Confirmatory Factor Analysis and Structural Equation Modeling Group Differences: Measurement Invariance.

As published in Benchmarks RSS Matters, March 2015

http://web3.unt.edu/benchmarks/issues/2015/03/rss-matters

Jon Starkweather, PhD

Jon Starkweather, PhD
jonathan.starkweather@unt.edu
Consultant
Research and Statistical Support

http://www.unt.edu

http://www.unt.edu/rss

RSS hosts a number of "Short Courses".
A list of them is available at:
http://www.unt.edu/rss/Instructional.htm

Those interested in learning more about R, or how to use it, can find information here:
http://www.unt.edu/rss/class/Jon/R_SC

Confirmatory Factor Analysis and Structural Equation Modeling Group Differences: Measurement Invariance.

This month's article focuses on an explanation of measurement invariance. This article is specifically oriented toward the context of detecting group differences among latent variables for confirmatory factor analysis (CFA) models or in a structural equation models (SEM). Social scientists are often concerned with identifying group differences (e.g. differences between genders, ethnicities, locations, etc.). SEM is often applied in an effort to model the complex relationships of latent variables between groups for CFA-type models. Therefore, it is likely that many social scientists would find this article useful as a means to evaluate group differences among complex latent variable model structures. Attempting to evaluate or discover group differences among latent variables is necessarily complex due to the underlying factor models which support the latent models (i.e. SEM). So, it is necessary to recognize such complexity and evaluate the sequentially imposed constraints on the group differences – which implicitly leads to a discussion of *measurement invariance*. An excellent reference for this material is a relatively new book by Beaujean (2014), particularly chapter 4.

Measurement invariance is not a single unified concept; although generally we can define measurement invariance as stable measurement parameters across multiple groups, settings, and time periods. Commonly, the parameters referred to in the previous sentence refer to the factor structure (i.e. specific observed variables to latent variables, etc.), factor loadings, intercepts, and the latent variable means of a measurement model (i.e. factor model). Typically, there are a series of sequentially imposed measurement constraints, ranked as level 1 (configural invariance), level 2 (weak invariance), level 3 (strong invariance), and level 4 (strict invariance). Configural invariance refers to the *configuration* or structure of the factor model (i.e. which observed variables go with which latent factors). Weak invariance refers to factor loadings (and configuration) being the same between two groups, settings, or time periods. Strong invariance refers to the intercepts (configuration, and loadings) of the factor model and strict invariance refers to the latent variable means (configuration, loadings, and intercepts) being the same between two groups, settings, or time periods.

Testing for measurement invariance consists of a series of statistical hypotheses that assume population group factor parameters are equal between the groups. Fortunately, there is (of course) a function in R for testing measurement invariance in CFA and SEM models. The package 'semTools' (Pornprasertmanit, et al., 2015) contains the function 'measurementInvariance' which will be demonstrated below. The 'measurementInvariance' function takes a 'lavaan' package (Rosseel, et al., 2015) model object and raw data and tests the fit of the object while checking for chi-square (and fit indices) differences between two (or more) groups.

1 The Examples

First, we import some (simulated) data. Keep in mind, the data is available for readers to duplicate what is done in this article by using the script shown in the article (script also available here[1]; data available here[2]). The data includes two groups ($n_1 = 500$ & $n_2 = 502$) with ($N_i = 1002$) responses on ($j = 24$)

[1] `http://www.unt.edu/rss/class/Jon/Benchmarks/BenchmarksFeb2015.R`
[2] `http://www.unt.edu/rss/class/Jon/ExampleData/measInvar_df.txt`

