Exploratory Factor Analysis

Image source: https://commons.wikimedia.org/wiki/File:Pleiades_large.jpg

Lecture 5

Survey Research & Design in Psychology

James Neill, 2018

Readings

1 Beavers et al. (2013). Practical considerations for using EFA in educational research. [Online]
2 Fabrigar et al. (1999). Evaluating the use of EFA in psychological research. [Online]
3 Floyd & Widaman (1995). Factor analysis in the development and refinement of clinical assessment instruments. [Online]
4 Howitt & Cramer (2014). Ch 31: Factor analysis: Simplifying complex data. [Textbook/UCLearn Reading List]
5 Streiner (1994). Figuring out factors: The use and misuse of factor analysis. [Online]
6 Tabachnick & Fidell (2007). Principal components and factor analysis. [UCLearn Reading List]
7 Williams, Brown, & Osman (2012). EFA: A five-step guide for novices. [Online]
8 Wikiversity (2017). EFA: Glossary. [Online]

2

Overview

1 Intro to factor analysis
2 EFA examples
3 Steps / process
4 Assumptions

3

Intro to Factor Analysis

4

Intro to factor analysis

1 What is it?
2 Purposes
3 History
4 Types
5 Models

5

Astronomy metaphor

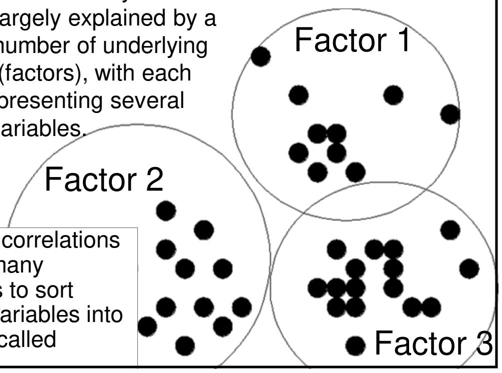
Factor analysis helps to identify the correlational structure amongst a set of variables.

Universe = **Topic** e.g., Time management
Galaxies = **Factors** e.g., Procrastination, Planning, Efficiency etc.
Solar systems = **Items** e.g., "I get easily distracted."

https://commons.wikimedia.org/wiki/File%3AEarth's_Location_in_the_Universe_SMALLER_(JPEG).jpg

Conceptual model of factor analysis

The variance of many variables may be largely explained by a smaller number of underlying clusters (factors), with each factor representing several related variables.



FA uses correlations among many variables to sort related variables into clusters called "factors"

Factor analysis is...

- a family of multivariate statistical techniques used for examining correlations amongst variables.
- for identifying clusters of inter-correlated variables (called 'factors').

8

Purposes

Main applications of factor analysis:

1. **Theory development**:
   Examine the hypothetical structure of relations between constructs, identify factors, and classify variables.
2. **Data reduction**:
   Reduce the number of 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.

9

Purposes: Theory development

- FA is used to test theoretical models by investigating the underlying correlational pattern shared by the variables.
- The goal is to address a theoretical question such as:
  - How many personality factors are there? (and what are they?)
  - Is intelligence general or multiple?

10

Example: Personality

How many dimensions of personality are there – and what are they?

e.g., FA can help to decide between 3 or 5 factor personality models:

Eysenck's 3?
- Extraversion
- Neuroticism
- Psychoticism

Big 5?
- Neuroticism
- Extraversion
- Agreeableness
- Openness
- Conscientiousness

11

Example: Intelligence

Is intelligence better described as:
- one global factor (g) or
- several specific factors (e.g., verbal, spatial, mathematical, social, kinaesthetic)?

FA can help decide which model is best supported by evidence.

12

Purposes: Data reduction

- In psychometric instrument development, FA is used to simplify the data structure by identifying a smaller number of underlying factors.
- FA then helps to identify items for improvement or removal because they are:
  - redundant, or
  - unclear/irrelevant, or
  - complex
- FA informs the calculation of factor scores, (composite scores combine a respondent's scores for several related items).

13

History of factor analysis

(Goldberg & Digman, 1994, cited in Fehriinger, 2004)

- Invented by Pearson (1901) and further developed by Spearman (1904)
- Usage hampered by onerousness of hand calculation
- Since the advent of computers, usage has thrived, especially for:
  - Theory e.g., determining the structure of psychological constructs such as personality or intelligence
  - Practice e.g., development of 10,000s+ of psychological screening & measurement tests

14

Types of factor analysis

EFA = **Exploratory Factor Analysis**

- explores & summarises underlying correlational structure for a data set

CFA = **Confirmatory Factor Analysis**

- tests correlational structure of a data set against a hypothesised structure and rates the “goodness of fit”

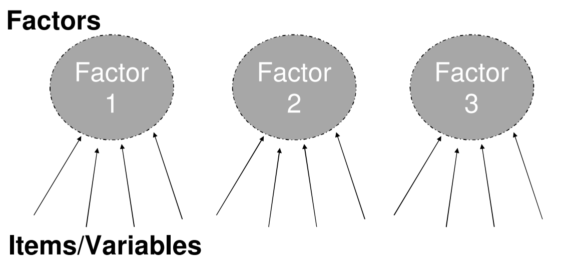
15

EFA vs. CFA

This (introductory) lecture focuses on **Exploratory Factor Analysis** (recommended for undergraduate level).
Confirmatory Factor Analysis is now generally preferred, but is more advanced (recommended for graduate/professional level).

