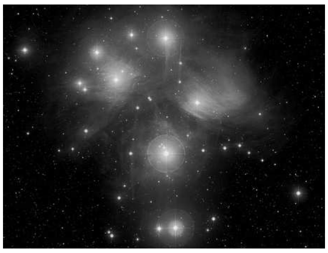
Exploratory Factor Analysis



Lecture 5

Survey Research & Design in Psychology

James Neill, 2012

Overview

1. What is factor analysis?
2. Assumptions
3. Steps / process
4. Examples
5. Summary

Readings: EFA

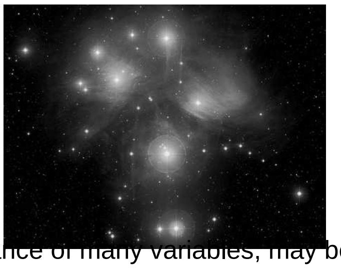
1. Fabrigar et al. (1999).
   Evaluating the use of exploratory factor analysis in psychological research.
   [article]
2. Tabachnick & Fidell (2001).
   Principal components and factor analysis.
   [chapter]

Available on e-reserve

What is factor analysis?

1. What is factor analysis?
2. Purpose
3. History
4. Types
5. Models

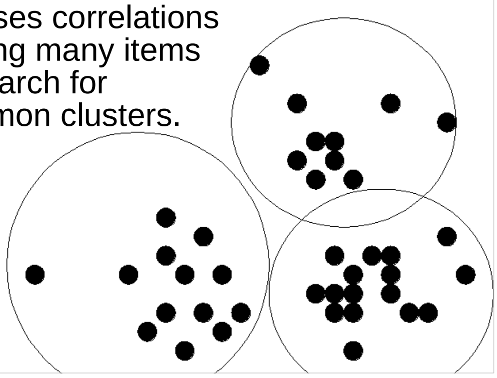
A galaxy is like a “factor” within the universe.



The variance of many variables, may be largely explained by some underlying factors (due to the co-relations (or clustering together)) of many variables.

Conceptual model of factor analysis

FA uses correlations among many items to search for common clusters.



Factor analysis...

- is used to identify clusters of inter-correlated variables (called 'factors').
- is a *family* of multivariate statistical techniques for examining correlations amongst variables.
- empirically tests theoretical data structures.
- is commonly used in psychometric instrument development.

7

Purposes

There are two main applications of factor analytic techniques:

1. **Theory development**: Detect structure in the relationships between variables, that is, to classify variables.
2. **Data reduction**: Reduce the number of variables to a smaller number of factors.

8

Purposes: Theory development

- Investigates the underlying correlational pattern shared by the variables in order to test theoretical models e.g., How many personality factors are there? Is intelligence general or multiple?
- The goal is to address a theoretical question (as opposed to calculating factor scores).

9

Purposes: Data reduction

- Simplifies data structure by revealing a smaller number of underlying factors (part of psychometrics)
- Helps to eliminate or identify items for improvement:
  - redundant variables
  - unclear variables
  - irrelevant variables
- Leads to calculating factor scores

10

History of factor analysis

- Invented by Charles Spearman (1904)
- Usage hampered by onerousness of hand calculation
- Since the advent of computers, usage has thrived, esp. to develop:
  - Theory e.g., determining the structure of personality or intelligence
  - Practice e.g., 10,000s+ of psychological screening & measurement tests

11

Two main types of FA: Exploratory vs. confirmatory factor analysis

EFA = **Exploratory Factor Analysis**

- explores & summarises underlying correlational structure for a data set

CFA = **Confirmatory Factor Analysis**

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

12

This lecture focuses on exploratory factor analysis

This (introductory) lecture focuses on **Exploratory Factor Analysis** (recommended for undergraduate level).

However, note that **Confirmatory Factor Analysis** (and Structural Equation Modeling) is generally preferred, but is more advanced, so is recommended for graduate/professional level.

