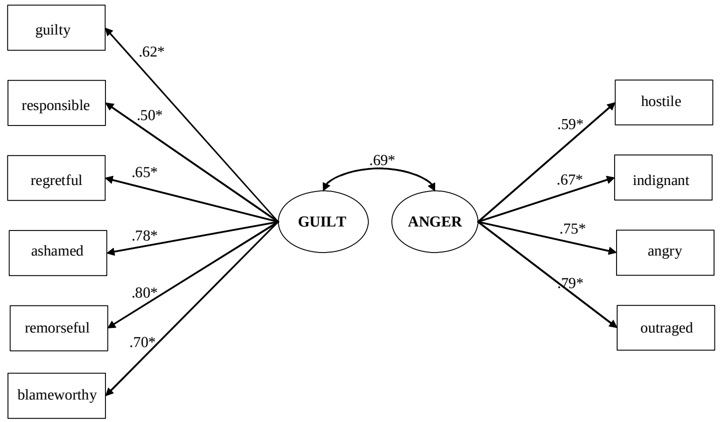
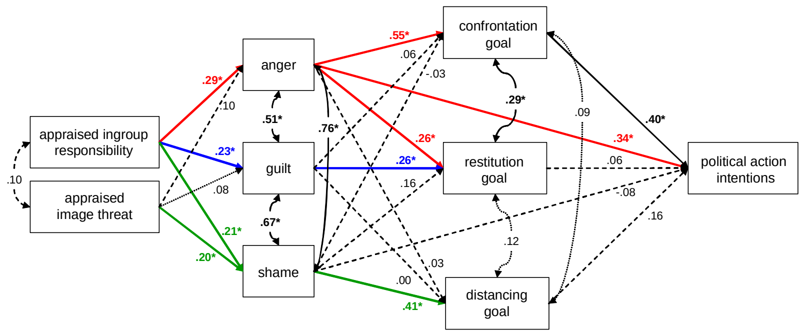
STATISTICS WORKSHOP



Confirmatory Factor Analysis (CFA)

&

Structural Equation Modelling (SEM)



overview

PART 1: INTRODUCTION TO CONCEPTS

1. CFA: purpose, key concepts, when to use it
2. SEM: purpose, key concepts, when to use it
3. how CFA and SEM work → *5 minute break*
4. steps in CFA & SEM
   (1) specifying a model
   (2) evaluating model fit
   (3) examining modification indices
   (4) testing alternative models → *10 minute break*

PART 2: TESTING MODELS IN AMOS

5. how to specify a model
6. how to test a CFA model
7. how to interpret output

purpose of CFA

- applied to single set of variables to test hypotheses about the relative independence of subsets of variables

