

測定不変性と構造不変性 (measurement and structural invariance)

京都大学大学院教育学研究科
D2 野崎 優樹

今日扱う内容

- 心理学研究の中で、ある集団に基づいて作成された尺度や課題を、別の集団に適用するということがしばしばある
 - その尺度や課題は、本当に異なる集団でも同等の内容を測定していると言えるだろうか？
 - 今日は、多母集団の同時分析を使って、上記の問いを定量的に検討する方法を紹介します

なぜ、このテーマが問題になるのか？

- 心理学では、しばしば潜在変数を比較する
 - 幸福感, 共感性, パーソナリティなど
- 同じ尺度や課題を使ったとしても、同じ内容を測定しているとは言いきれない！！
 - もし群間で違う内容を測定しているのなら、その比較は無意味
 - 測定値の差が、対象とする概念の度合いの差を真に反映しているかどうか重要

3

尺度や課題の測定内容の等質性が問題になる例

- 異なる集団間の比較
 - 西洋と東洋で等質な内容を測定している？
 - 熟達者-非熟達者で等質な内容を測定している？
 - 年長児と年少児で等質な内容を測定している？
- 縦断調査
 - Time 1~Time Nまで、同一の指標を使ったとして、その指標は等質の内容を測定している？

4

では、どのように検討するのか？

- 測定不変性について、いくつかのレベルに不変性を分解し、どのレベルの不変性が担保されているのかを調べる (Vandenberg & Lance, 2000)
 - configural invariance (因子数が不変)
 - metric invariance (因子負荷が不変)
 - scalar invariance (切片が不変)
 - residual invariance (誤差分散が不変)

では、どのように検討するのか？

- 潜在因子がどのように分布し、関連しているかについて、下記のレベルの不変性を調べることもできる (structural invariance; 構造不変性)
 - factor variance invariance (因子の分散が不変)
 - factor covariance invariance (共分散が不変)
 - Mean invariance (平均値が不変)

→これらの不変性の分析は、多母集団の同時分析を用いてできます！

実際の研究例

- 異なる集団間の比較
 - 情動知能の測定不変性について、性別、年代間で比較 (Tsaousis & Kazi, 2013)
 - 性別についてはscalar invariance, 年代についてはpartial scalar invarianceが満たされた
 - 性別、年代のどちらについてもfactor variance invarianceとfactor covariance invarianceが満たされた

7

実際の研究例

- 異なる集団間の比較
 - 情動知能の測定不変性について国際比較 (Ekermans et al., 2011)
 - metric invarianceまでが満たされた
- 縦断調査
 - 5時点での共感性の測定不変性について検討 (Allemand et al., in press)
 - 5時点間でのresidual invarianceが満たされた

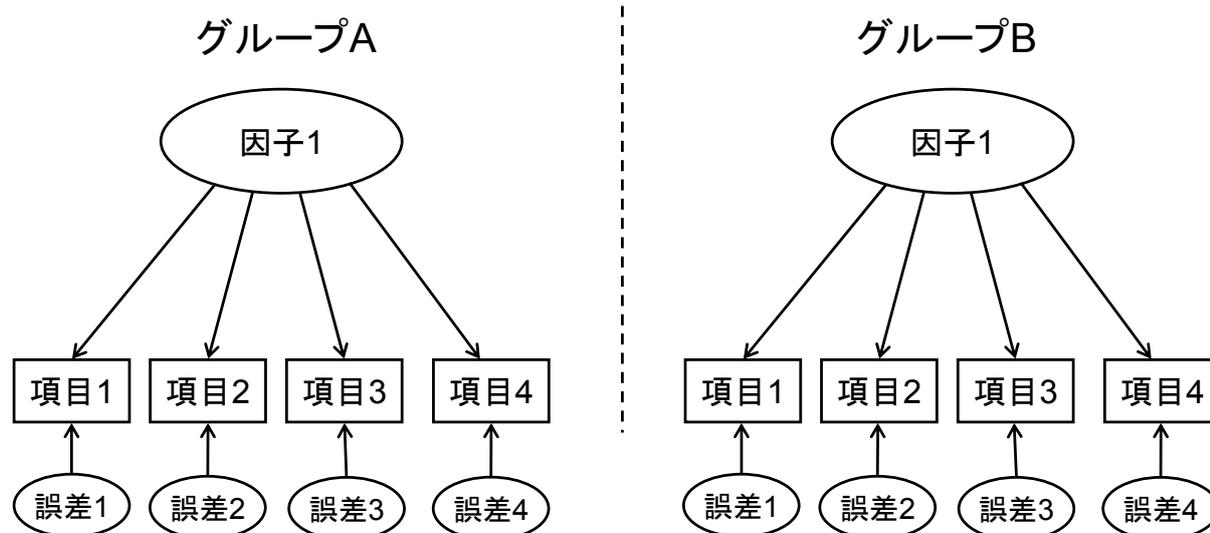
8

configural invariance

- グループ間で因子の構造は等しいか？
 - 因子数は等しいか？質問項目と因子の対応関係は等しいか？
- まず、グループごとに因子分析してみて、グループ間で因子構造が十分似ているかを検証
 - その後、2つのグループを組み合わせると同時に分析を実行

