

北村メンタルヘルス学術振興財団は定期的に統計学に関する研修会を行なっております。また、北村メンタルヘルス研究所からは「臨床で働きながら研究をしよう:統計の裏わざと SPSS の使い方」という書籍を発行し、ご好評を得ています。現在、この書籍は改訂を行なっており、新たに上級コースに該当する内容を加える予定です。その一部をPDF資料として皆様に提供しているのがこの資料です。

北村メンタルヘルス研究所の出版物にご興味にある方は ⇒

<http://www.institute-of-mental-health.jp/books.html>

北村メンタルヘルス学術振興財団の研修にご興味のある方は ⇒

<http://www.kitamura-foundation.org/training.html>

北村メンタルヘルス学術振興財団主催のセミナーにご興味のある方は ⇒

<http://www.kitamura-foundation.org/training.html>

その他のお問い合わせは ⇒ info@institute-of-mental-health.jp

〒151-0063 東京都渋谷区富ヶ谷 2-26-3 リバーランドハウス富ヶ谷A棟

12

複雑モデルを検討しよう (1)

配置不変・測定不変・構造不変

1 配置不変・測定不変・構造不変：序論

1. 測定法は同じ現象を測定しているのか

私たちが精神科医療やメンタルヘルス研究を行う際、症状について評価（アセスメント）を行う。評価は専門家が客観的に行うこともあり、患者・被検者が被検者が主観的に行うこともある。評価を複数の項目で行うことも日常的である。例えば、抑うつ状態の重症度を評価する場合に 6 つの症状（抑うつ感情、興味の減退、絶望感、不眠、食欲低下、便秘）を（例えば各項目 5 件法で）測定したとしよう。これらを実験項目という。実験項目は観測変数とも呼ばれる。そして因子分析の結果、2 つの因子（factor）が認められたとしよう。第 1 の因子に高い因子負荷量（factor loading）を示した項目は、抑うつ感情、興味の減退、絶望感で、これを抑うつの心理症状と命名したとしよう。第 2 の因子に高い因子負荷量を示した項目は、不眠、食欲低下、便秘で、これを抑うつの身体症状と命名したとしよう¹。

こうした報告があると私たちは、いきなり抑うつの 2 つの下位尺度（subscale）、例えば心理症状下位尺度と身体症状下位尺度を作り、この点数を用いる傾向がある。しかし、そうした使い方をする前に検討しなければならないことがある。

上記の抑うつ症状の尺度は

- (1) 男性患者（被検者）と女性患者（被検者）で同じ内容の現象を測定しているのだろうか²? 文化の違いで同一の尺度を用いて、その得点をそのまま比較してよいのだろうか³?

¹ 抑うつ症状の因子構造の優秀な総説としては Shafer, A. (2006). Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *Journal of Clinical Psychology*, 62, 123-146.

² Santos, L., Vagos, P., & Rijo, D. (2018). Dimensionality and measurement invariance of a brief form of the Young Schema Questionnaire for adolescents. *Journal of Child and Family Studies*, 27, 2100-2111.

³ Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, research & Evaluation*, 12, 3.

(2) 同じ集団（例えばうつ病で初めて来院した患者）でも測定する時期が異なっても（例えば初診時、治療が始まって1か月目）同じ内容の現象を測定しているのだろうか⁴？

もし、この評価法（面接やアンケート）が測定しているものが複数の群の間で異なっていたり、時期によって異なっていたりすれば、臨床や研究は成り立たなくなる。群間の比較や時期ごとの比較は無意味になるからだ。測定値の差が、対象とする概念（抑うつ心理症状など）の程度差を示すものではなくなるのである。項目得点の合算を「うつ病の重症度」とみなすことの合理的根拠がなくなってしまう。極端な例を挙げよう。「夜に尿を漏らす」という評価項目があったとしよう。被検者が3～4歳の場合と、70歳以上の場合で、測定している概念は当然に異なる。親の子どもに対する態度を測定する尺度があったとして（Parental Bonding Instrument など）、男親に使用した結果と女親に使用した結果を、単純に数値比較してよいのであろうか。複数項目による評価が異なる集団や異なる時点でも同じ内容の概念を評価していることを（広義に）測定不変性（measurement invariance, measurement equivalence）と呼ぶ。

こうした問題は次のようないくつかのステップを踏んで、検証される。重要なのは、この順に検討しなければいけないということである⁵。下記のステップのそれぞれで群間の不変性が却下された場合、次のステップを検討する意味はない。例えば、配置不変性と weak factorial invariance (metric invariance) が承認できたが、strong factorial invariance (scalar invariance) が承認できなかった、そのまま strict factorial invariance (residual invariance) の検討に進んではいけない。この場合、strong factorial invariance (metric invariance) を阻害している部分を確認し、その部分の改変を行なって、strong factorial invariance (metric invariance) が承認できるモデルを作ってから、次に進むことになる。なお、この領域の名称は必ずしも一致していないので、注意が必要である（表 12-1）。なお、factor covariance invariance は、因子が複数存在するときのみ計算される。

表 12-1. 測定不変の分類

測定不変	配置不変 configural invariance	配置不変 configural invariance	因子数と各因子の観測変数が群間で同じ
	測定不変 measurement invariance	weak factorial (metric) invariance	(上記に加えて) 各観測変数の因子負荷量が群間で同じ
		strong factorial (scalar) invariance	(上記に加えて) 各観測変数の切片(平均)が群間で同じ
構造不変	構造不変 structural invariance	strict factorial (residual) invariance	(上記に加えて) 各観測変数の誤差変数の分散が群間で同じ
		factor variance invariance	(上記に加えて) 因子の分散が群間で同じ
		factor covariance invariance	(上記に加えて) [複数の] 因子間の共分散が群間で同じ
		factor mean invariance	(上記に加えて) 因子の平均値が群間で同じ

2. 配置不変 (configural invariance)

⁴ Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factor invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, 4(1), 10-18.

⁵ 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 Method*, 3(1), 4-70.

さきほどの抑うつ症状の評価法は 6 項目から出来上がっている。これを、初診の男性患者と女性患者それぞれ 100 人に施行し、男女をわけて因子分析を行ったところ、いずれも 2 因子構造であることが分かった。これが配置不変である。しかし、もし男性では抑うつ感情と興味の減退が 1 因子を構成し、残りの 4 症状が別の因子を構成したとしよう。一方、女性患者は 3 項目ずつの 2 因子構造であった。これは配置不変とはいえない。そうであれば、この尺度を用いて抑うつ症状の男女差を比較することは不適切になる。男性用尺度と女性用尺度と考えるべきであり、両者を分けて使用することになる。

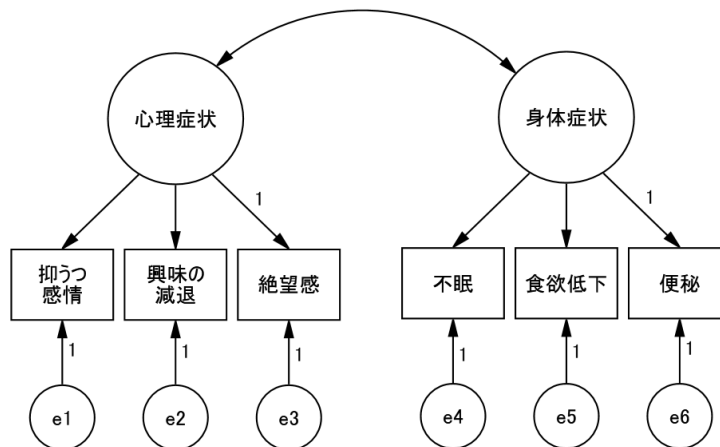


図 12-1. 抑うつ症状の仮定の因子構造

また、ひとつの集団での探索的因子分析で得られた因子構造を、もうひとつの集団に適応した確認的因子構造の適合度が悪い場合、測定不変性は否定される。この場合、下位尺度得点の絶対値をそのまま比較することはできない⁶。測定不変性が確認できない以上、次に述べるステップに進むこともできない。

男性

⁶ van de Vijver, F. J. R., & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross-cultural Psychology*, 31(1), 33-51.