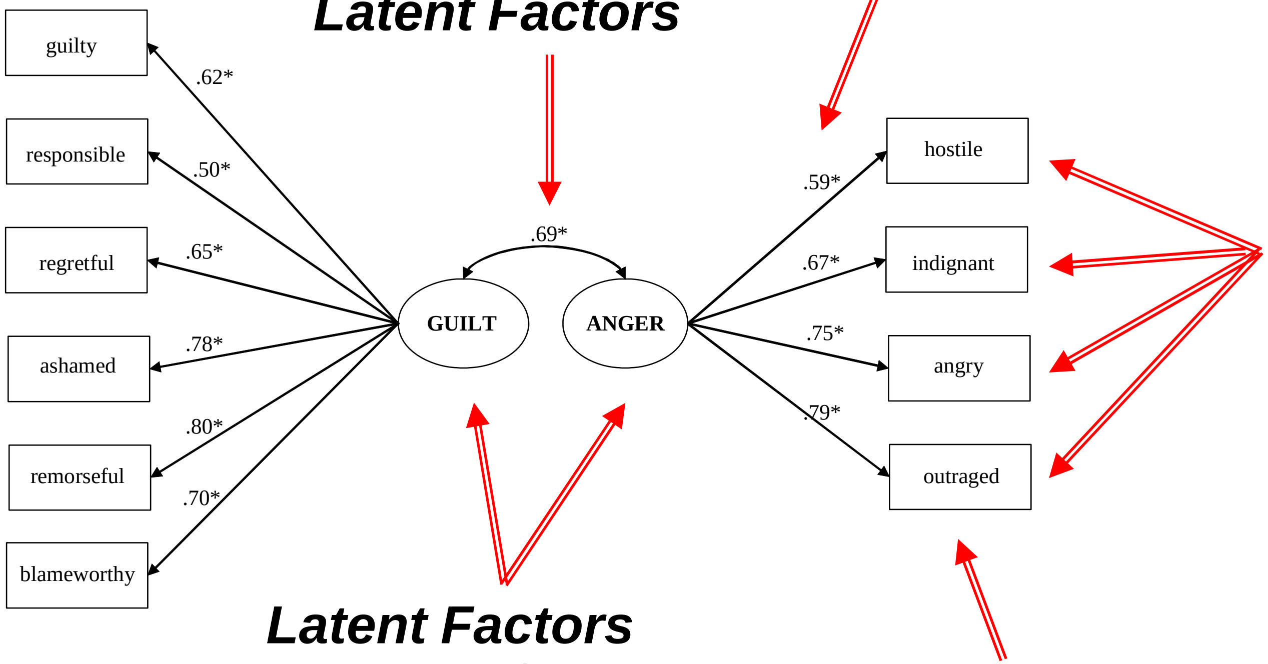
- similar aims to ***exploratory factor analysis (EFA)***:
  1. identify underlying constructs or factors that account for associations between subsets of variables
  2. identify how strongly each item is associated with one or more factors (factor loadings)

- key difference between EFA and CFA:
  **** EFA is data-driven***
  → *SPSS calculates all possible factor loadings – factor structure is interpreted post-hoc based on the results*
  **** CFA is theory-driven***
  → *program calculates only those factor loadings that we have hypothesized (all others are constrained to be "0")*
  → *factor structure is specified a priori*

key terms in CFA

Correlation between Latent Factors

Factor Loadings



Error Terms

- uniqueness variance of items that is not captured by the latent factor

→ *rarely shown in figures*

Latent Factors

- unmeasured constructs
- made up of common variance among items
- free of measurement error

→ *represented by circles or ellipses*

Manifest Variables* or *Items

- directly measured

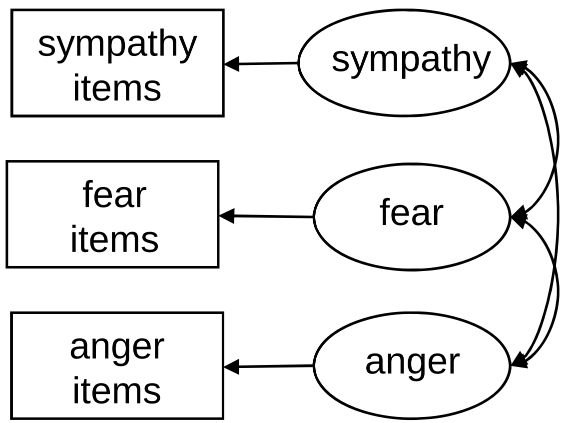
→ *represented by squares or rectangles*

when to use CFA

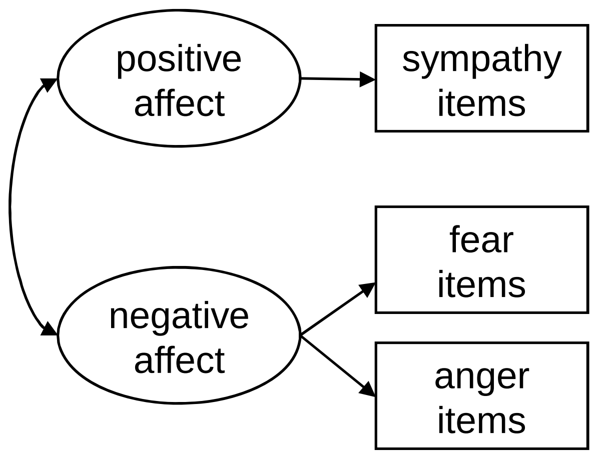
- CFA involves *a priori* hypotheses, and so is becoming preferred over EFA (at least in social psychology)

