

Power, Precision, and Sample Size Calculations

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

Power, Precision, and Sample Size Calculations

- 1 Introduction
- 2 Hypothesis Testing and Fit Evaluation: What, Where, How, and Why
 - Testing the Model for Perfect Fit
 - Testing for Close Fit
 - Testing for Not-Close Fit
 - Testing Individual Parameters
- 3 Power Charts for Tests of Overall Model Fit
- 4 Sample Size Calculations for Tests of Model Fit
 - Sample Size Tables
- 5 Interval Estimation Approaches
 - Changing the Emphasis – AIPE

Introduction

- One of the key decisions prior to initiating a research study using SEM is the choice of a sample size for the study.
- We've already seen how the wrong choice can lead to a high probability of failure in the case of confirmatory factor analysis, because of a lack of convergence.
- Besides issues of convergence, we also have the problems of power and precision, which arise in all areas of statistics.
- In this module, we'll review several approaches to power, precision, and sample size estimation in SEM, and conclude with some computational examples.
- We'll stick to the multivariate normal model with single samples, but many of the points and techniques discussed here will generalize to other situations.
- We'll begin with classical hypothesis testing considerations, then move on to the more modern confidence interval perspective.

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

- Let's begin by recalling some of the standard types of hypotheses that are tested in SEM.
 - 1 Evaluating whether a model fits perfectly.
 - 2 Evaluating whether it fits "significantly worse than good."
 - 3 Evaluating whether it fits "significantly better than bad."
 - 4 Testing whether one model fits better than another model it is nested within.
 - 5 Evaluating the badness of fit of a model with a point estimate and a confidence interval.
 - 6 Estimating model parameters and establishing a confidence interval.

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing the Model for Perfect Fit

- The standard χ^2 statistic tests for perfect model fit.
- If fit is perfect, this statistic has an asymptotic chi-square distribution with $p(p+1)/2 - t$ degrees of freedom, where p is the number of variables and t the number of (truly) free parameters in the model.
- This statistic is calculated in maximum likelihood estimation as

$$X = kF_{ML} \quad (1)$$

where F_{ML} is the standard Maximum Wishart Likelihood discrepancy function, and k is a scaling constant usually equal to $n - 1$.

- Other scaling constants — in particular one proposed by Swain in the context of covariance structure modeling and another by Bartlett in the context of factor analysis, may be used to attempt to improve performance at small sample sizes.
- Of course, if fit is not perfect in the population, then the test statistic will no longer have a χ^2 distribution.

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing the Model for Perfect Fit

- Strictly speaking, if the population covariance structure model does not fit the population Σ perfectly, then the distribution of the test statistic diverges.
- We might choose to measure how badly a model fits in the population by fitting the model to the population Σ with the method of maximum likelihood and using the “population discrepancy function” F^* as our measure of misfit.
- Steiger, Shapiro, and Browne (1985) invoked a special assumption called “population drift,” which imagines that the population parameters, rather than being stable, are drifting along with sample size toward a point at which model fit is perfect.
- This assumption models the situation in which fit is not too bad relative to sample size.
- This assumption leads to a distributional result, i.e., that the asymptotic distribution of the χ^2 statistic is noncentral χ^2 , with noncentrality parameter given by

$$(n - 1)F_{ML}^* \quad (2)$$

in the standard case in which $k = n - 1$ is used as the multiplier.

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing the Model for Perfect Fit

- The noncentral χ^2 approximation can furnish accurate estimates of the performance of the likelihood ratio test statistic when the null hypothesis is false.
- One employs the standard approach of first obtaining a rejection point from the central chi-square distribution, then computing the probability in the rejection region under the alternative hypothesis.

