# Measurement Invariance

Are we measuring the same thing in different groups?

# Intro

Today's goal:

If we run a CFA in multiple groups, find out whether we are measuring the same thing in each group

Outline:

- Intro to measurement invariance
- Partial invariance
- Practical example in Mplus (from a paper that is currently under review!)

# Types of invariance

Testing different levels of equivalence between groups

# Multiple groups

Let's say we run a CFA in multiple groups...

Are measuring the same thing in each group?

This is called measurement invariance

**Why do we want to know this?**

Out of fairness

e.g are some IQ-test questions biased against women?

# Multiple groups

To see if comparisons are warranted

e.g. can we compare measures of satisfaction between cultures?

To detect conceptual differences between groups

e.g. do privacy practices have similar meanings in different cultures?

To detect conceptual drift over time

e.g. does “information overload” mean the same thing now as it did 20 years ago?

# Types of invariance

Configural invariance (equal form invariance)

Does the same model work for both (all) groups?

Metric invariance (equal factor loadings)

Are the loadings similar* between groups?

Scalar invariance (equal intercepts/thresholds)

Are the intercepts/thresholds similar between groups?

(in most cases this is added on top of metric invariance)

# Types of invariance

Equivalence of construct variances/covariances

Are the factor variances and correlations the same?

(only makes sense on top of metric invariance)

Equivalence of residual variances/covariances

Are the uniquenesses and residual correlations the same?

(also only makes sense on top of metric invariance)

Full equivalence

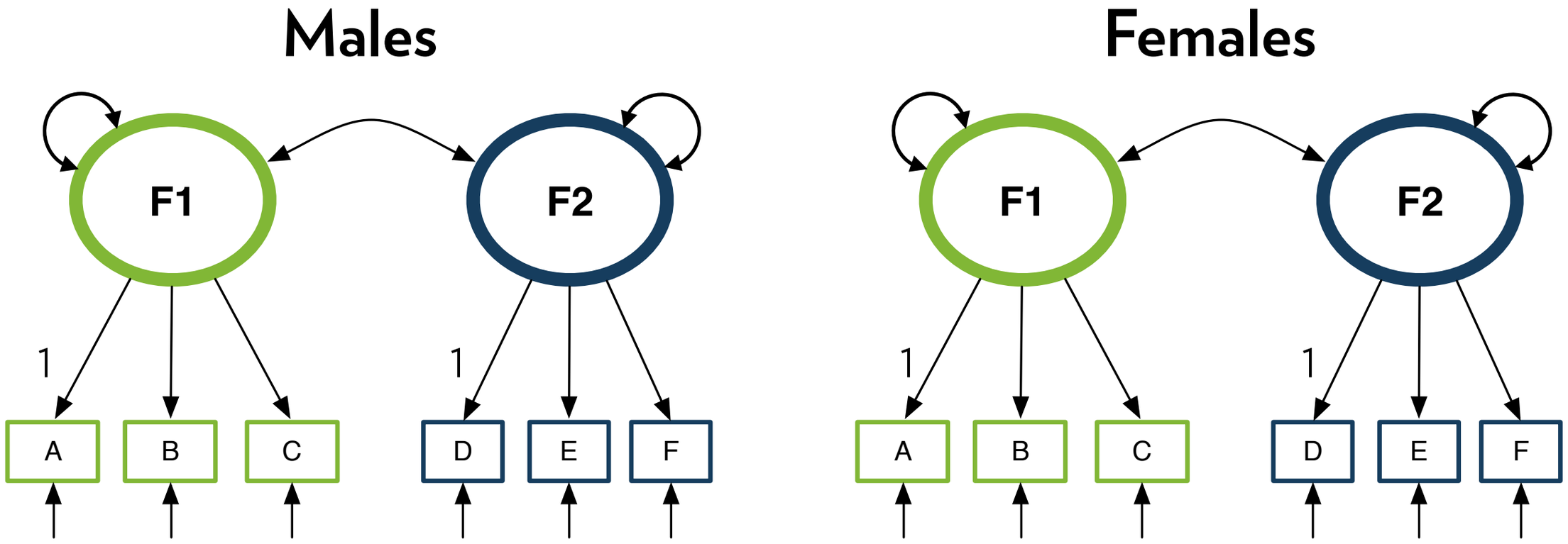
Can we combine the two groups and run a single model?

# Configural

Fit two models simultaneously

Allow estimates (loadings, uniqueness, etc.) to be different

Do these models both fit reasonably well?



