

Monthly User Guide from JMP Korea

제 22호 (2019년 5월)

DOE in JMP(1)

* 본 Guide 는 매월 두 번째 화요일에 발행됩니다

(2018년 7월호부터는 JMP 14 Version 기준입니다)

** Monthly User Guide 지난 호는 다음 Site 를 참조하세요(https://www.jmp.com/ko_kr/newsletters.html)

*** 본 Guide 의 내용과 관련한 문의는 ikju.Shin@jmp.com 으로 연락 바랍니다



DOE in JMP

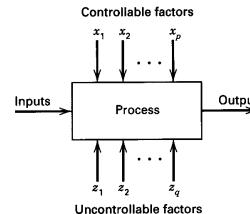
실험 계획법(DOE : Design of Experiment)은 실험(Design)의 두 가지 목적을 달성하기 위하여 최소의 실험 회수로부터 최대의 정보를 얻을 수 있는가 등에 대한 일련의 과정을 계획하는 것이라 할 수 있습니다.
이번 호부터 몇 개월 동안은 JMP의 최고의 장점중의 하나인 DOE에 대해 학습해 보겠습니다

실험의 목적

1. 주요 인자 선별
: Screening, Vital Few Xs의 선정
2. 최적 조건 선정
: Optimization, Modeling

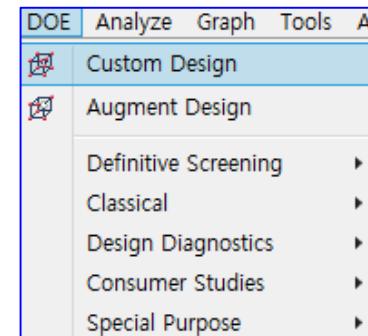
DOE의 핵심 내용

1. 실험의 설계(Design)



DOE in JMP

DOE Platform



2. 실험 결과의 분석(Analyze)

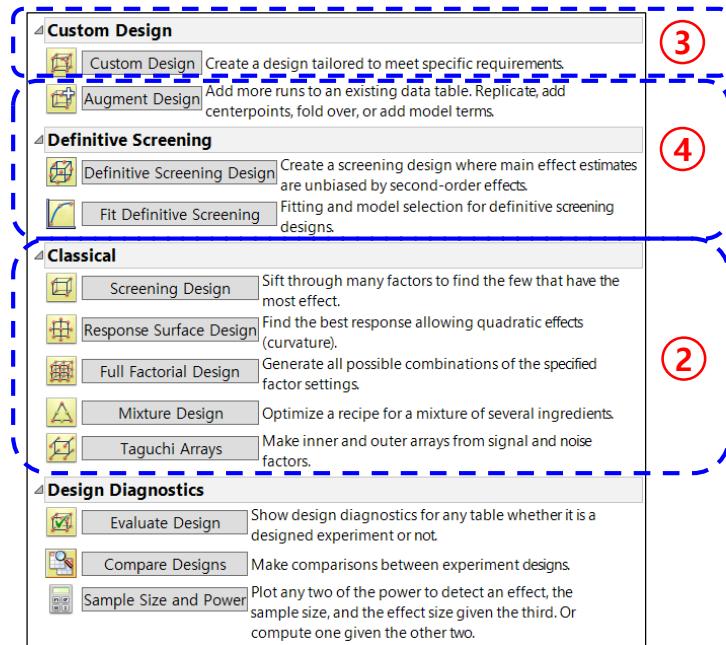
Analyze / Fit Model

DOE in JMP

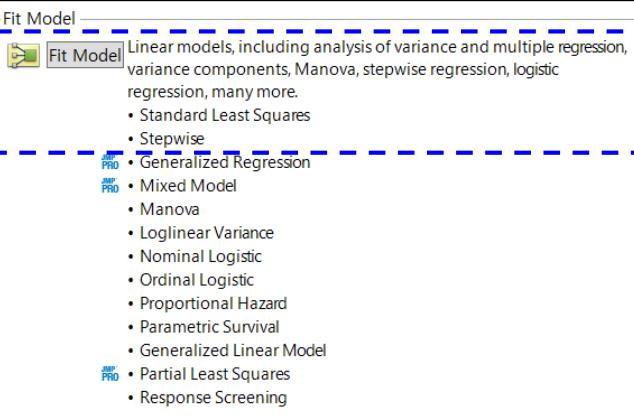
실험 계획법(DOE : Design of Experiment)의 주요 내용은 **실험의 설계(Design)**와 **분석(Analyze)**입니다.
이번 호부터 총 4회로 나누어 배워보도록 하겠습니다. 이번 호에서는 Fit Model의 기본적인 절차, 몇 가지 Option 및 두 가지 사례를 배워보도록 하겠습니다.

