




Professor Timothy Bates

tim_bates@ed.ac.uk
<http://timbates.wikidot.com/methodology.1>

Factor Analysis

Statistics to this point...

- t-Test
 - Observed data
 - Countable means and computable variance
 - Computed likelihood of observing a given difference in means, based on the variability of the data



- Nothing unobserved
- No variables created or destroyed

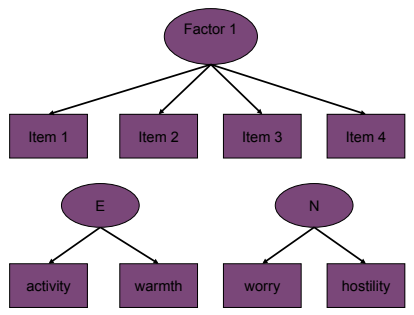
FACTOR ANALYSIS

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

- 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

Factor analysis of 210 subjects by Eyeball

→ 12*12 matrix

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

Questions

- Can you quantify these factors?
- Are they correlated?
- How about 100s of items?

	1	2	3	4	5	6	7	8	9	10	11
1	1										
2	0.27	1									
3	0.37	0.53	1								
4	0.40	0.30	0.38	1							
5	0.17	0.07	0.09	0.08	1						
6	0.17	0.05	0.06	0.10	0.59	1					
7	0.19	0.01	0.05	0.05	0.38	0.42	1				
8	0.06	0.02	0.02	0.02	0.51	0.54	0.48	1			
9	0.25	0.05	0.15	0.20	0.06	0.11	0.14	0.07	1		
10	0.24	0.10	0.09	0.10	0.63	0.62	0.13	0.68	0.39	1	
11	0.21	0.08	0.22	0.12	0.00	0.03	0.07	0.03	0.49	0.38	1
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

7

- Aim of factor analysis is to objectively detect natural groupings of variables (factors)
 - Can deal with large matrices
 - Objective statistical criteria.
 - Can obtain quantitative information
 - e.g. factor scores.
- Factors may 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 Analysis of Example Data

8

	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 responses*factor loadings

+ Factors in the Example Data

9

■ Factor loadings: Correlation of each variable with the underlying factor

	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 score:

- Responses * factor loadings
- $N = (0.75 * \text{Nervous})$
 - + (.73 * Irritable)
 - + (.73 * Insecure)
 - + (.72 * Self-pity)
 - + (-.10 * Extraverted)
 - + (-.21 * Assertive)

+ Requirements for a Valid And Useful Factor Analysis

10

- 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

+ Assumptions & Influences on Reliability

11

- 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.

+ Adequacy of sample size

12

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

+ Item-subject ratios.

13

- ▶ 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

+ Stages Of Analysis

14

1. Examine data for outliers and correlations
2. Choose number of factors
 - Scree plot
 - Parallel analysis (or in Confirmatory Analysis, fit statistics)
3. Rotate if desired
4. Interpret factors
5. Obtain scores
 1. Check reliability of scales defining factors
6. Further experiments to validate factors

+ Statistical Tests For Data Quality and Suitability for Factor Analysis

15

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

+ Bartlett's test of sphericity.

16

- Tests the hypothesis that correlations between variables are greater than would be expected by chance
 - Technically, tests if the matrix is significantly different from an identity matrix (no correlations between items)
- p-Value should be significant
 - i.e., the null hypothesis that all off-diagonal correlations are zero is falsified
 - Accept the alternative hypothesis, that some item-correlations are not 0

+ Two kinds of Matrices

17

- Identity matrix
 - One's 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.

+ KMO Sampling Adequacy

18

- 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

+ Item variance & FA

19

- ▶ Variance of each item can be thought of in three partitions:
 - I. Shared variance
 - ▶ 1. Common variance (accounted for by our factors)
- +
 - II. Unique variance: Not accounted for by other items
 - ▶ 2. Specific variance
 - ▶ 3. Error variance
- ▶ Community
 - ▶ 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

+ Computing a Factor Analysis

20

- Two main approaches
 - Differ in estimating communalities
- Principal components
 - Simplest computationally
 - Assumes all variance is common variance (implausible)
 - Often gives results similar to those of more sophisticated methods.
 - SPSS default!
- Principal factor analysis
 - Estimates communalities first
 - Models error: Much better if the job is to understand mechanisms in the world, rather than account for data (which contain error)

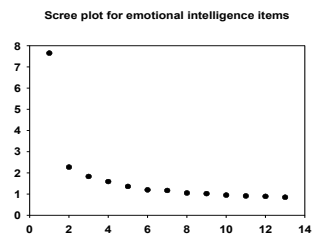
+ How many factors?