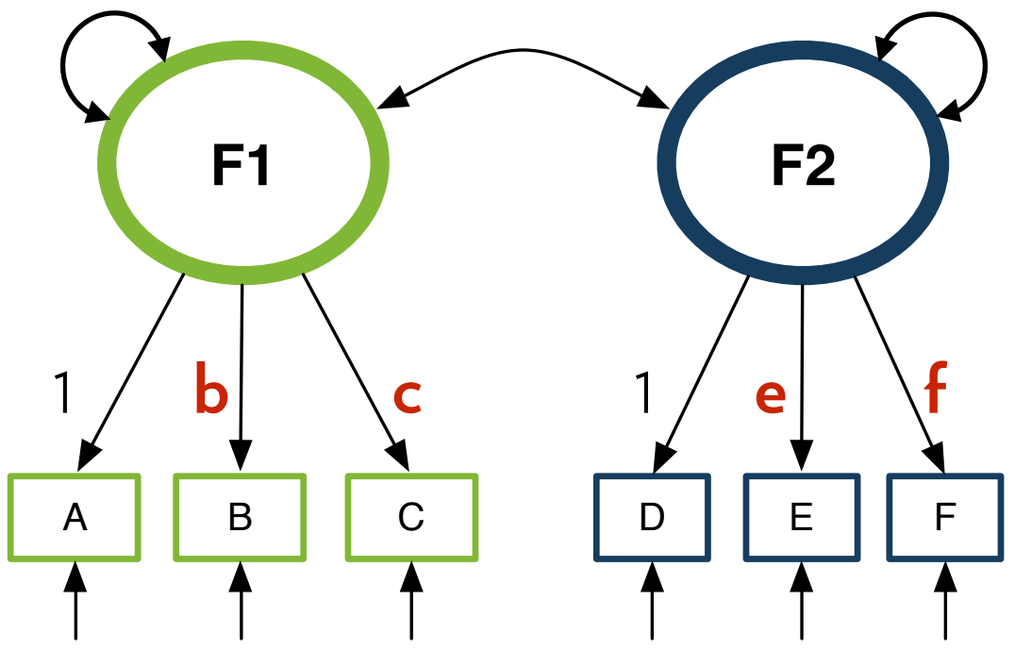
# Metric

Make the **loadings** equal between groups

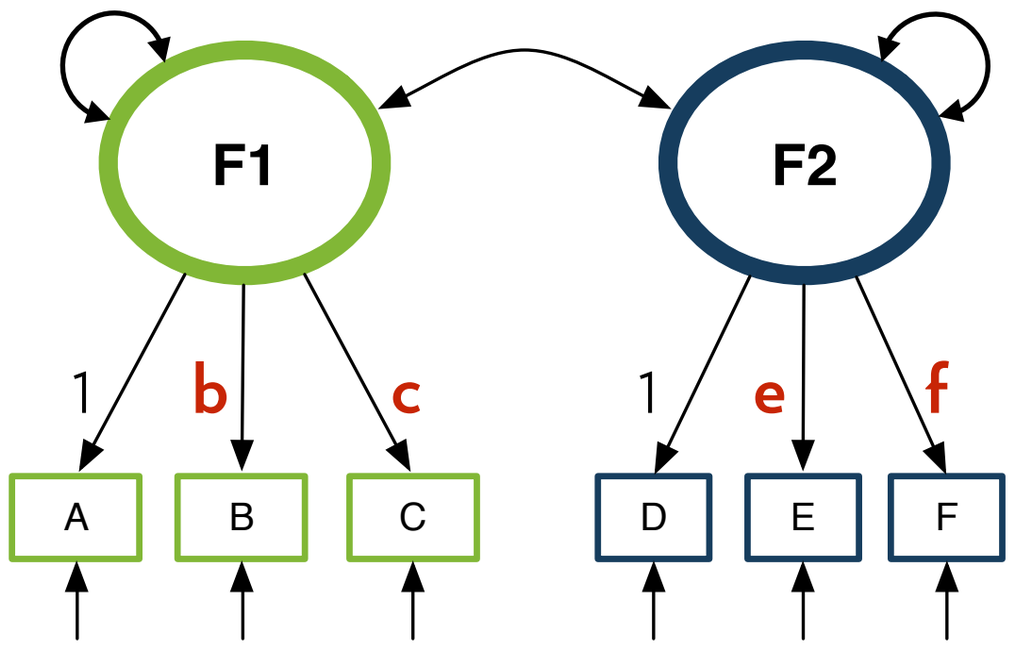
Does this fit significantly worse than the configural model?

If so, metric noninvariance: different contribution per item

## Males




## Females



# Scalar

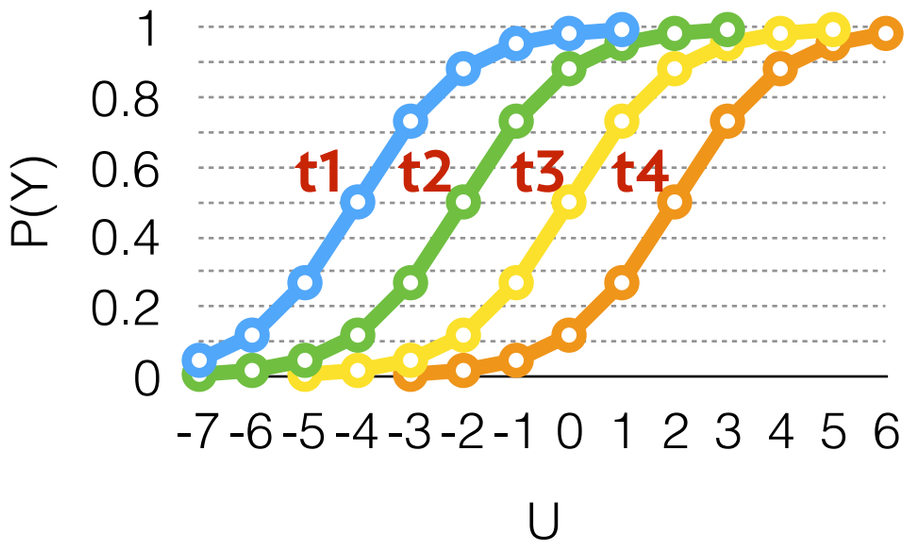
Make the **intercepts/thresholds** equal between groups

Does this fit significantly worse than the configural model?

