Update : 2025년 11월 12일

# Exploration, Experimentation & Interpretation : Social Science Data Story with JMP

## (JMP Application in Business Administration and Social Science)

신 익주(ikju.shin@jmp.com)
JMP Korea

STATISTICAL DISCOVERY

# 발표 내용

**Exploration, Experimentation & Interpretation**
**: Social Science Data Story with JMP**

**1. Exploration: Exploring Survey Data**
-Exploratory Data Analysis for Survey Results
-Reliability and Validity Analysis for Survey Data

**2. Experimentation: Application in Marketing**
-JMP Marketing Techniques to Support Decision Making
-JMP Analysis Case : Conjoint
-Conjoint in JMP

**3. Interpretation: Structural Equation Mode(SEM) Application**
-Confirmative Factor Analysis(CFA)
-Path Analysis, Mediating effect etc.

# Social Science Data,,,

**1. Data from 人 : 사람과 조직의 '생각/태도' 등**
1) 정성적(Qualitative) 데이터, 많은 경우 설문 데이터
2) 직접적인 측정이 불가능한 경우가 많음
3) 평균, 산포보다는 빈도(Frequency), 비율(Proportion)

**2. 다양한 데이터 형태(Data Type, Modeling Type)**
1) 데이터 형태가 다양함 : Nominal, Binary, Ordinal, Continuous,
Text(자유 응답), Click Log 등
2) 이러한 다양한 데이터 형태가 하나의 데이터셋에 공존

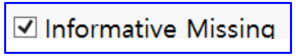
**3. 실험(Experiment)보다 관찰(Observation)이 많다**
1) 인과성(Causation)보다는 상관성(Correlation), 반복?
2) Random Design 어려움 → DOE 비스무리(Conjoint, MaxDiff 등)

**4. Sampling**
1) Random Sampling보다는 Stratified Sampling
2) 모집단과 맞추기 위해 층별 가중치(Weighting by Stratification) → 의도된 왜곡 가능성

# Social Science Data,,,

**5. 층(Layer, Hierarchy)을 가지는 경우도 많음**
1) 인적 층 : 개인 ⊂ 팀(단위 조직) ⊂ 상위 조직
2) 시간 층 : 패널 조사(Longitudinal Study)
3) ‘사람 안(within)의 변화’와 ‘사람 사이(between)의 차이’를 비교 분석

**6. 결측치(Missing Value)**




1) 결측치가 많음
2) 결측치(무응답,이탈 등)은 단순 결측(실수)가 아니라 의도, 행동 신호

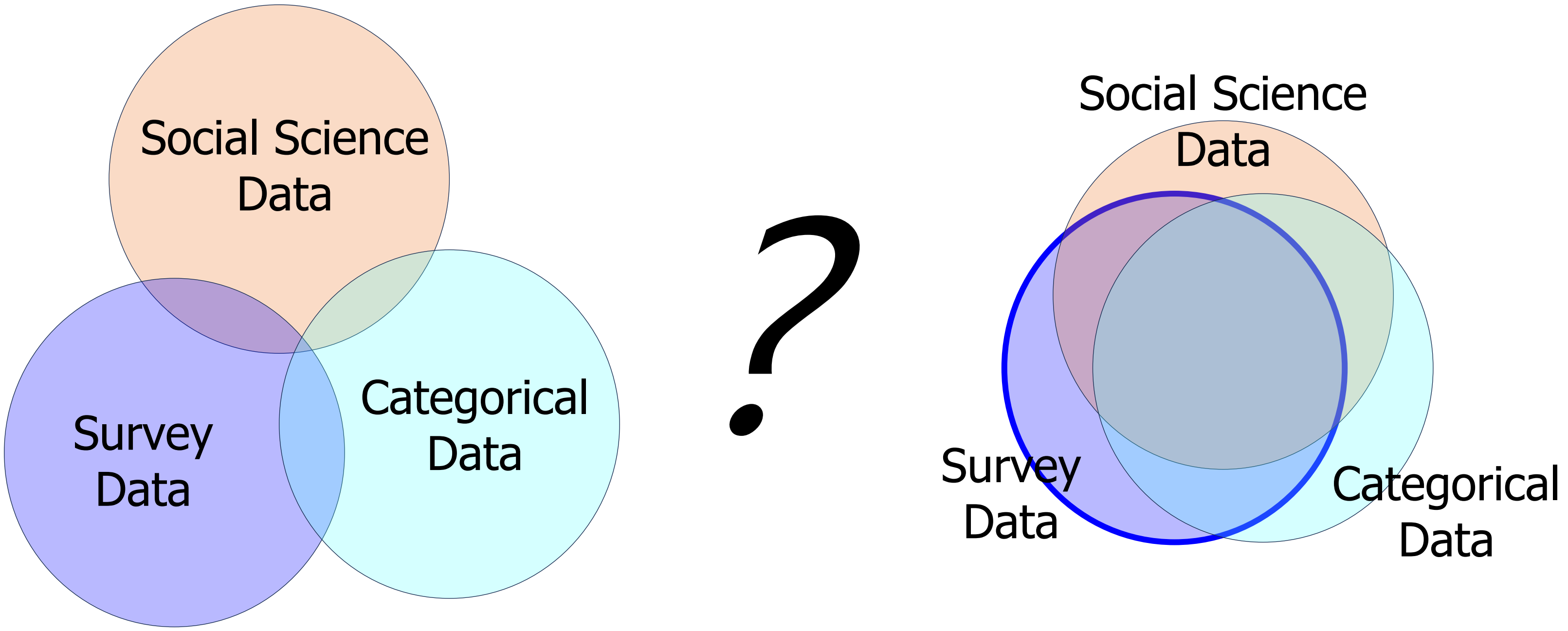
**7. 종종 GDPR* 등 특별한 Privacy, Ethics 규정에 따라야 함**

**8. 재현성과 소통이 중요**
1) 재현성이 낮음(사람, 상황에 따라 달라짐)
   유의성(Significance)보다 효과 크기(Effect Size), 신뢰구간이 더 중요
2) 소통을 잘하기 위한 시각화

* GDPR(General Data Protection Regulation)

신 익주(ikju.shin@jmp.com)

# Social Science Data,,,



# JMP Features for Social Science Data

## Design

### DOE / Consumer Studies

- Choice Design
- MaxDiff Design

## Analyze

### Analyze / Multivariate Methods

- Multivariate
- Principal Components
- Discriminant
- Partial Least Squares
- Multiple Correspondence Analysis
- Structural Equation Models
- Factor Analysis
- Multidimensional Scaling
- Multivariate Embedding
- Item Analysis

### Analyze / Screening

- Association Analysis

### Analyze / Clustering

- Hierarchical Cluster
- K Means Cluster
- Normal Mixtures
- Latent Class Analysis
- Cluster Variables

### Analyze / Consumer Research

- Categorical
- Choice
- MaxDiff
- Uplift
- Multiple Factor Analysis

## Utility

1. Data Table
1) Modeling Type
2) Header Graph/Statistics
3) Data Table Filter(JMP19)

2. Column Property
1) Multiple Response
2) Super-category
3) Recode
4) Value Label

3. Others
1) Data Filter
2) Cronbach's α
3) Exploring Missing Value
4) Freq, Weight
5) Forced Value(JMP19)
,,,

# 발표 내용

**탐색, 실험, 해석 : JMP로 풀어보는 사회과학 데이터 이야기**

**1. 탐색 : 설문 데이터 살펴보기**
1) 설문 데이터의 탐색(탐색적 데이터 분석)
2) 설문 데이터의 신뢰도와 타당도 분석

**2. 실험 : 마케팅 분야 적용**
-의사 결정을 지원하는 JMP의 마케팅 기법들
-JMP 분석 사례 : Conjoint
-JMP에서의 컨조인트 분석

**3. 해석 : 구조 방정식 적용 사례**
-확증적 요인 분석
-경로 분석, 매개(Mediating) 효과 등

# 1. 탐색 : 설문 데이터 살펴보기



**탐색적 데이터 분석(EDA : Exploratory Data Analysis)**
1) Summarize and Visualize Data for understanding and communicating
2) 좁은 의미 : Visualization + Summarization(통계량)
3) 넓은 의미 : 좁은 의미 + 관심 대상 변수에 대한 분석(주로, 유의성에 대한 확인)

**Survey Data에 대한 탐색적 데이터 분석**

**1) Visualization**
Single Variable : Pie Chart, Bar Chart, Tree Map 등
Two(& Multiple) Variables : Stacked Bar, Mosaic Plot 등

**2) Summarization**
Analyze / Distribution
Analyze / Tabulate

**3) Categorical Analysis**
Analyze / Fit Y by X
Analyze / Consumer Research / Categorical

# 1. 탐색 : 설문 데이터 살펴보기

## 예 1) <u>**Visualization**</u> for Single Variable*


*2. JMP를 알게(사용하게) 된 계기를 선택하여 주세요(복수 선택 가능)

- 인터넷 검색
- 동료 추천
- 업계 관계자 소개
- JMP 홈페이지, JMP 뉴스레터 등
- 학회, 협회, 전시회 등
- SNS(Social Media)
- 기타

| JMP를 알게(사용하게) 된 계기를 선택하여 주세요(복수 선택 가능) |
|---|
| 업계 관계자 소개 |
| 업계 관계자 소개 |
| 동료 추천, JMP(홈페이지/뉴스레터 등), 학회/협회/전시회 등 |
| 동료 추천, 업계 관계자 소개 |
| JMP(홈페이지/뉴스레터 등), 학회/협회/전시회 등 |
| 동료 추천 |
| 업계 관계자 소개, JMP(홈페이지/뉴스레터 등) |
| 동료 추천 |
| 회사에서 사용중 |
| 동료 추천 |
| 회사에서 사용중 |
| 업계 관계자 소개, JMP(홈페이지/뉴스레터 등), 학회/협회/전시회 등 |
| 업계 관계자 소개 |
| JMP(홈페이지/뉴스레터 등), SNS(Social Media) |
| 인터넷 검색, JMP(홈페이지/뉴스레터 등), SNS(Social Media) |

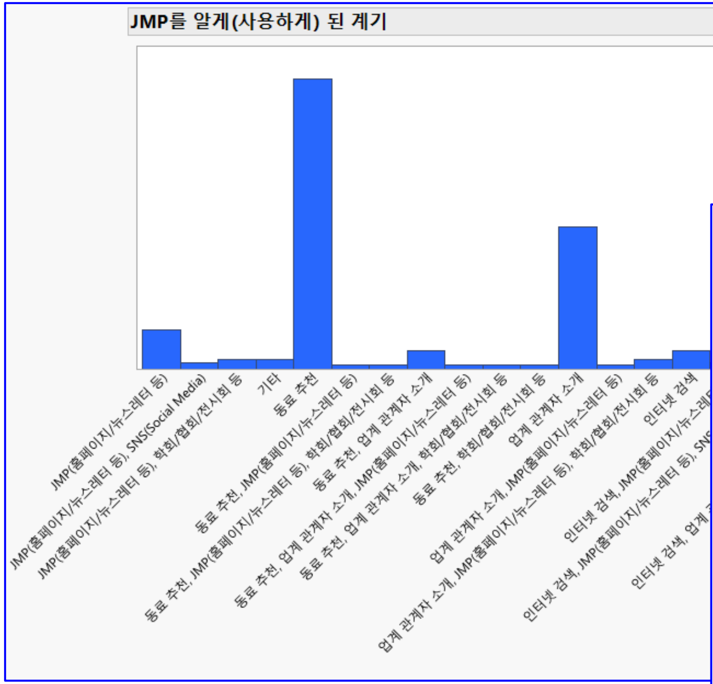
,,, N = 239, '기타' 중 '회사에서 사용 중 ' 을 별도 범주화

* 2025년 JMP Open Seminar(Webinar) 참석자 설문 중 일부

# 1. 탐색 : 설문 데이터 살펴보기



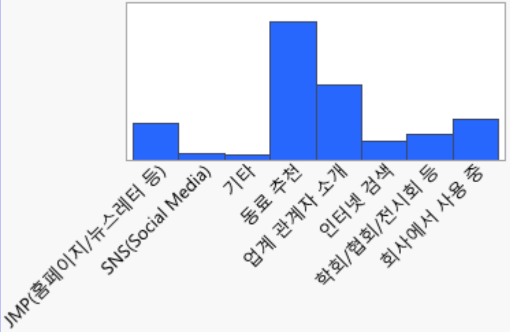
## 예 1) **Visualization** for Single Variable*




Multiple Response(다중 반응치)

| | | | |
|---|---|---|---|
| Data Type | Character | Subtype | Normal |
| Modeling Type | Multiple Response | | |

**JMP를 알게(사용하게) 된 계기(Multiple Response)**




**Frequencies**

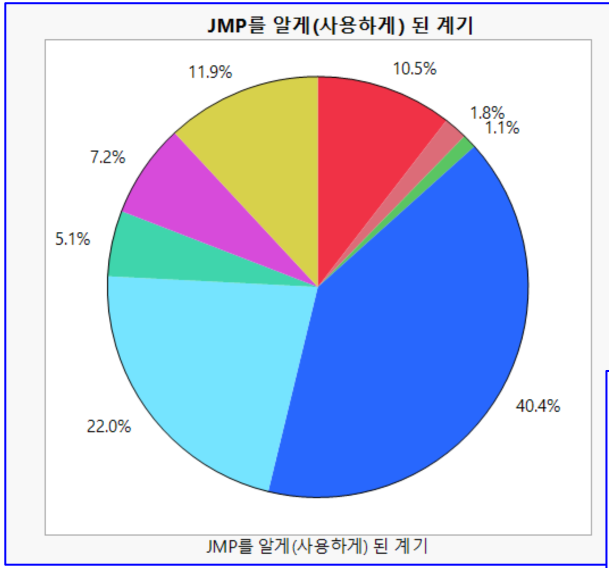
| Level | Count | Share of Responses | Rate Per Case |
|---|---|---|---|
| JMP(홈페이지/뉴스레터 등) | 29 | 0.1047 | 0.1213 |
| SNS(Social Media) | 5 | 0.0181 | 0.0209 |
| 기타 | 3 | 0.0108 | 0.0126 |
| 동료 추천 | 112 | 0.4043 | 0.4686 |
| 업계 관계자 소개 | 61 | 0.2202 | 0.2552 |
| 인터넷 검색 | 14 | 0.0505 | 0.0586 |
| 학회/협회/전시회 등 | 20 | 0.0722 | 0.0837 |
| 회사에서 사용 중 | 33 | 0.1191 | 0.1381 |

| | | | |
|---|---|---|---|
| Total Cases | 239 | Empty | 0 |
| Total Responses | 277 | Responding | 239 |
| Levels | 8 | Single Item | 210 |
| | | Multiple Item | 29 |