variables ($x1, x2, x3, ...x24$).

```
df.1 <- read.table(
       "http://www.unt.edu/rss/class/Jon/ExampleData/measInvar_df.txt",
        header = TRUE, sep = ",", na.strings = "NA", dec = ".")
summary(df.1)
     group               x1                   x2
 Min.   :1.000   Min.   :-3.703924   Min.   :-4.24310
 1st Qu.:1.000   1st Qu.:-0.843175   1st Qu.:-0.88901
 Median :2.000   Median : 0.076019   Median :-0.06182
 Mean   :1.501   Mean   : 0.001051   Mean   :-0.05534
 3rd Qu.:2.000   3rd Qu.: 0.853784   3rd Qu.: 0.80925
 Max.   :2.000   Max.   : 3.579749   Max.   : 3.77787
       x3                  x4                  x5
 Min.   :-4.015567   Min.   :-3.88353   Min.   :-3.86466
 1st Qu.:-0.886522   1st Qu.:-0.89705   1st Qu.:-0.85205
 Median : 0.046421   Median :-0.07672   Median :-0.02942
 Mean   : 0.004654   Mean   :-0.05154   Mean   :-0.02075
 3rd Qu.: 0.876326   3rd Qu.: 0.82199   3rd Qu.: 0.84022
 Max.   : 3.503825   Max.   : 3.60557   Max.   : 2.94853
       x6                  x7                   x8
 Min.   :-4.82883   Min.   :-3.415288   Min.   :-3.56686
 1st Qu.:-0.86454   1st Qu.:-0.847181   1st Qu.:-0.80358
 Median : 0.01619   Median : 0.042244   Median : 0.03872
 Mean   : 0.02247   Mean   : 0.005208   Mean   : 0.05161
 3rd Qu.: 0.90802   3rd Qu.: 0.853102   3rd Qu.: 0.89892
 Max.   : 4.06204   Max.   : 3.199517   Max.   : 4.16097
       x9               x10               x11               x12
 Min.   : 6.656   Min.   : 6.187   Min.   : 6.298   Min.   : 6.081
 1st Qu.: 9.251   1st Qu.: 9.261   1st Qu.: 9.213   1st Qu.: 9.257
 Median :10.085   Median :10.058   Median :10.041   Median :10.107
 Mean   :10.057   Mean   :10.038   Mean   :10.041   Mean   :10.059
 3rd Qu.:10.834   3rd Qu.:10.873   3rd Qu.:10.850   3rd Qu.:10.831
 Max.   :13.628   Max.   :13.615   Max.   :13.949   Max.   :13.481
      x13               x14               x15               x16
 Min.   : 6.077   Min.   : 6.471   Min.   : 6.450   Min.   : 6.463
 1st Qu.: 9.202   1st Qu.: 9.210   1st Qu.: 9.171   1st Qu.: 9.223
 Median :10.010   Median :10.049   Median :10.022   Median : 9.990
 Mean   :10.004   Mean   :10.008   Mean   : 9.979   Mean   : 9.991
 3rd Qu.:10.796   3rd Qu.:10.795   3rd Qu.:10.808   3rd Qu.:10.785
 Max.   :13.692   Max.   :13.386   Max.   :13.386   Max.   :14.251
      x17               x18               x19               x20
 Min.   : 6.154   Min.   : 6.854   Min.   : 6.687   Min.   : 5.959
 1st Qu.: 9.190   1st Qu.: 9.233   1st Qu.: 9.227   1st Qu.: 9.190
 Median :10.020   Median :10.033   Median : 9.988   Median : 9.945
 Mean   : 9.999   Mean   :10.019   Mean   :10.002   Mean   : 9.957
 3rd Qu.:10.729   3rd Qu.:10.795   3rd Qu.:10.784   3rd Qu.:10.741
```

```
 Max.   :13.122   Max.   :13.044   Max.   :13.510   Max.   :12.746  
      x21              x22              x23              x24        
 Min.   : 6.657   Min.   : 6.466   Min.   : 6.111   Min.   : 6.468  
 1st Qu.: 9.309   1st Qu.: 9.250   1st Qu.: 9.281   1st Qu.: 9.318  
 Median :10.036   Median :10.022   Median : 9.984   Median :10.040  
 Mean   :10.025   Mean   :10.002   Mean   :10.003   Mean   :10.050  
 3rd Qu.:10.742   3rd Qu.:10.735   3rd Qu.:10.760   3rd Qu.:10.787  
 Max.   :13.497   Max.   :13.164   Max.   :12.962   Max.   :13.449  
```

