Survey Research and Design in Psychology

Lecture 5 - Exploratory Factor Analysis

Dr Ananthan Ambikairajah

University of Canberra

A quick recap

- Covariation
- Purposes of correlation
- Linear correlation
- Types of correlation
- Interpreting correlation
- Assumptions and limitations

Focus/Overview for today

UNIVERSITY OF CANBERRA

- An introduction to factor analysis
- Example exploratory factor analyses
- The steps/process to conducting an exploratory factor analysis
- Assumptions to be aware of

What is factor analysis

- Factor analysis is a tool we can use to identify the structure of a set of items
- The variance of many variables might be explained by a smaller number of underlying clusters (factors), with each factor representing several related variables
- Factor analysis uses correlations among many variables to sort related variables into clusters called factors

Two main uses of factor analysis

Theory development

- Examine the hypothetical structure of relations between constructs, identify factors, and classify variables

Data reduction

- Reduce variables down to a smaller number of factors, leading to calculation of composite scores for each factor. The composite scores can be used in subsequent analyses

Theory development

UNIVERSITY OF CANBERRA

- We can test theoretical models by investigating the underlying correlational patterns
- We can address theoretical questions, such as:
  - How many personality factors are there? And what are they?
  - Is intelligence a general capacity, or are there multiple dimensions of intelligence?

Examples

- How many dimensions of personality are there? What are they?
- Using factor analysis can help you decide between three and five factor personality models
  - Eysenck's 3: Extraversion, neuroticism and psychoticism
  - Big 5: Openness, conscientiousness, extraversion, agreeableness and neuroticism

Examples

- Is intelligence best described as a global factor? Or several specific factors e.g. verbal, spatial, mathematical, social etc.
- Factor analysis can help decide which model best describes the data

More examples

- Depression scale - may have ‘clusters’ of items measuring different kinds of symptoms e.g. cognitive, somatic, emotional
- Wellbeing - factor analysis can pick out different ‘types’ e.g. feeling good (hedonic wellbeing) vs feeling life is meaningful (eudaimonic wellbeing)
- Emotions - experience of positive emotions cluster together, as do negative emotions

More examples

- Social support - can be separated into three factors: support from family, friends and romantic partners
- Sexism - hostile and benevolent sexism
- Social dominance orientation - one factor assessing preference for dominance, one on preference for inequality

Data reduction

- In psychometric instrument development, factor analysis is used to simplify the data structure by identifying a smaller number of underlying factors
- Factor analysis then helps to identify items for improvement or removal because they are:
  - redundant e.g., items which are highly correlated
  - unclear/irrelevant e.g., items which don't load cleanly on a single factor or have low loadings on all factors.
  - complex
- Factor analysis informs the calculation of factor scores (composite scores combine a respondent's scores for several related items)
- Factor analysis helps eliminate redundant variables (e.g., items which are highly correlated), unclear variables (e.g., items which don't load cleanly on a single factor) and irrelevant variables (e.g., variables which have low loadings on all factors)

History

- Factor analysis was invented by Pearson in 1901, then refined by Spearman in 1904
- Early factor analysis involved hand calculation, so it was not used often
- Factor analysis really took off once we had computers to do the analyses for us

Two types of factor analysis

Exploratory factor analysis (EFA)

- Explores and summmarises underlying correlational structure for a dataset

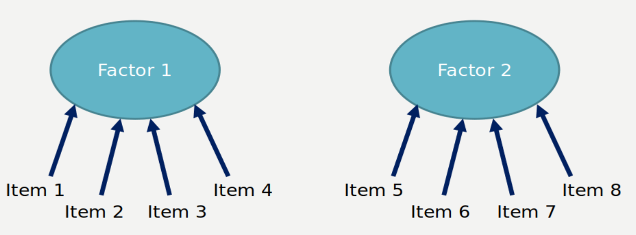
Confirmatory factor analysis (CFA)

- Tests correlational structure of a data set against a hypothesised structure and rates the 'goodness of fit'

What is the difference?