21

- There are potentially as many factors as items
 - But we don't want to retain factors which account for little variance.
- The number of factors to extract is based on the eigenvalues of the covariance matrix
 - An eigenvalue is simply the variance explained by a factor
- A common method to choose the number of factors is to plot these eigenvalues: A "scree" plot
 - We choose the number of factors prior to the eigenvalues becoming uniformly small
 - A point of inflection or kink or in the scree plot is a common method of making a cut-off

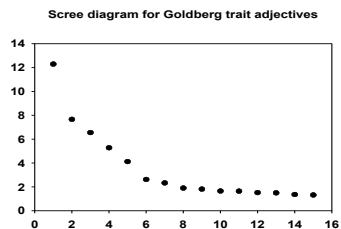
+ Scree plot for E-I: One-factor

22



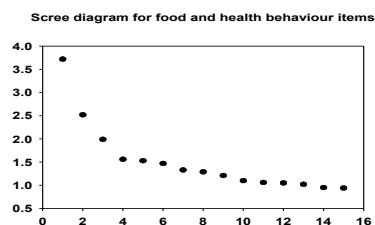
+ Scree plot for Personality: 5 factors

23



+ Food and Health Scree: 3 factors

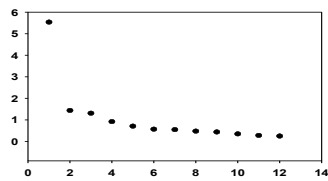
24



+ IQ Scree

25

Scree plot for ability test scores, Swedish Twin Study



+ Other Methods For Factor Numbers

26

- Eigenvalues > 1
 - Eigenvalues sum to the number of items
 - An eigenvalue of >1 is therefore a factor just more informative than a single average item
 - Not a useful guide in practice
- Parallel Analysis
 - Take eigenvalues above the value expected by chance
 - Repeatedly randomise the correlation matrix (1000s of times)
 - Determine how large each eigenvalue appears by chance
 - Excellent method
- Theory-driven
 - Extract a number of factors based on theoretical considerations
 - Explaining this approach requires SEM

+ How to align the factors?

27

- We've chosen how many, now where to put them?
- The initial solution is "un-rotated"
- Two undesirable features make this hard to interpret:
 - Maximises the loadings of all items on the first factor
 - Most items have large loadings on more than one factor
- Solution: Rotation
 - Reveals "Simple Structure"

+ Unrotated Factors For The Example Data

28

	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

Hides groupings in the data

+ Example Data Factors after rotation

29

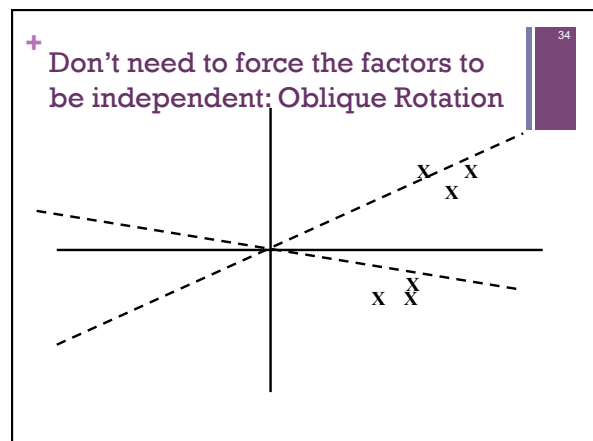
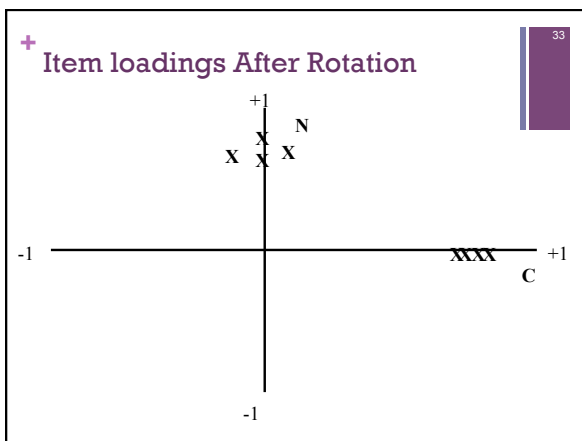
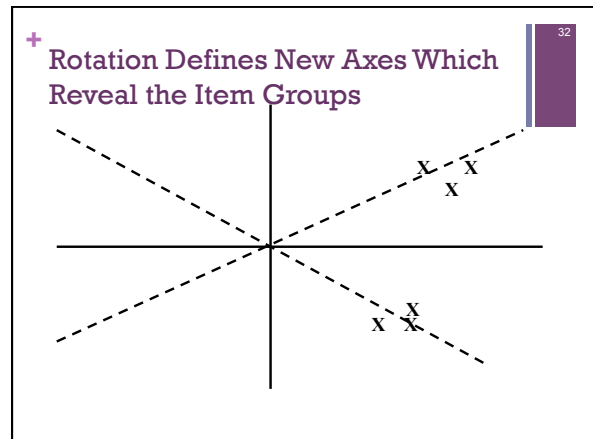
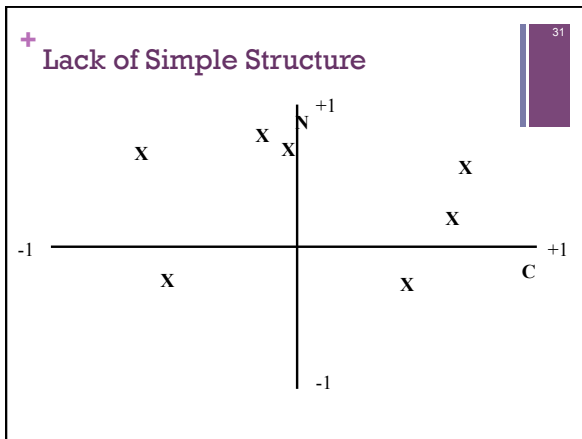
- Factors are now close to simple structure
- Are cross-loadings informative?

