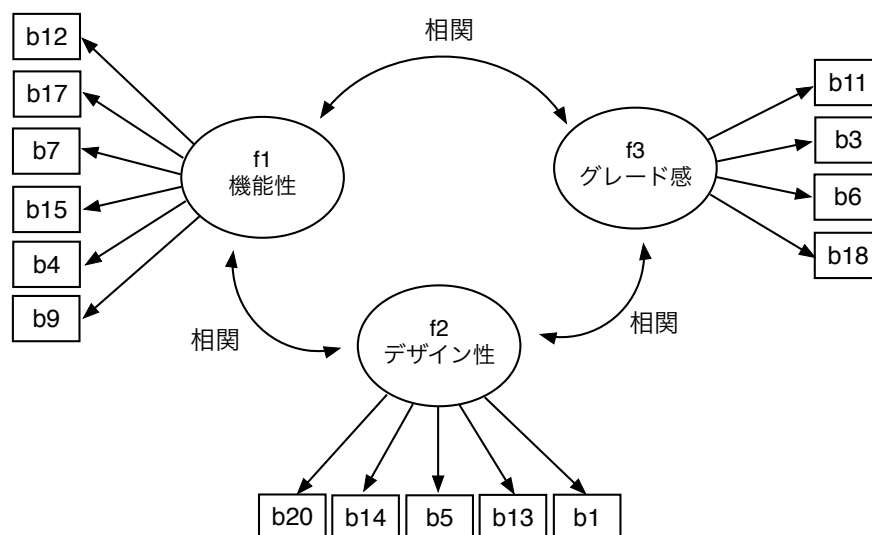


25 日目 : 確認的因子分析

「そこそこ」Rを使えるようになることを目指したこの資料に、この内容を入れるべきかどうか迷ったのですが、共分散構造分析を使った研究が増えていることもあって、ちょっとだけ触れておきます。

共分散構造分析に関しては、**sem**や**lavaan**といった代表的なパッケージがあります。確認的因子分析は共分散構造分析の利用法のひとつです。先行研究から尺度を借りてきた場合、オリジナルの尺度を作成する場合などに有益ですし、共分散構造分析の練習、パッケージを使う練習にもなると思うので、「sam2.xls」のデータを利用し、13 日目で作成した尺度について確認的因子分析を行ってみます。

10 日目から 13 日目にかけて、得られたデータの背後に、いくつかの、どのような因子が仮定されるのかを探索して確定し、それを測定する尺度を作成しました。そこで見いだされた因子構造、すなわちモデルは、相互に相関をもつ3つの因子があり、それぞれの因子は6項目、5項目、4項目からなる項目群にそれぞれ影響しているというものです。図にすれば以下のようなでしょう。このモデルの確からしさ、データとの当てはまり具合をチェックするのが確認的因子分析といってよいでしょう。



今回、パッケージは `lavaan` を使います。以下が命令ですが、まずモデルを組み、`cfa` 関数で計算し、`summary` で表示させるという単純な構造になっています、モデルは' 'で囲まれた部分のことで、以下のように式でそれを定義し、`f_model` という名前にします。`f_model` をデータ `x` に適用する計算を `cfa` 関数で行い、結果を `fit` に入れます。`summary` で `fit` の内容を表示します。ここまでが基本部分になります。

その後の `fitmeasures` は、各種の適合度指標の表示を指示する部分、`subset` の部分は、`modindices` によってもたらされる修正指標から、`mi>10` のものだけを抽出して表示させる部分です。修正指標は、適合度を向上させるための情報で、`mi` が大きいほど、より適合度指標の向上が見込める関連性を意味します。

```
library(lavaan)
f_model <- '
f1 =~ b12 + b17 + b7 + b15 + b4 + b9
f2 =~ br20 + b14 + b5 + b13 + b1
f3 =~ b11 + b6 + b3 + b18
'

fit <- cfa(f_model, data=x, estimator="ML")
summary(fit, standardized=TRUE, rsquare=TRUE )
fitmeasures(fit, c("chisq", "df", "pvalue", "gfi", "agfi", "nfi", "cfi",
"rmsea", "aic"))
subset(modindices(fit), mi >10)
```