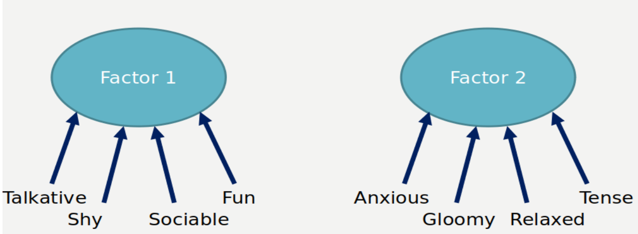
- EFA is used to *explore* the factor structure, and CFA is used to *confirm* it
- CFA is generally preferred, but is more advanced
- In this lecture, I introduce you to EFA as this is an important first step

Simple conceptual model

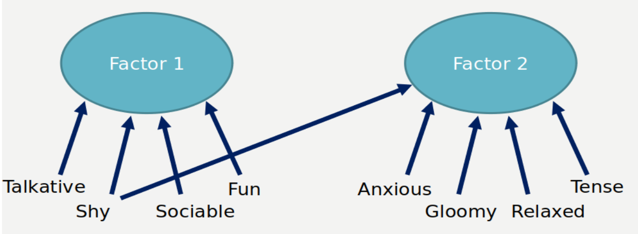


These items ‘load’ onto two underlying factors. The items represent/tap/measure what that construct/factor is. The factors are made up of relatively similar/related variables/items.

Simple conceptual model



Models can be simple or complex



- Simple - each item primarily loads onto one factor only
- Complex - crossloadings, load onto more than one factor

Key questions to ask

- How many factors are there?
- To which factor does each item best belong?

Example

- What is the underlying factor structure of students' attitudes towards debt?
  - Harrison et al. (2015) recruited indebted undergraduates to respond to 20 items on debt
  - Based on EFA (PCA, oblique rotation), four factors were identified, explaining 45% of the variance
    - Anxiety (negative affect associated with indebtedness)
    - Utility for lifestyle (debt useful for maintaining social life, normative)
    - Utility for investment (student debt is an investment in career)
    - Awareness (knowledge of debt and requirements for repayment)

Steps to conduct an EFA

8 key steps

1) Test the assumptions of EFA. Can you do the analysis on your data?
2) Select your method of extraction. Are factors likely to be related?
3) Determine the number of factors. What is your data telling you?
4) Select the items that fit each factor.
5) Name and define each factor. Do these make sense?
6) Examine correlations amongst factors. To what extent are they related to each other?
7) Analyse internal reliability. Do the items in each factor contribute to a reliable subscale?
8) Compute composite scores

Simulated Example

- Let's take a simulated survey dataset of a questionnaire on brain health awareness
- Sample size: $N = 350$ participants
- Items: 12 Likert-style items scored from 1 to 7 (strongly disagree to strongly agree)
- The questionnaire contains three sets of items:
  - Cardiometabolic risk awareness (Cardio1–Cardio4)
  - Cognitive symptom concern (Cogn1–Cogn4)
  - Risk-reduction self-efficacy (SelfE1–SelfE4)
- We will use exploratory factor analysis to see whether these items cluster into underlying factors

But before we even get there...

- Garbage In Garbage Out (GIGO)
- Remember back to previous lectures. What do we do before analysing any data?
  - Check univariate statistics (looking for data errors, outliers) and check assumptions (do we meet the requirements for this test?)

Assumptions

- The following are important to consider:
  - Theory
  - Sample size
  - Level of measurement
  - Normality
  - Linearity
  - Outliers
  - Factorability

Theory

- Your EFA should be driven by theory - avoid fishing expeditions!
- For example, you might be asking:
  - How many distinct dimensions (factors) of X are there what are they, and which items best represent these factors?

Sample size

- Factor analysis requires a bigger dataset than many other analyses
- How many cases of data do you need? This depends on the number of variables you have
  - At a minimum: more than 5 cases per variable (1:5 variables:cases ratio)
  - Ideal: more than 20 cases per variable (1:20 variables:cases ratio)
- The total number of cases should be at least 200, ideally more by Comrey & Lee (2013) guidelines:
  - 50 cases = very poor, 100 = poor, 200 = fair, 300 = good, 500 = very good, 1000+ = excellent

Sample size

- How many cases do people usually use?
- Fabrigar et al. (1999) looked into this and found that in two of the top journals, there was a range of sample sizes used

Variable	*Journal of Personality and Social Psychology*		*Journal of Applied Psychology*	
	N	%	*N*	%
Sample size				
100 or less	30	18.9	8	13.8
101–200	44	27.7	14	24.1
201–300	25	15.7	9	15.5
301–400	13	8.2	2	3.4
More than 400	47	29.6	25	43.1

Level of measurement

- In order to be included in an exploratory factor analysis, your variables must be suitable for Pearson product-moment correlational analyses
  - Interval or ratio level of measurement

Normality

- Factor analysis can cope with minor violations of normality (i.e. it is relatively robust)
- However, it works best if the variables are normally distributed

Outliers

- Factor analysis is sensitive to outliers including:
  - Bivariate outliers
  - Multivariate outliers (e.g. someone who is 18, post-graduate and has 3 children)
- Identify outliers (e.g. scatterplots, Mahalanobis distance), remove them or recode them if they are influential

Linearity

- Factor analysis is based on correlations between variables
- Remember that correlations are a measure of the linear association between variables
- It is important to check that the variables are linearly related
  - Check the scatterplots

Factorability

- Factorability assesses whether there are sufficient intercorrelations amongst the items to warrant factor analysis.
- Assess factorability via one or more of:
  - Correlation matrix correlations $> .3$?
  - Anti-image matrix diagonals $> .5$?
  - Measures of sampling adequacy (MSAs)?
    - Bartlett's significance?
    - KMO $> .5$ or $.6$?

