Multivariate Models: Assessment of Absolute Model Fit

EPSY 905: Fundamentals of Multivariate Modeling Online Lecture #11



In This Lecture

 How to assess the fit of a multivariate linear model to the data, in an absolute sense



EXAMPLE DATA SET



EPSY 905: Multivariate Models Absolute Model Fit

Data are simulated based on the results reported in:
 Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: a path analysis. *Journal of Educational Psychology, 86*, 193-203.

Sample of 350 undergraduates (229 women, 121 men)

 In simulation, 10% of variables were missing (using missing completely at random mechanism)

- Note: simulated data characteristics differ from actual data (some variables extend beyond their official range)
 - > Simulated using Multivariate Normal Distribution
 - Some variables had boundaries that simulated data exceeded

Results will not match exactly due to missing data and boundaries KU KANSAS

Variables of Data Example

- Female (sex variable: 0 = male; 1 = female)
- Math Self-Efficacy (MSE)
 - Reported reliability of .91
 - > Assesses math confidence of college students
- Perceived Usefulness of Mathematics (USE)
 - Reported reliability of .93
- Math Anxiety (MAS)
 - Reported reliability ranging from .86 to .90
- Math Self-Concept (MSC)
 - Reported reliability of .93 to .95
- Prior Experience at High School Level (HSL)
 - Self report of number of years of high school during which students took mathematics courses
- Prior Experience at College Level (CC)
 - Self report of courses taken at college level
- Math Performance (PERF)
 - Reported reliability of .788
 - > 18-item multiple choice instrument (total of correct responses)



METHODS OF EXAMINING MODEL FIT



EPSY 905: Multivariate Models Absolute Model Fit

Picking Up From Last Lecture...

- To demonstrate how models may vary in terms of model fit (and to set up a discussion of model fit and model comparisons) we will estimate a model where we set the covariance between PERF and USE to zero
 - > Zero covariance implies zero correlation which is unlikely to be true given our previous analysis
- You likely would not use this model in a real data analysis
 > If anything, you may start with a zero covariance and then estimate one
- But, this will help to introduce come concepts needed to assess the quality of the multivariate model



Model Notation

- The multivariate model for PERF and USE is given by two regression models, which are estimated simultaneously: $PERF_i = \beta_{0,PERF} + e_{i,PERF}$ $USE_i = \beta_{0,USE} + e_{i,USE}$
- As there are two variables, the error terms have a joint distribution that will be multivariate normal:

$$\begin{bmatrix} e_{i,PERF} \\ e_{i,USE} \end{bmatrix} \sim N_2 \left(\mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{R} = \begin{bmatrix} \sigma_{e,PERF}^2 & 0 \\ 0 & \sigma_{e,USE}^2 \end{bmatrix} \right)$$

 Each error term has its own variance and we do not estimate the covariance



• The lavaan() syntax for setting the covariance to zero is:

```
model03.syntax = "
#Variances:
    perf ~~ perf
    use ~~ use
#Covariance:
    perf ~~ 0*use
#Means:
    perf ~ 1
    use ~ 1
"
```

• The only difference is 0* in front of USE in the ~~ section

> You can set a parameter equal to a value this way



Model 3

Covariances:				
	Estimate	Std.Err	z-value	P(> z)
perf ~~				
use	0.000			
Intercepts:				
	Estimate	Std.Err	z-value	P(> z)
perf	13.966	0.174	80.397	0.000
use	52.500	0.874	60.047	0.000
Variances:				
	Estimate	Std.Err	z-value	P(> z)
perf	8.751	0.756	11.581	0.000
use	249.201	19.212	12.971	0.000



Examining Model/Data Fit





Methods of Model Fit

- Model-data fit is of utmost concern when building models with multivariate outcomes
- If a model does not fit the data:
 - Parameter estimates may be biased
 - Standard errors of estimates may be biased
 - Inferences made from the model may be wrong
 - > If the saturated model fit is wrong, then the LRTs will be inaccurate
- Examining model fit is the first step in multivariate models
- That said, not all "good-fitting" models are useful...
 - ...model fit just allows you to talk about your model...there may be nothing of significance (statistically or practically) in your results, though



