 대한 GLM-III 검정	
u	0	0	0	0	0	0	0	0
a1	0.6667	-0.3333	1	0	1	0	1	0
a2	-1	0.6667	0	1	0	1	0	1
a3	-1	-0.3333	-1	-1	-1	-1	-1	-1
a1b1	0.1667	-0.0833	0.2222	0	0.25	0	0.2424	0
a1b2	0.1667	-0.0833	0.3333	0	0.25	0	0.2727	0
a1b3	0.1667	-0.0833	0.2222	0	0.25	0	0.2424	0
a1b4	0.1667	-0.0833	0.2222	0	0.25	0	0.2424	0
a2b1	-0.1667	0.3333	0	0.5	0	0.5	0	0.5
a2b2	-0.1667	0.3333	0	0.5	0	0.5	0	0.5
a3b1	-0.1111	-0.1111	-0.3333	-0.3333	-0.3333	-0.3333	-0.3333	-0.3333
a3b2	-0.1111	-0.1111	-0.3333	-0.3333	-0.3333	-0.3333	-0.3333	-0.3333

표 A.10 GLM 및 JMP 가설 비교 (계속)

a3b3	-0.1111	-0.1111	-0.3333	-0.3333	-0.3333	-0.3333	-0.3333	-0.3333
a1b1c1	0.0833	-0.0417	0.1111	0	0.125	0	0.1212	0
a1b1c2	0.0833	-0.0417	0.1111	0	0.125	0	0.1212	0
a1b2c1	0.0556	-0.0278	0.1111	0	0.0833	0	0.0909	0
a1b2c2	0.0556	-0.0278	0.1111	0	0.0833	0	0.0909	0
a1b2c3	0.0556	-0.0278	0.1111	0	0.0833	0	0.0909	0
a1b3c1	0.0833	-0.0417	0.1111	0	0.125	0	0.1212	0
a1b3c2	0.0833	-0.0417	0.1111	0	0.125	0	0.1212	0
a1b4c1	0.0833	-0.0417	0.1111	0	0.125	0	0.1212	0
a1b4c2	0.0833	-0.0417	0.1111	0	0.125	0	0.1212	0
a2b1c1	-0.0833	0.1667	0	0.25	0	0.25	0	0.25
a2b1c2	-0.0833	0.1667	0	0.25	0	0.25	0	0.25
a2b2c1	-0.0833	0.1667	0	0.25	0	0.25	0	0.25
a2b2c2	-0.0833	0.1667	0	0.25	0	0.25	0	0.25
a3b1c1	-0.0556	-0.0556	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667
a3b1c2	-0.0556	-0.0556	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667
a3b2c1	-0.0556	-0.0556	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667
a3b2c2	-0.0556	-0.0556	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667
a3b3c1	-0.0556	-0.0556	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667
a3b3c2	-0.0556	-0.0556	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667	-0.1667

JMP 파라미터화 관점에서 A 에 대한 검정은 다음과 같습니다.

표 A.11 A 에 대한 검정

모수	GLM-IV 검정		JMP 검정	
m	0	0	0	0
a13	2	1	1	0
a23	1	2	0	1
a1:b14	0	0	0	0
a1:b24	0.11111	0	0	0
a1:b34	0	0	0	0
a2:b12	0	0	0	0
a3:b13	0	0	0	0
a3:b23	0	0	0	0
a1b1:c12	0	0	0	0
a1b2:c13	0	0	0	0
a1b2:c23	0	0	0	0
a1b3:c12	0	0	0	0
a1b4:c12	0	0	0	0
a2b1:c13	0	0	0	0
a2b2:c12	0	0	0	0
a3b1:c12	0	0	0	0
a3b2:c12	0	0	0	0
a3b3:c12	0	0	0	0

JMP 관점에서 보면 GLM 검정은 a1b24 모수에 계수를 적용하여 약간 다릅니다.

순서형 요인

순서형 모델링 유형으로 표시된 요인은 명목형 요인과 다르게 코드화됩니다. 모수 추정값이 다르게 해석되고 검정 및 최소 제곱 평균이 다릅니다.

순서형 요인의 첫 번째 수준은 대조 또는 기준 수준이며, 모수는 순서형 요인이 각 후속 수준으로 설정될 때 반응에 미치는 영향을 측정합니다. 순서형 요인 코딩은 첫 번째 용량이 0 일 때 다양한 용량을 나타내는 수준이 포함된 요인에 적합합니다. 다음 표에서는 3 수준 순서형 요인의 예를 보여 줍니다.

표 A.12 순서형 요인

항	코드화된 열		
A	a2	a3	
A1	0	0	대조 수준, 용량 0
A2	1	0	낮은 용량
A3	1	1	높은 용량

설계 패턴은 하위 삼각형 요소는 1이고 다른 요소는 0입니다.

단순 주효과 모형의 경우 다음과 같이 지정할 수 있습니다.

$$y = \mu + \alpha_2 X_{(a \leq 2)} + \alpha_3 X_{(a \leq 3)} + \varepsilon$$

여기서 μ 는 $A=1$ 에서의 기대 반응, $\mu + \alpha_2$ 는 $A=2$ 에서의 기대 반응, $\mu + \alpha_2 + \alpha_3$ 은 $A=3$ 에서의 기대 반응을 나타냅니다. 따라서 α_2 는 $A=1$ 에서 $A=2$ 로 이동 효과를 추정하고, α_3 은 $A=2$ 에서 $A=3$ 으로 이동 효과를 추정합니다.

순서형 주효과에 대한 모든 모수의 부호가 동일하면 순서형 수준에서 반응 효과는 단조 특성을 보입니다.

순서형 교호작용

순서형 교호작용은 명목형 효과와 마찬가지로 요인 열의 수평 직접곱으로 생성됩니다. 두 개의 순서형 요인 A와 B가 있고 각 요인의 수준이 세 개인 경우를 예로 들겠습니다. JMP의 순서형 코딩은 다음에 표시된 설계 행렬을 생성합니다. 교호작용의 패턴은 1로 구성된 하삼각 행렬의 블록 하삼각 행렬입니다.

표 A.13 순서형 교호작용

A	B	A2	A3	B2	B3	A*B			
						A2		A3	
						B2	B3	B2	B3
A1	B1	0	0	0	0	0	0	0	0
A1	B2	0	0	1	0	0	0	0	0
A1	B3	0	0	1	1	0	0	0	0
A2	B1	1	0	0	0	0	0	0	0
A2	B2	1	0	1	0	1	0	0	0
A2	B3	1	0	1	1	1	1	0	0
A3	B1	1	1	0	0	0	0	0	0
A3	B2	1	1	1	0	1	0	1	0
A3	B3	1	1	1	1	1	1	1	1

참고 : 효과가 없는지 여부를 검정할 때 단순 모형에 대한 명목형 요인과 순서형 요인 사이에는 큰 차이가 없습니다. 그러나 교호작용이 지정되면 큰 차이를 보입니다. 대부분의 모형에 순서형 요인보다는 명목형 요인을 사용하는 것이 좋습니다.

순서형 교차 모형에 대한 가설 검정

모수의 의미를 확인하려면 모수에 관한 셀 평균 기대값을 보여 주는 아래 표를 검토합니다. 여기서 μ 는 절편, α_2 는 A2 수준의 모수 등입니다.

표 A.14 셀 평균 기대값

	B1	B2	B3
A1	μ	$\mu + \beta_2$	$\mu + \beta_2 + \beta_3$
A2	$\mu + \alpha_2$	$\mu + \alpha_2 + \beta_2 + \alpha\beta_{22}$	$\mu + \alpha_2 + \beta_2 + \beta_3 + \alpha\beta_{22} + \alpha\beta_{23}$
A3	$\mu + \alpha_2 + \alpha_3$	$\mu + \alpha_2 + \alpha_3 + \beta_2 + \alpha\beta_{22} + \alpha\beta_{32}$	$\mu + \alpha_2 + \alpha_3 + \beta_2 + \beta_3 + \alpha\beta_{22} + \alpha\beta_{23} + \alpha\beta_{32} + \alpha\beta_{33}$

실제로 A에 대한 주효과 검정은 첫 번째 수준에서 B를 포함하는 A 수준을 검정하는 것입니다. 마찬가지로 B에 대한 주효과 검정은 첫 번째 수준에서 A를 포함하는 여러 B 수준의 상단 행을 검정하는 것입니다. 이 검정은 두 요인이 모두 처치 용량이 다른 실험에 적합합니다. 주 관심은 각 처치의 단독 효능이며, 두 약물의 용량을 모두 투여했을 때 **약물 교호작용**을 찾는 것은 부수적입니다. 경우에 따라 각 약물이 다량 투여되면 위험할 수도 있습니다.