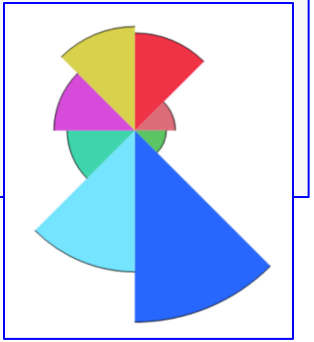
* 2025년 JMP Open Seminar(Webinar) 참석자 설문 중 일부

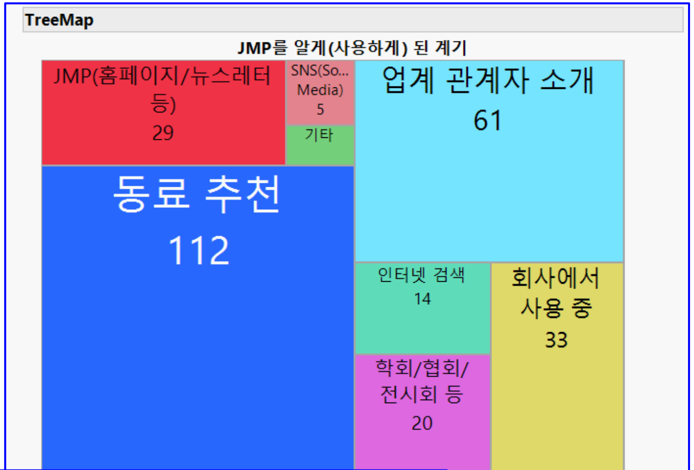
# 1. 탐색 : 설문 데이터 살펴보기

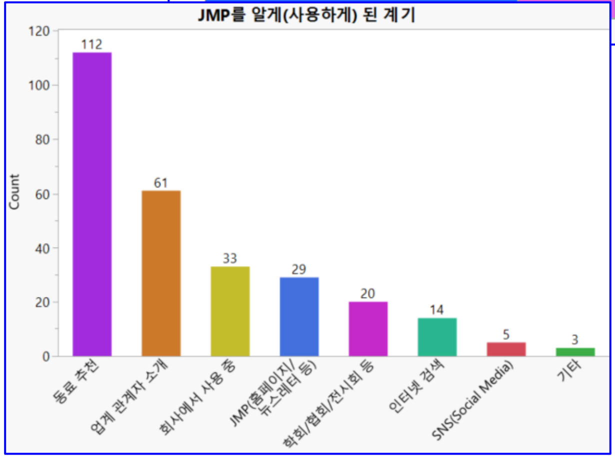
## 예 1) Visualization for Single Variable*




Coxcomb Chart






* 2025년 JMP Open Seminar(Webinar) 참석자 설문 중 일부

신 익주(ikju.shin@jmp.com)

# 1. 탐색 : 설문 데이터 살펴보기



예 2) **Visualization** for Two(& Multiple) Variables*

두 가지 Survey : JMP 기능 사용 빈도(30 문항), 통계 및 JMP 지식(21 문항), Likert 5

| 번호 | 번호 | JMP 기능에 대한 사용 빈도 조사(30문항) | 사용해 본 적이 전혀 없다 | 한두번 사용해 본 적이 있다 | 필요할 때 가끔 사용 | 자주 사용한다 | 매우 자주 사용 (업무에 필수적) |
|---|---|---|---|---|---|---|---|
| 1-1 | 1 | Header Graph(머리글 그래프) | 1 | 2 | 3 | 4 | 5 |
| 1-2 | 2 | Local Data Filter(로컬 데이터 필터) | | | | | |
| 1-3 | 3 | Column Property(열 특성) | | | | | |
| 1-4 | 4 | New Formula Column(새 계산식 열) | | | | | |
| 1-5 | 5 | Script를 활용한 분석 자동화 | | | | | |
| 2-1 | 6 | 분석/분포(Analyze / Distribution) | | | | | |
| 2-2 | 7 | 정규성 검정(Normality Test) | | | | | |
| 2-3 | 8 | 그래프 빌더(Graph Builder) | | | | | |
| 2-4 | 9 | 요약표 만들기(Analyze / Tabulate) | | | | | |
| 2-5 | 10 | T Test(1 Sample T 또는 2 Sample T) | | | | | |
| 2-6 | 11 | 분산분석(ANOVA) : Analyze / Fit Y by X | | | | | |
| 2-7 | 12 | 회귀분석(Regression)  : Analyze / Fit Y by X | | | | | |
| 2-8 | 13 | 상관분석 | | | | | |
| 2-9 | 14 | 예측 변수 선별(Predictor Screening) | | | | | |
| 2-10 | 15 | 군집 분석 | | | | | |

| | | |
|---|---|---|
| 3-1 | 16 | 모형 적합(최소 제곱법) : Analyze / |
| 3-2 | 17 | 단계별 회귀(Stepwise) |
| 3-3 | 18 | Prediction Profiler |
| 3-4 | 19 | Contour Profiler |
| 3-5 | 20 | 몬테 카를로 시뮬레이션 |
| 4-1 | 21 | 완전배치법(Full Factorial) |
| 4-2 | 22 | 부분배치법(Screening Design) |
| 4-3 | 23 | 사용자 설계(Custom Design) |
| 4-4 | 24 | 설계 확대(Augment Design) |
| 4-5 | 25 | 반응 표면 실험(Response Surface |
| 5-1 | 26 | 공정 능력 분석 |
| 5-2 | 27 | 측정 시스템 분석(Gauge R&R) |
| 5-3 | 28 | 계량형/계수형 차트(Variability Cha |
| 5-4 | 29 | 관리도 빌더(Control Chart Builder) |
| 5-5 | 30 | 공정 변수 선별(Process Screening) |

| 번호 | 번호 | 통계 및 JMP 활용 지식에 대한 조사(21문항) |
|---|---|---|
| 1-1 | 1 | JMP Data Table과 Excel Worksheet의 차이점을 알고 있다 |
| 1-2 | 2 | 변수별 모델링 타입을 확인하고, 필요한 경우 이를 변경할 수 있다. |
| 1-3 | 3 | JMP Script를 이용하여 반복적인 분석을 자동화할 수 있다. |
| 2-1 | 4 | 그래프 빌더를 이용하여 하나의 화면에 여러 개의 그래프를 그릴 수 있다 |
| 2-2 | 5 | 그래프 빌더를 이용하여 Color Map(Wafer Map 등)을 그릴 수 있다 |
| 2-3 | 6 | JMP를 이용하여 이상치(outlier)를 식별하고 조치할 수 있다. |
| 3-1 | 7 | JMP를 이용하여 특정한 변수의 평균값의 신뢰 구간을 구할 수 있다. |
| 3-2 | 8 | Boxplot의 중앙값과 평균값을 비교하여 설명할 수 있다. |
| 3-3 | 9 | 데이터가 정규분포하지 않을 경우 어떻게 해야 되는 지 이해하고 있다. |
| 4-1 | 10 | ANOVA(분산 분석)에서 P Value의 의미를 이해하고 해석할 수 있다. |
| 4-1 | 11 | 회귀 분석에서 R-Square의 의미를 이해하고 해석할 수 있다. |
| 4-3 | 12 | 변수의 모델링 타입에 따라 적합한 유의차 검정(가설 검정) 방법을 선택할 수 있다. |
| 5-1 | 13 | 최소 제곱(Least Square)의 의미를 이해하고 해석할 수 있다 |
| 5-2 | 14 | 여러 예측 변수(X 인자) 중에서 가장 유의한 변수를 찾는 방법을 아해하고 있다 |
| 5-3 | 15 | 모델링에서 직선이 아닌 곡선 효과를 적합하는 방법을 알고 있다. |
| 6-1 | 16 | 실험계획의 3대 원리(랜덤화, 반복 실험, 블록화)를 이해하고 있다 |
| 6-2 | 17 | JMP를 이용하여 반응 표면 실험을 디자인할 수 있다. |
| 6-3 | 18 | Prediction Profiler의 기능을 이해하고 활용할 수 있다. |
| 7-1 | 19 | Pp, Ppk의 차이를 이해하고 해석할 수 있다. |
| 7-2 | 20 | 계량형/계수형 차트(Variability Chart) 결과를 이해하고 해석할 수 있다 |
| 7-3 | 21 | 관리도 빌더(Control Chart Builder)에서 다양한 종류의 관리도를 만들 수 있다. |

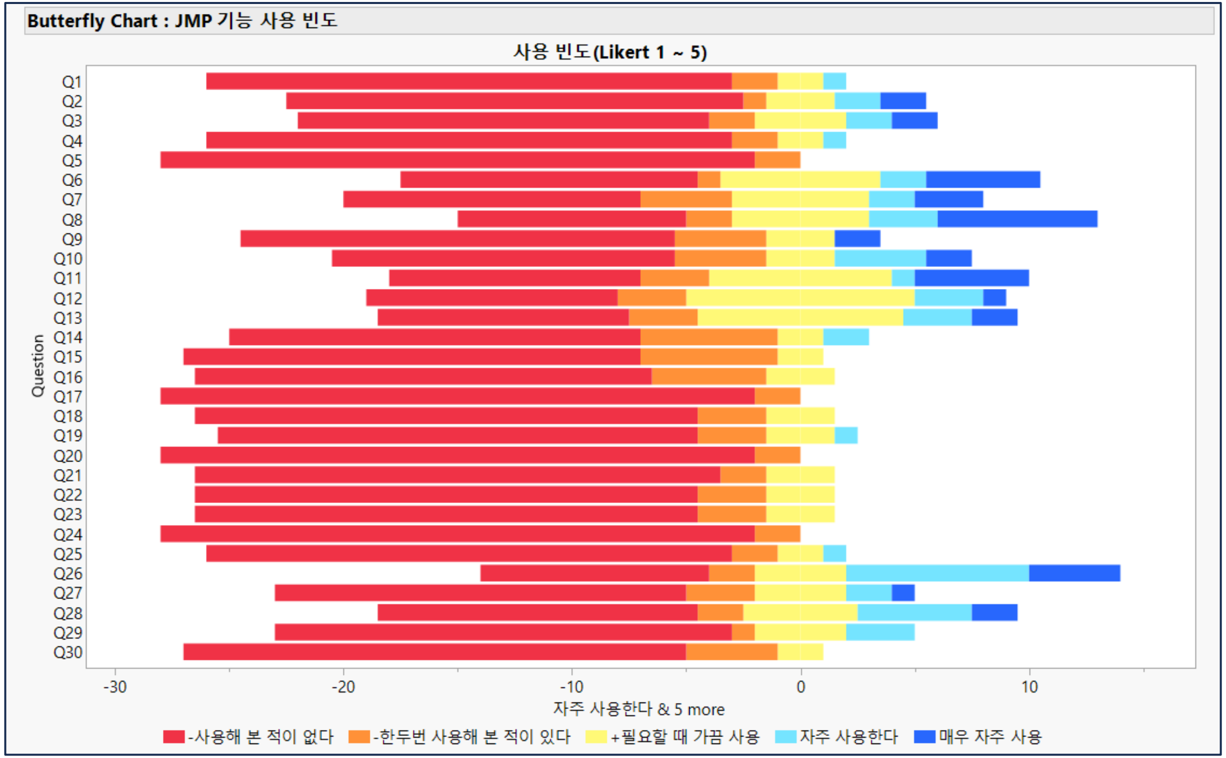
* 두 회사의 JMP 사용자 30명 대상 설문

신 익주(ikju.shin@jmp.com)

