

<JMP의 기술적 사양과 제약사항>

Monthly User Guide (37호) – 2020년 8월호

JMP Korea 신 익주 이사(ikju.shin@jmp.com)

김 상현 수석(sanghyun.kim@jmp.com)

모든 컴퓨터 하드웨어나 소프트웨어는 그 처리량에 한계가 있습니다. 다행히 우리가 하는 작업 대부분은 그 가용범위 안에서 이루어집니다(아니라면 컴퓨터나 소프트웨어를 바꿔야 할 시점이라는 뜻이 되겠지요) JMP와 같은 Data 분석 소프트웨어는 어떤 데이터를 취급하느냐에 따라 Data 처리 또는 저장 등에 있어서 제약의 범위가 정해질 수도 있는데요. 이번 달 User Guide에서는 JMP가 할 수 있는 일은 어디까지인지 살펴보겠습니다.

1. JMP가 처리할 수 있는 데이터 테이블 크기

2007년에 출시된 JMP Version 7부터 Data Table과 관련된 소프트웨어적인 제약은 모두 사라졌습니다. 남은 것은 사용중인 **운영 체제와 메모리의 용량**입니다. 윈도우 버튼에서 오른쪽 클릭 후 '시스템'을 선택하면 장치 사양(Device Specification)에서 프로세서, RAM 및 운영 체제 등을 확인할 수 있습니다

장치 사양	
디바이스 이름	DESKTOP-XXXXXX
프로세서	Intel(R) Core(TM) i7-8665U CPU @ 1.90GHz 2.11 GHz
설치된 RAM	32.0GB(31.8GB 사용 가능)
장치 ID	XXXXXXXXXX-XXXX-XXXX-XXXX-XXXXXXXXXXXX
제품 ID	XXXXXXXXXX-XXXX-XXXX-XXXX-XXXXXXXXXXXX
시스템 종류	64비트 운영 체제, x64 기반 프로세서

만약 여러분이 아직 32비트 윈도우 운영체제에서 JMP 14 이전 버전을 사용하고 있다면, 데이터 테이블의 행과 열은 32비트 정수로 정의됩니다. 부호를 갖는 32비트 정수 최대값은 2,147,483,647(2의 31제곱)으로, 최대 행(Column) 또는 열(Row)의 수는 이를 초과할 수 없습니다.

64비트 운영체제에서는 순수하게 메모리 용량에 따라 결정됩니다. JMP는 모든 데이터를 메모리에 올린 상태로 구동됩니다. 이를 **인메모리(In-Memory) 방식**이라고 하는데, 그래프와 데이터를 실시간으로 연동하여 빠르게 탐색할 수 있는 이유이기도 합니다.

일반적인 경우 JMP는 4GB 이상의 RAM을 권장하며, 작업하는 데이터 2배 정도의 메모리가 설치되어 있는 것이 좋습니다. 파일의 크기가 메모리보다 큰 경우에도 작업은 가능합니다. 운영체제의 가상 메모리 기술 덕분인데요. 단, 작업 속도가 느려지는 건 감수해야 합니다.

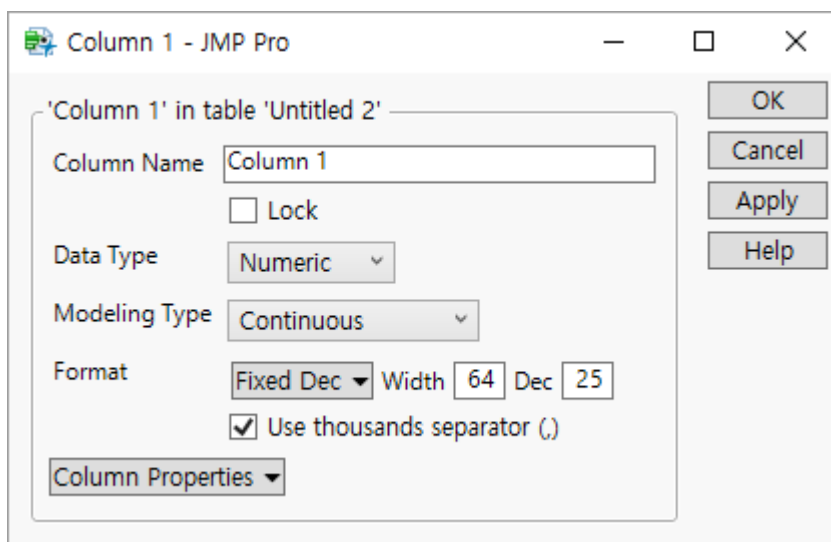
2. 하나의 셀에 저장할 수 있는 Data의 범위 및 크기

1) 숫자(Numeric)

JMP는 기본적으로 각 셀의 값에 따라 최적의 방법으로 표시합니다. 디폴트 상태에서 숫자를 입력하면 12자리, 즉 -999,999,999,999부터 999,999,999,999까지 정상 표시되는데, 여기서 9를 한번 더 입력하면 -1e+13 또는 1e+13처럼 지수 형태로 표기됩니다.