16

Simple conceptual model

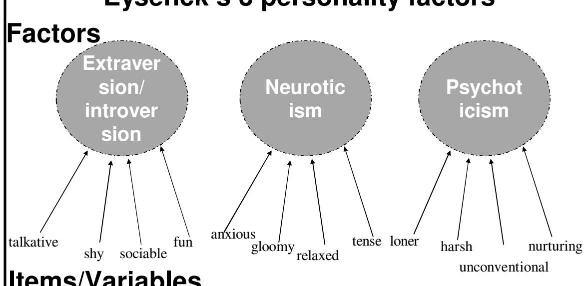


- e.g., 12 variables which "tap" (represent) 3 underlying factors
- Factors consist of relatively similar/related variables.

17

Simple conceptual model

Eysenck's 3 personality factors

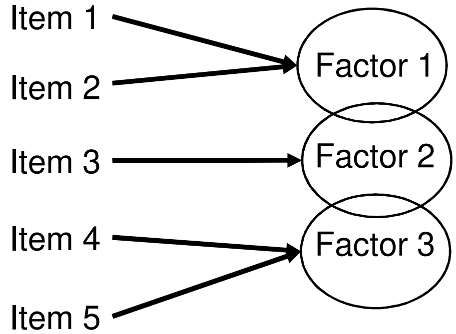


e.g., these 12 items measure 3 underlying dimensions of personality

18

Simple conceptual model

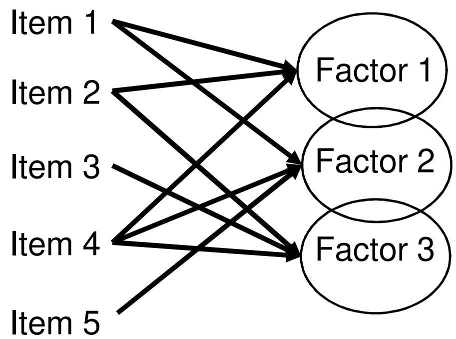
Each measurement item primarily loads onto only one factor.



19

Complex conceptual model

Each measurement item may load onto more than one factor.

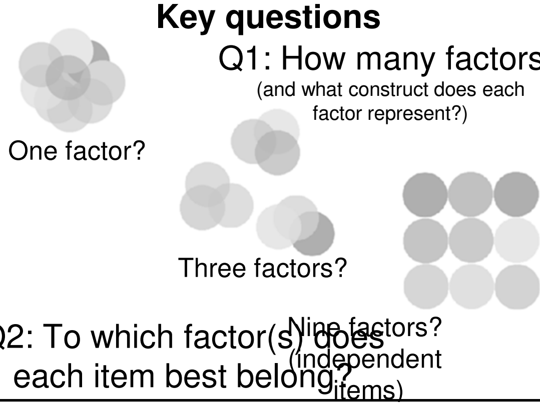


20

Exploratory factor analysis: Key questions

Q1: How many factors?

(and what construct does each factor represent?)



Q2: To which factor(s) does each item best belong?

EFA example 1: Essential facial features

22

Example: Essential facial features

What are the essential facial features for recognition of expression and communication?

(Ivancevic et al., 2003)

Example: Essential facial features

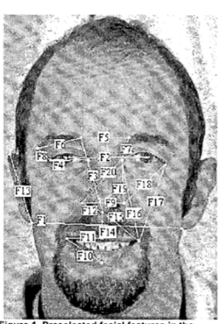
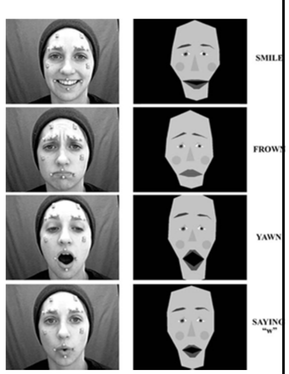


Figure 1. Preselected facial features in the sample of 80 facial images from the test database



(Ivancevic et al., 2003)

Example: Essential facial features

- The importance of 20 facial features in facial recognition was measured with 80 facial images.
- Based on EFA (PC, orthogonal), 6 factors were identified, representing 76.5% of the total variability in facial recognition:
  1. upper-lip
  2. eyebrow-position
  3. nose-width
  4. eye-position
  5. eye/eyebrow-length
  6. face-width

25

EFA example 2: Classroom behaviour

26

Classroom behaviour

Francis (2007) - based on the Victorian Quality Schools Project

- 15 classroom behaviours of high-school students were rated by teachers using a 5-point Likert scale.
- Task: Identify groups of variables (behaviours) that are strongly inter-related and represent underlying factors.