9

configural invariance



グループ間で、因子数と項目と因子の対応関係が同じ
パラメーターは、グループごとに自由推定

10

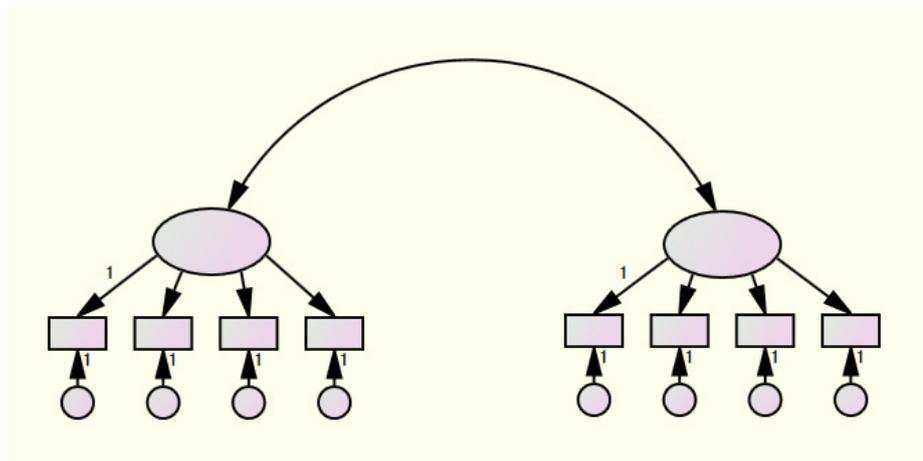
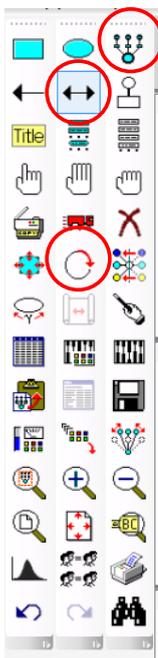
Amosを使って分析してみよう！

- 日本人300名, アメリカ人300名を対象としたダミーデータ
- 普段感じている情動の程度について質問
 - ポジティブ情動4項目 (item1-4)
 - ネガティブ情動4項目 (item5-8)
- Groupは日本:0, アメリカ:1

11

モデルの記述

- アイコンを使い, 因子分析のモデルを描く



12

モデルの記述

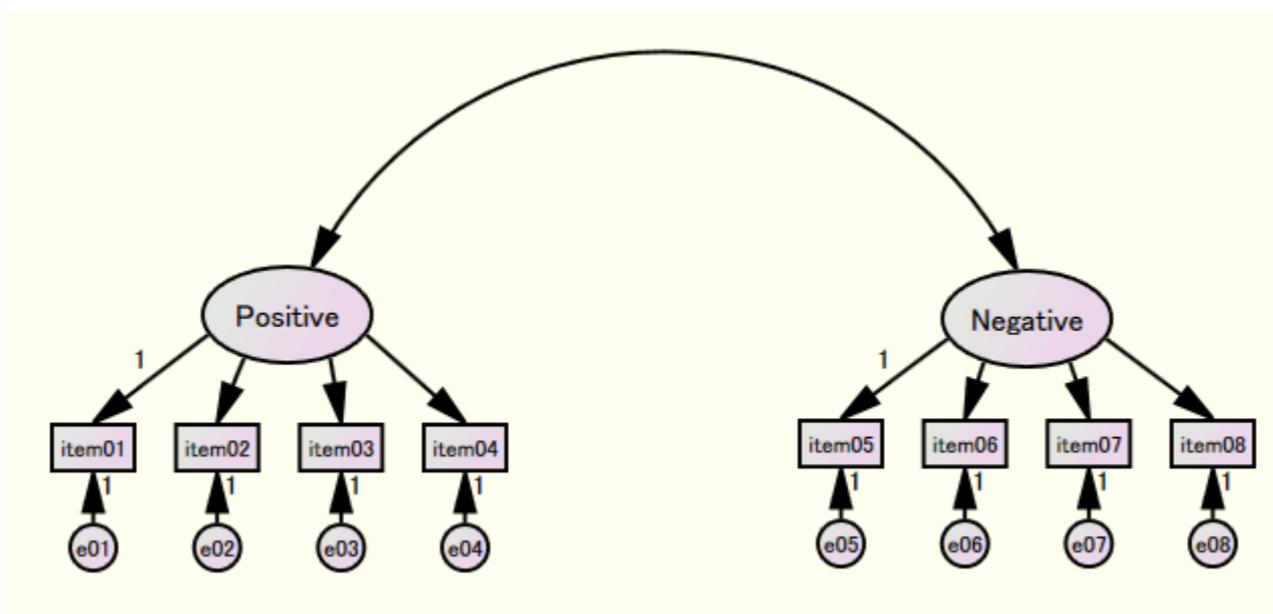
- 観測変数・潜在変数に名前をつける
 - オブジェクトの上で右クリック→オブジェクトのプロパティ→テキスト→変数名