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing the Model for Perfect Fit

- Suppose you have a structural equation model based on $p = 7$ variables with $t = 6$ free parameters. The population discrepancy function is 0.055. What is the power of the test of perfect fit if $\alpha = 0.05$, 1-sided and $n = 185$?
- Using R, we first establish the critical value (rejection point).

```
> p <- 7
> t <- 6
> F.ML <- 0.055
> df <- p*(p+1)/2
> df
[1] 28
> critical.value <- qchisq(0.95,df)
> critical.value
[1] 41.33714
```


Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing the Model for Perfect Fit

- Now we compute the power.
- First we calculate the noncentrality parameter.

```
> n <- 185
> lambda = (n-1)*F.ML
> lambda

[1] 10.12
```
- Next we compute the power.

```
> power <- 1 - pchisq(critical.value,df,lambda)
> power

[1] 0.3443796
```
- We find that the power is 0.344.

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing the Model for Perfect Fit

- Note how straightforward power calculation is. However, a key point was glossed over: How does one estimate the population discrepancy function?
- MacCallum, Browne, and Sugawara (1996) suggested using the RMSEA as a vehicle for estimating an alternative value of the discrepancy function. Recall that, in the population, the RMSEA is defined as

$$\epsilon = \sqrt{\frac{F_{ML}^*}{df}} \quad (3)$$

- Consequently,

$$F_{ML}^* = \epsilon^2 \times df \quad (4)$$

- If one can specify a “non-trivial” RMSEA cut-off, one can use that to compute a discrepancy function, and from that discrepancy function compute the noncentrality parameter, followed by power.

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing the Model for Perfect Fit

- Suppose, for example, the minimal RMSEA is 0.08. You want to have a guaranteed level of power to detect an RMSEA that small.
- In the present example, this RMSEA would translate into a noncentrality parameter as follows.

```
> epsilon <- 0.08
> F.ML <- epsilon^2 * df
> F.ML
[1] 0.1792
> lambda <- (n-1)*F.ML
> power <- 1 - pchisq(critical.value,df,lambda)
> power
[1] 0.9366569
```

- We find that power to reject the hypothesis of perfect fit is 0.937 if the RMSEA is 0.08 in the population.

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing for Close Fit

- In some cases, one may not wish to test that fit is perfect.
- For example, MacCallum, Browne, and Sugawara(1996) discussed two other kinds of tests, which they referred to as the **test of close fit** and the **test of not-close fit**, respectively.

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing for Close Fit

- The standard χ^2 test of perfect fit is also a test of the hypothesis that the RMSEA is equal to zero in the population.
- The test of close fit assumes as its statistical null hypothesis not that the population RMSEA is zero, but rather that it is less than or equal to some other reasonable value.
- For example, suppose the null hypothesis is

$$H_0 : \epsilon \leq 0.05 \quad (5)$$

- The alternative is

$$H_0 : \epsilon > 0.05 \quad (6)$$

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing for Close Fit

- This is a 1-sided test.
- The rejection region begins at a critical value at the $1 - \alpha$ quantile of the noncentral χ^2 distribution.
- Here is a sample problem. Suppose the test of close fit is performed as before with $n = 185$, $p = 7$, and $t = 6$, and $\alpha = 0.05$.
- What is the power to reject H_0 if the true population RMSEA is 0.08?

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing for Close Fit

- First we calculate the null distribution and rejection point.
- Although the null hypothesis specifies a *region* rather than a point, we can make our point of calculation the boundary of that region, at $\epsilon = 0.05$. Then power will be *at least as large* as the calculated value.

```
> epsilon0 <- 0.05; epsilon1 <- 0.08; df <- p * (p+1) /2 - t
> alpha <- 0.05; n<-185
> F0 <- epsilon0^2 *df
> lambda0 <- (n-1) * F0
> critical.value <- qchisq(1-alpha, df, lambda0)
> F1 <- epsilon1^2 * df
> lambda1 <- (n-1) * F1
> power <- 1 - pchisq(critical.value,df,lambda1)
> power
[1] 0.4504781
```

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing for Not-Close Fit

- The original χ^2 test of perfect fit suffered from several problems:
 - 1 It was unrealistic. The assumption of perfect fit is extremely unlikely to be correct.
 - 2 It is an Accept-Support test. Supporting a proposed model requires not rejecting the null hypothesis. Accept-Support testing suffers from a number of problems. In particular, it rewards low power and sloppy experimentation.