각 셀의 기대값은 현재 행과 열을 포함하여 왼쪽 및 위에 있는 각 셀과 관련된 모든 모수를 더하여 얻을 수 있습니다. 마지막 셀의 기대값은 모든 모수의 합입니다.

다른 효과에 포함된 효과에 대한 가설 검정은 순서형 코딩과 명목형 코딩에서 다르지만 다른 효과에 포함되지 않은 효과를 검정하는 것은 동일합니다. 위의 교차 설계에서 교호작용에 대한 검정은 A와 B의 적합이 명목형인지 아니면 순서형인지에 관계없이 동일합니다.

순서형 최소 제곱 평균

앞에서 설명했듯이 최소 제곱 평균은 다른 모든 요인을 중립 값으로 설정한 후 일부 수준 조합에 해당하는 예측값입니다. JMP는 연관이 없는 순서형 요인이 포함된 효과의 중립 값을 첫 번째 수준(대조 기준 수준)으로 정의합니다.

이와 같은 순서형 요인에 대한 최소 제곱 평균 정의는 포함된 효과의 가설 검정이 최소 제곱 평균의 동일 여부 검정과 동등하다는 개념을 유지합니다.

순서형 효과의 특이성 및 결측 셀

순서형 코딩을 사용하면 순서형 효과의 첫 번째 수준이 기준이라는 것을 의미합니다. 따라서 교호작용에 결측 셀이 있거나 교호작용의 데이터가 없는 경우에도 주효과에 대한 적합한 검정을 얻을 수 있습니다.

결측 셀이 있는 예

이 예는 A3B2 셀에 데이터가 없다는 점을 제외하고 셀당 두 개의 관측값이 있는 위의 예와 동일합니다. 이제 요인의 명목형 코드화 결과를 순서형 코드화 결과와 비교할 수 있습니다. 그럼 A.2에서 알 수 있듯이 모형 적합은 동일합니다.

표 A.15 관측값

Y	A	B
12	1	1
14	1	1
15	1	2
16	1	2
17	2	1

표 A.15 관측값 (계속)

17	2	1
18	2	2
19	2	2
20	3	1
24	3	1

그림 A.2 명목형 적합 (왼쪽) 및 순서형 적합 (오른쪽) 에 대한 요약 정보

적합 요약					적합 요약				
R ²		0.891732			R ²		0.891732		
Adj-R ²		0.805118			Adj-R ²		0.805118		
제곱근 평균 제곱 오차		1.48324			제곱근 평균 제곱 오차		1.48324		
반응 평균		17.2			반응 평균		17.2		
관측값 수(또는 가중합)		10			관측값 수(또는 가중합)		10		
분산 분석					분산 분석				
소스	DF	제곱합	평균 제곱	F 비	소스	DF	제곱합	평균 제곱	F 비
모형	4	90.60000	22.6500	10.2955	모형	4	90.60000	22.6500	10.2955
오차	5	11.00000	2.2000	Prob > F	오차	5	11.00000	2.2000	Prob > F
수정 합계	9	101.60000		0.0125*	수정 합계	9	101.60000		0.0125*

코딩이 다르기 때문에 모수 추정값은 매우 차이가 있습니다. 결측 셀은 일부 명목형 모수의 추정 가능성에 영향을 주지만 순서형 모수에는 영향을 주지 않습니다.

그림 A.3 명목형 적합 (왼쪽) 및 순서형 적합 (오른쪽) 에 대한 모수 추정값

모수 추정값					모수 추정값					
항	추정값	표준 오차	t 비	Prob> t	항	추정값	표준 오차	t 비	Prob> t	
절편	17.333333	0.60553	28.63	<.0001*	절편	13	1.048809	12.40	<.0001*	
A[1]	-4.333333	0.856349	-5.06	0.0039*	A[2-1]	4	1.48324	2.70	0.0429*	
A[2]	-0.333333	0.856349	-0.39	0.7131	A[3-2]	5	1.48324	3.37	0.0199*	
B[2-1]	편향	1.5	1.48324	1.01	0.3583	B[2-1]	2.5	1.48324	1.69	0.1527
A[1]*B[2-1]	편향	1	2.097618	0.48	0.6537	A[2-1]*B[2-1]	-1	2.097618	-0.48	0.6537
A[2]*B[2-1]	영점화됨	0	0	.	A[3-2]*B[2-1]	영점화됨	0	0	.	.

특이성 상세 정보는 선형 종속성을 보여 주며 값을 검토하여 결측 셀을 식별합니다.

그림 A.4 명목형 적합 (왼쪽) 및 순서형 적합 (오른쪽) 에 대한 특이성 상세 정보

특이성 상세 정보		특이성 상세 정보	
항	상세 정보	항	상세 정보
B[2-1]	=A[1]*B[2-1] + A[2]*B[2-1]	A[3-2]	B[2-1] = 0

명목형에 대한 효과 검정으로 인해 자유도가 손실됩니다. B의 경우 검정이 없습니다. 순서형의 경우 첫 번째 수준(기준)에 대한 결측 셀이 없으므로 손실이 없습니다.

그림 A.5 명목형 적합 (왼쪽) 및 순서형 적합 (오른쪽)에 대한 효과 검정

효과 검정							효과 검정						
소스	N모수	DF	제곱합	F 비	Prob > F		소스	N모수	DF	제곱합	F 비	Prob > F	
A	2	2	81.333333	18.4848	0.0049*		A	2	2	81.333333	18.4848	0.0049*	
B	1	0	0.000000	.	.	LostDFs	B	1	1	6.250000	2.8409	0.1527	
A*B	2	1	0.500000	0.2273	0.6537	LostDFs	A*B	2	1	0.500000	0.2273	0.6537	LostDFs

최소 제곱 평균도 다릅니다. 명목형 LSM은 모두 추정 가능한 것은 아니지만 순서형 LSM은 가능합니다. 셀 평균을 검토하여 값을 확인할 수 있습니다. 두 경우에 대한 A*B LSM은 동일합니다. 그림 A.6에서는 명목형 적합 및 순서형 적합의 최소 제곱 평균을 보여 줍니다.

그림 A.6 명목형 적합 (왼쪽) 및 순서형 적합 (오른쪽)에 대한 최소 제곱 평균

효과 상세 정보					효과 상세 정보				
A					A				
최소 제곱 평균 테이블					최소 제곱 평균 테이블				
수준	최소 제곱 평균	표준 오차	평균		수준	최소 제곱 평균	표준 오차	평균	
1	13.000000	1.0488088	14.2500		1	13.000000	1.0488088	14.2500	
2	17.000000	1.0488088	17.7500		2	17.000000	1.0488088	17.7500	
3	22.000000	1.0488088	22.0000		3	22.000000	1.0488088	22.0000	
B					B				
최소 제곱 평균 테이블					최소 제곱 평균 테이블				
수준	최소 제곱 평균	표준 오차	평균		수준	최소 제곱 평균	표준 오차	평균	
1	17.333333	0.60553007	17.3333		1	13.000000	1.0488088	17.3333	
2	0.000000	추정할 수 없음	17.0000		2	15.500000	1.0488088	17.0000	
A*B					A*B				
최소 제곱 평균 테이블					최소 제곱 평균 테이블				
수준	최소 제곱 평균	표준 오차			수준	최소 제곱 평균	표준 오차		
1,1	13.000000	1.0488088			1,1	13.000000	1.0488088		
1,2	15.500000	1.0488088			1,2	15.500000	1.0488088		
2,1	17.000000	1.0488088			2,1	17.000000	1.0488088		
2,2	18.500000	1.0488088			2,2	18.500000	1.0488088		
3,1	22.000000	1.0488088			3,1	22.000000	1.0488088		
3,2	0.000000	추정할 수 없음			3,2	0.000000	추정할 수 없음		

빈도

빈도 (정수가 아닌 값 포함)가 분석에 미치는 영향은 손실 함수에 미치는 영향으로 설명됩니다. 반응 값 y_i 와 예측 변수 $x_{i1}, x_{i2}, \dots, x_{in}$ 을 사용하여 θ 모수를 추정하려는 경우를 가정해 보겠습니다. 빈도 변수가 없다고 가정할 때 손실 함수는 다음과 같이 제공됩니다.

$$L(\theta|y) = \sum_{i=1}^n L(\theta|y_i, x_{i1}, x_{i2}, \dots, x_{in})$$

빈도 f_i 가 정의된 경우 손실 함수는 다음과 같습니다.

$$L(\theta|y, f) = \sum_{i=1}^n f_i L(\theta|y_i, x_{i1}, x_{i2}, \dots, x_{in})$$

모수 추정값, 표준 오차, 가설 검정, 신뢰 구간 등 추론의 기준이 되는 모든 통계량은 이러한 형식의 손실 함수를 기반으로 계산됩니다.

