Correlations

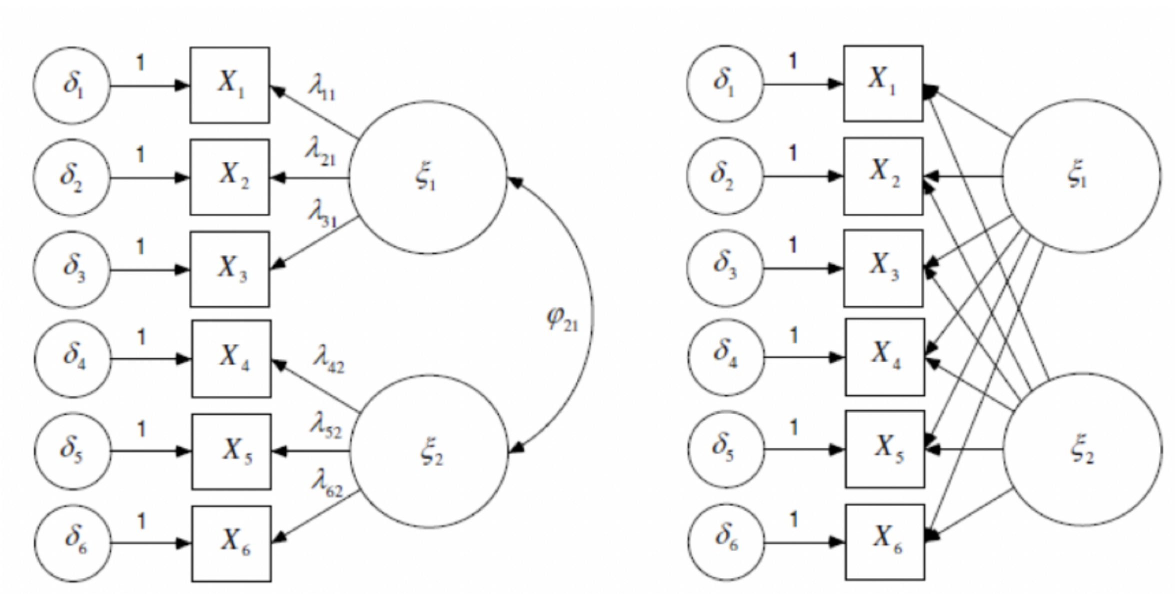
- Number of correlations for x number of variables: $\frac{x(x-1)}{2}$. i.e. 20 measures, $\frac{20(19)}{2} = 190$ correlations.These are a lot of correlations, so we need to use factor analysis to cut them down.
- $variance = \frac{\sum (x_i - \bar{x})^2}{N-1}$
- $covariance = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N-1}$
- $correlation = \frac{cov_{xy}}{\sqrt{var_x}\sqrt{var_y}}$

correlation matrix (1 on diagnals) is covariance matrix (varaince on diagnals) normalized, most times interchangable.

exception: when variances are meaningful, use covariance matrix

- longitudinal data, variance at different timepoints should not be the same, use covariance matrix
- multiple groups, different groups cannot have the same variance, use covariance matrix

Types of Factor Analysis



Confirmatory   Exploratory

- **Exploratory**: Let the data speak to itself, loadings (λ) exist for all measures, but some are small. Suppose there are 6 measures and 2 factors, the loading matrix with (column-factor, row-measures):

$$\begin{bmatrix} \lambda & \lambda \\ \lambda & \lambda \\ \lambda & \lambda \\ \lambda & \lambda \\ \lambda & \lambda \\ \lambda & \lambda \end{bmatrix}$$

- **Confirmatory**: verify the number of latent dimensions of the instrument (factors) and the pattern of item–factor relationships (loadings), restricted, some factor loadings (λ) are exactly 0.

$$\begin{bmatrix} \lambda & 0 \\ \lambda & 0 \\ \lambda & 0 \\ 0 & \lambda \\ 0 & \lambda \\ 0 & \lambda \end{bmatrix}$$

History of factor analysis

- factor analysis originated from intelligence tests.
- Use test to get measurement of intelligence (latent construct)
  - e.g. GPA - independent from students' relative performance or dependent?
  - is there one general intelligence or specific factors (g/s theory)
  - one negative affect construct vs several factors e.g. anxiety, depression etc.
- racism, some use factor analysis to show the superiority of some races
- a mathematical approach used poorly
- what is CFA good for?
  - strong falsification: put theories on danger zones
  - 'I am wrong if this does not hold'

factor analysis motivations

- finding the underlying psychometric structures
- one negative affect construct vs several factors e.g. anxiety, depression etc.
- scale development: developing a score from the structures
- e.g LIP, 20 items -> 18 items
- scoring methods: optimal score for a scale, deciding weights

steps in factor analysis

- how to extract factors
  - principle component analysis
  - principle axis factoring
- how many factors are there
  - very important
- rotation
  - improve interpretability of solutions