# 1. 탐색 : 설문 데이터 살펴보기



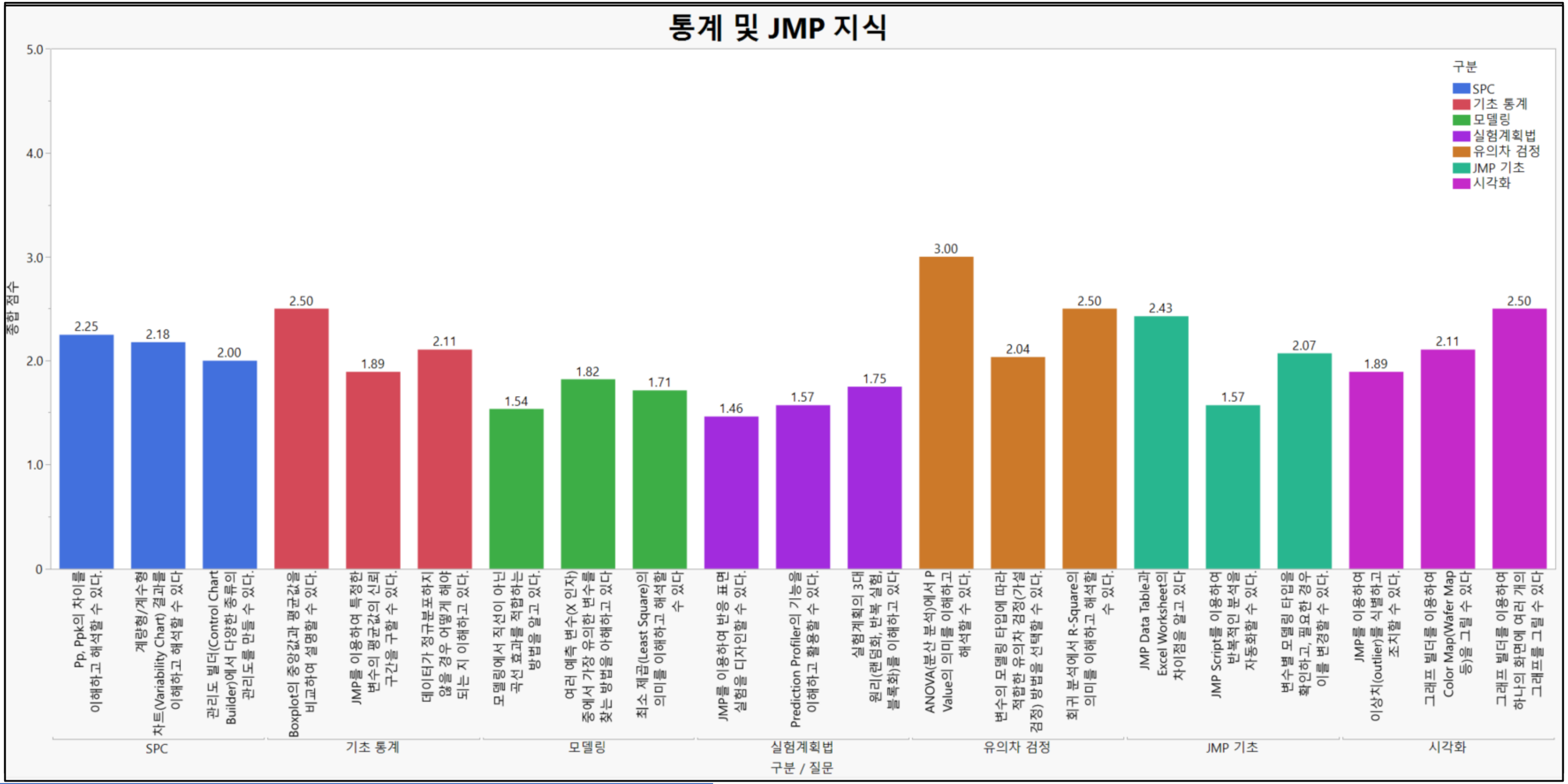
예 2) **Visualization** for Two(& Multiple) Variables




신 익주(ikju.shin@jmp.com)

# 1. 탐색 : 설문 데이터 살펴보기

## 예 2) **Visualization** for Two(& Multiple) Variables




신 익주(ikju.shin@jmp.com)

# 1. 탐색 : 설문 데이터 살펴보기



## 예 2) **Summarization** for Two(& Multiple) Variables

**Local Data Filter**

105 matching rows

☐ Inverse

**Category (2)**

| | |
|---|---|
| JMP기능 사용빈도 | 150 |
| 통계 및 JMP 지식 | 105 |

**Tabulate**

**Category** = 통계 및 JMP 지식

| | Level | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| 설문 문항 | N (Row %) | N (Row %) | N (Row %) | N (Row %) | N (Row %) |
| JMP기초-1 | 6 (21.43%) | 9 (32.14%) | 9 (32.14%) | 3 (10.71%) | 1 (3.57%) |
| JMP기초-2 | 12 (42.86%) | 8 (28.57%) | 3 (10.71%) | 4 (14.29%) | 1 (3.57%) |
| JMP기초-3 | 17 (60.71%) | 6 (21.43%) | 5 (17.86%) | 0 (0.00%) | 0 (0.00%) |
| 시각화-1 | 9 (32.14%) | 7 (25.00%) | 4 (14.29%) | 5 (17.86%) | 3 (10.71%) |
| 시각화-2 | 11 (39.29%) | 9 (32.14%) | 2 (7.14%) | 6 (21.43%) | 0 (0.00%) |
| 시각화-3 | 13 (46.43%) | 9 (32.14%) | 2 (7.14%) | 4 (14.29%) | 0 (0.00%) |
| 기초 통계-1 | 12 (42.86%) | 10 (35.71%) | 3 (10.71%) | 3 (10.71%) | 0 (0.00%) |
| 기초 통계-2 | 10 (35.71%) | 5 (17.86%) | 2 (7.14%) | 11 (39.29%) | 0 (0.00%) |
| 기초 통계-3 | 11 (39.29%) | 6 (21.43%) | 8 (28.57%) | 3 (10.71%) | 0 (0.00%) |
| 유의차 검정-1 | 5 (17.86%) | 4 (14.29%) | 9 (32.14%) | 6 (21.43%) | 4 (14.29%) |
| 유의차 검정-2 | 6 (21.43%) | 8 (28.57%) | 9 (32.14%) | 4 (14.29%) | 1 (3.57%) |
| 유의차 검정-3 | 9 (32.14%) | 11 (39.29%) | 6 (21.43%) | 2 (7.14%) | 0 (0.00%) |
| 모델링-1 | 15 (53.57%) | 9 (32.14%) | 2 (7.14%) | 1 (3.57%) | 1 (3.57%) |
| 모델링-2 | 13 (46.43%) | 9 (32.14%) | 4 (14.29%) | 2 (7.14%) | 0 (0.00%) |
| 모델링-3 | 16 (57.14%) | 10 (35.71%) | 1 (3.57%) | 1 (3.57%) | 0 (0.00%) |
| DOE-1 | 16 (57.14%) | 7 (25.00%) | 2 (7.14%) | 2 (7.14%) | 1 (3.57%) |
| DOE-2 | 19 (67.86%) | 7 (25.00%) | 0 (0.00%) | 2 (7.14%) | 0 (0.00%) |
| DOE-3 | 18 (64.29%) | 5 (17.86%) | 4 (14.29%) | 1 (3.57%) | 0 (0.00%) |
| SPC-1 | 11 (39.29%) | 3 (10.71%) | 10 (35.71%) | 4 (14.29%) | 0 (0.00%) |
| SPC-2 | 10 (35.71%) | 7 (25.00%) | 7 (25.00%) | 4 (14.29%) | 0 (0.00%) |
| SPC-3 | 13 (46.43%) | 7 (25.00%) | 3 (10.71%) | 5 (17.86%) | 0 (0.00%) |

150 rows have been excluded.

신 익주(ikju.shin@jmp.com)

# 1. 탐색 : 설문 데이터 살펴보기



## 예 3) Categorical Analysis(Analyze / Fit Y by X)

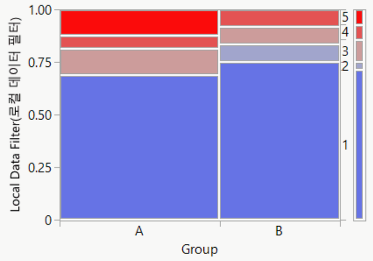
**Column Switcher**

30 Columns

- Header Graph(머리글 그래프)
- Local Data Filter(로컬 데이터 필터)
- Column Property(열 특성)
- New Formula Column(새 계산식 열)
- Script를 활용한 분석 자동화
- 분석/분포(Analyze / Distribution)
- 정규성 검정(Normality Test)
- 그래프 빌더(Graph Builder)
- 요약표 만들기(Analyze / Tabulate)
- T Test(1 Sample T 또는 2 Sample T)
- 분산분석(ANOVA...nalyze / Fit Y by X
- 회귀분석(Regress...nalyze / Fit Y by X
- 상관분석
- 예측 변수 선별(Predictor Screening)
- 군집 분석
- 모형 적합(최소 ...Analyze / Fit Model
- 단계별 회귀(Stepwise)
- Prediction Profiler
- Contour Profiler
- 몬테 카를로 시뮬레이션
- 완전배치법(Full Factorial)
- 부분배치법(Screening Design)
- 사용자 설계(Custom Design)
- 설계 확대(Augment Design)
- 반응 표면 실험(R... Surface Method)
- 공정 능력 분석
- 측정 시스템 분석(Gauge R&R)
- 계량형/계수형 차트(Variability Chart)

**Contingency Analysis of Local Data Filter(로컬 데이터 필터) By Group**

**Mosaic Plot**




**Contingency Table**

Local Data Filter(로컬 데이터 필터)

| Group | Count<br>Total %<br>Col %<br>Row % | 1 | 2 | 3 | 4 | 5 | Total |
|---|---|---|---|---|---|---|---|
| A | | 11<br>39.29<br>55.00<br>68.75 | 0<br>0.00<br>0.00<br>0.00 | 2<br>7.14<br>66.67<br>12.50 | 1<br>3.57<br>50.00<br>6.25 | 2<br>7.14<br>100.00<br>12.50 | 16<br>57.14 |
| B | | 9<br>32.14<br>45.00<br>75.00 | 1<br>3.57<br>100.00<br>8.33 | 1<br>3.57<br>33.33<br>8.33 | 1<br>3.57<br>50.00<br>8.33 | 0<br>0.00<br>0.00<br>0.00 | 12<br>42.86 |
| Total | | 20<br>71.43 | 1<br>3.57 | 3<br>10.71 | 2<br>7.14 | 2<br>7.14 | 28 |

**Tests**

| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 28 | 4 | 2.0628138 | 0.0755 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 4.126 | 0.3893 |
| Pearson | 3.024 | 0.5539 |

Warning: 20% of cells have expected count less than 5, ChiSquare suspect.

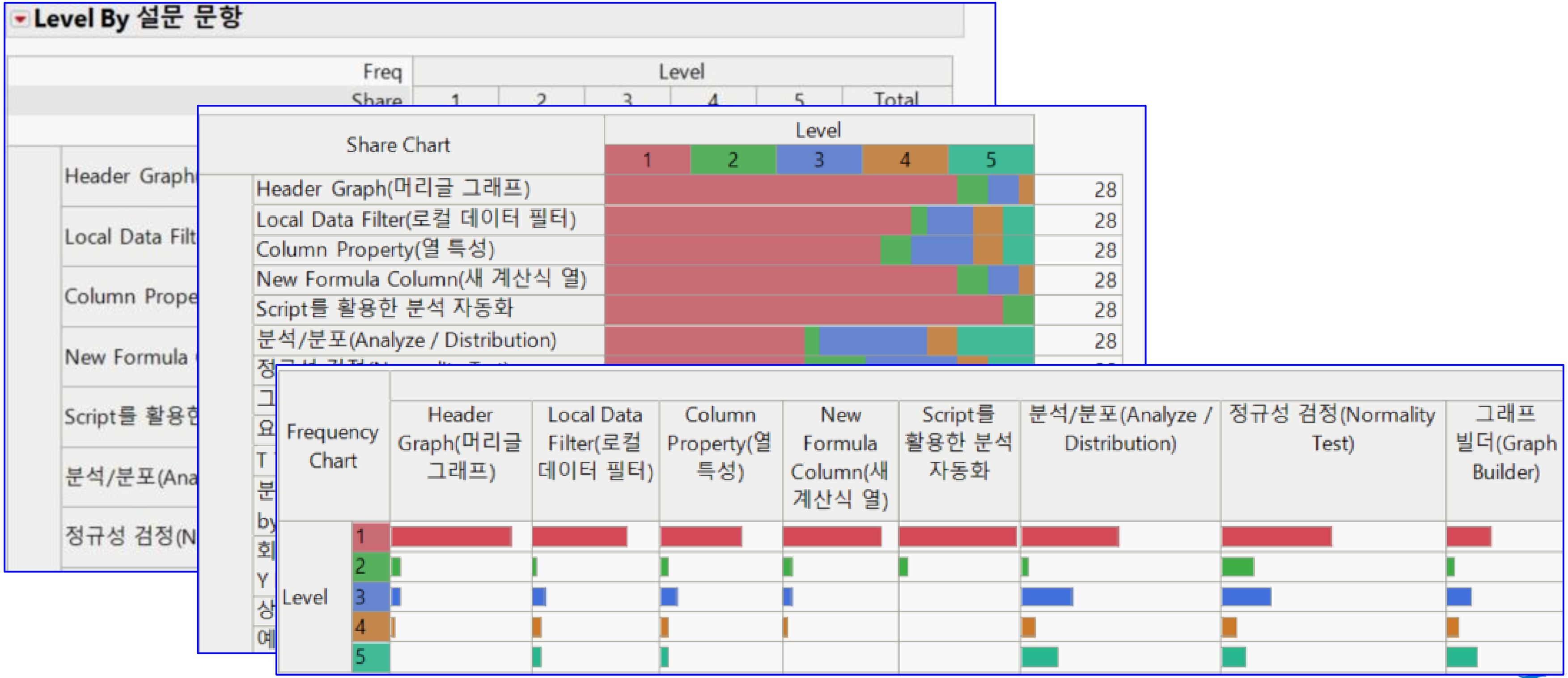
Warning: Average cell count less than 5, LR ChiSquare suspect.