일반적인 가정

통계량을 신뢰하려면 사용 기법에 대한 제한 사항과 값에 대해 잘 알고 있어야 합니다. 통계 방법은 도구일 뿐이므로 정확하지 않은 과학(근거 없는 통계적 가정) 또는 잘못된 데이터에 대비하거나 책임지지 않습니다.

가정된 모형

대부분의 통계량은 모형이 올바르다는 가정을 기반으로 합니다. 모형이 올바르지 않게 될 정도로 모형에서 생성되는 통계 보고서의 신뢰도를 줄여야 합니다.

상대적 유의성

대부분의 통계 검정에서는 모형을 절대적 의미로 평가하지 않습니다. 검정 통계량이 유의하면 모형이 평균과 같은 일부 축소 모형보다 더 잘 적합된다는 정도로만 해석할 수 있습니다. 모형이 데이터를 적합시킬 수 있지만 기본 물리적 모형을 완전히 잘 설명하지 못할 수 있습니다.

다중 추론

보통 통계 결과의 값은 그것을 있는 그대로 믿기보다는 어떤 발견의 실마리를 제공합니다. 그 발견을 사실로 확인하기 위해 추가 연구를 수행해야 할 수도 있습니다. 그렇지 않으면 그냥 데이터를 선별하는 것입니다.

예를 들어 충분한 분석을 수행하면 요인에 예측값이 없더라도 연구의 5%에서 5%의 유의한 효과를 발견할 수 있습니다. 마찬가지로 데이터에 대한 올바른 모형을 검정하는 대신 데이터를 사용하여 모형의 형태만 만들면 보고서의 유의 수준이 변질됩니다. 그러면 랜덤 오차가 모형 선택에 영향을 주고 모형이 실제보다 우수하다고 믿게 됩니다.

타당도 평가

다음과 같은 다양한 기법과 패턴을 사용하여 모형의 타당도를 평가할 수 있습니다.

- 적합 결여 검정을 사용하면 포화 상태의 요인과 비교하여 모형 타당도를 확인할 수 있습니다. 비포화 모형에 반복된 x 값이 포함된 경우 모형 적합 플랫폼에서 이러한 검정을 자동으로 제공합니다.
- 모형 적합 플랫폼에서 잔차 그림과 스튜던트화 잔차 그림을 검토하여 연속형 반응에 대한 분포 가정을 확인할 수 있습니다. 아니면 플랫폼 팝업 메뉴에서 **저장** 명령을 사용하여 데이터

테이블 옆에 잔차를 저장합니다. 그런 다음 해당 옆에서 **분석 > 분포**를 사용하여 정규 곡선이 있는 히스토그램 및 정규 분위수 그림을 검토합니다. 잔차가 완전히 독립적이지는 않지만 심각한 비정규 분포를 식별할 수 있습니다.

- 연속형 반응을 위한 최적의 종합 진단 도구는 각 점이 각 가설 검정에 미치는 영향을 보여 주는 레버리지 그림입니다. 데이터에 잘못된 값이 있다고 의심되는 경우 이 그림은 통계 검정이 단일 점의 영향을 심하게 받는지 여부를 판별하는 데 도움이 됩니다.
- 데이터에서 이상치를 스캔하여 올바른 관측값인지 여부를 검토하는 것이 좋습니다. 분포 플랫폼 보고서 및 그림에서 단변량 이상치를 발견할 수 있습니다. 이변량 이상치는 X로 Y적합 산점도 및 다변량 산점도 행렬에 나타납니다. 삼변량 이상치는 **그래프 > 3D 산점도**에서 생성된 3차원 그림에서 확인할 수 있습니다. 더 높은 차원의 이상치는 주성분 또는 3D 산점도를 사용하거나, 다변량 플랫폼에서 계산 및 표시된 Mahalanobis 거리와 잭나이프 거리를 사용하여 찾을 수 있습니다.

대체 방법

통계 문헌에 특수한 비모수 방법 및 로버스트 방법이 설명되어 있지만 현재 JMP에서는 일부만 구현됩니다. 이러한 방법은 분포 가정이 더 적어야 하고(비모수) 오염의 영향을 덜 받습니다(로버스트). 그러나 일반적인 방법론적 접근법에는 도움이 되지 않고 검정 통계량에 대한 표본 확률이 작아 계산에 시간이 많이 걸릴 수 있습니다.

선형 순위 검정에 관심이 있고 정규 대표본 유의성 근사만 필요한 경우 데이터 순위를 분석하여 Wilcoxon 순위합 검정 또는 Kruskal-Wallis 일원 검정과 동등한 작업을 수행할 수 있습니다.

연속형 반응이 정규 가정을 적절하게 충족하는지 확실치 않은 경우, 모델링 유형을 연속형에서 순서형으로 변경하면 풍부한 표현 방식과 통계 검정력을 일부 포기하더라도 안전하게 분석할 수 있습니다.

핵심 통계적 개념

전통적 통계를 통합하고 통계 특성 및 적합 원칙을 시각화 가능한 형식으로 캡슐화하는 두 가지 핵심 개념이 있습니다.

- 불확도 통합 개념
- 두 개의 기본 적합 기계

이 두 가지 아이디어는 수리 통계학의 기초에 근거한 직관적 개념으로 통계를 이해하는 데 도움이 됩니다.

통계학과 과학의 관계는 회계와 비즈니스의 관계를 생각하면 됩니다. 관측한 증거의 중요도를 따져 보고 균형을 맞추기 위한 기술입니다. 통계 검정은 신용 감사와 같습니다. 그러나 통계 도구는 그 이상의 것을 가능하게 합니다. 데이터에 대해 예상치 못한 상황을 보여 주고 흥미로운 새로운 아이디어를 이끌어 낼 수 있는 발견 도구입니다. 이러한 강력한 도구를 사용하려면 먼저 도구의 작동 방식을 이해해야 합니다.

불확도, 통합 개념

회계 작업 시 요약을 위해 자금을 합칩니다. 불확도 또는 잡음이 있는 상태에서 과학적 관측값을 검토하는 경우 데이터 요약을 위해 통계적 측정이 필요합니다. 돈이 가산되는 것처럼 적절한 측도를 선택하면 불확도는 가법성을 나타냅니다.

결합 확률을 얻으려면 관측값이 독립적이라고 가정한 후 확률을 더하는 것이 아니라 곱해야 하므로 직접 확률이 최적의 측도가 아닙니다. 각 확률의 로그를 사용하면 이를 합산할 수 있으며 합계는 결합 확률의 로그이므로 이 방법이 더 쉽습니다.

그러나 확률 로그는 0과 1 사이에 있는 숫자의 로그이므로 음수입니다. 양수를 유지하기 위해 JMP는 확률의 음의 로그를 사용합니다. 확률이 작아질수록 음의 로그가 커집니다. 이 측도를 불확도라고 하고 확률과 반대로 측정됩니다.

비즈니스에서는 수익을 최대화하고 비용을 최소화하려고 합니다. 과학에서는 불확도를 최소화하려고 합니다. 과학의 불확도는 비즈니스의 비용과 같은 역할입니다. 모든 통계 방법은 불확도가 최소화되도록 모형을 적합시킵니다.

불확도를 시각화하는 것은 어렵지 않습니다. 동전 던지기가 각각 독립적일 때 일련의 동전 던지기를 생각해 보십시오. 앞면을 던질 확률은 0.5 이고 밑이 2 인 로그의 $-\log(0.5)$ 는 1 입니다. 연속해서 h 번 앞면을 던질 확률은 다음과 같이 정의됩니다.

$$p = \left(\frac{1}{2}\right)^h$$

h 를 계산하면 다음과 같이 됩니다.

$$h = -\log_2 p$$

어떤 사건의 불확도를 연속으로 "앞면"을 던져야 하는 횟수로 생각하여 동일한 희귀 사건을 가져올 수 있습니다.