Check the correlation matrix

- Is the correlation matrix of our dataset factorable?
  - Correlation matrix correlations $> .3$

	Cardio1	Cardio2	Cardio3	Cardio4	Cogn1	Cogn2	Cogn3	Cogn4	SelfE1	SelfE2	SelfE3	SelfE4
Cardio1	1											
Cardio2	0.56	1										
Cardio3	0.52	0.57	1									
Cardio4	0.41	0.43	0.47	1								
Cogn1	-0.16	-0.16	-0.23	-0.41	1							
Cogn2	0.21	0.21	0.16	0.32	-0.42	1						
Cogn3	0.14	0.14	0.17	0.34	-0.46	0.49	1					
Cogn4	0.17	0.19	0.19	0.31	-0.47	0.45	0.45	1				
SelfE1	0.09	0.2	0.15	0.24	-0.23	0.19	0.17	0.11	1			
SelfE2	0.13	0.23	0.18	0.19	-0.2	0.1	0.14	0.14	0.5	1		
SelfE3	0.15	0.26	0.22	0.2	-0.17	0.13	0.12	0.11	0.5	0.45	1	
SelfE4	0.15	0.2	0.15	0.18	-0.13	0.1	0.01	-0.01	0.39	0.29	0.42	1

Check the anti-image correlation matrix

- Anti-image matrix diagonals $> .5$?

```
## Cardio1 Cardio2 Cardio3 Cardio4   Cogn1   Cogn2   Cogn3   Cogn4  SelfE1  SelfE2
##    0.81    0.81    0.83    0.89    0.84    0.83    0.83    0.83    0.79    0.82
##  SelfE3  SelfE4
##    0.82    0.81
```

Statistical measures

- The correlation matrix is factorable if:
  - The Bartlett's test of sphericity is significant OR
  - The Kaiser-Mayer Olkin (KMO) is greater than .5 or .6
- This is the fastest method, but least reliable

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: psych::KMO(r = X)
## Overall MSA =  0.83
## MSA for each item =
## Cardio1 Cardio2 Cardio3 Cardio4   Cogn1   Cogn2   Cogn3   Cogn4  SelfE1  SelfE2
##    0.81    0.81    0.83    0.89    0.84    0.83    0.83    0.83    0.79    0.82
##  SelfE3  SelfE4
##    0.82    0.81

## $chisq
## [1] 1218.852
##
## $p.value
## [1] 1.122989e-211
##
## $df
## [1] 66
```

Your correlation matrix is factorable if:

- Several correlations are above .3
- Anti-image correlation matrix diagonals are above .5
- Bartlett's test is significant
- KMO is greater than .5 or .6

Which extraction method do you use?

- There are two main approaches to extracting factors:
  - Principal axis factoring (PAF) analyses shared variance
  - Principal components analysis (PCA) analyses all variance

Principal axis factoring

- Principal axis factoring is used to discover the underlying structure of a set of variables
- It is theory driven
- It analyses only the common (shared) variance, so leaves out variance that is unique to each measurement item

Principal components

- Principal components analysis is more common - it is used to reduce many variables down to a smaller number of factor scores, which can then be used in other analyses
- It analyses all of the variance in each variable (common and unique)

Variance

- The variance of a variable is made up of:
  - Common variance that it shares with other variables
  - Unique variance that is not shared with other variables

Which should you use?

- Often there is little difference between principal components and principal axis factoring
- If they come to different solutions, try to work out why, and decide which is more appropriate for your data

The goal = explain the variance

- A good factor solution is one that explains the lion's share of the variance with the fewest factors
  - In general, we tend to be happy if we can explain 50 to 75% of the variance (but don't be too strict on this rule)

Example of total variance explained