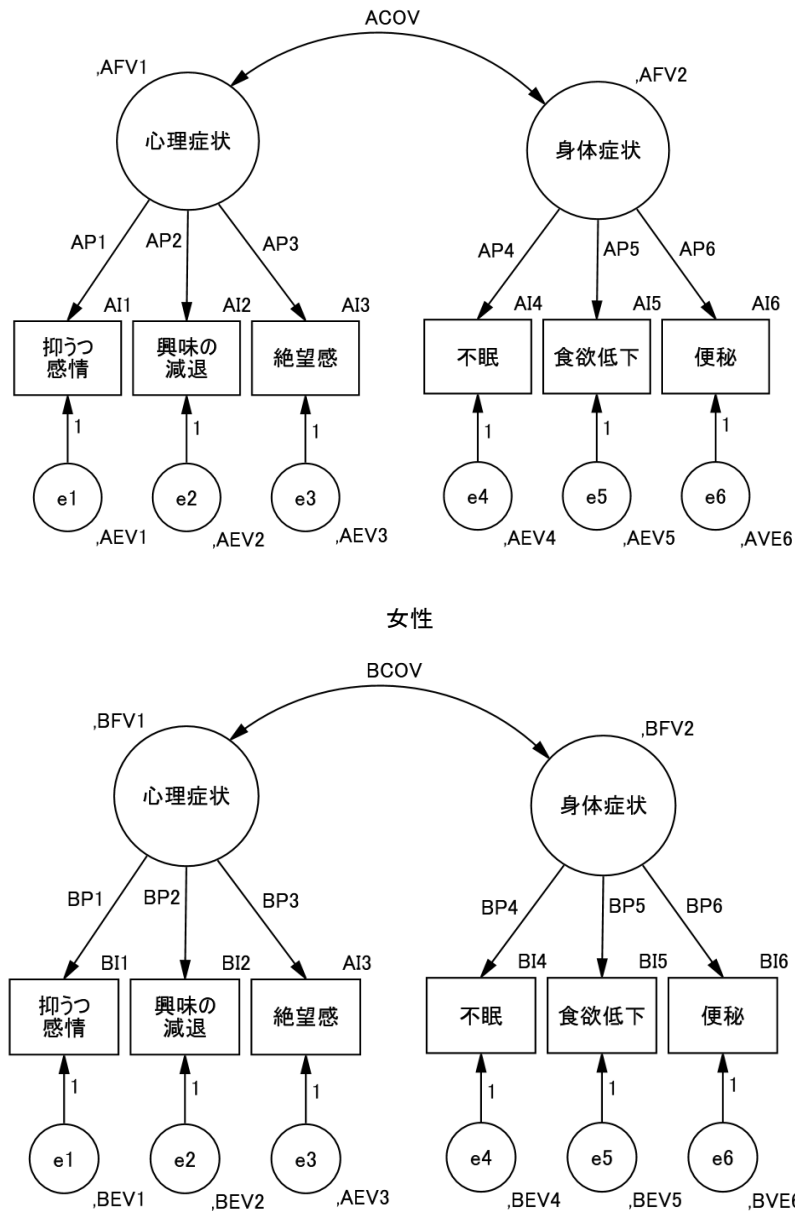


図 12-2. 男性と女性における抑うつ症状の仮定の因子構造

配置不変性を確認するには、図 12-2 のような男性と女性という 2 つに集団について、以下の制約のみを課して、同時に解析を行なう（多母集団同時解析）。

$$AP3 = BP3 = 1$$

$$AP6 = BP6 = 1$$

配置不変が確認されたら、つぎに測定不変の検討に進むことになる。ここでは、各パスが群間で同じ値であると仮定しても、全体の適合度が劣化しないことを確認してゆく。たとえば、さきほどの抑うつ症状の評価法の 6 項目の確認的因子分析を男性と女性のデータに分けた上で、一括して適合度の計算を行う。この際、（1）潜在因子から観測変数に走っているパス係数が、男性と女性で同値であると仮定する、（2）さらに各観測変数の切片（平均）が、男性と女性で同値であると仮定する、（3）さらに各観測変数の誤差変数の分散が、男性と女性で同値であると仮定するといったように、制約をだんだんきつくしてゆく。この制約は固定母数ではない。しかし、あるパス係数が男性と女性で同じ値を示すという条件で解を求めるのであるから、適合度は悪い方向に行く。

これが有意に悪くならなければ、儉約律にかなうと考えて、制約の強いほうを採用するというのが、基本的思考方法である。

3. weak factorial invariance (metric invariance)

上記の抑うつ症状の評価法の 6 項目の確認的因子分析について、潜在因子から観測変数に走っているパス係数が、男性と女性で同値であると仮定し、その適合度が配置不変モデルの適合度の適合度より 有意に悪くなければ こちらを「正しい」として採用する。もし、有意に悪くなるなら、配置不変モデルが正しいモデルである。測定不変モデルは weak factorial invariance (metric invariance) モデルとも呼ばれている。

測定不変性を確認する最初のステップでは、各観測変数（抑うつ感情、興味の減退、・・・）の因子負荷量が群と群を跨いで（測定時期を跨いで）同一であることを確認する。つまり以下の制約を掛けても適合度がよい（悪くならない）ことを確認するのである。その他のパス係数には制約を設定しない。

$$AP1 = BP1$$

$$AP2 = BP2$$

$$AP3 = BP3 = 1$$

$$AP4 = BP4$$

$$AP5 = BP5$$

$$AP6 = BP6 = 1$$

仮に weak factorial invariance (metric invariance) が却下されたとすると、どこかの因子負荷量が両群間で大きく異なっていることが原因と考えられる。そこで、この原因たる因子負荷量のパスがどれかを確認し、その観測変数を削除した上で、同じ計算をすると weak factorial invariance (metric invariance) が確認できるかもしれない。特定の因子負荷量を両群間で比較し、その z 値が最も有意に大きいものが、不変性を否定する原因の因子負荷量（とその観測変数）であると考えられる。また、それでも不変性が却下されれば、同じ作業を繰り返す。

4. Strong factorial invariance (Scalar invariance)

さらに、strong factorial invariance (scalar invariance) では各観測変数の切片 (intercept) つまり平均が群と群を跨いで（測定時期を跨いで）同一であることを確認する。つまり以下の制約を掛けても適合度がよいことを確認する。

$$AP1 = BP1$$

$$AP2 = BP2$$

$$AP3 = BP3 = 1$$

$$AP4 = BP4$$

$$AP5 = BP5$$

$$AP6 = BP6 = 1$$

$$AI1 = BI1$$

$$AI2 = BI2$$

$$AI3 = BI3$$

$$AI4 = BI4$$

$$AI5 = BI5$$

$$AI6 = BI6$$

測定不変モデルに比較すると、制約がずっと増えている。測定不変モデルの適合度より strong factorial invariance モデルの適合度は悪いが、これが有意に悪くなければこちらを「正しい」として採用する。もし、有意に悪くなるなら、weak factorial invariance が正しいモデルと判断し、それ以上の検討は行わない。しかし、モデルの改良を考えるのであれば、特定の観測変数の切片の値を両群間で比較し、その z 値が最も有意に大きいものが、不変性を否定する原因の切片（とその観測変数）であると考えられる。そして、その観測変数を削除したモデルを作り、再度、計算を行なう。また、それでも不変性が却下されれば、同じ作業を繰り返す。

5. Strict factorial invariance (Residual invariance)

最後に、各観測変数の誤差変数（因子分析でいえば独自因子）の分散（variance）が群と群を跨いで（測定時期を跨いで）同一であることを確認する。以下の制約を掛けても適合度がよいことを確認する。

AP1 = BP1
AP2 = BP2
AP3 = BP3 = 1
AP4 = BP4
AP5 = BP5
AP6 = BP6 = 1
A11 = B11
A12 = B12
A13 = B13
A14 = B14
A15 = B15
A16 = B16
AEV1 = BEV1
AEV2 = BEV2
AEV3 = BEV3
AEV4 = BEV4
AEV5 = BEV5
AEV6 = BEV6

モデルの改良を考えるのであれば、特定の観測変数の誤差分散を両群間で比較し、その z 値が最も有意に大きいものが、不変性を否定する原因の分散（とその観測変数）であると考えられる。そして、その観測変数を削除したモデルを作り、再度、計算を行なう。また、それでも不変性が却下されれば、同じ作業を繰り返す。

Weak factorial invariance から strong factorial invariance に行き、さらに strict factorial invariance に合致するほど、測定不変性が強固だと考えられる。しかし、実際の研究では strict factorial invariance までは要求されないであろう。Strong factorial invariance が満たされれば十分と考えられる。

6. 構造不変性 (structural invariance)

ここまでは、共分散構造分析の用語を用いれば測定モデルに関する群間の不変性を問題にしていた。そこで、weak factorial invariance (metric invariance)、strong factorial invariance (scalar invariance)、strict factorial invariance をまとめて測定不変とよぶこともあるようだ。また、豊田らは配置不変性と測定不変性を別項目としているが、研究者によって配置不変性をもっとも制約の少ない測定不変性として紹介しているものもある（表

12-1)。測定不変性が確認できたら次に、構造モデルの群間不変性を検討することになる。これが構造不変性である。

まず、factor variance の不変性を検討する。これは strict factorial invariance の制約に加えて以下の制約を与えても、適合度が有意に悪化しないことを照明することで確認できる。

$$\begin{aligned} AFV1 &= BFV1 \\ AFV2 &= BFV2 \end{aligned}$$

次の factor covariance の不変性はさらに以下の制約を与えても、適合度が有意に悪化しないことを照明することで確認できる。

$$ACOV = BCOV$$

最後の factor mean の不変性は、任意の一方グループ（例えば男性）の 2 つの因子の平均値を 0 に固定し、もう一方のグループの因子の平均値に任意の変数名を与える。本来、因子の平均値の絶対値は求められず相対的なものである。そこで、人為的に一方グループの因子平均をゼロとし、それを対照値としてもう一方グループのそれが大きく異なっているかを確認するのである。ある因子の平均値が両群間で有意の差を示すかを求める。

精神科臨床と研究において実証的症候評価はその基礎になっている。しかし、同じ項目を測定したからといって、母集団が異なれば同じ概念を測定している保証はない。その際に、配置不変・測定不変・構造不変を確認する作業は不可欠なのである。

7. 測定不変性における適合度

配置不変は各群で探索的因子分析を行う。測定不変・構造不変は Amos を用いて多母集団同時解析 (multigroup analysis) を実行し、確認的因子分析を行う。制約が強くなるにしたがって適合度は悪くなるが、これが有意に悪くならない限り制約を許容する。この時は、 χ^2 値の変化量を見る⁷ (likelihood ratio test, chi-squared difference test)。しかし、 χ^2 値の変化量の有意水準はサンプルの大きさ (検体の数) に依存している。サンプルが多数であるほど、 χ^2 値のモデル間の差 (変化量) が有意に傾く⁸。そのため、測定不変性が (おそらく過剰に) 強く否定されることになる。そこで、サンプルの大きさに影響されない、他の適合度指標 (alternative goodness-of-fit index: AGFI) を使うべきとの意見がある。例えば weak factorial invariance のモデルと strong factorial invariance のモデルに移行しても、CFI が 0.01 以上悪化しないことをもってその制約を許容し、strong factorial invariance が成立すると考えるのである。また、RMSEA の変化の程度が 0.01 未満であれば、不変性が確認できるとする意見もある。この章では、 χ^2 値、CFI、RMSEA の変化量を評価の指標とする。他にもいくつかの案が示されている (表 12-2)。

表 12-2. 測定不変性の評価の際の適合度指標

	ΔCFI	$\Delta RMSEA$
--	--------------	----------------

⁷ Meade, A. W., Johnson, E. C., & Braddy, P. W. (2006). The utility of alternative fit indices in tests of measurement invariance. Paper presented at the annual Academy of Management conference. Atlanta, GA.

⁸ Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, 39(3), 479-515.

Cheung, & Rensvold ⁹	< 0.01	—
Cheng ¹⁰ N ≤ 300	< 0.005	< 0.01
Cheng N > 300	< 0.01	< 0.015

⁹ Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.

¹⁰ Cheng, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464-504.