신 익주(ikju.shin@jmp.com)

# 1. 탐색 : 설문 데이터 살펴보기



예 3) Categorical Analysis(Analyze / Consumer Research / Categorical)




신 익주(ikju.shin@jmp.com)

# 1. 탐색 : 설문 데이터 살펴보기

## 2) 신뢰도와 타당도 분석

# 測 ≠ 定

재서 정한다 →

← 정해진 것을 잰다

측정
1) (사회과학적) 개념, 변수 등을 현실세계에서
   관찰 가능한 자료와 연결하는 과정
2) 일정한 절차에 따라 사물, 사건 등에 대해 숫자를 부여하는 과정 및 결과

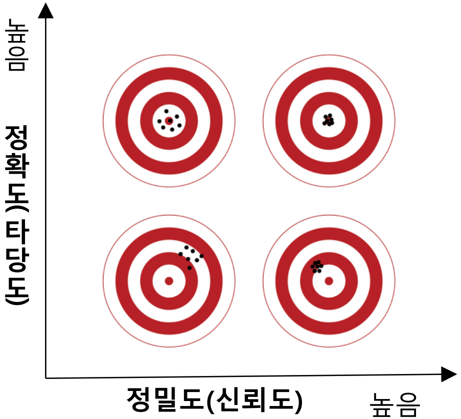
측정의 대상(사회 과학적 구분)
1) (구체적인) 변수(Variable) : 예) 성별, 직원 수, 길이 등
2) (추상적인) 개념(Construct) : 예) 만족도, 얼마나 사랑해 ?

추상적인 Construct에 대한 측정 도구
1) 예) 얼마나 사랑해 ? ← 일 통화회수, 카톡 답장 소요 시간
2) 측정 도구가 Construct를 얼마나 정확히, 일관성 있게 반영하는 지가 중요
   : 측정의 신뢰도와 타당도 문제 발생

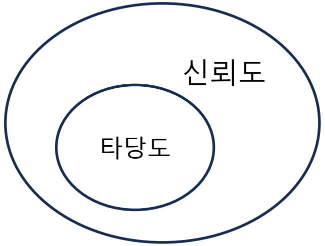
* 참고 : https://horizon.kias.re.kr/19879/ (KAIST 물리학과 최형순 교수)

**타당도(Validity, 타당성)**
1) 측정 도구가 측정하고자 하는 변수를 제대로 **정확하게** 측정하고 있는 정도
2) 척도가 그 목적에 맞는 정확한 정보를 제공하는 지를 표현
3) 잘못된 척도로 인해 발생하는 체계적 오류(Systematic Error)와 연관

**신뢰도(Reliability, 신뢰성)**
1) 측정 도구에 의해 도출되는 측정값들이 얼마나 **일정하게** 산출되는지를 나타내는 개념
2) 측정의 내적 일관성(Internal Consistency), 안정성, 예측성
3) 측정자 또는 측정 상황과 관련된 비체계적 오류(Nonsystematic Error, Random Error)와 연관




-타당도는 신뢰도의 충분조건, 신뢰도는 타당도의 필요 조건
-신뢰도가 없으면 타당도도 없다
-타당도가 있으면 반드시 신뢰도가 있다
-일반적으로 타당도 분석 후 신뢰도 분석




* 2202_측정이란 무엇인가2.pdf (kias.re.kr)-KAIST 물리학과 최형순 교수

신 익주(ikju.shin@jmp.com)

# 1. 탐색 : 설문 데이터 살펴보기

## 2) 신뢰도와 타당도 분석 : 타당도

| | |
|---|---|
| **외적 타당도(External Validity)** | 1) 실험, 연구 결과를 얼마만큼 일반화하여 적용할 수 있는가 ?<br>2) 외적 타당도는 일반화의 한계 사항으로 언급됨 |
| **내적 타당도(Internal Validity)** | 1) 실제로 측정하고자 하는 것을 얼마나 정확하게 측정하였는가?<br>(외부 요인의 통제와 관련, 측정 결과를 얼마나 믿을 수 있는가?)<br>2) 사회 과학 연구의 경우 내적 타당도가 분석의 중심 |

| (내적) 타당도(Validity) | 내용 |
|---|---|
| **내용 타당도<br>(Content Validity) : 대표성** | 1) 척도의 문항들이 측정하고자 하는 개념을 얼마나 잘 대표하는 지에 대한 논리적 타당도를 평가<br>2) 전문가 판단(Expert Reviw)에 의해 평가되는 경우가 많음 |
| **기준 관련 타당도(Criterion-Related Validity) : 효과성** | 1) 척도의 결과가 외부 기준(기준 변수)과 얼마나 관련이 있는지를 평가<br>2) 동시 타당도와 예측 타당도로 구분 |
| **동시 타당도<br>(Concurrent Validity)** | 1) 척도 결과와 기준 변수를 동시에 측정했을 때의 상관관계를 평가<br>2) 예 : 고교내신보다 수능점수가 고교졸업 학력을 더 잘 반영한다고 본다면<br>이는 수능점수가 고교졸업 학력에 대해 동시타당도가 더 높다고 본다는 뜻 |
| **예측 타당도<br>(Predictive Validity)** | 1) 척도의 결과가 미래의 기준 변수와 얼마나 관련이 있는지를 평가<br>2) 예 : 대학졸업성적과 고교내신, 수능점수 비교결과 수능점수와 대학졸업성적간의 상관성이 높으면 수능점수는 고교내신보다 대학졸업성적에 대해 예측 타당도가 높다고 할 수 있음 |
| **구성(개념) 타당도<br>(Construct Validity) : 논리성** | 1) 척도가 측정하고자 하는 개념(Construct)를 정확하게 측정하고 있는지를 평가<br>2) 집중 타당도와 판별 타당도로 구분<br>3) **구성(개념, Construct)들에 대한 타당도 검증은 보통 탐색적 요인분석(EFA : Exploratory Factor Analysis)**을 활용하여 분석하고<br>4) **확인적 요인분석(CFA : Confirmatory Factor Analysis)를 통해 집중 및 판별 타당성을 각각 분석**함 |
| **집중(수렴) 타당도<br>(Convergent Validity)** | 1) 같은 개념을 측정하는 유사한 문항들이 일관성이 있는가를 평가<br>2) **각 문항들이 같은 개념을 측정했을 때 상관관계가 높아야 함** |
| **판별 타당도<br>(Discriminant Validity)** | **서로 다른 개념을 측정했을 때 그 결과 간에 상관 관계가 낮아야 한다** |

# 1. 탐색 : 설문 데이터 살펴보기

## 2) 신뢰도와 타당도 분석 : 타당도

| 구성(개념) 타당도<br>(Construct Validity) | 평가 지표(Criterion) | 타당도 기준(일반적) | JMP |
|---|---|---|---|
| 집중(수렴) 타당도<br>(Convergent Validity) | 요인적재량<br>(Outer Loading) | > 0.7 | Factor Analysis |
| | 합성신뢰도<br>(Composite Reliability : CR) | > 0.7 | SEM |
| | 평균분산추출<br>(Average Variance Extracted : AVE) | > 0.5 | SEM |
| 판별 타당도<br>(Discriminant Validity) | Fornell-Larcker Criterion(각 변수의 AVE 제곱근 값)<br>각 변수의 AVE 제곱근과 변수 간 상관관계를 비교 | 변수의 평균분산추출 제곱근 값이<br>변수 간의 상관계수 값보다 높아야 함 | SEM |
| | 교차 적재치(Cross Loadings Analysis)<br>탐색적 요인분석의 일종으로 요인적재치와 교차적재치를 활용하여 평가하는 방법, | 특정한 변수와 관련된 한 측정변수의 값은<br>다른 변수들과 연관된 측정변수들의<br>모든 교차적재치보다 커야 함 | SEM |

신 익주(ikju.shin@jmp.com)

## 2) 신뢰도와 타당도 분석 : 신뢰도

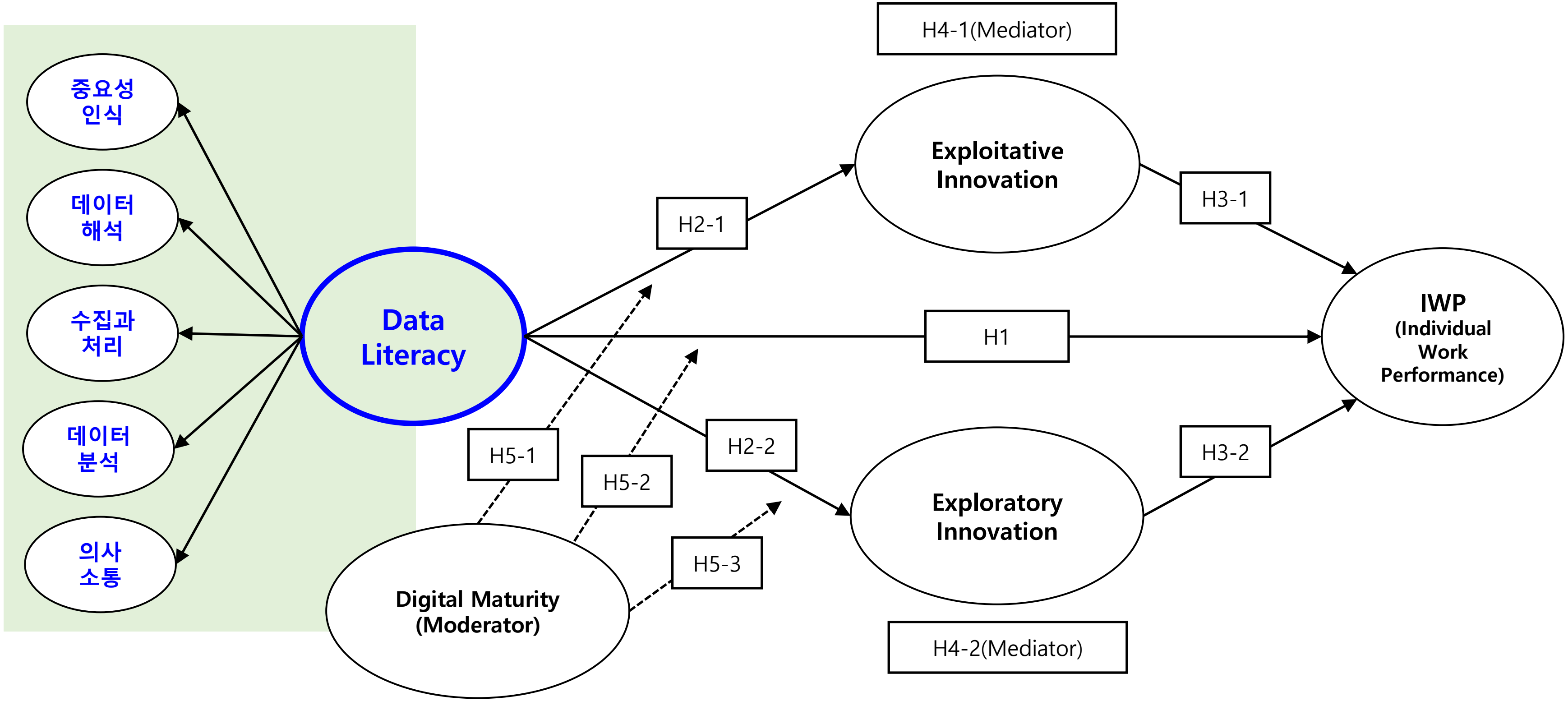
| 평가 방법 | 내용 |
|---|---|
| **재검사법 (Test-Retest Method)** | 1) 동일한 척도를 동일한 집단에 일정한 시간 간격을 두고 두 번 측정했을 때 결과가 얼마나 일관되게 나오는지를 평가<br>2) 두 번 측정 결과의 상관계수로서 신뢰도의 정도를 추정 |
| **내적 일관성 방법 (Internal Consistency Method)** | 1) 하나의 개념(Construct)에 대해 여러 항목으로 측정했을 때 그 문항들간의 일관성을 평가<br>2) 여러 항목들을 양분(split-halves)하는 상관 관계를 계산하고, 가능한 모든 상관관계 값들의 평균값인 크론바흐의 알파(Cronbach's Alpha) 계수가 대표적이며 실무적으로 가장 많이 활용된다 |
| **반분 신뢰도 (Split-Half Method)** | 1) 측정 항목(설문 문항 등)을 두 부분으로 나누어 각각의 부분에서 얻은 점수 간의 상관관계를 측정하여 평가<br>2) 반분된 측정 항목에 대한 동등성 확보가 어려움 |
| **복수 양식법 (Multiple Forms Method)** | 1) 두 개의 유사한 측정 도구를 사용하여 동일한 대상(표본)을 조사<br>2) 대안법(Alternative-form Method) 이라고도 함<br>3) 두 가지 측정도구의 유사성을 담보하기 어려우며, 측정결과의 신뢰도가 낮을 경우 신뢰도 문제인지, 측정 도구의 유사성 문제인지 확인 어려움 |

| 내적 일관성 평가 지표(Criterion) | 일반적인 기준 | JMP Menu |
|---|---|---|
| 크론바흐 알파 (Cronbach's Alpha) | > 0.7(최소 >0.6) | 다변량 분석(Analyze / Multivariate Methods / Multivariate) |
| 합성신뢰도 (Composite Reliability : CR) | > 0.7(최소 >0.6) | 구조 방정식(Analyze / Multivariate Methods / Structural Equation Model) |
| D-H rho_A (Dijkstra-Henseler's rho_A) | > 0.7(최소 >0.6) | 구조 방정식(Analyze / Multivariate Methods / Structural Equation Model) |

신 익주(ikju.shin@jmp.com)

# 1. 탐색 : 설문 데이터 살펴보기






* 기업 구성원의 데이터 리터러시 측정 도구 개발 및 타당성 탐색(신익주, 박승욱, 대한경영학회지, Vol.38 No.3 [2025])

신 익주(ikju.shin@jmp.com)

