Exploratory Factor Analysis: dimensionality and factor scores

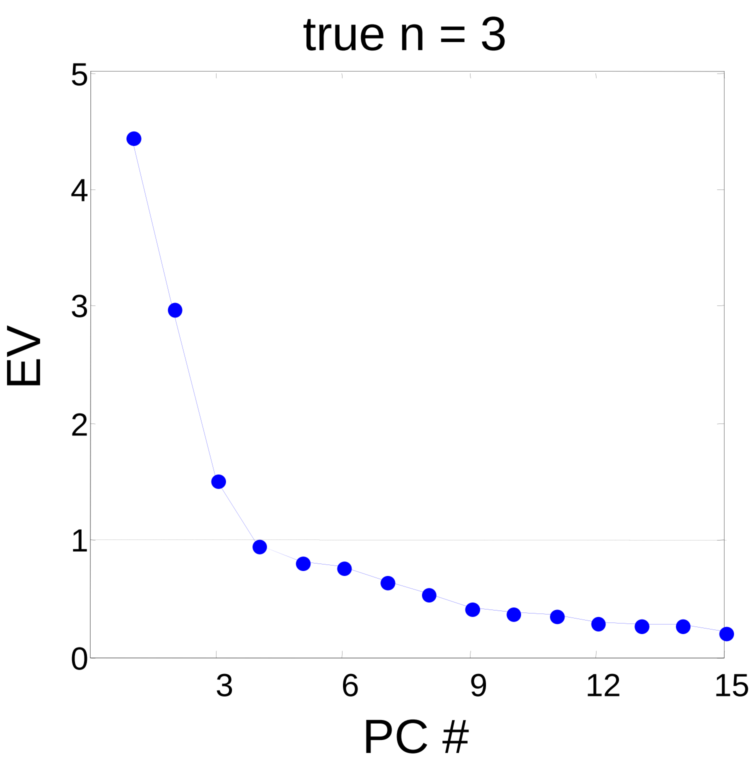
Psychology 588: Covariance structure and factor models

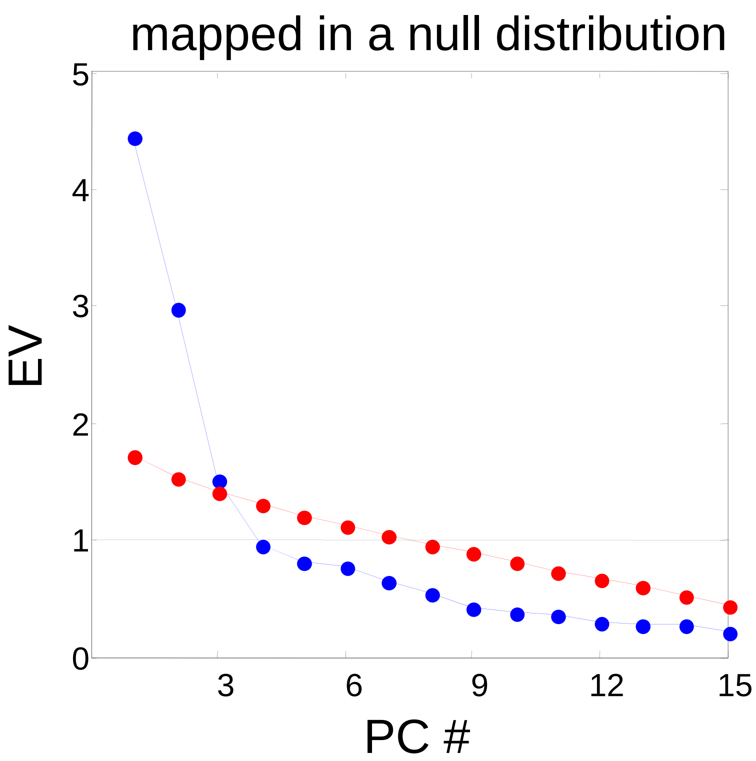
How many PCs to retain

- Unlike confirmatory FA, the number of factors to extract is not known in advance, or a presumed dimensionality need be empirically supported
- There is no generally acceptable single guideline to determine the dimensionality --- some are relevant to common factors and others to principal components, and mostly to both
- Many rules and tests are available, but unfortunately not necessarily suggesting the same number
- The most popular are Scree Test by Cattell and, so called, eigenvalue-greater-than-1 (or VAF per factor > average variance; a.k.a., Guttman-Kaiser rule)