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing for Not-Close Fit

- The test of close fit improves on the test of perfect fit, in that it tests a more reasonable null hypothesis.
- However, it is still an Accept-Support procedure, in that it requires not rejecting the null hypothesis to support a model.

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing for Not-Close Fit

- The test of not-close fit takes a different approach.
- The null hypothesis is now stated in the form

$$H_0 : \epsilon \geq \epsilon_0 \quad (7)$$

where ϵ_0 is a value that represents “reasonably good fit.”

- Since the alternative hypothesis is

$$H_1 : \epsilon < \epsilon_0 \quad (8)$$

rejection of the null implies that “fit is significantly better than reasonably good.”

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing for Not-Close Fit

- The test of not-close fit eliminates the problems associated with Accept-Support testing.
- This test is easy to perform, and to calculate power for.
- Suppose we wish to test the null hypothesis that $\epsilon \geq 0.05$, and the true state of the world is that $\epsilon = 0.01$.
- In that case, the power will be as calculated below. Note that now the rejection region is on the low side of the null hypothesized distribution.

```
> epsilon0 <- 0.05; epsilon1 <- 0.01
> F0 <- df * epsilon0^2
> lambda0 <- (n-1)*F0
> critical.value <- qchisq(0.05,df,lambda0)
> F1 <- df * epsilon1^2
> lambda1 <- (n-1)*F1
> power <- pchisq(critical.value,df,lambda1)
> power
```

```
[1] 0.3062439
```

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing Individual Parameters

- So far, we have been discussing tests on an entire model.
- Such tests can be useful, and power analysis for such tests is straightforward.
- Another kind of test performed frequently in the context of structural equation modeling is the test of significance on a parameter.
- Two types of tests can be performed:
 - ① The χ^2 difference test.
 - ② The Wald test.
- In the χ^2 difference test, two models are tested, with and without the path involving the parameter of interest.
- In the Wald test, the model is fit with the parameter of interest included, and an asymptotically normal test statistic is computed by dividing the parameter by its (estimated) standard error.

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing Individual Parameters

- A model without a path containing a parameter is nested within a model containing the parameter.
- Consequently, a result of Steiger, Shapiro, and Browne (1985) can be applied.
- Under fairly general conditions, the difference between the chi-square tests for two nested models is distributed as noncentral χ^2 with degrees of freedom equal to the difference in degrees of freedom for the two models (i.e., in this case 1), and a noncentrality parameter equal to the difference in noncentrality parameters for the two models.
- Consequently, one may estimate power to reject the null hypothesis for a single parameter by a two step process:
 - 1 Create the population covariance matrix corresponding to the model with the path containing the parameter.
 - 2 Fit this matrix to the model without the path containing the parameter, thereby obtaining a non-zero population discrepancy function, and through it, a noncentrality parameter.
 - 3 Use the non-central χ^2 approximation to compute the power.

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing Individual Parameters

- Here is an example. Suppose that the following factor pattern fit the population covariance matrix perfectly. Note there is one unanticipated “crossover loading” marked in red.

$$\Lambda = \begin{bmatrix} 0.7 & 0.0 & 0.0 \\ 0.7 & 0.0 & 0.0 \\ 0.7 & 0.0 & 0.0 \\ 0.0 & 0.7 & 0.0 \\ 0.0 & 0.7 & 0.0 \\ 0.0 & 0.7 & 0.0 \\ 0.0 & 0.2 & 0.7 \\ 0.0 & 0.0 & 0.7 \\ 0.0 & 0.0 & 0.7 \end{bmatrix} \quad (9)$$