	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	0.34	-0.21	0.66

+ ROTATION – DETAIL (1)

30

- Rotation shows up the groups of items in the data.
- Orthogonal rotation
 - All factors independent of each other
- Oblique rotation
 - Factors allowed (but not forced) to correlate
- Theoretical reasons to choose a type of rotation
 - (e.g. independent for personality traits);
- If in doubt, choose oblique
 - If there are large correlations between factors, it will reflect this
 - If not, the output will be orthogonal anyway.



- ### + 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

- ### + 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?

+ 1: Interpreting Factors

37

- 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.

+ 2: Factor Scores

38

- 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).

+ 3: Scale Reliability

39

- Assessing Reliability of Factor-Derived Scales
 - Factor Derived Scales can be used like any other scale
 - For instance using Cronbach's Alpha
- Cronbach's alpha of the scale
 - Adequate reliability is customarily set at 0.7 or above
 - Low reliability might indicate
 - under-factoring
 - too few items/bad construction/no construct
- Item alpha
 - Check item-alpha to identify poorly-functioning items
 - Delete these from the scale

+ Interpreting Factors

40

- Done on the basis of 'large' loadings
 - Often taken to be above 0.3.
 - Size of loading which should be considered depends 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-colinearity problems.

+ Factor Scores

41

- Factor scores
 - Estimate 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 E scores, farm annual profit, farm area).

+ External Validity

42

- 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

+ Example – Factor Structure Of Dietary Behaviour

43

- Research question: Is there a dimension of healthy vs. unhealthy diet preferences?
 - (MacNicol et al 2003)
- 451 schoolchildren completed a 38-item questionnaire mainly on regularly consumed foods
 - Subjects/variables = 13:1, Population not representative for SES.
- Scree suggested 3 factors, 2 diet-related
 - F1: Unhealthy foods (chips, fizzy drinks etc.)
 - F2: Healthy foods (fruit, vegetables etc.)
- Validation
 - Higher SES and better nutrition knowledge associated with healthier eating patterns.
- Factor reliabilities low
 - Problem of yes/no items
 - Sample heterogeneity (sex effects?)

+ Attitudes to chocolate: How many factors?

44

- Four ideas
 - 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
 - Prepared to work harder to get chocolate buttons

+ Attitudes to Chocolate

45

- 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
 - 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

+ Chocolate: 2 correlated Factors (Cramer & Hartleib, 2001)

46

- 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
- Both highly correlated.