GLOBAL MEASURES OF MODEL FIT



Types of Model Fit Information

- Model fit information for models where outcomes are <u>conditionally MVN*</u> come in several types, but all are based on the premise that any model mean and covariance structure must fit <u>as well as</u> the saturated mean vector and covariance matrix model
 *If model outcomes are not conditionally MVN, model fit is very different
- All possible models/structures **are nested within** the saturated mean vector and covariance matrix model
 - Most model fit statistics come from comparing any model/structure with the saturated model
- Indices shown first are called "global" model fit indices

> Report fit of model globally (as opposed to locally for specific parameters)



Example lavaan Model Fit Output

> summary(model03.fit, fit.measures=TRUE)

lavaan (0.5-23.1097) converged normally after 18 iterations

Number of observations	Used 348	Total 350	
Number of missing patterns	3		
Estimator Minimum Function Test Statistic Degrees of freedom P-value (Chi-square) Scaling correction factor for the Yuan-Bentler correction (Mplu	ML 6.064 1 0.014 us variant)	Robust 5.573 1 0.018 1.088	
Model test baseline model:			
Minimum Function Test Statistic Degrees of freedom P-value	6.064 1 0.014	5.573 1 0.018	
User model versus baseline model:			
Comparative Fit Index (CFI) Tucker-Lewis Index (TLI)	0.000 -0.000	0.000 0.000	
Robust Comparative Fit Index (CFI) Robust Tucker-Lewis Index (TLI)		0.000 0.000-0-	
Loglikelihood and Information Criteria:			
Loglikelihood user model (H0) Scaling correction factor for the MLR correction	-2088.064	-2088.064 1.012	
Loglikelihood unrestricted model (H1) Scaling correction factor for the MLR correction	-2085.032	-2085.032 1.028	
Number of free parameters Akaike (AIC) Bayesian (BIC) Sample-size adjusted Bayesian (BIC)	4 4184.128 4199.537 4186.848	4 4184.128 4199.537 4186.848	
Root Mean Square Error of Approximation:			
RMSEA 90 Percent Confidence Interval P-value RMSEA <= 0.05	0.121 0.044 0.220 0.063	0.115 0.040 0.073	0.210
Robust RMSEA 90 Percent Confidence Interval		0.120 0.039	0.224
Standardized Root Mean Square Residual:			



SRMR

The fit.measures=TRUE Model Fit Statistics

Unlabeled section

- > Likelihood ratio test versus the saturated model
- > Testing if your model fits as well as the saturated model
- Model test baseline model
 - > Likelihood ratio test pitting the saturated model against the independent variables model
 - > Testing whether any variables have non-zero covariances (significant correlations)
- User model versus baseline model
 - > CFI
 - > TLI
- Loglikelihood and Information Criteria
 - Likelihood ratio tests (nested models)
 - Information criteria comparisons (non-nested models)
- Root Mean Square Error of Approximation
 - > How far off a model is from the saturated model, per degree of freedom
- Standardized Root Mean Square Residual
 - How far off a model's correlations are from the saturated model correlations



Indices of Global Model Fit

- Primary: obtained model χ² (from Model test baseline model)

 here we use the MLR rescaled χ² from the "Robust" Column (next lecture)
 - > χ^2 is evaluated based on model df (difference in parameters between your CFA model and the saturated model)
 - > Tests null hypothesis that **this** model (H_0) fits equally to **saturated model** (H_1) so significance is undesirable (smaller χ^2 , bigger p-value is better)
 - Means saturated model is estimated automatically for each model analyzed
 - > Just using χ^2 is insufficient, however:
 - Distribution doesn't behave like a true χ² if sample sizes are small (or, if not using MLR, if items are non-normally distributed)
 - Obtained χ^2 depends largely on sample size
 - Some mention this is an unreasonable null hypothesis (perfect fit??)
- Because of these issues, alternative measures of fit are usually used in conjunction with the χ^2 test of model fit
 - > Absolute Fit Indices (besides χ^2)

Parsimony-Corrected; Comparative (Incremental) Fit Indices EPSY 905: Multivariate Models Absolute Model Fit