# 1. 탐색 : 설문 데이터 살펴보기

## 2) 신뢰도와 타당도 분석

| 구성 요소 | 설문 문항 | |
|---|---|---|
| 데이터 중요성에 대한 인식 | DL1-1 | 나는 나의 업무에 필요한 데이터가 무엇인지 이해하고 정의할 수 있다 |
| | DL1-2 | 나는 데이터의 역할과 활용 가능성을 잘 이해하고 있다 |
| | DL1-3 | 나는 데이터가 개인의 업무 성과에 중요한 영향을 미친다고 생각한다 |
| | DL1-4 | 나는 실제 상황에서 문제를 해결하기 위하여 데이터를 사용한다 |
| | DL1-5 | 나는 데이터와 분석 결과를 활용하여 문제를 정의할 수 있다 |
| | DL1-6 | 나는 문제 해결을 위하여 어떤 데이터가 필요한 지 판단할 수 있다 |
| 데이터 해석 | DL2-1 | 나는 표, 그래프(차트)에서 부정확한 표현, 오류를 식별할 수 있다 |
| | DL2-2 | 나는 데이터를 통해 특정한 경향 (트렌드) 및 패턴을 파악할 수 있다 |
| | DL2-3 | 나는 분석 결과를 해석하고 주요 의미를 도출할 수 있다 |
| | DL2-4 | 나는 데이터의 출처(소스)가 신뢰할 수 있는 지 판단할 수 있다 |
| | DL2-5 | 나는 분석 목적에 비추어 데이터가 충분한 품질을 확보하고 있는 지 판별할 수 있다 |
| | DL2-6 | 나는 데이터가 얻어진 맥락과 관련 정보를 연계하여 해석할 수 있다 |
| 데이터 수집과 처리 | DL3-1 | 나는 필요한 데이터를 수집하는 가장 적합한 방법을 결정하고 적용할 수 있다 |
| | DL3-2 | 나는 다양한 소스에서 필요한 데이터를 추출하고 통합할 수 있다 |
| | DL3-3 | 나는 실험, 설문 조사 등을 통해 필요한 데이터를 수집할 수 있다 |
| | DL3-4 | 나는 이상치를 식별하고 처리할 수 있다 |
| | DL3-5 | 나는 결측치를 식별하고 처리할 수 있다 |
| | DL3-6 | 나는 데이터의 형식을 변환하여 활용 가능한 형태로 만들 수 있다 |
| 데이터로 분석하기 | DL4-1 | 나는 다양한 분석 도구(소프트웨어)와 분석 기법에 대해 알고 있다 |
| | DL4-2 | 나는 분석 목적에 부합하는 적절한 데이터 분석 방법을 선택할 수 있다 |
| | DL4-3 | 나는 기본적인 데이터 분석을 수행할 수 있다 |
| | DL4-4 | 나는 선택한 분석 도구(소프트웨어)를 활용하여 필요한 분석을 할 수 있다 |
| | DL4-5 | 나는 필요한 분석 결과를 얻기 위해 다양한 분석 방법을 활용하여 결과를, 비교할 수 있다 |
| | DL4-6 | 나의 분석 결과와 다른 사람의 분석 결과를 비교, 평가할 수 있다 |
| 데이터로 의사소통 하기 | DL5-1 | 나는 데이터의 핵심 내용을 표, 그래프(차트)로 시각화할 수 있다 |
| | DL5-2 | 나는 데이터의 핵심 내용을 적절한 통계량으로 요약하여 표현할 수 있다 |
| | DL5-3 | 나는 여러 종류의 표, 그래프(차트) 각각의 특징을 이해하고 있다 |
| | DL5-4 | 나는 여러가지 표, 그래프(차트) 중에서 분석 목적에 적합한 것을 선택할 수 있다 |
| | DL5-5 | 나는 분석 결과를 활용하여 나의 주장을 논리적으로 설명할 수 있다 |
| | DL5-6 | 나는 분석 결과를 정리하여 보고서를 만들 수 있다 |

N 200, Likert 5 Scale

① 구성(개념) 타당도
: EFA(탐색적 요인 분석)

② 신뢰도
: Cronbach's Alpha

신 익주(ikju.shin@jmp.com)

# 1. 탐색 : 설문 데이터 살펴보기



## ① 구성(개념) 타당도 : EFA(탐색적 요인 분석)

**1차 EFA : Maximum Likelihood / Varimax**

| | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|---|---|---|---|---|
| DL3-4 | 0.7106027 | 0.1516241 | 0.1088697 | 0.3131901 |
| DL3-5 | 0.6891152 | 0.0809649 | 0.0802773 | 0.3651392 |
| DL2-5 | 0.6718044 | 0.3472169 | 0.2157194 | 0.2053770 |
| DL2-4 | 0.6576382 | 0.2837434 | 0.1694344 | 0.1997880 |
| DL2-6 | 0.5425643 | 0.3659034 | 0.3713172 | 0.2569452 |
| DL2-1 | 0.5418220 | 0.3920463 | 0.1963958 | 0.2271897 |
| DL2-3 | 0.5390255 | 0.3917690 | 0.3432746 | 0.1840568 |
| DL3-6 | 0.5144455 | 0.3447655 | 0.3127526 | 0.4630907 |
| DL2-2 | 0.4995090 | 0.3871791 | 0.3709882 | 0.1954012 |
| DL3-3 | 0.4952705 | 0.2533640 | 0.2362896 | 0.2824267 |
| DL3-1 | 0.4793620 | 0.3557677 | 0.3283694 | 0.2768223 |
| DL3-2 | 0.4373379 | 0.4297013 | 0.3430555 | 0.3328381 |
| DL5-1 | 0.2721336 | 0.7438667 | 0.2422394 | 0.1883006 |
| DL5-4 | 0.2973705 | 0.6597202 | 0.3200822 | 0.3263314 |
| DL5-2 | 0.3706329 | 0.6396066 | 0.2630681 | 0.3003598 |
| DL5-6 | 0.2092195 | 0.6297523 | 0.3386366 | 0.3459389 |
| DL5-3 | 0.3203154 | 0.6134670 | 0.2630918 | 0.3626435 |
| DL5-5 | 0.3608473 | 0.5748309 | 0.3179501 | 0.3096334 |
| DL1-1 | 0.1296896 | 0.1669721 | 0.8067845 | 0.1305218 |
| DL1-2 | 0.1485929 | 0.2335187 | 0.7848598 | 0.1458844 |
| DL1-6 | 0.3274649 | 0.1431795 | 0.6509896 | 0.2187594 |
| DL1-5 | 0.2670617 | 0.2299497 | 0.5777085 | 0.2147705 |
| DL1-3 | 0.0793000 | 0.1506993 | 0.5642196 | 0.0738067 |
| DL1-4 | 0.1193105 | 0.1833219 | 0.5531155 | 0.1884134 |
| DL4-2 | 0.3582726 | 0.2491024 | 0.2861560 | 0.6709632 |
| DL4-5 | 0.4008904 | 0.3440417 | 0.2631829 | 0.6458609 |
| DL4-4 | 0.3888109 | 0.2792221 | 0.1599064 | 0.6375833 |
| DL4-3 | 0.3154306 | 0.3575479 | 0.3137520 | 0.5879720 |
| DL4-1 | 0.3559102 | 0.3303773 | 0.1699060 | 0.5661233 |
| DL4-6 | 0.3718333 | 0.3751398 | 0.3296833 | 0.5302253 |

**2차 EFA : Maximum Likelihood / Varimax**

| | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|---|---|---|---|---|
| DL3-4 | 0.7307809 | 0.1762684 | 0.1272479 | 0.2810766 |
| DL3-5 | 0.7038185 | 0.0925631 | 0.0922357 | 0.3491590 |
| DL2-5 | 0.6539941 | 0.3570215 | 0.2288335 | 0.2218273 |
| DL2-4 | 0.6386631 | 0.3002375 | 0.1832703 | 0.2061470 |
| DL2-1 | 0.5346348 | 0.4018034 | 0.2080047 | 0.2291833 |
| DL3-6 | 0.5046324 | 0.3548303 | 0.3197395 | 0.4587479 |
| DL2-6 | 0.4953359 | 0.3726927 | 0.3757921 | 0.2937627 |
| DL2-3 | 0.4853369 | 0.3966618 | 0.3457521 | 0.2280084 |
| DL5-1 | 0.2447143 | 0.7540984 | 0.2486049 | 0.1982261 |
| DL5-4 | 0.2830909 | 0.6666144 | 0.3266976 | 0.3235423 |
| DL5-2 | 0.3474827 | 0.6511519 | 0.2697258 | 0.3073002 |
| DL5-6 | 0.1844298 | 0.6270660 | 0.3388948 | 0.3562108 |
| DL5-3 | 0.3077300 | 0.6113103 | 0.2667775 | 0.3705184 |
| DL5-5 | 0.3526927 | 0.5773590 | 0.3246279 | 0.3080789 |
| DL1-1 | 0.1121104 | 0.1618943 | 0.8076575 | 0.1442718 |
| DL1-2 | 0.1312717 | 0.2323333 | 0.7912554 | 0.1524777 |
| DL1-6 | 0.2993052 | 0.1494687 | 0.6524075 | 0.2428089 |
| DL1-5 | 0.2511661 | 0.2377008 | 0.5822006 | 0.2165361 |
| DL1-3 | 0.0824135 | 0.1534539 | 0.5704424 | 0.0627910 |
| DL1-4 | 0.0976459 | 0.1903925 | 0.5531483 | 0.1933354 |
| DL4-2 | 0.3452431 | 0.2477839 | 0.2829847 | 0.6830076 |
| DL4-5 | 0.3856566 | 0.3459142 | 0.2626465 | 0.6576618 |
| DL4-4 | 0.3903947 | 0.2889482 | 0.1617499 | 0.6253448 |
| DL4-3 | 0.2856035 | 0.3559643 | 0.3070987 | 0.6144618 |
| DL4-6 | 0.3397469 | 0.3744047 | 0.3267667 | 0.5609070 |
| DL4-1 | 0.3652108 | 0.3319886 | 0.1725649 | 0.5540190 |

**Measures of Fit**

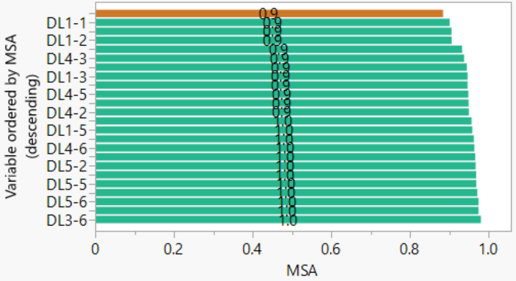
| Measures of Fit | Fit Index |
|---|---|
| Chi-Square without Bartlett's Correction | 311.867 |
| AIC | -60.133 |
| BIC | -673.620 |
| Tucker and Lewis's Index | 0.952 |
| Root Mean Square Error of Approximation | 0.058 |

**Bartlett's Test of Sphericity**

| ChiSquare | DF | Prob>ChiSq |
|---|---|---|
| 3635.921 | 276 | <.0001* |

**Kaiser-Meyer-Olkin Test**

| | MSA (Measure of Sampling Adequacy) |
|---|---|
| DL1-1 | 0.901 |
| DL1-2 | 0.905 |
| DL1-3 | 0.946 |
| DL1-4 | 0.932 |
| DL1-5 | 0.958 |
| DL1-6 | 0.965 |
| DL4-1 | 0.971 |
| DL4-2 | 0.949 |
| DL4-3 | 0.937 |
| DL4-4 | 0.946 |
| DL4-5 | 0.948 |
| DL4-6 | 0.963 |




,,, Overall MSA* 0.950

Legend
- 0.00 to 0.49 unacceptable
- 0.50 to 0.59 miserable
- 0.60 to 0.69 mediocre
- 0.70 to 0.79 middling
- 0.80 to 0.89 meritorious
- 0.90 to 1.00 marvelous

* MSA(Measures of Sampling Adequacy, 샘플링 적합도 측도)

신 익주(ikju.shin@jmp.com)

# 1. 탐색 : 설문 데이터 살펴보기



② 신뢰도 : Cronbach’s Alpha, 전체 & Construct 별

**Cronbach's α**

| | α |
|---|---|
| Entire set | 0.9683 |

| Excluded Col | α |
|---|---|
| DL1-1 | 0.9680 |
| DL1-2 | 0.9678 |
| DL1-3 | 0.9688 |
| DL1-4 | 0.9684 |
| DL1-5 | 0.9677 |
| DL1-6 | 0.9676 |
| DL2-1 | 0.9673 |
| DL2-2 | 0.9671 |
| DL2-3 | 0.9670 |
| DL2-4 | 0.9674 |
| DL2-5 | 0.9671 |
| DL2-6 | 0.9669 |
| DL3-1 | 0.9671 |
| DL3-2 | 0.9668 |
| DL3-3 | 0.9677 |
| DL3-4 | 0.9675 |
| DL3-5 | 0.9678 |
| DL3-6 | 0.9665 |
| DL4-1 | 0.9674 |
| DL4-2 | 0.9669 |
| DL4-3 | 0.9668 |
| DL4-4 | 0.9671 |
| DL4-5 | 0.9666 |
| DL4-6 | 0.9667 |
| DL5-1 | 0.9672 |
| DL5-2 | 0.9667 |
| DL5-3 | 0.9668 |
| DL5-4 | 0.9667 |
| DL5-5 | 0.9668 |
| DL5-6 | 0.9670 |

**Multivariate**

**Correlations**

| | DL1-1 | DL1-2 | DL1-3 | DL1-4 | DL1-5 | DL1-6 |
|---|---|---|---|---|---|---|
| DL1-1 | 1.0000 | 0.7594 | 0.4702 | 0.4880 | 0.5334 | 0.5913 |
| DL1-2 | 0.7594 | 1.0000 | 0.4793 | 0.4821 | 0.5465 | 0.6188 |
| DL1-3 | 0.4702 | 0.4793 | 1.0000 | 0.4388 | 0.5077 | 0.4528 |
| DL1-4 | 0.4880 | 0.4821 | 0.4388 | 1.0000 | 0.5725 | 0.4869 |
| DL1-5 | 0.5334 | 0.5465 | 0.5077 | 0.5725 | | |
| DL1-6 | 0.5913 | 0.6188 | 0.4528 | 0.4869 | | |

The correlations are estimated by Row-wise method.

