Latent Variables: Factor Analysis and Reliability

Jeremy Miles

Latent Variables

- Dictionary definition of latent:
  - Present or potential but not evident or active: e.g. *latent talent.*
- Something which is not measured directly
  - Existence of it is inferred in some way
- Examples:
  - Disease severity, quality of life, depression/anxiety

Quality of Life Example

- "Quality of Life"
  - Somewhat "fuzzy" concept
  - Does it really exist *as a unitary construct*?
- E.g. SF-36
  - Divides into domains
  - Mental functioning
  - Physical functioning
- Do these combine to form "quality of life"
  - Or are they two separate (but related) constructs

Depression / Anxiety

- Are depression and anxiety different things?
  - Two latent variables?
- Are they the same thing
  - One latent variable?

Where is it from?

- Spearman (1904)
  - General Intelligence, Objectively Measured and Defined (American Journal of Psychology)
- Lots of work by e.g. Spearman, Thurstone, Cattell, Jöreskog, many others

Why?

- To determine the nature of intelligence
  - one thing (*g*)
  - Several related things (*g* + *s*)
  - A few unrelated things
  - many unrelated things
- In terms of dimensionality

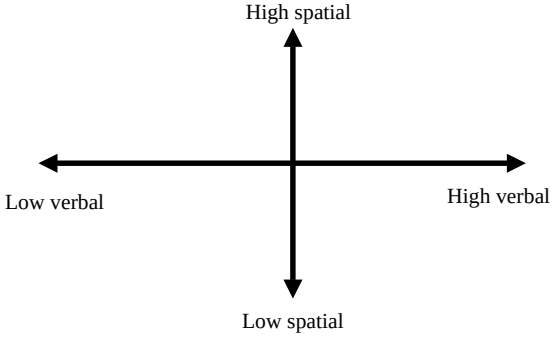
What?

- Dimensionality
  - the number of dimensions (scales) that you need to describe something
- A bottle:
  - size
- A Box
  - width, height, depth
- A Car
  - colour, price, size, speed, reliability, seats, doors, etc

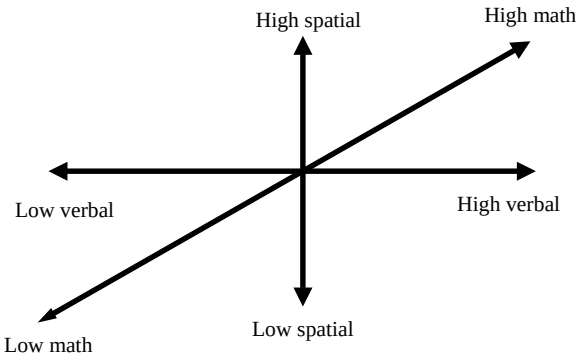
1 Dimensional Intelligence

Low High

2 Dimensional Intelligence



3 Dimensional Intelligence



Thurstone's Box Problem

- Sent students home to measure boxes, in different ways
- Added some error
- Factor analysed the results
- Found that boxes were
  - three dimensional
  - width, height, depth

Etc . . .

Use of Factor Analysis

- Three main (related) uses
  1. Is your scale measuring one thing, or more than one related thing
     - QoL – one thing, more than one thing?
  2. What is the nature of the thing that you are measuring?
     - Is it one thing, is it multiple things
     - Depression / anxiety
  3. Reducing the dimensionality of a set of variables
     - I've measured 50 variables, which ones do I include?
     - Factor analyse them to reduce to a smaller number of factors

How does it do that?

- Very, very mathematical
- Very, very hard work
  - early factor analysts were mad
  - an analysis took months of work
- We
  - don't care
  - don't need to know
  - wouldn't understand if we did know or care

So Simpler . . .

- Looks at correlations between items
- Tries to explain them more parsimoniously

.74				
.85	.87			
.04	.14	.11		
-.05	.10	.06	.89	
.09	.02	.03	.87	.91

.74				
.85	.87			
.04	.14	.11		
-.05	.10	.06	.89	
.09	.02	.03	.87	.91

- Two subscales - measuring similar things
  - a, b, c
  - d, e, f
- Called *FACTORS*

Labelling Factors

- Need to look at the variables that make up each factor
- Make sure they are psychologically meaningful
  - random numbers will produce factors

Labelling Factors

- Factor 1
  - Feel miserable, unhappy, lacking motivation
- Factor 2
  - Unable to walk up stairs, drive, run
- What are the factor names?
  - ..... , ...
- Could we label them as the opposite ends?

Another Example: I-GHQ

Interval general Health Questionnaire

1. been feeling unhappy and depressed?
2. been having restless and disturbed nights?
3. found everything getting 'on top' of you?
4. been thinking of yourself as a worthless person?
5. felt constantly under strain?
6. been feeling nervous and strung up all the time?
7. felt that life is entirely hopeless?
8. been losing confidence in yourself?
9. been getting scared or panicky for no good reason?
10. lost much sleep over worry?
11. felt that life isn't worth living?
12. found at times you couldn't do anything because your nerves were too bad?

