Assessing Fit

- Evaluating models
- χ^2
- Absolute and incremental fit indices
- Fit index formulae
- Indices based on residuals
- Hu & Bentler
- Downsides and caveats
- Comparing nested models
- Mathematically equivalent models

Evaluating Models: The First Step

- First, evaluate theory, constructs, and method
- Do the paths make theoretical sense?
  - » have any links been supported in the literature?
  - » have some links never been tested?
  - » have some links previously been disconfirmed?
- What variables have been left out?
  - » assume C causes both A and B but is omitted from the model
  - » then links from A to B or B to A are suspect
- Are measures reliable and valid?
- Any problems in procedure/administration?
- Only then consider model fit
  - » $\chi 2$ and fit indices

Fit: χ^2

- Most common index of fit is χ^2 goodness-of-fit index
- Derived directly from the “fitting function” (F)
  » fitting function is one number that represents the fit between the implied and observed covariance matrices
  » usually it is the sum of squared differences between the matrices (least squares method), but other functions exist
- $\chi^2 = F*(N-1)$
  » where F = the value of the fitting function
  » N = sample size (number of participants)
    - this is where sample size is used
- Called χ^2 because it’s distributed as χ^2 if
  » model is correct
  » endogenous variables have multivariate normal distribution

More About χ^2

- Really a “badness-of-fit” index
  - » so, small numbers mean better fit (0 would be best of all)
- Small p values (e.g. $< .05$) are bad
  - » p value tells us whether we should reject hypothesis that $\chi^2 = 0$
  - » we don’t want to reject that hypothesis
  - » ideally, would like to see $p = .30, .40, .50$
- Recall that χ^2 depends on sample size
  - » more participants means χ^2 will be bigger
  - » so, more likely to reject hypothesis that observed and implied covariance matrices are the same
- χ^2 has some drawbacks
  - » greatly affected by sample size
  - » assumption of mv normality often violated

Still More About χ^2

- Recall that more constraints means good fit is harder to achieve
  » solving for 2 variables with 3 equations is likely to produce better fit than solving for 2 vars with 20 equations
  » more places where misfit can occur
- To adjust or account for this, some people recommend dividing χ^2 by df
  » called *normed chi-square (NC)*
  » χ^2/df <= 2.0 (or 3.0 or even 5.0) considered acceptable
  » remember, df = # of observations - # parameters, NOT number of participants or cases
  » some say to only do this correction if number of participants is large
  » doesn't completely correct for influence of sample size

Example

- Consider two models with the same χ^2
- More parsimonious model (model with fewer parameters to estimate, therefore more df for error) is acceptable by this criterion
- Less parsimonious model not acceptable

χ^2	20	20
df	10	5
χ^2/df	2.0	4.0

Beyond χ^2: Fit Indices

- Descriptive; intuitively interpreted
- Not testing a null hypothesis
- Many different indices
  - » different programs print different ones
- Good to look at (and report) several
  - » but cf. McDonald & Moon Ho: "It is sometimes suggested that we should report a large number of these indices, apparently because we do not know how to use any of them."
- Some caveats:
  - » they measure overall or average fit; can be good even if fit in one portion of the model is quite bad
  - » good values do not guarantee that the model makes theoretical sense
  - » good values do not guarantee or prove that the model is correct

Categories of Fit Indices

- Two main types

1) Absolute and incremental (or relative) fit indices
   - » rough guideline: values above .90 usually considered adequate
   - » some suggest we should require > .95

2) Indices based on residuals matrix
   - » rough guideline: values below .10 or .05 usually considered adequate

- Kline also discusses predictive fit indices
   - » e.g. Akaike information criterion and its derivatives
   - » useful for comparing non-nested models
   - » see Kline p. 142, pp. 151-153

Absolute and Incremental Indices

- Represent how much of the variance in the covariance matrix has been accounted for
- Not testing a null hypothesis
- Examples:
  - » comparative fit index (CFI)
  - » normed fit index (NFI)
  - » non-normed fit index (NNFI) (aka Tucker-Lewis index, TLI)
  - » incremental fit index (IFI)
  - » goodness-of-fit index (GFI)
  - » adjusted goodness-of-fit index (AGFI)

Absolute vs. Incremental

- Absolute indices are computed using formulae that include discrepancies (matrix of residuals) and df and sample size
  - » not comparing this model to any model; hence the label "absolute"
- Formulae for incremental or relative indices also include discrepancies from a "null" model
  - » hence the labels "incremental" or "relative" or "comparative"