PCA vs EFA

1. let loading be 1.0 for each item, fix residual variance to 0

- The underlying factors account for 100% of measurement
- No measurement error
- $\hat{\eta}$ is the mean of all items

2. Let loading be freely estimated from data, fix residual to 0

- 5 items are differentially weighted
- $\hat{\eta}$ is the composite
- This is PCA
- All variance is true

3. Let loading be freely estimated from data, but also estimate residuals freely

- 5 items are differentially weighted
- error is free
- $\hat{\eta}$ is the composite
- This is PAF/Exploratory factor analysis
- Part of the variance is true, others is error
- the variance explained by factros is called "common variance"
- The other variance is "unique" variance
  - broken down to "specific variance" true to item but not associated with error, plus "error variance"

For 5 items

σ^2 of each item is 1.0

σ_t^2 of total items is 5.0

we need fewer than 5 factors for data reduction

orphan item: an item does not belong anywhere

EFA process Eigenvalues are generated for each factor, which corresponds to the proportion of variance. A factor with eigenvalue of 3 means the factor accounts for variance equavalent to 3 items.

Communality: explained variance = $R_m^2 ultiple$ - In factor analysis, communality is 1 - all variance can be used

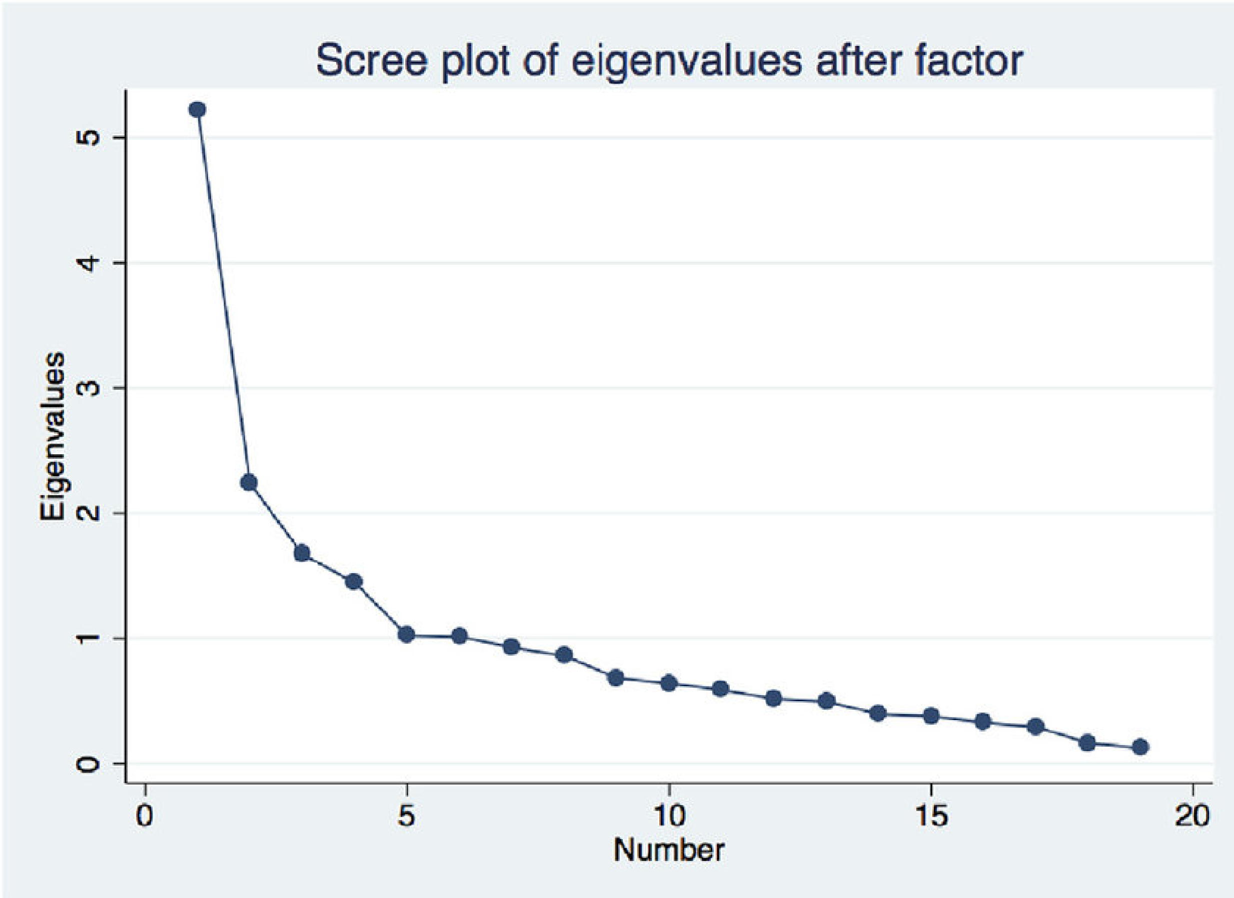
$$R = \begin{bmatrix} 1.0 & & & \\ r & 1.0 & & \\ r & r & 1.0 & \\ r & r & r & 1.0 \end{bmatrix}$$

1. PC_1= get a weighted composite of all items $= 0.1y_1 + 0.9y_2 + 0.3y_3....$

- eigenvalue is the variance of the composite.

