

# Confirmatory factor analysis

سید حمید شریف نیا  
هیات علمی دانشگاه علوم پزشکی مازندران

# Exploratory factor analysis

?

محقق به دنبال این است که بداند:

آیا همبستگی و یا ارتباطی بین متغیرهای حاضر وجود دارد یا نه؟

اگر وجود دارد، ارتباط بین کدام متغیرها بیشتر است؟

آیا می‌توان یک الگویی برای دسته‌بندی این متغیرها ایجاد کرد و آن‌ها را به گروه‌هایی مرتبط کرد؟

هر متغیر چقدر با گروهی که بدان اختصاص یافته مرتبط است؟

# روش استخراج داده‌ها

*Principal Components Analysis*

*Unweighted Least-Squares Method*

*Generalized Least-Squares Method*

*Maximum-Likelihood Method*

*Principal Axis Factoring*

*Alpha*

*Image Factoring*

# Structural Equation Modeling

چرا برخی از مردم علیرغم هشدارهای ستاد کرونا مبنی بر قرمز بودن برخی مناطق مانند شمال کشور، باز هم به سفر می روند؟ و یا اینکه علیرغم شدت گرفتن باز هم دور همی ها و مهمانی های شلوغ تشکیل می شود؟

افسردگی

اضطراب از کرونا

خستگی

مشکلات مالی

فشار روحی ناشی از قرنطینه

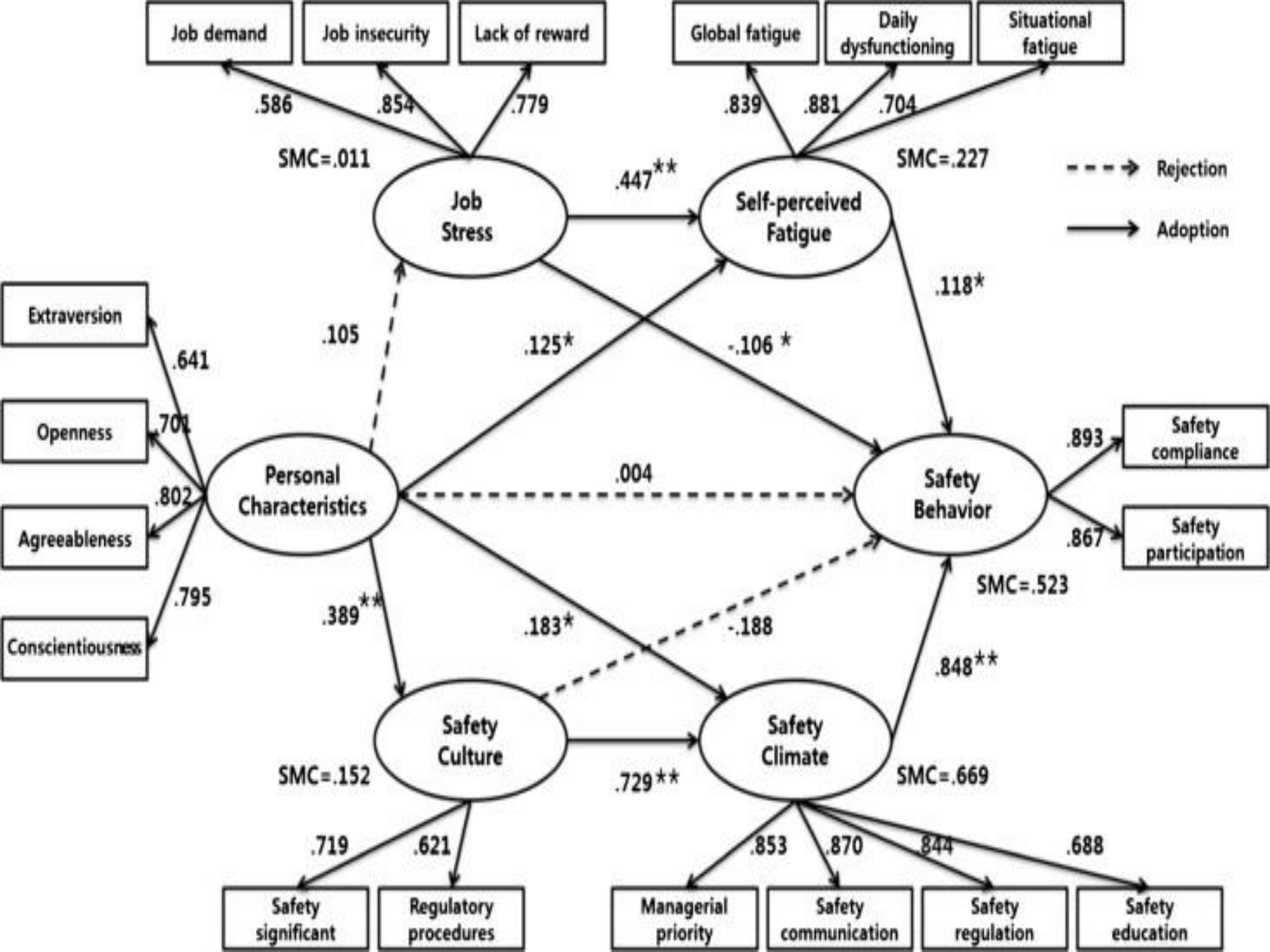
نگرش به رفتار بهداشتی

درک از ابتلا به بیماری عفونی

استعداد به ابتلا به بیماری های واگیر

درک از تعهد اجتماعی

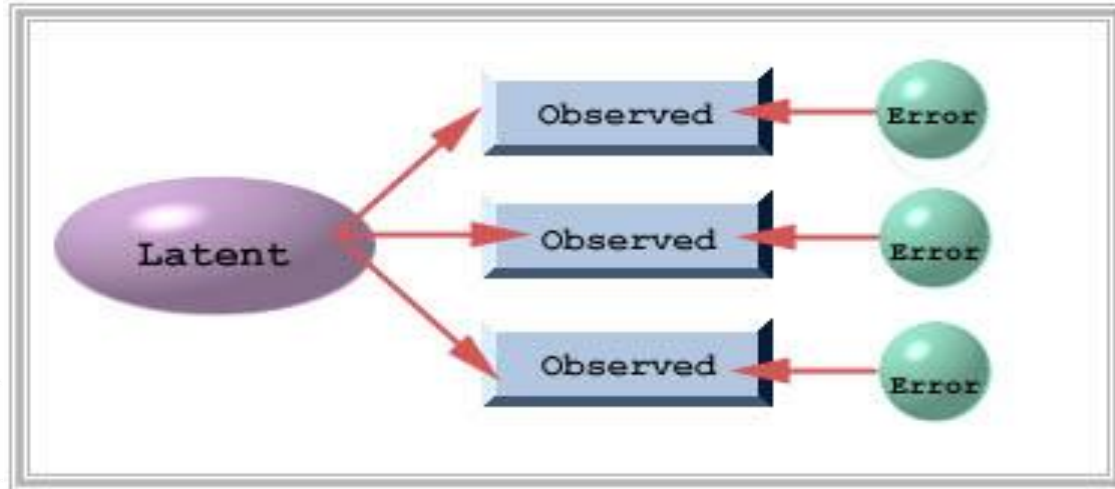
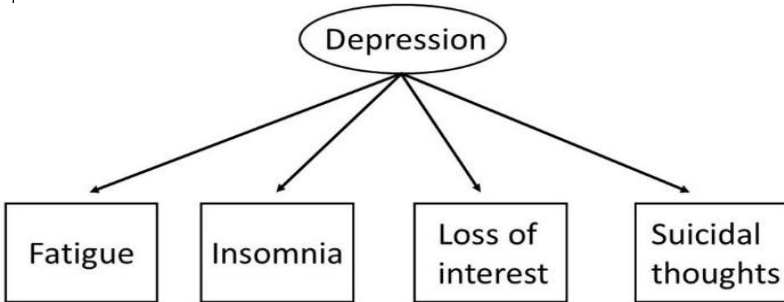
اختلال خلق



