Factor Analysis

Professor Timothy Bates
tim.bates@ed.ac.uk

- http://www.psy.ed.ac.uk/people/tbates/
  (lecture & tutorial are under Y3 teaching)
- Please attempt the exercise prior to the tutorial

- Bring your hand out to the tutorial

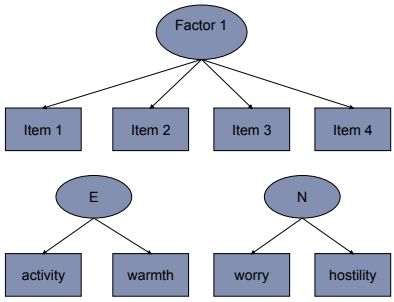
FACTOR ANALYSIS

"A statistical tool to account for variability in a set of measured items in terms of a smaller number of factors"

- Measured item = Observed variable
- Factor = Unobserved (latent) variable

- Values for an observation can be recovered (with some error) from a linear combination of a (usually much smaller) number of extracted factors.

Visually…



FA as data reduction

- **Simplify complex multivariate datasets by finding "natural" groupings within the data**
  - Subsets of variables that correlate strongly with each other and weakly with other variables in the dataset.
  - May correspond to underlying 'dimensions'

- **Why?**
  - Factors can assist the theoretical interpretation of complex datasets
  - Theoretical linkage of factors to underlying (latent) constructs, e.g. "extraversion", liberal attitudes, interest in ideas, ability

Factors by Eye

- 210 subjects → 12*12 matrix
- Clear structure in this case:
  - E: 1-4 Assertive, Talkative, Extraverted, Bold
  - C: 5-8 Organized, Efficient, Thorough, Systematic
  - N: 9-12 Insecure, Self-pitying, Nervous, Irritable
- How can you quantify these factors?
- Are they correlated?
- What if you had hundreds of items?

	1	2	3	4	5	6	7	8	9	10	11
2	**0.27**										
3	**0.37**	**0.53**									
4	**0.40**	**0.30**	**0.38**								
5	-0.17	-0.07	-0.09	-0.08							
6	-0.17	-0.05	-0.06	-0.10	**0.59**						
7	-0.19	-0.01	-0.05	-0.05	**0.38**	**0.42**					
8	-0.06	-0.02	-0.02	-0.02	**0.51**	**0.54**	**0.48**				
9	**0.25**	-0.05	-0.15	-0.20	-0.06	-0.11	-0.14	-0.07			
10	**0.24**	-0.10	-0.09	-0.10	-0.03	-0.02	-0.13	-0.08	**0.38**		
11	**0.21**	-0.08	**0.22**	-0.12	-0.00	-0.03	-0.07	-0.03	**0.49**	**0.38**	
12	-0.01	-0.02	-0.10	-0.04	-0.07	-0.09	-0.06	-0.04	**0.34**	**0.40**	**0.40**

Objectives And Outcomes Of Factor Analysis

- **Aim of factor analysis is to objectively detect natural groupings of variables (factors)**
- **Can deal with large matrices, uses (reasonably) objective statistical criteria.**
- **Can obtain quantitative information**
  - e.g. factor scores.
- **Factors are (should be) of theoretical interest.**
  - In the example the factors correspond to the personality traits of Extraversion, Neuroticism and Conscientiousness
- **Exploratory method, uncovering structure in data**
  - Confirmatory factor analysis (model testing) is also possible.

Factor Analysing the Example

	I (C)	II (N)	III (E)
EFFICIENT	0.82		
ORGANIZED	0.80		
SYSTEMATIC	0.79		
THOROUGH	0.71		
NERVOUS		0.75	-0.15
IRRITABLE	0.14	0.73	
INSECURE	-0.14	0.73	-0.16
SELF-PITYING		0.72	
EXTRAVERTED	-0.12	-0.10	0.79
TALKATIVE			0.75
BOLD			0.69
ASSERTIVE	0.24	-0.21	0.65

•**Factor loadings** = Correlation of each variable with the underlying factor

•**Factor score** = Subject's responsesxfactor loadings

7

Factors in the Example Data

- Factor loadings = Correlation of each variable with the underlying factor
- Factor score = Responses * factor loadings

N = (0.75*Nervous) + (.73*Irritable) + (.73*Insecure) + (.72*Self-pity) + (-.1*Extraverted) + (-.21*Assertive)

- Factors are close to simple structure
- Are cross-loadings informative?