Conceptual model - Simple model

Factor 1 Factor 2 Factor 3

- e.g., 12 items may 'tap' 3 underlying factors
- Factors consist of relatively homogeneous variables.

Eysenck's 3 personality factors

Extraversion/ introversion Neuroticism Psychoticism

talkative shy sociable fun anxious gloomy relaxed tense loner harsh unconventional nurturing

E.g., 12 items which measure 3 underlying dimensions of personality

Conceptual model - Simple model

Each question loads onto only one factor

Question 1

Question 2 Factor 1

Question 3 Factor 2

Question 4

Question 5 Factor 3

Conceptual model - Complex model

Each questions may load onto more than one factor

Question 1

Question 2 Factor 1

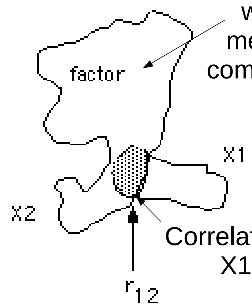
Question 3 Factor 2

Question 4 Factor 3

Question 5

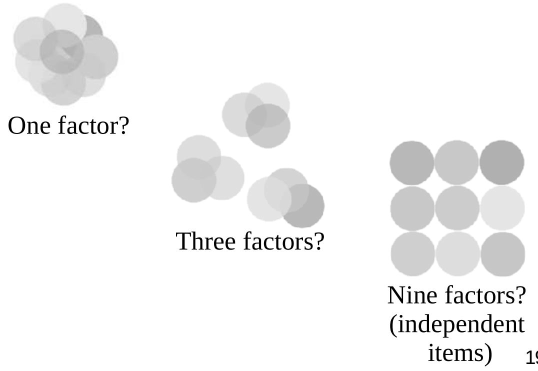
Conceptual model – Area plot

A theoretical factor which is partly measured by the common aspects of X1 and X2



Correlation between X1 and X2

How many factors?



19

Example: Personality

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

How could you decide between 3 or 5 personality factors?

Eysenck's 3?

- Extraversion
- Neuroticism
- Psychoticism

“Big 5”?

- Neuroticism
- Extraversion
- Agreeableness
- Openness
- Conscientiousness

20

Example: Intelligence

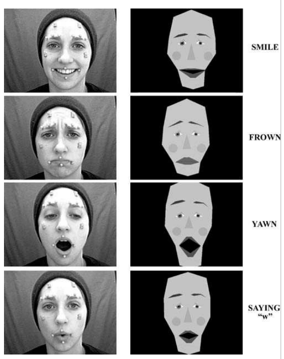
Is intelligence better described as:

- one global factor (g) or
- several specific factors (e.g., verbal, spatial, mathematical, social, kinaesthetic)?

How could you decide?

21

Example: What are the essential facial features for expression and communication?



(Ivancevic, 2003)

22

Example: Essential facial features

Six orthogonal (independent) factors, represent 76.5% of the total variability in facial recognition (in order of importance) (Ivancevic, 2003):

- upper-lip
- eyebrow-position
- nose-width
- eye-position
- eye/eyebrow-length
- face-width

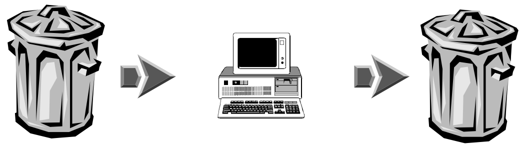
23

EFA assumptions

1. Garbage-In-Garbage-Out
2. Sample size
3. Levels of measurement
4. Normality
5. Linearity
6. Outliers
7. Factorability

24

Garbage. In. → Garbage. Out



- Screen the data
- Use variables that theoretically “go together”

Assumption testing: Sample size

Some guidelines:

- Min.: *N* > 5 cases per variable
  - e.g., 20 variables, should have > 100 cases (1:5)
- Ideal: *N* > 20 cases per variable
  - e.g., 20 variables, ideally have > 400 cases (1:20)
- Total *N* > 200 preferable

Assumption testing: Sample size

Comrey and Lee's (1992) guidelines:

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

Assumption testing: Sample size

	Journal of Personality and Social Psychology		*Journal of Applied Psychology*	
Variable	*N*	%	*N*	%
Ratio of variable to factors				
Less than 3:1	1	0.6	1	1.7
3:1	28	17.6	9	15.5
4:1	26	16.4	10	17.2
5:1	14	8.8	10	17.2
6:1	13	8.2	6	10.3
More than 6:1	74	46.5	18	31.0
Unknown	2	1.3	4	6.9

Assumption testing: Level of measurement

- All variables must be suitable for correlational analysis

i.e., they should be ratio/metric data or at least Likert data with several interval levels.

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.

Assumption Testing: Linearity

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

31

Assumption testing: Outliers

- FA is sensitive to outlying cases
  - –Bivariate outliers (e.g., check scatterplots)
  - –Multivariate outliers (e.g., Mahalanobis' distance)
- Identify outliers, then remove or transform

32

Example factor analysis: Classroom behaviour

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

33

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

34

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
14. Purposeful ↔ aimless
15. Cooperative ↔ disputes

35

Classroom behaviour items

Teachers, for each of the following paired behavioral statements, please mark a cross over the dot (e.g., ⊗) 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

36

Assumption testing: Factorability

Check the factorability of the correlation matrix (i.e., how suitable is the data for factor analysis?) by one or more of the following methods:

- Correlation matrix correlations > .3?
- Anti-image matrix diagonals > .5?
- Measures of sampling adequacy (MSAs)?
  - Bartlett's sig.?
  - KMO > .5 or .6?

37

Assumption testing: Factorability (Correlations)

Are there SOME correlations over .3? If so, proceed with FA

Correlation Matrix

		CONCEN TRATES	CURIOUS	PERSEV ERES	EVEN-TE MPERED	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 accurate

38

Assumption testing: Factorability: Anti-image correlation matrix

- Examine the diagonals on the anti-image correlation matrix
- Consider variables with correlations less than .5 for exclusion from the analysis – they lack sufficient correlation with other variables
- Medium effort, reasonably accurate

39

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]

Make sure to look at the anti-image CORRELATION matrix

40

Assumption testing: Factorability: Measures of sampling adequacy

- Global diagnostic indicators - correlation matrix is factorable if:
  - Bartlett's test of sphericity is significant and/or
  - Kaiser-Mayer Olkin (KMO) measure of sampling adequacy > .5 or .6
- Quickest method, but least reliable

41

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

42

Summary: Measures of sampling adequacy

Draw on one or more of the following to help determine the factorability of a correlation matrix:

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

Steps / process

1. Test assumptions
2. Select type of analysis
3. Determine no. 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 define factors
6. Examine correlations amongst factors
7. Analyse internal reliability
8. Compute composite scores

Type of EFA: Extraction method: PC vs. PAF

Two main approaches to EFA:

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

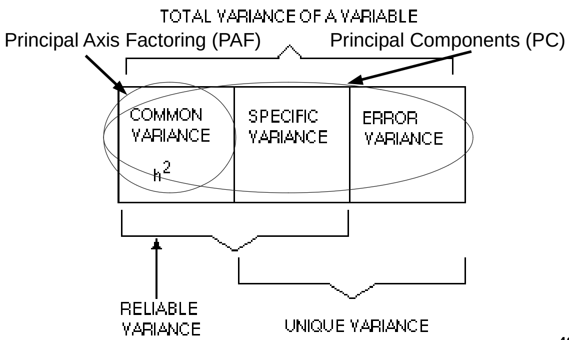
Principal axis factoring (PAF)

- Used to uncover the structure of an underlying set of *p* original variables
- More theoretical
- Analyses only shared variance (i.e. leaves out unique variance)

Principal components (PC)

- More common
- More practical
- Used to reduce data to a set of factor scores for use in other analyses
- Analyses all the variance in each variable