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing Individual Parameters

- We generate the covariance matrix corresponding to Λ . First we construct Λ in R.

```
> Lambda <- matrix(c(rep(.7,3),rep(0,9),rep(0.7,3)),  
+                   0.2,rep(0,8),rep(.7,3)),9,3)  
> Lambda
```

```
      [,1] [,2] [,3]  
[1,] 0.7  0.0  0.0  
[2,] 0.7  0.0  0.0  
[3,] 0.7  0.0  0.0  
[4,] 0.0  0.7  0.0  
[5,] 0.0  0.7  0.0  
[6,] 0.0  0.7  0.0  
[7,] 0.0  0.2  0.7  
[8,] 0.0  0.0  0.7  
[9,] 0.0  0.0  0.7
```

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing Individual Parameters

- Next, we apply one of the utility functions.

```
> MakeFactorCorrelationMatrix <- function(F){
+ r <-F %*% t(F)
+ h <- diag(r)
+ u <- diag(1 - h)
+ r <- r+u
+ r
+ }
> R <- MakeFactorCorrelationMatrix(Lambda)
> R
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	1.00	0.49	0.49	0.00	0.00	0.00	0.00	0.00	0.00
[2,]	0.49	1.00	0.49	0.00	0.00	0.00	0.00	0.00	0.00
[3,]	0.49	0.49	1.00	0.00	0.00	0.00	0.00	0.00	0.00
[4,]	0.00	0.00	0.00	1.00	0.49	0.49	0.14	0.00	0.00
[5,]	0.00	0.00	0.00	0.49	1.00	0.49	0.14	0.00	0.00
[6,]	0.00	0.00	0.00	0.49	0.49	1.00	0.14	0.00	0.00
[7,]	0.00	0.00	0.00	0.14	0.14	0.14	1.00	0.49	0.49
[8,]	0.00	0.00	0.00	0.00	0.00	0.00	0.49	1.00	0.49
[9,]	0.00	0.00	0.00	0.00	0.00	0.00	0.49	0.49	1.00

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing Individual Parameters

- We can save the correlation matrix to a text file for analysis by Mplus.

```
> Rout <- data.frame(R)
> write.table(Rout,"PowerCov.dat",row.names=FALSE,
+             col.names=FALSE,sep=" ")
```

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing Individual Parameters

- Next we analyze the data with Mplus, leaving out the path corresponding to the crossover loading.

```
TITLE: SETUP ANALYSIS FOR POWER CALCULATION;  
DATA: FILE IS PowerCov.dat;  
      TYPE IS FULLCOV;  
      NOBSERVATIONS = 1000000;  
VARIABLE: NAMES ARE Y1-Y9;  
MODEL: F1 BY Y1-Y3*;  
       F2 BY Y4-Y6*;  
       F3 BY Y7-Y9*;  
       F1-F3@1;
```

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing Individual Parameters

- This produces the following test statistic output:

Chi-Square Test of Model Fit

Value	39570.991
Degrees of Freedom	24
P-Value	0.0000

- With a “pretend” sample size of 10^6 , the discrepancy function calculates to about 0.03957.

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing Individual Parameters

- Suppose sample size was actually 250 with this population matrix.
- Then the population noncentrality parameter would be

```
> F <- 0.03957
> Lambda = (250-1) * F
> Lambda
[1] 9.85293
> epsilon = sqrt(F/24)
> epsilon
[1] 0.0406048
```
- The difference test has one degree of freedom, and we estimate the power as

```
> critical.value <- qchisq(0.95,1)
> power <- 1 - pchisq(critical.value, 1, Lambda)
> power
[1] 0.8807959
```
- Power is about 0.88.

Hypothesis Testing and Fit Evaluation: What, Where, How, and Why

Testing Individual Parameters

- On the other hand, the test of perfect model fit has 24 degrees of freedom, so the power to reject perfect model fit is lower.