13

モデルの記述

- 今回の分析で使うモデル



14

グループの作成

- 分析 (A) → グループ管理 (G) → グループ名のボックスに1つ目のグループ名を入力(日本) → 新規作成(A) をクリック → グループ名のボックスに2つ目のグループ名を入力(アメリカ) → 閉じる (C)



←2つのグループができたことを確認

15

データの読み込み

- データファイルを選択 → ファイル名 (N)
- グループ化変数 (G) → グループの変数を選択 (group) → OK, グループ値 (V) → 変数を選択
- もう片方のグループについても同じ方法を繰り返す

グループ名	ファイル	変数	数値	標本数
日本	サンプルデータ (XLS)	group	0	300/600
アメリカ	サンプルデータ (XLS)	group	1	300/600

Buttons: ファイル名(N), 実行中のファイル(W), ヘルプ(H), データの表示(D), グループ化変数(G), グループ値(V), OK, キャンセル

16

分析の実行

- 表示 (V) → 分析のプロパティ (A) → 出力 → 標準化推定値 (T) にチェック
- 分析のプロパティ (A) → 推定 → 平均値と切片を推定 (E) にチェック (後の分析で必要)
- 分析 (A) → 推定値を計算 (C)
- 「テキスト出力を表示」をクリック

17

分析の実行

データ分析.amw
分析の要約
グループについての注釈
変数の要約
パラメータの要約
モデルについての注釈
推定値
最小化履歴
モデル適合
実行時間

モデルについての注釈 (モデル番号 1)

自由度の計算 (モデル番号 1)

独立な標本積率の数: 88
独立な推定パラメータの数: 50
自由度 (88 - 50): 38

結果 (モデル番号 1)

最小値に達しました。
カイ2乗 = 119.336
自由度 = 38
有意確率 = .000

✓ モデルについての注釈
不適解になっていないことを確認

分析の要約
グループについての注釈
変数の要約
パラメータの要約
モデルについての注釈
推定値
最小化履歴
モデル適合
実行時間

標準化係数: (日本 - モデル番号 1)

	推定値
item04 <--- Positive	.831
item03 <--- Positive	.600
item02 <--- Positive	.582
item01 <--- Positive	.841
item08 <--- Negative	.463
item07 <--- Negative	.775
item06 <--- Negative	.910
item05 <--- Negative	.921

切片: (日本 - モデル番号 1)

✓ 推定値
標準化係数が因子負荷量
グループごとに推定値がある

18

分析の実行

CMIN

モデル	NPAR	CMIN	自由度	確率	CMIN/DF
モデル番号 1	50	138.058	38	.000	3.633
飽和モデル	88	.000	0		
独立モデル	32	2316.598	56	.000	41.368

基準比較

モデル	NFI	RFI	IFI	TLI	CFI
	Delta1	rho1	Delta2	rho2	
モデル番号 1	.940	.912	.956	.935	.956
飽和モデル	1.000		1.000		1.000
独立モデル	.000	.000	.000	.000	.000

RMSEA

モデル	RMSEA	LO 90	HI 90	PCLOSE
モデル番号 1	.066	.055	.078	.012
独立モデル	.260	.251	.269	.000

✓モデル適合

CFI, RMSEAなどを見て、モデル適合を評価

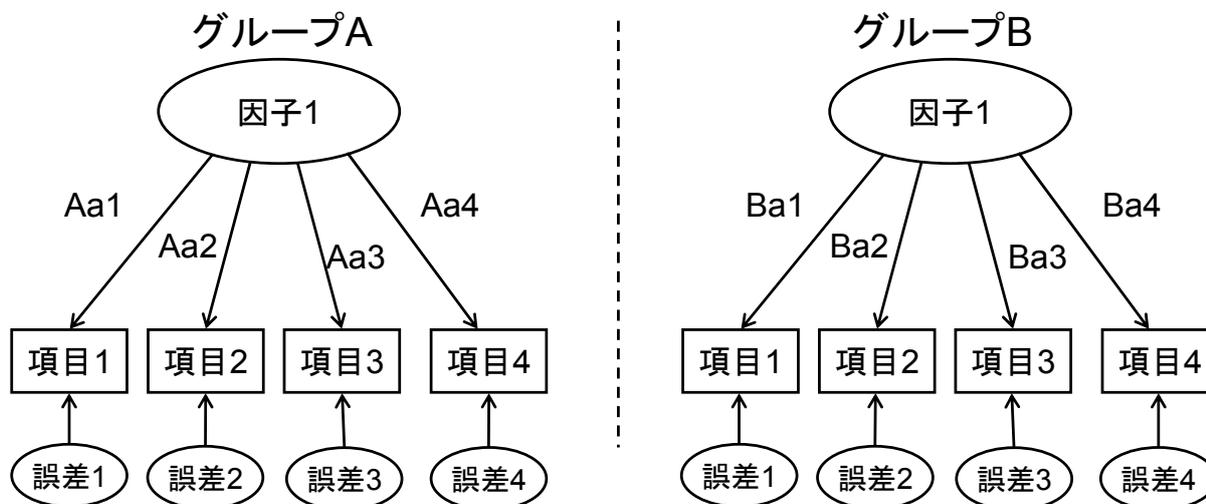
十分であれば、configural invarianceは満たされていると判断

metric invariance

- 各項目の因子負荷量はグループ間で等しいか？
– weak factorial invarianceとも呼ばれる
- 同じ項目の因子からの負荷量にグループ間で等値制約を置く