	I (C)	II (N)	III (E)
EFFICIENT	0.82		
ORGANIZED	0.80		
SYSTEMATIC	0.79		
THOROUGH	0.71		
NERVOUS		0.75	-0.15
IRRITABLE	0.14	0.73	
INSECURE	-0.14	0.73	-0.16
SELF-PITYING		0.72	
EXTRAVERTED	-0.12	-0.10	0.79
TALKATIVE			0.75
BOLD			0.69
ASSERTIVE	0.24	-0.21	0.65

8

Requirements for a Valid And Useful Factor Analysis

- **Assumptions met and data Replicable**
  - **Sample size and representativeness**
- **Factors retain most of the variance in the raw data**
  - **Parsimony compared to starting variables achieved without loss of explanatory power**
- **Factors have a meaningful interpretation**
  - **Simple structure**
    - **Each item loads highly on one factor and close to zero on all others**
  - **Rotation**

9

Stages Of Analysis

1. Examine data for outliers and correlations
2. Choose number of factors
   1. Scree plot
3. Rotate factors if necessary
4. Interpret factors
5. Obtain scores
   1. Check reliability of scales defining factors
6. Further experiments to validate factors

10

Assumptions& Influences on Reliability

- Large enough sample
  - So that the correlations are reliable
- Variables normally distributed
- No outliers
- No variables uncorrelated with any other
- No variables correlated 1.0 with each other
  - Remove one of each problematic pair, or use sum if appropriate.

11

Data Quality

- **Subjects/variable ratio**
  - **Values between 2:1 and 10:1 have been proposed as a minimums.**
- **Simulations suggest that overall sample size is more important**
- **Sample Size**
  - **Rough rule is that 300 is OK**
  - **Smaller numbers often used: unreliable**
- **Well-defined factors (large loadings) will replicate in smaller samples more often thanwill poorly-defined ones (small loadings)**

12

Item variance&FA

- Variance of each item can be thought of in three partitions:
  1. Shared variance
     - I. Common variance (accounted for by our factors)

  +

  II. Unique variance: Not accounted for by other items
     - 2. Specific variance
     - 3. Error variance
- Communality
  - The proportion of common variance for a given variable
    - = Sum of the squared item factor loadings
  - Large communalities are required for a valid and useful factor solution

13

Computing a Factor Analysis

- Two main approaches
  - Differ in estimating communalities
- **Principal components**
  - Simplest computationally
  - Assumes all variance is common variance (implausible) but gives similar results to more sophisticated methods.
  - SPSS default.
- **Principal factor analysis**
  - Estimates communalities first

14

How many Factors?

- Initially unknown
  - Needs to be specified by the investigator on the basis of preliminary analysis
  - No 100% foolproof statistical test for number of factors
  - Similar problems with other multivariate methods

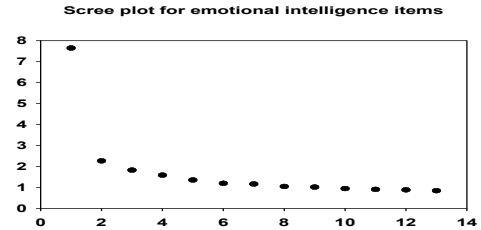
15

How many factors?

- There are potentially as many factors as items
- We don't want to retain factors which account for little variance.
- Most commonly-used method to decide the number of factors is the "scree" plot of the "E*igenvalues*"
  - Variance explained by each factor.
- A point of inflection or kink or in the scree plot is a good method of making a cut-off

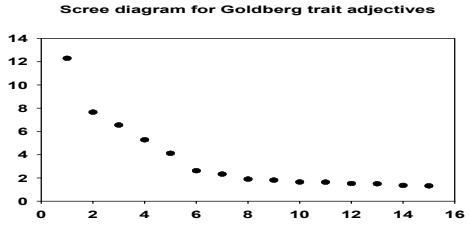
16

EQ Scree



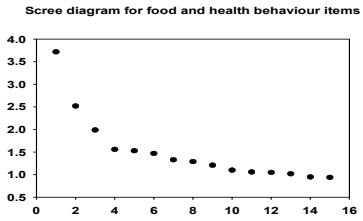
17

Goldberg Scree



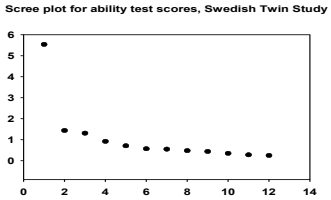
18

Food and Health Scree



19

IQ Scree



20

Other Methods For Factor Numbers

- **Eigenvalues> 1**
  - Eigenvalues sum to the number of items, so an eigenvalue of >1 is more informative than a single average item
  - Not a useful guide in practice
- **Parallel Analysis**
  - Repeatedly randomise the correlation matrix and determine how large an Eigenvalue appears by chance in many thousands of trials.
  - Excellent method
- **Theory-driven**
  - Extract a number of factors based on theoretical considerations
    - Justifying this approach requires structural modeling

21

How to align the factors?