```
> critical.value <- qchisq(0.95,24)
> power <- 1 - pchisq(critical.value, 24, Lambda)
> power
[1] 0.3614827
```

- Power for this test is estimated to be only 0.361

Power Charts for Tests of Overall Model Fit

- Constructing a power chart, displaying power as a function of sample size or some other quantity, is really straightforward in R.
- Let's begin by constructing a **power function**, that returns estimated power as a function of α , ϵ_0 , ϵ_1 , df , and n .

Power Charts for Tests of Overall Model Fit

```

> ## Calculate power as a function of
> ##   n           Sample Size
> ##   epsilon1    The true population RMSEA
> ##   epsilon0    The null RMSEA (0 for a test of perfect fit)
> ##   alpha       Type I Error Rate
> ##   tail        "lower" or "upper" (Where the rejection region is)
> SEMpower <- function(n,df,epsilon1,epsilon0=0,alpha=0.05,tail="upper"){
+   lambda0 <- (n-1)*df*epsilon0^2
+   lambda1 <- (n-1)*df*epsilon1^2
+   critical.p <- if(tail == "upper") 1-alpha else alpha
+   critical.value <- qchisq(critical.p,df,lambda0)
+   critical.prob <- pchisq(critical.value,df,lambda1)
+   power <- if (tail == "upper") 1-critical.prob else critical.prob
+   return(power)
+ }

```

Power Charts for Tests of Overall Model Fit

- Let's try out our function on the example we just worked.
- In that example, we had a population discrepancy function of $F = 0.03957$, which, combined with 24 degrees of freedom, translates into a population RMSEA of

$$\epsilon = \sqrt{\frac{0.03957}{24}} = 0.0406048 \quad (10)$$

- For a test of perfect fit, $\epsilon_0 = 0$.
- We can calculate power (using the default upper tail and $\alpha = 0.05$) as

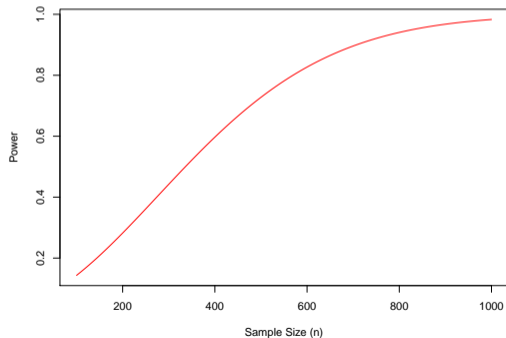
```
> SEMpower(250,24,sqrt(0.03957/24))
```

```
[1] 0.3614827
```


Power Charts for Tests of Overall Model Fit

- We can now display a **power chart** showing power as a function of n for this situation.