Formula elements

- Will present formulae for some fit indices
- First, define δ as the degree of misspecification of the model
  - » if $\delta = 0$, model is not misspecified
- Subtract df for the model from the χ^2 for the model
  - » but don't let δ be less than zero
- So, not requiring fit to be perfect (i.e., χ^2 = zero)

$$\hat{\delta}_{\mathrm{M}} = \max\left(\chi^2_{\mathrm{M}} - df_{\mathrm{M}}, 0\right)$$

CFI: Comparative Fix Index

$$\mathrm{CFI} = 1 - \frac{\hat{\delta}_{\mathrm{M}}}{\hat{\delta}_{\mathrm{B}}}$$

- Compares performance on your model to performance on baseline ("null" or "independence") model
  - » baseline model assumes zero correlation between all observed variables
- Proposed by Bentler
  - » hence, it first appeared in EQS, but is printed by Lisrel and Amos now, also
- Recommended by Kline and Hu & Bentler
  - » better than some other indices for small samples
  - » criticism: baseline model implausible

NFI: Normed Fix Index

$$\text{NFI} = 1 - \frac{\chi^2_{\text{M}}}{\chi^2_{\text{B}}}$$

- Also comparing model to baseline
- Not adjusting for df of either model
  - » i.e., no bonus for parsimony
  - » sensitive to sample size (N)
- Proposed by Bentler & Bonett

NNFI: Non-normed Fix Index

$$\text{NNFI} = 1 - \frac{\chi^2_{\text{M}} / df_{\text{M}}}{\chi^2_{\text{B}} / df_{\text{B}}}$$

- Proposed by Bentler & Bonett
  - » also called the Tucker-Lewis index
- Still comparing model to baseline
- Does adjust for df
- Can fall outside 0.0-1.0 range
- Recommended by Hu & Bentler

Absolute Fit Indices: GFI & AGFI

- Other indices do not compare to a baseline model -- "absolute" indices
- Jöreskog proposed two measures based on % variance explained

$$\mathrm{GFI} = 1 - \frac{v_{\mathrm{residual}}}{v_{\mathrm{total}}}$$

- v_{residual} = residual variance in covariance matrix (variance that can't be explained by the model)
- v_{total} = total variance in the covariance matrix
- GFI: goodness of fit
- AGFI adjusts for number of parameters

More on GFI & AGFI

- Can fall outside 0.0-1.0 range
- Hu & Bentler
  - » GFI and AGFI do not perform as well with latent variable models as with path models
  - » too many Type I errors when N is large (accepting models that should be rejected)
- Kline
  - » AGFI has not performed well in simulation studies
- GFI still often reported, perhaps because it was the first fit index proposed

Residuals-Based Indices

- Indices based on residuals matrix
  - » looks at discrepancies between observed and predicted covariances
  - » e.g., root mean square error of approximation (RMSEA), standardized root mean squared residual (SRSR)
  - » values below .05 (or, sometimes, .10) considered adequate

RMSEA

- RMSEA: Root Mean Square Error of Approximation
- Error of Approximation
  - » lack of fit of our model to population data, when parameters are optimally chosen
- vs. Error of Estimation
  - » lack of fit of our model (with parameters chosen via fitting to the sample data) to population data (with optimally-chosen parameters)
- Error of estimation is affected by sample size, but error of approximation is not

RMSEA

$$RMSEA = \sqrt{\frac{\hat{\delta}_{\mathrm{M}}}{df_{\mathrm{M}}(N-1)}}$$

- Recall that δ = degree of misspecification of the model
- Kline:
  - » <= .05 "close approximate fit"
  - » > .05 but < .08 "reasonable approximate fit"
  - » > = .10 "poor fit"
  - » also suggests a method for using confidence intervals (see pp. 139-140)

RMR and Standardized RMR

- RMR = root mean square residual
- Simply the mean absolute value of the covariance residuals
- Standardized formula puts it into a metric that is easily interpretable
- A "badness of fit" index (in that higher numbers mean worse fit)
- Values of standardized RMR < .10 generally considered adequate

Hu & Bentler (1995)