If so, scalar noninvariance: item is biased

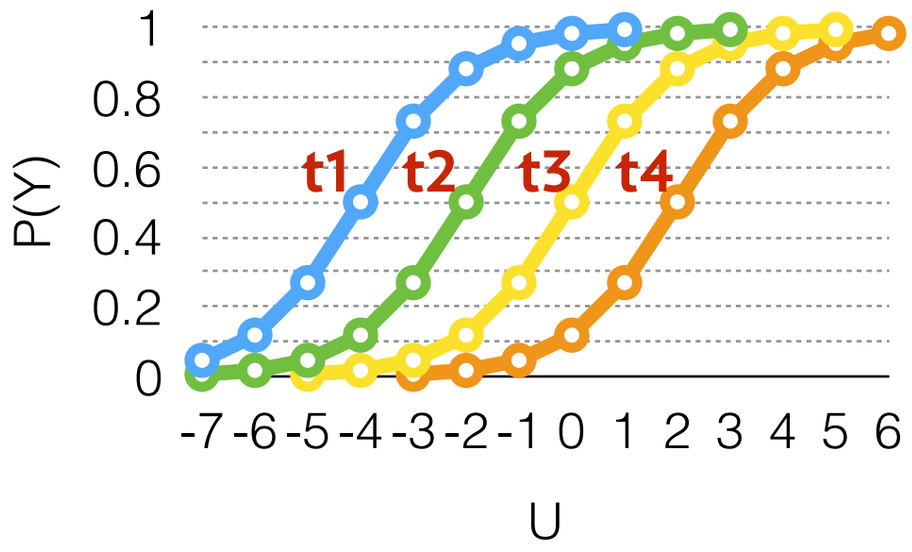
## Males

**for each item:**




## Females

**for each item:**



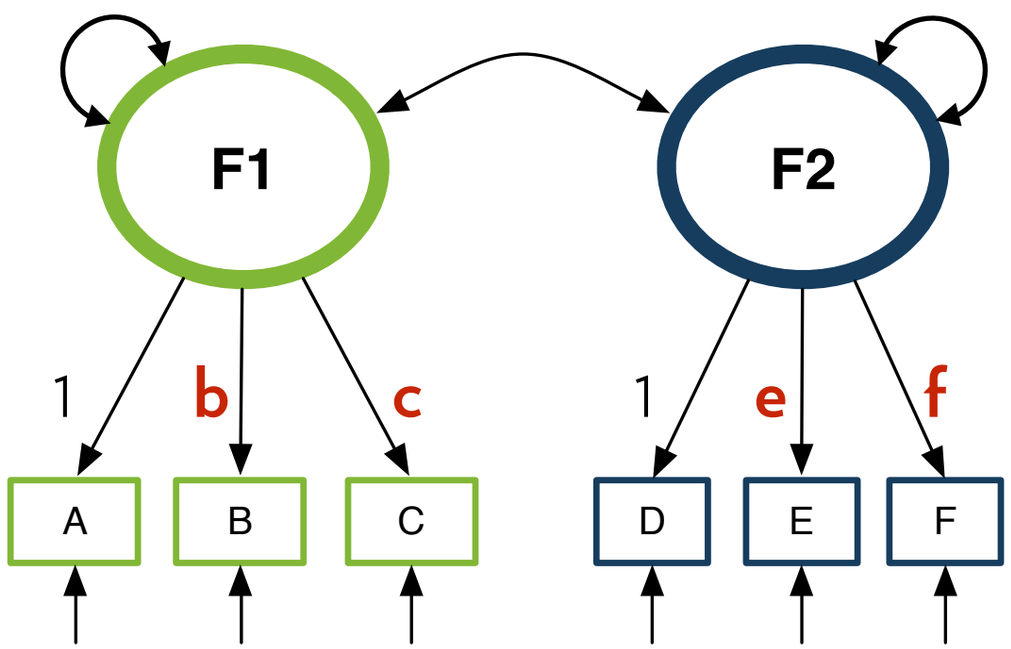
# Scalar+Metric

Make **loadings and intercepts/thresholds** equal

More common (why?)

