Psychology 454: Latent Variable Modeling

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

October, 2012

Outline

Does a test measure the same thing

1. Across groups
   - Different schools
   - Different groups (e.g., ethnicity, age, gender)
2. Across time
   - Is todays' measure the same as next year's measure?
3. Types of invariance
   - Configural: Are the arrows the same
   - Weak invariance: Are the loadings the same across groups
   - Strong invariance: Loadings and intercepts are equal across groups
   - Super strong: Loadigs, intercepts and means are equal across groups

Consider the Holzinger Swineford data set

1. 9 ability measures from two schools
   - 145 from Grant-White
   - 156 from Pasteur
2. Are the factor structures the same across schools
   - Although lavaan does this in one call, lets do it part by part
   - over all factor structure
   - factors with schools
   - constrain factors to have the same loadings, etc.

First, some descriptive statistics

```
describe(HolzingerSwineford1939,skew=FALSE)
```

	var	n	mean	sd	median	trimmed	mad	min	max	range	se
id	1	301	176.55	105.94	163.00	176.78	140.85	1.00	351.00	350.00	6.11
sex	2	301	1.51	0.50	2.00	1.52	0.00	1.00	2.00	1.00	0.03
ageyr	3	301	13.00	1.05	13.00	12.89	1.48	11.00	16.00	5.00	0.06
agemo	4	301	5.38	3.45	5.00	5.32	4.45	0.00	11.00	11.00	0.20
school*	5	301	1.52	0.50	2.00	1.52	0.00	1.00	2.00	1.00	0.03
grade	6	300	7.48	0.50	7.00	7.47	0.00	7.00	8.00	1.00	0.03
x1	7	301	4.94	1.17	5.00	4.96	1.24	0.67	8.50	7.83	0.07
x2	8	301	6.09	1.18	6.00	6.02	1.11	2.25	9.25	7.00	0.07
x3	9	301	2.25	1.13	2.12	2.20	1.30	0.25	4.50	4.25	0.07
x4	10	301	3.06	1.16	3.00	3.02	0.99	0.00	6.33	6.33	0.07
x5	11	301	4.34	1.29	4.50	4.40	1.48	1.00	7.00	6.00	0.07
x6	12	301	2.19	1.10	2.00	2.09	1.06	0.14	6.14	6.00	0.06
x7	13	301	4.19	1.09	4.09	4.16	1.10	1.30	7.43	6.13	0.06
x8	14	301	5.53	1.01	5.50	5.49	0.96	3.05	10.00	6.95	0.06
x9	15	301	5.37	1.01	5.42	5.37	0.99	2.78	9.25	6.47	0.06

describeBy each group

```
> describeBy(HolzingerSwineford1939,group=HolzingerSwineford1939$school,skew=FAL
group: Grant-White
      var   n  mean   sd median trimmed  mad   min   max range   se
sex     2 145  1.50 0.50   2.00    1.50 0.00  1.00  2.00  1.00 0.04
ageyr   3 145 12.72 0.97  13.00   12.67 1.48 11.00 16.00  5.00 0.08
...
grade   6 144  7.45 0.50   7.00    7.44 0.00  7.00  8.00  1.00 0.04
x1      7 145  4.93 1.15   5.00    4.96 1.24  1.83  8.50  6.67 0.10
x2      8 145  6.20 1.11   6.25    6.14 1.11  2.25  9.25  7.00 0.09
...
x8     14 145  5.49 1.05   5.50    5.45 0.89  3.05 10.00  6.95 0.09
x9     15 145  5.33 1.03   5.31    5.33 1.15  3.11  9.25  6.14 0.09
------------------------------------------------------------------------------
group: Pasteur
      var   n  mean   sd median trimmed  mad   min   max range   se
sex     2 156  1.53 0.50   2.00    1.53 0.00  1.00  2.00  1.00 0.04
ageyr   3 156 13.25 1.06  13.00   13.15 1.48 12.00 16.00  4.00 0.09
...
grade   6 156  7.50 0.50   7.50    7.50 0.74  7.00  8.00  1.00 0.04
x1      7 156  4.94 1.19   5.00    4.97 1.24  0.67  7.50  6.83 0.09
x2      8 156  5.98 1.23   5.75    5.89 1.11  3.50  9.25  5.75 0.10
...
```