27

Classroom behaviour: Items

Teachers, for each of the following paired behavioral statements, please mark a cross over the dot (e.g. ~~o~~) which is **nearest** the statement that **best** describes the TYPICAL behavior of THIS student at school

1. Cannot concentrate on any particular task; easily distracted	o	o	o	o	o	Can concentrate on any task; not easily distracted
2. Perseveres in the face of difficult or challenging tasks	o	o	o	o	o	Lacks perseverance; is impatient with difficult or challenging tasks
7. Persistent, sustained attention span	o	o	o	o	o	Easily frustrated; short attention span
10. Purposeful activity	o	o	o	o	o	Aimless; impulsive activity

Classroom behaviour: Items

1. Cannot concentrate ↔ can concentrate
2. Curious & enquiring ↔ little curiousity
3. Perseveres ↔ lacks perseverance
4. Irritable ↔ even-tempered
5. Easily excited ↔ not easily excited
6. Patient ↔ demanding
7. Easily upset ↔ contented

29

Classroom behaviour: Items

8. Control ↔ no control
9. Relates warmly to others ↔ disruptive
10. Persistent ↔ frustrated
11. Difficult ↔ easy
12. Restless ↔ relaxed
13. Lively ↔ settled

30

Classroom behaviour

- Results are embedded in subsequent slides
- See also: Tutorial 03: Psychometrics: EFA Exercise 2: Classroom behaviour
  - https://en.wikiversity.org/wiki/Survey_research_and_design_in_psychology/Tutorials/Psychometrics/Exploratory_factor_analysis#Exercise_2:_Classroom_behaviour

31

Steps / process

32

Steps / process

1 Test assumptions
2 Select extraction method
3 Determine # of factors
(Eigen Values, % variance explained, scree plot)
4 Select items
(check factor loadings to identify which items belong best in which factor; drop items one by one; repeat)
5 Name and define factors
6 Examine correlations amongst factors
7 Analyse internal reliability
8 Compute composite scores

33

Garbage. In. →
Garbage. Out

… screen the data ...

34

Assumption testing

1 Theory
2 Sample size
3 Level of measurement
4 Normality
5 Linearity
6 Outliers
7 Factorability

35

Assumption testing: Theory

EFA should be driven by a theoretically-driven research question e.g.,

“How many distinct dimensions (factors) of X are there, what are they, and which items best represent these factors?”

36

Assumption testing: Sample size

- FA is "data hungry"
- Some guidelines:
  - Minimum:
    N > 5 cases per variable
    e.g., 12 variables, should have > 60 cases (1:5)
  - Ideal:
    N > 20 cases per variable
    e.g., 12 variables, ideally have > 240 cases (1:20)
  - Total:
    N > 200 preferable

37

Assumption testing: Sample size

Comrey and Lee's (1992) guidelines:

50 = very poor
100 = poor
200 = fair
300 = good
500 = very good
1000+ = excellent

38

Assumption testing: Sample size

Fabrigar et al. (1999)

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

Assumption testing: Level of measurement

- All variables must be suitable for Pearson product-moment correlational analysis

  i.e., the variables should have interval or ratio levels of measurement.

40

Assumption testing: Normality

- FA is generally robust to minor violation of assumptions of normality.
- If the variables are normally distributed then the solution is enhanced.

41

Assumption testing: Outliers

- FA is sensitive to outlying (unusual) cases, including:
  - Bivariate outliers
    (e.g., check scatterplots)
  - Multivariate outliers
    (e.g., Mahalanobis' distance)
- Identify outliers, then remove or recode if they are influential

42

Assumption testing: Linearity

- FA is based on correlations between variables, so it is important to check there are linear relations amongst the variables (i.e., check scatterplots)

43

Assumption testing: 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 sig.?
  - KMO > .5 or .6?

44

Assumption testing: Factorability (Correlations)

To be factorable: Are there SEVERAL correlations over .3?
If so, proceed with EFA.

Correlation Matrix

		CONCENTRATES	CURIOUS	PERSEVERES	EVEN-TEMPERED	PLACID
Correlation	CONCENTRATES	1.000	.717	.751	.554	.429
	CURIOUS	.717	1.000	.826	.472	.262
	PERSEVERES	.751	.826	1.000	.507	.311
	EVEN-TEMPERED	.554	.472	.507	1.000	.610
	PLACID	.429	.262	.311	.610	1.000

Takes some effort with a large number of variables, but is the most accurate

Assumption testing: Factorability

Anti-image correlation matrix

- Examine the diagonal values on the **anti-image correlation** matrix
- Variables with AI correlations less than .5 should be noted for possible exclusion because they may lack sufficient correlation with other variables
- Medium amount of effort, and reasonably accurate

46

Anti-image correlation matrix

.973[a]	-.141	-.180	.001	.002
-.141	.937[a]	-.452	.018	.052
-.180	-.452	.941[a]	-.028	.034
.001	.018	-.028	.945[a]	-.200
.002	.052	.034	-.200	.944[a]