# SEM

Observed variables

Latent variables



**Exogenous and endogenous variables**

# CB-SEM vs. VB-SEM

## CB-SEM

جهت آزمایش و تایید یا رد کردن نظریه و ارزیابی و مقایسه مدل‌های مختلف

به کل مدل با هم و به صورت یکجا نگاه می‌کند و امکان ارزیابی برازش مدل (Model fit) وجود دارد.

نیاز به حجم نمونه‌های بزرگ است

نرم‌افزارهای Amos و LISREL

## VB-SEM

جهت استفاده در تحقیقات اکتشافی و حداکثر کردن  $R^2$

امکان کار با مدل‌های دارای متغیرهای پنهان ترکیبی

امکان کار با حجم نمونه‌های کوچک وجود دارد

نرم‌افزارهای SmartPLS و WrapPLS



# Confirmatory Factor Analysis

در حقیقت CFA به دنبال این مطلب است که با ایجاد تأییدیه ای از نیکویی برآزش، مدل ارائه شده را مشابه مدل واقعی در جمعیت مورد مطالعه خلق کند. به زبان بسیار ساده تر، CFA می‌کوشد تا مدلی را که بر اساس EFA بدست آوردیم، تأیید کند.

۲- CFA محقق را قادر می‌سازد تا پایایی ابزار را ارزیابی کند. تفاوت عمده ارزیابی ثبات داخلی به کمک CFA نسبت به روش‌های سنتی مثل آلفا کرونباخ در این است که CFA قادر به اندازه‌گیری خطای اندازه‌گیری است.

۳- CFA قادر است تا ساختارهای عاملی را در گروه‌ها مقایسه کند تا مشخص شود که آیا ابزار در آزمودنی‌های مختلف، متفاوت عمل می‌کند؟

# Multitriat-multimethod

Confirmatory factor analysis

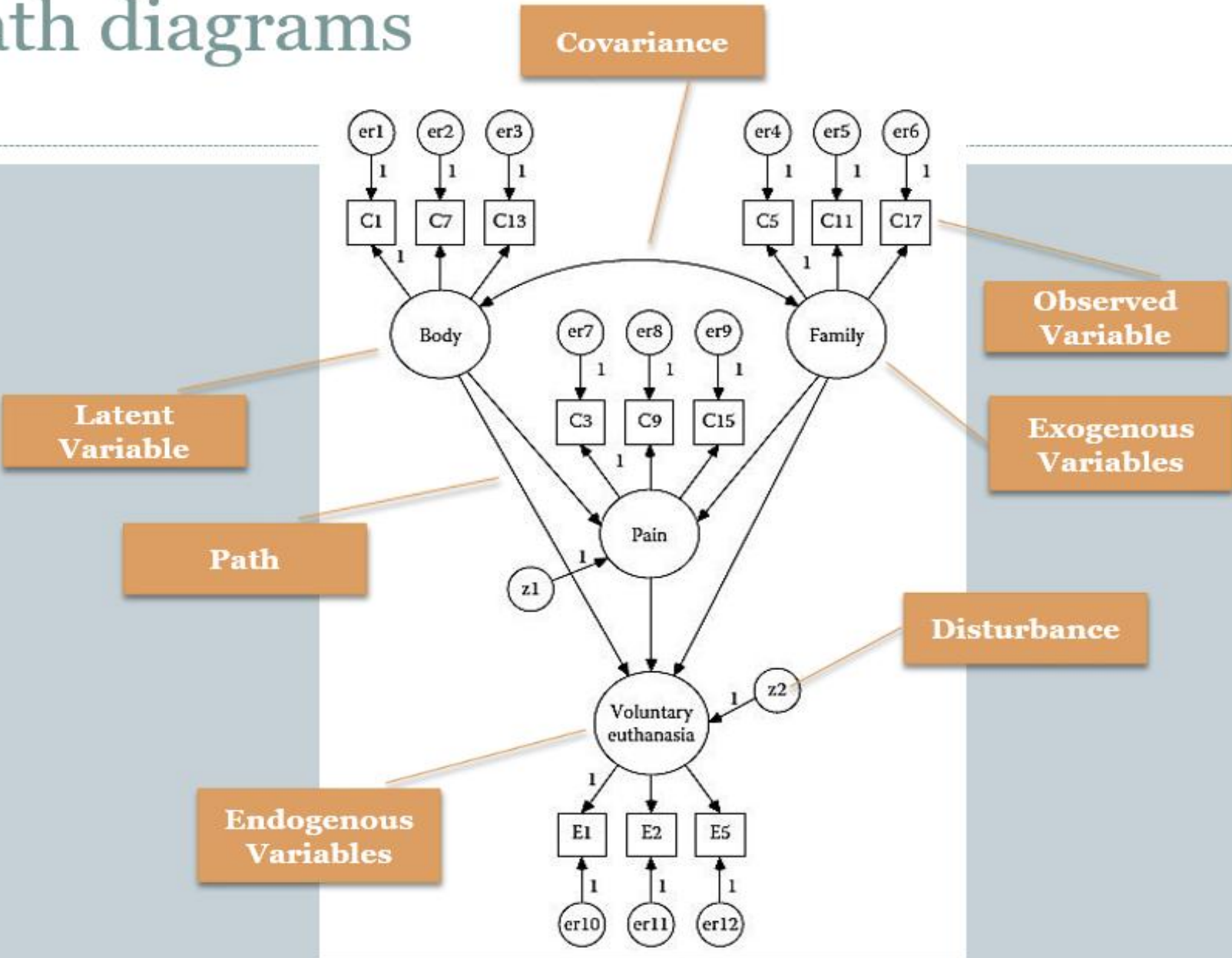
Discriminant validity

Convergent validity

Construct reliability

Heterotrait-monotrait ratio of correlations (**HTMT**)