	Column 1
1	999999999999
2	1e+13

더 큰 숫자를 모두 표시하려면 변수명(Column Header)를 더블클릭하거나 변수명 위에서 우측 마우스 클릭 Column Info 선택하여, Format을 Fixed Dec로 변경하시면 됩니다. 이를 통해 **최대 64자리, 소수점 이하 25자리까지를 표시**할 수 있습니다.



그러면 64자리를 모두 9로 채운 숫자가 최대/최소값일까요?

JMP는 IEEE 754 배정밀도 부동소수점 방식(double-precision floating-point format)에 따라 실수(實數)를 표현합니다. 부동소수점이란 컴퓨터에서 실수를 다룰 때, 유효숫자를 나타내는 가수(假數)와 소수점의 위치를 풀이하는 지수(指數)로 나누어 표현하는 방식을 말

합니다. 해당 방식으로 나타낼 수 있는 숫자는 -1.7976931348623157e+308부터 1.7976931348623157e+308까지입니다.

2) 문자(Text, Character)

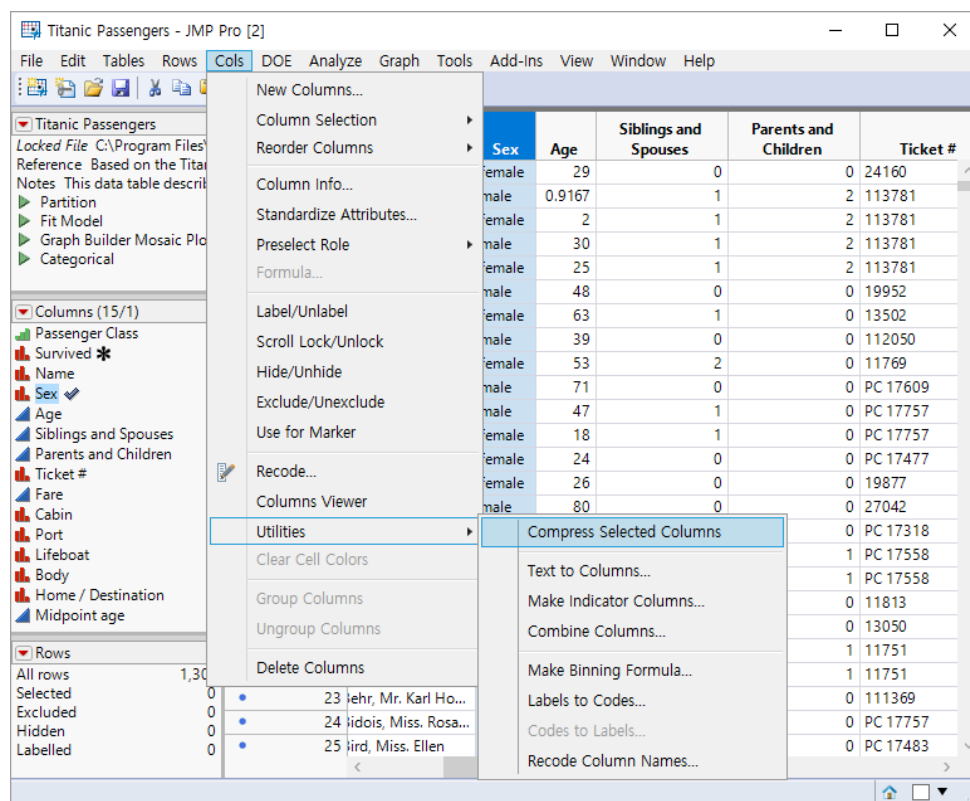
메모리 용량이 허용하는 범위 내에서 문자의 길이는 제약이 없습니다.

Column 1	Text
70	70 Debate. Discuss topics in a group. Each person should choose a viewpoint (even if you don't agree with it) and debate it within the group. Make sure you get your point across. Learn to listen actively. Active listening will help in the classroom and it will help you get more out of, and contribute more to, group study sessions. Focus on the person who is talking. Don't fidget or become distracted by other people or events. Concentrate on the speaker with your ears and eyes. Follow the movements the speaker makes in an effort to hear more. It may help to repeat what you hear others say in an effort to understand their thoughts.