Check anti-image CORRELATION (not COVARIANCE) matrix

Assumption testing: Factorability

Measures of sampling adequacy

- The correlation matrix is factorable if either of these global indicators:
  - Bartlett's test of sphericity is significant and/or
  - Kaiser-Mayer Olkin (KMO) > .5 or .6
- Quickest method, but least reliable

48

Assumption testing: Factorability

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.956
Bartlett's Test of Sphericity	Approx. Chi-Square	19654.15
	df	105
	Sig.	.000

> .5 or .6

Significant $p < .05$

Summary: Measures of factorability

Use any of the following to determine the factorability of a correlation matrix:

1 Several correlations > .3?
2 Anti-image correlation matrix diagonals > .5?
3 Bartlett's test significant?
4 KMO > .5 to .6?
(depends on whose rule of thumb)

50

Extraction method

Two main approaches to EFA:

- Principal Axis Factoring (PAF)
  Analyses **shared** variance
- Principal Components (PC)
  Analyses **all** variance

51

Principal axis factoring (PAF)

- Purpose: Discover the underlying structure of a set of variables
- Theory-driven
- Analyses only common (shared) variance
  (i.e., leaves out variance that is unique to each measurement item)

52

Principal components (PC)

- More commonly used
- Purpose: Reduce many variables down to a smaller number of factor scores. These scores can be used in other analyses (e.g., for hypothesis testing).
- Analyses all the variance in each variable (common and unique)

53

Variance components

Total variance of a variable

Common variance (shared with other variables)	**Unique variance** (not shared with other variables)
PAF	PC

54

PC vs. PAF

- In practice, try both PC and PAF.
- Often there is little difference between PC and PAF solutions.
- If you get different solutions, try to work out why and decide on which solution is more appropriate.

55

Explained variance

- A good factor solution is one that explains the lion's share of the variance with the fewest factors
- Realistically, researchers are happy with 50 to 75% of the variance explained

56

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	9.355	62.366	62.366	9.094	60.628	60.628	7.801
2	1.532	10.216	72.583	1.294	8.625	69.253	7.261
3	.933	6.220	78.802	.635	4.232	73.485	5.732
4	.467	3.113	81.915				
5	.378	2.519	84.434				
6	.344	2.295	86.729				
7	.305	2.032	88.761				
8	.285	1.902	90.663				
9	.262	1.745	92.408				
10	.229	1.525	93.933				
11	.219	1.459	95.392				
12	.201	1.340	96.732				
13	.184	1.227	97.959				
14	.159	1.059	99.018				
15	.147	.982	100.000				

Extraction Method: Principal Axis Factoring.

a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

3 factors explain 73.5% of the variance in the 15 classroom behaviour items – very useful!

Communalities

- Each variable has a communality
  - which indicates the proportion of the variable's variance explained by the extracted factors
- Communalities can range between
  - 0 (no variance explained)
  - 1 (all variance explained)

58

Communalities

- High communalities (> .5): Extracted factors explain most of the variance in the variable
- Low communalities (< .5): A variable has considerable variance unexplained by the extracted factors. Consider:
  - Extracting more factors
  - Eliminating the item

59

Communalities > .5 for all variables

	Initial	Extraction
behav1 CONCENTRATES	.713	.746
behav2 CURIOUS	.743	.788
behav3 PERSEVERES	.766	.811
behav4 EVEN-TEMPERED	.729	.747
behav5 PLACID	.609	.664
behav6 COMPLIANT	.687	.710
behav7 SELF-CONTROLLED	.730	.749
behav8 RELATES-WARMLY	.605	.660
behav9 SUSTAINED ATTENTION	.776	.803
behav10 COMMUNICATIVE	.657	.674
behav11 RELAXED	.786	.820
behav12 CALM	.737	.786
behav13 PURPOSEFUL ACTIVITY	.764	.798
behav14 COOPERATIVE	.626	.647
behav15 CONTENTED	.595	.621

Extraction Method: Principal Axis Factoring.

Eigen Values (EVs)

- 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 EV indicates the amount of overall variance that each factor accounts for.
- Rule of thumb: Eigen values over 1 are "stable" (Kaiser's criterion).
- EVs for successively extracted factors have lower values.
- EVs can be usefully expressed as %s of explained variance.
- Total of all EVs = the number of variables = or 100%.

61

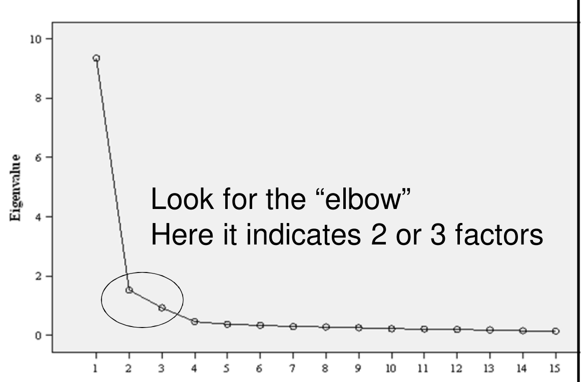
Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	9.355	62.366	62.366	9.094	60.628	60.628	7.801
2	1.532	10.216	72.583	1.294	8.625	69.253	7.261
3	.933	6.220	78.802	.635	4.232	73.485	5.732
4	.467	3.113	81.915				
5	.378	2.519	84.434				
6	.344	2.295	86.729				
7	.305	2.032	88.761				
8	.285	1.902	90.663				
9	.262	1.745	92.408				
10	.229	1.525	93.933				
11	.219	1.459	95.392				
12	.201	1.340	96.732				
13	.184	1.227	97.959				
14	.159	1.059	99.018				
15	.147	.982	100.000				

EVs range between 9.36 and 0.15. Two factors satisfy Kaiser's criterion (EVs > 1) but the third EV is .93 (and turns out to be a useful factor). There is a drop to the 4th factor's EV.

