GC-3 Workshop: Measurement Invariance Testing in R

Yichi Zhang

11/03/2022

Outline

- Brief overview of measurement invariance (MI)
  - Definition of MI, measurement noninvariance, partial invariance
  - Importance of validating the MI assumption
  - Levels of MI
- Illustrative example
  - Investigate MI of the mini-IPIP (International Personality Item Pool) scale across gender.
  - Small Exercise

Motivating Example

- The Positive and Negative Affect Schedule (PANAS) scale has been found to function differently for participants from different cultural backgrounds.

Table 1. Positive and Negative Affect Schedule (PANAS) Scorecard [27].

1 Very Slightly or Not at All	2 A Little	3 Moderately	4 Quite a Bit	5 Extremely

______ 1. Interested	______ 11. Irritable
______ 2. Distressed	______ 12. Alert
______ 3. Excited	______ 13. Ashamed
______ 4. Upset	______ 14. Inspired
______ 5. Strong	______ 15. Nervous
______ 6. Guilty	______ 16. Determined
______ 7. Scared	______ 17. Attentive
______ 8. Hostile	______ 18. Jittery
______ 9. Enthusiastic	______ 19. Active
______ 10. Proud	______ 20. Afraid

- Are the observed differences in scale scores caused by real differences in negative affect or the different cultural meanings of scale items?

What Is Measurement Invariance (MI)?

- Formal definition:
- Implied meaning: Using the same questionnaire in different groups (such as countries or at various points in time, or under different conditions) does measure the same construct in the same way.

What Is Measurement Noninvariance (MNI)?

- If there is a violation of MI => Measurement noninvariance/Item bias/Differential item functioning
  - The scale has different measurement properties across groups for individuals with the same latent construct level
  - E.g., Male consistently reported a higher score than female in mathmatics self-efficacy scale.
- If MI only holds for a subset of items => Partial invariance

How To Test For MI?

- Multi-Group Confirmatory Factor Analysis (MGCFA)
  - Likelihood ratio test to compare nested models, which means one model with equality constraints of a particular parameter and the other without such constraints.
  - E.g., the invariance of loadings can be tested by comparing the model with freely estimated loadings and the model with has equality constraints of loadings across groups
- Item Response Theory

Multi-Group Confirmatory Factor Analysis (MGCFA)

- CFA assumes the observed items load on a latent factor that represents the construct
- A single-factor MG-CFA

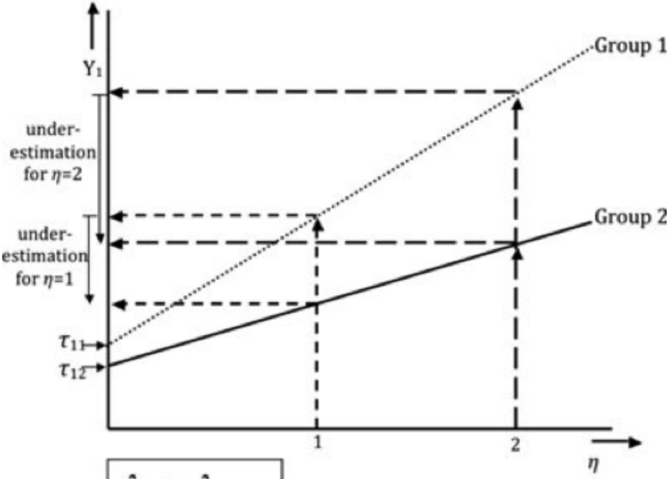
$$y_{ik} = \nu_k + \lambda_k \eta_{ik} + \epsilon_{ik}$$

- y_{ik} and η_{ik} are the observed continuous response and the latent construct score for the ith person in the kth group
- ν_k represents intercepts, λ_k represents factor loadings, and the unique factor variables represents ϵ_{ik}.

Levels Of MI Testing

- Configural Invariance: the same model holds for all the groups
- Metric/Weak Invariance: factor loadings (slopes) are the same across the groups
- Scalar/Strong Invariance: intercepts and loadings are the same across the groups
- Strict Invariance: unique factor invariance, intercepts and loadings are the same across the groups