```
##                         PA1  PA2  PA3
## SS loadings            3.32 1.31 1.11
## Proportion Var         0.28 0.11 0.09
## Cumulative Var         0.28 0.39 0.48
## Proportion Explained   0.58 0.23 0.19
## Cumulative Proportion  0.58 0.81 1.00
```

Communalities

- The communality of a variable is the proportion of the variable's variance explained by the extracted factors
- Communalities range from 0 (no variance explained) to 1 (all variance explained). To interpret this value:
  - If it is greater than .5, this indicates high communality -> the extracted factors explain most of the variance in the variable
  - If it is lower than .5, this indicates low communality -> considerable variance in the variable is unexplained by the extracted factors.
    - Either extract more factors or eliminate the item.

Example

	Communality
Cardio1	0.52
Cardio2	0.60
Cardio3	0.54
Cardio4	0.45
Cogn1	0.48
Cogn2	0.43
Cogn3	0.50
Cogn4	0.45
SelfE1	0.57
SelfE2	0.40
SelfE3	0.50
SelfE4	0.30

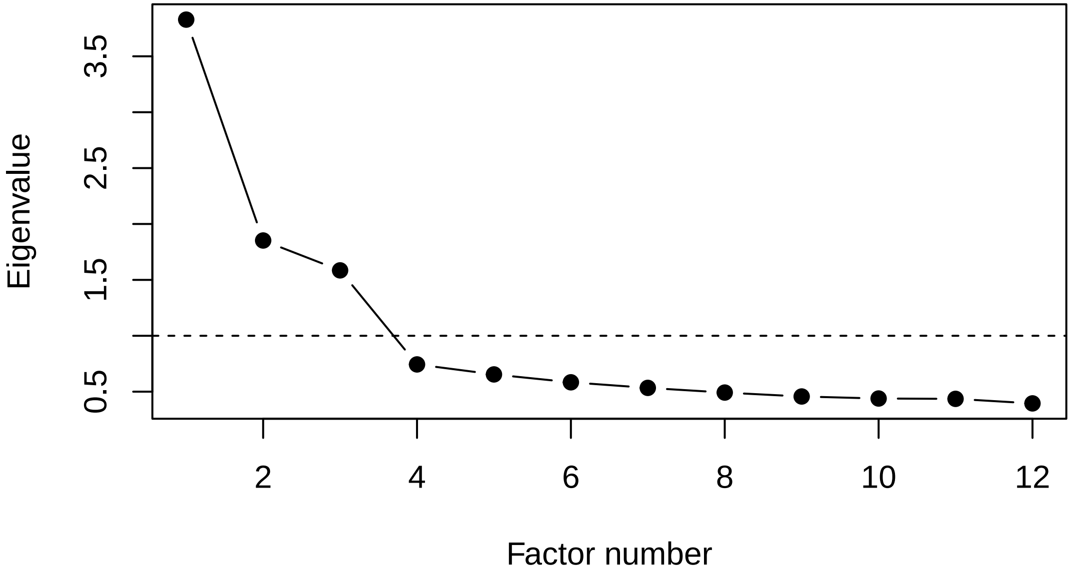
Eigenvalues

- Each variable contributes to the variance that needs to be explained
- Each factor tries to explain as much of the total variance as possible
- An eigenvalue indicates the amount of overall variance that each factor accounts for
- Rule of thumb: Eigenvalues over 1 are “stable” (Kaiser’s criterion)
- Eigenvalues for successively extracted factors have lower values
- Eigenvalues can be usefully expressed as percentages of explained variance
- Total of all eigenvalues = the number of variables = or 100%

Example of total variance explained

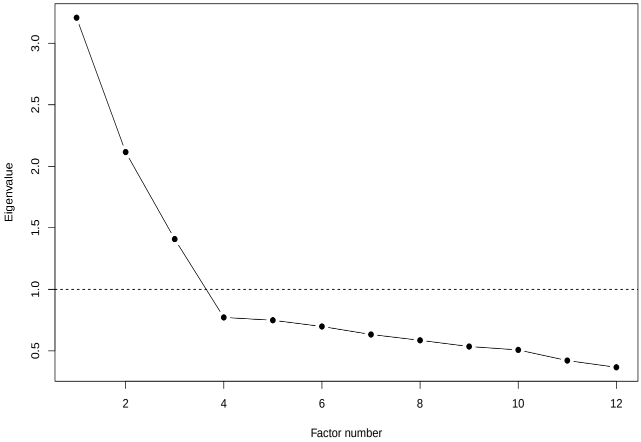
Factor	Eigenvalue
1	3.83
2	1.85
3	1.59
4	0.74
5	0.65
6	0.58
7	0.53
8	0.49
9	0.46
10	0.44
11	0.44
12	0.40

Check the scree plot



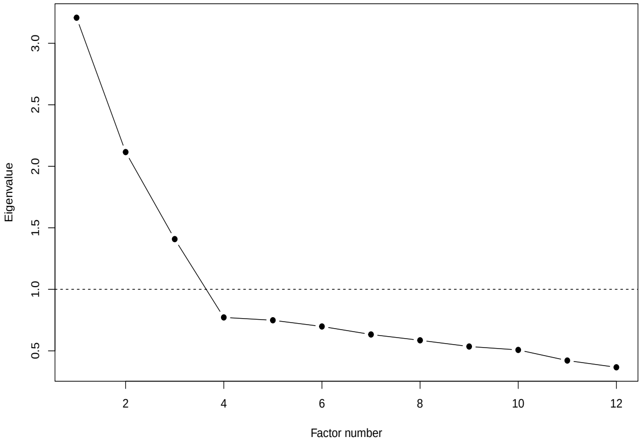
Example

- Based on this scree plot, how many factors would you retain?
  - 2 factors?
  - 3 factors?
  - 4 factors?



Example

- 3 factors
  - There is a clear elbow after the third factor
  - The first three factors explain most of the common variance
  - Factors after that contribute relatively little



Scree plot

- A scree plot is a line graph of all eigenvalues
- It shows how much variance is explained by each factor, from the most to the least
- You use it to decide what the best number of factors is
  - Look for the elbow - the point where additional factors do not account for much additional variance

Practice question

- An exploratory factor analysis of 20 variables indicates that 4 factors explain 60% of the variance. What do the eigenvalues of factors 5 to 20 add up to?

a) impossible to tell
b) 8
c) 12
d) 20

Practice question

- An exploratory factor analysis of 20 variables indicates that 4 factors explain 60% of the variance. What do the eigenvalues of factors 5 to 20 add up to?

a) impossible to tell
b) **8**
c) 12
d) 20

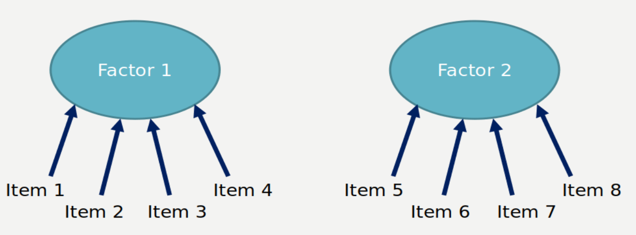
How many factors do I have?

- Jamovi will give you the same number of factors as variables that you have
- There is no definitive statistic that tells you. Make your decision based on:
  - Which solution explains the largest amount of variance with the smallest number of factors
  - What does the theory say? (predictions/expectations)
  - Which Eigenvalues are greater than 1?
  - Where is the elbow in the scree plot?
  - How interpretable is your last factor?
  - Which solution fits best? (type of EFA, rotation, number of factors)
- Your factors must be interpretable and make theoretical sense.

Also keep in mind

- Remember than an exploratory factor analysis will keep going until it has extracted the same number of factors as you have variables
- You should be aiming for a solution that explains 50 - 75% of the variance explained by $\frac{1}{4}$ to $\frac{1}{3}$ as many factors as variables
- Stop extracting factors when they no longer represent useful/meaningful clusters of variables

Simple conceptual model



These items 'load' onto two underlying factors. The items represent/tap/measure what that construct/factor is. The factors are made up of relatively similar/related variables/items.

Factor loading matrix

- A factor loading is the relationship each variable has with the factor

The first solution is unrotated

- In the initial solution, each factor “selfishly” grabs maximum unexplained variance.
- 1^{st} factor extracted:
  - Best possible line of best fit through the original variables
  - Seeks to explain lion’s share of all variance
  - Gives the best single factor summary of the variance in the whole set of items
  - All variables will tend to load strongly on the 1^{st} factor
- 2^{nd} factor extracted:
  - Seeks to maximise its own Eigenvalue by trying to explain as much of the remaining unexplained variance as possible.
- Each subsequent factor does the same, until all variance is explained

Rotating the solution

- A rotated solution is generally easier to interpret
- Rarely see a simple unrotated factor structure
- Many variables will load on to two or more factors

Orthogonal

- Varimax in jamovi
- Minimises the factor covariation
- Produces uncorrelated factors

Oblique

- Oblimin in jamovi
- Allows factors to covary
- Allows correlations between factors

Orthogonal vs oblique

- Are you expecting factors to be related or unrelated?
  - If related, choose oblique
  - If unrelated, choose orthogonal
  - If unsure, choose oblique, then switch to orthogonal if the correlations between factors are smaller than .3
  - You can try both, then see which solution is most interpretable

Interpretability matters

- Please don’t just go with the solution that has the ‘best’ factor loadings
- Think it through
- Be guided by theory and common sense
- Every factor needs to be interpretable

A word of caution

- Because it is a subjective process, this means we can also be swayed to ‘see what we want to see’
- Be aware of this, and avoid it
- If the evidence is suggesting a different solution that fits the data better, try and see if this also makes sense
- It is possible that there is more than one good solution e.g. this is what we get in personality (3 and 5 factor models, even 16 factor models that work)