Chi-Square Test of Model Fit

- The Chi-Square Test of Model Fit provides a likelihood ratio test comparing the current model to the saturated (unstructured) model:
 - > The value is -2 times the difference in log-likelihoods (rescaled if MLR)
 - The degrees of freedom is the difference in the number of estimated model parameters
 - > The p-value is from the Chi-square distribution

• If this test has a significant p-value:

- The current model (H₀) is rejected the model fit is significantly worse than the full model
- > In latent variable models, this test is usually ignored
 - Said to be overly sensitive

If this test does not have a significant p-value:

> The current model (H_0) is not rejected – **fits equivalently to full model**

Where the Saturated Model Test Comes From

- The saturated model LRT comes from a likelihood ratio test of the current model with the saturated model
- If using MLR (Robust method), then this LRT is rescaled based on the estimated scaling factors of both models
- This same information can be obtained from:
 - > Loglikelihood model output section
 - anova() function comparing fit for current and saturated models



Calculating the LRT for Global Fit Test for Model 03

• From the lavaan output:

Estimator	ML
Minimum Function Test Statistic	6.064
Degrees of freedom	1
P-value (Chi-square)	0.014
Scaling correction factor	
for the Yuan-Bentler correction (Mplus	variant)

Robust Loglikelihood and Information Criteria:

88.064
1.012
85.032
1.028

- Calculation:
 - > 14 parameters in our model; 20 in saturated model
 - Scaling correction factor:

$$c_{LR} = \left| \frac{(q_{restricted})(c_{restricted}) - (q_{full})(c_{full})}{(q_{restricted} - q_{full})} \right| = 1.088$$

> $\chi^2 = \frac{6.064}{1.008} = 5.573$
> DF = 1

Conclusion: this model fit significantly worse than the saturated model



Saturated Model LRT and Loglikelihood Output

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-2088.064	-2088.064
Scaling correction factor		1.012
for the MLR correction		
Loglikelihood unrestricted model (H1)	-2085.032	-2085.032
Scaling correction factor		1.028
for the MLR correction		

- If the loglikelihoods of the current model ("User model" or H₀) are equal to the loglikelihoods of the saturated model ("Unrestriced model" or H₁), then you are running a model that is equivalent to the saturated model
 - No other model fit will be available or useful



The fit.measures=TRUE Model Fit Statistics

Unlabeled section

> Likelihood ratio test versus the saturated model

> Testing if your model fits as well as the saturated model

Model test baseline model

> Likelihood ratio test pitting the saturated model against the independent variables model

- > Testing whether any variables have non-zero covariances (significant correlations)
- User model versus baseline model
 - > CFI
 - > TLI
- Loglikelihood and Information Criteria
 - Likelihood ratio tests (nested models)
 - > Information criteria comparisons (non-nested models)
- Root Mean Square Error of Approximation
 - > How far off a model is from the saturated model, per degree of freedom
- Standardized Root Mean Square Residual
 - How far off a model's correlations are from the saturated model correlations



Model Test Baseline Model

• The "model test baseline model" section provides a LRT:

 Comparing the saturated (unstructured) model with an independent variables model (zero covariance)

Model test baseline model:

Minimum Function Test Statistic	6.064	5.573
Degrees of freedom	1	1
P-value	0.014	0.018

- Here, the "null" model is the baseline
 - If the test is significant, this means that at least one (and likely more than one) variable has a significant covariance (and correlation)
 - If the test is not significant, this means that the independence model is appropriate
 - This is not likely to happen
 - But if it does, there are virtually no other models that will be significant
- Not often reported as it is likely variables are correlated



The fit.measures=TRUE Model Fit Statistics

Unlabeled section

> Likelihood ratio test versus the saturated model

Testing if your model fits as well as the saturated model

Model test baseline model

Likelihood ratio test pitting the saturated model against the independent variables model
 Testing whether any variables have non-zero covariances (significant correlations)

User model versus baseline model

- > CFI
- > TLI
- Loglikelihood and Information Criteria
 - Likelihood ratio tests (nested models)
 - Information criteria comparisons (non-nested models)
- Root Mean Square Error of Approximation
 - > How far off a model is from the saturated model, per degree of freedom
- Standardized Root Mean Square Residual
 - \succ How far off a model's correlations are from the saturated model correlations \mathbf{v}