2 解析の実際

1. データの準備¹¹

Hospital Anxiety and Depression Scale (HADS) の因子構造について、年齢のグループで配置・測定・構造不変性を検討してみよう。まず、サンプルデータの W1Q10（被験者の年齢）という変数の値の再割付を行なって、年齢の上下で分ける新しい変数 AGEGROUP を作る。つまり 30 歳未満の妊婦（AGEGROUP = 1）と 30 歳以上の妊婦（AGEGROUP = 2）で、HADS の測定不変性および構造不変性が確認できるかを検討する。

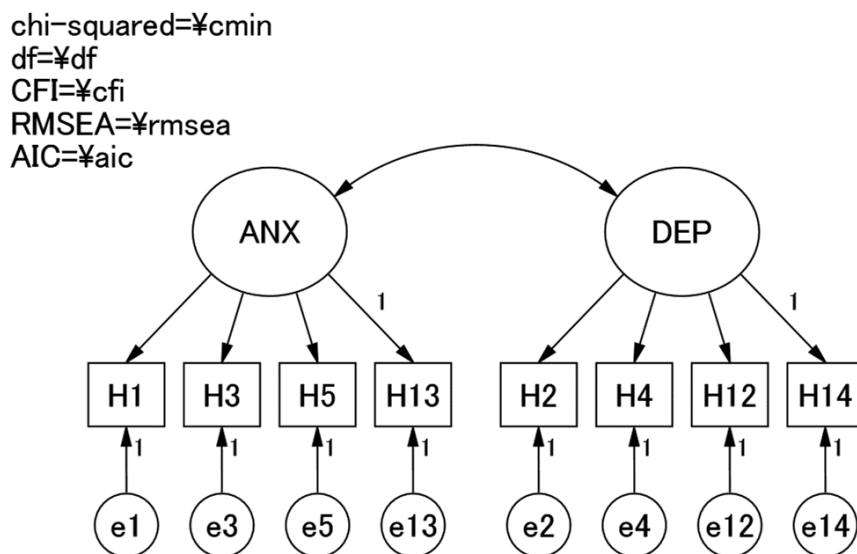
```

FREQUENCIES  W1Q10.
RECODE       W1Q10(LOW THRU 29=1)(30 THRU HI=2)INTO AGEGROUP.
FREQUENCIES  AGEGROUP.

```

1

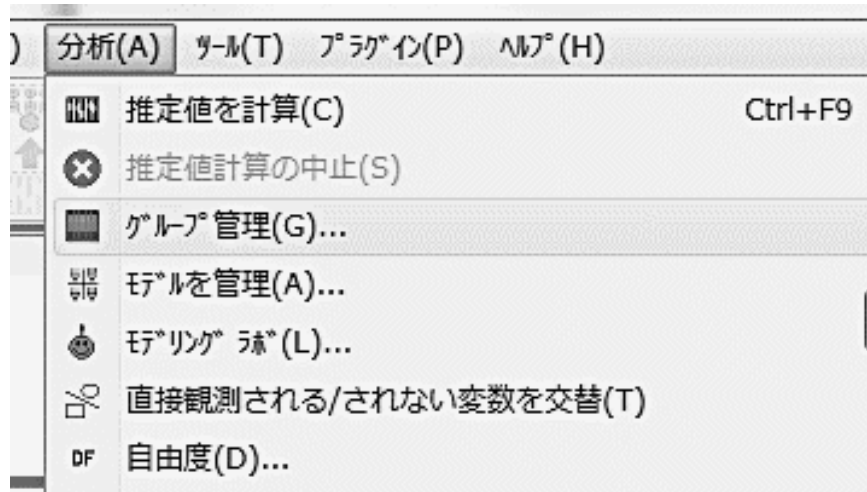
ここでは計算実例を簡略にするために、ANXIETY (ANX) の 4 項目 (H1, H3, H5, H13)、DEPRESSION (DEP) の 4 項目 (H2, H4, H12, H14) で尺度が構成されているとして計算する。まず、以下のような通常の確認的因子分析のモデルを AMOS 上に描きあげる。解を求められるだけの自由度を確保するために、固定母数を 2 つ設定する。各因子に属する観測変数のうち任意のひとつへの因子負荷量を “1” に設定する。



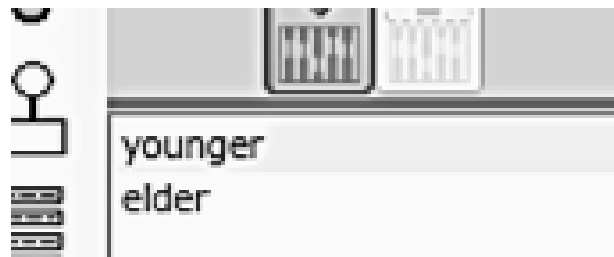
次に「分析 (A)」のプルダウンメニューから「グループ管理 (A)」を選ぶ。

¹¹ AMOS 等を利用した測定・構造不変性の実務の解説として

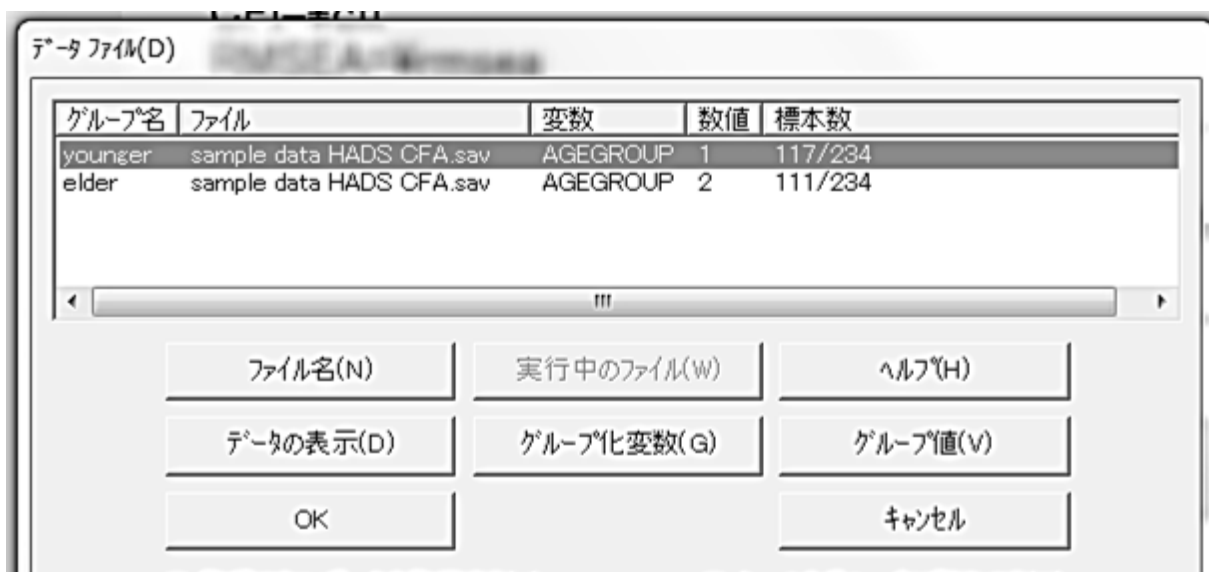
Yu, L., & Shek, D. T. L. (2014). Testing factorial invariance across groups: An illustration using AMOS. *International Journal of Disability and Human Development*, 13(2), 205-216.



ここで、younger と elder という 2 つのグループ名を書き込む。



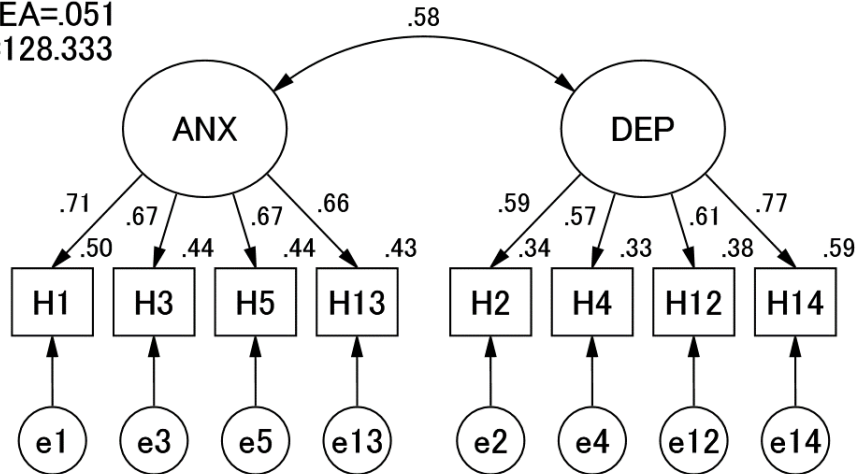
次にデータを AMOS に取り込む。「データファイルを選択」のアイコンを開き、次のようなダイアログボックスから、まず「ファイル名 (N)」をクリックする。そして対象とするデータファイルを選択し、「開く (O)」をクリックする。次に「グループ化変数 (G)」をクリックし、ここから先ほど作成した AGEGROUP という変数を選ぶ。そして「グループ値 (V)」のアイコンをクリックし、ここから “1” を選ぶ。さらに、再び「ファイル名 (N)」⇒ AGEGROUP ⇒ 「開く (O)」⇒ 「グループ値 (V)」⇒ “2” を選択する。結果は、「データファイル (D)」のダイアログボックスに以下のように表示される。



2. Configural invariance の判定

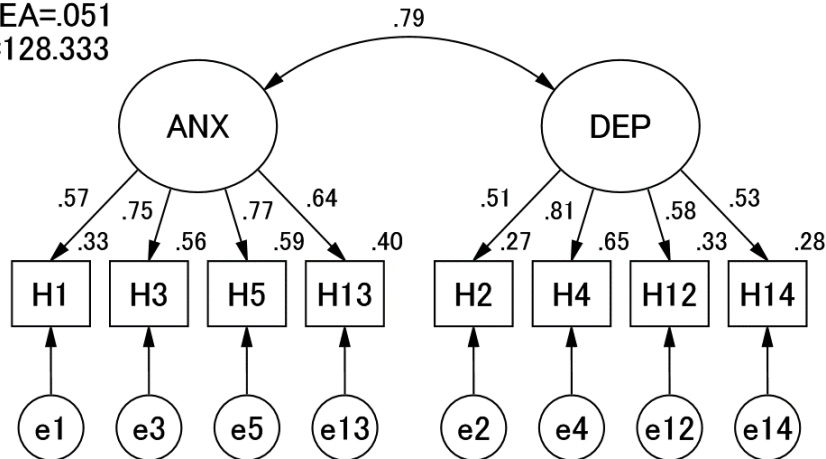
まず configural invariance を確認する。今回の HADS のデータが、30 歳未満の妊婦と 30 歳以上の妊婦で同一の因子数、同一の観測変数がそれぞれの因子にぶら下がっていると推定しても、データに適合しているかを確認するのである。先ほどの AMOS を実行する。すると次の結果（標準化解）が現れる。

chi-squared=60.333
df=38
CFI=.953
RMSEA=.051
AIC=128.333



AGEGROUP = Younger

chi-squared=60.333
df=38
CFI=.953
RMSEA=.051
AIC=128.333



AGEGROUP = Elder

次にこのモデルの適合度を見る。適合度はグラフにも表示するようにしてある。30 歳未満の妊婦の図と 30 歳以上の妊婦の図で、適合度が同じであることに注目しよう。また、「テキスト出力の表示」から「モデル適合」を選