Keep it simple

UNIVERSITY OF CANBERRA

- The simpler your factor structure, the easier it will be to interpret
- Some more criteria:
  - Each variable should load strongly (about .40 or higher) on only one factor
  - Each factor should have three or more strong loadings - the more it has, the more reliable it will be
  - Most loadings should either be high or low (few intermediate values), so they either load or they don't

Refining your factors

- You may need to eliminate some items to refine your factors
- Again, this is a subjective process
- Some points to consider for each item:
  - How big is the item's main loading? (should be at least .4)
  - How big is its cross loadings? (should be lower than .3)
  - What is the meaning of the item? What contribution does it make to the factor? Does it make sense that it would belong to this factor?
- Eliminate one variable at a time, then re-run the exploratory factor analysis

Refining your factors

- Comrey & Lee (2013) offer the following guidelines for assessing factor loadings of an item
  - Less than .32 = poor
  - Greater than .45 = fair
  - Greater than .55 = good
  - Greater than .63 = very good
  - Greater than .70 = excellent
- It is also important to judge this in relation to other items in the factor - e.g. if most are loading excellently and the item in question is fair, that's a big gap in the loadings
- Is the factor interpretable without the item?

Steps

1) Try both principal components and principal axis factoring methods of extractions
2) Run orthogonal and oblique rotations
3) Test out a range of possible factor structures (2, 3, 4 . . . N factors)
4) Remove poor items one at a time, then re-run the analyses
5) Check reliability analyses (next lecture)

Example

- Student adjustment to university survey
- Participants responded to 12 items
- Responses were on a 7-point Likert scale from strongly disagree to strongly agree
- The items were designed to capture:
  - Academic stress
  - Sense of belonging
  - Self-regulated learning

Assumptions

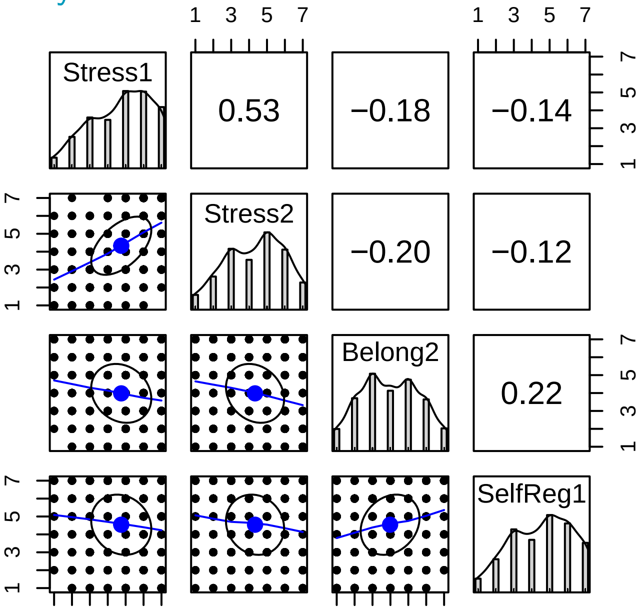
- The following are important to consider:
  - Theory: is there a good reason to ask this research question?
  - Sample size: N = 1200. Is the cases:variables ratio sufficient?
  - Level of measurement: Likert scale with 7 response options - is that sufficient?

Assumptions - normality

- Factor analysis can cope with minor violations of normality (i.e. it is relatively robust)
- However, it works best if the variables are normally distributed

	Mean	SD	Skew	Kurtosis
Stress1	4.75	1.68	-0.40	-0.83
Stress2	4.33	1.63	-0.21	-0.85
Stress3	4.01	1.63	0.03	-0.97
Stress4	3.93	1.57	0.08	-0.81
Belong1	3.33	1.69	0.40	-0.84
Belong2	3.98	1.65	0.03	-0.93
Belong3	3.64	1.60	0.16	-0.89
Belong4	3.96	1.71	0.02	-0.97
SelfReg1	4.54	1.68	-0.25	-0.90
SelfReg2	4.06	1.73	0.02	-0.99
SelfReg3	4.11	1.61	-0.13	-0.87
SelfReg4	3.55	1.62	0.24	-0.81

Outliers and Linearity



Factorability

- Bartlett's significance?
- KMO $>$ .5 or .6?