- you might use CFA instead of EFA if you have clear hypotheses (based on previous theory and/or research) about the factors underlying a set of items

- CFA is also useful if you want to compare different factor structures that are theoretically viable



vs.



purpose of SEM

- **used to test hypothesized relationships between variables**
  - * assumes linear relationships
  - * assumes multivariate normality
  - * variables can be continuous or categorical (not in AMOS though)
  - * can be used with correlational or experimental data

- **theory-driven approach**
  - * researcher specifies order of association between variables
  - * researcher specifies which relationships should be tested (all other relationships constrained to be "0")

- **comprehensive approach**
  - ***** can subsume standard techniques of regression and ANOVA

advantages of SEM

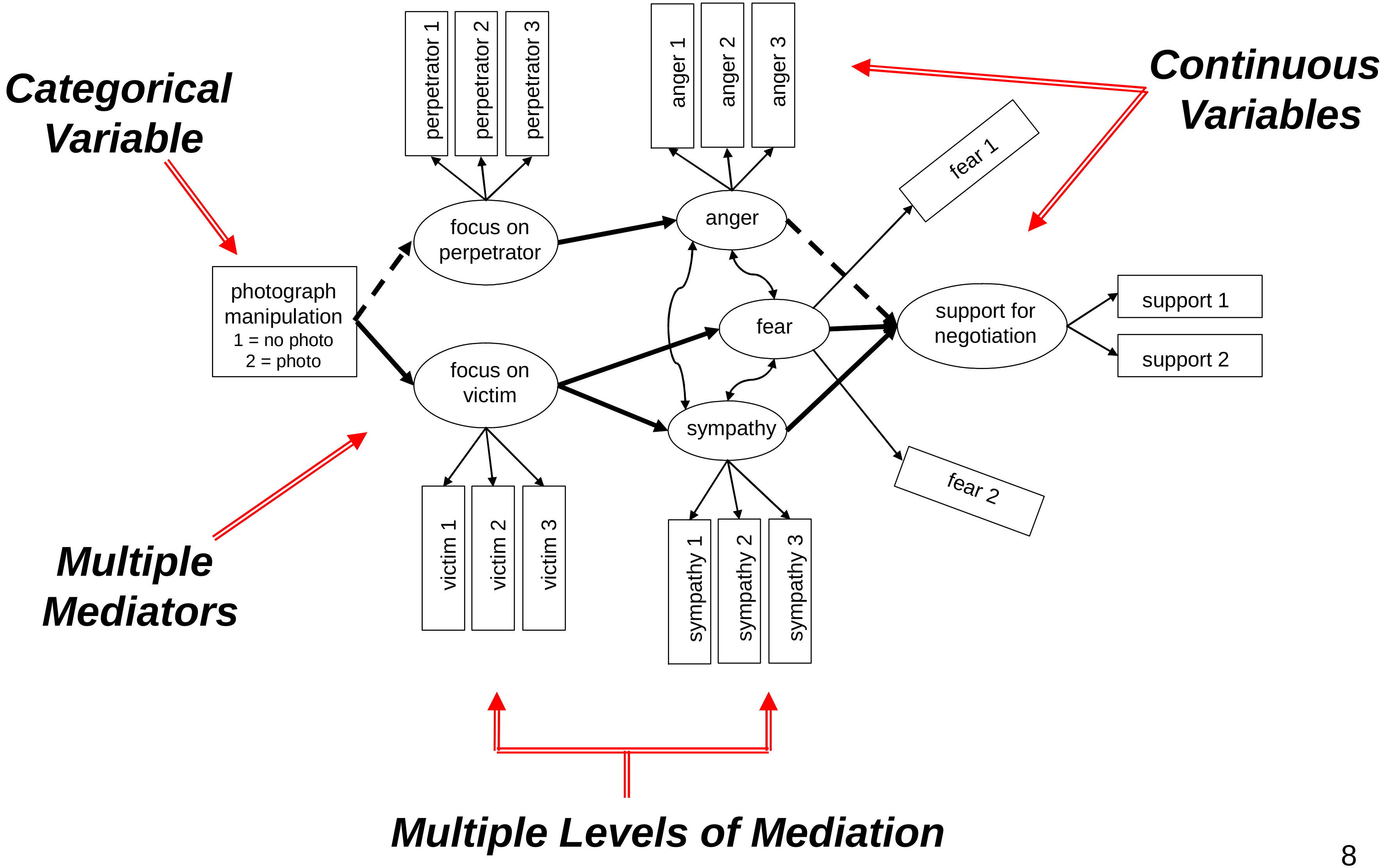
1. can provide more accurate estimate of relationships