+ Example – Abnormal Personality

47

- 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

+ Good place to take a break

+ A worked example in SPSS

49

- FA is found in the menus under: Analyse: data reduction: 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

+ Example raw data

50

Descriptive Statistics			
	Mean ^a	Std. Deviation ^b	Analysis N ^c
INSTRUCTOR WELL PREPARED	4.46	.729	1365
INSTRUCTOR 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	1.118	1365
COMPARED TO OTHER INSTRUCTORS, THIS INSTRUCTOR IS	3.81	.957	1365
COMPARED TO OTHER COURSES THIS COURSE WAS	3.67	.926	1365

+ Example correlations

51

Correlation Matrix												
	INSTRUCTOR WELL PREPARED	INSTRUCTOR SCHOLARLY GRASP	INSTRUCTOR CONFIDENCE	INSTRUCTOR FOCUS LECTURES	INSTRUCTOR USES CLEAR RELEVANT EXAMPLES	INSTRUCTOR SENSITIVE TO STUDENTS	INSTRUCTOR ALLOWS ME TO ASK QUESTIONS	INSTRUCTOR IS ACCESSIBLE TO STUDENTS OUTSIDE CLASS	INSTRUCTOR AWARE OF STUDENTS UNDERSTANDING	I AM SATISFIED WITH STUDENT PERFORMANCE EVALUATION	COMPARED TO OTHER INSTRUCTORS, THIS INSTRUCTOR IS	COMPARED TO OTHER COURSES THIS COURSE WAS
INSTRUCTOR WELL PREPARED	1.000											
INSTRUCTOR SCHOLARLY GRASP	.691	1.000										
INSTRUCTOR CONFIDENCE	.605	.645	1.000									
INSTRUCTOR FOCUS LECTURES	.500	.500	.500	1.000								
INSTRUCTOR USES CLEAR RELEVANT EXAMPLES	.517	.542	.547	.588	1.000							
INSTRUCTOR SENSITIVE TO STUDENTS	.455	.433	.437	.483	.505	1.000						
INSTRUCTOR ALLOWS ME TO ASK QUESTIONS	.500	.530	.500	.535	.449	.437	1.000					
INSTRUCTOR IS ACCESSIBLE TO STUDENTS OUTSIDE CLASS	.514	.516	.556	.511	.411	.511	.485	1.000				
INSTRUCTOR AWARE OF STUDENTS UNDERSTANDING	.476	.444	.408	.451	.504	.484	.459	.478	1.000			
I AM SATISFIED WITH STUDENT PERFORMANCE EVALUATION	.479	.577	.498	.493	.486	.488	.461	.467	.488	1.000		
COMPARED TO OTHER INSTRUCTORS, THIS INSTRUCTOR IS	.504	.500	.500	.483	.513	.503	.481	.473	.505	.483	1.000	
COMPARED TO OTHER COURSES THIS COURSE WAS	.484	.443	.438	.430	.511	.474	.374	.387	.500	.484	.503	1.000

+ Example: KMO & Bartlett's

52

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

+ Example: Eigenvalues

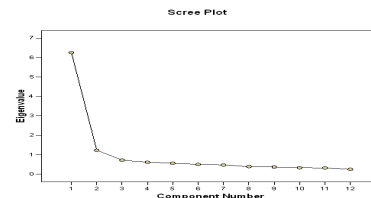
53

Component	Total Variance Explained			Extraction Sums of Squared Loadings ^a		
	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	.368	3.040	89.258			
9	.368	3.040	92.298			
10	.328	2.736	95.034			
11	.317	2.645	97.679			
12	.252	2.095	100.000			

Extraction Method: Principal Component Analysis.

+ Example: Scree plot

54



+ Example components

55

Component Matrix^{a,b}

	Component ^c	
	1	2
Item13 INSTRUCTOR WELL PREPARED	.727	-.449
Item14 INSTRUCTOR 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.

+ How to Assess Factor Analysis in research reports

56

- ▶ Sample size:
 - ▶ Total sample size
 - ▶ Ratio of subjects to items
 - ▶ 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.
 - ▶ Three or more strong loadings per factor
- ▶ 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 exploit sample-dependent noise
 - ▶ Use of unit weightings of items is more replicable in practice

+ Alternatives To Factor Analysis

- Alternative Analytic Tools
 - Spectral Clustering
 - Example from HEXACO vs 5 factor model
- Structural Equation Modeling
 - Examples from Abnormal Psychology

+ Spectral Clustering

- Factor analysis (FA) and more confirmatory factor analysis (CFA; Jöreskog, 1969) remain the primary techniques for exploring structure in questionnaire data
 - Foundation for the FFM
- Attempts to confirm the FA derivation of the FFM had mixed success
 - (Marsh et al, 2010; Gignac, Bates & Jane, 2007)
- Attempts to use non FA-based methods are in their infancy
 - (Tiliopoulos, Pallier and Coxon, 2010)
- Brocklebank and Bates (2011)
 - Responses from 20,993 subjects to a 300-item questionnaire chosen to assess the Five Factor Model

+ Spectral Clustering

- Developed in multi-object automated scene analysis
- Projects items onto an n-dimensional "sphere"
- Builds clusters to maximise the variance accounted for of the location of items in this n-dimensional space