통계적으로 수행하는 거의 모든 일에는 핵심에 불확도($-\log p$)가 있습니다. 통계 문헌에서는 불확도를 음의 로그 가능도라고 합니다.

두 개의 기본 적합 기계

통계적 적합에 대한 놀라운 사실은 대부분의 전통적 설계 방법이 스프링과 압력 실린더라는 두 가지 간단한 기계를 사용하는 것으로 축소된다는 것입니다.

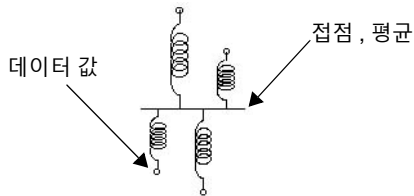
스프링

먼저, 스프링은 연속형 반응 모형에 대한 적합 기계입니다(Farebrother 1987). n 개의 점이 있고 점의 기대값(평균)을 알고 싶다고 가정해 보겠습니다. 점을 배치한 후 스프링으로 공통 접합부에 점을 연결하면 어떻게 될지 예상해 봅니다(그림 A.7). 손을 놓으면 스프링이 점점을 위아래로 움직이다가 중간 지점에서 멈춥니다. 이것은 물리학에 의해 발생하는 현상입니다.

스프링이 부착된 접점을 평균으로 데이터가 정규 분포를 따르는 경우 각 점의 스프링에서 발생하는 물리적 에너지는 데이터 점의 불확도에 비례합니다. 스프링의 에너지, 즉 불확도를 계산하려면 각 점에서 평균까지의 거리에 대한 제곱합을 계산해야 합니다.

관측된 데이터에 대한 불확도를 최소화하는 추정값을 선택하기 위해 스프링 안정점이 평균 추정값으로 선택됩니다. 이 점은 스프링을 늘리기 위해 최소 에너지가 필요한 지점이며 최소 제곱 적합과 동등합니다.

그림 A.7 데이터 점에 스프링 연결

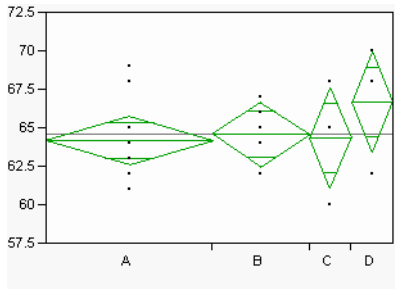


이 방법으로 하나 이상의 평균을 적합시킵니다. 또한 선, 평면 또는 초평면을 적합시킵니다. 이 방법은 연속형 데이터에 대한 거의 모든 모형을 적합시킬 수 있습니다. 스프링을 늘려야 하는 거리의 제곱합으로 에너지 또는 불확도를 측정합니다.

통계학자들은 최소한의 신뢰를 요구하는 정규 분포를 믿습니다. 어떤 의미에서는 가장 랜덤하다고 할 수 있습니다. 정규 분포의 모양은 분포에 대해 가장 무정보적입니다. 즉, 주어진 분산에 대해 예상되는 불확도가 가장 높은 분포입니다. 이 분포의 불확도는 제곱 거리로 측정됩니다. 대부분의 경우 분포가 혼합되어 있거나 독립 통계량의 합이 있을 때 이 분포가 극한 분포입니다. 이 분포를 사용하면 검정 통계량을 상당히 쉽게 측정할 수 있습니다.

가설에 의해 적합이 제약되는 경우 동일한 스프링 에너지를 측정하여 가설을 검정합니다. 실험에서 네 가지 서로 다른 처리의 반응이 있고 평균이 유의하게 다른지 검정하려는 경우를 가정해 보겠습니다. 먼저 그림 A.8에 나오는 것처럼 데이터가 그룹으로 표시되었지만 각 처리에 대한 개별 평균에 스프링이 연결되어 있다고 생각해 봅시다. 그런 다음 스프링력에 압력을 가하여 개별 평균을 공통 평균으로 이동합니다. 결과는 어떻게 될까요? 평균이 같아지도록 누르는 에너지의 양이 바로 필요한 검정 통계량입니다. 해당 에너지는 평균이 같은지 여부를 검정하는 가설에 대한 F -검정의 주요 성분입니다.

그림 A.8 연속형 반응 변수에 대한 일원 그림



압력 실린더

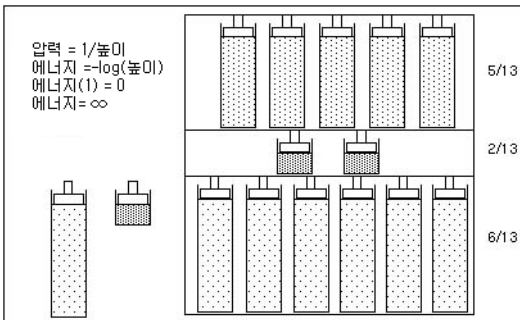
반응이 연속형 대신 범주형이면 어떻게 될까요? 예를 들어 반응이 자동차 표본의 생산 국가인 경우를 가정해 보겠습니다. 표본에 대해 American, European, Japanese 라는 세 가지 반응 수준의 확률이 있습니다. 생산 국가에 대한 이러한 확률을 몇 가지 추정값으로 설정한 후 데이터의 불확도를 평가할 수 있습니다. 이 불확도는 데이터에 의해 제공된 반응 확률의 음의 로그를 합산하여 구합니다. 이 행렬은 다음과 같이 정의됩니다.

$$H = \sum h_{y(i)} = -\sum \log p_{y(i)}$$

스프링 개념은 평균이 연속형 데이터에 어떻게 적합되는지 보여 줍니다. 반응이 범주형인 경우 통계 방법은 반응 확률을 직접 추정하고 데이터의 전체 불확도를 최소화하는 추정값을 선택합니다. 확률 추정값은 음수가 아니고 합이 1이어야 합니다. 반응 확률을 총 길이가 1인 척도의 구성 요소로 생각할 수 있습니다. 각 반응 관측값의 경우, 타이어 펌프와 같은 가스 압력 실린더를 반응 영역에 넣습니다. 최저 위치 에너지의 평형에 도달할 때까지 반응 수준 간의 파티션을 다르게 변경합니다. 그러면 결과 파티션의 크기를 사용하여 반응 확률이 추정됩니다.

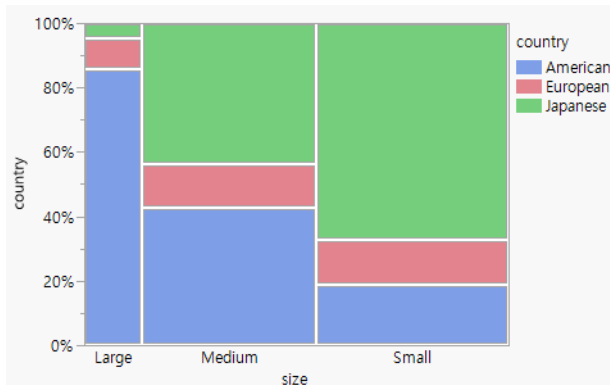
그림 A.9에서는 중형 (medium) 자동차와 같은 단일 범주의 상황을 보여 줍니다 (그림 A.10에 medium 라벨이 지정된 Carpoll.jmp의 모자이크 열 참조). 13개의 반응 (자동차)이 있다고 가정합니다. 첫 번째 수준 (American)의 반응은 6개이고 다음 수준과 마지막 수준의 반응은 각각 2개, 5개입니다. 총 에너지를 최소화하도록 반응 파티션에 대한 압력이 균형을 이루기 때문에 반응 확률은 각각 6/13, 2/13, 5/13가 됩니다.

그림 A.9 파티션에서 압력 실린더의 효과



연속형 데이터에 대한 스프링과 마찬가지로 일부 요인에 따라 표본을 나누고 개별 파티션 집합을 적합시킬 수 있습니다. 그런 다음 파티션을 동일하게 만드는 데 필요한 추가 에너지 양을 측정하여 반응물이 그룹 전체에서 같은지 검정합니다. 자동차 크기별로 그룹화된 자동차 생산 국가 확률에 대한 압력 실린더를 상상해 봅니다. 각 그룹의 파티션을 수평으로 정렬하는 데 필요한 에너지로 변수의 확률이 같은지 여부를 검정합니다. 그림 A.10에서는 이러한 파티션을 보여 줍니다.