* usually has two components
  (1) measurement = items loading on to latent factors
  (2) structural = relationships between factors
* use of latent factors partials out items' error variance (uniqueness)

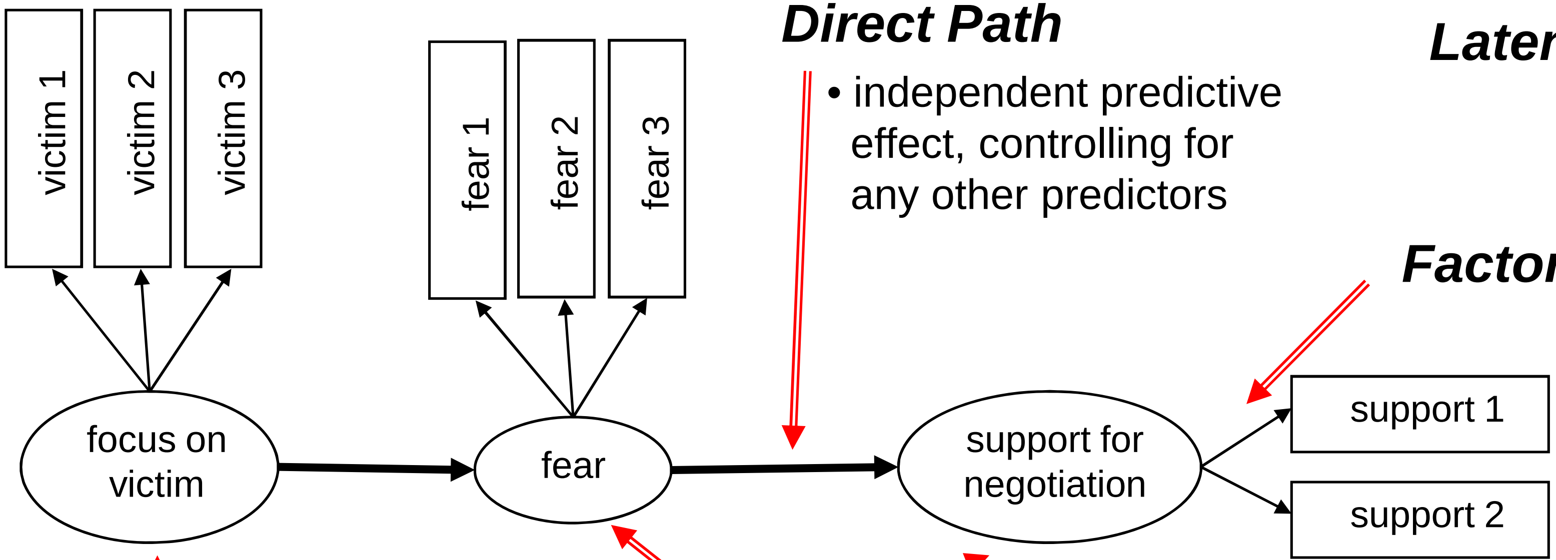
2. can test complex sets of relationships between variables