EFA for both groups

by(HolzingerSwineford1939[,7:15],HolzingerSwineford1939[,5],fa,nfactors=

```
HolzingerSwineford1939[, 5]: Grant-White
Factor Analysis using method =  minres
Call: FUN(r = data[x, , drop = FALSE], nfactors = 3)
Standardized loadings (pattern matrix) based upon correlation matrix
     MR1   MR2   MR3   h2   u2
x1  0.09  0.06  0.64 0.50 0.50
x2  0.02 -0.03  0.51 0.26 0.74
x3  0.11 -0.03  0.64 0.47 0.53
x4  0.86 -0.04  0.04 0.76 0.24
x5  0.82  0.09 -0.03 0.70 0.30
x6  0.81 -0.04  0.05 0.68 0.32
x7  0.14  0.78 -0.19 0.60 0.40
x8 -0.11  0.79  0.18 0.69 0.31
x9  0.08  0.46  0.40 0.54 0.46

                       MR1  MR2  MR3
SS loadings           2.23 1.53 1.44
Proportion Var        0.25 0.17 0.16
Cumulative Var        0.25 0.42 0.58
Proportion Explained  0.43 0.29 0.28
Cumulative Proportion 0.43 0.72 1.00

 With factor correlations of
     MR1  MR2  MR3
MR1 1.00 0.25 0.41
MR2 0.25 1.00 0.31
MR3 0.41 0.31 1.00
```

```
HolzingerSwineford1939[, 5]: Pasteur
Factor Analysis using method =  minres
Call: FUN(r = data[x, , drop = FALSE], nfactors = 3)
Standardized loadings (pattern matrix) based upon corr
     MR1   MR2   MR3   h2   u2
x1  0.27  0.59  0.00 0.51 0.49
x2  0.03  0.49 -0.16 0.25 0.75
x3 -0.08  0.73  0.01 0.50 0.50
x4  0.80  0.02  0.06 0.68 0.32
x5  0.92 -0.07 -0.06 0.79 0.21
x6  0.78  0.13  0.06 0.70 0.30
x7  0.06 -0.14  0.71 0.52 0.48
x8 -0.02  0.13  0.60 0.39 0.61
x9 -0.02  0.37  0.40 0.34 0.66

                       MR1  MR2  MR3
SS loadings           2.22 1.36 1.10
Proportion Var        0.25 0.15 0.12
Cumulative Var        0.25 0.40 0.52
Proportion Explained  0.48 0.29 0.23
Cumulative Proportion 0.48 0.77 1.00

 With factor correlations of
     MR1  MR2  MR3
MR1 1.00 0.27 0.26
MR2 0.27 1.00 0.14
MR3 0.26 0.14 1.00
```

How similar are the solutions: factor congruence

Factor congruence is the cosine of the angle between two vectors:

$$Congruence = (X'X)^{.5} Y'X (Y'Y)^{.5}$$

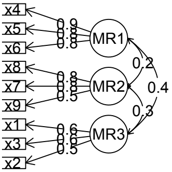
```
> f3.pasteur <- fa(HolzingerSwineford1939[1:156,7:15],3)
> f3.grant <- fa(HolzingerSwineford1939[157:301,7:15],3)
>factor.congruence(f3.pasteur,f3.grant)
```

```
     MR1  MR2  MR3
MR1 0.97 0.03 0.10
MR2 0.12 0.11 0.99
MR3 0.08 0.97 0.05
```

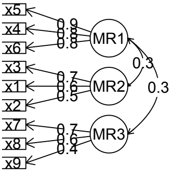
```
  cross <- t(y) %*% x
  sumsx <- sqrt(1/diag(t(x) %*% x))
   sumsy <- sqrt(1/diag(t(y) %*% y))
```

Do they look alike?

Factor Analysis



Factor Analysis



Test if the model fits the combined data

```
HS.model <-  visual   =~ x1 + x2 + x3
              textual =~ x4 + x5 + x6
              speed   =~ x7 + x8 + x9

 fit <- cfa(HS.model, data=HolzingerSwineford1939,std.lv=TRUE)
summary(fit, fit.measures=TRUE)
```

```
lavaan (0.4-14) converged normally after 41 iterations

  Number of observations                           301

  Estimator                                         ML
  Minimum Function Chi-square                   85.306
  Degrees of freedom                                24
  P-value                                        0.000

Chi-square test baseline model:

  Minimum Function Chi-square                  918.852
  Degrees of freedom                                36
  P-value                                        0.000

Full model versus baseline model:

  Comparative Fit Index (CFI)                    0.931
  Tucker-Lewis Index (TLI)                       0.896

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -3737.745
  Loglikelihood unrestricted model (H1)      -3695.092
```