- Reviewed research on fit indices
  - » usually, simulation studies
- Found that fit statistics were not robust across
  - » sample size
  - » estimation method (e.g., maximum likelihood vs. generalized least squares)
  - » distribution of latent variables (e.g., violations of multivariate normality)
- For maximum likelihood, large samples (N > 250) and independent latent variables, H&B recommend
  - » NNFI, IFI, CFI, MacDonald (MFI)
- Argue against .90 benchmark
  - » instead, use indices to compare/contrast competing models

Interpreting & Reporting

- Don't rely on just one fit index
- Don't pick the indices to interpret/report based on the ones that are most supportive of your model
- Decide on a set of fit indices you will generally rely on and go with those every time
  - » suggestions in Kline, Hu & Bentler, Hoyle & Panter, McDonald & Moon Ho
- Ideally, different fit indices will point to the same conclusion
- If fit indices lead to different conclusion, conservative choice is to reject the model

Down Side of Fit Indices

- Fit indices are useful because they summarize fit
  - » reduce it to one number
- The flip side of this is that the fit indices obscure the reason for good (or bad) fit
- Is misfit scattered about most or all of the cells in the residual matrix?
  - » if so, this is probably the best we can do
- Or is misfit limited to a few cells?
  - » if so, model may be misspecified
  - » may be able to improve model dramatically by adding one or two paths
- Good overall fit may hide the fact that one or two key theoretical relationships are not being accurately reproduced
- Moral: always look at your residual matrix

Model Fit: Caveats

- Good (or perfect) fit does not ensure that model is correct, only that it is plausible
  - » be careful of your language, when presenting results
  - » good to test alternative models
    - if you can reject these, the argument for your model is strengthened
  - » good to also discuss equivalent models (more in a moment)
- Fit different than predictive power
  - » the "% variance" indices represent how much of the variance/covariance matrix you can explain
    - **not** how much of the variance in latent or observed endogenous variables is explained
  - » could have perfect fit but not explain very much variance in any of the endogenous variables

Model Fit: Caveats

- Values of fit indices depend critically on how many degrees of freedom were available to estimate fit
  - » no degrees of freedom available
    - "saturated model"
    - fit is <u>always</u> perfect (for recursive models)
  - » few degrees of freedom available
    - not much "room" for disconfirming
    - fit is likely to be very good
  - » many df available
    - fit could be quite bad
    - thus, good fit is more impressive
  - » this is why the various "adjusted" fit indices were developed

Comparing Models: Nested Models

- A nested model is a logical subset of another model
  - » obtained by estimating fewer parameters (i.e., dropping paths)
  - » also called “hierarchical models”
- When fewer parameters are estimated (i.e., more df), fit will always be worse
  - » the question is, how much worse?
  - » if fit is almost as good, the nested model (with fewer parameters) is preferred because it’s more parsimonious
  - » what is “almost as good”?
- One of the few places in SEM where we do classic significance testing

Nested Models: Comparing χ^2s

- Compute the difference between the χ^2 values for the two models
  - » $\chi2$ for the reduced model <u>cannot</u> be smaller than the $\chi2$ for the model it is nested under
  - » logically, it must be larger (or the same) because fewer parameters are being estimated -- impossible for fit to be better
- Compute the difference between the df for the two models
- Look up *p* value in $\chi2$ table
- Example
  - » Baseline (or full) model: $\chi^2(4) = 3.21$
  - » Reduced model: $\chi^2(6) = 11.41$
  - » $\Delta\chi^2 = 8.2$; for 2 df, $p < .05$ (χ^2 for 1 df = 3.84)
  - » fit is significantly worse without those two paths

Comparing Models: Equivalent Models

- There are (almost) always alternate models that are mathematically equivalent
  - » will have identical fit
- There are mathematical rules for generating equivalent models
  - » Lee-Herschberger replacing rules
  - » e.g., X-->Z can be replaced by
    - Z-->X
    - D_X <--> D_Z
    - X-->Z and Z-->X (with equality constraint)
- Equivalent models not given enough attention in the literature

Comparing Models

- Useful to generate at least some equivalent models
  - » some can be rejected on basis of logic, theory or prior data
  - » others cannot be rejected; this is good to know
- Non-equivalent, non-nested competing models
  - » there may be models that are not mathematically equivalent, but have similar fit to your preferred model
  - » if the models aren't nested, cannot compare $\chi 2$ directly (although this is often done)
  - » rely on the "art" of comparing fit indices
  - » also rely on theory, previous research, logic
  - » also, AIC-type indices for non-nested models