# Path diagrams



# پذیره های مورد نیاز برای CFA

۱- تخمین برآورد پارامتر با Maximum likelihood

۲- توزیع طبیعی دادهها

۳- دادههای پرت

۴- دادههای فراموش شده

یکی از نکات مهمی که در ارائه‌ی نتایج مطالعات روان‌سنجی به کمک تحلیل عامل تأییدی باید گزارش شود، چگونگی بررسی توزیع طبیعی داده‌ها، داده‌های پرت و در نهایت داده‌های فراموش‌شده است که در عمده موارد گزارش نمی‌شود! به‌طور کلی باید به خاطر داشت که داده‌ها با توزیع طبیعی چندمتغیره یک پیش‌فرض بسیار مهم و حیاتی برای اجرای تحلیل‌های مدل‌سازی معادلات ساختاری است. جهت بررسی توزیع طبیعی چندمتغیره (Multivariate) باید ابتدا توزیع طبیعی تک‌متغیره (univariate) بررسی شود. پیش‌فرض توزیع طبیعی تک‌متغیره داده‌ها بر اساس شاخص چولگی (Skewness)  $\pm 3$  و کشیدگی (Kurtosis)  $\pm 7$  بررسی می‌شود (West et al., 1995). توزیع طبیعی چندمتغیره داده‌ها به کمک ضریب Mardia بررسی می‌شود. ضریب Mardia بالاتر از ۸ بیانگر نقض کشیدگی است، اگرچه بنتلر (۲۰۰۵) بیان می‌کند مقادیر بالاتر از ۵ را باید به‌عنوان داده‌هایی با توزیع غیرطبیعی در نظر گرفت. یکی از راه‌هایی که جهت تحلیل مدل‌سازی معادلات ساختاری داده‌هایی با توزیع غیرطبیعی چندمتغیره پیشنهاد می‌شود، استفاده از Asymptomatic distribution free است. اخیراً دانشمندان آماری نشان داده‌اند که حداقل اندازه‌ی نمونه‌ها برای استفاده از این روش باید بزرگ‌تر از ۱۰ برابر تعداد پارامترهای برآوردشده باشد، در غیر این صورت نتایج این روش معتبر نیست. داده‌های پرت چندمتغیره به کمک آزمون d-squared Mahalanobis مورد بررسی قرار می‌گیرد. Mahalanobis d-squared با  $p < 0.001$  به‌عنوان وجود داده‌های پرت در نظر گرفته می‌شود.



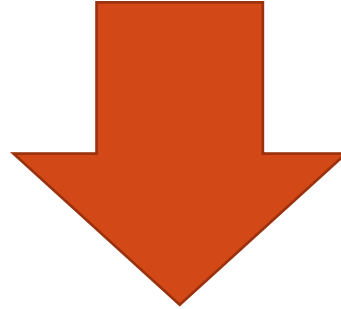
# مراحل انجام تحلیل عامل تاییدی

- **تعیین کردن مدل:** کاملاً نظری است و با دیاگرام مسیر نشان داده می شود
- **شناسایی:** یک موضوع نظری نیست بلکه یک موضوع آماری محسوب می شود. یعنی مقادیر شناخته شده واریانس و کواریانس گویه های ابزار که از CFA بدست می آید. محقق سعی می کند تا حداقل سه متغیر مشاهده ای به ازای هر متغیر پنهان داشته باشد.
- **برآورد کردن پارامترهای مدل:** توسط کامپیوتر انجام می شود.
- **ارزشیابی برازش مدل با داده ها:** به کمک شاخص های استاندارد برازش مدل ارزیابی می شود.
- **اصلاح یا تعیین مجدد مدل جهت بهبود در برازش:** اگر شاخص ها نشان دهند که مدل از برازش مناسبی برخوردار نیست می توان مدل را اصلاح کرد

$$X + Y = 10$$

# ارزیابی برازش مدل توسط تحلیل عامل تاییدی؟

Hypothesized Model.....Observation model



Goodness-of-fit chi-square

*P* value?

H0

H1

نقطه ضعف...

# Test of Fit



- How good was it for you?
  - Signs of the coefficients
  - Statistically significant coefficients
  - $\chi^2$  GoF test to look at the model as a whole
  - So many because none is ideal
- ✦ Three Groups:
  - **Absolute Fit**
    - How well the hypothesized model fits observed data
    - $\chi^2$ , GFI, RMSR, RMSEA, etc.
  - **Incremental Fit**
    - The relative position of the model between worst fit to perfect fit
    - TLI, NFI, RFI, IFI, CFI, etc.
  - **Parsimonious Fit**
    - Compare models with differing number of parameters
    - AIC, AGFI, PNFI, PCFI, etc.

# Recommended Criteria for Fit Indices

Fit Indices	Authors	Recommended Value
$\chi^2$	Meyers et al, 2005	P-value > .05
CMIN/DF	Marsh & Hocevar, 1985; Bentler, 1990 ; Hair et al., 2009	< 5.0 < 5.0 Reported if n > 200 < 3.0 Good; <5.0 Sometimes Permissible
GFI	Chau, 1997; Segars & Grover, 1993	> .90 > .90
CFI	Bentler, 1990 Hatcher, 1994	> .90 > .90
RMSEA	Byrne, 2001 Hu & Bentler, 1999 Meyers et al, 2005	< .08 < .05 < .08 : good fit; .08 to .1 : moderate fit; > .1 : poor fit.
SRMR	Hair et al., 2009	<0.09
NFI	Bentler & Bonett, 1980	> .90
CMIN	Tabachnik & Fidell, 1996	Reported if n between 100 and 200
RFI	Meyers et al, 2005	> .90
IFI	Meyers et al, 2005	> .90
RMSR	Meyers et al, 2005	< .05
PNFI	Meyers et al, 2005	> .5
PCFI	Meyers et al, 2005	> .5
PCLOSE	Hair et al, 2009	> 0.05
AGFI	Hair et al., 2009	> 0.80

## Model Fit Summary

### CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	83	505.997	267	.000	1.895
Saturated model	350	.000	0		
Independence model	50	4561.474	300	.000	15.205

### Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.889	.875	.944	.937	.944
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

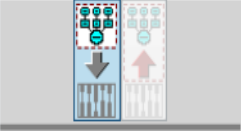
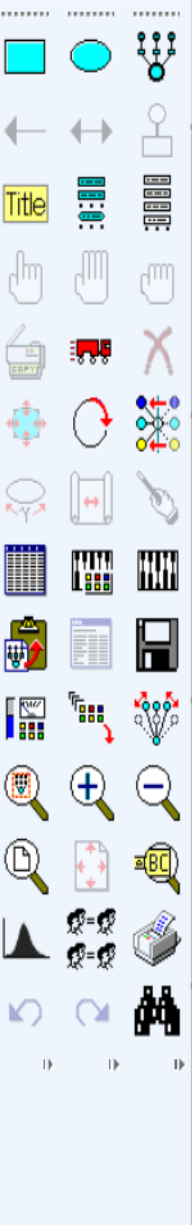
### Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	.890	.791	.840
Saturated model	.000	.000	.000
Independence model	1.000	.000	.000

### NCP

Model	NCP	LO 90	HI 90
Default model	238.997	179.374	306.431
Saturated model	.000	.000	.000
Independence model	4261.474	4046.526	4483.699