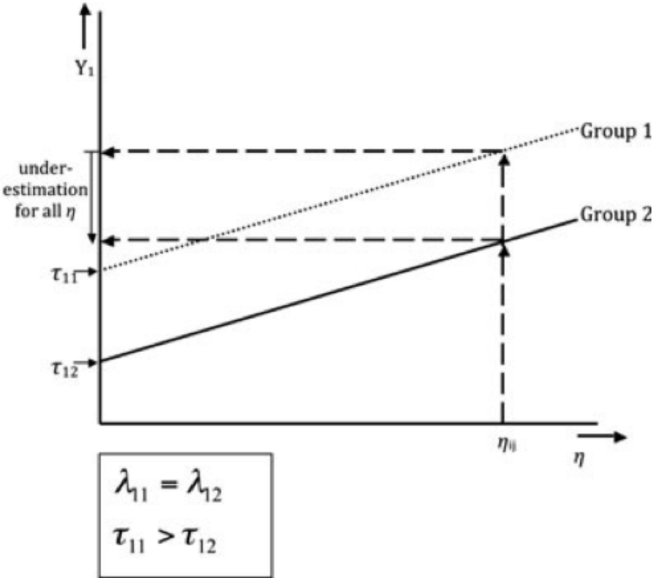
Configural Invariance



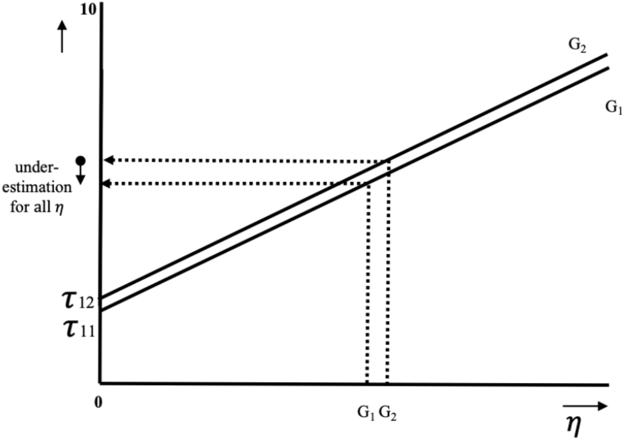
$\lambda_{11} > \lambda_{12}$

$\tau_{11} > \tau_{12}$

Metric Invariance



Scalar Invariance



Illustrative Example

- Goal: Examine measurement invariance of the mini-IPIP across gender.
- Data: Collected from 1994 Spring to 1996 Fall (Ock et al., 2020)
  - 564 participants (239 males, 325 females)
  - 20 items in total, four items per dimension (Agreeableness, Conscientiousness, Extraversion, Neuroticism, Openness to Experience)
  - 5-point Likert-type scale from 1 (*very inaccurate*) to 5 (*very accurate*).

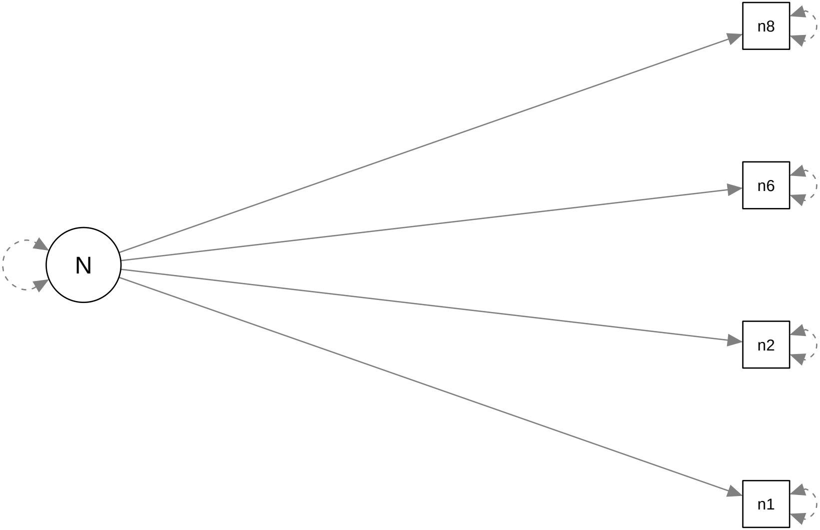
Model Specification

```r
data <- read.table(here::here("example", "IPIPFFM.dat"),
                   header = TRUE)%>%
  select("sex", "e1", "e4", "e6", "e7", "n1", "n2", "n6", "n8")
head(data)
```

```
##   sex e1 e4 e6 e7 n1 n2 n6 n8
## 1   1  4  3  4  4  4  3  3  3
## 2   1  3  2  2  2  2  4  4  4
## 3   1  4  5  4  4  1  1  2  2
## 4   1  2  4  4  3  3  4  4  4
## 5   1  2  4  4  3  3  2  3  2
## 6   1  2  4  4  4  2  2  2  1
```

```r
mod_n <- 'N =~ n1 + n2 + n6 + n8'
# check model fit
fit_n <- cfa(mod = mod_n, data = data)
# summary(fit_e, fit.measures = TRUE)
```

```r
semPlot::semPaths(lavaanify(mod_n), rotation = 2, intercepts = TRUE)
```



Testing Metric Invariance

- Metric Invariance add equality constraints on factor loadings across gender
- The Likelihood Ratio Test (LRT) suggests the metric invariance model has similar fit as the configural invariance model

```r
# fit configural invariance model
fit_con <- cfa(mod = mod_n, data = data,
               group = "sex", std.lv = TRUE)
# fit metric invariance model (loadings are constrained to be equal across groups)
fit_met <- cfa(mod = mod_n, data = data, group = "sex",
               group.equal = "loadings", std.lv = TRUE)
# Likelihood Ratio Test
lavTestLRT(fit_con, fit_met)
```

```
## Chi-Squared Difference Test
##
##         Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## fit_con  4 6547.7 6651.8 5.9861
## fit_met  7 6542.5 6633.6 6.8001    0.81397       3     0.8461
```