- **The initial solution is "un-rotated"**
- **Two undesirable features make it hard to interpret:**
  - Designed to maximise the loadings of all items on the first factor
  - Most items have large loadings on more than one factor
- **Hides groupings in the data**

22

Unrotated Factors For The Example Data

	I	II	III
EFFICIENT	0.45	0.69	0.02
ORGANIZED	0.37	0.71	-0.04
SYSTEMATIC	0.37	0.70	0.04
THOROUGH	0.45	0.55	-0.02
NERVOUS	-0.56	0.33	0.40
IRRITABLE	-0.34	0.37	0.56
INSECURE	-0.62	0.21	0.38
SELF-PITYING	-0.52	0.28	0.42
EXTRAVERTED	0.46	-0.41	0.51
TALKATIVE	0.36	-0.31	0.58
BOLD	0.48	-0.24	0.45
ASSERTIVE	0.64	-0.10	0.33

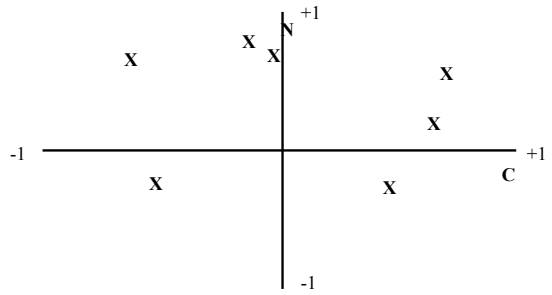
23

ROTATION – DETAIL (1)

- **Rotation shows up the groups of items in the data.**
- **Orthogonal rotation**
  - Factors remain independent
- **Oblique rotation**
  - Factors allowed to correlate
- **Theoretical reasons to choose a type of rotation**
  - (e.g. independent for personality traits);
- **May explore both types**
  - Choose oblique if there are large correlations between factors, orthogonal otherwise.

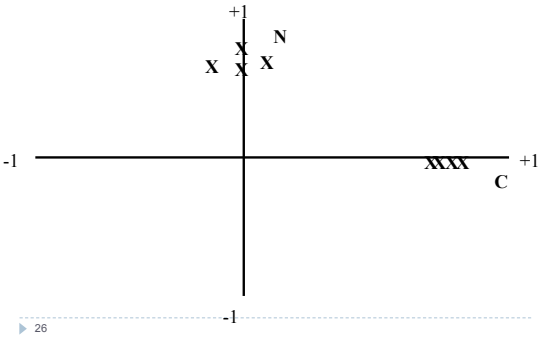
24

Lack of Simple Structure



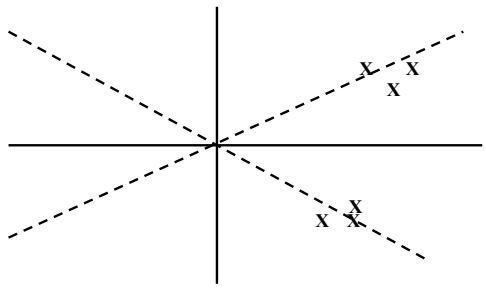
25

Item loadings After Rotation



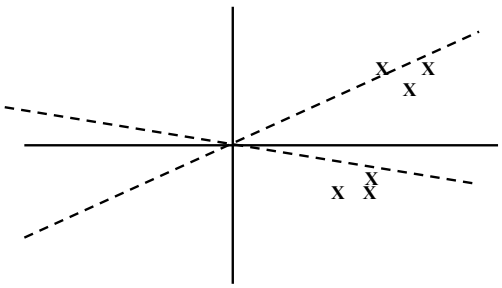
26

Rotation Defines New Axes Which Reveal the Item Groups



27

Oblique Rotation



28

Rotation -Detail (2)

- **Rotated and un-rotated solutions are mathematically equivalent**
  - Rotation is performed for purposes of interpretation.
- **Most common types:**
  - **Orthogonal**
    - Varimax (maximizes squared column variance)
      - *Most common*
  - **Oblique**
    - Direct oblimin