With values of

```
Number of free parameters                          21
  Akaike (AIC)                                7517.490
  Bayesian (BIC)                              7595.339
  Sample-size adjusted Bayesian (BIC)         7528.739

Root Mean Square Error of Approximation:

  RMSEA                                          0.092
  90 Percent Confidence Interval          0.071  0.114
  P-value RMSEA <= 0.05                          0.001

Standardized Root Mean Square Residual:

  SRMR                                           0.065

Parameter estimates:

  Information                                 Expected
  Standard Errors                             Standard

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  visual =~
    x1                0.900    0.081   11.127    0.000
    x2                0.498    0.077    6.429    0.000
    x3                0.656    0.074    8.817    0.000
  textual =~
    x4                0.990    0.057   17.474    0.000
    x5                1.102    0.063   17.576    0.000
    x6                0.917    0.054   17.082    0.000
  speed =~
    x7                0.619    0.070    8.903    0.000
```

Now do it for both groups one analysis

```
> fit2 <- cfa(HW.model, data=HolzingerSwineford1939, group="school",std.lv=TRUE)
> summary(fit2)

  Number of observations per group
  Pasteur                                           156
  Grant-White                                       145

  Estimator                                          ML
  Minimum Function Chi-square                   115.851
  Degrees of freedom                                 48
  P-value                                         0.000

Chi-square for each group:

  Pasteur                                        64.309
  Grant-White                                    51.542

Parameter estimates:

  Information                                  Expected
  Standard Errors                              Standard
```

Results for Pasteur

```
Group 1 [Pasteur]:

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  visual =~
    x1                1.047    0.132    7.934    0.000
    x2                0.412    0.110    3.753    0.000
    x3                0.597    0.108    5.525    0.000
  textual =~
    x4                0.946    0.079   11.927    0.000
    x5                1.119    0.089   12.604    0.000
    x6                0.827    0.068   12.230    0.000
  speed =~
    x7                0.591    0.106    5.557    0.000
    x8                0.665    0.102    6.531    0.000
    x9                0.545    0.097    5.596    0.000

Covariances:
  visual ~~
    textual           0.484    0.086    5.600    0.000
    speed             0.299    0.109    2.755    0.006
  textual ~~
    speed             0.325    0.100    3.256    0.001

Intercepts:
    x1                4.941    0.095   52.249    0.000
    x2                5.984    0.098   60.949    0.000
    x3                2.487    0.093   26.778    0.000
    x4                2.823    0.092   30.689    0.000
    x5                3.995    0.105   38.183    0.000
    x6                1.922    0.079   24.321    0.000
    x7                4.432    0.087   51.181    0.000
```

Results for Grant-White

```
Latent variables:
  visual =~
    x1                 0.777    0.103    7.525    0.000
    x2                 0.572    0.101    5.642    0.000
    x3                 0.719    0.093    7.711    0.000
  textual =~
    x4                 0.971    0.079   12.355    0.000
    x5                 0.961    0.083   11.630    0.000
    x6                 0.935    0.081   11.572    0.000
  speed =~
    x7                 0.679    0.087    7.819    0.000
    x8                 0.833    0.087    9.568    0.000
    x9                 0.719    0.086    8.357    0.000

Covariances:
  visual ~~
    textual            0.541    0.085    6.355    0.000
    speed              0.523    0.094    5.562    0.000
  textual ~~
    speed              0.336    0.091    3.674    0.000

Intercepts:
    x1                 4.930    0.095   51.696    0.000
    x2                 6.200    0.092   67.416    0.000
    x3                 1.996    0.086   23.195    0.000
    x4                 3.317    0.093   35.625    0.000
    x5                 4.712    0.096   48.986    0.000
    x6                 2.469    0.094   26.277    0.000
    x7                 3.921    0.086   45.819    0.000
    x8                 5.488    0.087   63.174    0.000
    x9                 5.327    0.085   62.571    0.000
    visual             0.000
```

Constrain the loadings to be the same

```
> fit2e <- cfa(HW.model, data=HolzingerSwineford1939, group="school",std.lv=TRUE,group.equal="loadings")
> summary(fit2e)

lavaan (0.4-14) converged normally after 30 iterations

  Number of observations per group
  Pasteur                                          156
  Grant-White                                      145

  Estimator                                         ML
  Minimum Function Chi-square                  127.834
  Degrees of freedom                                57
  P-value                                        0.000

Chi-square for each group:

  Pasteur                                       71.064
  Grant-White                                   56.770

Parameter estimates:

  Information                                 Expected
  Standard Errors                             Standard
```