View / JMP Starter

실험의 설계(Design) : DOE Platform



결과의 분석(Analyze) : Analyze / Fit Model



학습 내용

1. 기본적인 절차(다중 회귀 분석)
2. Fit Model의 몇 가지 Option
3. Fit Model 사례

기본적인 절차(다중 회귀 분석)

변수들 간의 관련성을 규명하기 위하여 어떤 수학적 모델을 가정하고, 이 모델을 측정된 데이터로부터 추정하는 통계적 분석 방법을 회귀(Regression) 분석이라 말합니다.

- 1) 회귀 분석 중에서 하나의 X 인자와 하나의 Y 인자 간의 관련성을 파악하는 것을 단순 회귀 또는 이변량(Bivariate) 회귀 분석이라 하고,
- 2) X 인자가 두 개 이상인 경우에는 다중(Multiple) 또는 다변량(Multivariate) 회귀 분석이라 부름.
- 3) JMP에서는 일반적으로, X 인자가 하나인 경우에는 Analyze / Fit Y by X platform에서 회귀 분석을 수행하고, X 인자가 두 개 이상일 경우에는 Analyze / Fit Model Platform에서 회귀 분석을 한다.
- 4) 일반적으로, 실험의 경우에는 X인자가 두 가지 이상이고, 인자 간의 교호 작용(Interaction)을 고려하므로 Analyze / Fit Model Platform을 이용하여 그 결과를 분석한다.

Sample Data : Help / Sample Data Library / fitness.jmp

1. Analyze / Fit Model에서 Oxy 변수를 Y로 선택하고, Name을 제외한 6개 변수를 X인자로 선택한다(6개 변수를 선택한 다음 'Construct Model Effects'에서 Add를 클릭), 그런 다음, 다른 Option은 그대로 두고 Run을 클릭하면, 우측과 같은 결과가 생성된다.

- 1) Summary Fit의 RSquare 와 RSquare Adj의 값이 0.8 보다 크고, ANOVA 테이블에서도 Prob > F값이 매우 작으므로 도출된 회귀 모델은 유의하다고 말할 수 있다.
- 2) Parameter Estimates의 각 요인들의 P-Value를 보면 유의한 인자도 있고, 유의하지 않는 인자도 있음을 알 수 있다.

Summary of Fit				
RSquare		0.850676		
RSquare Adj		0.80523		
Root Mean Square Error		2.351053		
Mean of Response		47.37581		
Observations (or Sum Wgts)		31		
Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	724.25014	103.464	18.7183
Error	23	127.13140	5.527	Prob > F
C. Total	30	851.38154		<.0001*

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	97.294858	14.04425	6.93	<.0001*
Sex[F]	0.424729	0.598153	0.71	0.4848
Age	-0.20234	0.102168	-1.98	0.0597
Weight	-0.04492	0.069756	-0.64	0.5260
Runtime	-2.579015	0.398721	-6.47	<.0001*
RunPulse	-0.358804	0.122297	-2.93	0.0075*
RstPulse	-0.009308	0.068437	-0.14	0.8930
MaxPulse	0.2984435	0.138821	2.15	0.0423*

- 3) 최종적으로 회귀 모델을 완성하기 위해서는 유의하지 않는 인자를 제거해야 한다. 제거하는 방법은 아래 Effect Summary에서 유의하지 않는 (P Value가 큰) 순서대로 하나씩 제거해 나가야 한다.

Effect Summary		
Source	LogWorth	PValue
Runtime	5.874	0.00000
RunPulse	2.127	0.00746
MaxPulse	1.373	0.04232
Age	1.224	0.05973
Sex	0.314	0.48480
Weight	0.279	0.52598
RstPulse	0.049	0.89300

[Remove](#) [Add](#) [Edit](#) FDR

기본적인 절차(다중 회귀 분석)

4) 유의하지 않는 인자를 모델에서 제거한 후의 결과는 아래와 같다.