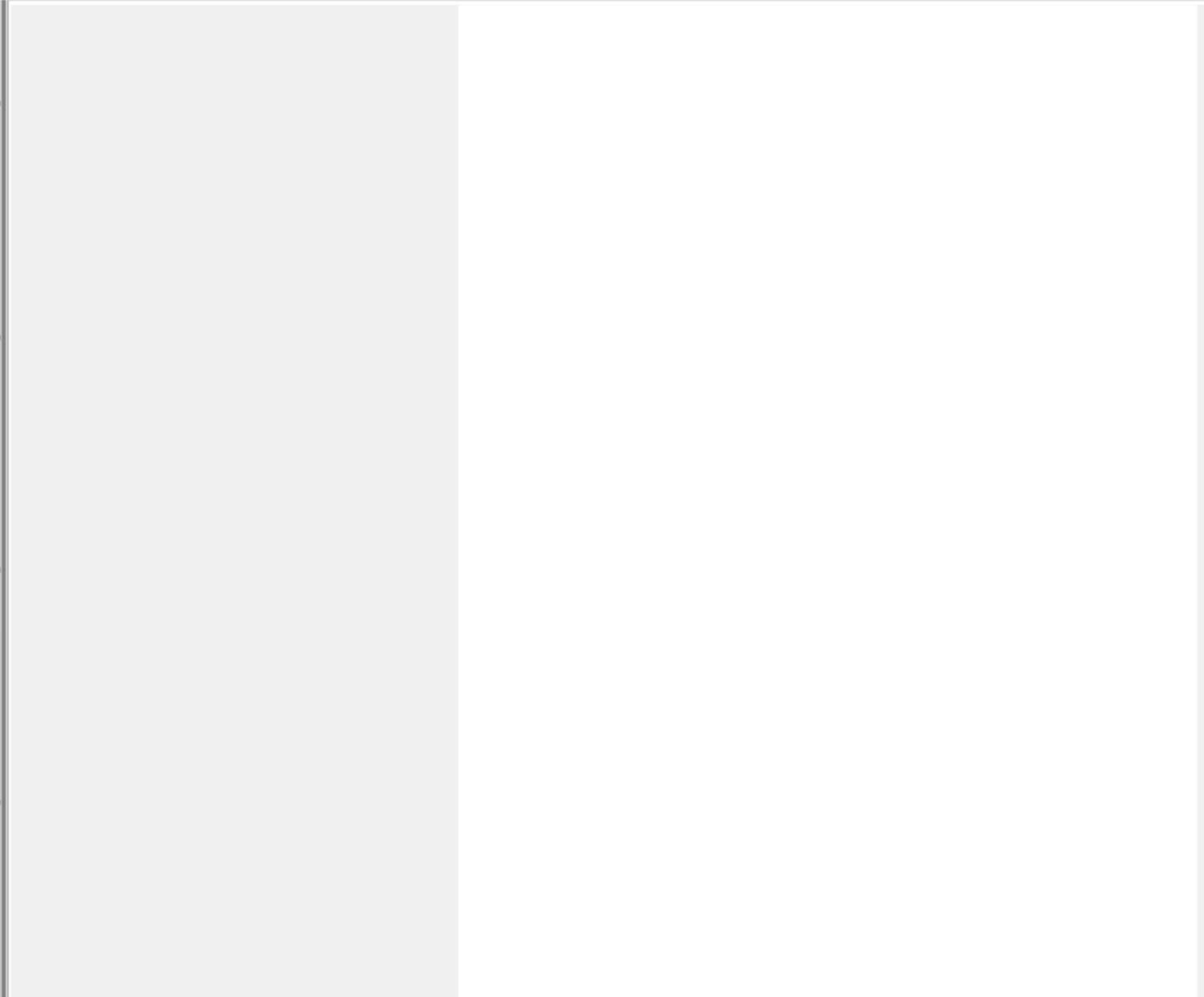
### RMSEA



Group number 1

XX: Default model

Unstandardized estimates  
Standardized estimates



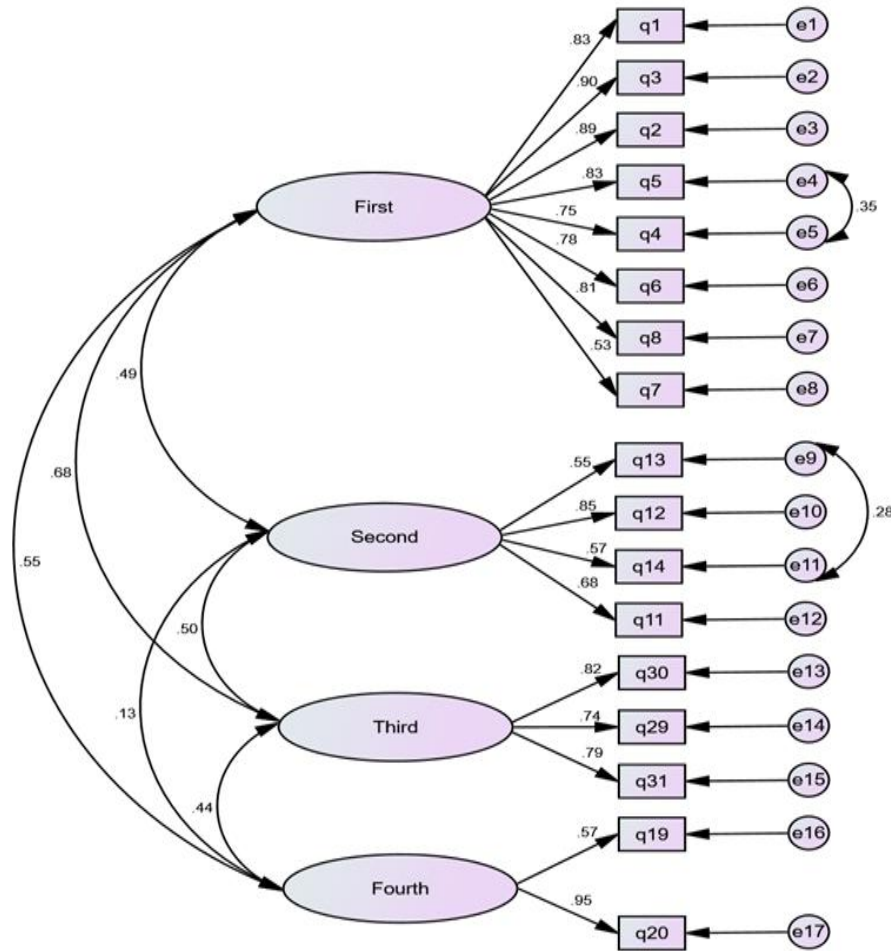


Figure 2. Final Model of the First Order CFA of the End-of-life Caregiving Experience Appraisal Scale

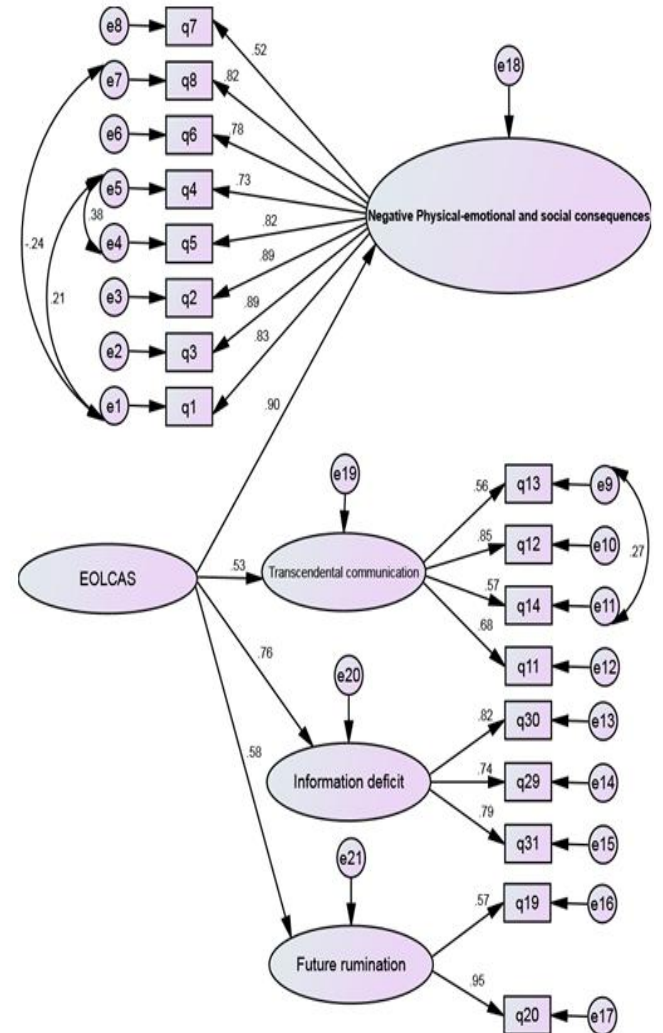


Figure 1. Final Model of the Second Order CFA of the End-of-life Caregiving Experience Appraisal Scale