그림 A.10 범주형 데이터에 대한 모자이크 그림



가능도, AICc 및 BIC

JMP의 많은 통계 모형은 최대 가능도라는 기법을 사용하여 적합됩니다. 이 기법은 가능도 함수를 최대화하여 모형의 모수 (일반적으로 β 로 표시)를 추정합니다. $L(\beta)$ 로 표시되는 가능도 함수는 관측된 데이터 값에서 평가된 확률 밀도 함수 또는 확률 질량 함수 (이산형 분포의 경우)의 곱입니다. 최대 가능도 추정은 관측된 데이터가 주어지면 모수 β 에 대해 $L(\beta)$ 을 최대화하는 값을 찾습니다.

가능도 함수 $L(\beta)$ 을 최대화하는 것보다 가능도 함수의 음의 자연 로그 $-\log L(\beta)$ 을 사용하는 것이 더 편리합니다. $L(\beta)$ 최대화 문제는 음의 로그 가능도를 최소화하려는 최소화 문제로 다시 작성됩니다 ($-1 \times \log \text{가능도} = -\log L(\beta)$). 따라서 음의 로그 가능도 값이 더 작거나 음의 로그 가능도의 두 배 ($-2 \times \log \text{가능도}$)인 경우 모형이 더 적합하다는 것을 나타냅니다.

음의 로그 가능도 값을 사용하여 모형을 선택하고 JMP의 여러 플랫폼에서 생성된 모형 적합을 비교하는 사용자 가설 검정을 수행할 수 있습니다. 이 작업은 가능도비 검정을 사용하여 수행됩니다. 여러 JMP 플랫폼에서 $-2 \times \log \text{가능도}$ 가 보고되는 한 가지 이유는 전체 모형과 축소 모형 $-2 \times \log \text{가능도}$ 값의 차이가 점근적으로 카이제곱 분포를 따르기 때문입니다. 이 가능도비 검정과 관련된 자유도는 두 모형의 모수 수 차이와 같습니다 (Wilks 1938).

수정된 AICc(Akaike 정보 기준) 및 BIC(Bayes 정보 기준)는 정보를 기반으로 모형 적합을 평가하는 기준입니다. 둘 다 $-2 \times \log \text{가능도}$ 를 기반으로 합니다.

AICc 는 다음과 같이 정의됩니다 .

$$AICc = -2\text{LogLikelihood} + 2k + 2k(k+1)/(n-k-1)$$

여기서 k 는 모형의 추정된 모수 수를 나타내고 n 은 모형에 사용된 관측값 수를 나타냅니다. 이 값은 동일한 데이터 집합에 대한 여러 모형을 비교하여 가장 적합한 모형을 결정하는 데 사용할 수 있습니다. Akaike(1974)에 설명된 대로 값이 가장 작은 모형이 일반적으로 선호됩니다.

BIC 는 다음과 같이 정의됩니다 .

$$BIC = -2\text{LogLikelihood} + k\ln(n)$$

여기서 k 는 모형의 추정된 모수 수를 나타내고 n 은 모형에 사용된 관측값 수를 나타냅니다. 두 모형의 BIC 값을 비교할 때 BIC 값이 작은 모형이 더 나은 것으로 간주됩니다.

일반적으로 BIC는 모수 수가 많은 모형에 AICc보다 벌점을 더 많이 부과합니다. 이러한 이유로 AICc와 비교하여 더 간결한 모형, 즉 모수 수가 더 적은 모형을 선택하게 됩니다. AICc와 BIC의 자세한 비교는 Burnham and Anderson 연구 자료(2004)에서 확인하십시오.

최소 제곱 회귀의 단순화된 AICc 및 BIC 계산식

최소 제곱 회귀의 경우 AICc 및 BIC 를 SSE(오차 제곱합) 에 따라 계산할 수도 있습니다 . SSE 를 기준으로 AICc 및 BIC 는 다음과 같이 정의됩니다 .

$$AICc = n\ln\left(\frac{SSE}{n}\right) + 2k + 2k(k+1)/(n-k-1) + n\ln(2\pi) + n$$

$$BIC = n\ln\left(\frac{SSE}{n}\right) + k\ln(n) + n\ln(2\pi) + n$$

여기서 k 는 모형의 추정된 모수 수, n 은 모형에 사용된 관측값 수, SSE는 모형의 오차 제곱합을 나타냅니다.

검정력 계산

다음 섹션에서는 LSN(최소 유의 개수), LSV(최소 유의값), 검정력 및 조정 검정력을 계산하는 데 사용되는 계산식을 제공합니다. LSV 를 제외하고 각 효과 및 사용자 지정 대비 모음 (" 사용자 지정 " 및 " 최소 제곱 평균 대비 ") 에 대해 이러한 계산이 제공됩니다. LSV 는 단일 선형 대비에 대해서만 계산됩니다. 아래 상세 정보에서 가설은 원하는 대비의 모음을 나타냅니다.

- "LSN 계산"
- "LSV 계산"
- "검정력 계산"
- "조정 검정력 계산"

LSN 계산

LSN(최소 유의 개수) 은 다음 방정식에서 N 값을 구합니다 .

$$\alpha = 1 - FDist \left[\frac{N\delta^2}{df_{Hyp}\sigma^2}, df_{Hyp}, N - df_{Hyp} - 1 \right]$$

다음은 각 요소에 대한 설명입니다 .

$FDist$ - 중심 F 분포의 누적 분포 함수

df_{Hyp} - 가설의 자유도

σ^2 - 오차 분산

δ^2 - 효과 크기 제곱

후향적 분석의 경우 δ^2 은 가설의 제곱합을 n , 즉 현재 표본 크기로 나눈 값으로 추정됩니다. 효과 검정인 경우 δ^2 은 해당 효과의 제곱합을 현재 연구의 관측값 수로 나눈 값으로 추정됩니다. 후향적 연구의 경우 오차 분산 σ^2 은 평균 제곱 오차로 추정됩니다. 이러한 추정값은 α 값 0.05와 함께 "검정력 상세 정보 창"에 기본값으로 입력됩니다.

향후 연구를 계획하기 위해 전향적 분석을 수행할 때 지정된 검정력을 얻을 수 있는 표본 크기를 결정해야 합니다("검정력 계산"(517페이지) 참조).

LSV 계산

LSV(최소 유의값)는 단일 선형 대비에 대해서만 계산됩니다.

단일 선형 대비 검정

자유도가 1 인 $L\beta=0$ 검정을 가정해 봅니다 . 여기서 L 은 상수로 구성된 행 벡터입니다 . 이 가설에 대한 t -검정의 검정 통계량은 다음과 같습니다 .

$$\frac{Lb}{s\sqrt{L(X'X)^{-1}L'}}$$

여기서 s 는 제곱근 평균 제곱 오차입니다. 검정 통계량의 절대값이 자유도가 오차의 자유도와 같은 t 분포의 $1 - \alpha/2$ 분위수 ($t_{1-\alpha/2}$)를 초과하는 경우 유의 수준 α 에서 가설을 기각합니다.

$(Lb)^{LSV}$ 로 표시된 최소 유의값을 찾으려면 Lb 값을 구합니다 .

$$(Lb)^{LSV} = t_{1-\alpha/2} s \sqrt{L(X'X)^{-1}L'}$$

단일 모수 검정

선형 대비가 단일 β_i 를 0 으로 설정하는 가설을 검정하는 특수한 경우에는 계산식이 다음과 같이 축소됩니다.

$$b_i^{LSV} = t_{1-\alpha/2} s \sqrt{(X'X)^{-1}_{ii}} = t_{1-\alpha/2} StdError(b_i)$$

평균의 차이 검정

두 그룹의 평균을 비교하는 검정에 관심이 있는 경우 문헌에서 **LSD(최소 유의차)**에 대한 설명을 참조하십시오. 모형에 명목형 변수 하나만 포함된 특수한 경우 단일 선형 대비 검정을 위한 계산식이 LSD 계산식으로 축소됩니다.

$$LSD = t_{1-\alpha/2} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

그러나 JMP에서 명목형 효과의 수준과 관련된 모수는 해당 수준의 평균과 모든 수준의 평균 간의 차이를 측정합니다. 따라서 이러한 비교에 대한 LSV는 평균 차이에 대한 LSD의 절반입니다.