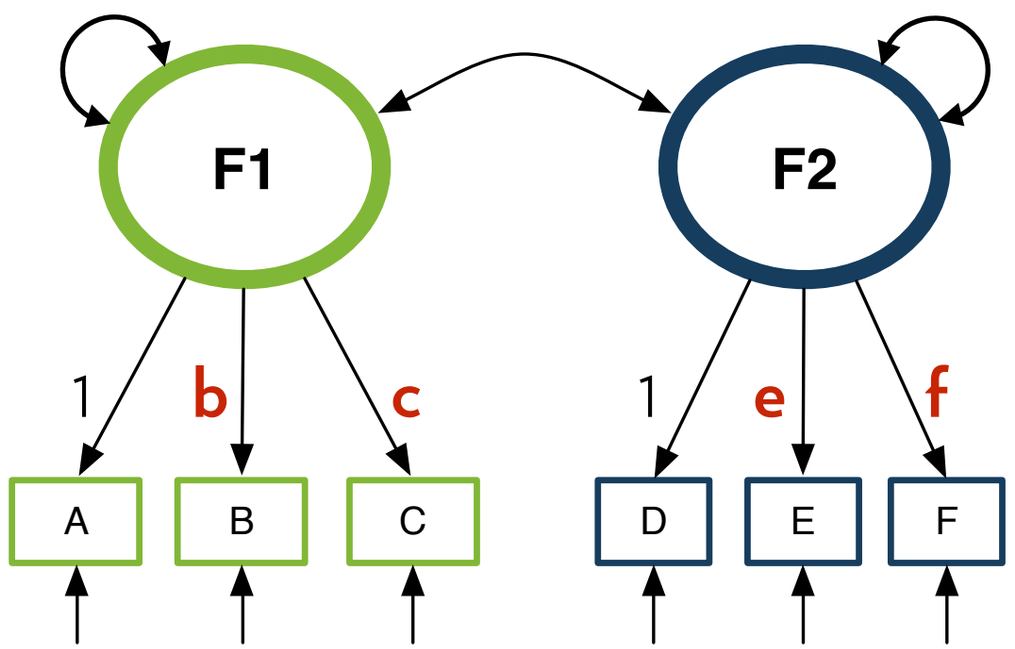
Usually tested against the metric model

## Males




+t1, t2, t3, t4 for each item

## Females



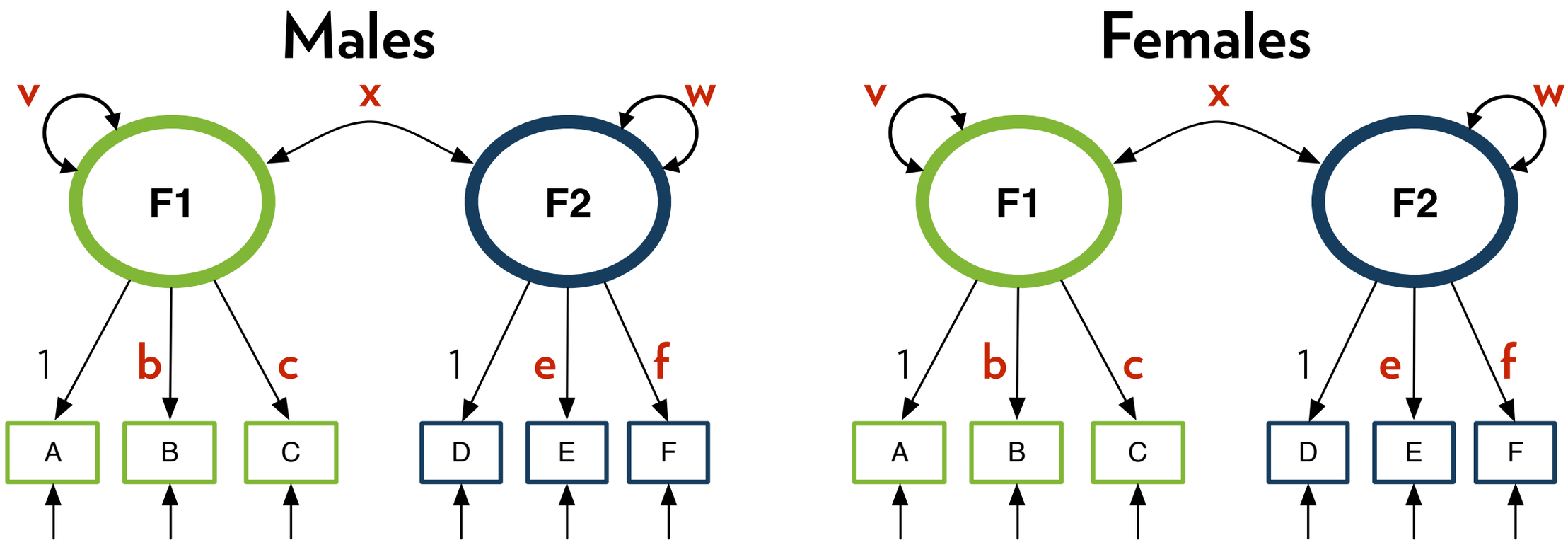

+t1, t2, t3, t4 for each item

# Construct var/cov

Make **construct variances and covariances** equal

Usually done on top of (and tested against) metric model

Do the groups equally (co-)vary on the construct?