Total variance of a variable



PC vs. PAF

- Often there is little difference in the solutions for the two procedures.
- If unsure, check your data using both techniques
- If you get different solutions for the two methods, try to work out why and decide on which solution is more appropriate

Communalities

- Each variable has a communality =
  - the proportion of its variance explained by the extracted factors
- Ranges between 0 and 1
- If communality for a variable is low (e.g., < .5, consider extracting more factors or removing the variable)

Communalities

- High communalities (> .5): Extracted factors explain most of the variance in the variables being analysed
- Low communalities (< .5): A variable has considerable variance unexplained by the extracted factors
  - May then need to extract MORE factors to explain the variance or remove these items from the EFA

Communalities

	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.

Explained variance

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

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 items – very useful!

Eigen values (EVs)

- Each factor has an EV which indicates the amount of variance each factor accounts for.
- EVs for successive factors have lower values.
- Rule of thumb: Eigen values over 1 are ‘stable’ (Kaiser's criterion).
- EVs can also be expressed as %s.
- Total of all EVs is the number of variables. Each variable contributes a variance of one. EVs are then allocated to factors according to amount of variance explained.

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.

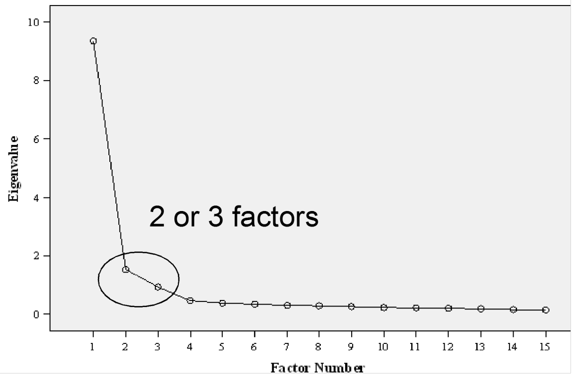
a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

The EVs ranged between .16 and 9.35. Two factors satisfied Kaiser's criterion (EVs > 1) but the third EV is .93 and it also appears to be a useful factor.

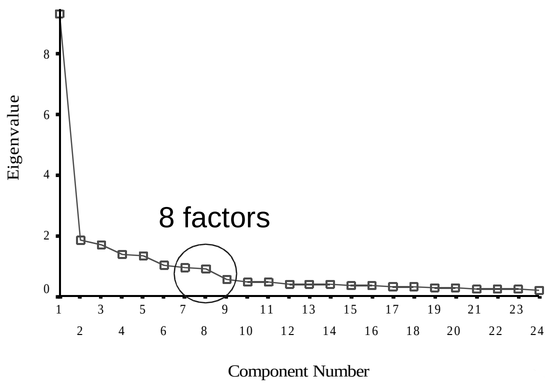
Scree plot

- A line graph of EVs.
- Depicts amount of variance explained by each factor.
- Cut-off: Look for where additional factors fail to add appreciably to the cumulative explained variance.
- 1st factor explains the most variance.
- Last factor explains the least amount of variance.

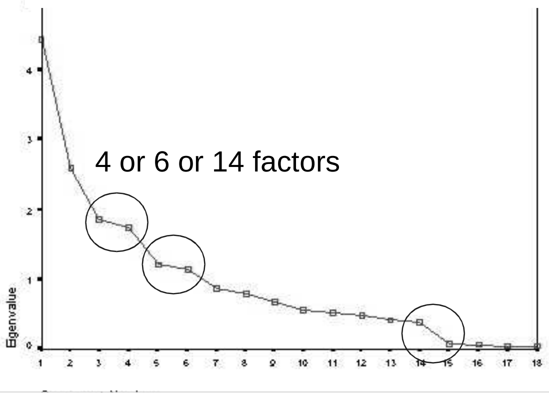
Scree plot



Scree plot



Scree plot



How many factors?

A *subjective* process ... Seek to explain maximum variance using fewest factors, considering:

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 FA type, rotation, # of factors)
6. Factors must be able to be meaningfully interpreted & make theoretical sense?