참고: 명목형 효과 수준에서 대비를 검정하는 경우 JMP가 명목형 효과를 코드화하는 방식에 유의하십시오. 즉, 주어진 수준과 관련된 모수는 모든 수준의 평균에 대한 차이를 측정합니다.

검정력 계산

유의 수준 α 및 표본 크기 N 을 기반으로 선형 가설 검정의 검정력을 계산하는 데 관심이 있다고 가정해 보겠습니다. δ 크기의 효과를 감지하려고 합니다.

검정력을 계산하려면 먼저 선형 가설의 α 수준 F -검정에 대한 임계값을 찾습니다. 이 값은 방정식에서 F_C 값을 구하여 제공됩니다.

$$\alpha = 1 - FDist[F_C, df_{Hyp}, N - df_{Model} - 1]$$

여기서 df_{Hyp} 는 가설의 자유도, df_{Model} 은 모형의 자유도, N 은 제안된 표본 크기 또는 실제 표본 크기를 나타냅니다.

그런 다음 원하는 효과 크기와 관련된 비중심성 모수를 계산합니다. 비중심성 모수는 다음과 같이 정의됩니다.

$$\lambda = (N\delta^2)/\sigma^2$$

여기서 σ^2 은 제안된 또는 추정된 오차 분산 값입니다.

δ 크기의 효과가 주어지면 검정 통계량은 비중심성 모수 λ 가 있는 비중심 F 분포를 따릅니다. 분포 함수는 아래에 $FDist$ 로 표시됩니다. 검정력을 구하려면 검정 통계량이 임계값을 초과할 확률을 계산합니다.

$$\text{검정력} = 1 - FDist\left[F_C, df_{Hyp}, N - df_{Model} - 1, \frac{N\delta^2}{\sigma^2}\right]$$

관측값이 n 개인 연구의 후향적 검정력을 구할 때 JMP는 비중심성 모수 $\lambda = (n\delta^2)/\sigma^2$ 을 $\lambda = SS_{Hyp}/\hat{\sigma}^2$ 으로 추정합니다. 여기서 SS_{Hyp} 는 가설로 인한 제곱합을 나타냅니다.

조정 검정력 계산

조정 검정력 계산(Wright and O'Brien 1988)은 후향적 검정력 분석에만 관련됩니다. 조정 검정력은 모수가 표본 추정값으로 대체될 때 발생하는 양의 편향을 제거하도록 조정된 비중심성 모수 추정값을 사용하여 검정력을 계산합니다.

표본 추정값으로 δ 와 σ 를 추정하여 구한 비중심성 모수 λ 의 추정값은 다음과 같이 계산됩니다.

$$\hat{\lambda} = SS_{Hyp}/MSE$$

Wright and O'Brien(1988)에서는 다음 식을 사용하여 비중심성 모수의 비편향 추정값을 구할 수 있다고 설명합니다.

$$[\hat{\lambda}(df_{Error} - 2)/df_{Error}] - df_{Hyp} = \frac{\hat{\lambda}(N - df_{Model} - 1 - 2)}{N - df_{Model} - 1} - df_{Hyp}$$

오른쪽 표현식에서는 후향적 검정력 분석을 위해 제안된 표본 크기 N 이 연구 크기 n 과 다를 때 비편향 비중심성 모수의 계산을 보여 줍니다. 여기서 df_{Hyp} 는 가설의 자유도를 나타내고 df_{Model} 은 전체 모형의 자유도를 나타냅니다.

이렇게 비중심성 추정값을 조정하면 음수 값이 생성될 수 있습니다. 음수 값은 0으로 설정되고 약간의 편향이 다시 나타납니다. 조정된 비중심성 추정값은 다음과 같습니다.

$$\hat{\lambda}_{adj} = \text{최대값}\left[0, \frac{\hat{\lambda}(N - df_{Model} - 1 - 2)}{N - df_{Model} - 1} - df_{Hyp}\right]$$

조정 검정력은 다음과 같습니다.

$$\text{검정력}_{adj} = 1 - FDist[F_C, df_{Hyp}, N - df_{Model} - 1, \hat{\lambda}_{adj}]$$

비중심성 모수의 신뢰 한계는 Dwass(1955)에 설명된 대로 구성됩니다.

$$\lambda \text{의 신뢰 하한} = \text{최대값}\left[0, \left[\sqrt{SS_{Hyp}/MSE} - \sqrt{df_{Hyp}F_C}\right]^2\right]$$

$$\lambda \text{의 신뢰 상한} = \left[\left(\sqrt{SS_{Hyp}/MSE} \right) - \sqrt{df_{Hyp} F_C} \right]^2$$

검정력의 신뢰 한계는 λ 에 대한 이 신뢰 한계를 다음 방정식에 대입하여 구합니다.

$$\text{검정력} = 1 - FDist[F_C, df_{Hyp}, N - df_{Model} - 1, \lambda]$$

신뢰 한계를 사용한 역추정 예측

역추정 예측은 반응 값에서 독립 변수의 값을 추정합니다. 생물학적 분석 문제의 경우 신뢰 한계를 사용한 역추정 예측이 특히 유용합니다. JMP에서는 연속형 및 이항 반응 모형에 대한 역추정 예측 추정값을 요청할 수 있습니다. 반응이 연속형인 경우 개별 반응 또는 기대 반응에 대한 신뢰 한계를 요청할 수 있습니다.

신뢰 한계는 다음 논리를 기반으로 한 Fieller의 정리 (1954)를 사용하여 계산됩니다. 다른 회귀 변수 및 반응 값이 주어지면 단일 회귀변수의 값과 신뢰 한계를 예측하는 것이 목표입니다.

- $\mathbf{b}$ 가 모수 β 를 추정하여 $\mathbf{b}$ 가 $N(\beta, V)$ 분포를 따르도록 합니다.
- $\mathbf{x}$ 를 관심 있는 회귀변수 값으로 합니다 (i 번째 값 추정).
- $\mathbf{y}$ 를 반응 값으로 합니다.

다른 모든 $\mathbf{x}$ 값이 주어질 때 $\beta'x = y$ 가 되도록 $\mathbf{x}[i]$ 값에 대한 신뢰 영역을 원합니다.

역추정 예측은 다음과 같습니다.

$$\mathbf{x}[i] = \frac{\mathbf{y} - \beta'_{(i)} \mathbf{x}_{(i)}}{\beta[i]}$$

여기서 괄호 안의 서브스크립트 (i)는 i 번째 성분이 생략되었음을 나타냅니다. 다음과 같은 관계에서 신뢰 구간을 구성할 수 있습니다.

$$(\mathbf{y} - \mathbf{b}'\mathbf{x})^2 < t^2 \mathbf{x}'\mathbf{V}\mathbf{x}$$

여기서 t 는 지정된 신뢰 수준에 대한 t 값입니다.

다음 방정식은

$$(\mathbf{y} - \mathbf{b}'\mathbf{x})^2 - t^2 \mathbf{x}'\mathbf{V}\mathbf{x} = 0$$

$\mathbf{z} = \mathbf{x}[i]$ 에 따라 2차로 작성할 수 있습니다.

$$g\mathbf{z}^2 + h\mathbf{z} + \mathbf{f} = 0$$

다음은 각 요소에 대한 설명입니다.

$$\mathbf{g} = \mathbf{b}[i]^2 - t^2 \mathbf{V}[i, i]$$

$$h = -2yb[i] + 2b[i]b'_{(i)}x_{(i)} - 2t^2V[i, (i)]'x_{(i)}$$

$$f = y^2 - 2yb'_{(i)}x_{(i)} + (b'_{(i)}x_{(i)})^2 - t^2x_{(i)}'V_{(i)}x_{(i)}$$

g, h 및 f 값에 따라 부등식을 충족하는 값 집합 및 역추정 예측의 신뢰 구간은 다음과 같은 여러 형식을 가질 수 있습니다.

- $\phi_1 < \phi_2$ 조건에서 (ϕ_1, ϕ_2) 형식의 구간
- $\phi_1 < \phi_2$ 조건에서 $(-\infty, \phi_1) \cup (\phi_2, \infty)$ 형식의 겹치지 않는 두 구간
- 확장된 실직선 $(-\infty, \infty)$
- $(-\infty, \phi)$ 또는 (ϕ, ∞) 중 하나만

Fieller 신뢰 구간이 확장된 실직선인 경우 Wald 구간이 제공됩니다.