Effect Summary

Source	LogWorth	PValue
Runtime	7.879	0.00000
RunPulse	2.147	0.00713
MaxPulse	1.269	0.05381
Age	1.256	0.05552

2. 유의한 인자만을 대상으로 회귀 모형을 만들어 예측하기 위해서는

▼ Response Oxy / Factor Profiling / Profiler 를 실행해야 한다.

X 인자의 값을 조정해 보면(붉은 색 라인을 옮겨보면)

그에 따른 Y값을 추정할 수 있다

Prediction Profiler



3. 최종적인 모델의 식을 보고자 한다면 ▼Response Oxy / Estimates / show prediction formula에서 확인할 수 있다.

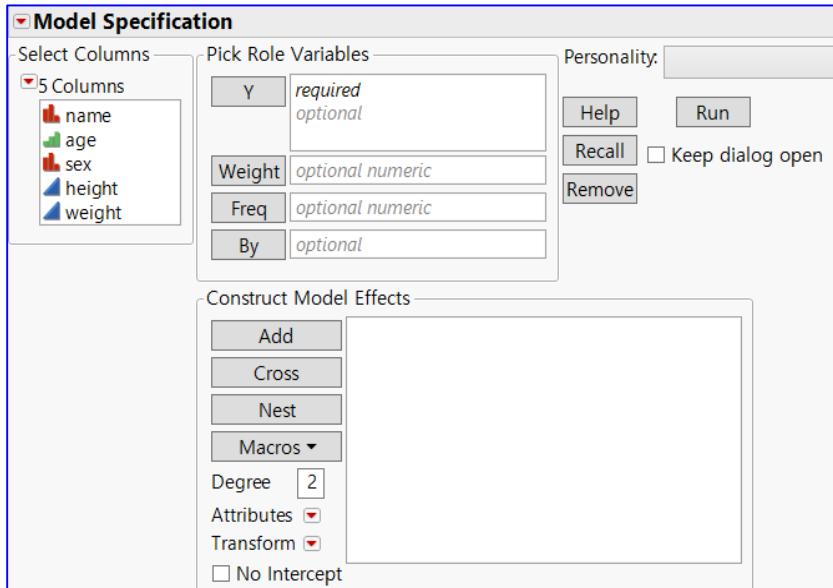
Prediction Expression

```
97.185201625  
+ -0.189218389 •Age  
+ -2.775605587 •Runtime  
+ -0.345271669 •RunPulse  
+ 0.2714363697 •MaxPulse
```

Fit Model의 몇 가지 Option

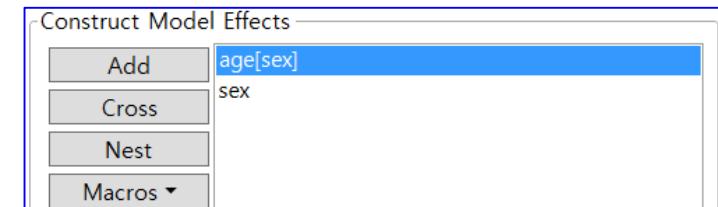
Sample Data : Help / Sample Data Library / big class.jmp

Fit Model 의 Launching Platform에는 다양한 Option이 있는 데,
이 중 중요한 것 몇 가지를 살펴 보도록 한다



1) **Construct Model Effects** : 모델안에 어떤 효과(주 효과, 교호 작용 등)를 포함할 것인지를 결정하는 Option이다.

- 1) Add : 좌측의 Select Columns에 있는 효과를 추가(Add)하는 기능이다.
변수에 대해 사전에 변환 등을 하지 않았다면, Select Columns에는 보통 해당 변수만 있으므로, 주효과만을 모델에 포함한다는 뜻이다
- 2) Cross : 특정한 교호 작용을 모델에 반영할 때 활용된다.
예를 들어 네 개의 변수를 선택하고 Cross를 클릭하면
4인자 교호작용이 모델에 반영된다
- 3) Nest : 내포된(Nested) 효과를 반영할 때 활용된다.
만약 나이(Age) 변수가 특정한 성별(Sex)에서만 효과를 가진다면
나이 변수가 성별 변수에 내포되었다고 말하고,
입력 방법은 두 변수를 모두 Construct Model Effects에서 Add 한 다음,
Construct Model Effects에서 나이(age)을 선택하고,
Select Columns에서 성별(sex)을 선택한 뒤 Nest 를 클릭하면 된다
그 결과는 아래와 같다. Nest 기능은 범주형 범주만이 고려될 수 있다



Fit Model의 몇 가지 Option

4) Macros ; 일반적으로 많이 사용되는 모델링 방법을 모아 놓은 것이다

-**Full Factorial** : 주 효과와 모든 교호 작용을 모델에 포함한다

-**Factorial to Degree** : 주 효과와 하단 Degree의 숫자만큼의 교호 작용을 포함한다.