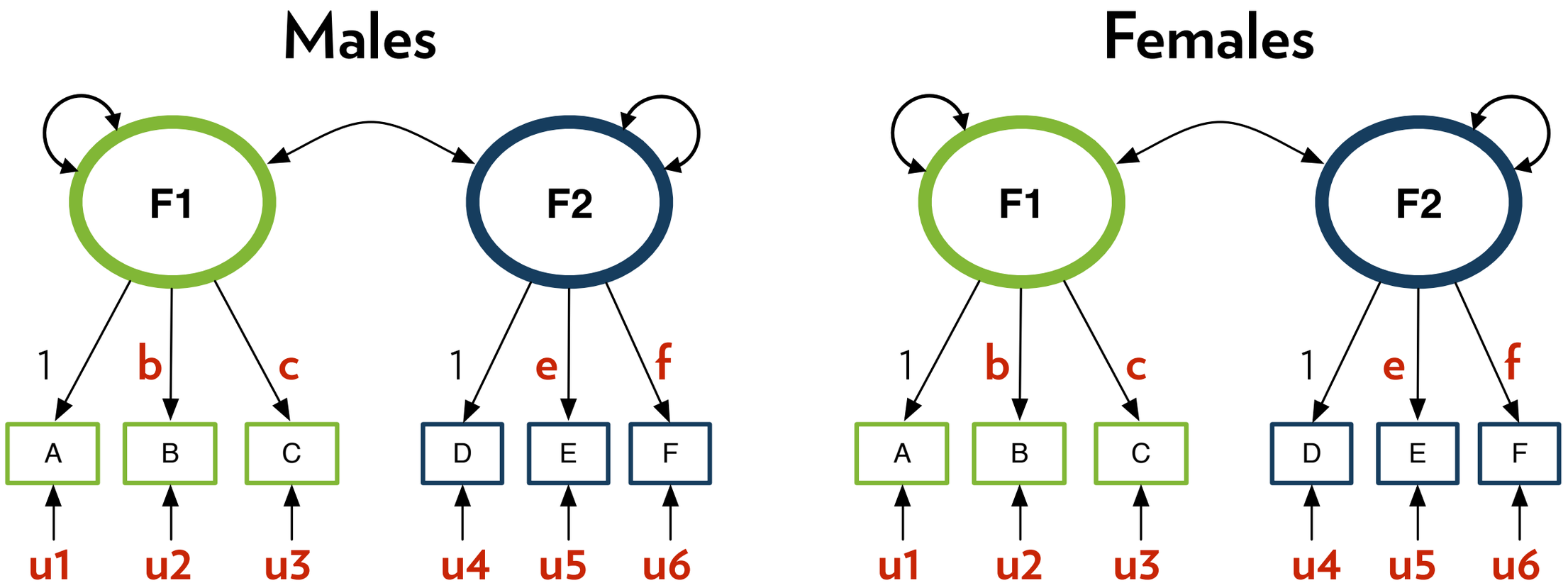


# Residual var/cov

Make **residual variances (and covariances)** equal

Usually done on top of (and tested against) metric model

Are the uniquenesses equal across groups?