The total of these EVs is 15. There are 15 measurement items. If 15 factors are extracted, 100% of the variance is explained.

Scree plot

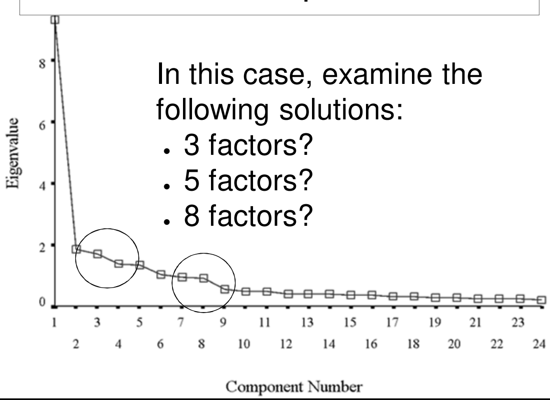


Scree plot

- A cumulative line graph of eigen values (EVs).
- Depicts amount of variance explained by each factor.
  - 1st factor explains the most variance.
  - Last factor explains least amount of variance.
- To determine the optimal # of factors: Look for where additional factors fail to add appreciably to the cumulative explained variance (where the “cliff” turns into “scree”).

64

Scree plot



In this case, examine the following solutions:

- 3 factors?
- 5 factors?
- 8 factors?

Practice quiz question:

EVs and % of variance explained

An EFA of 20 variables indicates that 4 factors explain 60% of the variance. What do the EVs of factors 5 to 20 add up to?

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

All EVs add up to 20 (100%). 4 factors explain 60%, so the other 16 factors explain 40% or 8 EVs.

66

How many factors?

- A *subjective* decision.
- Aim to explain most of the variance using a small number of factors.
- Take into account:
  1 Theory – what is predicted/expected?
  2 Eigen Values > 1? (Kaiser's criterion)
  3 Scree plot – where does it drop off?
  4 Interpretability of last factor?
  5 Try several different solutions?
  (consider EFA type, rotation, # of factors)
  6 Factors must be meaningfully interpretable and make theoretical sense.

67

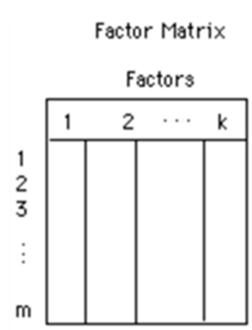
How many factors?

- Aim for 50 to 75% of variance explained by ¼ to ⅓ as many factors as variables.
- Stop extracting factors when they no longer represent useful/meaningful clusters of variables.
- Keep checking/clarifying the meaning of each factor – make sure to examine the wording of each item.

68

Factor loading matrix

- Factor loadings (FLs) indicate the relative importance of each item to each factor.
- A factor matrix shows variables in rows and factors in columns.
- Factors are weighted combinations of variables.



69

Initial solution: Unrotated factor structure

- In the initial solution, each factor “selfishly” grabs maximum unexplained variance.
- 1st 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 1st factor.

70

Initial solution: Unrotated factor structure

- Each subsequent factor tries to explain the remaining unexplained variance.
- Second factor is orthogonal to first factor - seeks to maximise its own Eigen Value (i.e., tries to gobble up as much of the remaining unexplained variance as possible), etc.

71

Vectors (lines of best fit)

Image source (left): http://www.xs-stock.co.uk/shopimages/products/normal/Kerplunk.jpg
Image source (right): http://mayhem-chaos.net/photoblog/images/super_sized_kerplunk.jpg

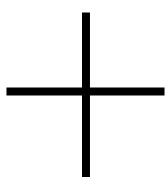
72

Factor rotation

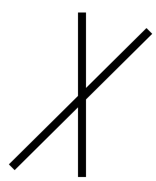
- However, until the factor loadings are rotated, they are difficult to interpret.
  - Seldom see a simple unrotated factor structure
  - Many variables will load on two or more factors
- Rotation of the factor loading matrix helps to find a more interpretable factor structure.

73

Factor rotation: Types



Orthogonal
(SPSS Varimax)
minimises factor covariation, produces factors which are uncorrelated



Oblique
(SPSS Oblimin)
allows factors to covary, allows correlations between factors

74

Factor rotation: Orthogonal vs. oblique

- Theory? (expecting related or unrelated factors?)
- Start with oblique rotation, then check correlations between factors:
  - If > ~.3 then with oblique rotation (> 10% shared variance between factors)
- Try both orthogonal and oblique rotations and assess which set of factor loadings are most interpretable? (i.e., which makes most sense?)