29

Interpreting Factors

- **Done on the basis of 'large' loadings**
  - Often taken to be above 0.3.
  - Size of loading which should be considereddepends on sample-size
    - For large samples loadings of 0.1 or below may be significant but do not explain much variance
- **Well-defined factor should have at least three high-loading variables**
  - Existence of factors with only one or two large loadings indicates factors over-extracted, or multi colinearityproblems.
- **Assigning meaning to factors.**

30

Factor Scores

- **Factor scores**
  - Estimate of each subject's score on the underlying latent variable
  - Calculated from the factor loadings of each item and the subject's responses
- **Simple scoring: unit loadings for large items**
  - Often used for, e.g., personality questionnaires is to sum the individual item scores (reverse-keying where necessary).
  - This method is reasonable when all variables are measured on the same scale;
  - What if you have a mix of items measured on different scales?
    - **(e.g. farmer's extraversion score, farm annual profit, farm area).**

31

Example – Factor Structure Of Dietary Behaviour

- **Research question: Is there a dimension of healthy vs. unhealthy diet preferences?**
  - (MacNicolet al 2003)
- **451 schoolchildren completed a 35-item questionnaire mainly on food items regularly consumed (also some general health behaviour items)**
  - Subjects/variables = 13:1, Population not representative for SES.
- **Scree suggested three factors, two diet related**
  - F1: Unhealthy foods (chips, fizzy drinks etc)
  - F2 Healthy foods (fruit, vegetableetc)
- **Validation**
  - Higher SES and better nutrition knowledge associated with healthier eating patterns.
- **Factor reliabilities low**
  - Problem of yes/no items
  - Sampleheterogeneity (sex effects?)

32

Attitudes to chocolate: How many factors?

- Craving
  - I like to indulge in chocolate
  - I often go into a shop for something else and end up buying chocolate),
- Guilt
  - I feel guilty after eating chocolate
- Functional approach
  - I eat chocolate to keep my energy levels up when doing physical exercise.
- Craving
  - Consuming more bars per month
  - Were prepared to work harder to get chocolate buttons

33

Attitudes to Chocolate

- **80-items constructed using interviews & theory (approach, avoidance, guilt theory) measuring:**
  - Difficulty controlling consumption, positive attitudes, negative attitudes, craving, self-report chocolate consumption
- **Participants also performed a bar-pressing task with chocolate reinforcements delivered on a progressive ratio schedule.**
- **Factor analysis gave 3 factors (eigenvalue>1)**
  - 33.2%, 14.1% & 6.1% of the variance
  - Over-factored: 3rd scale had low reliability

34

Two correlated Factors of Chocolate (Cramer & Hartleib, 2001)

- Craving
  - I like to indulge in chocolate
  - I often go into a shop for something else and end up buying chocolate)
- Guilt
  - I feel guilty after eating chocolate
- Craving associated more with depression, Guilt more with anxiety, but both highly correlated.

35

Example– Abnormal Personality

- **How does personality disorder relate to normal personality?**
  - **Deary et al. (1998).**
  - Scale-level analysis of DSM-III-R personality disorders & EPQ-R
  - Sample = 400 students
- **Joint analysis gives four factors:**
  - N+ Borderline, Self-defeating, Paranoid
  - P+ Antisocial, Passive-aggressive, Narcissistic
  - E+ avoidant(-), histrionic
  - P(-) Obsessive-compulsive, Narcissistic

36

End of Lecture I

- See you next time!

37

Using Factors

1. **Interpretation (Naming)**
   - Use content of high-loading items as a guide
2. **Scoring**
   - Factor scores
   - 'unit weighting' often used in practice
3. **Reliability**
   - Assess internal reliability for each factor
4. **Validation**
   - Convergent Validity
     - Do factor scores correlate as expected with other tests?
   - Construct Validity
     - Does the test predict a range of behaviors as you expect?

38

1: Interpreting Factors

- **Done on the basis of 'large' loadings**
  - Often taken to be above 0.3.
  - Size of loading which should be considered substantive is sample-size dependent.
  - For large samples loadings of 0.1 or below may be significant but do not explain much variance.
- **Well-defined factor should have at least three high-loading variables**
  - Existence of factors with only one or two large loadings indicates factors over-extracted, or multi-colinearity problems.
- **Assigning meaning to factors.**

39

Using Factors

1. **Interpretation (Naming)**
   - Use content of high-loading items as a guide
2. **Scoring**
   - Factor scores
   - 'unit weighting' often used in practice
3. **Reliability**
   - Assess internal reliability for each factor
4. **Validation**
   - Convergent Validity
     - Do factor scores correlate as expected with other tests?
   - Construct Validity
     - Does the test predict a range of behaviors as you expect?