User Model Versus Baseline Model Section

 The "User model versus baseline model" section provides two additional measures of model fit comparing the current (user) model to the baseline (independent variables) model

User model versus baseline model:

Comparative Fit Index (CFI)	0.000	0.000
Tucker-Lewis Index (TLI)	-0.000	-0.000
Robust Comparative Fit Index (CFI)		0.000
Robust Tucker-Lewis Index (TLI)		-0.000

• CFI stands for Comparative Fit Index

> Higher is better (above .95 indicates good fit)

- TLI stands for Tucker Lewis Index
 - > Higher is better (above .95 indicates good fit)

Comparative (Incremental) Fit Indices

• Fit evaluated relative to a 'null' model (of 0 covariances)

> Relative to that, your model should be great!

• CFI: Comparative Fit Index

T = target (current/estimated) model N = null (baseline/independent variables) model

> Based on idea of the chi-square non-centrality parameter: $(\chi^2 - df)$

>
$$CFI = 1 - \frac{\max(\chi_T^2 - df_T, 0)}{\max(\chi_T^2 - df_T, \chi_N^2 - df_N, 0)}$$

> From 0 to 1: bigger is better, > .90 = "acceptable", > .95 = "good"

TLI: Tucker-Lewis Index (= Non-Normed Fit Index)

$$\succ TLI = \frac{\frac{\chi_N^2}{df_N} - \frac{\chi_T^2}{df_T}}{\frac{\chi_N^2}{df_N} - 1}$$

> From <0 to >1, bigger is better, >.95 = "good"



Information Criteria Output

• The information criteria output provides relative fit statistics:

Number of free parameters	4	4
Akaike (AIC)	4184.128	4184.128
Bayesian (BIC)	4199.537	4199.537
Sample-size adjusted Bayesian (BIC)	4186.848	4186.848

- > AIC: Akaike Information Criterion
- > BIC: Bayesian Information Criterion (also called Schwarz's criterion)
- Sample-size Adjusted BIC
- These statistics weight the information given by the parameter values by the parsimony of the model (the number of model parameters)
 - > For all statistics, the smaller number is better
- The core of these statistics is -2*log-likelihood



The fit.measures=TRUE Model Fit Statistics

Unlabeled section

> Likelihood ratio test versus the saturated model

> Testing if your model fits as well as the saturated model

Model test baseline model

Likelihood ratio test pitting the saturated model against the independent variables model
 Testing whether any variables have non-zero covariances (significant correlations)

- User model versus baseline model
 CFI
 TH
- Loglikelihood and Information Criteria
 - > Likelihood ratio tests (nested models)
 - > Information criteria comparisons (non-nested models)
- Root Mean Square Error of Approximation
 - > How far off a model is from the saturated model, per degree of freedom
- Standardized Root Mean Square Residual
 - How far off a model's correlations are from the saturated model correlations



Comparing Information Criteria

Information criteria are relative tests of fit

Number of free parameters	4	4
Akaike (AIC)	4184.128	4184.128
Bayesian (BIC)	4199.537	4199.537
Sample-size adjusted Bayesian (BIC)	4186.848	4186.848

- The are calculated based on the log-likelihood of the model, factoring in a penalty for number of parameters (plus other things)
- They should never be used to compare nested models
 - > The likelihood ratio test is the most powerful test statistic to use for nested models
- When comparing non-nested models, first choose a statistic
 > AIC, BIC, or Sample-size Adjusted BIC are what are given by default
- The preferred model is the one with the lowest value of that statistic



The fit.measures=TRUE Model Fit Statistics

Unlabeled section

> Likelihood ratio test versus the saturated model

> Testing if your model fits as well as the saturated model

Model test baseline model

Likelihood ratio test pitting the saturated model against the independent variables model
 Testing whether any variables have non-zero covariances (significant correlations)

- Loglikelihood and Information Criteria
 - > Likelihood ratio tests (nested models)
 - Information criteria comparisons (non-nested models)
- Root Mean Square Error of Approximation
 - > How far off a model is from the saturated model, per degree of freedom
- Standardized Root Mean Square Residual
 - \succ How far off a model's correlations are from the saturated model correlations \mathbf{v}