- Scree Test is criticized for its graphical nature (subjective and non-statistical) --- parametric (Bentler & Yuan, 1998) and non-parametric (Hong et. al, 2006) scree tests are available

- Parallel Analysis (Horn, 1965) improves the G-K rule for data size, yet non-statistical





Large sample inference

- With a random sample $\mathbf{x}$ from a normal distribution, the sampling distributions of eigenvalues, eigenvectors, and equality of the last $q-n$ eigenvalues are known --- but with <u>large N</u>, these properties also hold for non-normally distributed $\mathbf{x}$ (due to central limit theorem), allowing for parametric statistical testing

- Sample eigenvalues $\mathbf{e}$ are distributed as $N_q\left(\mathbf{e}, 2\mathbf{E}^2/N\right)$, so that we can test $H_0 : e_k = 0,\ k = 1,\ldots,n,$ with associated $100(1-\alpha)\%$ CI as:

$$\frac{\hat{e}_k}{1 + z(\alpha/2)\sqrt{2/N}} \le e_k \le \frac{\hat{e}_k}{1 - z(\alpha/2)\sqrt{2/N}}$$

- Another parametric test for principal components is Bartlett test for $H_0 : e_{n+1} = \cdots = e_q$, which is known to suggest too many PCs in practice

$$\hat{\chi}^2 = -(N - q/3 - 5/6 - 2n/3)\log R_{q-n},$$

$$df = (q-n)(q-n-1)/2, \quad R_{q-n} = \frac{|\mathbf{R}|}{\left(\prod_{k=1}^{n} e_k\right)\left(\frac{q - \sum_{k=1}^{n} e_k}{q-n}\right)^{q-n}}$$

- Sample size has a direct consequence on the statistic with no adjustment in the df --- large N causing too many components retained

- Anderson provides more generally applicable χ^2 statistic than Bartlett's (e.g., not necessarily the last $q-n$ roots), which is widely used in practice and isn't sensitive to too large N

$$\hat{\chi}^2 = -\nu\left(\sum_{k=n+1}^{q}\log e_k - (q-n)\log\frac{\sum_{k=n+1}^{q} e_k}{q-n}\right),$$

$$df = (q-n-1)(q-n-2)/2, \qquad \nu = q(q+1)/2$$

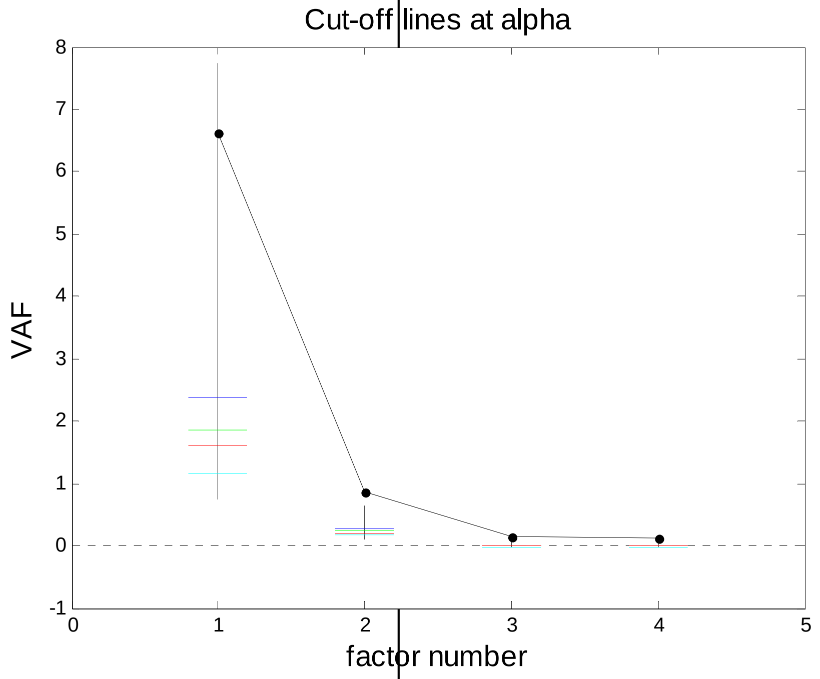
- In most cases, this improved test behaves reasonably while it suggests too many components to retain when the 1st PC is dominantly large