40

2: Factor Scores

- **Factor scores**
  - Estimate of each subject's score on the underlying latent variable
  - Calculated from the factor loadings of each item.
- **Simple scoring methods**
  - Often used for, e.g., personality questionnaires is to sum the individual item scores (reverse-keying where necessary).
  - This method is reasonable when all variables are measured on the same scale;
  - What if you have a mix of items measured on different scales?
    - **(e.g. farmer's extraversion score, farm annual profit, farm area).**

41

Using Factors

1. **Interpretation (Naming)**
   - Use content of high-loading items as a guide
2. **Scoring**
   - Factor scores
   - 'unit weighting' often used in practice
3. **Reliability**
   - Assess internal reliability for each factor
4. **Validation**
   - Convergent Validity
     - Do factor scores correlate as expected with other tests?
   - Construct Validity
     - Does the test predict a range of behaviors as you expect?

42

3: Scale Reliability

- **Assessing Reliability of Factor-Derived Scales**
  - Factor Derived Scales can be used like any other scale
  - For instance using Cronbach's Alpha
- **Scale alpha**
  - Adequate reliability is customarily 0.7 or above
  - If too low, might indicate under-factoring
  - Too few items
- **Item alpha**
  - Check item-alpha to identify poorly-functioning items
  - Delete these from the scale

Buss & Perry (1992)

- 4 factors
- 29 items

Table 1
Four Aggression Factors

Factor	Factor loadings
Physical Aggression	
1. Once in a while I can't control the urge to strike another person.	.66, .55, .62
2. Given enough provocation, I may hit another person.	.79, .84, .80
3. If somebody hits me, I hit back.	.60, .65, .60
4. I get into fights a little more than the average person.	.44, .52, .58
5. If I have to resort to violence to protect my rights, I will.	.63, .68, .58
6. There are people who pushed me so far that we came to blows.	.60, .62, .65
7. I can think of no good reason for ever hitting a person.*	.47, .53, .51
8. I have threatened people I know.	.45, .48, .65
9. I have become so mad that I have broken things.	.47, .57, .47
Verbal Aggression	
1. I tell my friends openly when I disagree with them.	.41, .41, .48
2. I often find myself disagreeing with people.	.38, .49, .35
3. When people annoy me, I may tell them what I think of them.	.45, .45, .40
4. I can't help getting into arguments when people disagree with me.	.38, .41, .36
5. My friends say that I'm somewhat argumentative.	.37, .36, .46
Anger	
1. I flare up quickly but get over it quickly.	.53, .49, .49
2. When frustrated, I let my irritation show.	.47, .45, .37
3. I sometimes feel like a powder keg ready to explode.	.60, .35, .35
4. I am an even-tempered person.*	.64, .62, .69
5. Some of my friends think I'm a hotbead.	.63, .51, .64
6. Sometimes I fly off the handle for no good reason.	.75, .64, .70
7. I have trouble controlling my temper.	.74, .66, .69
Hostility	
1. I am sometimes eaten up with jealousy.	.41, .43, .49
2. At times I feel I have gotten a raw deal out of life.	.61, .58, .52
3. Other people always seem to get the breaks.	.65, .65, .63
4. I wonder why sometimes I feel so bitter about things.	.48, .45, .59
5. I know that "friends" talk about me behind my back.	.55, .37, .47
6. I am suspicious of overly friendly strangers.	.42, .35, .43
7. I sometimes feel that people are laughing at me behind my back.	.66, .64, .70
8. When people are especially nice, I wonder what they want.	.55, .50, .47

* The scoring of these items is reversed.

Bryant & Smith (2001)

- 12 items, with better fit and cross-cultural reliability

TABLE 3
CFA Factor Loadings for the Refined 12-Item, Four-Factor Measurement Model of the AQ

	PA sample			VA sample			ANG sample			HO sample		
AQ items	1	2	3	1	2	3	1	2	3	1	2	3
2. Given enough provocation, I may hit another person.	76	70	58									
6. There are people who pushed me so far that we came to blows.	72	73	65									
8. I have threatened people I know.	80	82	68									
11. I often find myself disagreeing with people.				80	75	70						
13. I can't help getting into arguments when people disagree with me.				82	71	68						
14. My friends say that I'm somewhat argumentative.				58	61	76						
15. I flare up quickly but get over it quickly.							50	62	69			
20. Sometimes I fly off the handle for no good reason							81	83	57			
21. I have trouble countrolling my temper.							71	71	34			
23. At times I feel I have gotten a raw deal out of life.										65	76	45
24. Other people always seem to get the breaks.										77	75	64
25. I wonder why sometimes I feel so bitter about things.										68	68	52