-**Factorial Sorted** : Full Factorial과 동일하나 인자의 배열 순서만 다르다. 모든 주 효과가 먼저 배열되고, 그 다음 2차의 교호 작용, 3차의 교호 작용 순으로 배열된다.

-**Response Surface** : 주 효과, 2차 교호 작용 및 2차(quadratic) 함수 효과를 포함한다

-**Mixture Response Surface** : 혼합물에 대한 Response Surface 모형이라 할 수 있다

-**Polynomial to Degree** : Degree 숫자만큼 고차의 효과를 모델에 반영한다. 네 개의 인자를 선택하고, Degree 의 숫자가 3 이라면, 네 개의 인자에 대해 3차 함수까지의 모형을 포함한다는 뜻이다.

-**Scheffe Cubic** : 혼합물 실험에서 활용되는 모형이다

5) Attributes : Random Effect 등의 속성을 반영하는 Option 이다

6) Transform : 선택된 Y 또는 Construct Model Effects 에 반영된 인자에 대해 변환을 할 수 있는 Option 이다

2. **Fitting Personality** : 적절한 모델링 방법을 결정하는 Option 이다.
간략히 설명하면 우측과 같다.

Personality Option	설명
Standard Least Square	최소 제곱법에 의한 모델 추정
Stepwise	단계적 회귀 분석을 통한 변수 선별
MANOVA	다변량 분산 분석
Loglinear Variance	데이터의 분산을 반응 변수로 하여 모델 추정
Nominal Logistic	명목형 반응 변수에 대한 로지스틱 회귀 분석
Ordinal Logistic	서열형 반응 변수에 대한 로지스틱 회귀 분석
Proportional Hazard	신뢰성, 생존 분석을 위한 모델
Parametric Survival	신뢰성, 생존 분석에서의 모수적 수명 추정 방법
Generalized Linear Model	반응 변수가 정규성, 잔차의 등분산성 등을 만족하지 않을 때
Response Screening	반응 변수가 여러 개인 경우의 회귀 분석

Fit Model 사례(1) : 최소 제곱법

Sample Data : Help / Sample Data Library / design experiment / Custom RSM.jmp

Y 값에 영향을 주는 3개 Factor에 대하여 실험한 결과에 대해

- Personality는 Standard Least Squares로,

- Model Effects 는 Response Surface로 하여

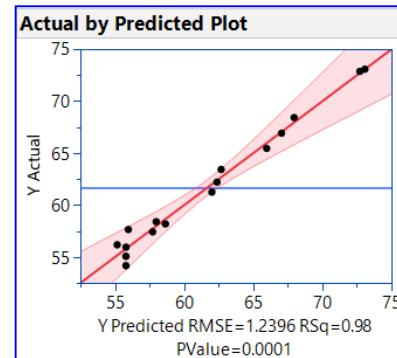
3개 Factor의 Y에 대한 영향도를 파악하고 Y를 최적화하는 X의 세부 조건을 확인하고자 한다.

(이 예제 파일은 JMP의 Custom Design 기능을 이용하여 미리 실험 설계한 것으로, Model Specification에 인자 및 Model Effect 등이 이미 입력되어 있으나, 입력되어 있지 않은 것으로 가정하여 설명한다)

1. Analyze / Fit Model에서 Y 변수를 Y로 선택하고, 나머지 3개 X 인자를 선택하여 'Construct Model Effects'의 Macros 클릭 후 Response Surface 를 선택한다. 그런 다음 Personality는 Standard Least Square로, Emphasis에서는 Effect Screening, Run을 클릭한다.

2. 분석 결과를 하나씩 살펴보면

- 1) Actual by Predicted Plot 을 보았을 때 이 모델이 통계적으로 유의함을 알 수 있다



2) Effect Summary를 보면 유의하지 않는 변수들이 있으므로 이를 제거하여야 한다. 모델에서 이를 제거하고자 할 경우에는 반드시 가장 유의하지 않는 효과부터 하나씩 제거해야 한다.