ぶと、以下のように、 $\chi^2/df=1.588$, CFI=0.953, RMSEA=0.051 であり¹² (当然に、AMOS の描画部分の表示と同一)、このモデルの適合度は比較的良好なものであった (特に CFI が 0.95 を越えている)。従って、configural invariance は容認できる。

CMIN

モデル	NPAR	CMIN	自由度	確率	CMIN/DF
モデル番号 1	34	60.333	38	.012	1.588
飽和モデル	72	.000	0		
独立モデル	16	527.286	56	.000	9.416

基準比較

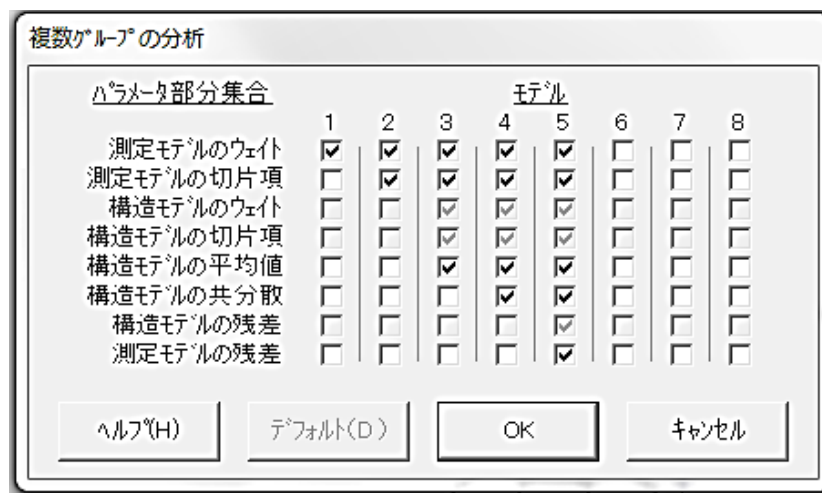
モデル	NFI	RFI	IFI	TLI	CFI
	Delta1	rho1	Delta2	rho2	
モデル番号 1	.886	.831	.954	.930	.953
飽和モデル	1.000		1.000		1.000
独立モデル	.000	.000	.000	.000	.000

RMSEA

モデル	RMSEA	LO 90	HI 90	PCLOSE
モデル番号 1	.051	.024	.074	.447
独立モデル	.193	.178	.208	.000

3 Weak factorial (metric) invariance の判定

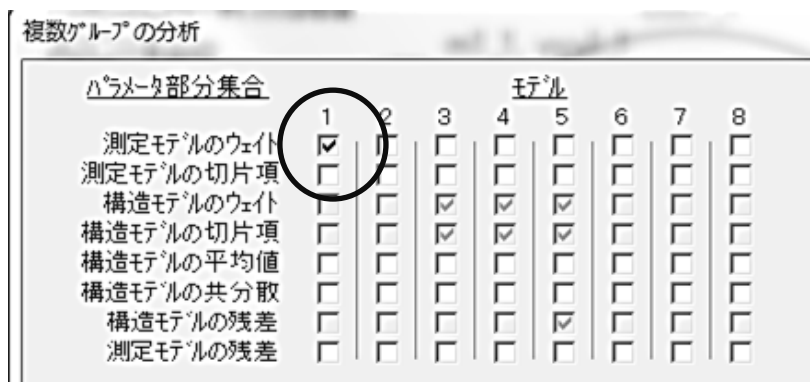
Configural invariance は確認できた。そこで、weak factorial (metric) invariance を確認する。「複数グループの分析」のアイコンをクリックすると、次のダイアログボックスが現れる。



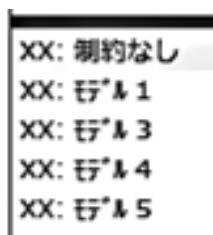
¹² Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, research & Evaluation, 12*(3), 1-25

「複数グループの分析」のダイアログボックスには、すでに default menu として 5 つのモデルが設定されている。左から右の順で、モデル 1 からモデル 5 まで並んでいる。例えば、モデル 3 は、「測定モデルのウェイト」、「測定モデルの切片項」、「構造モデルの平均値」の 3 つに制約がかけられモデルである。「測定モデルのウェイト」は今回の場合、因子負荷量を意味し、「測定モデルの切片項」は観測変数の切片を意味し、「構造モデルの平均値」は因子（潜在変数）の平均値を意味する。

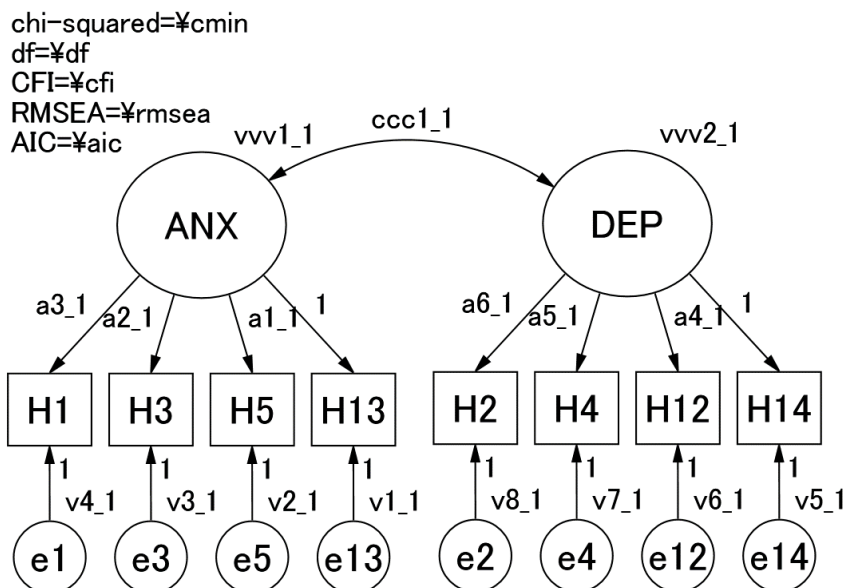
しかし、別に述べる理由から、これらモデルをにわかには採用しない。まず、第 2 モデルから第 5 モデルまでの下にある濃いティックをすべてクリックして外す。このことで「測定モデルのウェイト」（因子負荷量）のみ両群間で制約を課する第 1 モデルのみ実行することになる。

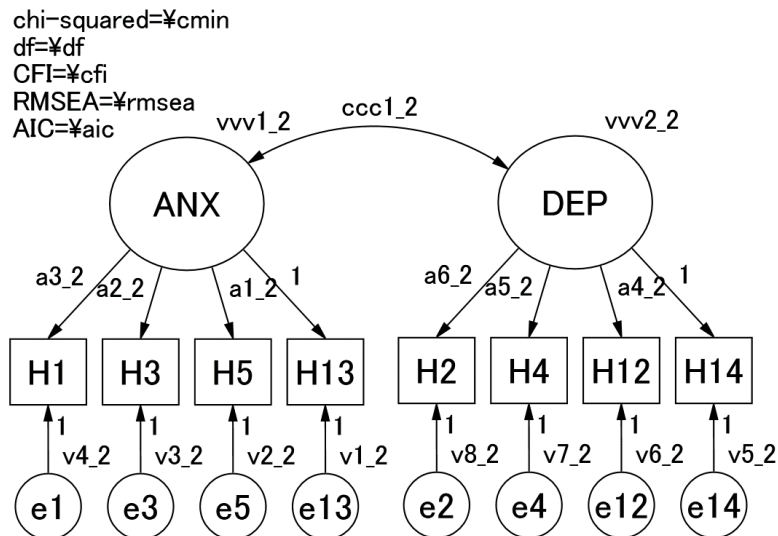


そうすると、AMOS 画面の情報領域のグループの真下に次のようなモデル情報が現れる。今回は 1 つのモデルのみ作ってあるので、モデル 3、モデル 4、モデル 5 は内容が空のモデルである。



グラフを見ると各母数に以下のような名称が自動的につけられている。

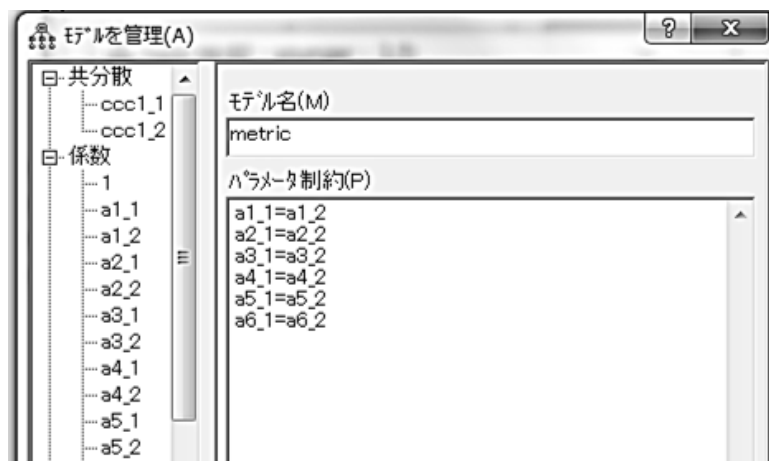




次に、情報領域のモデルを表示してある部分から「制約なし」をクリックし、「モデルを管理 (A)」というダイアログボックスを表示する。ここで「モデル名 (M)」を「制約なし」から configural と書き直す。パラメータ制約はなにもない。「閉じる」をクリックするとこのダイアログボックスが消える。



次に、「Model 1」を扱う。すでに自動的に「パラメータ制約 (P)」が記載されている。各因子負荷量が両群で同じ値であるというものである。例えば、ANX という因子から H% に走っている因子負荷量は 30 歳未満の妊婦では a1_1 , 30 歳以上の妊婦では a1_2 と名前がついていて、この値は同地であるという制約が a1_1 = a1_2 という式で示されている。さらに、「モデル名 (M)」を metric と書き直す。



これで計算を実行する。結果、以下に示すように metric モデルの適合度は $\chi^2/df=1.704$, CFI=0.935, RMSEA=0.056 であった。

CMIN

モデル	NPAR	CMIN	自由度	確率	CMIN/DF
configural	50	60.333	38	.012	1.588
metric	44	74.963	44	.002	1.704
モデル 3	50	60.333	38	.012	1.588
モデル 4	50	60.333	38	.012	1.588
モデル 5	50	60.333	38	.012	1.588
飽和モデル	88	.000	0		
独立モデル	32	527.286	56	.000	9.416