61

How many factors?

- Aim for 50-75% of variance explained by 1/4 to 1/3 as many factors as variables/items.
- 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.

62

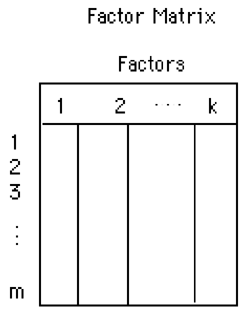
Initial solution: Unrotated factor structure

- Factor loadings (FLs) indicate relative importance of each item to each factor.
  - In the initial solution, each factor tries "selfishly" to grab maximum unexplained variance.
  - All variables will tend to load strongly on the 1st factor

63

Initial solution - Unrotated factor structure

- Factors are weighted combinations of variables
- A factor matrix shows variables in rows and factors in columns



64

Initial solution - Unrotated factor structure

1st factor extracted:

- Best possible line of best fit through the original variables
- Seeks to explain lion's share of all variance
- A single factor, best summary of the variance in the whole set of items

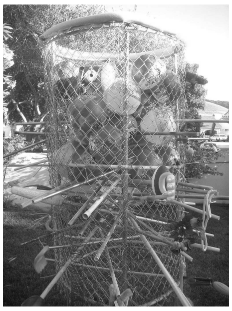
65

Initial solution - Unrotated factor structure

- Each subsequent factor tries to explain the maximum amount of 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)

66

Vectors (Lines of best fit)



67

Initial solution: Unrotated factor structure

- Seldom see a simple unrotated factor structure
- Many variables load on 2 or more factors
- Some variables may not load highly on any factors (check: low communality)
- Until the FLs are rotated, they are difficult to interpret.
- Rotation of the FL matrix helps to find a more interpretable factor structure.

68

Two basic types of factor rotation

Orthogonal
(SPSS Varimax)

Oblique
(SPSS Oblimin)

69

Two basic types of factor rotation

1. Orthogonal
   minimises factor covariation, produces factors which are uncorrelated
2. Oblimin
   allows factors to covary, allows correlations between factors

70

Why rotate a factor loading matrix?

- After rotation, the vectors (lines of best fit) are rearranged to optimally go through clusters of shared variance
- Then the FLs and the factor they represent can be more readily interpreted

71

Orthogonal vs. oblique rotations

- Consider purpose of factor analysis
- If in doubt, try both
- Consider interpretability
- Look at correlations between factors in oblique solution
  - if >.3 then go with oblique rotation (>10% shared variance between factors)

72

Interpretability

- It is dangerous to be driven by factor loadings only – think carefully - be guided by theory and common sense in selecting factor structure.
- You must be able to understand and interpret a factor if you're going to extract it.

Interpretability

- However, 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

Factor loadings & item selection

A factor structure is most interpretable when:

1. Each variable loads strongly (> $\pm$.40) on only one factor
2. Each factor shows 3 or more strong loadings; more loadings = greater reliability
3. Most loadings are either high or low, few intermediate values.
4. These elements give a 'simple' factor structure.

Rotated factor matrix - PC Varimax

Rotated Component Matrix[a]

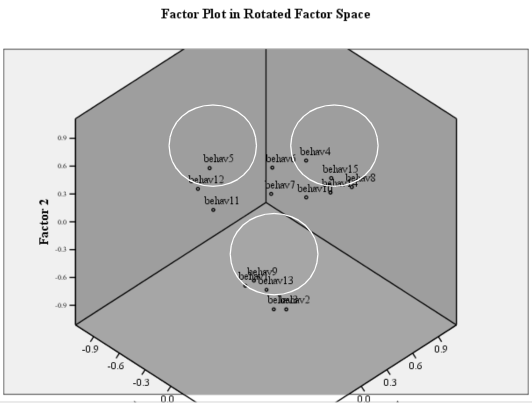
	Component		
	1	2	3
PERSEVERES	.861	.158	.288
CURIOUS	.858		.310
PURPOSEFUL ACTIVITY	.806	.279	.325
CONCENTRATES	.778	.373	.237
SUSTAINED ATTENTION	.770	.376	.312
PLACID		.863	.203
CALM	.259	.843	.223
RELAXED	.422	.756	.295
COMPLIANT	.234	.648	.526
SELF-CONTROLLED	.398	.593	.510
RELATES-WARMLY	.328	.155	.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