Effect Summary		
Source	LogWorth	PValue
X2	4.848	0.00001
X2*X2	3.298	0.00050
X1	3.013	0.00097
X1*X1	2.994	0.00101
X3	0.352	0.44415
X1*X2	0.336	0.46112
X3*X3	0.158	0.69549
X1*X3	0.117	0.76434
X2*X3	0.002	0.99563

유의하지 않은 효과를 제거한 후의 결과는 아래와 같다.
X1, X2 두 변수의 주효과와 제곱 효과가 모델에 포함되었음을 알 수 있다.

Effect Summary		
Source	LogWorth	PValue
X2	8.063	0.00000
X2*X2	5.419	0.00000
X1*X1	4.864	0.00001
X1	4.773	0.00002 ^

Fit Model 사례(1) : 최소 제곱법

- 3) 최종적인 모델(함수)를 확인하고자 한다면
Estimates / Show Prediction Expression에서 확인할 수 있다.

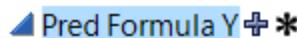
Prediction Expression

```
55.721136364  
+ 2.349 • X1  
+ 5.003 • X2  
+ X1 • (X1 • 4.4365909091 )  
+ X2 • (X2 • 5.0765909091 )
```

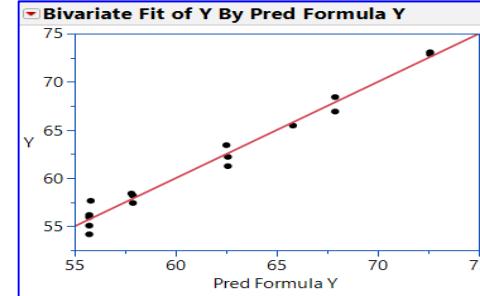
- 4) Save Columns / Prediction Formula 을 활용하여
실제의 예측 값을 Data Table에 저장할 수도 있다.

	X1	X2	X3	Y	Pred Formula Y
1	-1	-1	-1	57.42	57.882318182
2	0	0	0	55.07	55.721136364
3	0	-1	0	57.65	55.794727273
4	-1	0	0	58.40	57.808727273
5	1	1	-1	72.87	72.586318182

또한, 좌측 Column Panel에서 'Pred Formula Y' 우측의 아이콘을 클릭해서 Formula를 확인할 수도 있다

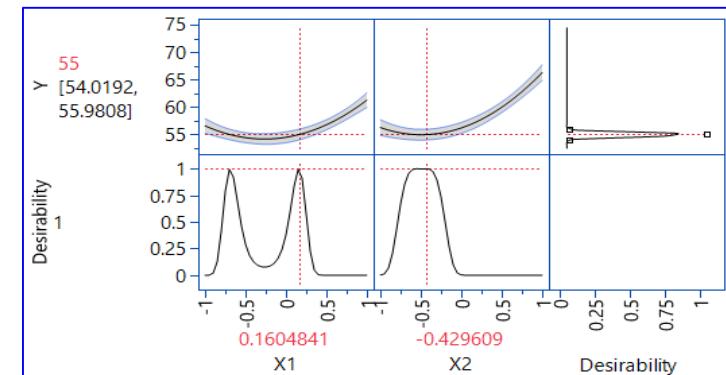


- 5) Y 와 Pred Formula Y 변수를 가지고 산점도를 그린 것이
Actual by Predicted Plot 이다.



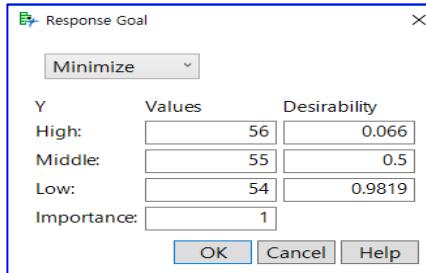
3. Prediction Profiler를 이용한 최적화

- 1) 현재의 조건에서 최적화를 하고자 한다면 ▼Prediction Profiler / Desirability Function 및 Maximize Desirability를 클릭하면 된다.
아래의 결과가 최적 조건이다.