```
## [1] 0.827098
```

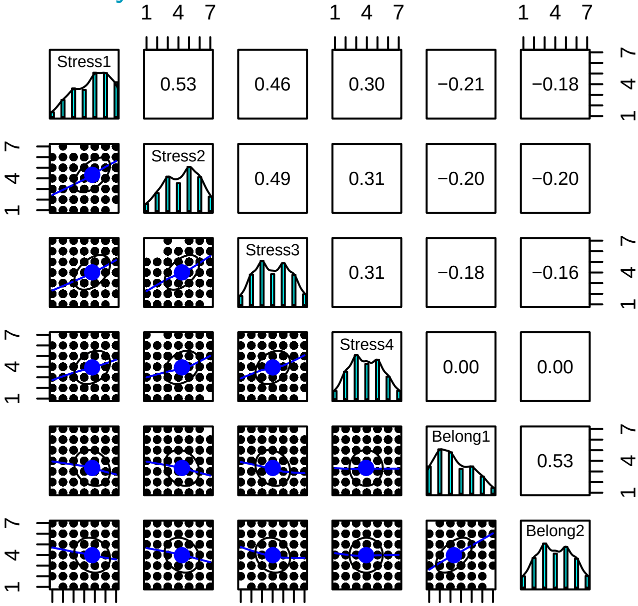
```
## $chisq
## [1] 3642.709
##
## $p.value
## [1] 0
##
## $df
## [1] 66
```

Factorability

- Correlation matrix correlations $> .3$?

	Stress1	Stress2	Stress3	Stress4	Belong1	Belong2	Belong3	Belong4	SelfReg1	SelfReg2	SelfReg3	SelfReg4
Stress1	1											
Stress2	0.53	1										
Stress3	0.46	0.49	1									
Stress4	0.3	0.31	0.31	1								
Belong1	-0.21	-0.2	-0.18	0	1							
Belong2	-0.18	-0.2	-0.16	0	0.53	1						
Belong3	-0.2	-0.22	-0.18	-0.01	0.46	0.44	1					
Belong4	0.25	0.28	0.21	0.03	-0.45	-0.45	-0.43	1				
SelfReg1	-0.14	-0.12	-0.1	-0.04	0.22	0.22	0.18	-0.21	1			
SelfReg2	-0.19	-0.16	-0.13	-0.02	0.32	0.29	0.26	-0.27	0.52	1		
SelfReg3	-0.13	-0.05	-0.1	-0.02	0.14	0.13	0.12	-0.12	0.46	0.44	1	
SelfReg4	-0.09	-0.09	-0.1	-0.01	0.17	0.15	0.11	-0.14	0.44	0.37	0.34	1

Factorability and linearity



The key steps

1) Test the assumptions of exploratory factor analysis. Can you do the analysis on your data?

- Yes - the items are sufficiently correlated and the sample size is adequate

2) Select your method of extraction. Are factors likely to be related?

- Factors could be related - oblique/oblimin rotation
- If the goal is to identify underlying structure, principal axis factoring is a reasonable choice

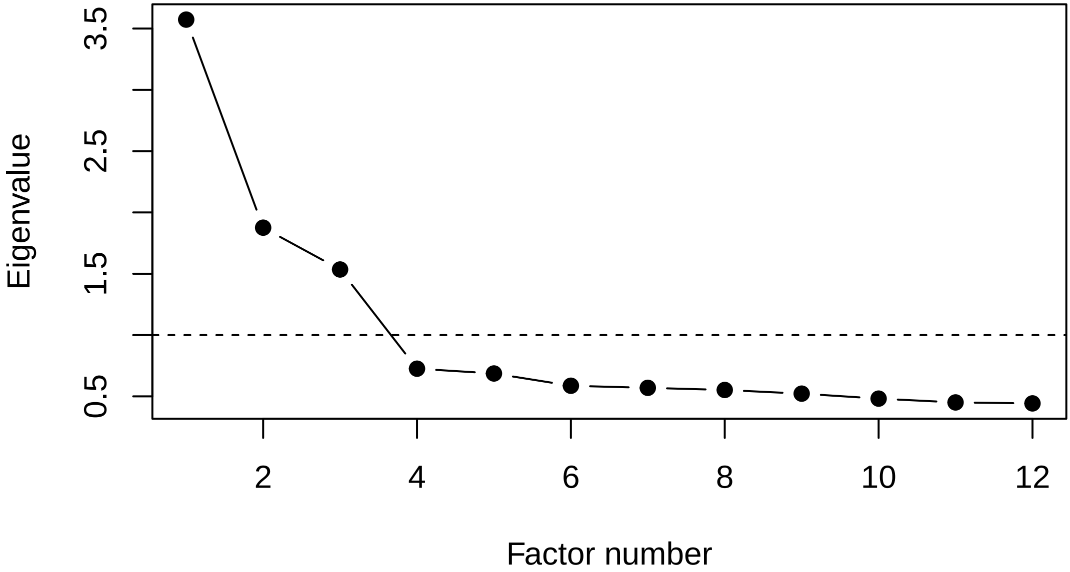
The key steps

3) Determine the number of factors

- What is your data telling you?

4) Select the items that fit each factor
5) Name and define each factor. Do these make sense?

Scree plot



Eigenvalues

Factor	Eigenvalue
1	3.57
2	1.88
3	1.53
4	0.73
5	0.69
6	0.59
7	0.57
8	0.55
9	0.52
10	0.48
11	0.45
12	0.44

Table of factor loadings

UNIVERSITY OF CANBERRA

	PA1	PA2	PA3
Stress1	-0.02	0.69	-0.04
Stress2	-0.06	0.72	0.03
Stress3	0.00	0.66	-0.01
Stress4	0.18	0.51	0.00
Belong1	0.72	0.02	0.02
Belong2	0.72	0.03	0.01
Belong3	0.64	-0.03	-0.01
Belong4	-0.60	0.10	-0.01
SelfReg1	-0.01	0.00	0.76
SelfReg2	0.15	-0.01	0.64
SelfReg3	-0.08	0.00	0.65
SelfReg4	-0.02	0.00	0.56

The key steps

3) Determine the number of factors

- What is your data telling you?
  - A 3-factor solution appears most interpretable

4) Select the items that fit each factor

- Stress1–Stress4 mainly fit factor 2
- Belong1–Belong4 mainly fit factor 1
- SelfReg1–SelfReg4 mainly fit factor 3

5) Name and define each factor. Do these make sense?

- Factor 1 represents sense of belonging
- Factor 2 represents academic stress
- Factor 3 represents self-regulated learning

The key steps

6) Examine the correlations amongst factors. To what extent are they related to each other