**Cronbach's α**

| | α |
|---|---|
| Entire set | 0.8720 |

| Excluded Col | α |
|---|---|
| DL1-1 | 0.8420 |
| DL1-2 | 0.8386 |
| DL1-3 | 0.8650 |
| DL1-4 | 0.8595 |
| DL1-5 | 0.8469 |
| DL1-6 | 0.8474 |

**Multivariate**

**Correlations**

| | DL4-1 | DL4-2 | DL4-3 | DL4-4 | DL4-5 | DL4-6 |
|---|---|---|---|---|---|---|
| DL4-1 | 1.0000 | 0.6604 | 0.6127 | 0.6756 | 0.6314 | 0.5600 |
| DL4-2 | 0.6604 | 1.0000 | 0.7013 | 0.6341 | 0.7610 | 0.6740 |
| DL4-3 | 0.6127 | 0.7013 | 1.0000 | 0.6855 | 0.6716 | 0.7235 |
| DL4-4 | 0.6756 | 0.6341 | 0.6855 | 1.0000 | 0.7083 | 0.6087 |
| DL4-5 | 0.6314 | 0.7610 | 0.6716 | 0.7083 | 1.0000 | 0.7622 |
| DL4-6 | 0.5600 | 0.6740 | 0.7235 | 0.6087 | 0.7622 | 1.0000 |

The correlations are estimated by Row-wise method.

**Cronbach's α**

| | α |
|---|---|
| Entire set | 0.9234 |

| Excluded Col | α |
|---|---|
| DL4-1 | 0.9183 |
| DL4-2 | 0.9063 |
| DL4-3 | 0.9080 |
| DL4-4 | 0.9105 |
| DL4-5 | 0.9030 |
| DL4-6 | 0.9108 |

신 익주(ikju.shin@jmp.com)

# 2. 실험 : 마케팅 분야 적용

## 1) 의사 결정을 지원하는 JMP의 마케팅 기법들

### Design

**DOE / Consumer Studies**

- Choice Design
- MaxDiff Design

### Analyze

**Analyze / Multivariate Methods**

- Multivariate
- Principal Components
- Discriminant
- Partial Least Squares
- Multiple Correspondence Analysis
- Structural Equation Models
- Factor Analysis
- Multidimensional Scaling
- Multivariate Embedding
- Item Analysis

**Analyze / Screening**

- Association Analysis

**Analyze / Clustering**

- Hierarchical Cluster
- K Means Cluster
- Normal Mixtures
- Latent Class Analysis
- Cluster Variables

**Analyze / Consumer Research**

- Categorical
- Choice
- MaxDiff
- Uplift
- Multiple Factor Analysis

### Utility

1. Data Table
1) Modeling Type
2) Header Graph/Statistics
3) Data Table Filter(JMP19)

2. Column Property
1) Multiple Response
2) Super-category
3) Recode
4) Value Label

3. Others
1) Data Filter
2) Cronbach's α
3) Exploring Missing Value
4) Freq, Weight
5) Forced Value(JMP19)
,,,

신 익주(ikju.shin@jmp.com)

# 2. 실험 : 마케팅 분야 적용

## 1) 의사 결정을 지원하는 JMP의 마케팅 기법들

| 메뉴 | | 내용 |
|---|---|---|
| DOE / Consumer Studies | Choice Design | 1) 컨조인트 분석을 위한 디자인<br>2) 다양한 속성을 가진 잠재적 제품 집합(Product Profile)의 선택 |
| | MaxDiff Design | 1) 최대 차이 분석을 위한 설계<br>2) 소수의 선택 집합 중에서 가장 선호하는 옵션과 가장 선호하지 않는 옵션 디자인 |
| Analyze / Screening | Association Analysis | 1) 장바구니(Market Basket) 분석<br>2) 품목간의 인접성(친화성, Affinity)을 분석하여 상품의 Display, Package 상품의 구성, 할인 품목 선정 등의 용도로 많이 활용 |
| Analyze / Consumer Research | Categorical | 1) 일반적인 설문 데이터에 대한 분석 기능을 제공<br>2) 범주형 데이터를 다양한 측면에서 요약/시각화 할 수 있음 |
| | Choice | 1) 컨조인트(Conjoint) 분석<br>2) 제품 속성의 변화에 따른 소비자의 최적 선택을 측정하여 속성의 값어치를 측정 |
| | MaxDiff | 1) 표준 선호도 척도를 사용하여 평가 항목의 상대 중요도를 결정하는 방법의 대안<br>2) 응답자는 최선과 최악을 선택하여야 함(순위 결정) |
| | Uplift | 1) 호의적으로 반응할 가능성이 높은 개인에게만 제안을 보내 마케팅 예산의 영향을 최대화 할 수 있음<br>2) 데이터 집합이 크고 , 가능한 행동 또는 인구 통계 예측 변수가 많은 경우에 많이 활용 |
| | Multiple Factor Analysis | 1) 다중그룹에 의한 다중변수에 따른 요인분석을 실시함<br>2) 감각 데이터 분석에서 참가자 간의 합치도를 분석할 수 있음 |

# 2. 실험 : 마케팅 분야 적용



## 컨조인트 분석 : 돼지고기와 술 속성에 대한 가상 사례


1. 4 Attributes
1) 돼지고기 : 삼겹살, 목살, 다리살
2) 술 : 소주, 맥주, 막걸리
3) 후식 : 된장밥, 냉면, 국수
4) 김치 : 일반김치, 묵은지

**Choice Design**

**Attributes**

| Name | Role | Values | | |
|---|---|---|---|---|
| 돼지고기 | Categorical | 삼겹살 | 목살 | 다리살 |
| 술 | Categorical | 소주 | 맥주 | 막걸리 |
| 후식 | Categorical | 된장밥 | 냉면 | 국수 |
| 김치 | Categorical | 일반김치 | 묵은지 | |

2. 실험 비교
1) 완전 배치 : 54 회
2) 부분 배치 : 18 회 등

| Number Of Runs | Design Type |
|---|---|
| 18 | L18 - John |
| 18 | L18 - Taguchi |

**Design Generation**

| | |
|---|---|
| 4 | Number of attributes that can change within a choice set |
| 3 | Number of profiles per choice set |
| 12 | Number of choice sets per survey |
| 1 | Number of surveys |
| 20 | Expected number of respondents per survey |

3) Conjoint Design
-12개(←54)의 Choice Set 구성
-Choice Set 당 3가지 Profile 중 택 1
-20명 응답자

# 2. 실험 : 마케팅 분야 적용



컨조인트 분석 : 돼지고기와 술 속성에 대한 가상 사례

| | Respondent | Choice Set | Response Indicator | 돼지고기 | 술 | 후식 | 김치 | Choice Simulated |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | • | 목살 | 소주 | 냉면 | 일반김치 | 0 |
| 2 | 1 | 1 | • | 삼겹살 | 맥주 | 된장밥 | 묵은지 | 0 |
| 3 | 1 | 1 | • | 목살 | 막걸리 | 국수 | 묵은지 | 1 |
| 4 | 1 | 2 | • | 목살 | 맥주 | 국수 | 묵은지 | 0 |
| 5 | 1 | 2 | • | 삼겹살 | 소주 | 된장밥 | 묵은지 | 1 |
| 6 | 1 | 2 | • | 다리살 | 소주 | 냉면 | 일반김치 | 0 |

*Profile*

*Choice Set*

''''''

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 34 | 1 | 12 | • | 삼겹살 | 소주 | 냉면 | 묵은지 | 1 |
| 35 | 1 | 12 | • | 목살 | 맥주 | 된장밥 | 일반김치 | 0 |
| 36 | 1 | 12 | • | 삼겹살 | 막걸리 | 국수 | 일반김치 | 0 |

''''''

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 718 | 20 | 12 | • | 삼겹살 | 소주 | 냉면 | 묵은지 | 1 |
| 719 | 20 | 12 | • | 목살 | 맥주 | 된장밥 | 일반김치 | 0 |
| 720 | 20 | 12 | • | 삼겹살 | 막걸리 | 국수 | 일반김치 | 0 |

20 Respondents
12 Choice Sets
3 Profiles
→ 720 Row Data

신 익주(ikju.shin@jmp.com)

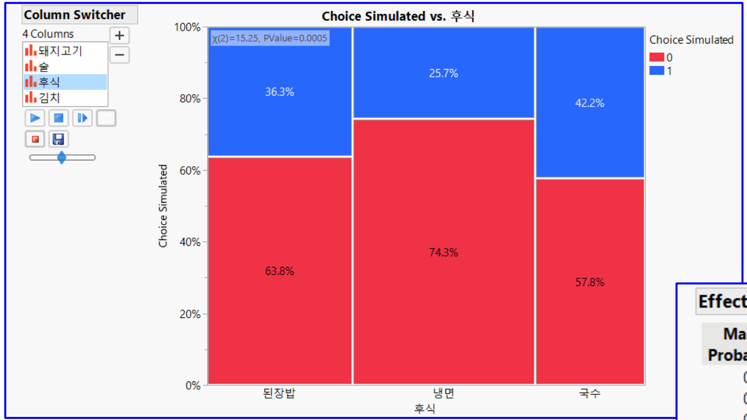
# 2. 실험 : 마케팅 분야 적용

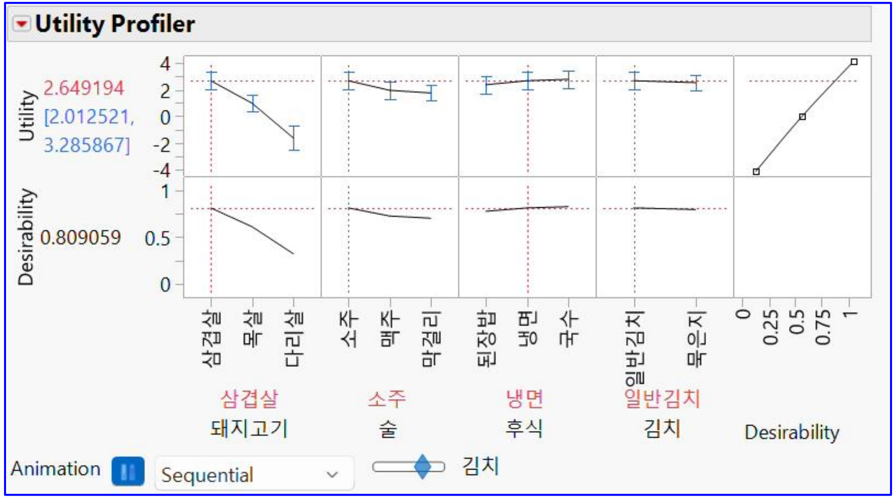


1) Effect Summary : 돼지고기와 술은 유의, 후식과 김치는 종류는 유의하지 않음
2) Marginal Utility(Parameter Estimates, Part-worth Utility)
  -돼지고기의 경우 삼겹살이 가장 선호, 다리살은 비선호, 술은 소주가 선호, 막걸리가 가장 비선호
  -후식과 김치의 한계 효용 차이는 아주 제한적

**Effect Summary**

| Source | Logworth | | PValue |
|---|---|---|---|
| 돼지고기 | 53.046 | | 0.00000 |
| 술 | 4.660 | | 0.00002 |
| 후식 | 0.593 | | 0.25543 |
| 김치 | 0.321 | | 0.47800 |




**Effect Marginals <--부분 가치(part-worths)**

| Marginal Probability | Marginal Utility | | 돼지고기 |
|---|---|---|---|
| 0.8323 | 1.9774 | | 삼겹살 |
| 0.1559 | 0.3027 | | 목살 |
| 0.0118 | -2.2801 | | 다리살 |

| Marginal Probability | Marginal Utility | | 술 |
|---|---|---|---|
| 0.5285 | 0.54155 | | 소주 |
| 0.2591 | -0.17134 | | 맥주 |
| 0.2124 | -0.37021 | | 막걸리 |

| Marginal Probability | Marginal Utility | | 후식 |
|---|---|---|---|
| 0.2606 | -0.23209 | | 된장밥 |
| 0.3498 | 0.06208 | | 냉면 |
| 0.3896 | 0.17001 | | 국수 |

| Marginal Probability | Marginal Utility | | 김치 |
|---|---|---|---|
| 0.5340 | 0.06815 | | 일반김치 |
| 0.4660 | -0.06815 | | 묵은지 |

* Marginal Probability(한계 확률) : 해당 수준이 선택될 예상 확률, 0~1
* Marginal Utility(한계 효용) : 해당 수준의 선호 강도(utility)