With parameter values of

```
Group 1 [Pasteur]:

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  visual =~
    x1                0.866    0.078   11.149    0.000
    x2                0.523    0.076    6.916    0.000
    x3                0.683    0.071    9.689    0.000
  textual =~
    x4                0.954    0.056   17.002    0.000
    x5                1.033    0.061   17.012    0.000
    x6                0.870    0.052   16.750    0.000
  speed =~
    x7                0.630    0.066    9.500    0.000
    x8                0.752    0.065   11.586    0.000
    x9                0.650    0.064   10.205    0.000

Covariances:
  visual ~~
    textual           0.485    0.087    5.555    0.000
    speed             0.341    0.109    3.126    0.002
  textual ~~
    speed             0.336    0.094    3.590    0.000

Intercepts:
    x1                4.941    0.092   53.661    0.000
    x2                5.984    0.099   60.420    0.000
    x3                2.487    0.093   26.734    0.000
    x4                2.823    0.093   30.400    0.000
    x5                3.995    0.100   39.756    0.000
    x6                1.922    0.081   23.732    0.000
```

Now, compare the fit of the two group model to the one with equal parameters

```
>  anova(fit2,fit2e)

Chi Square Difference Test

      Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
fit2  48 7484.4 7706.8 115.85
fit2e 57 7478.4 7667.4 127.83     11.982       9     0.2143
```

But, another way to specify the fit2e model – without making the latent variables standardized

```
fit2e <- cfa(HW.model, data=HolzingerSwineford1939, group="school",group.equal=c("loadings"))
 Number of observations per group
  Pasteur                                           156
  Grant-White                                       145

  Estimator                                          ML
  Minimum Function Chi-square                   124.044
  Degrees of freedom                                 54
  P-value                                         0.000

Chi-square for each group:
  Pasteur                                        68.825
  Grant-White                                    55.219
Parameter estimates:
  Information                                  Expected
  Standard Errors                              Standard

Group 1 [Pasteur]:
                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  visual =~
    x1                1.000
    x2                0.599    0.100    5.979    0.000
    x3                0.784    0.108    7.267    0.000
  textual =~
    x4                1.000
    x5                1.083    0.067   16.049    0.000
    x6                0.912    0.058   15.785    0.000
  speed =~
    x7                1.000
```

frame

```
> anova(fit2,fit2e)

Chi Square Difference Test

      Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
fit2  48 7484.4 7706.8 115.85
fit2e 54 7480.6 7680.8 124.04     8.1922       6     0.2244
```

Continue this logic, of successive tests with more constraints, but do it automatically

```
mi <- measurementInvariance(HW.model, data=HolzingerSwineford1939, group="school")
Measurement invariance tests:
Model 1: configural invariance:
   chisq        df    pvalue       cfi     rmsea       bic
 115.851    48.000     0.000     0.923     0.097  7706.822
Model 2: weak invariance (equal loadings):
   chisq        df    pvalue       cfi     rmsea       bic
 124.044    54.000     0.000     0.921     0.093  7680.771
[Model 1 versus model 2]
  delta.chisq      delta.df delta.p.value     delta.cfi
        8.192         6.000         0.224         0.002
Model 3: strong invariance (equal loadings + intercepts):
   chisq        df    pvalue       cfi     rmsea       bic
 164.103    60.000     0.000     0.882     0.107  7686.588
[Model 1 versus model 3]
  delta.chisq      delta.df delta.p.value     delta.cfi
       48.251        12.000         0.000         0.041
[Model 2 versus model 3]
  delta.chisq      delta.df delta.p.value     delta.cfi
       40.059         6.000         0.000         0.038
Model 4: equal loadings + intercepts + means:
   chisq        df    pvalue       cfi     rmsea       bic
 204.605    63.000     0.000     0.840     0.122  7709.969
[Model 1 versus model 4]
  delta.chisq      delta.df delta.p.value     delta.cfi
       88.754        15.000         0.000         0.083
[Model 3 versus model 4]
  delta.chisq      delta.df delta.p.value     delta.cfi
       40.502         3.000         0.000         0.042
>
```