Parsimony-Corrected: RMSEA

- Root Mean Square Error of Approximation
- Uses comparison with CFA model and saturated model

> χ^2 listed here from first part of lavaan output

- Relies on a non-centrality parameter (NCP)
 - > Indexes how far off your model is $ightarrow \chi^2$ distribution shoved over
 - > NCP → d = (χ^2 df) / (N-1) Then, RMSEA = SQRT(d/df)
 - df is difference between # parameters in CFA model and saturated model
 - > RMSEA ranges from 0 to 1; smaller is better
 - < .05 or .06 = "good", .05 to .08 = "acceptable", .08 to .10 = "mediocre", and >.10 = "unacceptable"
 - In addition to point estimate, get 90% confidence interval
 - RMSEA penalizes for model complexity it's discrepancy in fit per df left in model (but not sensitive to N, although CI can be)
 - > Test of "close fit": null hypothesis that RMSEA \leq .05

RMSEA (Root Mean Square Error of Approximation)

 The RMSEA is an index of model fit where 0 indicates perfect fit (smaller is better):

Root Mean Square Error of Approximation:

RMSEA		0.121	0.115	
90 Percent Confidence Interval	0.044	0.220	0.040	0.210
P-value RMSEA <= 0.05		0.063	0.073	
Robust RMSEA			0.120	
90 Percent Confidence Interval			0.039	0.224

- RMSEA is based on the approximated covariance matrix
- The goal is a model with an RMSEA less than .05
 > Although there is some flexibility
- The result above indicates our model fits poorly (Robust RMSEA of .120)



The fit.measures=TRUE Model Fit Statistics

Unlabeled section

> Likelihood ratio test versus the saturated model

> Testing if your model fits as well as the saturated model

Model test baseline model

Likelihood ratio test pitting the saturated model against the independent variables model
 Testing whether any variables have non-zero covariances (significant correlations)

User model versus baseline model CFI TII

Loglikelihood and Information Criteria

- Likelihood ratio tests (nested models)

> Information criteria comparisons (non-nested models)

Root Mean Square Error of Approximation How far off a model is from the saturated model, per degree of freedom

Standardized Root Mean Square Residual

How far off a model's correlations are from the saturated model correlations

Standardized Root Mean Squared Residual

- The SRMR (standardized root mean square residual) provides the average standardized difference between:
 - > The estimated covariance matrix of the saturated model
 - > The estimated covariance matrix of the current model

Standardized Root Mean Square Residual:

SRMR

0.066 0.066

• Lower is better (some suggest less than 0.08)



LOCAL MODEL FIT MEASURES



"Local" Measures of Model (Mis)Fit

- Local measures of model (mis)fit are statistics that point to the location (typically of a covariance matrix) where a model may not fit well
 - As opposed to "global" measures that indicate a model fit overall
- Local measures of model (mis)fit are typically of two types:
 - > Residual covariance matrices (unstandardized, standardized, or normalized)
 - The difference between the model's estimated covariance matrix and the saturated model's estimated covariance matrix
 - These were used for the SRMR
 - > Model "modification indices"
 - 1-degree of freedom hypothesis tests for the improvement of the model LRT if one more parameter was allowed to be estimated



Residual Covariance Matrices

- Residual covariance matrices are used to figure out how to best improve model misfit
- The "raw" or "unstandardized" residual covariance matrix for the model literally takes the difference between model implied and saturated model covariance matrices
- I often prefer "normalized" versions of these matrices
 - > We can inspect the normalized residual covariance matrix (like z-scores) to see where our biggest misfit occurs > residuals(model03.fit, type = "normalized")

-0.037 -0.069

\$type
[1] "normalized"
\$cov
 perf use
perf -0.012
use 2.403 0.002
\$mean
 perf use



Modification Indices: More Help for Fit

- As we used Maximum Likelihood to estimate our model, another useful feature is that of the modification indices
 - Modification indices, also called Score or LaGrange Multiplier tests, attempt to suggest the change in the log-likelihood for adding a given model parameter (larger values indicate a better fit for adding the parameter)

> modindices(model03.fit)
 lhs op rhs mi mi.scaled epc sepc.lv sepc.all sepc.nox
3 perf ~~ use 6.239 5.734 7.125 7.125 0.153 0.153