```
> curve(SEMpower(x,24,0.0406),100,1000,col="red",  
+       xlab="Sample Size (n)",  
+       ylab="Power")
```



Power Charts for Tests of Overall Model Fit

- The SEM_{power} function can be adapted to handle situations in which a test of close-fit, or a test of not-close fit is being performed.
- It can also be used to estimate the sample size required to achieve a guaranteed level of power under specified conditions, as shown in the next section.

Sample Size Calculations for Tests of Model Fit

- In the just-completed analysis, we found of power of 0.36, certainly not adequate.
- The question immediately arises: What would sample size n have to be to yield an acceptable level of power?
- In this situation, there is no direct analytic solution to the required sample size. The solution can be approximated reasonably well, or calculated precisely by iterative methods.
- The **method of bisection** works quite well in this case, and is extraordinarily simple.
- The calculation is hampered slightly by the fact that, in the final analysis, n must be an integer.

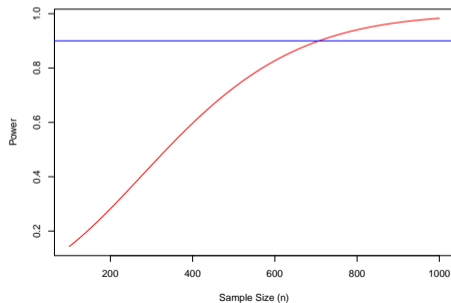
Sample Size Calculations for Tests of Model Fit

- A function calculating required sample size is a good one to have, and we'll develop one eventually.
- However, for one specific situation, one doesn't need to go to this trouble.
- Instead, we use a simple graphical approach.
- Remember this approach well, as you may find it useful in the future. It is, in general, much easier to construct a power calculation function than it is to construct a sample size calculating function!

Sample Size Calculations for Tests of Model Fit

- What sample size would be required to guarantee a power of 0.90? Looking at the previously generated graph, we can begin by adding a horizontal line at 0.90.
- Simple inspection tells us that the required n is between 650 and 750.

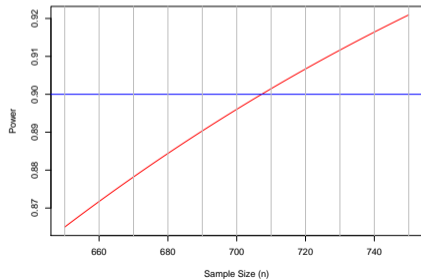
```
> curve(SEMpower(x,24,0.0406),100,1000,col="red",
+       xlab="Sample Size (n)",
+       ylab="Power")
> abline(h = 0.90, col="blue")
```



Sample Size Calculations for Tests of Model Fit

- We redraw the graph with narrower limits of 650 and 750.
- We add gridlines in grey to help us.
- We can see that the required n is clearly between 705 and 710, and is almost certainly either 707 or 708.

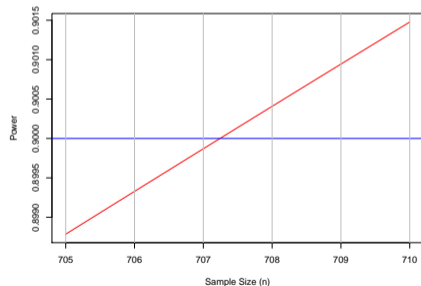
```
> curve(SEMpower(x,24,0.0406),650,750,col="red",
+       xlab="Sample Size (n)",
+       ylab="Power")
> abline(h = 0.90, col="blue")
> abline(v = seq(650,750,10), col="grey")
```



Sample Size Calculations for Tests of Model Fit

- We home in on one more time and nail down the correct answer.