Using sim.structural to create models

```
fx <-matrix(c( .9,.8,.6,rep(0,4),.6,.8,-.7),ncol=2)
fy <- matrix(c(.6,.5,.4),ncol=1)
rownames(fx) <- c("V","Q","A","nach","Anx")
rownames(fy)<- c("gpa","Pre","MA")
Phi <-matrix( c(1,0,.7,.0,1,.6,.7,.6,1),ncol=3)
gre.gpa <- sim.structural(fx,Phi,fy,n=1000)
gre.gpa
```

```
Call: sim.structural(fx = fx, Phi = Phi, fy = fy)

 $model (Population correlation matrix)
        V    Q     A  nach   Anx   gpa   Pre    MA
V    1.00 0.72  0.54  0.00  0.00  0.38  0.32  0.25
Q    0.72 1.00  0.48  0.00  0.00  0.34  0.28  0.22
A    0.54 0.48  1.00  0.48 -0.42  0.47  0.39  0.31
nach 0.00 0.00  0.48  1.00 -0.56  0.29  0.24  0.19
Anx  0.00 0.00 -0.42 -0.56  1.00 -0.25 -0.21 -0.17
gpa  0.38 0.34  0.47  0.29 -0.25  1.00  0.30  0.24
Pre  0.32 0.28  0.39  0.24 -0.21  0.30  1.00  0.20
MA   0.25 0.22  0.31  0.19 -0.17  0.24  0.20  1.00

$reliability (population reliability)
   V    Q    A nach  Anx  gpa  Pre   MA
0.81 0.64 0.72 0.64 0.49 0.36 0.25 0.16
```

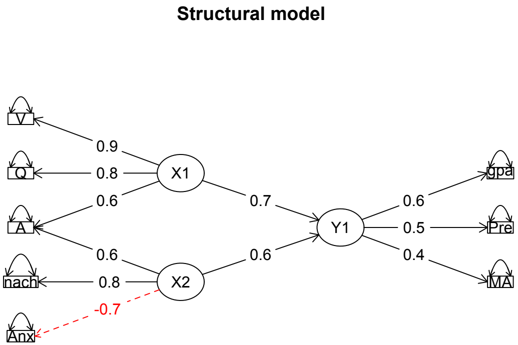

```
$reliability (population reliability)
   V    Q    A nach  Anx  gpa  Pre   MA
0.81 0.64 0.72 0.64 0.49 0.36 0.25 0.16
```


```
> fx
> fy
> Phi
```

```
>  fx
     [,1] [,2]
V     0.9  0.0
Q     0.8  0.0
A     0.6  0.6
nach  0.0  0.8
Anx   0.0 -0.7
> Phi
     [,1] [,2] [,3]
[1,]  1.0  0.0  0.7
[2,]  0.0  1.0  0.6
[3,]  0.7  0.6  1.0
> fy
    [,1]
gpa  0.6
Pre  0.5
MA   0.4
```

Create a figure (and write sem code)

```
mod4 <- structure.diagram(fx,Phi,fy,errors=TRUE,e.size=.3)
```



Show the sem code

```
> mod4

      Path           Parameter Value
 [1,] "X1->V"        "F1V"     NA
 [2,] "X1->Q"        "F1Q"     NA
 [3,] "X1->A"        "F1A"     NA
 [4,] "X2->A"        "F2A"     NA
 [5,] "X2->nach"     "F2nach"  NA
 [6,] "X2->Anx"      "F2Anx"   NA
 [7,] "V<->V"        "x1e"     NA
 [8,] "Q<->Q"        "x2e"     NA
 [9,] "A<->A"        "x3e"     NA
[10,] "nach<->nach"  "x4e"     NA
[11,] "Anx<->Anx"    "x5e"     NA
[12,] "Y1->gpa"      "Fygpa"   NA
[13,] "Y1->Pre"      "FyPre"   NA
[14,] "Y1->MA"       "FyMA"    NA
[15,] "gpa<->gpa"    "y1e"     NA
[16,] "Pre<->Pre"    "y2e"     NA
[17,] "MA<->MA"      "y3e"     NA
[18,] "X2<->X1"      "rF2F1"   NA
[19,] "X1->Y1"       "rX1Y1"   NA
[20,] "X2->Y1"       "rX2Y1"   NA
[21,] "X1<->X1"      NA        "1"
[22,] "X2<->X2"      NA        "1"
[23,] "Y1<->Y1"      NA        "1"
attr(,"class")
[1] "mod"
```