75

Interpretability

- Avoid being guided by factor loadings only – *think carefully* - be guided by theory and common sense in selecting the final factor structure.
- You must be able to understand and interpret each factor that you choose to extract.

76

Interpretability

- Watch out for "seeing what you want to see" when evidence might suggest a different, better solution.
- There may be more than one good solution! e.g., in personality:
  - 2 factor model
  - 5 factor model
  - 16 factor model

77

Factor loadings & item selection

A simple factor structure is most interpretable:

1. Each variable loads strongly (> $\pm$.40) on only one factor
2. Each factor has 3 or more strong loadings; more strongly loading variables = greater reliability
3. Most loadings are
   high (towards -1 or +1) or
   low (towards 0)
   (i.e., few intermediate values).

78

Rotated factor matrix - PC Varimax

Rotated Component Matrix[a]

	Component 1	2	3
PERSEVERES	.861	.158	.288
CURIOUS	.858		.310
PURPOSEFUL ACTIVITY	.806	.279	.325
CONCENTRATES	.778	.373	.237
SUSTAINED ATTENTION	.770	.376	.312
PLACID		.863	.203
CALM	.259	.843	.223
RELAXED	.422	.756	.295
COMPLIANT	.234	.648	.526
SELF-CONTROLLED	.398	.593	.510
RELATES-WARMLY	.328	.155	.797
CONTENTED	.268	.286	.748
COOPERATIVE	.362	.258	.724
EVEN-TEMPERED	.240	.530	.662
COMMUNICATIVE	.405	.396	.622

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 6 iterations.

Rotated factor matrix - PC Oblimin

ability

sk
tation

tled-
ess

	Component 1	2	3
RELATES-WARMLY	.920		.153
CONTENTED	.845		
COOPERATIVE	.784	-.108	
EVEN-TEMPERED	.682		-.338
COMMUNICATIVE	.596	-.192	-.168
PERSEVERES		-.938	
CURIOUS		-.933	.171
PURPOSEFUL ACTIVITY		-.839	
CONCENTRATES		-.831	-.201
SUSTAINED ATTENTION		-.788	-.181
PLACID			-.902
CALM		-.131	-.841
RELAXED		-.314	-.686
COMPLIANT	.471		-.521
SELF-CONTROLLED	.400	-.209	-.433

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.
a. Rotation converged in 13 iterations.

Rotated factor matrix - PC Oblimin

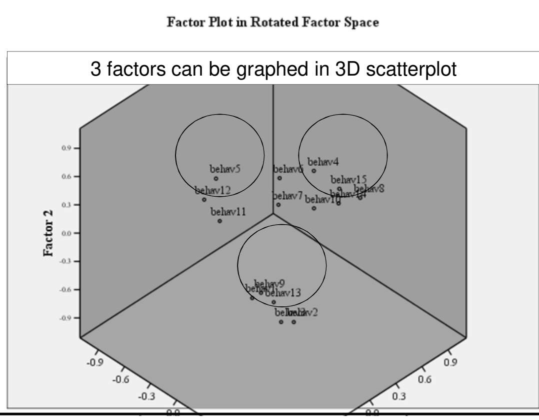
sk
tation

tled-
ess

ability

	Factor 1	2	3
PERSEVERES	.918		
CURIOUS	.897		
PURPOSEFUL ACTIVITY	.805		
CONCENTRATES	.778	.202	
SUSTAINED ATTENTION	.753		
CALM		.836	
PLACID		.759	
RELAXED	.282	.689	
RELATES-WARMLY			.851
CONTENTED			.800
COOPERATIVE			.735
EVEN-TEMPERED		.269	.717
COMMUNICATIVE			.599

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.



How many items per factor?

- Bare min. = 1
- Practical min. = 2
- Recommended min. = 3
- Max. = unlimited
- More items:
  - → ↑ reliability
  - → ↑ 'roundedness'
  - → Law of diminishing returns
- Typically 4 to 10 items per factor is reasonable

83

How to eliminate items

A subjective process; consider:

1 Size of item's main loading (min. > .4)
2 Size of cross loadings (max. < .3?)
3 Meaning of item & contribution it makes to the factor (face validity)
4 Eliminate 1 variable at a time, then re-run, before deciding which/if any items to eliminate next
5 Number of items already in the factor

84

Factor loadings & item selection

Comrey & Lee's (1992) guideline for primary (target) factor loadings:
> .70 - excellent
> .63 - very good
> .55 - good
> .45 - fair
< .32 - poor

85

Factor loadings & item selection

Cut-off for item loadings within a factor:

- Look for gap in loadings - e.g.,

.8
.7
.6
___
.3
.2

- Also consider: can the factor can be interpreted (i.e., does it make sense?) using items above but not below cut-off?

86

Factor analysis in practice

- To find a good EFA solution, try:
  - PC and PAF methods of extraction
  - Orthogonal and Oblique rotation
  - A range of possible factor structures, e.g., for 2, 3, 4, 5, 6, and 7 factors
- i.e., conduct *many* EFAs before deciding on a final solution.