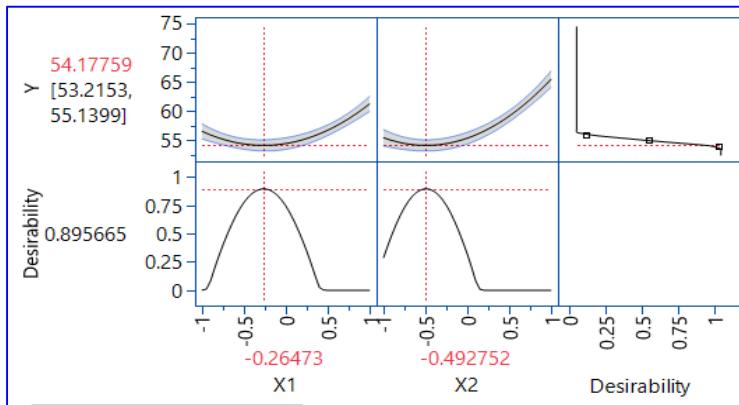


Fit Model 사례(1) : 최소 제곱법

2) 본 예제에서 Y 변수는 그 특성이 Match Target으로 설정되어 있는 데, 만약 Minimize로 변경하고자 하다면 Profiler 그래프의 맨 오른쪽 위에서 더블 클릭하거나 Ctrl 키를 누른 상태에서 클릭하면, 아래와 같은 화면이 Display된다.



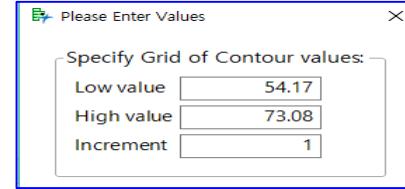
3) Minimize로 변경한 다음, 3-1) 번 과정을 다시 실행하면 된다



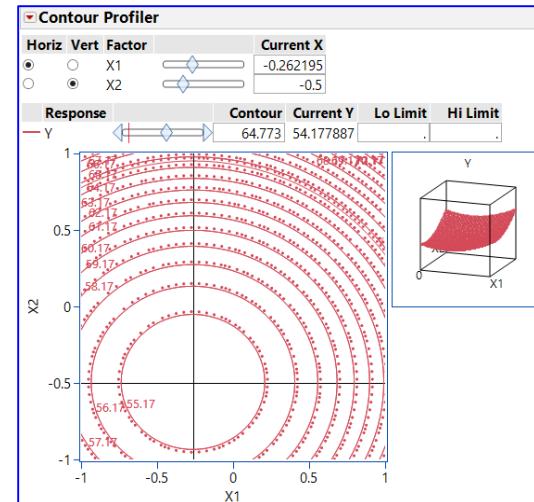
4. Contour Tour를 활용할 수도 있는데, 이를 위해서는

▼ Response Y / Factor Profiling / Contour Profiler을 활용한다

1) 등고선의 간격을 조정하기 위해서는 ▼Contour Profiler / Contour Grid에서 Increment를 조정한다(여기서는 1)



2) 아래와 같은 Contour Profiler가 생성된다. Contour Plot의 Option 기능을 이용하여 X인자 값의 변화에 따른 Y값을 추정할 수 있다