metric invariance

$$Aa1 = Bb1, Aa2 = Bb2, Aa3 = Bb3, Aa4 = Bb4$$

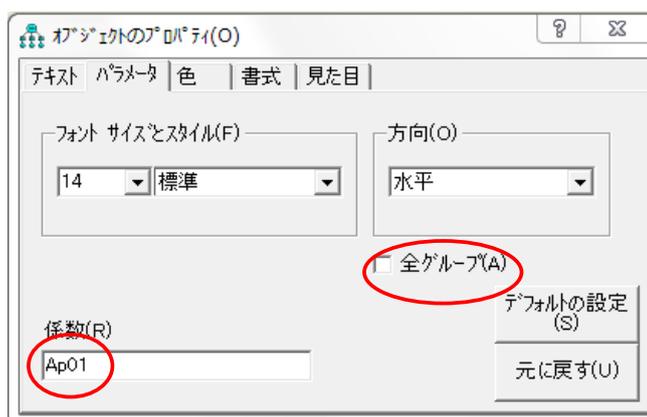


configural invariance + 因子から項目への負荷量が群間で同じ

21

Amosを使って分析してみよう！

- パスを右クリックし、オブジェクトのプロパティ (O) をクリック
- パラメータ→係数の欄に名前をつける
- 全グループのチェックを外す

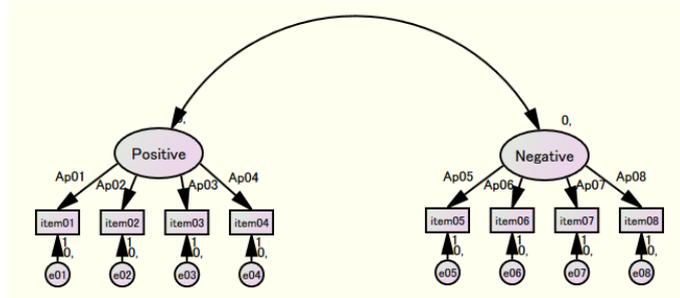


22

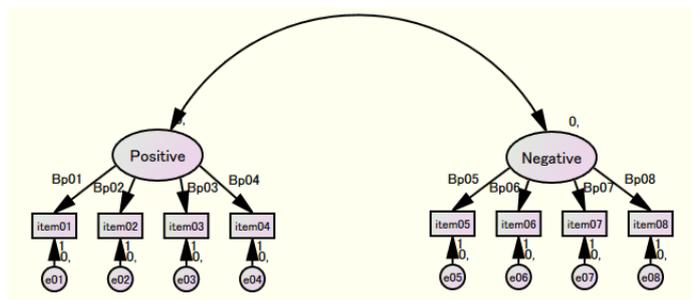
パラメーターの命名

- すべての因子負荷量のパスに名前をつける(日本とアメリカ両方)