Upon initial inspection, the two groups appear to be virtually identical in terms of how the factor model fits each group's data.

```
factanal(df.1[1:500, 2:9], factors = 2)    # Group 1.

Call:
factanal(x = df.1[1:500, 2:9], factors = 2)

Uniquenesses:
   x1    x2    x3    x4    x5    x6    x7    x8 
0.338 0.401 0.323 0.348 0.507 0.485 0.556 0.572 

Loadings:
   Factor1 Factor2
x1  0.812         
x2  0.774         
x3  0.823         
x4  0.807         
x5          0.702 
x6          0.716 
x7          0.666 
x8          0.654 

               Factor1 Factor2
SS loadings      2.588   1.882
Proportion Var   0.323   0.235
Cumulative Var   0.323   0.559

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 21.21 on 13 degrees of freedom.
The p-value is 0.0689 

factanal(df.1[501:1002,2:9], factors = 2)  # Group 2.

Call:
factanal(x = df.1[501:1002, 2:9], factors = 2)

Uniquenesses:
```

```
   x1    x2    x3    x4    x5    x6    x7    x8
0.371 0.359 0.363 0.317 0.519 0.515 0.541 0.498

Loadings:
   Factor1 Factor2
x1  0.793
x2  0.801
x3  0.798
x4  0.826
x5          0.691
x6          0.696
x7          0.677
x8          0.708

               Factor1 Factor2
SS loadings      2.594   1.923
Proportion Var   0.324   0.240
Cumulative Var   0.324   0.565

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 16.2 on 13 degrees of freedom.
The p-value is 0.238
```

Next, we load the ‘lavaan’ and ‘semTools’ packages in order to specify the CFA model and test for the levels of measurement invariance formally.

```
library(lavaan)
This is lavaan 0.5-17
lavaan is BETA software! Please report any bugs.
library(semTools)

###############################################################################
This is semTools 0.4-6
All users of R (or SEM) are invited to submit functions or ideas for functions.
###############################################################################
cfa.model <- '
  f1 =~ x1 + x2 + x3 + x4
  f2 =~ x5 + x6 + x7 + x8
  f1 ~~ 0*f2
  '
measurementInvariance(cfa.model, data = df.1, group = "group")

Measurement invariance tests:

Model 1: configural invariance:
    chisq        df    pvalue       cfi     rmsea       bic
   48.209    40.000     0.175     0.997     0.020 19980.029
```

```
Model 2: weak invariance (equal loadings):
    chisq        df    pvalue       cfi     rmsea       bic
   51.489    46.000     0.268     0.998     0.015 19941.851

[Model 1 versus model 2]
  delta.chisq      delta.df delta.p.value     delta.cfi
        3.280         6.000         0.773        -0.001

Model 3: strong invariance (equal loadings + intercepts):
    chisq        df    pvalue       cfi     rmsea       bic
   56.353    52.000     0.315     0.999     0.013 19905.257

[Model 1 versus model 3]
  delta.chisq      delta.df delta.p.value     delta.cfi
        8.145        12.000         0.774        -0.001

[Model 2 versus model 3]
  delta.chisq      delta.df delta.p.value     delta.cfi
        4.864         6.000         0.561         0.000

Model 4: equal loadings + intercepts + means:
    chisq        df    pvalue       cfi     rmsea       bic
 1222.336    54.000     0.000     0.622     0.208 21057.420

[Model 1 versus model 4]
  delta.chisq      delta.df delta.p.value     delta.cfi
     1174.127        14.000         0.000         0.375

[Model 3 versus model 4]
  delta.chisq      delta.df delta.p.value     delta.cfi
     1165.983         2.000         0.000         0.376
```