- Similar χ^2 tests are available for the common factor model, due to Anderson, Lawley and Rubin

- National track records data for men tested by the Bartlett test (using Anderson's formula) and the bootstrap Scree test (MATLAB codes available in my netfiles, "bscree.m" with its syntax "[dim,pvalues] = bscree(X, alpha, nBs, flagm)"

```
[ndim,prob,chisquare] = barttest(zscore(trackm),.05)

ndim = 8
prob =  0                chisquare = 768.9522
        0                            256.6401
   0.0000                             59.0284
   0.0003                             39.6948
   0.0074                             22.5032
   0.0080                             15.6116
   0.0326                              6.8479
```




```
dim = bscree(zscore(trackm), [.25 .10 .05 .01], 1000, 0)
    = 4    3    2    2
```

Indeterminacy of factor scores

- Once all parameters of the common factor model ($\mathbf{\Lambda}$, $\mathbf{\Phi}$ and $\mathbf{\Theta}$) are obtained, we may sometimes want to know factor scores of "subjects" --- quantities that are not considered as a part of model parameters, instead, as some values useful to know afterwards

- Factor scores are not uniquely determinable since there are $n+q$ unknown factors, given only q data variables

$$\underset{q\times 1}{\mathbf{x}_i} = \left[\underset{q\times n}{\mathbf{\Lambda}}, \underset{q\times q}{\mathbf{I}} \right] \underset{(n+q)\times 1}{\begin{bmatrix} \boldsymbol{\xi}_i \\ \boldsymbol{\delta}_i \end{bmatrix}}, \quad i = 1,\ldots,N, \quad E\left(\boldsymbol{\xi}\boldsymbol{\delta}'\right) = \mathbf{0}, \quad E\left(\boldsymbol{\delta}\boldsymbol{\delta}'\right) = \mathbf{\Theta}$$

- Two approaches are considered here to overcome this problem --- weighted least squares and regression methods

What if PCs are extracted

- Since the PC model "ignores" the existence of specific factors, the estimation of factor scores simply reduces to the OLS, given (mean-centered) data and the loading matrix as:

$$\mathbf{x}_i = \boldsymbol{\Lambda}\boldsymbol{\xi}_i + \boldsymbol{\delta}_i, \quad \hat{\boldsymbol{\xi}}_i = \left(\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\Lambda}'\mathbf{x}_i = \mathbf{E}^{-1}\boldsymbol{\Lambda}'\mathbf{x}_i$$

  which exactly determines the least-squares estimate of $\boldsymbol{\xi}$ --- accordingly, the indeterminacy of factor scores doesn't apply to the PC model

- This LS property holds for any rotated $\boldsymbol{\Lambda}$

$$\hat{\tilde{\boldsymbol{\xi}}}_i = \left(\tilde{\boldsymbol{\Lambda}}'\tilde{\boldsymbol{\Lambda}}\right)^{-1}\tilde{\boldsymbol{\Lambda}}'\mathbf{x}_i = \left(\mathbf{T}'^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\mathbf{T}^{-1}\right)^{-1}\mathbf{T}'^{-1}\boldsymbol{\Lambda}'\mathbf{x}_i$$

$$= \mathbf{T}\mathbf{E}^{-1}\boldsymbol{\Lambda}'\mathbf{x}_i = \mathbf{T}\hat{\boldsymbol{\xi}}_i$$