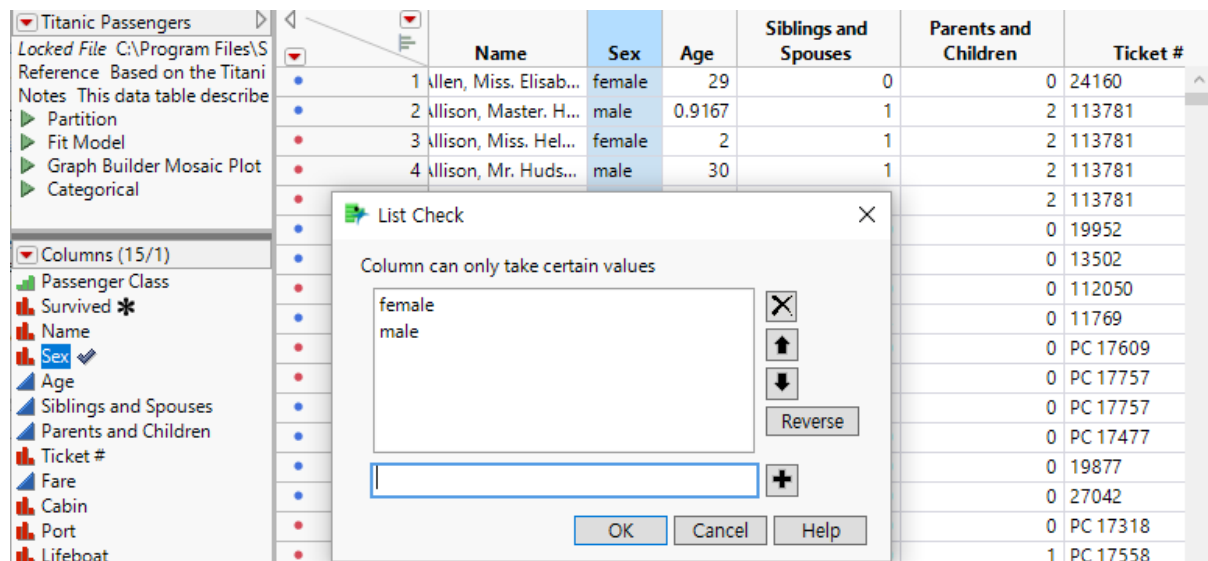
3) 메모리 용량이 부족한 경우 팁

작은 정수가 많이 포함된 숫자열이나, 고유 값의 수가 255개 미만인 열을 압축하여 데이터 테이블 용량을 줄일 수 있습니다. 예를 들어, 389개의 열과 85,000개의 행이 있는 데이터에서 이 기능을 사용하면 250 MB파일이 33MB 정도까지 줄어들기도 합니다.

사용법은 간단합니다. 대상 열(Column)을 하나 이상 선택한 후, **Cols / Utilities / Compress Selected Columns**를 실행하면 됩니다.



문자 열에서 고유 값의 개수가 255개 미만일 경우는 List Check(목록 확인)가 표시됩니다.



숫자 열의 경우, Data Format이 Best, Fixed Dec 또는 Data 형식으로 된 경우만 압축되며, 데이터에 따라 1 byte, 2 bytes 또는 4 bytes 정수로 압축됩니다.

3. JMP와 Multi-threading(멀티 쓰레딩)

Multi-threading이란 멀티코어 CPU 환경에서 처리속도 향상을 위해 연산을 분할하여 동시에 처리하는 기술입니다. AMD의 최신 라이젠 CPU 같은 경우, 최대 32코어 64쓰레드 까지 지원하고 있습니다. 이런 CPU를 구매하게 되면 JMP 처리속도도 비약적으로 향상될 거라 기대하시는 경우가 종종 있는데요. 결론부터 말씀드리면 '향상될 수도 있고 아닐 수도 있다'입니다.