How to Assess FA in papers

- **Sample size:Two things matter:**
  - Ratio of subjects to Items
  - Total sample size
  - Item to subject ratio is important
  - Can get away with smaller numbers when communalities are high (i.e. factors well-defined)
- **Restriction of range (subjects too similar)**
  - Reduces correlations
- **Items per factor.**
  - Need at least three per factor, four is better. Some published analyses discuss factors with only one item loading!
- **Use of eigenvalue>1.**
  - Often seen in papers where factor number comes out implausibly high.
- **Rotation.**
  - Orthogonal forced when oblique should have been tried.
- **Scores.**
  - SPSS and other packages give scores which are sample-dependent.
  - Use of unit weighting of items is better practice.

Adequacy of sample size

- 50 – very poor
- 100 – poor
- 200 – fair
- 300 – good
- 500 – very good
- >1000 – excellent
  - Comfrey and Lee (1992, p. 217)

Item-subject ratios.

- With too many items and too few subjects, the data are "over-fitted"
  - Unreplicable results
    - Bobko & Schemmer, 1984
- Subjects to items
  - 5:1 (Gorsuch, 1983, p.332; Hatcher, 1994, p. 73)
  - 10:1 (Nunnally, 1978, p. 421)
- Subjects to parameters measures
  - MacCallum, Widaman, Preacher, & Hong (2001)
    - Subject: factor ratio
    - Item communalities
    - Item loadings

Advanced FA: Statistical Tests For Data Quality

- Bartlett's test of sphericity
- Kaiser-Meyer-Olkin test of sampling adequacy

49

Bartlett's test of sphericity.

- Tests the hypothesis that correlations between variables are greater than would be expected by chance
  - Technically, tests if the matrix is an identity matrix
- p-value should be significant
  - i.e., the null hypothesis that all off-diagonal correlations are zero is falsified

50

Two kinds of Matrices

- Identity matrix
  - Is on the diagonal and zeros elsewhere.
  - Each item correlates only with itself
  - Bartlett's test tests that the matrix is significantly different from an identity matrix.
- Singular matrix
  - A matrix in which one or more off-diagonal elements = 1
  - Cannot be factor analysed
  - Solution = remove duplicate items.

51

KMO Sampling Adequacy

- Kaiser-Mayer-Olkin Test of Sampling Adequacy
  - Tests whether there are a significant number of factors in our dataset
  - Technically, tests the *ratio* of item-correlations to partial item-correlations: If the partials are similar to the raw correlations, it means the item doesn't share much variance with other items.
  - Range = 0.0-1.0; Desired values are > 0.5
    - Low values indicate diffuse correlations with no substantive groupings.
- KMO statistics for each item
  - Item values below 0.5 indicate item does not belong to a group and may be removed

52

A worked example in SPSS

- **FA is found in the menus under: Analyse: data reductio: factor**
- **Tasks are:**
- **EXTRACTION**
  - **Select scree plot for initial run.**
  - **Choose number of factors.**
- **ROTATION**
  - **Select rotation method**
  - **Increase number of iterations for rotation if necessary (default 25)**
- **DESCRIPTIVES**
  - **Check KMO and Bartlett tests**
  - **Examine reproduced correlations and residuals**
- **SCORES**
  - **Save as variables**

53

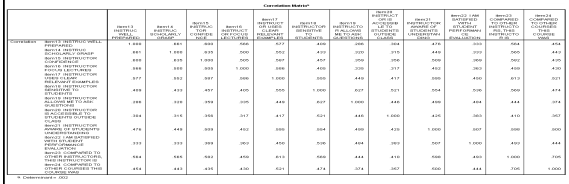
Example raw data

Descriptive Statistics

	Mean [a]	Std. Deviation [b]	Analysis N [c]
INSTRUC WELL PREPARED	4.46	.729	1365
INSTRUC SCHOLARLY GRASP	4.53	.700	1365
INSTRUCTOR CONFIDENCE	4.45	.732	1365
INSTRUCTOR FOCUS LECTURES	4.28	.829	1365
INSTRUCTOR USES CLEAR RELEVANT EXAMPLES	4.17	.895	1365
INSTRUCTOR SENSITIVE TO STUDENTS	3.93	1.035	1365
INSTRUCTOR ALLOWS ME TO ASK QUESTIONS	4.08	.964	1365
INSTRUCTOR IS ACCESSIBLE TO STUDENTS OUTSIDE CLASS	3.78	.909	1365
INSTRUCTOR AWARE OF STUDENTS UNDERSTANDING	3.77	.984	1365
I AM SATISFIED WITH STUDENT PERFORMANCE EVALUATION	3.61	1.116	1365
COMPARED TO OTHER INSTRUCTORS, THIS INSTRUCTOR IS	3.81	.957	1365
COMPARED TO OTHER COURSES THIS COURSE WAS	3.67	.926	1365