* ***regression*** can test:
  (a) multiple predictors of 1 outcome OR
  (b) 1 mediator between 1 predictor and 1 outcome
* ***SEM*** can test:
  (a) multiple predictors of multiple outcomes AND
  (b) multiple mediators between multiple predictors and multiple outcomes

what SEM can do



key terms in SEM



Direct Path

- independent predictive effect, controlling for any other predictors

Correlation between Latent Factors

Factor Loading

Exogenous Factor

- not predicted by anything else in the model
  → *i.e., an IV or predictor*

Endogenous Factor

- predicted by something else in model
- can predict other variables
- has an associated ***Disturbance*** (all variance that predictors did not explain)

Manifest Variables

- directly measured

Error Term

- uniqueness variance of items that is not captured by the latent factor
  → *rarely shown in figures*

when to use SEM

- ***SEM can be used to…***

  * control for measurement error

  * test a complex set of relationships (i.e., multiple mediators and/or moderators) between a large number of variables (i.e., > 4 or 5)

  → especially useful in longitudinal research

- ***SEM should not be used to…***

  * make causal claims from correlational data

  * test very simple models that regression or ANOVA can do:

1. manipulation → outcome(s)
2. 1 predictor → 1 mediator → 1 outcome
3. multiple predictors → outcome(s)

relationship between CFA & SEM

- both CFA and SEM use structural equation modeling procedures
  - * specific mathematical model
  - * theory-driven approach
- CFAs can be conducted alone – focus solely on underlying factor structure
- SEM usually includes CFA (measurement part of the model)
- key difference between the two:
  - * ***CFA*** specifies correlational associations between factors in the model (i.e., bidirectional)
  - * ***SEM*** specifies “causal” associations between factors in the model (i.e., unidirectional) as well as correlations

how CFA and SEM work – (i)

- we supply two things to the software package
  - * data set
  - * model: statement of relationships between variables

- the software first calculates a ***variance-covariance matrix***
  - * observed variances and covariances among variables

- it then estimates ***parameters*** in the model
  - * parameters indicate the nature and size of relationships between variables in the population (correlations or direct paths)
  - * we can never know the true value of a parameter, but statistics help estimate it
  - * parameters are ***fixed*** (i.e., set to be “0”) or ***free*** (to be estimated from the data)

how CFA and SEM work – (ii)

- based on the parameter estimates, the software computes an ***estimated variance-covariance matrix***

- it then compares the ***estimated*** and ***actual*** variance-covariance matrices

- in the end, the software produces two things:

  (a) information regarding the similarity between the estimated and actual variance-covariance matrices
  → *how well the model "fits" the data*

  (b) parameter estimates
  → *nature of relationships between variables*

********** 5 minute break **********

summary of steps in CFA/SEM

1. specify a model

A. decide on the order of association between variables

B. decide whether each parameter should be free or constrained

C. consider the size of your sample

2. evaluate model fit

A. $\chi 2$ test

B. absolute and incremental fit indices

C. residual indices

3. examine modification indices

A. WALD test *(only in Lisrel & EQS, not AMOS)*

B. Lagrange Multiplier test *(in EQS; Modification Indices in AMOS)*

4. test alternative models

A. different orders of association (SEM only)

B. nested models

step 1: specifying a model

A. decide on the order of association between variables

- which variables are exogenous (predictors) and which are endogenous (mediators or outcomes)?
  * if cross-sectional design, you have to decide
- how to decide the order of association:
  * does previous theory and/or research suggest (or dictate) a particular order of association?
  * what is your research question?
- can compare models with different orders of association…
- …***BUT*** it's hard to data-fish in SEM

→ ***should have clear idea of model(s) you want to test***
→ ***should be able to defend your choice(s)***

step 1: specifying a model

B. decide whether each parameter should be free or constrained

- **constrained** = parameter set to be “0”
  (i.e., left out of the model)
- **free** = parameter allowed to be estimated
  (i.e., included in the model)

- need at least one constrained parameter for software to be able to estimate var-cov matrix and assess model fit

- more free parameters = model fits data better
  (fewer constraints = fewer places where “mis-fit” can occur)

- BUT more free parameters = less parsimony & fewer d.f.