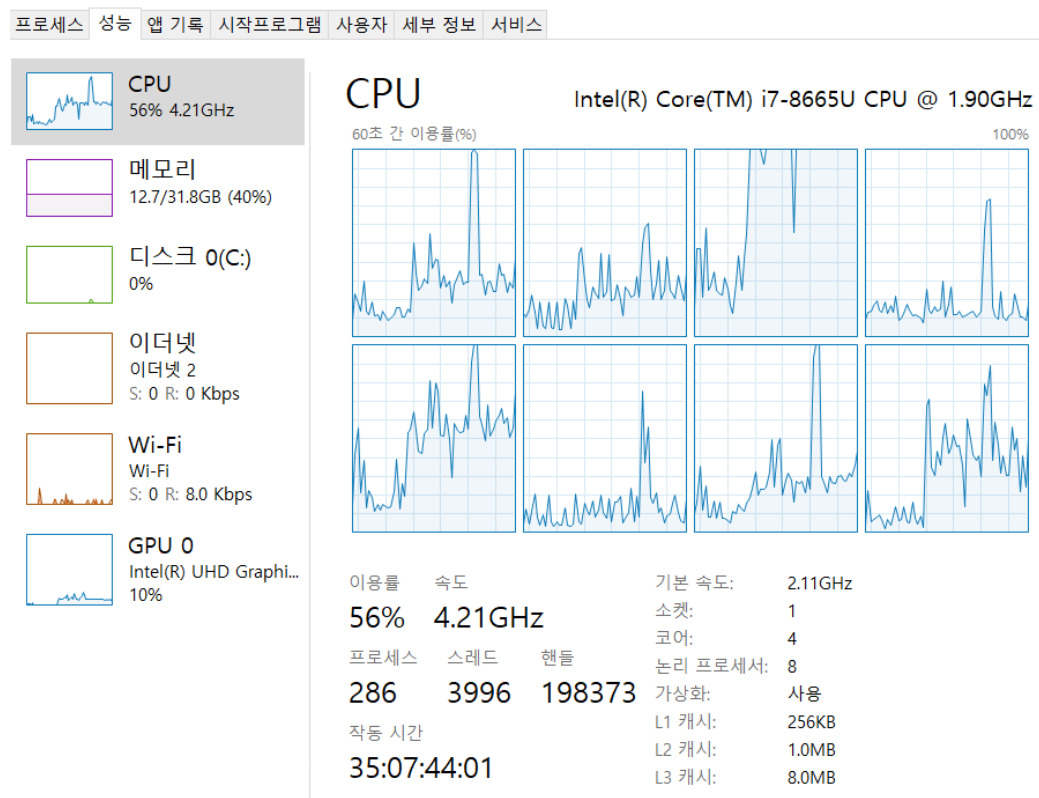
Multi-threading을 적용하려면 연산을 각각의 독립적인 계산으로 분할할 수 있어야 합니다. 그래서 JMP의 기능 중에는 Multi-threading에 적합한 연산도 있고, 그렇지 않은 연산도 있습니다. Multi-threading을 지원하는 기능은 아래와 같습니다. (밑줄은 JMP Pro)

- Boosted Trees
- Bootstrap Forest
- Choice
- Covering Arrays
- Capability Analysis in Distribution
- Factor Analysis
- Fit Life by X
- Fit Model: Parametric Survival, Mixed Model, Generalized Regression, and Response Screening
- Functional Data Explorer
- Gaussian (Fast GASP and the use of Categorical variables are only in JMP Pro.)

- Latent Class Analysis
- Life Distribution
- Multivariate
- Neural (Some features are only in JMP Pro.)
- Nominal Logistic
- Nonlinear and Nonlinear Curve
- Normal Mixtures in Cluster Analysis
- Partial Least Squares (Some features are only in JMP Pro.)
- Partition
- Predictor Screening
- Principal Components
- Process Screening
- Profiler (Does not apply to Profilers launched from within other platforms.)
- Reliability Forecast
- Repairable Systems Simulation
- Response Screening
- Sparse Principal Components
- Structural Equation Models
- Support Vector Machines
- SVD Imputation in the Missing Values Screening
- Text Explorer (SVD and Latent Class Analysis are only in JMP Pro.)
- Time Series Forecast
- Uplift
- Variable Clustering

CPU가 잘 사용되고 있는지 확인하고자 한다면 윈도우의 작업관리자 또는 리소스 모니터 프로그램을 실행시켜 확인하실 수 있습니다. 윈도우 화면 작업표시줄 오른쪽 클릭 후 작업관리자를 선택하거나 윈도우 검색에서 이름으로 검색 후 실행하면 됩니다.

아래는 조금 큰 용량의 Data를 가지고 Bootstrap Forest를 실행했을 때의 상황입니다



* 참고자료(링크)

[What is the largest table size that JMP® can handle?](#)

[Features That Support Multithreading](#)

[Column 및 Data Table 크기\(용량\) 줄이기](#)