WLS for common factor scores

- Under the common factor model, q observed variables have a varying contribution to the n common factors (i.e., different communalities) --- taking this into account, a weighted sum of squared errors would provide a better prediction of factor scores (due to Bartlett) as:

$$f_{\text{WLS\_}i} = \sum_{j=1}^{q} \frac{\delta_{ij}^2}{\phi_{jj}} = \boldsymbol{\delta}_i' \boldsymbol{\Theta}^{-1} \boldsymbol{\delta}_i = \left(\mathbf{x}_i - \boldsymbol{\Lambda\xi}_i \right)' \boldsymbol{\Theta}^{-1} \left(\mathbf{x}_i - \boldsymbol{\Lambda\xi}_i \right)$$

- Accordingly, the WLS estimator is:

$$\hat{\boldsymbol{\xi}}_{\text{WLS\_}i} = \left(\boldsymbol{\Lambda}' \boldsymbol{\Theta}^{-1} \boldsymbol{\Lambda} \right)^{-1} \boldsymbol{\Lambda}' \boldsymbol{\Theta}^{-1} \mathbf{x}_i$$

- Would be the weight θ_{jj}^{-1} larger or smaller with larger communality h_{jj}^2 ?

- A rotated version of $\hat{\boldsymbol{\xi}}_{\text{WLS}}$ is obtained by simply replacing $\boldsymbol{\Lambda}$ in the formula by $\tilde{\boldsymbol{\Lambda}}$

- The WLS estimate satisfies only the 0-mean property of $\boldsymbol{\xi}$ but not the others (i.e., unit-variance & orthogonality if orthogonally rotated), though the deviations tend ignorable --- alternatively, Anderson-Rubin's modified estimation provides all satisfied results

$$\hat{\boldsymbol{\xi}}_{\text{A-R\_}i} = \left(\boldsymbol{\Lambda}'\boldsymbol{\Theta}^{-1}\mathbf{S}\boldsymbol{\Theta}^{-1}\boldsymbol{\Lambda}\right)^{-0.5}\boldsymbol{\Lambda}'\boldsymbol{\Theta}^{-1}\mathbf{x}_i$$

Regression on data variables

- Consider a partitioned vector of the data variables and common factors $[\mathbf{x}', \boldsymbol{\xi}']'$, with all entries mean-centered and normalized to unit variance, then we have the expectation of its cross-products as:

$$E\left(\begin{bmatrix}\mathbf{z}\\ \boldsymbol{\xi}\end{bmatrix}\begin{bmatrix}\mathbf{z}' & \boldsymbol{\xi}'\end{bmatrix}\right) = E\left(\begin{bmatrix}\mathbf{zz}' & \mathbf{z}\boldsymbol{\xi}'\\ \boldsymbol{\xi}\mathbf{z}' & \boldsymbol{\xi\xi}'\end{bmatrix}\right) = \begin{bmatrix}\mathbf{R} & \mathbf{P}\\ \mathbf{P}' & \boldsymbol{\Phi}\end{bmatrix}, \qquad \mathbf{P} = \boldsymbol{\Lambda\Phi}$$

- And if we set up a regression equation for common factors predicted by the scaled data

$$\boldsymbol{\xi}_{N\times n} = \mathbf{Z}_{N\times q}\mathbf{B}_{q\times n} + \boldsymbol{\varepsilon}$$

Then, the OLS estimator of $\mathbf{B}$ is:

$$\hat{\mathbf{B}} = \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\boldsymbol{\xi} = \mathbf{R}^{-1}\mathbf{P} = \mathbf{R}^{-1}\boldsymbol{\Lambda\Phi}$$