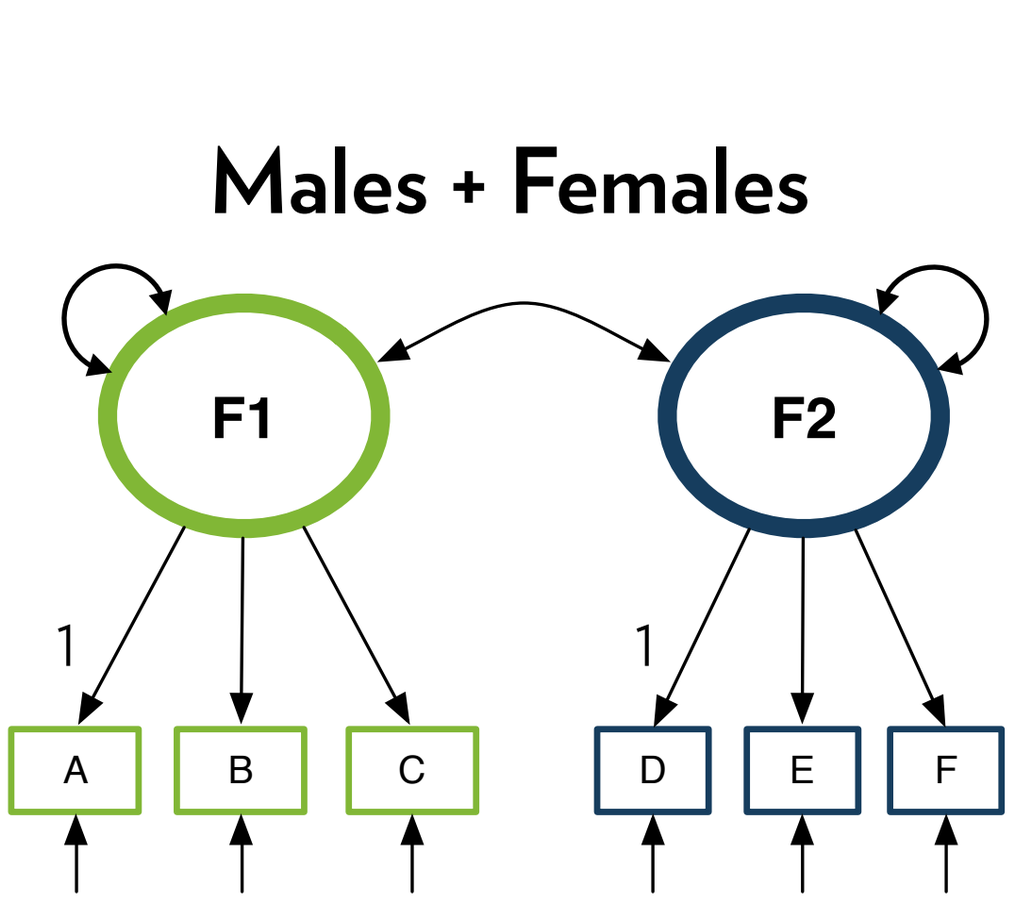


# Full equivalence

Make everything equal between groups

Essentially, fit a single model on both groups

Test vs. more relaxed model (e.g. metric+scalar+construct)



# Tests

Tests between different levels of equivalence are conducted as chi-square model comparisons

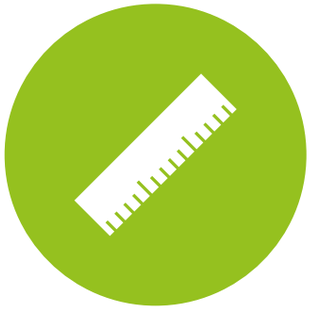
- With large N, likely significant!

Solution: also look at approximate fit statistics, especially CFI

- If CFI differs < .002, essentially no effect
- If CFI differs < .01, likely no effect

# Partial invariance

What if some (but not all) parameters are different?



# Partial invariance

Question: What if we don't have metric invariance?

i.e. the metric model is significantly worse than the configural model

Answer: Go for partial metric invariance!

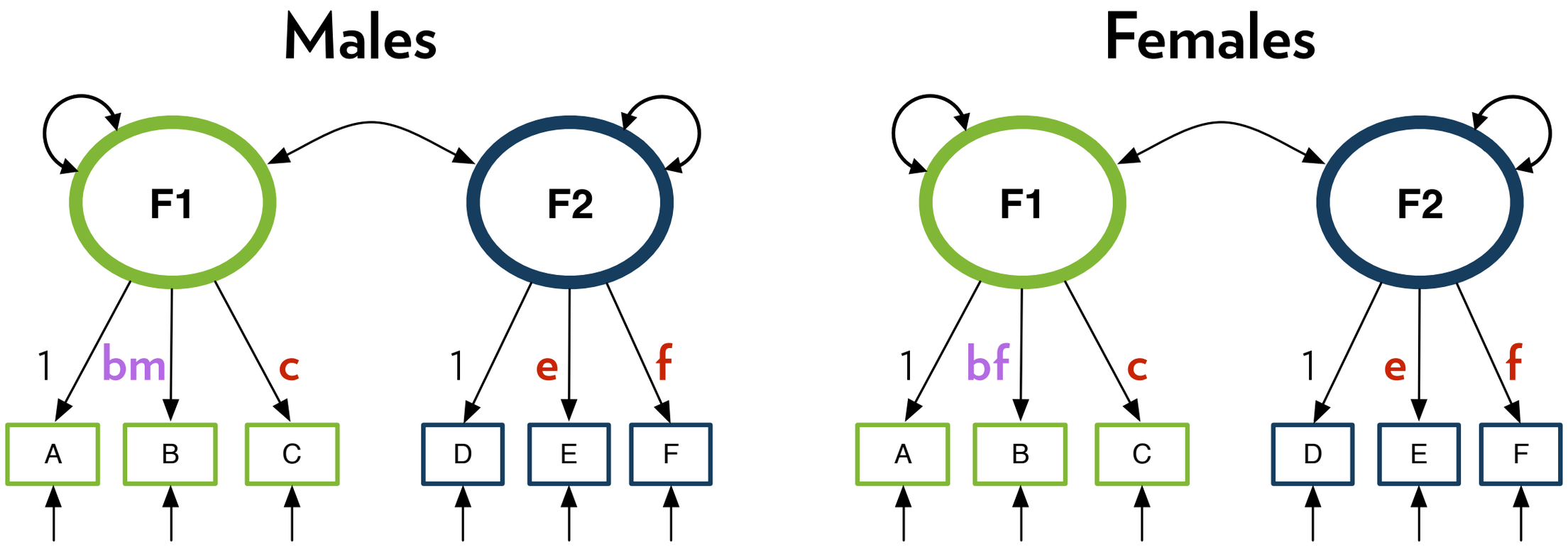
Inspect the loadings in the configural model

Which loadings differ the most? These are "differential functioning indicators"

# Partial invariance

Run the metric model with one parameter relaxed

- Test against configural model
- Still significant? Repeat until no longer significant!