54

Example correlations



55

Example: KMO & Bartlett's

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.934
Bartlett's Test of Sphericity	Approx. Chi-Square	8676.712
	df	66
	Sig.	.000

56

Example: Eigenvalues

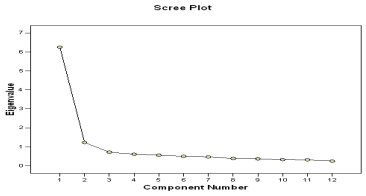
Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.249	52.076	52.076	6.249	52.076	52.076
2	1.229	10.246	62.322	1.229	10.246	62.322
3	.719	5.992	68.313			
4	.613	5.109	73.423			
5	.561	4.676	78.099			
6	.503	4.192	82.291			
7	.471	3.927	86.218			
8	.389	3.240	89.458			
9	.368	3.086	92.524			
10	.328	2.735	95.259			
11	.317	2.645	97.904			
12	.253	2.096	100.000			

Extraction Method: Principal Component Analysis.

57

Example: Scree plot



58

Example components

Component Matrix

	Component	
	1	2
item13 INSTRUC WELL PREPARED	.727	-.449
item14 INSTRUC SCHOLARLY GRASP	.724	-.408
item15 INSTRUCTOR CONFIDENCE	.746	-.308
item16 INSTRUCTOR FOCUS LECTURES	.685	
item17 INSTRUCTOR USES CLEAR, RELEVANT EXAMPLES	.806	
item18 INSTRUCTOR SENSITIVE TO STUDENTS	.755	.366
item19 INSTRUCTOR ALLOWS ME TO ASK QUESTIONS	.641	.497
item20 INSTRUCTOR IS ACCESSIBLE TO STUDENTS OUTSIDE CLASS	.593	.378
item21 INSTRUCTOR AWARE OF STUDENTS UNDERSTANDING	.763	
item22 I AM SATISFIED WITH STUDENT PERFORMANCE EVALUATION	.651	.364
item23 COMPARED TO OTHER INSTRUCTORS, THIS INSTRUCTOR IS	.819	
item24 COMPARED TO OTHER COURSES THIS COURSE WAS	.714	

Extraction Method: Principal Component Analysis.
a. 2 components extracted.

59

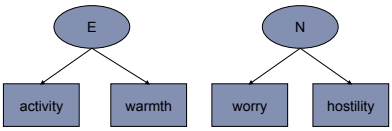
External Validity

- **Factor scores can be used in further analyses**
  - e.g. are there M/F differences in scores on N, E, C?
- **Do the factor scores correlate with other measures**
  - Exam anxiety, subjective reports of life quality, number of friends, exam success…
- **Biological Validity**
  - Map onto brain structures, neurotransmitters, genes

60

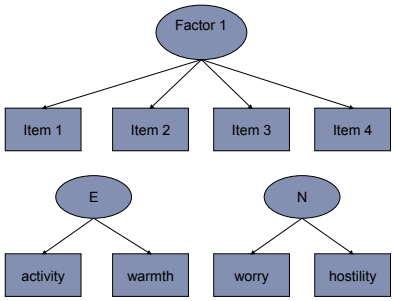
Structural Equation Modeling & Factor Analysis

- SEM incorporates factor analysis and also path analysis
- Confirmatory factor analysis
  - An SEM model in which each factor (latent variable) has multiple indicators but there are no direct effects (arrows connecting the observed variables)



61

Visually…



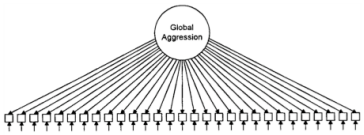
62

Factor Analysis & Path Analysis

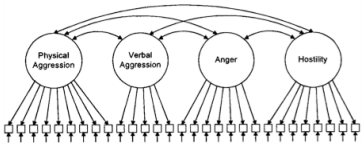
- SEM can be extended to models where each latent variable has several indicators, *and* there are paths specified connecting the latent variables.