基準比較

モデル	NFI	RFI	IFI	TLI	CFI
	Delta1	rho1	Delta2	rho2	
configural	.886	.831	.954	.930	.953
metric	.858	.819	.936	.916	.934
モデル 3	.886	.831	.954	.930	.953
モデル 4	.886	.831	.954	.930	.953
モデル 5	.886	.831	.954	.930	.953
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
configural	.051	.024	.074	.447
metric	.056	.033	.077	.310
モデル 3	.051	.024	.074	.447
モデル 4	.051	.024	.074	.447
モデル 5	.051	.024	.074	.447
独立モデル	.193	.178	.208	.000

問題は, configural モデルに比較して metric モデルの適合が有意に悪化していなければ metric モデルを採用し, 有意に悪化しているなら metric モデルは却下するということである。「テキスト出力の表示」のアイコンをクリックし, 「Amos 出力」のダイアログボックスを開ける。ここで「モデル比較」をクリックし, 「モデル configural は正しいという仮定の下で:」という部分に注目する。すると, モデル metric はモデル configural に比較して, 自由度が 6 (= 44 - 38) 増加し, χ^2 値は 14.630 (= 74.963 - 60.3333) 増加している。自由度 6 の臨界値からす

ると、この増加は $p < 0.023$ で有意であった。加えて CFI の低下も 0.019 と 0.01 を越えていた。従って、metric モデルは却下される。

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	14.630	.023	.028	.030	.012	.014

2 つのモデル (configural と metric) の適合度は有意に異なり、metric モデルは却下された。そこで、metric モデルの不変性を阻止している観測項目を発見する作業を行なう。おそらくいずれかの因子負荷量 (因子から観測変数へのパス) の値を同値として制約したことが原因であろう。そうであれば、configural モデルにおいてその因子負荷量は両群間で著しく異なっていたはずである。

そこで、configural モデルに戻る。「Amos 出力」からダイアログボックスの「パラメータの一対比較」のうち「パラメータ間の差に対する検定統計量」を選択し、以下の情報を入力する。これは今回の計算で使用したすべての変数のうち 2 変数の組み合わせを作り、両者の値の差 (critical ratio) が有意であるかを z 値で見るものである。もしこの値が 1.96 を越えれば $p < 0.05$ で有意である。

Amos出力

HADS MI 02.amw

パラメータ間の差に対する検定統計量 (configural)

	a1_1	a2_1	a3_1	a4_1	a5_1	a6_1	ccc1_1	vvv1_1	v1_1	v2_1
a1_1	.000									
a2_1	-.051	.000								
a3_1	.331	.358	.000							
a4_1	-1.678	-1.730	-1.974	.000						
a5_1	-2.549	-2.689	-2.977	-1.154	.000					
a6_1	-1.087	-1.112	-1.354	.881	1.834	.000				
ccc1_1	-3.928	-4.234	-4.447	-3.175	-2.975	-3.645	.000			
vvv1_1	-3.133	-3.361	-3.578	-2.597	-2.145	-3.143	1.297	.000		
v1_1	-3.796	-3.970	-4.279	-2.211	-1.571	-2.816	2.011	.783	.000	

この部分を詳しく見てみる。

Amos出力

HADS MI 02.amw

パラメータ間の差に対する検定統計量 (configural)

	a1_1	a2_1	a3_1	a4_1
a1_1	.000			
a2_1	-.051	.000		
a3_1	.331	.358	.000	
a4_1	-1.678	-1.730	-1.974	.000
a5_1	-2.549	-2.689	-2.977	-1.154
a6_1	-1.087	-1.112	-1.354	.881

12 うつ病の原因を人生史から探る — 非逐次モデルと状態・特性モデル

この行列表は極端に見づらい。そこで、これをコピーアンドペーストにて EXCEL に移動する。このほうが、表が見やすい。さらに EXCEL の「表示」から「ウインド枠の固定」を利用してみたい部分を確認する。



因子負荷量が両群間でどのくらい異なっているかの部分に注目する。

	A	B	C	D	E	F	G	H	I
1	ハ°ラメータ間の差に対する検定統計量 (configural)								
2									
3		a1_1	a2_1	a3_1	a4_1	a5_1	a6_1	ccc1_1	vvv1_1
28	i8_1	-0.513	-0.507	-0.844	1.885	3.935	0.947	8.891	6
29	a1_2	0.23	0.277	0.015	1.965	2.929	1.358	4.689	.
30	a2_2	0.302	0.355	0.082	2.136	3.191	1.491	5.109	.
31	a3_2	-1.228	-1.26	-1.533	0.441	1.417	-0.19	3.531	2
32	a4_2	-0.26	-0.237	-0.471	1.203	1.95	0.691	3.372	2
33	a5_2	0.042	0.078	-0.17	1.638	2.486	1.081	4.06	3
34	a6_2	0.777	0.821	0.633	1.957	2.504	1.566	3.488	.

1.94 を越えた値が、a5_1 と a5_2 の間の際にみられた (2.486 > 1.94)。これは、HADS の第 4 項目への因子付加量であった。そこで、この項目を削除したモデルを作る。再度、同じことを行なう。この際、「複数グループの分析」のアイコンをクリックする作業が必要になる。

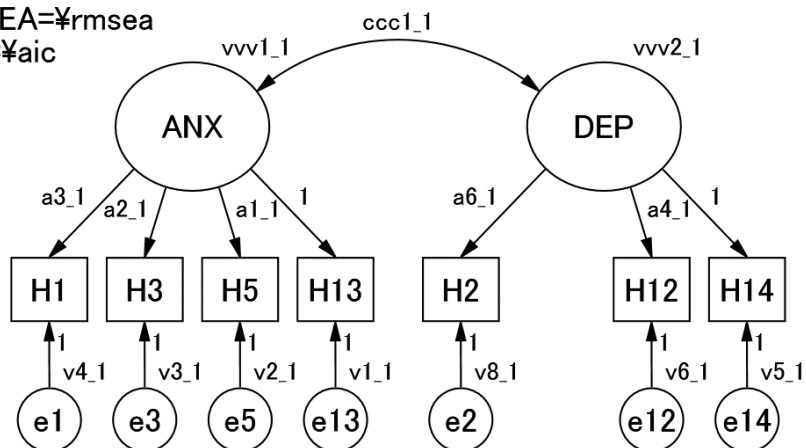
chi-squared=¥cmin

df=¥df

CFI=¥cfi

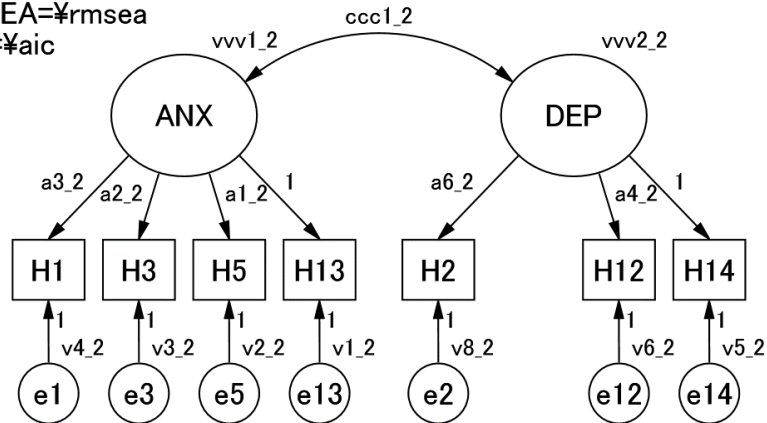
RMSEA=¥rmsea

AIC=¥aic



AGEGROUP = younger

chi-squared= χ^2 cmin
df= χ^2 df
CFI= χ^2 cfi
RMSEA= χ^2 rmsea
AIC= χ^2 aic



AGEGROUP = elder

この新モデルで先ほどと同様の計算を行なう。今回は、モデル metric はモデル configural に比較して、自由度が 5 (=30 - 25) 増加し、 χ^2 値は 6.438 (=48.575 - 42.136) 増加している。自由度 5 の臨界値からすると、この増加は $p=0.266$ で有意ではない。また CFI の減少も 0.004 (=0.957 - 0.953) でわずかなものであった。RMSEA は増加せずむしろ低下していた (0.052 が 0.050 に)。従って、metric モデルは容認されたのである。観測変数がひとつ削除され、モデルが変更されているので、念のため configural モデルそのものの適合度を再確認する。 $\chi^2/df=1.621$, CFI=0.957, RMSEA=0.052 であった。当初の 8 項目 HADS における configural モデルの適合度より多少よくなっている。

CMIN

モデル	NPAR	CMIN	自由度	確率	CMIN/DF
configural	30	42.136	26	.024	1.621
metric	25	48.575	31	.023	1.567
モデル 2	30	42.136	26	.024	1.621
モデル 3	30	42.136	26	.024	1.621
飽和モデル	56	.000	0		
独立モデル	14	415.446	42	.000	9.892

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	IFI	RFI	TLI
				Delta-1	Delta-2	rho-1	rho2
metric	5	6.438	.266	.015	.017	-.005	-.006

基準比較

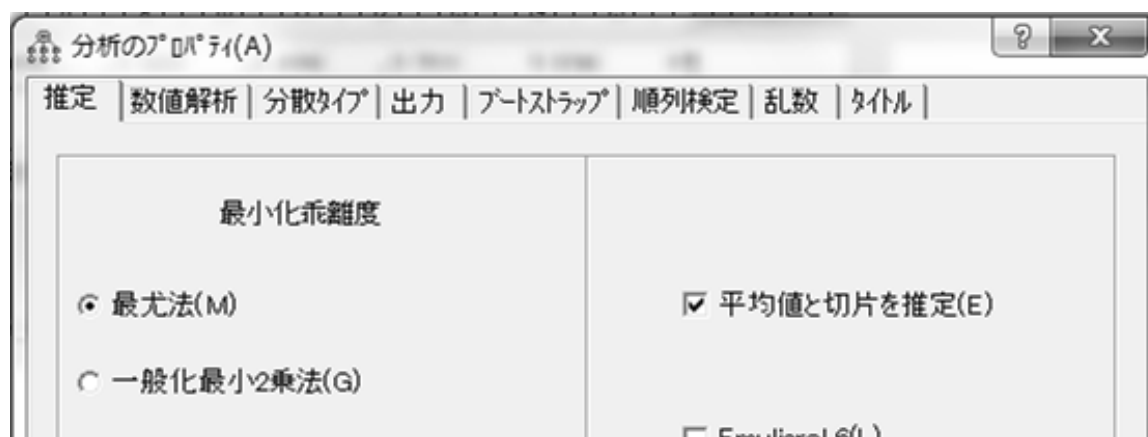
モデル	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
configural	.899	.836	.959	.930	.957
metric	.883	.842	.954	.936	.953
モデル 2	.899	.836	.959	.930	.957
モデル 3	.899	.836	.959	.930	.957
飽和モデル	1.000		1.000		1.000
独立モデル	.000	.000	.000	.000	.000