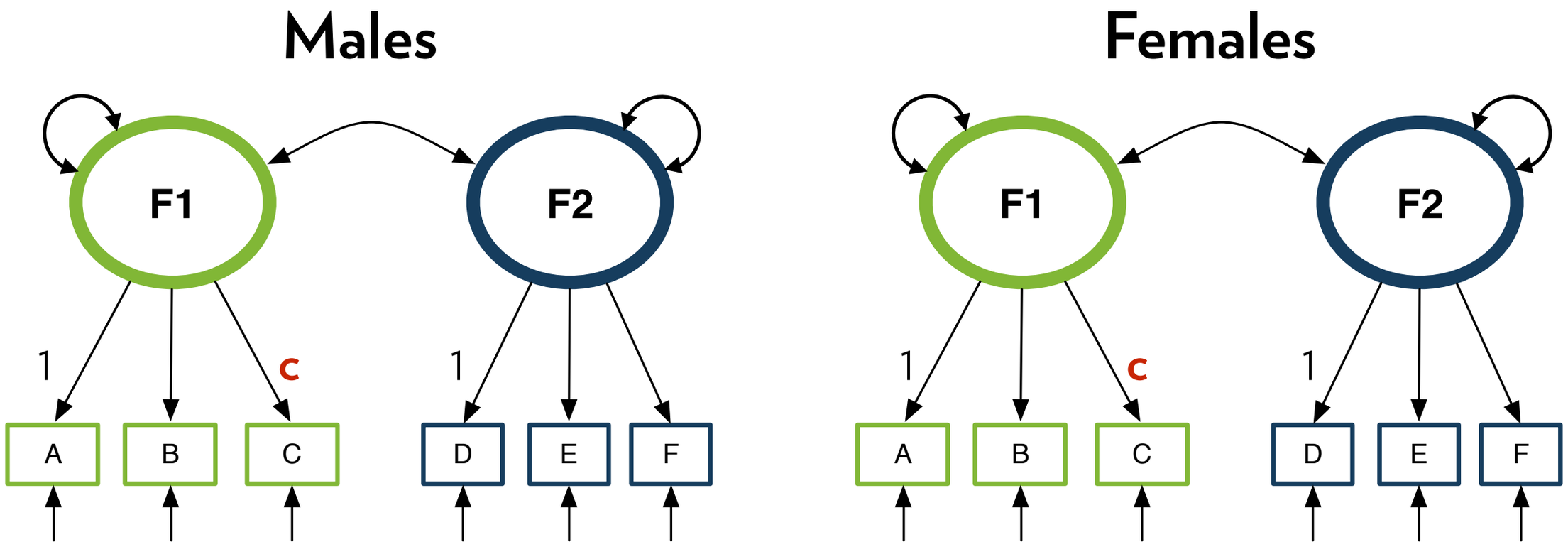


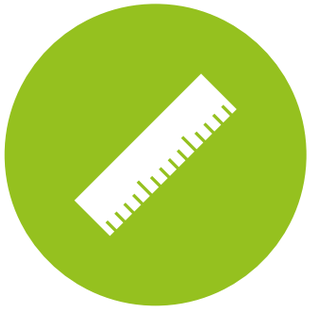
# Partial invariance

You can also start from the configural model

Constrain one parameter and test against configural

If not significant, continue until it is (then one step back)





# Partial invariance

Notes:

- You can also do this for partial scalar invariance, and even partial equivalence of construct or residual variances and covariances
- You can combine, and have a model that is e.g. fully metric invariant + partially scalar invariant

# Theoretical value

## The constraints you relax may have theoretical value:

### Conceptual differences

If certain items load differentially, this tells us how the concept differs between groups

e.g. surveillance-related privacy items load stronger on privacy concern for women than for men

# Theoretical value

**The constraints you relax may have theoretical value:**

Biased items

If certain items have higher/lower intercepts/thresholds, this shows us how the scale may be biased

e.g. IQ questions that use football metaphors may have a higher intercept for US participants, so such items should be avoided

# Theoretical value

**The constraints you relax may have theoretical value:**

Differences in discriminant validity

If certain factors have much higher correlation in one group, this shows a difference in the complexity of users' attitudes/perceptions/behaviors

e.g. preventative and collaborative privacy management strategies have a much higher correlation in the US than in Korea and Singapore (to the point where there is a lack of discriminant validity)

# Practical example

in Mplus

# Dataset

**500_dataset.csv:** survey of collective privacy management strategies in three countries (the US, Singapore, and Korea)

Columns:

- cntry: 1=US, 2=SG, 3=KR
- male: gender (0=female, 1=male)
- age

# Dataset

Columns (continued):

- csuntag-csauntag: corrective strategies (7 items)
- icconaud-iclimsha: information control strategies (3 items)
- psrespos-pstimrvw: preventive strategies (4 items)
- clnegrul-cledupri: collaborative strategies (7 items)
- nminvite-nmcontac: network control strategies (3 items)
- psuprof-psuaddtg: audience control strategies (6 items)

All 7-point scales; pre-trimmed (all items fit)

# Invariance tests

**invariance.inp**

Grouping parameter to indicate the 3 groups

R will automatically run a configural model as specified under “model”

We added some residual covariances, and specified them configurally (separate parameters) for each country

You can do this under model US, model SG, model KR

Note: items not specified as ordered categorical!