日本



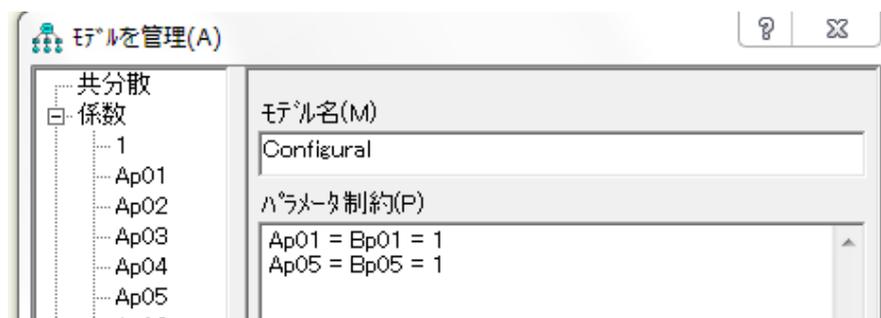
アメリカ



23

モデルの管理

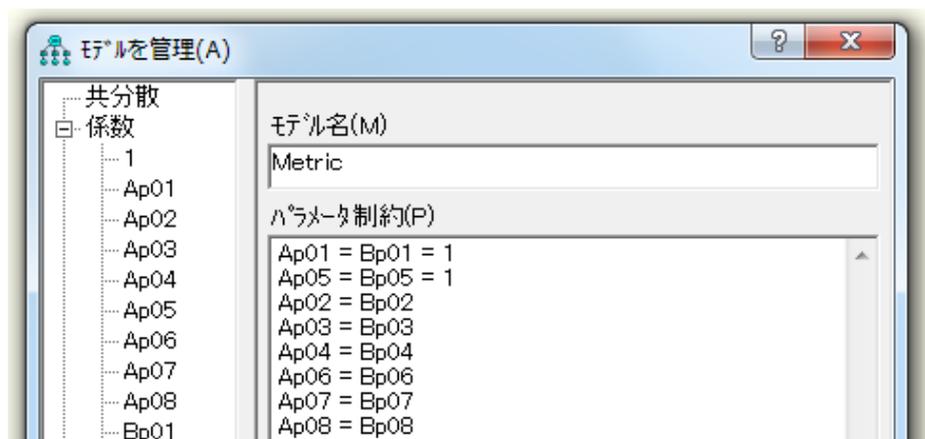
- 左のXX: モデル番号1をダブルクリック
- モデル名に名前をつける→Configural
- 識別のため、パラメーターに制約を入れる



24

モデルの管理

- 新規作成をクリック
- モデル名に名前をつける→Metric
- 因子負荷量が群間で等値という制約を入れる



25

分析の実行

- 分析 (A) → 推定値を計算 (C)
- 「テキスト出力を表示」をクリック
- 適合度指標やモデル比較の結果を見て、制約を入れたことによるモデル適合度の変化を確認

基準比較

モデル	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Configural	.940	.912	.956	.935	.956
Metric	.938	.921	.956	.944	.956
飽和モデル	1.000		1.000		1.000
独立モデル	.000	.000	.000	.000	.000

ネストしたモデルの比較

モデル Configural は正しいという仮定の下で:

モデル	自由度	CMIN	確率	NFI Delta-1	IFI Delta-2	RFI rho-1	TLI rho2
Metric	6	5.218	.516	.002	.002	-.009	-.009

26

適合度の変化の基準

- χ^2 分析→有意に悪くならなければOK
 - ただしサンプルサイズの影響を受けるので、他の基準が使われることも多い
- Δ CFIがよく用いられている
 - Δ CFI \leq .01 (Cheung & Rensvold 2002)
 - Δ CFI \leq .002 (Meade et al., 2008)
 - ただし、これらの基準はGolden Ruleというわけではない (Kline, 2010)

27

metric invariance

- すべての因子負荷量に制約を入れると、適合度が悪くなってしまった時、partial metric invarianceを検討できる (Byrne et al., 1989)
 - 一部の因子負荷量のみ制約を入れ、どの部分が群間で異なるのかを検証する
- どれぐらいの一致が見られれば partial metric invarianceと言えるのかについて、一致した見解はないが、少なくとも1因子あたり1つ以上

28

scalar invariance

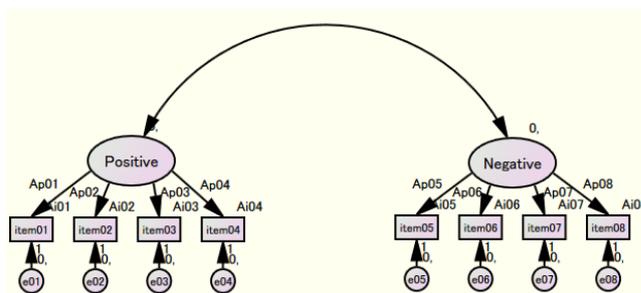
- 各項目の切片はグループ間で等しいか？
 - strong factorial invarianceとも呼ばれる
- Scalar invarianceが満たされる場合，項目レベルでの平均値の差は，因子レベルでの平均値の差であると考えることができる
 - 項目に対する反応の仕方の違いではなく，因子によって説明される部分での平均値の差が反映されている

29

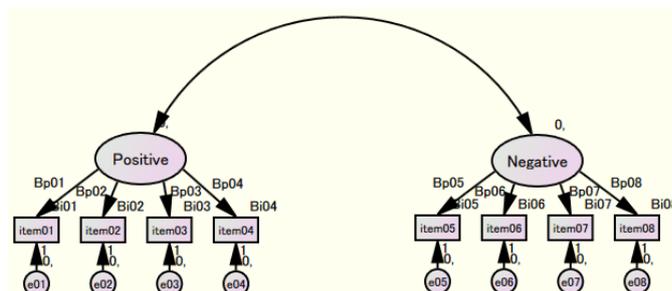
パラメーターの命名

- 因子負荷量と同様にすべての項目の切片に名前をつける(日本とアメリカ両方)