87

Factor analysis in practice

- Eliminate poor items one at a time, re-examining results each time.
- You may come up with a different solution from someone else.
- Advanced: Check final model across sub-groups (e.g., gender) if there is sufficient data.
- Check reliability analysis (next lecture)

88

EFA example 3: Condom Use Self-Efficacy Scale

89

Example: Condom use

- Condom Use Self-Efficacy Scale
- 10-item measure administered to 447 multicultural college students (Barkley & Burns, 2000).
- EFA PC with a Varimax rotation.
- Three factors extracted:
  1 Appropriation
  2 Sexually Transmitted Diseases
  3 Partner's Disapproval

90

Factor analysis loadings & item selection

Factor 1: Appropriation - Acquisition and use of a condom (**α** = .76)	FL
I feel confident in my ability to put a condom on myself or my partner.	.75
I feel confident I could purchase condoms without feeling embarrassed.	.65
I feel confident I could remember to carry a condom with me should I need one.	.61
I feel confident I could gracefully remove and dispose of a condom after sexual intercourse.	.56

91

Factor analysis loadings & item selection

Factor 2: Sexually Transmitted Diseases - Stigma associated with STDs (**α** = .83)	FL
I would not feel confident suggesting using condoms with a new partner because I would be afraid he or she would think I've had a past homosexual experience.	.72
I would not feel confident suggesting using condoms with a new partner because I would be afraid he or she would think I have a sexually transmitted disease.	.86
I would not feel confident suggesting using condoms with a new partner because I would be afraid he or she would think I thought they had a sexually transmitted disease.	.80

92

Factor analysis loadings & item selection

Factor 3: Partner's reaction - students' partners' feelings about condoms (**α** = .66)	FL
If I were to suggest using a condom to a partner, I would feel afraid that he or she would reject me.	.73
If I were unsure of my partner's feelings about using condoms I would not suggest using one.	.65
If my partner and I were to try to use a condom and did not succeed, I would feel embarrassed to try to use one again (e.g. not being able to unroll condom, putting it on backwards or awkwardness).	.58

93

Example:
Zimbardo Time Perspective Inventory

- 56 items administered to 606 US college students (Zimbardo & Boyd, 1999).
- EFA PC with a Varimax rotation
  KMO = .83
- Five factors explained 36% of the variance
- Scree plot showed big drop between the 5th and 6th factors.

94

Example:
Zimbardo Time Perspective Inventory

1. **Past – Negative**
   "I think about the bad things that have happened to me in the past."
2. **Present – Hedonistic**
   "I do things impulsively."
3. **Future**
   "I am able to resist temptations when I know that there is work to be done."
4. **Past – Positive**
   "I get nostalgic about my childhood."
5. **Past – Negative**
   "My life path is controlled by forces I cannot influence."

95

Example: Zimbardo Time Perspective Inventory

Table 3
Intercorrelations Between Zimbardo Time Perspective Inventory Factors: Samples 1–4 (n = 606)

Factor	1	2	3	4	5
1. Past-Negative	—				
2. Present-Hedonistic	.16***	—			
3. Future	−.13**	−.29***	—		
4. Past-Positive	−.24***	.18***	.12**	—	
5. Present-Fatalistic	.38***	.32***	−.26***	−.09*	—

* $p < .05$. ** $p < .01$. *** $p < .001$.

Test–Retest Reliability

Test–retest reliabilities of the five subscales of the ZTPI were established with 58 Stanford introductory psychology students over a 4-week period. Reliabilities ranged from .70 to .80. The Future scale demonstrated the best test–retest reliability (.80), followed by Present-Fatalistic (.76), Past-Positive (.76), Present-Hedonistic (.72), and Past-Negative (.70). All correlations were

Image source: http://www.thetimeparadox.com/wp-content/uploads/2012/09/1999PuttingTimeinPerspective.pdf