RMSEA

モデル	RMSEA	LO 90	HI 90	PCLOSE
configural	.052	.019	.080	.415
metric	.050	.019	.076	.469
モデル 2	.052	.019	.080	.415
モデル 3	.052	.019	.080	.415
独立モデル	.198	.181	.216	.000

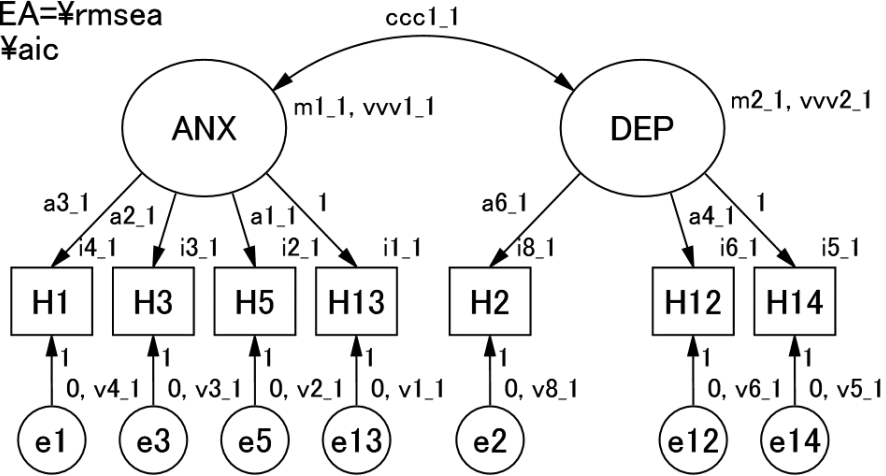
4 Strong factorial (scalar) invariance の判定

Metric モデルは今回は許容された。では scalar モデルの不変性を検討する。ここではじめて、観測変数に切片 (intercept) の計算が必要になる。そこで「分析のプロパティ (A)」のダイアログボックスの「推定」の頁を選び、ここで「平均値と切片を推定 (E)」に印をいれる、



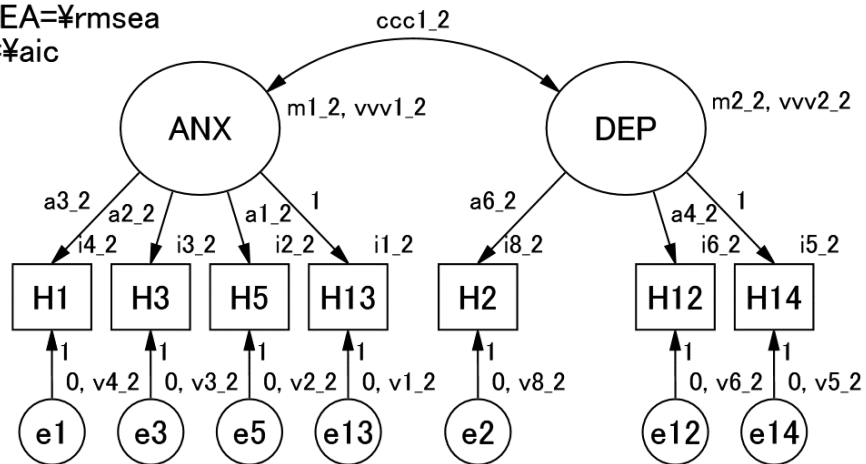
すると観測変数の切片に自動的に変数名が与えられる。新しく、各観測変数の切片の変数名として “i” で始まる変数名が宛てられている。

chi-squared= χ^2 cmin
df= χ^2 df
CFI= χ^2 cfi
RMSEA= χ^2 rmsea
AIC= χ^2 aic



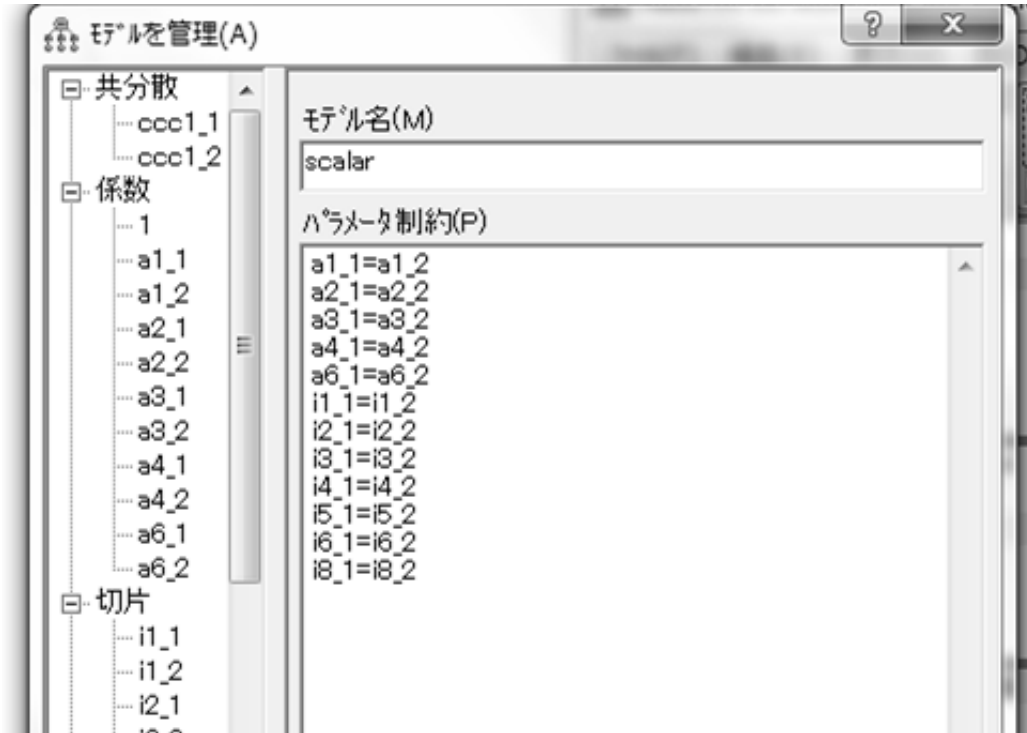
AGEGROUP = younger

chi-squared= χ^2 cmin
df= χ^2 df
CFI= χ^2 cfi
RMSEA= χ^2 rmsea
AIC= χ^2 aic



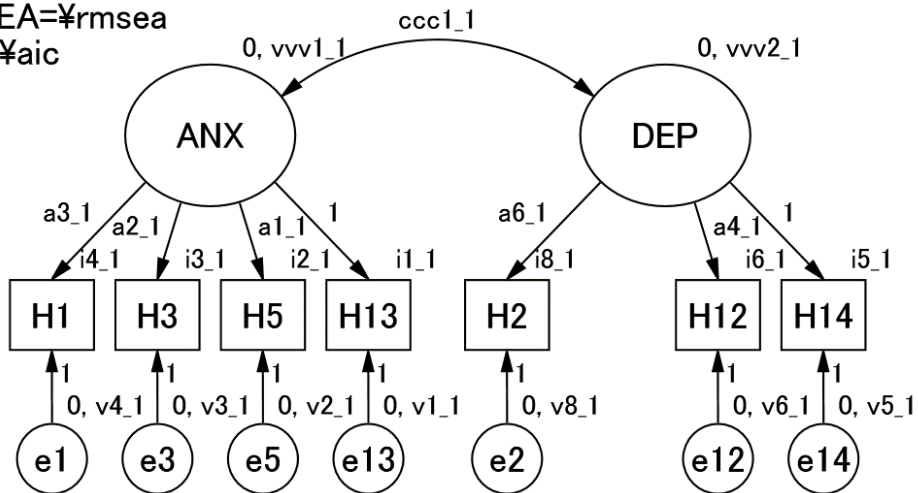
AGEGROUP = elder

次に、「モデル管理 (A)」のダイアログボックスの「モデル名 (M)」で、新しいモデルとして scalar というものを作り、次のようなパラメータ制約を記入する。手入力がわずらわしい場合は、「モデルを管理 (A)」のダイアログボックスの左の欄に、共分散、係数、切片などの変数名が出ているので、希望する変数にカーソルを合わせ、左ダブルクリックをすると、自動的にその変数名が「パラメータ制約」に移動するので便利である。



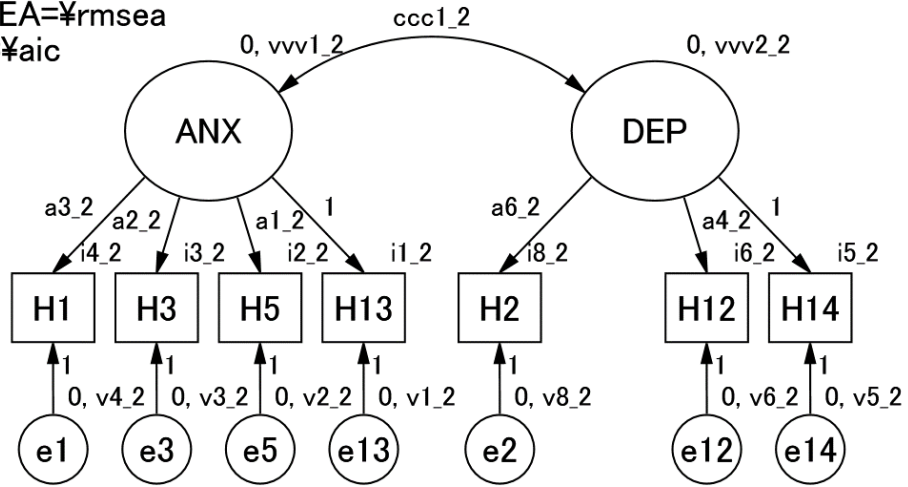
加えて因子 (ANX と DEP) の平均をすべて “0” に設定する。

chi-squared= χ^2 cmin
 df= χ^2 df
 CFI= χ^2 cfi
 RMSEA= χ^2 rmsea
 AIC= χ^2 aic



AGEGROUP = younger

chi-squared=¥cmin
df=¥df
CFI=¥cfi
RMSEA=¥rmsea
AIC=¥aic



AGEGROUP = elder

この計算を実行する。今回は問題なく scalar モデルが許容された。

CMIN

モデル	NPAR	CMIN	自由度	確率	CMIN/DF
configural	44	42.136	26	.024	1.621
metric	39	48.575	31	.023	1.567
scalar	32	54.218	38	.043	1.427
モデル 3	44	42.136	26	.024	1.621
モデル 4	44	42.136	26	.024	1.621
モデル 5	44	42.136	26	.024	1.621
飽和モデル	70	.000	0		
独立モデル	28	415.446	42	.000	9.892