日本



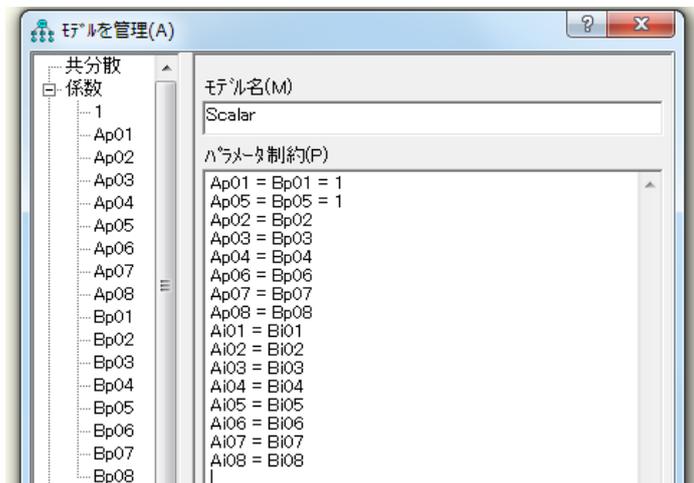
アメリカ



30

モデルの管理

- 左のXX: Metricをクリック→新規作成
- モデル名に名前をつける→Scalar
- 切片が群間で等値という制約を入れる



31

分析の実行

- 分析 (A) → 推定値を計算 (C)
- 「テキスト出力を表示」をクリック
- 適合度指標やモデル比較の結果を見て、制約を入れたことによるモデル適合度の変化を確認

基準比較

モデル	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Configural	.940	.912	.956	.935	.956
Metric	.938	.921	.956	.944	.956
Scalar	.935	.931	.957	.954	.957
飽和モデル	1.000		1.000		1.000
独立モデル	.000	.000	.000	.000	.000

nested models comparison

Model Configural is correct under the assumption:

モデル	自由度	CMIN	確率	NFI Delta-1	IFI Delta-2	RFI rho-1	TLI rho2
Metric	6	5.218	.516	.002	.002	-.009	-.009
Scalar	14	11.434	.652	.005	.005	-.018	-.019

Model Metric is correct under the assumption:

モデル	自由度	CMIN	確率	NFI Delta-1	IFI Delta-2	RFI rho-1	TLI rho2
Scalar	8	6.216	.623	.003	.003	-.009	-.009

32

scalar invariance

- すべての切片に制約を入れると、適合度が悪くなってしまった時、partial scalar invarianceを検討できる (Byrne et al., 1989)
 - 一部の因子負荷量のみ制約を入れ、どの部分が群間で異なるのかを検証する
- どれぐらいの一致が見られれば partial scalar invarianceと言えるのかについて、一致した見解はないが、少なくとも1因子あたり1つ以上

33

residual invariance

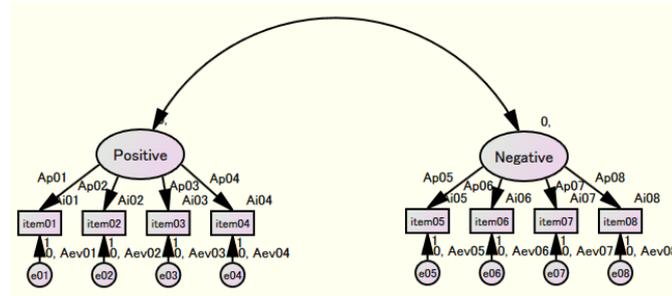
- グループ間で、各項目の誤差分散は等しいか？
 - strict factorial invarianceとも呼ばれる
- 測定不変性を検証する最終ステップ
 - 強い仮定であり、満たされていないなくても群間の比較は可能

34

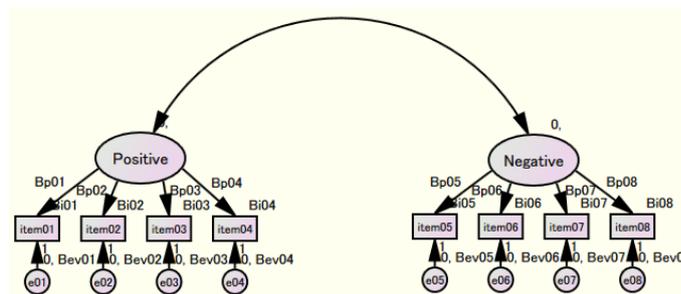
パラメーターの命名

- これまでと同様にすべての項目の誤差分散に名前をつける(日本とアメリカ両方)

日本



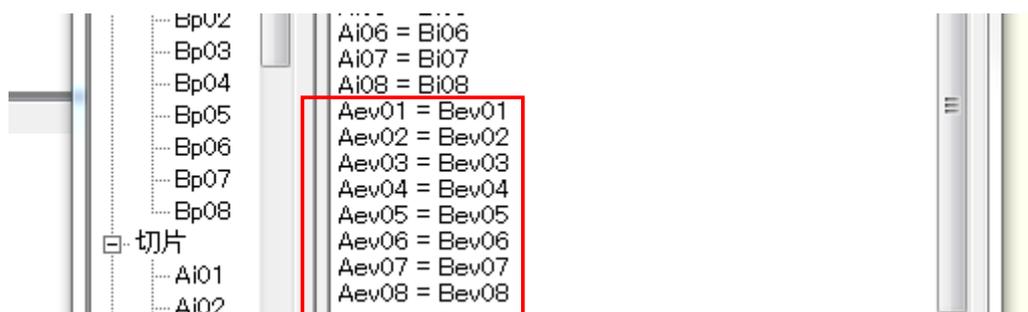
アメリカ



35

モデルの管理

- 左のXX: Scalarをクリック→新規作成
- モデル名に名前をつける→ Residual
- 誤差分散が群間で等値という制約を入れる



36

分析の実行

- 分析 (A) → 推定値を計算 (C)
- 「テキスト出力を表示」をクリック
- 適合度指標やモデル比較の結果を見て, 制約を入れたことによるモデル適合度の変化を確認