```
> curve(SEMpower(x,24,0.0406),705,710,col="red",
+       xlab="Sample Size (n)",
+       ylab="Power")
> abline(h = 0.90, col="blue")
> abline(v = 705:710, col="grey")
```



- Note that the correct answer is not an integer! The convention is to pick the next highest integer, which is the lowest integer n for which power is greater than or equal to 0.90. We choose $n = 708$.

Sample Size Calculations for Tests of Model Fit

Sample Size Tables

- Using an iterative algorithm, it is easy to construct tables or charts of sample sizes required to achieve a given level of power.
- Here are some examples from MacCallum, Browne, and Sugawara (1996).
- What are some “take home” messages from the tables? (C.P.)

Sample Size Calculations for Tests of Model Fit

Sample Size Tables

Table 5

Minimum Sample Sizes for Test of Exact Fit for Selected Levels of Degrees of Freedom (df) and Power

<i>df</i>	Minimum <i>N</i> for power = 0.80	Minimum <i>N</i> for power = 0.50
2	1,926	994
4	1,194	644
6	910	502
8	754	422
10	651	369
12	579	332
14	525	304
16	483	280
18	449	262
20	421	247
25	368	218
30	329	196
35	300	180
40	277	167
45	258	157
50	243	148
55	230	140
60	218	134
65	209	128
70	200	123
75	193	119
80	186	115
85	179	111
90	174	108
95	168	105
100	164	102

Note. The $\alpha = .05$, $\epsilon_0 = 0.0$, and $\epsilon_a = 0.05$, where ϵ_0 is the null value of the root-mean-square error of approximation (RMSEA) and ϵ_a is the alternative value of RMSEA.

Sample Size Calculations for Tests of Model Fit

Sample Size Tables

Table 4

Minimum Sample Size to Achieve Power of 0.80 for Selected Levels of Degrees of Freedom (df)

df	Minimum N for test of close fit	Minimum N for test of not-close fit
2	3,488	2,382
4	1,807	1,426
6	1,238	1,069
8	954	875
10	782	750
12	666	663
14	585	598
16	522	547
18	472	508
20	435	474
25	363	411
30	314	366
35	279	333
40	252	307
45	231	286
50	214	268
55	200	253
60	187	240
65	177	229
70	168	219
75	161	210
80	154	202
85	147	195
90	142	189
95	136	183
100	132	178

Note. For all analyses, $\alpha = .05$. For the test of close fit, $\epsilon_0 = 0.05$ and $\epsilon_1 = 0.08$, where ϵ_0 is the null value of the root-mean-square error of approximation (RMSEA) and ϵ_1 is the alternative value of RMSEA. For the test of not-close fit, $\epsilon_0 = 0.05$ and $\epsilon_1 = 0.01$.

Interval Estimation Approaches

- An alternative to a hypothesis-testing framework is based on confidence interval estimation.
- One can perform all the standard hypothesis tests discussed above simply by constructing a confidence interval on the RMSEA.
- Since the tests are one-sided, one constructs a $1 - 2\alpha$ confidence interval and examines it to see if the confidence interval excludes the critical value of the RMSEA.

Interval Estimation Approaches

- Constructing a confidence interval on the RMSEA requires an iterative procedure.
- The method is discussed in detail in the article by Steiger and Fouladi (1997) available on my website in the Publications section.
- Briefly — One solves inversely for those values of λ that would place the observed data at the 95th and 5th percentile of the non-central chi-square distribution. These values are the endpoints of a 90% confidence interval for λ , which may be converted into a confidence interval for the RMSEA ϵ .

Interval Estimation Approaches

- Suppose, for example, one calculates the 90% CI for ϵ , and it has endpoints 0 and 0.0491.
- Could one reject the hypothesis of perfect fit? How about the hypothesis of close fit, with $H_0 : \epsilon \leq 0.05$?
- Could one reject the hypothesis of not-close fit, if H_0 is

$$H_0 : \epsilon \geq 0.05 \quad (11)$$

Interval Estimation Approaches

- Suppose the confidence interval on the RMSEA has endpoints of 0.0471 and 0.0752.
- Could one reject the hypothesis of perfect fit?
- Could one reject the hypothesis of close fit?
- Could one reject the hypothesis of not-close fit?

Interval Estimation Approaches

Changing the Emphasis – AIPE

- In the preceding section, we examined several types of hypothesis tests that can be performed simultaneously with a confidence interval.
- Only one of the hypothesis tests — the test of not-close fit — involves a Reject-Support strategy.
- The original sample size table from MacCallum, Browne, and Sugawara (1996) involves a null hypothesis that $\epsilon \geq 0.05$. Power is calculated under the very stringent alternative hypothesis that $\epsilon = 0.01$.
- Clearly, to get a rejection using this strategy requires a narrow confidence interval, because the interval has to fit between 0 and 0.05 in order for a rejection to occur!
- If your confidence interval is wider than 0.05, this can never happen!

Interval Estimation Approaches

Changing the Emphasis – AIPE

- As sample size increases, the average width of a confidence interval for ϵ generally decreases.
- The expected value of the width of a confidence interval is a complex function of the sample size and the RMSEA itself.
- However, one can estimate the sample size required to produce a “narrow-enough” confidence interval for the RMSEA under a reasonable set of assumptions.
- This approach of calculating sample size in terms of the expected width of a confidence interval has been called AIPE (Accuracy in Parameter Estimation) by Ken Kelley of Notre Dame, who has a number of interesting articles on the topic.
- Some of these techniques are implemented in his MBESS package for R. MBESS also implements confidence interval calculation for the RMSEA.

Interval Estimation Approaches

Changing the Emphasis – AIPE

- Here is an example (the package is currently broken)
 - > `##library(MBESS)`
 - > `##ss.aipe.rmsea(0.03,24,0.08,0.9)`