		Component		
		1	2	3
Sociability	RELATES-WARMLY	.920		.153
	CONTENTED	.845		
	COOPERATIVE	.784	-.108	
	EVEN-TEMPERED	.682		-.338
	COMMUNICATIVE	.596	-.192	-.168
Task Orientation	PERSEVERES		-.938	
	CURIOUS		-.933	.171
	PURPOSEFUL ACTIVITY		-.839	
	CONCENTRATES		-.831	-.201
	SUSTAINED ATTENTION		-.788	-.181
Settledness	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.



How many items per factor?

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

79

How do I eliminate items?

A subjective process; consider:

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

80

Factor loadings & item selection

Comrey & Lee (1992) guideline for primary (target) factor loadings:

> .70 - excellent

> .63 - very good

> .55 - good

> .45 - fair

> .32 - poor

81

Factor loadings & item selection

Cut-off for acceptable loadings:

- Look for gap in loadings - e.g.,
  - .8
  - .7
  - .6
  - .3
  - .2
- But also consider whether factor can be interpreted above but not below cut-off.

82

Other considerations: Normality of items

- Check the item descriptives.
- The more normally distributed the item scores, the better the distribution of the composite scores.
  - e.g. if two items have similar Factor Loadings and Reliability analysis, consider selecting items which will have the least skew and kurtosis.

83

Factor analysis in practice

- To find a good solution, consider:
  - PC/PAF
  - Varimax/oblimin
- Range of possible factor structures, e.g., for 2, 3, 4, 5, 6, and 7 factors
- Thus, the researcher would normally conduct many initial EFAs before deciding on a probable structure.

84

Factor analysis in practice

- Eliminate poor items one at a time, retesting the possible solutions
- Check factor structure across sub-groups (e.g., gender) if there is sufficient data
- You will probably come up with a different solution from someone else!
- Check/consider reliability analysis (next lecture)

Example: Condom use

- The Condom Use Self-Efficacy Scale (CUSES) was administered to 447 multicultural college students (Barkley & Burns, 2000).
- PC EFA with a varimax rotation.
- Three factors were extracted:
  1. Appropriation
  2. Sexually Transmitted Diseases
  3. Partners' Disapproval

Factor 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

Factor 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

Factor 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

Summary

1. Introduction
2. Assumptions
3. Steps/Process

Introduction: Summary

- 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) are summary indicators of underlying fuzzy constructs.

Assumptions: Summary

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

Summary: Steps / process

1. Test assumptions
2. Select type of analysis
3. Determine no. 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 define factors
6. Examine correlations amongst factors
7. Analyse internal reliability
8. Compute composite scores

Next lecture

Summary: Types of FA

- PAF: Theoretical data exploration
  - uses shared variance
- PC: Data reduction
  - uses all variance
- Consider trying both ways
  - Are solutions different? Why?

Summary: Rotation

- Orthogonal (varimax)
  - perpendicular vectors
- Oblique (oblimin)
  - angled vectors
- Consider trying both ways
  - Are solutions different? Why?

Summary: Factor extraction

No. of factors to extract?

- Inspect EVs
  - look for > 1 or sudden drop (inspect scree plot)
- % of variance explained
  - aim for 50 to 75%
- Interpretability / theory

References

1. Barkley, T. W. Jr., & Burns, J. L. (2000). Factor analysis of the Condom Use Self-Efficacy Scale among r . *Health Education Research*, *15*(4), 485-489.
2. 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.
3. 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.
4. 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.

Open Office Impress

- This presentation was made using Open Office Impress.
- Free and open source software.
- http://www.openoffice.org/product/impress.html