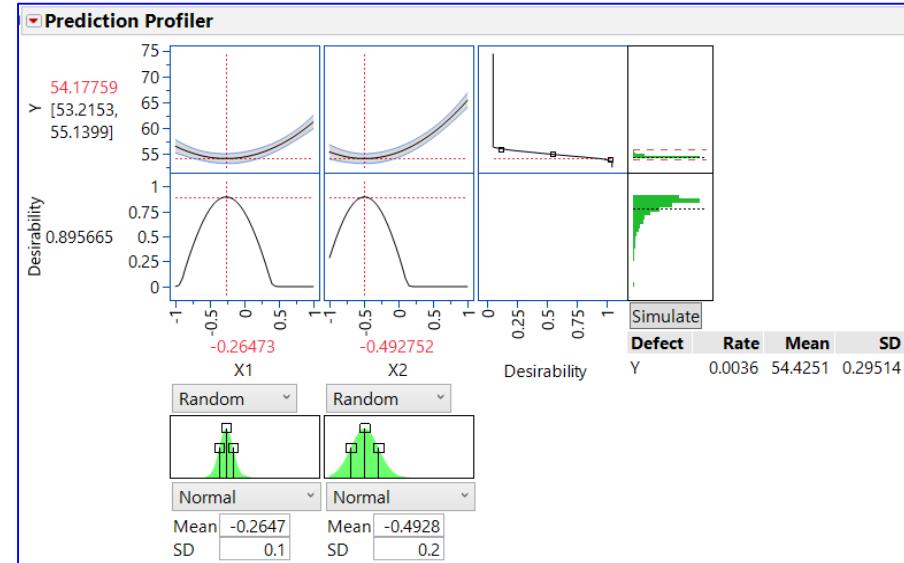


Fit Model 사례(1) : 최소 제곱법

5. Simulation 기능을 활용하여 모델에 포함된 Factor의 산포 (정확하는 Random Variation) 등을 감안하여 결과를 추정할 수 있다.

만약 위의 3-2) 결과에서 Y의 Spec이 54 ~ 56이고, X1과 X2가 설정된 최적 조건에서 각각 표준편차 0.1과 0.2 만큼의 산포를 가지는 상황에서 Y 값을 추정해 보고자 한다면

- 1) ▶ Prediction Profiler / Simulator 클릭, 변수 명의 하단에서 Random 선택 후 X1 및 X2 Factor에 대해 표준편차 0.1 및 0.2 입력
- 2) 하단 ▶ Simulator / Spec Limit에서 Spec 입력 후 Save 클릭
- 3) Profiler Graph 우측 하단에서 Simulator 버튼을 클릭하면 우측과 같이 Y값에 대해 그 결과를 추정할 수 있다.



Fit Model 사례(2) : Stepwise(단계별 회귀 분석)

Sample Data : Help / Sample Data Library / fitness.jmp

회귀 분석 방법으로 많이 활용되는 방법중의 하나로 단계별 회귀 분석(Stepwise)이라는 방법이 있다.

Stepwise는 회귀 모델을 구축하기 위해서 한꺼번에 모든 인자를 선택하는 것이 아니라 단계적으로 접근하기 때문에 붙여진 이름이다.

세부적으로 Forward(전진 선택), Backward(후방 제거), Mixed(혼합법)의 세 가지 방법이 있다.

A, B, C, D 네 개의 인자가 있다고 가정하면

1) Forward는 $A \rightarrow A+B \rightarrow A+B+C$ 의 순서처럼 영향이 큰 인자 순으로 하나씩 추가하는 방법이고,

2) Backward 는 $A+B+C+D \rightarrow A+B+C \rightarrow A+B$ 의 순서처럼 처음에 모든 인자를 선택한 다음 영향이 작은 인자부터 하나씩 제거하는 방법이다. 3) 반면에 Mixed 방법은 $A \rightarrow A+B \rightarrow A+B+C \rightarrow B+C$ 처럼

Forward와 Backward 가 혼합된 방법이다.

예제를 통해 살펴보자.

1. Analyze / Fit Model 에서 Oxy 변수를 Y로 선택하고, Runtime, Weight, Runpulse, RstPulse, MaxPluse를 X인자로 선택한다. (5개 변수를 선택한 다음 'Construct Model Effects'에서 Add를 클릭), 그런 다음 Personality에서 Stepwise를 선택한 후 Run 을 클릭하면, 우측과 같은 결과가 생성된다.

The screenshot shows the 'Stepwise Fit for Oxy' dialog box. Under 'Stepwise Regression Control', the Stopping Rule is set to 'Minimum BIC'. The Direction is 'Forward'. Buttons for 'Enter All' and 'Make Model' are available. Below are buttons for 'Go', 'Stop', and 'Step'. A table of 'Current Estimates' lists variables Intercept, Runtime, Weight, RunPulse, RstPulse, and MaxPulse with their respective estimates, standard errors, and p-values. The 'Step History' section is also visible.

Stopping Rule에 대해서 살펴보면 Stopping Rule은 변수 선별에 사용하는 평가지표로서 P-Value threshold, Minimum AIC, Minimum BIC가 있다.

1) P-Value threshold는 P값과 유의 수준(α)을 비교하여 모델을 추정하는 방법이다. 이를 선택했을 경우 입력해야 하는 Prob to Enter는 Forward 방법에서 모델에 포함시키는 최대의 P Value를 말하고, Prob to Leave는 Backward 방법에서 모델에서 제거하는 최소의 P Value를 말한다.

The screenshot shows a 'P-value Threshold' dialog box with two fields: 'Prob to Enter' set to 0.25 and 'Prob to Leave' set to 0.1.