- <u>mi column</u>: the expected value of the LRT of the current model and a model where this parameter was added
- mi.scaled column: the scaled (robust) LRT
 - Should be bigger than 3.84 for 1 df
 - Practice is to find values that are much higher (say 10 or more)
- <u>epc column</u>: expected value of the parameter in the model where this parameter was added

COMPARING INDICES



Model Fit for the Model with Covariance Estimated

> summary(model02.fit, fit.measures=TRUE)

lavaan (0.5-23.1097) converged normally	after 29 iteratio	ons	Loglikelihood and Information Criteria:			
	Used	Total	Loglikelihood user model (H0)	-2085.032	-2085.032	
Number of observations	348	350	Loglikelihood unrestricted model (H1)	-2085.032	-2085.032	
Number of missing patterns	3		Number of free parameters	5	5	
51			Akaike (AIC)	4180.064	4180.064	
Estimator	ML	Robust	Bayesian (BIC)	4199.325	4199.325	
Minimum Function Test Statistic	0.000	0.000	Sample-size adjusted Bayesian (BIC)	4183.464	4183.464	
Degrees of freedom	0	0				
Minimum Function Value	0.000000000000		Root Mean Square Error of Approximation:			
Scaling correction factor		NA		0.000	0 000	
for the Yuan-Bentler correction (Mp	lus variant)		KMSEA	0.000	0.000	0 000
			B value BMSEA - 0.05	0.000 0.000	0.000	0.000
Model test baseline model:			P-Value RMSEA <= 0.05	NA	NA	
			Robust RMSEA		0.000	
Minimum Function Test Statistic	6.064	5.573	90 Percent Confidence Interval		0.000	0.000
Degrees of freedom	1	1				
P-value	0.014	0.018	Standardized Root Mean Square Residual:			
User model versus baseline model:			SRMR	0.000	0.000	
	4 . 000					
Comparative Fit Index (CFI)	1.000	1.000				
Tucker-Lewis Index (TLI)	1.000	1.000				
Robust Comparative Fit Index (CFI)		NA				
Robust Tucker-Lewis Index (TLI)		NA				

40

Model Fit for the Univariate Model

> summary(model01.fit, fit.measures=TRUE)

lavaan (0.5-23.1097) converged normally after 8 iterations

	llead	Total				
Number of observations	290	350	Loglikelihood and Information Criteria:			
Number of missing patterns	1		Loglikelihood user model (H0) Loglikelihood unrestricted model (H1)	-726.014 -726.014	-726.014 -726.014	
Estimator	ML	Robust	Number of free parameters	2	2	
Minimum Function Test Statistic	0.000	0.000	Akaike (AIC)	1456.028	1456.028	
Degrees of freedom	0	0	Bayesian (BIC)	1463.367	1463.367	
Scaling correction factor		NA	Sample-size adjusted Bayesian (BIC)	1457.025	1457.025	
for the Yuan-Bentler correction (Mplus	s variant)		Root Mean Square Error of Approximation:			
Model test baseline model:			RMSEA	0.000	0.000	0 000
Minimum Function Test Statistic	0.000	0.000	$P_{\rm V}$ Percent Confidence Interval P_Value RMSEA <= 0.05	0.000 0.000 NA	0.000 NA	0.000
Dearees of freedom	0	0				
P-value	NA	NA	Robust RMSEA		0.000	
			90 Percent Confidence Interval		0.000	0.000
User model versus baseline model:			Standardized Root Mean Square Residual:			
Comparative Fit Index (CFI)	1.000	1.000	CDMD	0,000	0 000	
Tucker-Lewis Index (TLI)	1.000	1.000	SKMK	0.000	0.000	
Robust Comparative Fit Index (CFI)		NA				
Robust Tucker-Lewis Index (TLI)		NA				



WRAPPING UP



Wrapping Up

- The biggest difference between univariate and multivariate models is with respect to model fit
 > Univariate models are saturated models that fit perfectly
- Global model fit helps determine absolute model fit overall
- Local model fit helps determine where model fit can be improved by adding additional parameters
- But what about model comparisons? That's next...