基準比較

モデル	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Configural	.940	.912	.956	.935	.956
Metric	.938	.921	.956	.944	.956
Scalar	.935	.931	.957	.954	.957
Residual	.932	.937	.957	.960	.957
飽和モデル	1.000		1.000		1.000
独立モデル	.000	.000	.000	.000	.000

ネストしたモデルの比較

モデル Configuralは正しいという仮定の下で:

モデル	自由度	CMIN	確率	NFI Delta-1	IFI Delta-2	RFI rho-1	TLI rho2
Metric	6	5.218	.516	.002	.002	-.009	-.009
Scalar	14	11.434	.652	.005	.005	-.018	-.019
Residual	22	18.712	.663	.008	.008	-.025	-.025

モデル Metricは正しいという仮定の下で:

モデル	自由度	CMIN	確率	NFI Delta-1	IFI Delta-2	RFI rho-1	TLI rho2
Scalar	8	6.216	.623	.003	.003	-.009	-.009
Residual	16	13.494	.636	.006	.006	-.016	-.016

モデル Scalarは正しいという仮定の下で:

モデル	自由度	CMIN	確率	NFI Delta-1	IFI Delta-2	RFI rho-1	TLI rho2
Residual	8	7.277	.507	.003	.003	-.006	-.006

37

Factor variance/covariance invariance

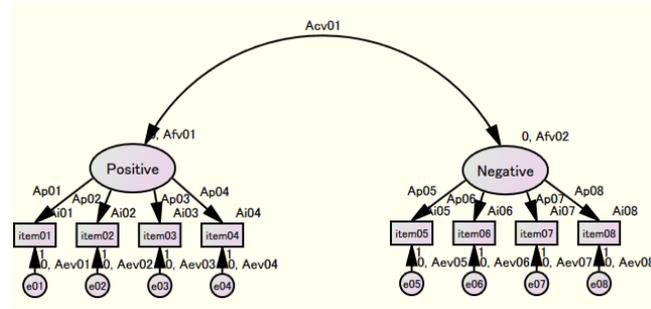
- ここからは, 構造不変性の検討
- 因子の分散, 因子間の共分散は等しいか?
 - 測定不変性の時と同様に等値制約を置くことで検討

38

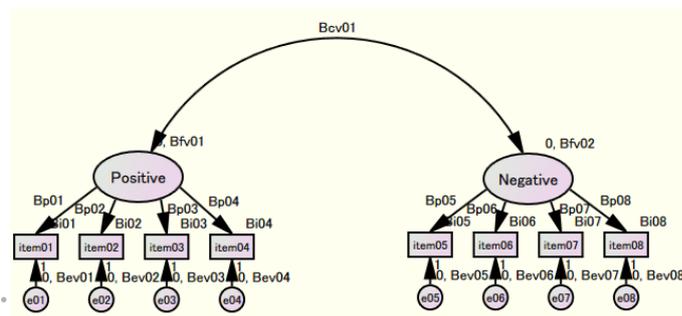
パラメーターの命名

- これまでと同様に因子の共分散と分散に名前をつける(日本とアメリカ両方)

日本



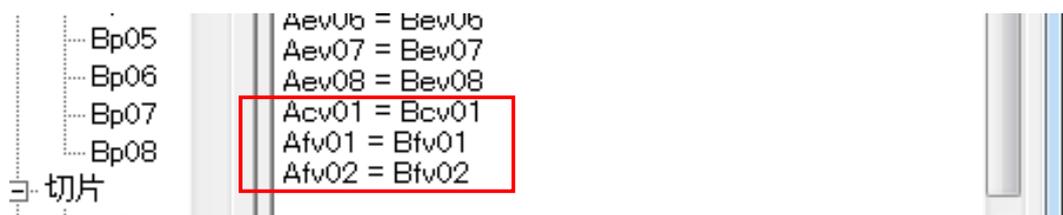
アメリカ



39

モデルの管理

- 左のXX: Residualをクリック→新規作成
- モデル名に名前をつける→ Structural
- 因子の分散, 共分散が群間で等値という制約を入れる



40

分析の実行

- 分析 (A) → 推定値を計算 (C)
- 「テキスト出力を表示」をクリック
- 適合度指標やモデル比較の結果を見て、制約を入れたことによるモデル適合度の変化を確認

基準比較

モデル	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Configural	.940	.912	.956	.935	.956
Metric	.938	.921	.956	.944	.956
Scalar	.935	.931	.957	.954	.957
Residual	.932	.937	.957	.960	.957
Structural	.930	.937	.956	.961	.956
飽和モデル	1.000		1.000		1.000
独立モデル	.000	.000	.000	.000	.000