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	IFI	RFI	TLI
				Delta-1	Delta-2	rho-1	rho2
scalar	7	5.643	.582	.014	.015	-.014	-.016

基準比較

モデル	NFI	RFI	IFI	TLI	CFI
	Delta1	rho1	Delta2	rho2	
configural	.899	.836	.959	.930	.957
metric	.883	.842	.954	.936	.953

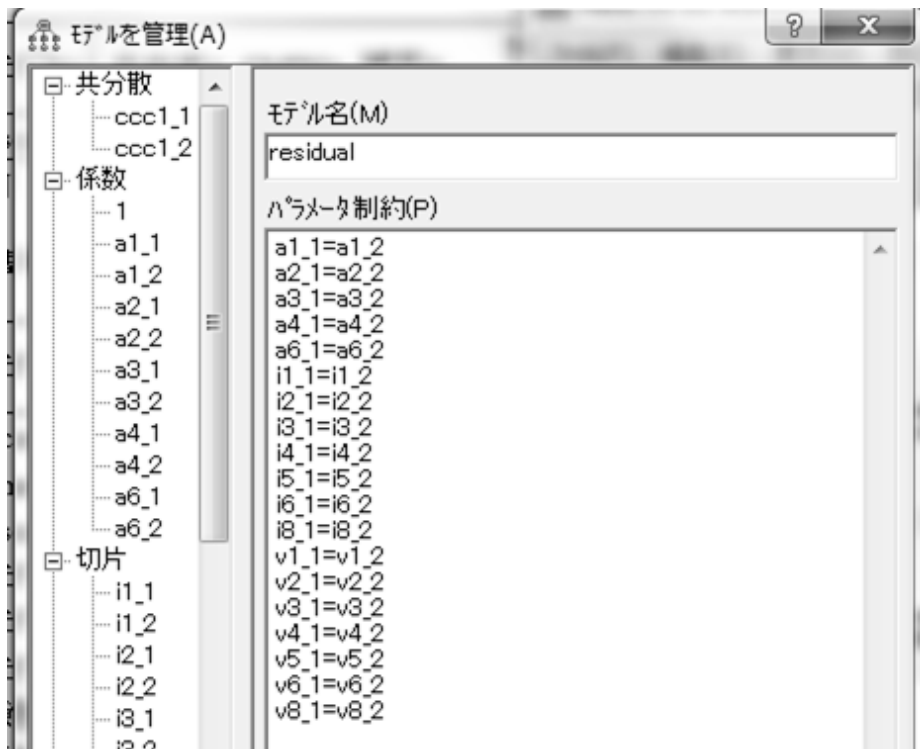
モデル	NFI	RFI	IFI	TLI	CFI
	Delta1	rho1	Delta2	rho2	
scalar	.869	.856	.957	.952	.957
モデル 3	.899	.836	.959	.930	.957
モデル 4	.899	.836	.959	.930	.957
モデル 5	.899	.836	.959	.930	.957
飽和モデル	1.000		1.000		1.000
独立モデル	.000	.000	.000	.000	.000

RMSEA

モデル	RMSEA	LO 90	HI 90	PCLOSE
configural	.052	.019	.080	.415
metric	.050	.019	.076	.469
scalar	.043	.008	.068	.640
モデル 4	.052	.019	.080	.415
モデル 5	.052	.019	.080	.415
独立モデル	.198	.181	.216	.000

5 Strict factorial (residual) invariance の判定

次に residual model をつくる。



この計算を実行する。唯一 CFI の値が metric モデルから residual モデルに移行することで 0.018 (= 0.957 - 0.939) 悪化していた。他の2つの指標は residual モデルを支持していたので、residual モデルが許容されたと解釈する。

CMIN

モデル	NPAR	CMIN	自由度	確率	CMIN/DF
configural	44	42.136	26	.024	1.621
metric	39	48.575	31	.023	1.567
scalar	32	54.218	38	.043	1.427
residual	25	67.661	45	.016	1.504
モデル 4	44	42.136	26	.024	1.621
モデル 5	44	42.136	26	.024	1.621
飽和モデル	70	.000	0		
独立モデル	28	415.446	42	.000	9.892

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	7	13.443	.062	.032	.036	.008	.009

基準比較

モデル	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
configural	.899	.836	.959	.930	.957
metric	.883	.842	.954	.936	.953
scalar	.869	.856	.957	.952	.957
residual	.837	.848	.939	.943	.939
モデル 4	.899	.836	.959	.930	.957
モデル 5	.899	.836	.959	.930	.957
飽和モデル	1.000		1.000		1.000
独立モデル	.000	.000	.000	.000	.000

RMSEA

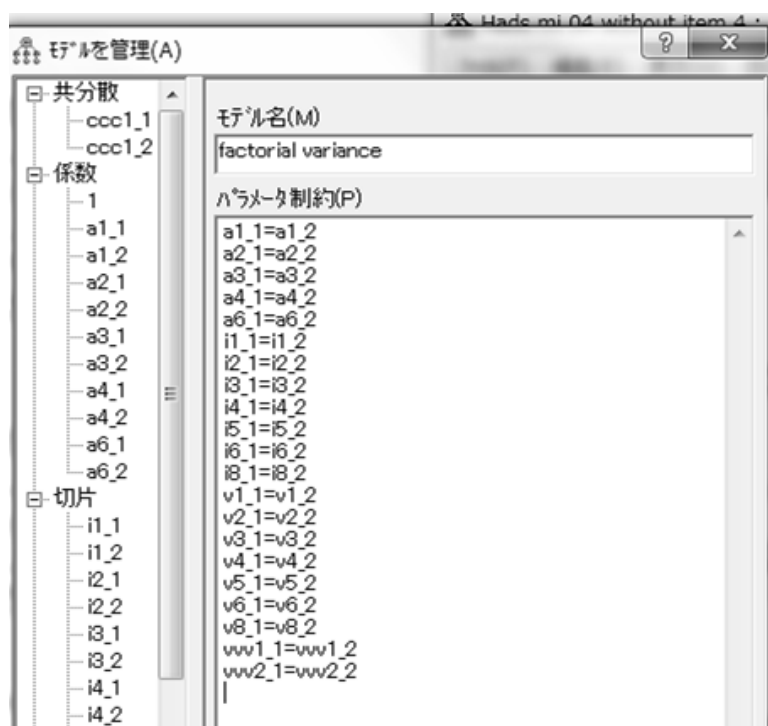
モデル	RMSEA	LO 90	HI 90	PCLOSE
configural	.052	.019	.080	.415
metric	.050	.019	.076	.469
scalar	.043	.008	.068	.640
residual	.047	.021	.069	.556
モデル 4	.052	.019	.080	.415

モデル	RMSEA	LO 90	HI 90	PCLOSE
モデル 5	.052	.019	.080	.415
独立モデル	.198	.181	.216	.000

Residual モデルは許容された。そこで factorial variance モデルに進む。

6 Factor variance invariance の判定

新しいモデルを次のように作る。



この計算を実行する。問題なくこのモデルが許容された。

CMIN

モデル	NPAR	CMIN	自由度	確率	CMIN/DF
configural	44	42.136	26	.024	1.621
metric	39	48.575	31	.023	1.567
scalar	32	54.218	38	.043	1.427
residual	25	67.661	45	.016	1.504
factorial variance	23	69.302	47	.019	1.475
モデル 5	44	42.136	26	.024	1.621
飽和モデル	70	.000	0		
独立モデル	28	415.446	42	.000	9.892

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
factorial variance	2	1.641	.440	.004	.004	-.003	-.003

基準比較

モデル	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
configural	.899	.836	.959	.930	.957
metric	.883	.842	.954	.936	.953
scalar	.869	.856	.957	.952	.957
residual	.837	.848	.939	.943	.939
factorial variance	.833	.851	.939	.947	.940
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
configural	.052	.019	.080	.415
metric	.050	.019	.076	.469
scalar	.043	.008	.068	.640
residual	.047	.021	.069	.556
factorial variance	.046	.019	.068	.597
独立モデル	.198	.181	.216	.000

7 Factor covariance invariance の判定

Factorial variance model の不変性は許容された。次に factorial covariance model を作る。



このモデルも以下のように許容された。

CMIN

モデル	NPAR	CMIN	自由度	確率	CMIN/DF
configural	44	42.136	26	.024	1.621
metric	39	48.575	31	.023	1.567
scalar	32	54.218	38	.043	1.427
residual	25	67.661	45	.016	1.504
factorial variance	23	69.302	47	.019	1.475
factorial covariance	22	70.655	48	.018	1.472
飽和モデル	70	.000	0		
独立モデル	28	415.446	42	.000	9.892

モデル **factorial variance** は正しいという仮定の下で:

モデル	自由度	CMIN	確率	NFI	IFI	RFI	TLI
				Delta-1	Delta-2	rho-1	rho2
factorial covariance	1	1.352	.245	.003	.004	.000	.000

基準比較

モデル	NFI	RFI	IFI	TLI	CFI
	Delta1	rho1	Delta2	rho2	
configural	.899	.836	.959	.930	.957
metric	.883	.842	.954	.936	.953
scalar	.869	.856	.957	.952	.957
residual	.837	.848	.939	.943	.939
factorial variance	.833	.851	.939	.947	.940
factorial covariance	.830	.851	.938	.947	.939
飽和モデル	1.000		1.000		1.000
独立モデル	.000	.000	.000	.000	.000

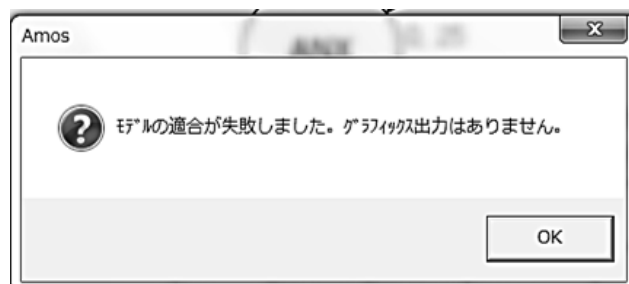
RMSEA

モデル	RMSEA	LO 90	HI 90	PCLOSE
configural	.052	.019	.080	.415
metric	.050	.019	.076	.469
scalar	.043	.008	.068	.640
residual	.047	.021	.069	.556
factorial variance	.046	.019	.068	.597
factorial covariance	.046	.019	.067	.602
独立モデル	.198	.181	.216	.000