Evaluating the output of the ‘measurementInvariance’ function necessarily starts with configual invariance (model 1) which assumes the factor pattern is equal for both groups. Next, the second hypothesis is evaluated; weak invariance (model 2) which evaluates the chi-square change (or delta: Δ) and associated *p*-value; as well as the change in the Comparative Fit Index (CFI). The output for the comparison between model 1 and model 2 indicates no statistically significant change in the chi-square value, and the CFI does not change very much either - which indicates the loadings of the two groups are *close enough*. When the loadings are essentially the same, then weak measurement invariance is supported. The next hypothesis, strong invariance (model 3), is then evaluated. Model 3 involves testing the hypothesis that the loadings *and intercepts* are the same, or statistically equivalent, for both groups. The output shows that the first comparison, model 1 to model 3, is not statistically significant ($p = 0.774$); meaning the chi-square value is not significantly different between those two models. The second comparison, model 2 to model 3, also is not statistically significant ($p = 0.561$). In other words, when the loadings and intercepts are constrained to be equal, the model fit is not significantly different than the actual model fit across the two groups. Therefore, strong measurement invariance is supported. However, when we evaluate the final hypothesis of measurement invariance, strict invariance (model 4), we find that the latent variable *means* appear to be different – based on the chi-square change; indicating a significant difference between the groups' fit. There are several pieces of output which show this difference. First, numerically / visually compare the chi-square values for model 3 ($\chi^2 = 56.353, df = 52, p = 0.315$) and model 4 ($\chi^2 = 1222.336, df = 54, p < 0.000$); which is a substantial change in chi-square. Also, notice how much the CFI changed from model 3 ($cfi = 0.999$) to model 4 ($cfi = 0.622$); while model 2 ($cfi = 0.998$) and model 1 ($cfi = 0.997$) are both very close to model 3. These differences (in chi-square & CFI) are also revealed in the two model comparisons. Comparing the change in fit between model 1 and model 4, we observe a significant chi-square change ($\chi^2_\Delta = 1174.127, df_\Delta = 14, p_\Delta < 0.000$). Furthermore, comparing the change in fit between model 3 and model 4, we observe another significant chi-square change ($\chi^2_\Delta = 1165.983, df_\Delta = 2, p_\Delta < 0.000$). The appropriate conclusion is; we do not have strict measurement invariance.

The utility of the 'measuremenInvariance' function extends beyond straightforward CFA and it can be applied to SEM settings as well. For instance, following the Anderson and Gerbing (1988) two stage approach to SEM, we can specify the measurement model of a SEM and use the 'measurementInvariance' function to check the levels (or models) of measurement invariance.

```
cfa.model <- '
  f1 =~ x1 + x2 + x3 + x4
  f2 =~ x5 + x6 + x7 + x8
  f3 =~ x9 + x10 + x11 + x12 + x13 + x14 + x15
  f4 =~ x16 + x17 + x18 + x19 + x20
  f5 =~ x21 + x22 + x23 + x24
  f1 ~~ 0*f2
  f1 ~~ f3
  f1 ~~ f4
  f1 ~~ f5
  f2 ~~ f3
  f2 ~~ f4
  f2 ~~ f5
  f3 ~~ f4
  f3 ~~ f5
  f4 ~~ f5
  '
measurementInvariance(cfa.model, data = df.1, group = "group")

Measurement invariance tests:

Model 1: configural invariance:
    chisq        df    pvalue       cfi     rmsea       bic
  492.800   486.000     0.406     0.999     0.005 61241.402

Model 2: weak invariance (equal loadings):
    chisq        df    pvalue       cfi     rmsea       bic
  508.302   505.000     0.450     1.000     0.004 61125.619
```

```
[Model 1 versus model 2]
  delta.chisq      delta.df delta.p.value     delta.cfi
       15.502        19.000         0.690         0.000

Model 3: strong invariance (equal loadings + intercepts):
    chisq        df     pvalue        cfi     rmsea        bic
  528.129   524.000      0.441     1.000     0.004 61014.161

[Model 1 versus model 3]
  delta.chisq      delta.df delta.p.value     delta.cfi
       35.329        38.000         0.594         0.000

[Model 2 versus model 3]
  delta.chisq      delta.df delta.p.value     delta.cfi
       19.827        19.000         0.405         0.000

Model 4: equal loadings + intercepts + means:
    chisq        df     pvalue        cfi     rmsea        bic
 1732.314   529.000      0.000     0.855     0.067 62183.796

[Model 1 versus model 4]
  delta.chisq      delta.df delta.p.value     delta.cfi
     1239.513        43.000         0.000         0.145

[Model 3 versus model 4]
  delta.chisq      delta.df delta.p.value     delta.cfi
     1204.184         5.000         0.000         0.145
```