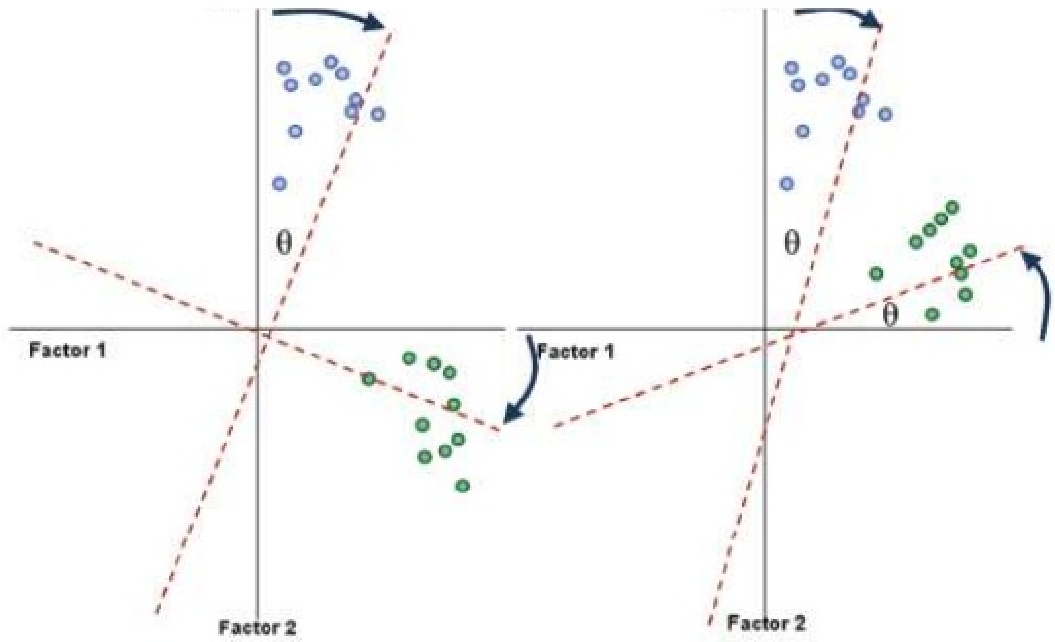
2. PC_2 calculated same was as PC_1, but need to exclude variance of PC_1
3. PC_3 exclude varaince of PC_1, PC_2
4. … follow until all variance is explained **Where do we stop? How to choose factors?**

- **Kaiser-Guttman Rule**: factors in general need to have eigenvalues larger than 1, so a vector represents more variance than 1 measured variables
- **Scree Plot**: A plot of eigenvalues, when the plot slope change, stop choosing factors.



Factor Rotation

- "simple structure": items need to load large on one factor and small on other factors
  - large vs small loading are subjective
- rotation helps interpretability, but it will not change fit of the model
- **Orthogonal Rotation**: the factors have a 90 degree, thus orthogonal. $cos(\gamma)$ where gamma is angle between factors is the correlation between axis representing factors, when γ is 0, $cos(\gamma)$ (correlation) is 0.
  - varimax is one type of orthogonal rotation sometimes referenced as part of the "little Jiffy" process which also includes using PCA and the Kaiser-Gutman rule.
- **Oblique Rotation**: the factors do not retain a 90 degree angle between each other. There are correlation among factors. γ (correlation) is not 0.
  - promax is one type of oblique rotation.
- cross loading: loads across factors



Orthogonal Rotation Oblique Rotation

EFA in SAS The `reorder` argument reorders factor loading matrices.

```
/* PCA with scree plot: nfactors set to 1 because we are not really going to do PCA on data, just using
proc factor data=[data_name] method=principal scree nfactors=1 simple corr; run;
/* PCA with 4 factors using varimax rotation - not an appropriate subsituted of EFA because it ignores i
proc factor data=[data_name] method=principal nfactors=4 rotate=varimax reorder; run;
/* EFA with 4 factors using promax rotation*/;
proc factor data=[data_name] method=ML priors=smc nfactors=4 rotate=promax reorder; run;
```

EFA in Mplus Plot2 in the plot statement specifies Mplus to generate a scree plot.

```
Data:
  File is "[path_to_data_file]" ;
Variable:
  Names are
    [variable_names_following_format: name1 name2 name3 etc];
    Missing are all (-999) ;
Analysis:
    type = efa [lower_factor_number] [higher_factor_number];
    rotation=promax;
Plot:
    type=Plot2;
```