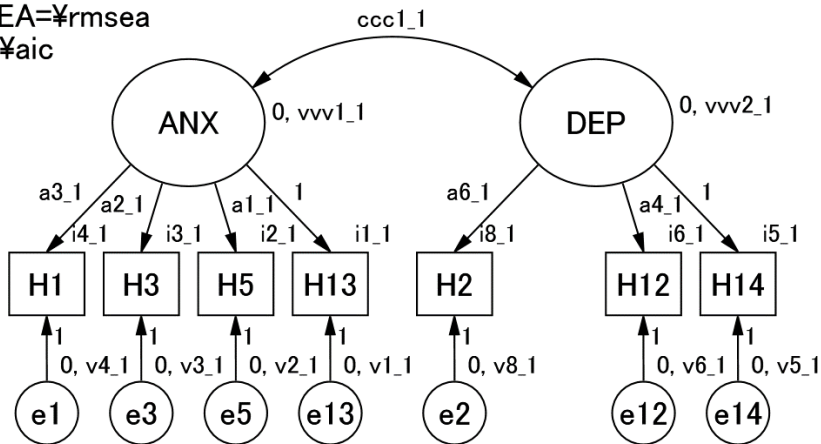
Factor covariance モデルの普遍性は許容された。次に factor mean モデルに進む。

8 Factor mean invariance の判定

Factorial covariance model の不変性は許容された。次に factorial mean model モデルを作る。先ほどの factor covariance モデルのまでの描画を見ると、両群のすべての因子の平均値が“0”と設定されていた。ここでは、どちらかひとつのグループでのみ因子平均値を“0”として、それを参照グループとする。他のグループの因子に適宜名称をつける。ここでは30歳以上の妊婦群の ANX の平均値を fmANX, DEP の平均値を fmDEP と命名した。その上で計算をする。しかし、このモデルは factor mean モデルのみ検定できる。上記操作をすることで、これまでの configural モデルから factor covariance モデルまでのすべての計算はできなくなる。たとえば、情報領域から configural あるいは metric にカーソルを合わせてクリックすると、次の警告が現れる。

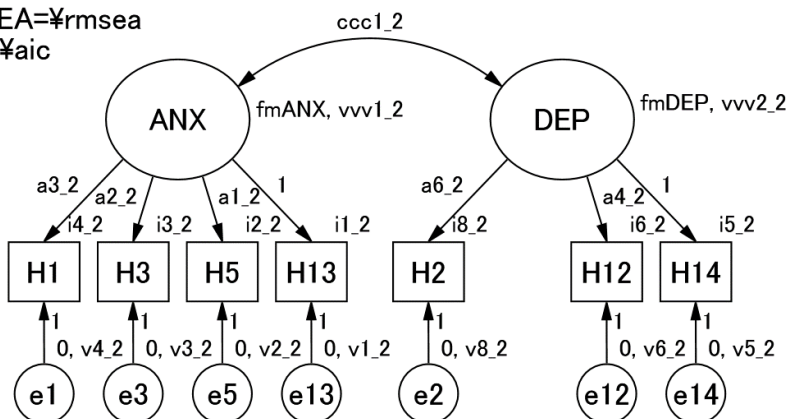


chi-squared=¥cmin
 df=¥df
 CFI=¥cfi
 RMSEA=¥rmsea
 AIC=¥aic



AGEGROUP = younger

chi-squared=¥cmin
 df=¥df
 CFI=¥cfi
 RMSEA=¥rmsea
 AIC=¥aic



AGEGROUP = elder

警告を気にせず、「テキスト出力の表示」から、「推定値」を選び、Elder の「平均値」を求める。すると以下のようなデータが出て来る。ANX の因子の平均値は 20 歳以上の妊婦では 20 歳未満の妊婦より -0.036 低く、また DEP は -0.001 低いと示されている。しかし、いずれの統計検定量は低く $p = 0.624$ 、 $p = 0.994$ であった。両群間で 2 つの因子の平均値に差はなかったのである。

平均値 : (elder - factor mean)

	推定値	標準誤差	検定統計量	確率	ラベル
ANX	-0.036	.074	-0.490	.624	fmANX
DEP	-0.001	.073	-.007	.994	fmDEP

5 作表のしかた

これまでの結果を表にすると以下ようになる。

表 12-3. 配置不変性・構造不変性のまとめ

	χ^2	<i>df</i>	χ^2/df	$\Delta\chi^2$ (<i>df</i>)	CFI	Δ CFI	RMSEA	Δ RMSEA	Judgement
Younger vs. Elder									
Configural	60.333	38	1.588	ref	0.953	Ref	0.051	Ref	ACCEPT
Metric	74.963	44	1.704	14.630 (6) *	0.934	0.019	0.056	0.005	REJECT
Younger vs. Elder without item 4									
Configural	42.136	26	1.621	Ref	0.957	Ref	0.052	Ref	ACCEPT
Metric	48.575	31	1.567	6.438 (5)	0.953	0.004	0.050	+0.002	ACCEPT
Scalar	54.218	38	1.427	5.643 (7)	0.957	+0.004	0.043	+0.007	ACCEPT
Residual	67.661	45	1.504	13.443 (7)	0.939	0.018	0.047	0.004	ACCEPT
Factor variance	69.302	47	1.475	1.641 (2)	0.940	+0.001	0.046	+0.001	ACCEPT
Factor covariance	70.655	48	1.472	1.352 (1)	0.939	0.001	0.046	0.000	ACCEPT

* $p < .05$; ** $p < 0.01$; *** $p < 0.001$

latent mean difference: Elder as compared with the younger

Factors	Differences of latent mean	SE
ANX	-0.036 ^{NS}	0.074
DEP	-0.001 ^{NS}	0.073

* $p < .05$; ** $p < .01$; *** $p < .001$; NS, not significant

3 今後の展開

1. 異なるアプローチ

この章で述べてきた測定・構造不変性があまりに厳密すぎて、多くの尺度が不変性のないものと判断される可能性がある。そこで、すべての因子負荷量ではなく、一部のみ不変性が満たされれば不変性を認めるべきとの意見もある¹³。しかし、これには批判もある。

2. さまざまな研究領域

実例として見たのは、年齢群で不安・抑うつ状態の因子構造が一致しているか否かを見たものである。実際の研究では、例えば男性と女性で、うつ病の症状¹⁴、創造的思考¹⁵、パーソナリティ¹⁶、体型のこだわりの評価方法¹⁷、あるいはインターネット情報についての認識の評価¹⁸の構造が同一であることを示す研究がある。当然、同時点で男性群と女性群を測定しているという前提で議論したものである。属性によって因子構造は変わりうる。こうした属性として、性別、結婚歴、子どもの有無（初産婦・経産婦）、教育歴などさまざまなものが考えられる。

さらに、同じ評価法を用いて多文化における研究をする場合には重大な研究課題となる¹⁹。たとえ翻訳作業が正確であっても、文化によって項目の受け取り方やイメージの違いなどは十分に想定できる。従って、測定不変性・構造不変性を確認しないまま、単に合計点や下位尺度得点を国別に比較し、その意味を検討することは誤った解釈を導く危険性があるといえよう。

一方、同じ評価方法を同じサンプルに適用するが、一定間隔を置いて数回測定する場合も、配置不変・測定不変・構造不変を検討する必要がある²⁰。例えば児童の行為障害を複数の項目で測定することを想像してみよう。同じ項目でも（例えばほかの子どもと喧嘩する）、その意味するところは、就学前、小学校低学年、小学校高学年、中学生で異なることが直感的にわかる。そこで同一の尺度を用いるだけで、同一の概念を評価していると結論するわけにはいかない。スイスの研究グループ²¹は、子ども共感性を評価する尺度を同じ集団に対し、12歳、13歳、14歳、15歳、16歳、35歳の6時点で配布し、年齢が上がってもその因子構造は変わらないことを報告している。こうした測定方法の確認は、メンタルヘルス領域の研究では必須の基礎研究になっている。

¹³ Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456-466.

¹⁴ González-Blanch, C., Medrano, L. A., Muñoz-Navarro, R., Ruiz-Rodríguez, P., Moriana, J. A., Limonero, J. T., Schmitz, F., Cano-Vindel, A. (2018). Factor structure and measurement invariance across various demographic groups and over time for the PHQ-9 in primary care patients in Spain. *PLoS ONE*, 13(2), e0193356.

¹⁵ Kuhn, J.-T., & Holling, H. (2009). Measurement invariance of divergent thinking across gender, age, and school forms. *European Journal of Psychological Assessment*, 25(1), 1-7.

¹⁶ Hong, S., Malik, M. L., & Lee, M.-K. (2003). Testing configural, metric, scalar, and latent mean invariance across genders in sociotropy and autonomy using a non-Western sample. *Educational and Psychological Measurement*, 63(4), 636-654.

¹⁷ Tanaka, M., Tayama, J., & Arimura, T. (2015). Factor structure of the Body Image Concern Inventory in a Japanese sample. *Body Image*, 13, 18-21.

¹⁸ Chiu, Y.-L., Tsai, C.-C., & Liang, J.-C. (2015). Testing measurement invariance and latent mean differences across gender groups in college students' Internet-specific epistemic beliefs. *Austrasian Journal of Educational Technology*, 31(4), 2015.

¹⁹ 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.

van de Vijver, F. J. R., & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross-cultural Psychology*, 31(1), 33-51.

²⁰ Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factor invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, 4(1), 10-18.

²¹ Allemand, M., Steiger, A. E., & Fend, H. A. (2015). Empathy development in adolescence predicts social competencies in adulthood. *Journal of Personality*, 83, 229-241.

なお、名義尺度あるいは順序尺度を用いた因子構造の測定不変性²²、二次因子構造の測定不変性²³、非常に多数のグループ間で測定不変性を確認する手法についても最近開発が進んでいる²⁴。

²² Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, 39(3), 479-515.

Jang, Y. J., Lee, J., Puig, A., & Lee, S. M. (2012). Factorial invariance and latent mean differences for the Five factor Wellness Inventory with Korean and American Counselors. *Measurement and Evaluation in Counseling and Development*, 45(2), 71-83.

²³ Rudnev, M., Lytkina, E., Davidov, E., Schmidt, P., & Zick, A. (2018). Testing measurement invariance for a second-order factor: A cross-national test of the alienation scale. *Methods, Data, Analyses*, 12(1), 47-76.

²⁴ Zercher, F., Schmidt, P., Cieciuch, J., & Davidov, E. (2015). The comparability of the universalism value over time and across countries in the European Social Survey: Exact vs. appropriate measurement invariance. *Frontiers in Psychology*, 6, 733.