どのように結果が表示されるのかは、実際にやってみてもらおうのがよいでしょう。かなり長い結果が示されます。

以下はその一部で、因子と項目の関係（因子から項目への影響）の部分です。先のモデル定義の部分と対照させてください。一番右の `Std.all` が因子負荷量に対応するパスの係数です。`P` の部分が有意性の検定結果です。

Latent Variables:	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
f1 =~						
b12	1.000				0.671	0.802
b17	1.135	0.074	15.430	0.000	0.761	0.822
b7	0.954	0.067	14.335	0.000	0.640	0.774
b15	0.922	0.074	12.396	0.000	0.619	0.633
b4	0.954	0.068	14.024	0.000	0.640	0.761

さて、定義されたモデルとデータの適合の程度は、適合度指標とよばれる様々な指標を総合的に判断します。今回参照した指標の中で、 χ^2 値は適合が悪いほど値が大きくなり、「データはモデルに適合している」という帰無仮説を検定するために使われます。棄却されないことが望ましいのですが、この検定結果のみを用いて判断することはほとんどありません。GFI, AGFI, NFI, CFI といった指標は1に近いほど適合が良く、.900を超えることが望ましいとされます。RMSEAは0に近いほど適合が良いことを示し、.050を下回ることを望ましく、.100を超えるとモデルを再検討すべきと考えられるものです。

今回得られた適合度指標は、 $\chi^2(87) = 300.5$ で、1%水準で有意になります。GFI = .880, AGFI = .835, NFI = .866, CFI = .900, RMSEA = .091 でした。これらを総合的に判断すると、当てはまりがよいとは言い難いもの、悪いともいえない、許容範囲内にある…くらいでしょうか。換言すれば、「改善の余地あり」ということになります。

次に、一部を上の結果の図に示した、因子から項目へのパス係数に注目すると、統計的にはすべてのパスが有意なので、因子と項目の対応関係は無関係とはいえないことになります。このレベルでの問題はないでしょう。しかし、その数値を全体的にながめると低いものもいくつか目に付きます。つまり、モデル的には対応する因子から影響を受けるはずなのに、因子以外の何らかからの影響が強い項目がモデルに組み込まれてしまっているということです。b9 はもちろん、b1 などにも気になります。これらの項目は改善が必要といえるかもしれません。

modindices によって修正指標が得られますが、mi が一番大きかったのは、f3 ~ b15, すなわち「f3 から b15 へのパスを追加すべき」というものでした。こうなると、計算上の指摘をふまえて、「グレード感」の因子が「安心感のある」という項目に影響しているというパスを加えるというモデルの修正が妥当かどうかという判断になります。また、測定という点から、この結果をどう扱うべきかという問題も考慮に入れなければなりません。

ダミーデータである今回の結果を検討しても意味がないので、結果を読み、検討するポイント、方向性がわかるようになってもらえればいいかなと思います。また先行研究や、そこで作成された尺度を評価する目も養えるかと思えます（13 日目の探索的因子分析の結果と、確認的因子分析の結果の差、特に因子と項目の関連性の強さに注目してください）。先行研究で作成された尺度を自分の研究で使う際には、その研究で示された因子分析の結果、負荷量の程度などをしっかり吟味しておかないと、データをとった後で困惑する場合があります。確認的因子分析をおこなって、十分な適合度が得られるような尺度かどうか…

今回は、ちょっとだけ、概略のみを紹介しました。共分散構造分析はいろいろと難しい部分がありますが、興味深い情報も得られます。この資料だけではまったくたりませんので、おもしろいと感じたらいろいろと調べてみてください。

lavaan の概略を把握するには以下などがよいと思います。

「lavaan チュートリアル」

<http://www.ec.kansai-u.ac.jp/user/arakit/documents/>

lavaanTutorial20170124.pdf

本日はここまでです。