참고 : X 로 Y 적합 로지스틱 플랫폼 및 모형 적합 명목형 로지스틱 분석법에서는 역추정 예측을 위한 신뢰 구간을 계산할 때 t 값을 사용합니다 . 모형 적합 일반화 선형 모형 분석법 및 SAS/STAT 의 PROC PROBIT 에서는 서로 다른 결과를 제공하는 z 값을 사용합니다 .

- Aitken, M. (1987). "Modelling Variance Heterogeneity in Normal Regression Using GLIM." *Journal of the Royal Statistical Society, Series C* 36:332–339.
- Akaike, H. (1974). "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control* AC-19:716–723.
- Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley & Sons.
- Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons.
- Benjamini, Y., and Hochberg, Y. (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society, Series B* 57:289–300.
- Box, G. E. P. (1954). "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, Part 2: Effects of Inequality of Variance and of Correlation between Errors in the Two-Way Classification." *Annals of Mathematical Statistics* 25:484–498.
- Box, G. E. P., and Cox, D. R. (1964). "An Analysis of Transformations." *Journal of the Royal Statistical Society, Series B* 26:211–243.
- Box, G. E. P., and Meyer, R. D. (1986). "An Analysis for Unreplicated Fractional Factorials." *Technometrics* 28:11–18.
- Box, G. E. P., and Meyer, R. D. (1993). "Finding the Active Factors in Fractionated Screening Experiments." *Journal of Quality Technology* 25:94–105.
- Burnham, K. P., and Anderson, D. R. (2004). "Multimodel Inference: Understanding AIC and BIC in Model Selection." *Sociological Methods and Research* 33:261–304.
- Burnham, K. P., Andersen, D. R., and Huyvaert, K. P. (2011). "AIC Model Selection and Multimodel Inference in Behavioral Ecology: Some Background, Observations, and Comparisons." *Behavioral Ecology and Sociobiology* 65:23–35.
- Candes, E., and Tao, T. (2007). "The Dantzig Selector: Statistical Estimation when p is Much Larger than n ." *The Annals of Statistics* 35:2313–2351.
- Carroll, R. J., and Ruppert, D. (1988). *Transformation and Weighting in Regression*. London: Chapman & Hall.
- Chilès, J.-P., and Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty*. 2nd ed. New York: John Wiley & Sons.
- Cobb, G. W. (1998). *Introduction to Design and Analysis of Experiments*. New York: Springer-Verlag.

- Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.
- Cole, J. W. L., and Grizzle, J. E. (1966). "Applications of Multivariate Analysis of Variance to Repeated Measures Experiments." *Biometrics* 22:810–828.
- Conover, W. J. (1999). *Practical Nonparametric Statistics*. 3rd ed. New York: John Wiley & Sons.
- Cook, R. D., and Weisberg, S. (1982). *Residuals and Influence in Regression*. New York: Chapman & Hall.
- Cook, R. D., and Weisberg, S. (1983). "Diagnostics for Heteroscedasticity in Regression." *Biometrika* 70:1–10.
- Cornell, J. A. (1990). *Experiments with Mixtures*. 2nd ed. New York: John Wiley & Sons.
- Cox, D. R. (1972). "Regression Models and Life-Tables." *Journal of the Royal Statistical Society, Series B* 34:187–220.
- Cox, D. R., and Snell, E. J. (1989). *The Analysis of Binary Data*. 2nd ed. London: Chapman & Hall.
- Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Rev. ed. New York: John Wiley & Sons.
- Daniel, C. (1959). "Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments." *Technometrics* 1:311–341.
- Dunnnett, C. W. (1955). "A Multiple Comparisons Procedure for Comparing Several Treatments with a Control." *Journal of the American Statistical Association* 50:1096–1121.
- Dwass, M. (1955). "A Note on Simultaneous Confidence Intervals." *Annals of Mathematical Statistics* 26:146–147.
- Efron, B. (1977). "The Efficiency of Cox's Likelihood Function for Censored Data." *Journal of the American Statistical Association* 72:557–565.
- Farebrother, R. W. (1987). "Mechanical Representations of the L1 and L2 Estimation Problems." In *Statistical Data Analysis Based on L_1 Norm and Related Methods*, edited by Y. Dodge, 455–464. Amsterdam: North-Holland.
- Fieller, E. C. (1954). "Some Problems in Interval Estimation." *Journal of the Royal Statistical Society, Series B* 16:175–185.
- Firth, D. (1993). "Bias Reduction of Maximum Likelihood Estimates." *Biometrika* 80:27–38.
- Fleming, T. R., and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: John & Sons.
- Goodnight, J. H. (1978). *Tests of Hypotheses in Fixed Effects Linear Models*. Technical Report R-101, SAS Institute Inc., Cary, NC.
- Goodnight, J. H., and Harvey, W. R. (1978). *Least-Squares Means in the Fixed-Effects General Linear Models*. Technical Report R-103, SAS Institute Inc., Cary, NC.
- Goos, P., and Jones, B. (2011). *Optimal Design of Experiments: A Case Study Approach*. Chichester, UK: John Wiley & Sons.
- Greenhouse, S. W., and Geisser, S. (1959). "On Methods in the Analysis of Profile Data." *Psychometrika* 32:95–112.

- Harrell, F. E. (1986). "The LOGIST Procedure." In *SUGI Supplemental Library Guide, Version 5 Edition*. Cary, NC: SAS Institute Inc.
- Harvey, A. C. (1976). "Estimating Regression Models with Multiplicative Heteroscedasticity." *Econometrica* 44:461–465.
- Hastie, T. J., Tibshirani, R. J., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer-Verlag.
- Hayter, A. J. (1984). "A Proof of the Conjecture That the Tukey-Kramer Method Is Conservative." *Annals of Statistics* 12:61–75.
- Heinze, G., and Schemper, M. (2002). "A Solution to the Problem of Separation in Logistic Regression." *Statistics in Medicine* 21:2409–2419.
- Hocking, R. R. (1985). *The Analysis of Linear Models*. Monterey, CA: Brooks/Cole.
- Hoening, J. M., and Heisey, D. M. (2001). "The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis." *American Statistician* 55:19–24.
- Hoerl, A. (1962). "Application of Ridge Analysis to Regression Problems." *Chemical Engineering Progress* 58:54–59.
- Hoerl, A., and Kennard, R. (1970). "Ridge Regression: Biased Estimation for Non-orthogonal Problems." *Technometrics* 12:55–67.
- Hsu, J. C. (1992). "The Factor Analytic Approach to Simultaneous Inference in the General Linear Model." *Journal of Computational and Graphical Statistics* 1:151–168.
- Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods*. London: Chapman & Hall.
- Hu, W., Laber, E. B., Barker, C., and Stefanski, L. A. (2019). "Assessing Tuning Parameter Selection Variability in Penalized Regression." *Technometrics* 61:154–164.
- Huber, P. J., and Ronchetti, E. M. (2009). *Robust Statistics*. 2nd ed. John Wiley & Sons.
- Hui, F., Warton, D., and Foster, S. (2015). "Tuning Parameter Selection for the Adaptive Lasso Using ERIC." *Journal of the American Statistical Association* 110:262–269.
- Huynh, H., and Feldt, L. S. (1970). "Conditions Under Which Mean Square Ratios in Repeated Measurements Designs Have Exact F-Distributions." *Journal of the American Statistical Association* 65:1582–1589.
- Huynh, H., and Feldt, L. S. (1976). "Estimation of the Box Correction for Degrees of Freedom from Sample Data in the Randomized Block and Split Plot Designs." *Journal of Educational Statistics* 1:69–82.
- Kackar, R. N., and Harville, D. A. (1984). "Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models." *Journal of the American Statistical Association* 79:853–862.
- Kalbfleisch, J. D., and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. 2nd ed. Hoboken, NJ: John Wiley & Sons.
- Kenward, M. G., and Roger, J. H. (1997). "Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood." *Biometrics* 53:983–997.
- Koenker, R., and Hallock, K. (2001). "Quantile Regression: An Introduction." *Journal of Economic Perspectives* 15:143–156.