- Unweighted least squares
  - No assumptions about the variables distributions
  - Results are dependent on the scale of measurement
- Weighted least squares
  - Large sample sizes
- Maximum Likelihood
  - Requires multivariate normality
  - No ordinal scales or highly skewed measures

Analysis Properties

Title | Estimation | Numerical | Bias | Output | Bootstrap | Permutations | Random #

Discrepancy

- Maximum likelihood
- Generalized least squares
- Unweighted least squares
- Scale-free least squares
- Asymptotically distribution-free

**Normality Assumption**

- Estimate means and intercepts
- Emulisrel6
- Chicorrect

For the purpose of computing fit measures with incomplete data:

- Fit the saturated and independence models
- Fit the saturated model only
- Fit neither model

# در گام اول چه کنیم؟

## Necessary Data Screening



- Handle Missing Data
- Address outliers and influential cases.
- Meet multivariate statistical assumptions for alternative tests

Your data should be  
“clean”

# Necessary Data Screening



- Handle Missing Data
- Address outliers and influential cases.
- Meet multivariate statistical assumptions for alternative tests



- Reports
- Descriptive Statistics**
- Tables
- Compare Means
- General Linear Model
- Generalized Linear Models
- Mixed Models
- Correlate
- Regression
- Loglinear
- Neural Networks
- Classify
- Dimension Reduction
- Scale
- Nonparametric Tests
- Forecasting
- Survival
- Multiple Response
- Missing Value Analysis...
- Multiple Imputation
- Complex Samples
- Quality Control
- ROC Curve...
- IBM SPSS Amos...



- 1** Frequencies...
- Descriptives...
- Explore...
- Crosstabs...
- Ratio...
- P-P Plots...
- Q-Q Plots...

**Frequencies**

Variable(s):

- Image1
- Image2
- Image3
- Image4
- Image5
- Image6
- Job.Relev1
- Job.Relev2
- Job.Relev3

**2**

➔

Variable(s):

- Subj.Norms1
- Subj.Norms2
- Subj.Norms3
- Subj.Norms4
- Subj.Norms5

Display frequency tables

Statistics...

Charts...

Format...

Bootstrap...

OK Paste Reset Cancel Help

**3**

		Subj.Norms1	Subj.Norms2	Subj.Norms3	Subj.Norms4	Subj.Norms5
N	Valid	300	301	299	301	299
	Missing	1	0	2	0	2

# Handling Missing Data



## *Hair et al.'s (2009) Rules of Thumb:*

- Missing data under 10% for an individual case or observation can generally be ignored, except when the missing data occurs in a specific nonrandom fashion.
- The number of cases with no missing data must be sufficient for the selected analysis technique if replacement values will not be substituted (imputed) for the missing data.

# Imputation Methods

(Hair et al. (2009), table 2-2)



- Use only valid data
  - No imputation, just use valid cases or variables
  - In SPSS: Exclude Pairwise (variable), Listwise (case)
- Use known replacement values
  - Match missing value with similar case's value
- Use calculated replacement values
  - Use variable mean, median, or mode
  - Regression based on known relationships
- Model based methods
  - Iterative two step estimation of value and descriptives to find most appropriate replacement value



# Imputation in SPSS

1. **Replace Missing Values...**

2. Include each variable that has values that need imputing

3. For each variable you can choose the new name (for the imputed column) and the type of imputation

4. OK

Subj.Norms1  
Subj.Norms2  
Subj.Norms3  
Subj.Norms4  
Subj.Norms5  
Image1  
Image2  
Image3  
Image4  
Image5  
Image6  
Job.Relev2

New Variable(s):  
Subj.Norms1\_1 = SMEAN(Subj.Norms1)  
Subj.Norms2\_1 = SMEAN(Subj.Norms2)  
Subj.Norms3\_1 = SMEAN(Subj.Norms3)  
Subj.Norms4\_1 = SMEAN(Subj.Norms4)  
Subj.Norms5\_1 = SMEAN(Subj.Norms5)

Name and Method  
Name: Subj.Norms1\_1  
Method: Series mean  
Snap of  
Number: 2 All

Imputation Method	Advantages	Disadvantages	Best Used When:
Mean Substitution	<ul style="list-style-type: none"> <li>• Easily implemented</li> <li>• Provides all cases with complete information</li> </ul>	<ul style="list-style-type: none"> <li>• Reduces variance of the distribution</li> <li>• Distorts distribution of the data</li> <li>• Depresses observed correlations</li> </ul>	<ul style="list-style-type: none"> <li>• Relative low levels of missing data</li> <li>• Relatively strong relationships among variables</li> </ul>
Regression Imputation	<ul style="list-style-type: none"> <li>• Employs actual relationships among the variables</li> <li>• Replacement values calculated based on an observation's own values on other variables.</li> <li>• Unique set of predictors can be used for each variable with missing data.</li> </ul>	<ul style="list-style-type: none"> <li>• Reinforces existing relationships and reduces generalizability</li> <li>• Must have sufficient relationships among variables to generate valid predicted values.</li> <li>• Understates variance unless error term added to replacement value.</li> <li>• Replacement values may be "out of range"</li> </ul>	<ul style="list-style-type: none"> <li>• Moderate to high levels of missing data</li> <li>• Relationships sufficiently established so as to not impact generalizability</li> </ul>
Model-Based Methods	<ul style="list-style-type: none"> <li>• Accommodates both nonrandom and random missing data processes</li> <li>• Best representation of original distribution of values with least bias.</li> </ul>	<ul style="list-style-type: none"> <li>• Complex model specification by researcher</li> <li>• Requires specialized software</li> <li>• Typically not available in software programs (except EM method in SPSS)</li> </ul>	<ul style="list-style-type: none"> <li>• Only method that can accommodate nonrandom missing data process</li> <li>• High levels of missing data require least biased method to ensure generalizability</li> </ul>



# Best Method – Prevention!



- Short surveys
- Easy to understand and answer survey items
- Force completion (incentives, technology)
- Digital surveys (rather than paper)
- Bribe/motivate (iPad drawing)
- **Put dependent variables at the beginning** of the survey!