モデル Residual は正しいという仮定の下で:

モデル	自由度	CMIN	確率	NFI Delta-1	IFI Delta-2	RFI rho-1	TLI rho2
Structural	3	6.140	.105	.003	.003	-.001	-.001

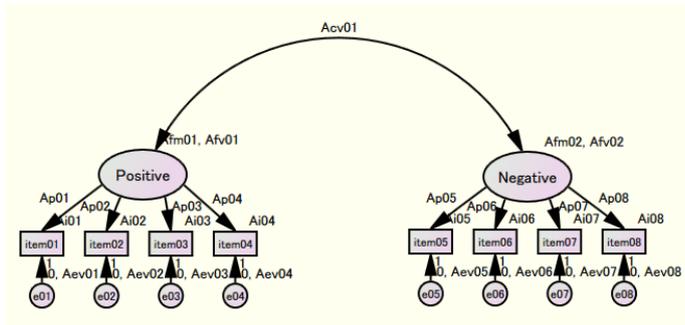
潜在変数の平均値差の分析

- さらに、潜在変数の平均値差の分析も可能
- 1つの集団の因子の平均値を固定
 - 通常は平均0に固定
 - 固定した集団を基準として、他の集団の因子平均の違いを考察する

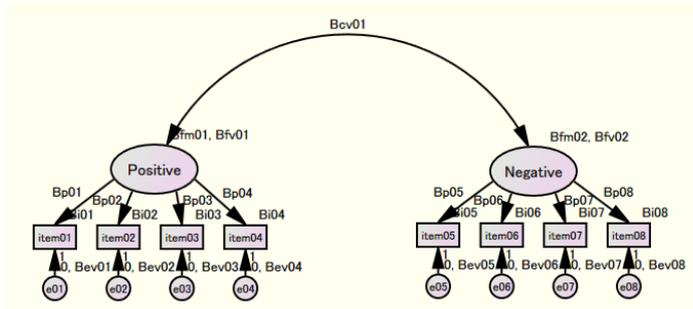
パラメーターの命名

- これまでと同様に因子平均に名前をつける

日本



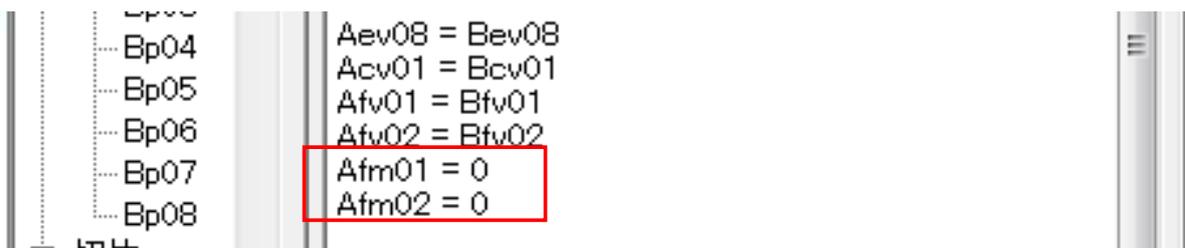
アメリカ



43

モデルの管理

- 左のXX: Structuralをクリック→新規作成
- モデル名に名前をつける→ Mean
- 1つの集団の因子平均を0に固定



44

分析の実行

- 分析 (A) → 推定値を計算 (C)
- 「テキスト出力を表示」→ 推定値をクリック
- アメリカの平均値の欄を見て、有意に日本と異なるかどうかを検討可能

平均値 : (アメリカ - Mean)

	推定値	標準誤差	検定統計量	確率	ラベル
Positive	-.053	.060	-.887	.375	Bfm01
Negative	-.010	.057	-.179	.858	Bfm02

45

まとめ

- 測定不変性や構造不変性は、いくつかのステップに分割することで、どのレベルでの不変性が担保されているのかを検討することができる
- 測定不変性は暗に仮定されがち
→ 今回紹介した手法を用いることで、実証的に確かめることが可能

46

引用文献

- Allemand, M., Steiger, A. E., & Fend, H. A. (in press). Empathy development in adolescence predicts social competencies in adulthood. *Journal of Personality*.
- Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105, 456-466.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness of fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9, 233-255.
- Ekermans, G., Saklofske, D. H., Austin, E., & Stough, C. (2011). Measurement invariance and differential item functioning of the Bar-On EQ-i: S measure over Canadian, Scottish, South African and Australian samples. *Personality and Individual Differences*, 50, 286-290.

引用文献

- Kline, R. B. (2010). *Principles and practice of structural equation modeling, Third Edition*. The Guilford Press, New York, NY
- Meade, A., Johnson, E., & Braddy, P. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, 93, 568-592.
- 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, 4-70.
- Tsaousis, I., & Kazi, S. (2013). Factorial invariance and latent mean differences of scores on trait emotional intelligence across gender and age. *Personality and Individual Differences*, 54, 169-173.