+ HEXACO

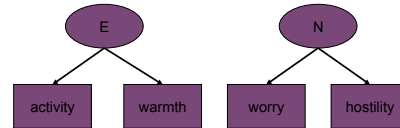
- An actively researched alternative to the 5 factor-model of personality
- Suggests that personality consists of, not five, but six factors
 - (Ashton & Lee, 2007)
- Honesty
- Emotionality
- eXtraversion
- Agreeableness
- Conscientiousness
- Openness

+ 5FM and HEXACO

- Humility: Gains morality, modesty, honesty
- Emotion: Gains openness to emotion
- eXtraversion: Unchanged
- A: Gains empathy; Loses morality and modesty
- C: loses honesty
- O loses empathy and feeling of emotions

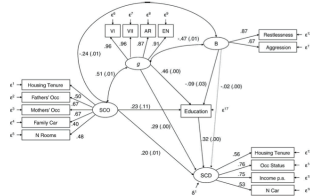
+ Structural Equation Modeling & Factor Analysis

- 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)



+ Factor Analysis & Path Analysis

- SEM can be extended beyond factor analysis, where measurement instruments are modeled, to path analysis, where causal inferences are made, and models with both latent variables (factors) and paths specified connecting the latent variables



+ Buss & Perry (1992)

- 4 factors
- 29 items

Table 1
Four Aggression Factors

Item	Factor	Factor loading
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, .62, .60
4. I get into fights a little more than the average person.		.61, .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, .40, .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.		.51, .49, .49
2. When frustrated, I get my irritation show.		.47, .45, .37
3. I sometimes feel like a powder keg ready to explode.		.60, .35, .35
4. I am an air-ventilating person.		.64, .62, .69
5. Some of my friends think I'm a hothead.		.63, .51, .64
6. Sometimes I fly off the handle for no good reason.		.71, .66, .70
7. I have trouble controlling my temper.		.74, .66, .69
Hostility		
1. I am sometimes eaten up with jealousy.		.41, .41, .49
2. At times I feel I have gotten a raw deal out of life.		.61, .58, .55
3. Other people always seem to get the breaks.		.63, .63, .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

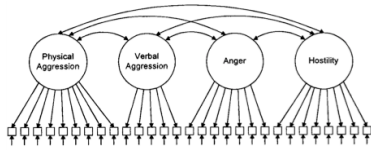
TABLE 3
CFA Factor Loadings for the Refined 12-Item, Four-Factor Measurement Model of the AQ

AQ items	PA sample			VA sample			ANG sample			HO sample		
	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 controlling 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

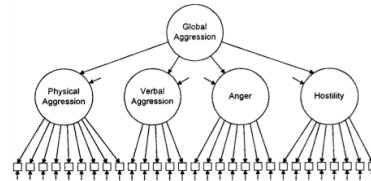
+ Bryant & Smith (2001)



+ Bryant & Smith (2001)



+ Bryant & Smith (2001)

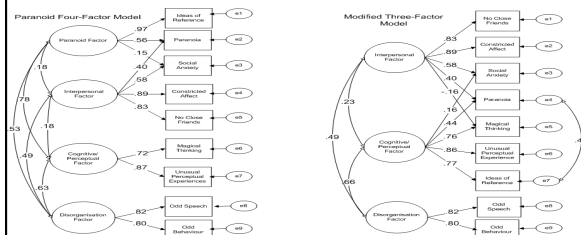


+ Bryant & Smith (2001)

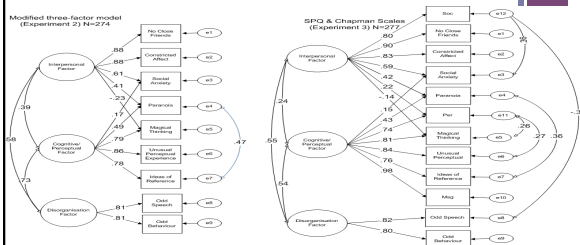
TABLE 1
Goodness-of-Fit Statistics for Various Measurement Models of the AQ Imposed on Samples 1-3

Model	No. items	Sample	Absolute fit measures					Relative fit measures	
			χ^2	df	χ^2/df	GFI	RMSEA	CFI	NNFI
One-factor (total score)	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	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	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	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	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	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	

+ Schizotypal Personality (Wuthrich & Bates, 2006)



+ SPQ Wuthrich & Bates (2006)



+ Study Summary

- What is factor analysis?
 - Statistical method compactly accounting for variance 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 variables and items as dependent variables

+ Summary cont.

73

- 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? (how?)
- Does FA test causes?
- How can we model and test causes and (model latent structure?)