63

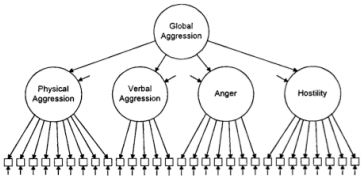
Bryant & Smith (2001)



Bryant & Smith (2001)



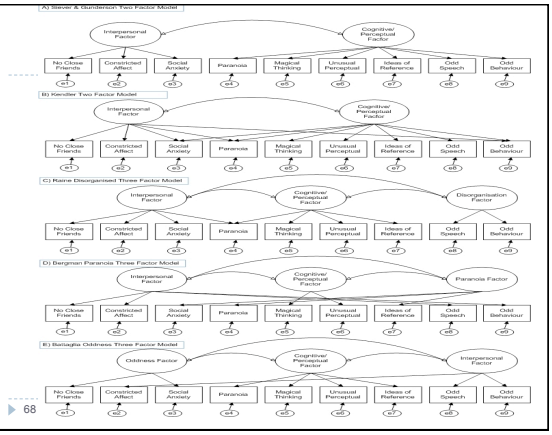
Bryant & Smith (2001)



Bryant & Smith (2001)

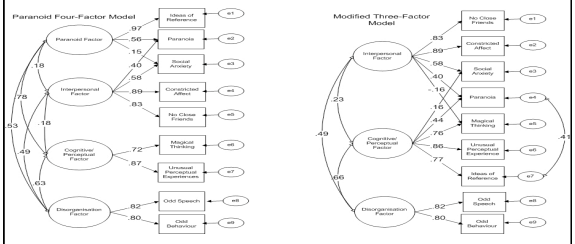
TABLE 1
Goodness-of-Fit Statistics for Various Measurement Models of the AQ Imposed on Samples 1–3

Model	No. items	Sample	Absolute fit measures χ^2	*df*	χ^2/df	GFI	RMSEA	Relative fit measures CFI	NNFI
One-factor (total score)	29	1	1567.9	377	4.2	.70	.102	.66	.64
		2	1267.8	377	3.4	.65	.109	.62	.59
		3	1469.1	377	3.9	.68	.098	.66	.63
Buss & Perry's four factors: PA, VA, ANG, HO	29	1	1042.8	371	2.8	.81	.077	.81	.79
		2	886.4	371	2.4	.76	.084	.78	.76
		3	950.3	371	2.6	.81	.072	.82	.80
Buss & Perry's hierarchical model: one second-order factor	29	1	1046.4	373	2.8	.81	.077	.81	.79
		2	888.5	373	2.4	.76	.083	.78	.76
		3	969.6	373	2.6	.81	.072	.81	.80
Buss & Perry's PA, VA, ANG, & Harris's HO factor	27	1	881.9	318	2.9	.82	.076	.83	.82
		2	734.2	318	2.3	.78	.081	.81	.79
		3	806.6	318	2.5	.83	.071	.83	.81
Four refined factors: PA, VA, ANG, HO	12	1	105.7	48	2.2	.94	.063	.96	.94
		2	92.4	48	1.9	.93	.068	.95	.93
		3	121.7	48	2.5	.94	.071	.91	.87
Refined hierarchical model: one second-order factor	12	1	108.5	50	2.2	.94	.062	.96	.94
		2	94.4	50	1.9	.93	.067	.95	.93
		3	133.6	50	2.7	.93	.074	.90	.86



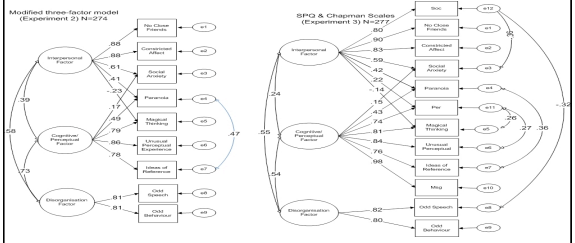
68

SPQ (Wuthrich & Bates, 2006)



69

SPQ Wuthrich & Bates (2006)



70

Summary

- What is factor analysis?
  - Statistical method
  - Accounting for variability in observed traits
    - ("observed random variables")
  - In terms of a smaller number of factors
    - ("unobserved random variables")
  - Allows recovery of values for a subject from a linear combination of the extracted factors.
    - (with some error)
- Can think of the factors as Independent and items as dependent variables

71

Summary cont.

- **What is a scree plot?**
- **What is an identity matrix?**
- **What are communalities?**
- **What is a factor loading?**
- **What is a factor score?**
- **Bartlett's test of sphericity?**
- **KMO?**
- **What is a good number of subjects?**
- **Why do we rotate factors?**
- **Does FA test causes?**
- **How can we model and test causes and (model latent structure?)**

72