Testing Scalar Invariance

- Scalar Invariance add equality constraints on factor loadings and intercepts across groups
- The LRT suggests the scalar invariance model fit significantly worse than the metric invariance model

```
# fit scalar invariance model
fit_sca <- cfa(mod = mod_n, data = data, group = "sex",
               group.equal = c("loadings", "intercepts"),
               std.lv = TRUE)
# Likelihood Ratio test of scalar invairance model and metric invariance model
lavTestLRT(fit_met, fit_sca)
```

```
## Chi-Squared Difference Test
##
##         Df    AIC    BIC   Chisq Chisq diff Df diff Pr(>Chisq)
## fit_met  7 6542.5 6633.6  6.8001
## fit_sca 10 6553.1 6631.1 23.3668     16.567       3  0.0008676 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Option 1: `pinSearch` Function From `pinsearch` Package

- This function automates the iterative search for noninvariant parameters (loadings, intercepts, residuals)
- Item `n1` has noninvariant intercepts across groups

```
res <- pinsearch::pinSearch(config_mod = mod_n, data = data, effect_size = TRUE,
                     group = "sex",  type = "intercepts")
res$`Non-Invariant Items`
```

```
##   lhs rhs group       type
## 1  n1         2 intercepts
```

Option 2: Sequential MI Testing By Scale Items

- Manually searching for noninvariant parameters (loadings, intercepts, residuals)
- Find the item associated with the largest change in χ^2 => item `n1`

```
# fit scalar invariance model
fit_sca1 <- cfa(mod_n, data = data, group = "sex",
                group.equal = c("loadings", "intercepts"),
                group.partial = "n1 ~ 1", std.lv = TRUE)
lavTestLRT(fit_sca1, fit_sca)
```

```
## Chi-Squared Difference Test
##
##          Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## fit_sca1  9 6545.5 6627.8 13.736
## fit_sca  10 6553.1 6631.1 23.367     9.6311       1   0.001913 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
fit_sca2 <- cfa(mod_n, data = data, group = "sex",
                group.equal = c("loadings", "intercepts"),
                group.partial = "n2 ~1", std.lv = TRUE)
lavTestLRT(fit_sca2, fit_sca)
```

```
## Chi-Squared Difference Test
##
##          Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## fit_sca2  9 6552.5 6634.9 20.753
## fit_sca  10 6553.1 6631.1 23.367     2.6133       1      0.106
```

```r
fit_sca6 <- cfa(mod_n, data = data, group = "sex",
                group.equal = c("loadings", "intercepts"),
                group.partial = "n6 ~1", std.lv = TRUE)
lavTestLRT(fit_sca6, fit_sca)
```

```
## Chi-Squared Difference Test
##
##          Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## fit_sca6  9 6554.0 6636.4 22.308
## fit_sca  10 6553.1 6631.1 23.367     1.0589       1     0.3035
```

```r
fit_sca8 <- cfa(mod_n, data = data, group = "sex",
                group.equal = c("loadings", "intercepts"),
                group.partial = "n8 ~1", std.lv = TRUE)
lavTestLRT(fit_sca8, fit_sca)
```

```
## Chi-Squared Difference Test
##
##          Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## fit_sca8  9 6546.4 6628.7 14.629
## fit_sca  10 6553.1 6631.1 23.367     8.7383       1   0.003116 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testing Strict Invariance

- The partial strict invariance model fit significantly worse than the partial scalar invariance model

```
fit_str <- cfa(mod_n, data = data, group = "sex",
               group.equal = c("loadings", "intercepts", "residuals"),
               group.partial = "n1 ~ 1", std.lv = TRUE)
lavTestLRT(fit_sca1, fit_str)
```

```
## Chi-Squared Difference Test
##
##          Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## fit_sca1  9 6545.5 6627.8 13.736
## fit_str  13 6555.8 6620.9 32.101     18.365       4   0.001047 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testing Strict Invariance

- Items `n1`,`n2` has noninvariant unique factor variances across gender
- Effect size (Standardized mean difference) is 0.239

```
res_str <- pinsearch::pinSearch(config_mod = mod_n, data = data, effect_size = TRUE,
                     group = "sex",  type = "residuals")
res_str$`Non-Invariant Items`
```

```
##   lhs rhs group       type
## 1  n1         2 intercepts
## 2  n2  n2     1  residuals
## 3  n1  n1     1  residuals
```

```
res_str$effect_size
```

```
##            n1-N
## dmacs 0.2389943
```

Partial Strict Invariance Model

- Conclusion: items `n6`, `n8` are strict invariant, item `n2` is strong/scalar invariant, item `n1` is weak/metric invariant
- Final model: partial strict invariance model with freely estimated intercept for item `n1` and unique factor variances for items `n1` and `n2`.

```
final_fit <- cfa(mod_n, data = data, group = "sex",
                 group.equal = c("loadings", "intercepts", "residuals"),
                 group.partial = c("n1 ~ 1", "n1 ~~ n1", "n2 ~~ n2"), std.lv = TRUE)
```

Exercise time

Please complete the file “exercise.RMD”.

Additional Topic:

- MI testing for categorical data
- Approximate invariance for many groups
- Practical significance of MI