# Outliers and Influentials



- Outliers can influence your results, *pulling the mean away from the median.*
- Outliers also affect distributional assumptions and often reflect false or mistaken responses
- Two type of outliers:
  - *outliers for individual variables (univariate)*
    - ✦ Extreme values for a single variable
  - *outliers for the model (multivariate)*
    - ✦ Extreme (uncommon) values for a correlation

# Detecting Univariate Outliers



**1** Legacy Dialogs > **Boxplot...**

**2** Simple  
Data in Chart Are:  
 Summaries for groups of cases  
 Summaries of separate variables

**3** Define Simple Boxplot: Summaries of Separate Variables  
Boxes Represent:  
Subj.Norms1  
Subj.Norms2  
Subj.Norms3  
Subj.Norms4  
Subj.Norms5

SEM and AMOS - PowerPoint (Product Activation Failed)

File Home Insert Design Transitions Animations Slide Show Review View Tell me what you want to do... Sign in Share

Clipboard Font Paragraph Drawing Editing

39 Best Method - Prevention!

- Short surveys
- Easy to understand and answer survey items
- Prior completion (incentives, technology)
- Digital surveys (rather than paper)
- Incentivize (P&M drawing)
- Put dependent variables at the beginning of the survey!

40 Outliers and Influentials

- Outliers can influence your results pulling the mean away from the median
- Outliers also affect distributional assumptions and often reflect false or mistaken responses

41 Detecting Univariate Outliers

42 Detecting Univariate Outliers

43 The Outlier Labeling Rule

### Detecting Univariate Outliers

Mean

50% should fall within the box

99% should fall within this range

Outliers!

107

107

60

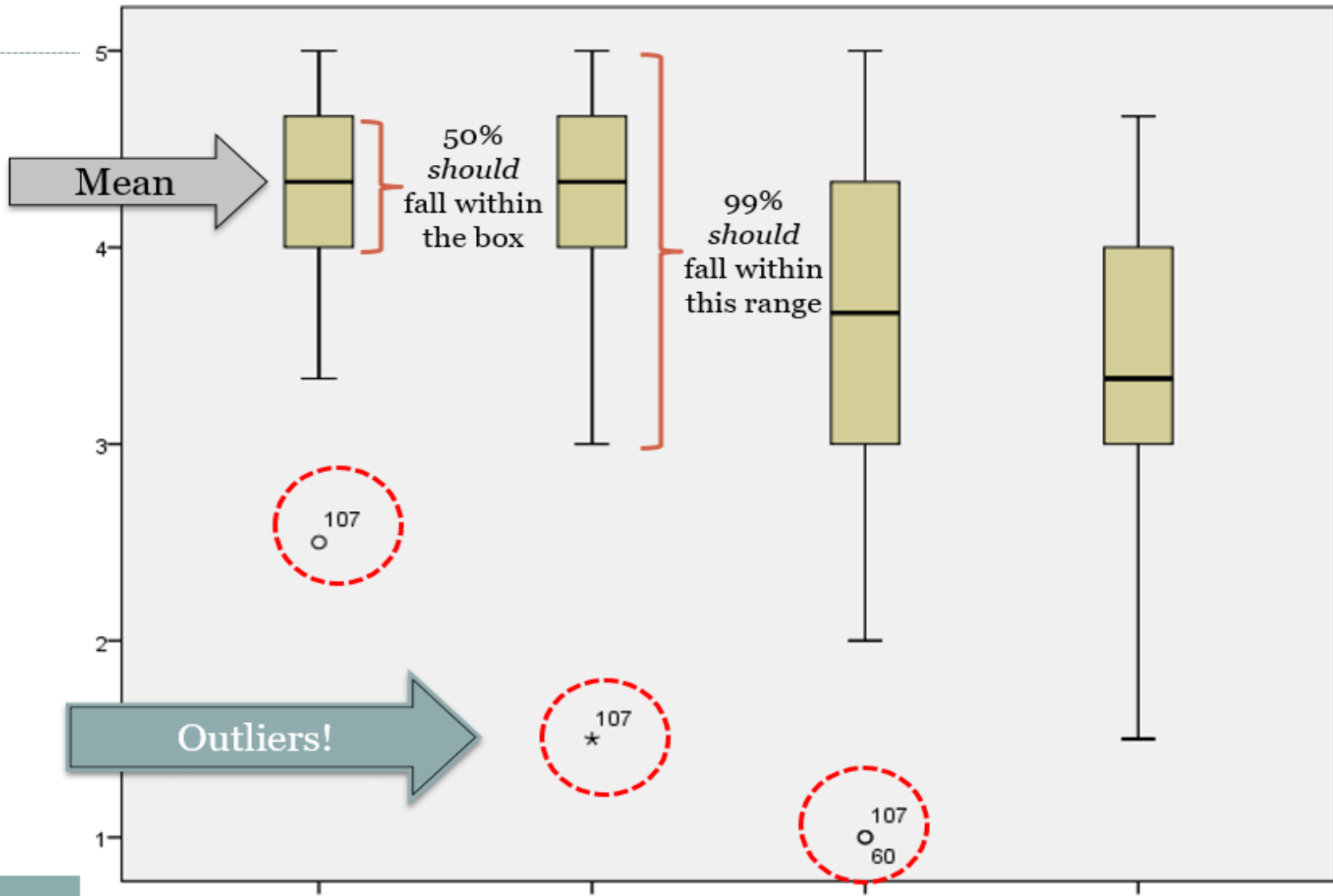
Click to add notes

Activate Windows  
Go to Settings to activate Windows.

Slide 42 of 91 English (United States) Notes Comments 83%

12:59 AM 5/21/2021

# Detecting Univariate Outliers



# Handling Univariate Outliers



- Should be examined on a case by case basis.
- If the outlier is truly abnormal, and not representative of your population, then it is okay to remove. However, this requires careful examination of the data points
  - e.g., you are studying dogs, but somehow a cat got ahold of your survey
  - e.g., someone answered “1” for all 75 questions on the survey

# Detecting Multivariate Outliers



- Multivariate outliers refer to sets of data points that do not fit the standard sets of correlations exhibited by the other data points in the dataset with regards to your causal model.
  - Exercise and Weight loss
- **Mahalanobis** d-squared.



Analysis Properties

Estimation | Numerical | Bias | **Output** | Bootstrap | Permutations | Random # | Title

Minimization history       Indirect, direct & total effects

Standardized estimates       Factor score weights

Squared multiple correlations       Covariances of estimates

Sample moments       Correlations of estimates

Implied moments       Critical ratios for differences

All implied moments       Tests for normality and outliers

Residual moments       Observed information matrix

Modification indices       Threshold for modification indices

Outliers.amw

- [-] Analysis Summary
  - Notes for Group
- [-] Variable Summary
  - Parameter summary
  - Assessment of normality
  - Observations farthest from the centroid (Mahalanobis distance)**
- [-] Notes for Model
- [-] Estimates
  - Minimization History
- [-] Model Fit
  - Execution Time

These are **row** numbers from SPSS

Anything less than .05 in the p1 column is abnormal, and is candidate for inspection

Observation number	Mahalanobis d-squared	p1	p2
7	39.225	.000	.000
58	23.518	.000	.000
92	19.238	.001	.000
134	17.758	.001	.000
131	16.426	.002	.000
74	16.292	.003	.000
126	15.419	.004	.000
107	14.891	.005	.000
102	13.809	.008	.000
150	13.786	.008	.000
55	13.701	.008	.000
66	13.510	.009	.000
33	13.351	.010	.000
72	13.265	.010	.000
53	12.764	.012	.000
112	11.571	.021	.000
90	11.402	.022	.000
20	8.634	.071	.027
52	8.564	.073	.018
108	8.495	.075	.012
76	8.393	.078	.010