```
##       PA1   PA3   PA2
## PA1  1.00  0.41 -0.37
## PA3  0.41  1.00 -0.22
## PA2 -0.37 -0.22  1.00
```

The key steps

7) Analyse internal reliability.

- Do the items in each factor contribute to a reliable subscale?
- Need to first ensure all variables are coded in the same direction

Scale	Alpha
Academic stress	0.73
Sense of belonging	0.77
Self-regulated learning	0.75

Summary

What is factor analysis?

- Factor analysis is a family of multivariate correlational data analysis methods for summarising clusters of covariance
- Factor analysis summarises correlations amongst items
- The common clusters (called factors) indicate underlying fuzzy constructs

How do I do it?

1) Examine assumptions
2) Choose extraction method and rotation
3) Determine number of factors
4) Select items
5) Name and describe factors
6) Examine correlations amongst factors
7) Analyse internal reliability (next week!)
8) Compute composite scores (next week!)

What are the assumptions?

UNIVERSITY OF CANBERRA

- Check your sample size in relation to the number of variables you have (at least 5 cases per variable, ideally 20)
- Check for outliers
- Check that your correlation matrix is factorable
- Check for normality

Types of...

Factor analysis

- Principal axis factoring uses shared variance for theoretical exploration
- Principal components uses all variance for data reduction

Rotation

- Orthogonal for uncorrelated factors
- Oblique for correlated factors

The number of factors depends on...

- The eigenvalues (look at values above 1 and for the elbow in the scree plot)
- The percentage of variance explained (aim for 50 - 75%)
- How interpretable each factor is
- How the factors fit with theory

Selecting items depends on...

- An exploratory factor analysis of a good measurement instrument ideally has:
  - A simple factor structure (each variable loads strongly ($>$ .50) on only one factor)
  - Each factor has multiple loading variables (more loadings -> greater reliability)
  - Target factor loadings are high ($>$ .5) and cross-loadings are low ($<$.3), with few intermediate values (.3 to .5)

Next week - psychometric instrument development

- What concepts are and how to measure them
- Measurement error
- Psychometrics
- Reliability and validity
- Creating composite scores

Contributions to this course

Dr James Neill

Dr Samantha Stanley

Dr Jeroen van Boxtel

References I

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*. Psychology press. https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9781315827506&type=googlepdf

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Harrison, N., Agnew, S., & Serido, J. (2015). Attitudes to debt among indebted undergraduates: A cross-national exploratory factor analysis. *Journal of Economic Psychology*, *46*, 62–73. https://doi.org/10.1016/j.joep.2014.11.005