# Invariance tests

**invariance.inp**

ANALYSIS: model = configural metric scalar

This automatically tests these three types of invariance!

(the latter is metric+scalar)

# Configural

Outcome:

- $\chi^2(1156) = 2099.409$, $p < .0001$
- CFI = .935
- TLI = .927
- RMSEA = .070 [.065, .075], p(RMSEA<=.05) < .001

# Metric invariance

Outcome:

- $\chi^2(1204) = 2192.533$
- Against configural: $\chi^2(48) = 93.124$, p = .0001
- CFI = .932 (decreases .003)
- TLI = .927
- RMSEA = .070 [.066, .075], p(RMSEA<=.05) < .001

No full metric invariance!

But pretty close

# Scalar+metric

Outcome:

- $\chi^2(1252) = 2319.573$
- Against metric: $\chi^2(48) = 127.040$, $p < .0001$
- CFI = .927 (decreases .006)
- TLI = .924
- RMSEA = .072 [.067, .076], p(RMSEA<=.05) < .001

# Partial metric

**metric.inp**

The metric model, with modification indices

Result: relax constraint on iclimsha

Determined after some trial and error

# Partial metric

## partial metric 1.inp

MODEL SG: infoctrl BY iclimsha*;

MODEL KR: infoctrl BY iclimsha*;

The metric model, with relaxed constraint on iclimsha

- $\chi^2(1202) = 2169.304$
- Against configural: $\chi^2(46) = 69.895$, p = .0131
- Modification indices suggest relaxin constraints on icadjcon

# Partial metric

## partial metric 2.inp

MODEL SG: infoctrl BY iclimsha* icadjcon*;

MODEL KR: infoctrl BY iclimsha* icadjcon*;

The metric model, with relaxed constraints on iclimsha and icadjcon

- $\chi^2(1200) = 2156.372$
- Against configural: $\chi^2(44) = 56.963$, p = .09
- CFI: .934, TLI: .929 (better!), RMSEA: .069 [.065, .074]

# Which items?

**iclimsha:** I limit what I share on Facebook to only what is appropriate for all of my friends to see

Loading US: 0.921, SG: 1.416, KR: 0.650

**icadjcon:** I adjust the content of my post based on who I think will see it

Loading US: 1.048, SG: 1.523, KR: 1.008

**icconaud:** Before posting on Facebook I consider the audience that will read my post

Loading US: 1.000, SG: 1.000, KR: 1.000

# Scalar+partial metric

**partial metric full scalar.inp**

Change model to scalar;

MODEL SG: infoctrl BY iclimsha* icadjcon*; [icadjcon]; [iclimsha];

MODEL KR: infoctrl BY iclimsha* icadjcon*; [icadjcon]; [iclimsha];

This frees the intercepts for the freed loadings

Why are we doing that?

# Scalar+partial metric

Outcome:

- $\chi^2(1244) = 2235.565$
- Against partial metric: $\chi^2(44) = 79.193$, p = .0009
- CFI = .932 (decreases .002)
- TLI = .929
- RMSEA = .069 [.065, .074]
- Modification indices suggest relaxing intercepts of cldispri and cledupri

# Both partial

## partial metric partial scalar 2.inp

Add [cldispri] and [cledupri] to MODEL SG and KR

Results:

- $\chi^2(1240) = 2212.000$
- Against partial metric: $\chi^2(40) = 55.628$, p = .05
- CFI: .933, TLI: .930 (better!), RMSEA: .069 [.064, .073]

# Which items?

**cldispri:** Prior to disclosing content, my friends and I discuss the appropriate privacy settings

Intercept US: 3.620, SG: 3.740, KR: 3.950

**cledupri:** I educate my friends about privacy issues

Intercept US: 3.849, SG: 3.926, KR: 3.621

# Correlations

**cors.inp**

Add a test for the correlation between preventive and collaborative strategies

model US: preventive with colstgy (p1);

model SG: preventive with colstgy (p2);

model KR: preventive with colstgy (p3);

Also get the standardized output

# Correlations

Outcome:

- Wald test: $\chi^2(2) = 15.412$, $p = .0005$
- Correlation in US: 0.804 ($\sqrt{}$AVE of preventive is 0.726!)
- Correlation in SG: 0.584
- Correlation in KR: 0.527

Meaning: US participants do not distinguish between collaborative and preventive strategies!

# Many groups

What if I have a lot of groups (with low N in each group?)

Multiple group modeling becomes impossible!

Solution: use a random effect!

Treat your data like repeated measures

Create **random** slopes and intercepts for each indicator

Test whether the variance of these effects is significantly larger than zero

This is called the **alignment** method

“It is the mark of a truly intelligent person
to be moved by statistics.”

T H A N K S !

George Bernard Shaw