Table 2
Exploratory Principal-Components Analysis: Varimax-Rotated Factor Matrix

ZTPI item	Past-Negative	Present-Hedonistic	Future	Past-Positive	Present-Fatalistic
1	.07	.42	−.02	.14	−.10
2	−.08	.18	.06	.62	.02
3	.24	.19	.09	.14	.44
4	.66	−.01	−.07	.05	.15
5	.41	.00	.02	.23	.18
6	.08	−.16	.46	.10	.02
7	−.25	.14	.01	.68	−.02
8	.03	.51	−.27	−.10	.05
9	−.09	.21	−.33	−.08	.12
10	−.16	.13	.56	−.03	−.09
11	−.41	.06	.03	.63	−.12
12	.09	.32	−.04	.13	.22
13	−.08	−.17	.63	.04	.10
14	.10	.04	−.15	−.07	.64
15	.18	.09	.09	.63	.06
16	.69	.16	−.01	−.18	.06
17	−.20	.50	.19	.11	−.06
18	.11	.04	.48	−.06	−.04
19	.05	.38	.12	.10	.07
20	−.24	.24	.11	.64	−.03
21	−.12	.04	.46	.17	−.04
22	.49	.24	.07	−.20	−.04
23	.07	.51	−.25	−.12	.13
24	.06	.28	−.49	−.11	.20
25	.55	−.02	.02	−.52	.21
26	.05	.56	.05	.18	−.14
27	.55	.03	−.18	.05	.02
28	.00	.36	−.30	.06	.33
29	.04	.06	−.02	.64	.21
30	.08	.03	.37	.16	−.29
31	−.00	.70	−.02	−.00	.03
32	−.13	.45	−.08	.08	.15
33	.43	.04	−.17	−.08	.29
34	.67	−.01	.05	−.25	.07
35	.20	.16	−.20	−.09	.42
36	.47	.08	.06	.24	.21
37	.14	.17	−.12	−.04	.59
38	.17	−.02	.06	.02	.73
39	.04	−.02	−.01	−.10	.68
40	−.17	−.02	.61	−.01	.04
41	−.00	.00	−.00	−.45	.25
42	.00	.71	−.01	−.04	.08
43	−.05	.07	.45	.07	−.05
44	.18	.45	−.10	.07	.12
45	−.16	−.09	.61	−.06	−.06
46	.16	.44	−.22	.23	.10
47	.20	−.09	−.00	.09	.42
48	−.04	.45	−.16	−.10	.18
49	.10	−.06	.11	.47	−.03
50	.76	.06	.06	−.08	.05
51	.09	−.07	.51	.01	−.08
52	−.05	.28	−.18	−.04	.34
53	.08	.14	−.11	.02	.45
54	.63	−.07	−.13	.01	.21
55	.20	.44	−.00	.07	−.02
56	−.11	.29	−.36	.09	.10

Note. ZTPI = Zimbardo Time Perspective Inventory.

Summary

97

Summary: Intro to factor analysis

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

98

Summary: Steps / process

1 Examine assumptions
2 Choose extraction method and rotation
3 Determine # of factors
(Eigen Values, Scree plot, % variance explained)
4 Select items
(check factor loadings to identify which items belong in which factor; drop items one by one; repeat)
5 Name and describe factors
6 Examine correlations amongst factors
7 Analyse internal reliability
8 Compute composite scores

Next lecture

99

Summary: Assumptions

- Sample size
  - Min: 5+ cases per variables
    Ideal: 20+ cases per variable)
  - Or $N > 200$
- Bivariate & multivariate outliers
- Factorability of correlation matrix (Measures of Sampling Adequacy)
- Normality enhances the solution

100

Summary: Types of factor analysis

- **PAF** (Principal Axis Factoring): For theoretical data exploration
  - uses shared variance
- **PC** (Principal Components): For data reduction
  - uses all variance

101

Summary: Rotation

- Orthogonal (Varimax)
  - perpendicular (uncorrelated) factors
- Oblique (Oblimin)
  - angled (correlated) factors
- Consider trying both ways
  - Are solutions different? Why?

102

Summary: Factor extraction

How many factors to extract?

- Inspect EVs
  - look for EVs > 1 or sudden drop (inspect scree plot)
- % of variance explained
  - aim for 50 to 75%
- Interpretability
  - does each factor "make sense"?
- Theory
  - do the factors fit with theory?

103

Summary: Item selection

An EFA 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).

104

References

1 Barkley, T. W. Jr., & Burns, J. L. (2000). Factor analysis of the Condom Use Self-Efficacy Scale among multicultural college students. *Health Education Research*, *15*(4), 485-489.

2 Comrey, A. L. & Lee, H. B. (1992). *A first course in factor analysis.* Hillsdale, NJ: Erlbaum.

3 Fehriinger H. M.(2004). *Contributions and limitations of Cattell's sixteen personality factor model*.

4 Howitt, D. & Cramer, D. (2011). Chapter 30: Factor analysis: Simplifying complex data. In *Introduction to statistics in psychology* (pp. 362-379) (5th ed.). Harlow, UK: Pearson.

5 Ivancevic, V., Kaine, A. K., MCLindin, B .A, & Sunde, J. (2003). Factor analysis of essential facial features. In the *Proceedings of the 25th International Conference on Information Technology Interfaces* (ITI), pp. 187-191, Cavtat, Croatia.

6 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.

7 Francis, G. (2007). *Introduction to SPSS for Windows: v. 15.0 and 14.0 with Notes for Studentware* (5th ed.). Sydney: Pearson Education.

8 Tabachnick, B. G. & Fidell, L. S. (2001). Principal components and factor analysis. In *Using multivariate statistics*. (4th ed., pp. 582 - 633). Needham Heights, MA: Allyn & Bacon.

9 Zimbardo, P. G., & Boyd, J. N. (1999). Putting time in perspective: A valid, reliable individual-differences metric. *Journal of Personality and Social Psychology*, *77*, 1271-1288.

105

Next lecture

Psychometric instrument development
- Concepts & their measurement
- Measurement error
- Psychometrics
- Reliability & validity
- Composite scores
- Writing up

106