- Kramer, C. Y. (1956). "Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications." *Biometrics* 12:307–310.
- Lenth, R. V. (1989). "Quick and Easy Analysis of Unreplicated Factorials." *Technometrics* 31:469–473.
- Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. (2006). *SAS for Mixed Models*. 2nd ed. Cary, NC: SAS Institute Inc.
- Mallows, C. L. (1973). "Some Comments on C_p ." *Technometrics* 15:661–675.
- Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press.
- McClave, J. T., and Dietrich, F. H. (1988). *Statistics*. San Francisco: Dellen.
- McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*. 2nd ed. London: Chapman & Hall.
- McCulloch, C. E., Searle, S. R., and Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models*. New York: John Wiley & Sons.
- Meeker, W. Q., and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. New York: John Wiley & Sons.
- Miller, A. J. (1990). *Subset Selection in Regression*. New York: Chapman & Hall.
- Montgomery, D. C. (1991). "Using Fractional Factorial Designs for Robust Process Development." *Quality Engineering* 3:193–205.
- Muller, K. E., and Barton, C. N. (1989). "Approximate Power for Repeated-Measures ANOVA Lacking Sphericity." *Journal of the American Statistical Association* 84:549–555. Also see "Correction to 'Approximate Power for Repeated-Measures ANOVA Lacking Sphericity'," *Journal of the American Statistical Association* 86 (1991): 255–256.
- Nagelkerke, N. J. D. (1991). "A Note on a General Definition of the Coefficient of Determination." *Biometrika* 78:691–692.
- Nelder, J. A., and Wedderburn, R. W. M. (1972). "Generalized Linear Models." *Journal of the Royal Statistical Society, Series A* 135:370–384.
- Nelson, F. D. (1976). "On a General Computer Algorithm for the Analysis of Models with Limited Dependent Variables." *Annals of Economic and Social Measurement* 5:493–509.
- Nelson, P. R., Wludyka, P. S., and Copeland, K. A. F. (2005). *The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions*. Philadelphia: SIAM.
- Patterson, H. D., and Thompson, R. (1974). "Maximum Likelihood Estimation of Components of Variance." In *Proceedings of the Eighth International Biometric Conference*, 197–207. Washington, DC: International Biometric Society.
- Portnoy, S., and Koenker, R. (1997). "The Gaussian Hare and the Laplacian Tortoise: Computation of Squared-Error vs. Absolute-Error Estimators." *Statistical Science* 12:279–300.
- Rawlings, J. O. (1988). *Applied Regression Analysis: A Research Tool*. Pacific Grove, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

- Ries, P. N., and Smith, H. (1963). "The Use of Chi-Square for Preference Testing in Multidimensional Problems." *Chemical Engineering Progress* 59:39–43.
- Sall, J. P. (1990). "Leverage Plots for General Linear Hypotheses." *American Statistician* 44:308–315.
- SAS Institute Inc. (2020a). "The GENMOD Procedure." In *SAS/STAT 15.2 User's Guide*. Cary, NC: SAS Institute Inc.
<https://go.documentation.sas.com/api/docsets/statug/15.2/content/genmod.pdf>.
- SAS Institute Inc. (2020b). "The GLM Procedure." In *SAS/STAT 15.2 User's Guide*. Cary, NC: SAS Institute Inc.
<https://go.documentation.sas.com/api/docsets/statug/15.2/content/glm.pdf>.
- SAS Institute Inc. (2020c). "Introduction to Statistical Modeling with SAS/STAT Software." In *SAS/STAT 15.2 User's Guide*. Cary, NC: SAS Institute Inc.
<https://go.documentation.sas.com/api/docsets/statug/15.2/content/intromod.pdf>.
- SAS Institute Inc. (2020d). "The MIXED Procedure." In *SAS/STAT 15.2 User's Guide*. Cary, NC: SAS Institute Inc.
<https://go.documentation.sas.com/api/docsets/statug/15.2/content/mixed.pdf>.
- Satterthwaite, F. E. (1946). "An Approximate Distribution of Estimates of Variance Components." *Biometrics Bulletin* 2:110–114.
- Scheffé, H. (1958). "Experiments with Mixtures." *Journal of the Royal Statistical Society, Series B* 20:344–360.
- Schuirman, D. J. (1987). "A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability." *Journal of Pharmacokinetics and Biopharmaceutics* 15:657–680.
- Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components*. New York: John Wiley & Sons.
- Seber, G. A. F. (1984). *Multivariate Observations*. New York: John Wiley & Sons.
- Singer, J. D. (1998). "Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models." *Journal of Educational and Behavioral Statistics* 23:323–355.
- Snedecor, G. W., and Cochran, W. G. (1967). *Statistical Methods*. 6th ed. Ames: Iowa State University Press.
- Spiller, S. A., Fitzsimons, G. J., Lynch, J. G., Jr., and McClelland, G. (2013). "Spotlights, Floodlights, and the Magic Number Zero: Simple Effects Tests in Moderated Regression." *Journal of Marketing Research* 50:277–288.
- Stone, C., and Koo, C. Y. (1985). "Additive Splines in Statistics." In *Proceedings of the Statistical Computing Section*, 45–48. Alexandria, VA: American Statistical Association.
- Sullivan, L. M., Dukes, K. A., and Losina, E. (1999). "An Introduction to Hierarchical Linear Modelling." *Statistics in Medicine* 18:855–888.
- Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society, Series B* 58:267–288.

- Tukey, J. W. (1953). "The Problem of Multiple Comparisons." In *Multiple Comparisons*, 1948–1983, edited by H. I. Braun, vol. 8 of *The Collected Works of John W. Tukey* (published 1994), 1–300. London: Chapman & Hall. Unpublished manuscript.
- Walker, S. H., and Duncan, D. B. (1967). "Estimation of the Probability of an Event as a Function of Several Independent Variables." *Biometrika* 54:167–179.
- Westfall, P. H., Tobias, R. D., and Wolfinger, R. D. (2011). *Multiple Comparisons and Multiple Tests Using SAS*. 2nd ed. Cary, NC: SAS Institute Inc.
- Wilks, S. S. (1938). "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses." *Annals of Mathematical Statistics*. 9:60–62.
- Wolfinger, R. D., Tobias, R. D., and Sall, J. (1994). "Computing Gaussian Likelihoods and Their Derivatives for General Linear Mixed Models." *SIAM Journal on Scientific Computing* 15:1294–1310.
- Wright, S. P., and O'Brien, R. G. (1988). "Power Analysis in an Enhanced GLM Procedure: What it Might Look Like." In *Proceedings of the Thirteenth Annual SAS Users Group International Conference*, 1097–1102. Cary, NC: SAS Institute Inc.
<http://www.sascommunity.org/sugi/SUGI88/Sugi-13-220%20Wright%20OBrien.pdf>.
- Zou, H. (2006). "The Adaptive Lasso and Its Oracle Properties." *Journal of the American Statistical Association* 101:1418–1429.
- Zou, H., and Hastie, T. (2005). "Regularization and Variable Selection via the Elastic Net." *Journal of the Royal Statistical Society, Series B* 67:301–320.

기술 라이선스 표시

- Scintilla is Copyright © 1998–2017 by Neil Hodgson <neilh@scintilla.org>.

All Rights Reserved.

Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

- Progress® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is Copyright © 2009–2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES

(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software is Copyright © 1995–2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- This program, “bzip2”, the associated library “libbzip2”, and all documentation, are Copyright © 1996–2019 Julian R Seward. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
2. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.
3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- R software is Copyright © 1999–2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
- Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF

USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

Copyright © 1998–2003 Daniel Veillard. All Rights Reserved.

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

Copyright © 1985, 1986, 1992, 1993

The Regents of the University of California. All rights reserved.

THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball is Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako is Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities are Copyright 2006–2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS “AS IS” WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at

<http://www.apache.org/licenses/LICENSE-2.0>

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
3. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

- You must give any other recipients of the Work or Derivative Works a copy of this License; and
 - You must cause any modified files to carry prominent notices stating that You changed the files; and
 - You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
 - If the Work includes a “NOTICE” text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
 - You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.
- LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign. Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at: <http://www.apache.org/licenses/LICENSE-2.0>
Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.
 - clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign. Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at: <http://www.apache.org/licenses/LICENSE-2.0>
Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS”, WITHOUT WARRANTIES OR CONDITIONS OF

ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- lld is Copyright © 2011–2019 by the University of Illinois at Urbana-Champaign.

Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

<http://www.apache.org/licenses/LICENSE-2.0>

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS”, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libcurl is Copyright © 1996–2020, Daniel Stenberg, daniel@haxx.se, and many contributors, see the THANKS file. All rights reserved.

Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