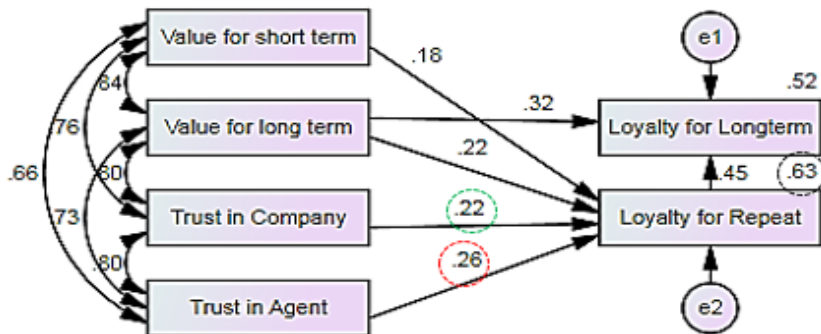
# Handling Multivariate Outliers



- Create a new variable in SPSS called “Outlier”
  - Code 0 for Mahalanobis > .05
  - Code 1 for Mahalanobis < .05
- AMOS: “Outlier” as a grouping variable
  - This then runs your model with only non-outliers

# Before and after removing outliers

N=340



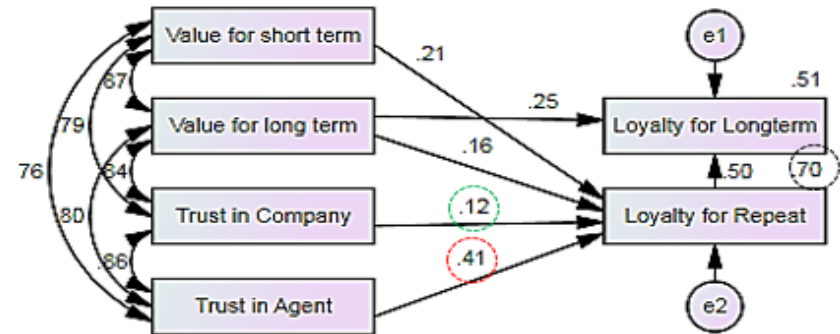
Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	18	4.238	3	.237	1.413

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.035	.000	.104	.546

Default model  
Standardized RMR = .0122

**BEFORE**

N=295



Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	18	2.949	3	.399	.983

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.000	.000	.098	.669

Default model  
Standardized RMR = .0091

**AFTER**

Even after you remove outliers, the Mahalanobis will come up with a whole new set of outliers, so these should be checked on a case by case basis, using the Mahalanobis as a guide for inspection.

# “Best Practice” for outliers



- It is a bad idea to remove outliers, unless they are truly “abnormal” and do not represent accurate observations from the population.
- **Removing outliers is risky**
  - Generalizability

# Normality in AMOS

- Refer to the “Assessment of normality” in the Text View output
- Data is considered to be normal if:

:: Skewness is between -3 to +3

:: Kurtosis is between -7 to +7

Barbara G. Tabachnick

Amos Output

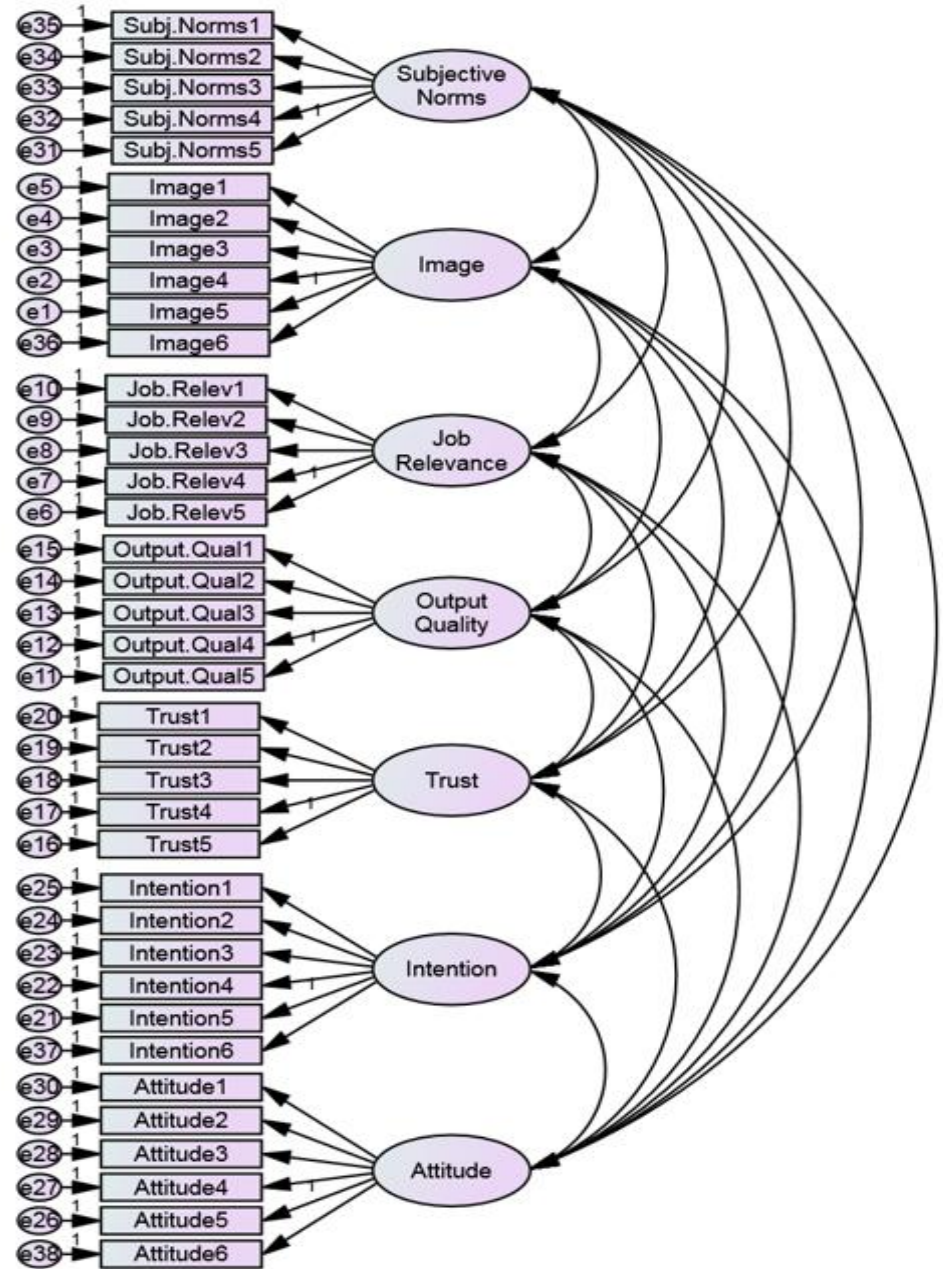
Ruziah Model.amw

- Analysis Summary
- Notes for Group
- Variable Summary
- Parameter summary
- Assessment of normality
- Observations farthest from th
- Sample Moments
- Notes for Model
- Estimates
- Notes for Group/Model
- Modification Indices
- Minimization History
- Model Fit
- Execution Time