# 3. 해석 : 구조 방정식 적용 사례

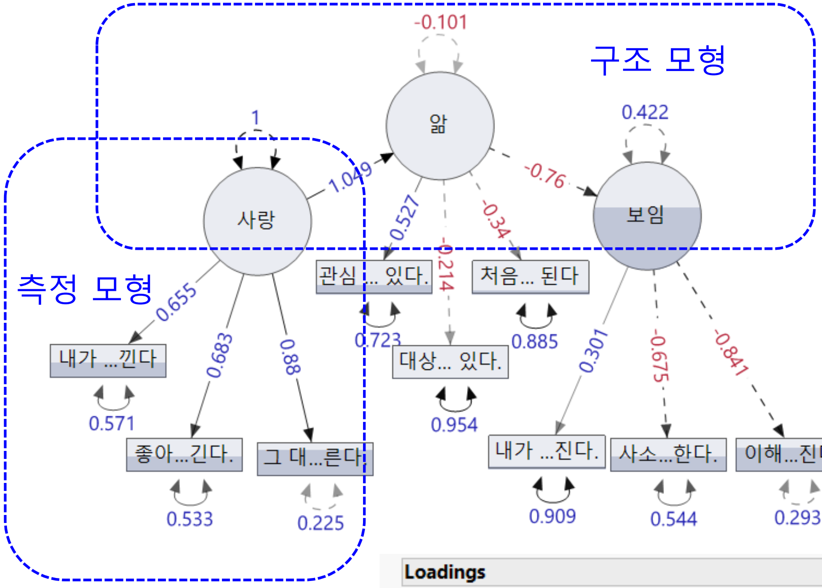


**구조 방정식(SEM : Structural Equation Model)**

1. 현실에서의 변수들간의 관계는 매우 복잡
1) 직접 관계(A→B) 뿐만 아니라 간접(A→B→C)적인
매개(mediating) 관계, 조절(moderating) 관계 등 다양
2) 개념(Construct)를 정확히 측정하기가 매우 어려움
예) 사랑하면 알게 되고, 알면 보이나니
그때 보이는 것은 전과 같지 않으리라*

2. 원리
1) 측정 모형(Measurement Model)
설문 문항들이 '개념'을 얼마나 잘 측정하는 지 확인
2) 구조 모형(Structural Model)
각 개념들 사이의 인과 관계를 추정
3) 요인 분석(측정 모형) + 회귀 분석(구조 모형)

3. CB(Covariance-based) vs PLS(Partial Least Square-based)
1) CB(Covariance-based) SEM : JMP, AMOS, LISREL 등
2) PLS(Partial Least Square-based) SEM : SmartPLS 등



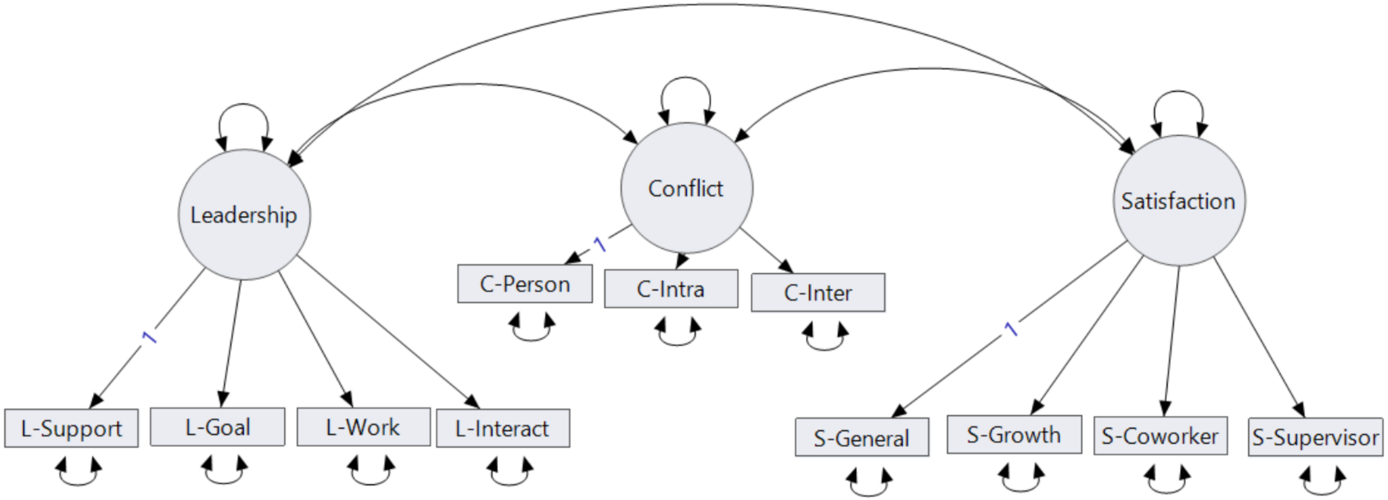

**Loadings**
사랑 → 내가 관심 있는 대상에 진심으로 애정을 느낀다(1)
사랑 → 좋아하는 대상에 대해 더 알고 싶다는 열망이 생긴다.
사랑 → 그 대상에 몰입하면 시간 가는 줄 모른다.
앎 → 관심 있는 대상의 특징과 원리를 이해하고 있다.(1)
앎 → 대상을 객관적으로 바라보고 설명할 수 있다.
앎 → 처음엔 몰랐던 의미를 점점 더 깊이 깨닫게 된다
보임 → 내가 보는 세상이 이전과 다르게 느껴진다.(1)
보임 → 사소한 것에서도 새로운 의미를 발견한다.
보임 → 이해를 통해 시야가 넓어지고 관점이 달라진다.

* 조선 정조 때 유 한준의 글 '知則爲眞愛 愛則爲眞看 看則畜之而非徒畜也'을 유 홍준 교수가 '나의 문화 유산답사기' 1권에서 수정하여 인용

신 익주(ikju.shin@jmp.com)

# 3. 해석 : 구조 방정식 적용 사례

Job Satisfaction.jmp
: 3개 Construct, 11개 측정 변수, N 200, Likert 7 Scale



# 3. 해석 : 구조 방정식 적용 사례



1) Unrestricted(Saturated) : 포화 모형으로 모든 측정변수들의 관계성을 허용, 잠재 변수 없음, DF 0
2) Independence : 독립모형으로 측정변수들의 관계성은 모두 독립으로 가정, 잠재 변수 없음
3) 위의 두 모형 대비 연구자가 가정한 잠재 변수가 적절한지 분석한 것이 확증적 요인 분석

**Model Comparison**

| | Model Name | -2 Log Likelihood | Number of Parameters | AICc | AICc Weight | .2 .4 .6 .8 | BICu | ChiSquare | DF | Prob>ChiSq | CFI | RMSEA | Lower 90% | Upper 90% | SRMR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Unrestricted (Saturated) | 6708.0203 | 77 | 6960.4793 | 0.0000 | | 0.0000 | 0.0000 | 0 | . | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 2 | Independence | 7278.1696 | 22 | 7327.8871 | 0.0000 | | 278.7418 | 570.1493 | 55 | <.0001* | 0.0000 | 0.2164 | 0.2005 | 0.2327 | 0.2859 |
| 3 | 3 Factor CFA | 6742.2869 | 36 | 6830.6305 | 1.0000 | | -182.9644 | 34.2666 | 41 | 0.7624 | 1.0000 | 0.0000 | 0.0000 | 0.0350 | 0.0308 |

**Summary of Fit**

Maximum Likelihood. Converged in Gradient.

| | |
|---|---|
| Sample Size | 200.0000 |
| Rows with Missing | 0.0000 |
| -2 Log Likelihood | 6742.2869 |
| Iterations | 9.0000 |
| Number of Parameters | 36.0000 |
| AICc | 6830.6305 |
| BICu | -182.9644 |
| ChiSquare | 34.2666 |
| DF | 41.0000 |
| Prob>ChiSq | 0.7624 |
| CFI | 1.0000 |
| RMSEA | 0.0000 |
| Lower 90% | 0.0000 |
| Upper 90% | 0.0350 |

### 1) Prob>Chisq

-유의하지 않을수록 확증적 요인 분석이 해석 가능함을 의미
-Ho : "지정한 모형(구조)이 관측된 공분산 구조와 통계적으로 유의미하게 다르지 않다"

### 2) 상대적 적합도 지수(CFI : Comparative Fit Index)

-모든 변수들이 독립이라고 가정한 모형보다 얼마나 더 적합한지를 평가하는 지수이다.
-1에 근접할수록 우수, 보통 0.9 이상이면 양호하다고 해석

### 3) RMSEA(Root Mean Square Error of Approximation

-수집된 공분산 구조와 이론적 모형 사이의 근사 적합도를 측정하며 값이 작을수록 적합하다고 해석
-0.05 이하이면 매우 우수, 0.1보다 크면 부적합하다고 평가

신 익주(ikju.shin@jmp.com)

# 3. 해석 : 구조 방정식 적용 사례

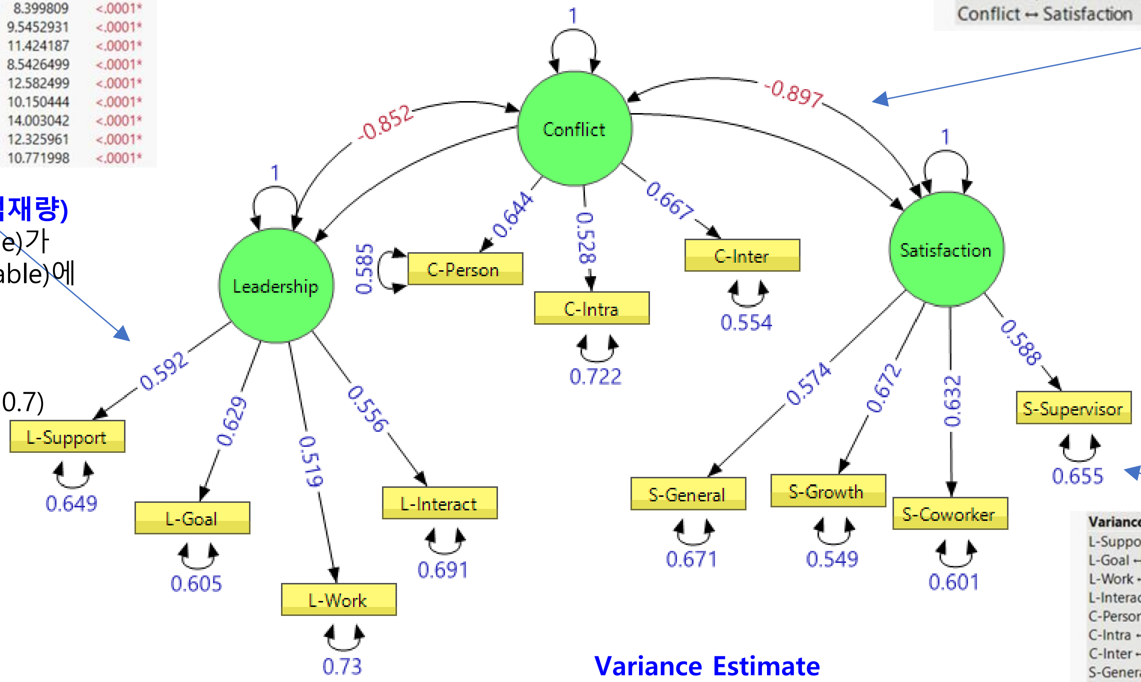
## 2) 확정적 요인 분석

| Loadings | Estimate | Std Error | Wald Z | Prob>\|Z\| |
|---|---|---|---|---|
| Leadership → L-Support | 0.5920495 | 0.0567471 | 10.433127 | <.0001* |
| Leadership → L-Goal | 0.6288604 | 0.0535719 | 11.738625 | <.0001* |
| Leadership → L-Work | 0.5192589 | 0.0618179 | 8.399809 | <.0001* |
| Leadership → L-Interact | 0.5560815 | 0.0582571 | 9.5452931 | <.0001* |
| Conflict → C-Person | 0.6439065 | 0.0563634 | 11.424187 | <.0001* |
| Conflict → C-Intra | 0.5275505 | 0.0617549 | 8.5426499 | <.0001* |
| Conflict → C-Inter | 0.6674685 | 0.0530474 | 12.582499 | <.0001* |
| Satisfaction → S-General | 0.5739112 | 0.0565405 | 10.150444 | <.0001* |
| Satisfaction → S-Growth | 0.6719171 | 0.0479837 | 14.003042 | <.0001* |
| Satisfaction → S-Coworker | 0.6316909 | 0.0512488 | 12.325961 | <.0001* |
| Satisfaction → S-Supervisor | 0.5875893 | 0.0545478 | 10.771998 | <.0001* |

**Loading Estimate(요인 적재량)**
1) 잠재 변수(latent variable)가 관측 변수(measured variable)에 미치는 영향력, 설명력 (잠재 변수의 관측 변수에 대한 측정 적합도)
2) 권고(>0.5), 매우 좋음(>0.7)

| Covariances | Estimate | Std Error | Wald Z | Prob>\|Z\| |
|---|---|---|---|---|
| Leadership ↔ Conflict | -0.852045 | 0.0682055 | -12.49232 | <.0001* |
| Leadership ↔ Satisfaction | 0.9380916 | 0.0534906 | 17.53751 | <.0001* |
| Conflict ↔ Satisfaction | -0.897115 | 0.0598796 | -14.98198 | <.0001* |