It is also possible to specify a structural model of a SEM and check for measurement invariance; as show below.

```
str.model <- '
  f1 =~ x1 + x2 + x3 + x4
  f2 =~ x5 + x6 + x7 + x8
  f3 =~ x9 + x10 + x11 + x12 + x13 + x14 + x15
  f4 =~ x16 + x17 + x18 + x19 + x20
  f5 =~ x21 + x22 + x23 + x24
  f4 ~ f1
  f3 ~ f2
  f5 ~ f2 + f3
  f1 ~~ 0*f2
  f1 ~~ f3
  f1 ~~ f5
  f2 ~~ f4
  f3 ~~ f4
  f4 ~~ f5
  '
```

```
measurementInvariance(str.model, data = df.1, group = "group")

Measurement invariance tests:

Model 1: configural invariance:
    chisq        df    pvalue       cfi     rmsea       bic
  492.800   486.000     0.406     0.999     0.005 61241.402

Model 2: weak invariance (equal loadings):
    chisq        df    pvalue       cfi     rmsea       bic
  508.302   505.000     0.450     1.000     0.004 61125.619

[Model 1 versus model 2]
  delta.chisq      delta.df delta.p.value     delta.cfi
       15.502        19.000         0.690         0.000

Model 3: strong invariance (equal loadings + intercepts):
    chisq        df    pvalue       cfi     rmsea       bic
  528.129   524.000     0.441     1.000     0.004 61014.161

[Model 1 versus model 3]
  delta.chisq      delta.df delta.p.value     delta.cfi
       35.329        38.000         0.594         0.000

[Model 2 versus model 3]
  delta.chisq      delta.df delta.p.value     delta.cfi
       19.827        19.000         0.405         0.000

Model 4: equal loadings + intercepts + means:
    chisq        df    pvalue       cfi     rmsea       bic
 1732.314   529.000     0.000     0.855     0.067 62183.796

[Model 1 versus model 4]
  delta.chisq      delta.df delta.p.value     delta.cfi
     1239.513        43.000         0.000         0.145

[Model 3 versus model 4]
  delta.chisq      delta.df delta.p.value     delta.cfi
    1204.184         5.000         0.000         0.145
```

The output above for both the measurement model and the structural model of the SEM show very similar results to what was observed with the initial CFA measurement invariance results. This is because only the first two latent factors (f1 & f2) contain group differences; while the remaining elements in the SEM do not display group differences (i.e. f3, f4, & f5 measurement structures). For those interested in duplicating everything done in this article (and seeing the results of the SEM fit with groups specified);

please see the RSS Do-it-yourself Introduction to R web site[3] and specifically here[4] in Module 9.

Lastly, it is very important to realize the example above used simulated data in order to demonstrate many aspects of measurement invariance. The examples above used a relatively small data set (n = 1002). Large sample sizes typically seen when conducting SEM are likely to provide statistically significant chi-square change statistics (chi-square is very sensitive to large sample sizes). Large sample sizes reduce the utility of the chi-square test. The implication being, that with large samples it would be very unlikely to establish measurement invariance using the chi-square change statistics. Therefore, Vandenber and Lance (2000) recommend using a CFI change of 0.2 as representative of a meaningful difference between models fit (p. 47).

Until next time; *"have I told you about Sammy Jankis?"*

2 References and Resources

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*, 411 – 423.

Beaujean, A. A. (2014). *Latent variable modeling in R: A step by step guide*. New York: Routledge.

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research, 3*(1), 111 – 121.

Pornprasertmanit, S., et al. (2015). Package 'semTools'. Documentation available at CRAN: http://cran.r-project.org/web/packages/semTools/index.html

Rosseel, Y., et al. (2015). Package 'lavaan'. Documentation available at CRAN: http://cran.r-project.org/web/packages/lavaan/index.html

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18*, 210 – 222. DOI:10.1016/j.hrmr.2008.03.003

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4 – 70.

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 1*, 1 – 7. DOI:10.1080/17405629.2012.686740

This article was last updated on March 10, 2015.

This document was created using LaTeX

[3] `http://www.unt.edu/rss/class/Jon/R_SC/`
[4] `http://www.unt.edu/rss/class/Jon/R_SC/Module9/MeasurementInvariance.R`