- From the OLS estimator of regression weights, we have

$$\hat{\xi}_{\mathrm{REG}\_i} = \hat{\mathbf{B}}'\mathbf{z}_i = \mathbf{\Phi}\mathbf{\Lambda}'\mathbf{R}^{-1}\mathbf{z}_i, \quad \hat{\mathbf{B}} = \mathbf{R}^{-1}\mathbf{\Lambda}\mathbf{\Phi}$$

Note that this estimator is applicable also to unstandardized, mean-centered data by replacing $\mathbf{Z}, \mathbf{R}, \mathbf{\Lambda}_{\mathrm{R}}, \mathbf{\Phi}_{\mathrm{R}},$ respectively, by $\mathbf{X}, \mathbf{S}, \mathbf{\Lambda}_{\mathrm{C}}, \mathbf{\Phi}_{\mathrm{C}}$ (subscripts R and C represent the correlation and covariance data)

- Factor scores (for factors 1 and 2) are estimated based on ML extraction and Oblimin rotation, once by regression and once by Anderson-Rubin WLS:

Rank	Reg1	Reg2	A-R1	A-R2	Reg1	Reg2	A-R1	A-R2
1	UK	USA	Portugl	Dom Rep	-0.933	-1.688	-1.067	-1.991
2	Kenya	Italy	Kenya	USA	-0.928	-1.465	-0.976	-1.687
3	USA	USSR	NewZlnd	Bermuda	-0.864	-1.259	-0.911	-1.532
4	Portugl	UK	Norway	Italy	-0.855	-1.158	-0.846	-1.467
5	E Ger	W Ger	Nethrlnd	Tailand	-0.853	-1.015	-0.844	-1.308

Reliability of model parameters

- Under normality, the ML method and PCA provide a basis for parametric testing

- For other methods, with large N, split-half analysis can be performed to see whether an optimal factor solution derived from one random half of the data "agree" with results from the other by the same method (same n, factoring and rotation) --- using congruent coefficient to see how the two sets agree

- Alternatively, bootstrapping could be used to empirically create sampling distribution of parameter estimates (e.g., Ichikawa & Konish, 1995)

  - Note that this type of non-parametric approach do not require any of the usual parametric assumptions

References for EFA

- Anderson, T.W., & Rubin, H. (1956). Statistical inference in factor analysis. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 5, 111-150.
- Bartlett, M.S. (1950). Tests of significance in factor analysis. British Journal of Mathematical and Statistical Psychology, 3, 77-85.
- Bentler, P.M., & Yuan, K.H. (1998). Tests for linear trend in the smallest eigenvalues of the correlation matrix. Psychometrika, 63, 131–144.
- Carroll, J.B. (1953). An analytic solution for approximating simple structure in factor analysis. Psychometrika, 18, 23-38.
- Carroll, J.B. (1960). unpublished manuscript for Oblimin
- Cattell, R.B. (1966). The Scree test for the number of factors. Multivariate Behavioral Research, 1, 245-276.
- Guttman-Kaiser test: Kaiser, H.F. (1970). A second generation of Little Jiffy. 35, 401-415.
- Hong, S., Mitchell, S.K, & Harshman, R.A. (2006). Bootstrap scree tests: A Monte Carlo simulation and applications to published data. British Journal of Mathematical and Statistical Psychology, 59, 35-57.
- Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. Psychometrika, 30, 179-185.
- Ichikawa, M., & Konish, S.(1995). Application of the bootstrap methods in factor analysis. Psychometrika, 60, 77-93 .
- Jennrich, R.I., & Sampson, P.F. (1966). Rotation for simple loadings. Psychometrika, 31, 313-323.
- Kaiser, H.F. (1958). The Varimax criterion for analytic rotation in factor analysis. Psychometrika, 23, 187-200.
- MINRES: Harman, H.H. (1976). Modern factor analysis (3rd ed.). Chicago: University of Chicago Press.
- ML factoring: Lawley, D.N., & Maxwell, A.E. (1971). Factor analysis as a statistical method (2nd ed.). New York: Elsevier.
- Simple structure: Thurstone, L.L. (1947). Multiple factor analysis. Chicago: University of Chicago Press.