→ ***have to balance these two issues***

step 1: specifying a model

B. decide whether each parameter should be free or constrained

<u>HOW TO DECIDE WHICH PARAMETERS TO INCLUDE</u>:

* which relationships do you hypothesize to be important?

* which relationships do you have to control for?

* which variables actually have an association?
  (i.e., correlation > .20)

* which relationships do you want to show to be “0”?

→ ***need to consider theoretical, empirical, and rhetorical questions***

step 1: specifying a model

C. consider the size of your sample

- **issue #1: statistical stability of model**
  * if too few participants, mathematical basis of analysis is unsound, and output should not be trusted
- **issue #2: statistical power**
  * if too few participants, may not be able to detect small effects

<u>SO…HOW MANY PARTICIPANTS DO I NEED</u>?

* ideally, 10+ participants for every estimated parameter
  → *includes factor loadings, direct effects, and correlations*

* if between 5 and 10 participants per estimated parameter
  → *may compromise statistical power*

* if < 5 participants per estimated parameter
  → *will compromise statistical stability of model*

step 2: evaluating model fit

- model fit refers to how similar the ***estimated*** variance-covariance matrix is to the ***actual*** variance-covariance matrix
  → ***more similarity between the two matrices = good fit***

- good fit means that the hypothesized model provides a good account for the actual relationships in the dataset

- good fit does **NOT** mean that the model is “correct”
  → only that it is plausible, and so cannot be rejected

- good fit does **NOT** mean that the model explains a large percentage of variance in the endogenous variables

step 2: evaluating model fit

<u>FIT DEPENDS ON SPECIFIED MODEL, BUT ALSO ON D.F.</u>

* fewer degrees of freedom = fewer constraints = better chance of good fit

* fewer constraints can be due to simplicity of model (i.e., fewer variables)

* fewer constraints can be due to more estimated parameters (not fixed at “0”)

* therefore, models that are simple and/or have more free parameters have a better chance of fitting the data well

→ ***good fit is not necessarily impressive – need to look at model complexity and the # of fixed parameters***

step 2: evaluating model fit

A. χ^2 test

- tests degree of similarity between the estimated variance-covariance matrix and actual variance-covariance matrix

- really a “badness-of-fit” index:
  large χ^2 value and small *p* value means that there ***is*** a significant difference between estimated and actual matrices

* rejecting the null hypothesis = model ***does not*** fit well
* accepting the null hypothesis = model ***does fit*** well

→ ***want a small and non-significant χ^2 value***

step 2: evaluating model fit

A. χ^2 test

DRAWBACKS OF χ^2 TEST:

* very sensitive to sample size
  (larger *N* = more chance of finding significant differences)
* assumption of multivariate normality is often violated

→ ***MUST report χ^2, whether it's good or bad***

→ *if χ^2 test looks bad, you have two options:*

(1) can calculate the χ^2 / degrees of freedom ratio:
    * divide χ^2 value by degrees of freedom
    * if < 2, indicates good fit
(2) if other fit indices suggest good fit, downplay χ^2

step 2: evaluating model fit

B. absolute and incremental fit indices

- represent how much of the variance in the covariance-matrix ***has*** been accounted for by the model
- ***NOT*** testing a null hypothesis
- software calculates the following indices in this category:
  * normed fit index (NFI) * non-normed fit index (NNFI)
  * incremental fit index (IFI) * comparative fit index (CFI)
  * goodness-of-fit index (GFI) * adjusted goodness-of-fit index (AGFI)
- range from 0 to 1, with ***higher*** values indicating better fit
- general standard for good fit = .95 or higher (when $N < 250$)
  * some debate about whether this is too strict