**Covariance Estimate**
1) 잠재변수간 관계의 크기
2) 인과성이 아니라 연관성




**Variance Estimate**
1) 해당 변수가 가진 변량의 크기
(설명되지 않는 변량-고유 변량-포함)
2) 요인 적재량이 높으면 이 값이 낮음

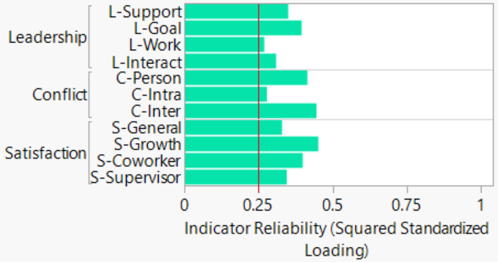
| Variances | Estimate | Std Error | Wald Z | Prob>\|Z\| |
|---|---|---|---|---|
| L-Support ↔ L-Support | 0.6494774 | 0.0671942 | 9.6656837 | <.0001* |
| L-Goal ↔ L-Goal | 0.6045346 | 0.0673785 | 8.9722202 | <.0001* |
| L-Work ↔ L-Work | 0.7303702 | 0.064199 | 11.376652 | <.0001* |
| L-Interact ↔ L-Interact | 0.6907733 | 0.0647914 | 10.66149 | <.0001* |
| C-Person ↔ C-Person | 0.5853844 | 0.0725856 | 8.0647475 | <.0001* |
| C-Intra ↔ C-Intra | 0.7216905 | 0.0651576 | 11.076068 | <.0001* |
| C-Inter ↔ C-Inter | 0.5544858 | 0.0708149 | 7.8300713 | <.0001* |
| S-General ↔ S-General | 0.6706259 | 0.0648985 | 10.333465 | <.0001* |
| S-Growth ↔ S-Growth | 0.5485274 | 0.0644821 | 8.5066655 | <.0001* |
| S-Coworker ↔ S-Coworker | 0.6009666 | 0.0647468 | 9.2817908 | <.0001* |
| S-Supervisor ↔ S-Supervisor | 0.6547388 | 0.0641035 | 10.213784 | <.0001* |
| Leadership ↔ Leadership | 1 | 0 | . | . |
| Conflict ↔ Conflict | 1 | 0 | . | . |
| Satisfaction ↔ Satisfaction | 1 | 0 | . | . |

# 3. 해석 : 구조 방정식 적용 사례



**Indicator Reliability**

| Latent Variable | Manifest Variable | Indicator Reliability |
|---|---|---|
| Leadership | L-Support | 0.3505 |
| Leadership | L-Goal | 0.3955 |
| Leadership | L-Work | 0.2696 |
| Leadership | L-Interact | 0.3092 |
| Conflict | C-Person | 0.4146 |
| Conflict | C-Intra | 0.2783 |
| Conflict | C-Inter | 0.4455 |
| Satisfaction | S-General | 0.3294 |
| Satisfaction | S-Growth | 0.4515 |
| Satisfaction | S-Coworker | 0.3990 |
| Satisfaction | S-Supervisor | 0.3453 |

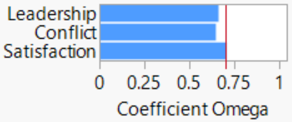



## Indicator Reliability(표시자 신뢰도)

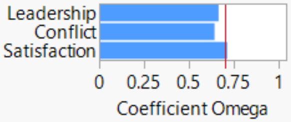
1) 표준화된 Loading 값의 제곱과 제안된 최소 허용 가능 신뢰도 임계치(0.25, 0.5의 제곱)를 나타냄
2) 모든 값이 임계치를 초과하므로 이 설문 결과를 신뢰할 수 있다고 판정할 수 있음

**Composite Reliability**

| Latent Variable | Standardized Omega |
|---|---|
| Leadership | 0.6618 |
| Conflict | 0.6474 |
| Satisfaction | 0.7077 |




| Latent Variable | Unstandardized Omega |
|---|---|
| Leadership | 0.6645 |
| Conflict | 0.6421 |
| Satisfaction | 0.7124 |

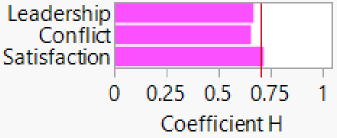



## Composite Reliability(종합 신뢰도, Macdonald Omega)와 Construct Maximal Reliability(구성 최대 신뢰도, Hancock H)

1) Omega 계수 및 H 계수는 보통 0.7 이상이 권장
2) Omega는 관측된 종합 스코어에서 잠재 변수의 분산 비율을
   H는 표시자(indicator)가 나타내는 잠재 변수의 분산 비율을 나타냄
3) Omega는 신뢰도 중 내적 일관성에 대한 평가 지표로 Cronbach Alpha와 함께 사용됨
   H는 AVE 값처럼 집중(수렴, convergent) 타당도를 평가
4) 지금은 제안된 임계값 0.7 보다 약간 낮지만 크게 문제될 수준은 아니라고 할 수 있음

**Construct Maximal Reliability**

| Latent Variable | H |
|---|---|
| Leadership | 0.6678 |
| Conflict | 0.6549 |
| Satisfaction | 0.7147 |



# 3. 해석 : 구조 방정식 적용 사례

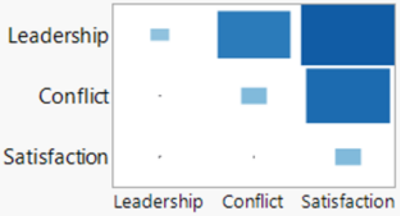


**Construct Validity Matrix(구성 타당도 행렬)**

사용자가 생각하는 대상을 잠재 변수가 측정하고 있는지 여부를 판별하는 **판별 타당성(discriminant validity)**을 나타냄

**Construct Validity Matrix**

| | Leadership | Conflict | Satisfaction |
|---|---|---|---|
| Leadership | 0.3312 | 0.7260 | 0.8800 |
| Conflict | -0.8520 | 0.3795 | 0.8048 |
| Satisfaction | 0.9381 | -0.8971 | 0.3813 |




*Latent variable (LV) correlations are below the diagonal, average variance extracted by the LVs are in the diagonal, and squared factor correlations of LVs are above the diagonal.*

상관 관계²

평균 분산 추출(AVE)

상관 관계

1) 왼쪽 하단 데이터는 잠재 변수간 상관 관계(계수)

| Covariances | Estimate |
|---|---|
| Leadership ↔ Conflict | -0.852045 |
| Leadership ↔ Satisfaction | 0.9380916 |
| Conflict ↔ Satisfaction | -0.897115 |

2) 오른쪽 상단 데이터는 잠재 변수간 상관 계수의 제곱

3) 대각선(붉은 선 표시) 값은 **평균 분산 추출(AVE : Average Variance Extracted) 값**. 이 값은 각 잠재 변수에 대한 Indicator Reliability(표시자 신뢰도) 값의 평균과 동일.
예) Leadership : (0.3505+0.3955+0.2696+0.3092) / 4 = 0.3312

**Indicator Reliability**

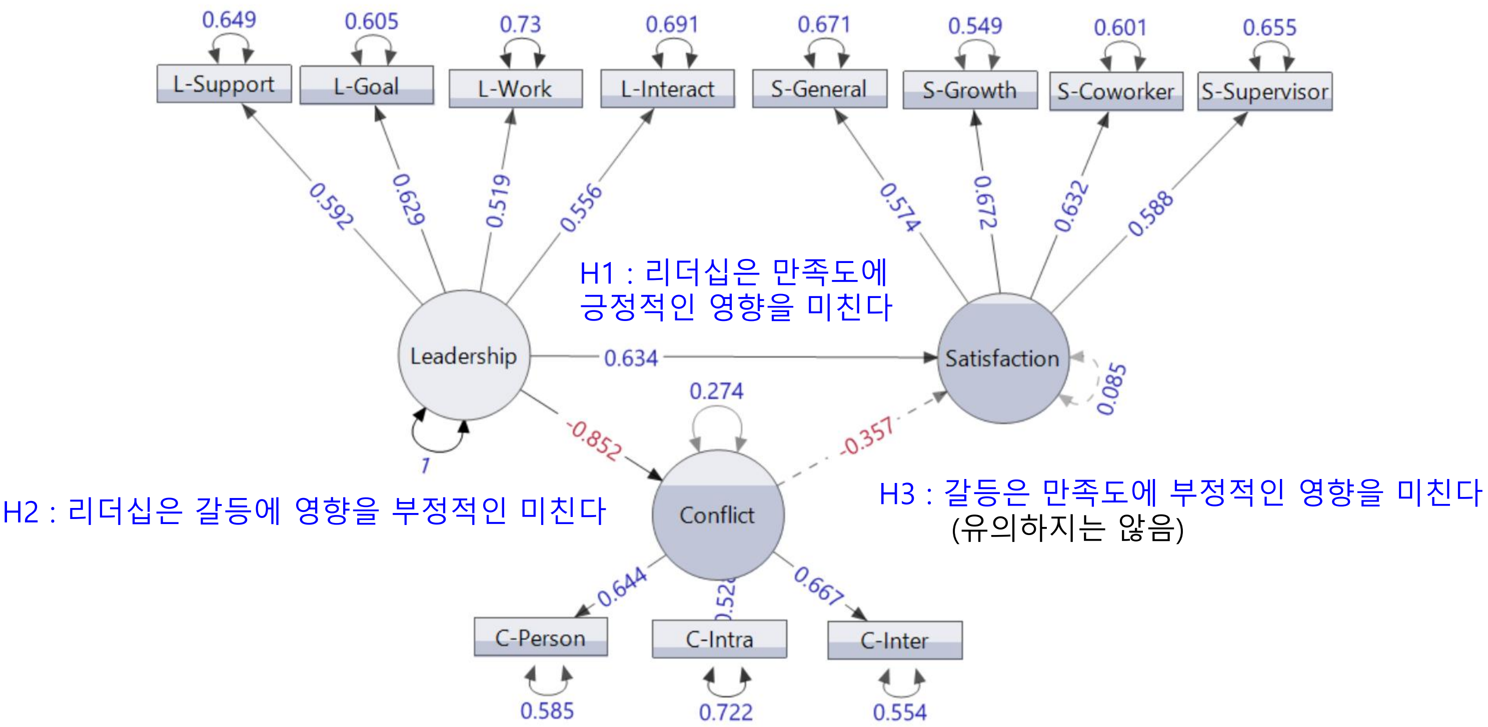
| Latent Variable | Manifest Variable | Indicator Reliability |
|---|---|---|
| Leadership | L-Support | 0.3505 |
| Leadership | L-Goal | 0.3955 |
| Leadership | L-Work | 0.2696 |
| Leadership | L-Interact | 0.3092 |

4) **수렴 타당성 기준 : AVE 값 ≥ 0.5**
**판별 타당성 기준(Fornell-Larcker 기준)**
**: AVE 값이 잠재 변수간 상관 계수의 제곱보다 커야 함**
→ 설문 항목의 수정, 보완 등의 조치가 필요한 상황

신 익주(ikju.shin@jmp.com)

# 3. 해석 : 구조 방정식 적용 사례



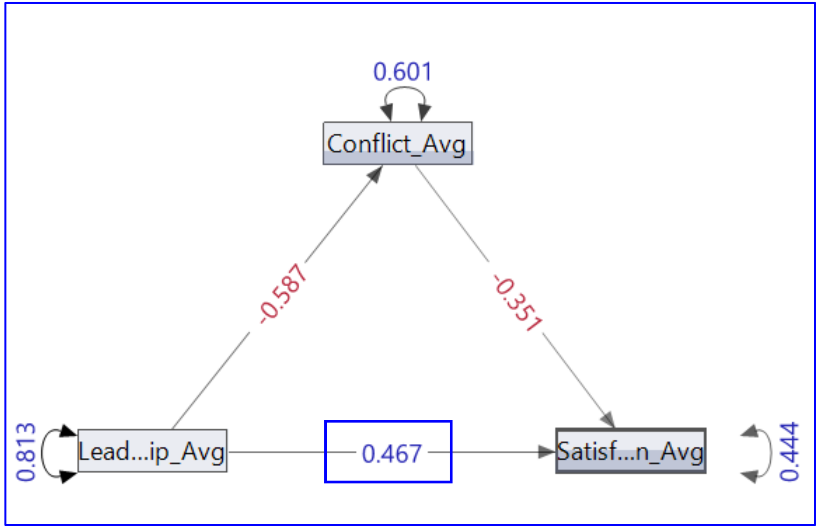



| Regressions | Estimate | Std Error | Wald Z | Prob>\|Z\| | Lower 95% | Upper 95% |
|---|---|---|---|---|---|---|
| Leadership → Conflict | -0.852045 | 0.0682055 | -12.49232 | <.0001* | -0.985725 | -0.718365 |
| Leadership → Satisfaction | 0.6339296 | 0.2475568 | 2.5607445 | 0.0104* | 0.1487273 | 1.119132 |
| Conflict → Satisfaction | -0.356979 | 0.2528214 | -1.41198 | 0.1580 | -0.8525 | 0.1385421 |

신 익주(ikju.shin@jmp.com)

# 3. 해석 : 구조 방정식 적용 사례






Leadership → Satisfaction

Total Effect(0.673)
= Direct Effect(0.467) + Indirection Effect(0.206)

**Total Effects**

| Predictor | Outcome | Estimate | Std Error | Wald Z | Prob>\|Z\| | Standardized Estimate |
|---|---|---|---|---|---|---|
| Leadership_Avg | Conflict_Avg | -0.587049 | 0.0608013 | -9.655198 | <.0001* | -0.563848 |
| Leadership_Avg | Satisfaction_Avg | 0.6730094 | 0.0564294 | 11.926572 | <.0001* | 0.6446861 |
| Conflict_Avg | Satisfaction_Avg | -0.351107 | 0.0607489 | -5.779645 | <.0001* | -0.350169 |

**Indirect Effects**

| Predictor | Outcome | Estimate | Std Error | Wald Z | Prob>\|Z\| | Standardized Estimate |
|---|---|---|---|---|---|---|
| Leadership_Avg | Satisfaction_Avg | 0.2061168 | 0.0415637 | 4.9590552 | <.0001* | 0.1974425 |

신 익주(ikju.shin@jmp.com)

# 마무리

**탐색, 실험, 해석**
**: JMP로 풀어보는 사회과학 데이터 이야기**

**1. Exploration(EDA) for Survey Data**
1) Visualization, Summarization,
Categorical Analysis
2) Validity, Reliability

**2. Experimentation**
**: Application in Marketing**
(Conjoint Analysis)

**3. Interpretation : Structural Equation Model**
1) Confirmative Factor Analysis
2) Path Analysis, Mediating effect

***From Discovery to Insight***
***– All in JMP***


신 익주(ikju.shin@jmp.com)