Assessment of normality (Group number 1)

Variable	min	max	skew	c.r.	kurtosis	c.r.
tingt	14.000	120.000	-.870	-6.091	.630	2.204
comchang	37.000	109.000	-.666	-4.663	.473	1.654
OPB	7.000	67.000	-.370	-2.591	-.052	-.183
commuctn	14.000	100.000	-.837	-5.857	.827	2.894
procfair	8.000	80.000	-.919	-6.433	.927	3.246
pos	16.000	96.000	-.882	-6.173	.892	3.121
trleader	15.000	146.000	-1.030	-7.208	1.135	3.972
svalues	39.000	90.000	-.823	-5.759	.612	2.143
feedback	17.000	90.000	-.948	-6.637	.774	2.710
jobaut	15.000	65.000	-.978	-6.846	.942	3.296
emppart	19.000	90.000	-.682	-4.776	.436	1.525
gc	8.000	80.000	-1.214	-8.501	1.935	6.773
Multivariate					100.266	46.895

# CFA, Measurement Model



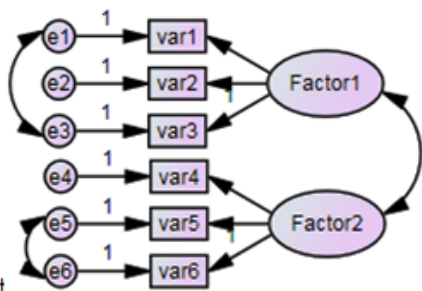




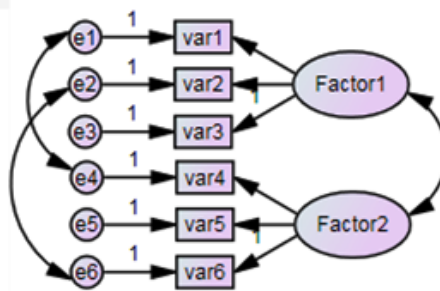
DAS.amw

- Analysis Summary
- Notes for Group
- Variable Summary
- Parameter Summary
- Assessment of normality
- Observations farthest from the cent
- Notes for Model
- Estimates
- Modification Indices**
  - Covariances:
  - Variances:
  - Regression Weights:
- Minimization History
- Model Fit
  - CMIN
  - RMR, GFI
  - Baseline Comparisons
  - Parsimony-Adjusted Measures
  - NCP
  - FMIN
  - RMSEA
  - AIC
  - ECVI
  - HOELTER

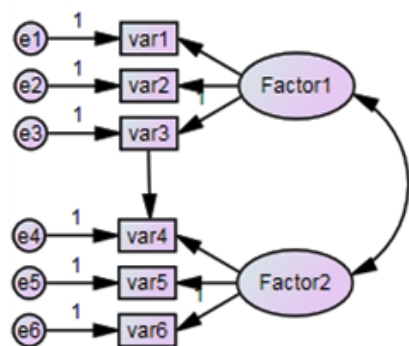
Acceptable



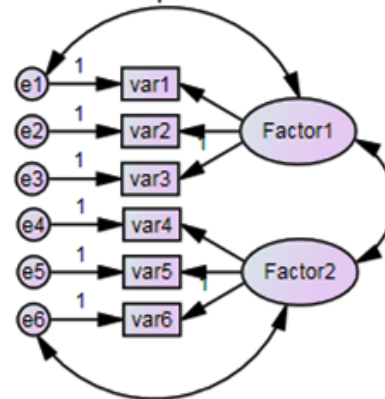
Inacceptable



Inacceptable



Inacceptable



Modification Indices (Group number 1 - Default model)

Covariances: (Group number 1 - Default model)

	M.I.	Par Change
e9 <--> e13	5.460	-.136
e7 <--> Schond	6.219	-.098
e7 <--> First	8.936	.159
e7 <--> e15	13.346	.233
e7 <--> e12	5.986	.178
e7 <--> e9	8.927	-.183
e6 <--> e9	9.768	.169
e4 <--> Fourth	13.817	.149
e4 <--> e14	5.722	.163
e4 <--> e5	5.039	-.175
e3 <--> e13	6.077	-.149
e3 <--> e6	5.520	-.131
e2 <--> e11	5.380	-.116
e2 <--> e9	9.418	.165
e2 <--> e8	12.347	-.210

# Reliability and Validity

## Construct Validity



- If you have **convergent validity** issues, then your variables do not correlate well with each other within their parent factor; i.e. the latent factor is not well explained by its observed variables.
- If you have **discriminant validity** issues, then your variables correlate more highly with variables outside their parent factor than with the variables within their parent factor; i.e., the latent factor is better explained by some other variables (from a different factor), than by its own observed variables.



# Validity and Reliability



- It is absolutely necessary to establish convergent and discriminant validity, as well as reliability, when doing a CFA. If your factors do not demonstrate adequate validity and reliability, moving on to test a causal model will be useless - garbage in, garbage out!

## Reliability

- $CR > 0.7$

## Convergent Validity

- **Factor Loading:**

- All factor loadings should be statistically significant .
- Standardized loadings estimates should be 0.5 or higher, and ideally 0.7 or higher
- $CR > AVE$
- $AVE > 0.5$

## Discriminant Validity

- $MSV < AVE$
- $ASV < AVE$

<b>CR</b>	: Construct Reliability
<b>AVE</b>	: Average Variance Extracted
<b>MSV</b>	: Maximum Shared Squared Variance
<b>ASV</b>	: Average Shared Squared Variance

File Home Insert Page Layout Formulas Data Review View Tell me what you want to do...

Clipboard Font Alignment Number Styles Cells

Calibri 11 A A Wrap Text General Input Linked Cell

Conditional Formatting Table AutoSum Fill Clear

A2

A B C D E F G H I J K L M N O P Q R S T U V

1 Paste Correlations table in A2 Paste Standardized Regression Weights table in F2

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

Read Caveats and Assumptions below, and then click me.

#### Caveats and Assumptions:

1. Your **latent** variable names **do not** end in numbers (bad: F1, Factor12). It is okay to have **observed** variables named whatever you want.
2. Your error/residual names **do** end with numbers (good: e1, res12)
3. Your variable names are not any of the following: AVE, Max, CR, MSV
4. You have more than 2 latent variables.
5. You have no keywood cases

$$CR = \frac{\left( \sum_{j=1}^i \lambda_j \right)^2}{\left( \sum_{j=1}^i \lambda_j \right)^2 + \left( \sum_{i=1}^i 1 - \lambda_j^2 \right)}$$

$$AVE = \frac{\sum_{i=1}^n L_i^2}{n}$$

$$H = \frac{1}{1 + \frac{1}{\sum_{i=1}^n \lambda_i^2}}$$



# اشتباه های رایجی در گزارش یافته های روایی سازه

در مقدمه، روش کار(عدم گزارش پیش فرضهای ضروری)، نتایج و بحث (عدم گزارش e)

گزارش مقدار ویژه و واریانس استخراج شده

کاربرد نادرست مولفه / عامل (PCA/EFA)

گزارش اشتباه سازه تایید شده در گروه ها

به کارگیری نمره حاصل از پرسشنامه و انجام آزمون فرضیه

اصلاح نامناسب مدل و یا اصلاح بیش از اندازه مدل

نتیجه گیری

# کار عملی