→ ***should report: NFI, IFI, CFI, GFI***

step 2: evaluating model fit

C. residual indices

- represent the ***discrepancies*** ("residuals") between estimated and observed covariances
- ***NOT*** testing a null hypothesis
- software calculates the following indices in this category:
  * SRMR = standardized root mean squared residual
  * RMSEA = root mean square error of approximation
- range from 0 to 1, with ***lower*** values indicating better fit
- general standards for acceptable fit:
  * SRMR = .08 or lower * RMSEA = .06 or lower

→ ***should report both SRMR (use and RMSEA***

→ *RMSEA can be sensitive to Type 1 errors (if N < 250) and outliers*

step 2: evaluating model fit

<u>*COMPARING THE FIT OF DIFFERENT MODELS*</u>

- all three sets of fit indices assess absolute, rather than relative, fit

- ***NEVER*** compare incremental (CFI, GFI, etc.) or residual fit indices (SRMR, RMSEA) between models

  → *there is no way to test whether the difference in fit indices is statistically reliable/significant*

- ***CAN*** use χ^2 test to compare fit of 2 models in **one** case: when two models are nested within each other

  → *more on this in Step 4 (testing alternative models)*

what if your model has bad fit?

- **bad fit =** model does not account for all relationships in data
  → *1 or more fixed parameter needs to be freed*

- **what is wrong?**
  * CFA: an item should load onto other factors
  * SEM: two possible problems…
    (1) measurement: an item should load onto other factors
    (2) structural: a relationship between factors should be added

- **what you need to do:**
  * add paths to model (modification indices can help – see Step 3)

- **BUT remember:**
  * testing repeated models increases Type 1 errors
  * post-hoc modification moves away from *a priori* approach

→ ***keep the model-tweaking to a minimum***
→ ***make sure that your changes make theoretical sense***

step 3: modification indices

A. WALD test → EQS only; not available in AMOS

- tests whether you can ***drop*** any paths you have estimated
  → to help improve parsimony of model and free up d.f.

- using a χ^2 distribution, WALD tests whether ***dropping*** each parameter would significantly ***worsen*** the overall fit of the model
  * modified model = fewer estimated parameters than original

- fewer estimated parameters = worse fit (b/c more constraints)
  * so, the modified model (with fewer estimated parameters) will NEVER fit better than the original
  * the best the modified model can do is to fit just as well as the original

*** *significant*** χ^2 = dropping parameter WOULD worsen model fit
→ ***keep parameter***

*** *non-sig.*** χ^2 = dropping parameter WOULD NOT worsen fit
→ ***can drop parameter***

step 3: modification indices

B. LaGrange Multiplier test → "Modification Indices" in Amos

- tests whether you need to ***add*** any of the paths you left out
  → to better account for relationships in the data

- using a χ^2 distribution, LM tests whether ***adding*** each parameter would significantly ***improve*** overall model fit
  * modified model = more estimated parameters than original

- more estimated parameters = better fit (b/c fewer constraints)
  * modified model (with more estimated parameters) will NEVER fit worse than the original
  * the worst the modified model can do is to fit just as well as the original

**** significant*** χ^2 = adding the parameter WOULD improve model fit
→ ***add parameter***

**** non-sig.*** χ^2 = adding the parameter WOULD NOT sig. improve fit
→ ***parameter not needed***

step 4: testing alternative models

A. different orders of association

- good fit does NOT necessarily mean that your model wins
  * does not discount other models that are as plausible
- a different order of association may be theoretically viable

hypothesized:

perception of discrimination → **ingroup identification** → **identity maint. strategies**

plausible alternative:

ingroup identification → **perception of discrimination** → **identity maint. strategies**

step 4: testing alternative models

A. different orders of association

<u>WHAT TO DO</u>?