Fit Model 사례(2) : Stepwise(단계별 회귀 분석)

2) Minimum AIC 및 Minimum BIC는 각각 그 값이 가장 작아질 때까지 변수를 추가 또는 제거하는 방법이다. AIC(Akaike Information Criterion)와 BIC(Baysian Information Criterion)은 회귀 모형의 적합성을 비교하기 위해 사용되는 지표로써 둘 모두 그 값이 적을수록 좋은 모형이라 판단한다.

3) P-Value threshold, Minimum AICc, Minimum BICc 세 가지 모두 Forward 및 Backward 방법에서 활용 가능하나, Mixed 방법에서는 P-Value threshold만 사용 가능하다.

2. Stopping Rule에서 Minimum BIC를 선택하면 아래와 같은 결과가 생성된다. 다섯 개의 Factor 중 세 개가 중요한 Factor로 선정되었음을 알 수 있다.

Stepwise Fit for Oxy

Stepwise Regression Control

Stopping Rule: Minimum BIC Enter All Make Model

Direction: Forward Remove All Run Model

Go Stop Step

SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC
161.77233	27	2.4477679	0.8100	0.7889	2.8284105	4	151.5925	156.3624

Current Estimates

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Intercept	80.9007896	1	0	0.000	1
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Runtime	-2.9701867	1	443.2028	73.971	3.25e-9
<input type="checkbox"/>	<input type="checkbox"/>	Weight	0	1	4.989591	0.827	0.37137
<input type="checkbox"/>	<input checked="" type="checkbox"/>	RunPulse	-0.3751142	1	55.14175	9.203	0.00529
<input type="checkbox"/>	<input type="checkbox"/>	RstPulse	0	1	0.350744	0.056	0.81399
<input type="checkbox"/>	<input checked="" type="checkbox"/>	MaxPulse	0.35421891	1	41.34703	6.901	0.01403

3. Run Model을 실행한 결과는 다음과 같다
(Make Model을 실행하면 선정된 변수만을 대상으로 Model Launch 화면을 다시 보여준다. 이 때, 다른 변수를 포함하여 다시 분석할 수도 있다)

Response Oxy

Effect Summary

Source	LogWorth	PValue
Runtime	8.489	0.00000
RunPulse	2.277	0.00529
MaxPulse	1.853	0.01403

Remove Add Edit FDR

Summary of Fit

RSquare	0.809988
RSquare Adj	0.788876
Root Mean Square Error	2.447768
Mean of Response	47.37581
Observations (or Sum Wgts)	31

Analysis of Variance

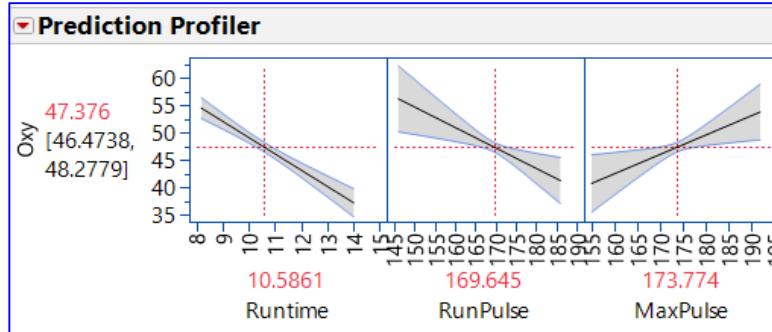
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	689.60921	229.870	38.3655
Error	27	161.77233	5.992	Prob > F
C. Total	30	851.38154		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	80.90079	8.816186	9.18	<.0001*
Runtime	-2.970187	0.345345	-8.60	<.0001*
RunPulse	-0.375114	0.12365	-3.03	0.0053*
MaxPulse	0.3542189	0.13484	2.63	0.0140*

Fit Model 사례(2) : Stepwise(단계별 회귀 분석)

4. 최소 제곱법(Standard Least Squares)에서 설명한 것처럼 Prediction Profiler, Contour Profiler 또는 Surface Profiler를 활용하여 추가적인 분석을 할 수 있다.
- 1) Prediction Profiler의 결과(▼ Fit Group / Prediction Profiler)



- 2) 만약, Surface Profiler을 사용하고자 한다면, 그 결과는 다음과 같다(▼ Fit Group / Surface Profiler)