→ **test the alternative model**

* examine fit indices: does it have good absolute fit?
* if alternative model has the same # of degrees of freedom as the original, cannot directly compare fit

→ **if** alternative model meets the absolute thresholds for good fit, you **HAVE** to say that both are viable alternatives…

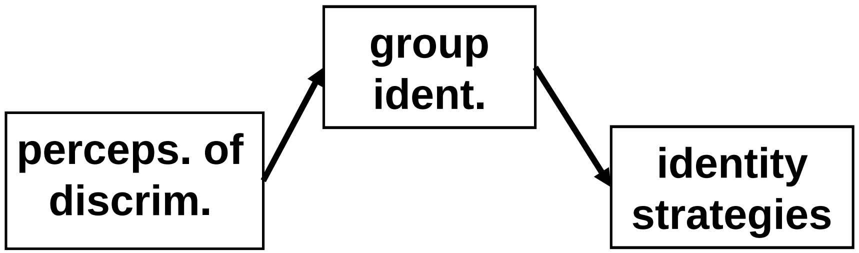
→ **…but you can still compare relative fit:**

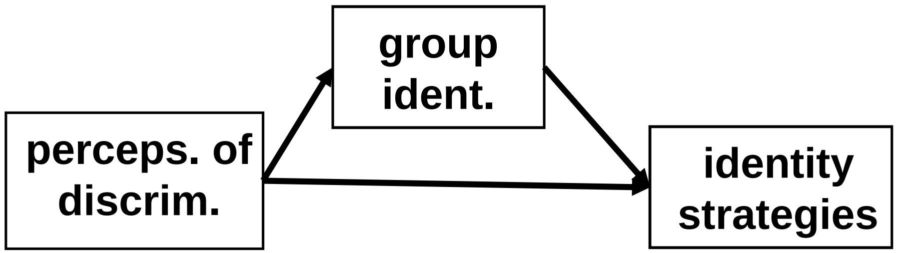
* model with lower AIC value is *relatively* better fitting
* though cannot test reliability or magnitude of this difference

step 4: testing alternative models

B. nested models

- nested model = logical subset of another model
- obtained by changing number of parameters (adding or dropping paths)
- model with fewer parameters is nested within the model with more parameters





→ *models include the same variables, but have different d.f.*

step 4: testing alternative models

B. nested models

<u>WHAT TO DO</u>?

compare χ^2 values to see which model fits better:

1. subtract smaller χ^2 value from larger one
2. subtract smaller degrees of freedom from larger one
3. look up corresponding *p* value in $\chi 2$ table

* ***non-signficant*** χ^2 = no difference in fit of the two models

→ more parsimonious nested model fits as well as larger one

→ larger model offers no advantage, so nested model is better

*** *significant*** χ^2 = sig. difference in the fit of the two models

→ nested model fits significantly worse than larger model

→ larger model is better

variations of SEM

- **problems that may prevent you from using SEM:**
  1. your sample is too small
  2. the model does not fit well because of measurement component
     (e.g., identification items want to load onto well-being factor)

- but you may still be interested in testing a complex model
  → can use a variation of “proper” SEM

- **variation #1: use measured variables instead of latent factors**
  * create scales as you would for regression or ANOVA
  * specify model using measured variables, instead of building factors
  * no longer accounts for measurement error, but can still test complex m’s

- **variation #2: use mix of measured variables and latent factors**
  * maybe only a subset of your measures are problematic
  * or, factors don’t make sense for all variables (e.g., categorical)
  * can test a model with some measured variables and some factors
  * account for only some measurement error, but can still test complex m’s

more advanced possibilities…

(1) compare models between groups

- does a pattern of relationships holds for different groups?
- what to do: specify the model in the two groups, constrain them to be equal, and examine the fit of this constrained model
- **good fit** = no significant differences between the groups
- **bad fit** = there is at least one significant difference
  → *can then establish which parameter estimates are different*

(2) compare two paths in the same model

- is one association/relationship stronger than another?
- what to do: constrain the relevant parameters to be equal, and test the overall fit of the model
- **good fit** = no significant differences between the parameters
- **bad fit** = there is a significant difference between the parameters

********** 10 minute break **********

Kline, R. B. (2004). *Principles and practices of structural equation modeling (2nd Ed)*. New York: Guilford.

Maruyama, G. M. (1998). *Basics of structural equation modeling*. Thousand Oaks, CA: Sage.

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models.* Thousand Oaks, CA: Sage.