Method

- Administered by post to ~400 people
  - Sufferers from psoriasis
- In the last 4 weeks, how often have you felt ...
- Response scale:
  - 0: Never
  - 1: Sometimes
  - 2: Frequently
  - 3: Nearly always

Correlation Matrix

2.	.49									
3.	.61	.53								
4.	.47	.36	.44							
5.	.57	.47	.69	.44						
6.	.62	.45	.66	.44	.69					
7.	.46	.33	.50	.63	.44	.53				
8.	.47	.33	.47	.61	.49	.56	.			
9.	.49	.33	.47	.45	.45	.61	.56	.59		
10.	.49	.64	.56	.36	.58	.54	.36	.42	.47	
11.	.41	.29	.41	.56	.39	.45	.78	.53	.53	.39
12.	.37	.28	.41	.43	.36	.46	.51	.48	.47	.35	.55
	1	2	3	4	5	6	7	8	9	10	11

Correlation Matrix

2.	.49									
3.	**.61**	.53								
4.	.47	*.36*	.44							
5.	.57	.47	**.69**	.44						
6.	**.62**	.45	**.66**	.44	**.69**					
7.	.46	*.33*	.50	**.63**	.44	.53				
8.	.47	*.33*	.47	**.61**	.49	.56	.			
9.	.49	*.33*	.47	.45	.45	.61	.56	.59		
10.	.49	**.64**	.56	.36	.58	.54	*.36*	.42	.47	
11.	.41	.29	.41	.56	.39	.45	**.78**	.53	.53	*.39*
12.	*.37*	*.28*	.41	.43	.36	.46	.51	.48	.47	*.35*
	1	2	3	4	5	6	7	8	9	10

.55

11

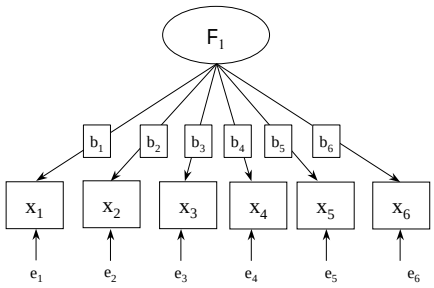
Factor Analysis of the I-GHQ

- Gives us two factors
- There are two “things” being measured
- Results of factor analysis
  - Factor loading matrix
- A series of multiple regressions
  - Your score on an item is determined by:
    - Your score on the factor(s)
    - Error

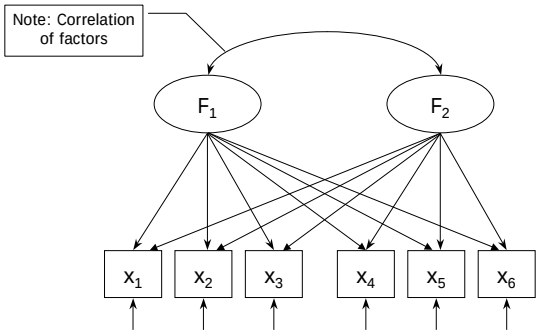
Factor Loadings as Regressions

- Series of regressions
  - $x_1 = b_{1.1}F_1 + b_{1.2}F_2 \ldots + b_{1.k}F_k + e_1$
  - $x_2 = b_{2.1}F_1 + b_{2.2}F_2 \ldots + b_{2.k}F_k + e_2$
- Much easier to think of it diagrammatically
  - Factors (latent variables) circles
  - Measured variables (items) squares

Path Diagram – 1 Factor



Path Diagram: 2 Factors



Item	F1	F2
1. been feeling unhappy and depressed?	.63	-.15
2. been having restless and disturbed nights?	.72	.08
3. found everything getting 'on top' of you?	.77	-.07
4. been thinking of yourself as a worthless person?	.10	-.64
5. felt constantly under strain?	.78	-.02
6. been feeling nervous and strung up all the time?	.66	-.21
7. felt that life is entirely hopeless?	-.08	-.93
8. been losing confidence in yourself?	.20	-.60
9. been getting scared or panicky for no good reason?	.29	-.49
10. lost much sleep over worry?	.78	.05
11. felt that life isn't worth living?	-.12	-.90
12. found at times you couldn't do anything because your nerves were too bad?	.08	-.59

- Simplify results in two ways:
  - Small loadings are boring
    - Remove <0.10
  - Sort the items by the size of the loadings

Item	F1	F2
10. lost much sleep over worry?	.78	
5. felt constantly under strain?	.78	
3. found everything getting 'on top' of you?	.77	
2. been having restless and disturbed nights?	.72	
6. been feeling nervous and strung up all the time?	.66	.21
1. been feeling unhappy and depressed?	.63	.15
7. felt that life is entirely hopeless?		.93
11. felt that life isn't worth living?	-.12	.90
4. been thinking of yourself as a worthless person?		.64
8. been losing confidence in yourself?	.20	.60
12. found at times you couldn't do anything because your nerves were too bad?		.59
9. been getting scared or panicky for no good reason?	.29	.49

Label Factors

- Can think of it as:
  - Item 10 = F1 * 0.78
  - Item 5 = F1 * 0.78
    - Etc
- What is Factor 1 measuring?
  - .....
- What is factor 2 measuring?
  - .....

The Process of Factor Analysis

- Stage 1:
  - Extract appropriate number of factors
- Stage 2:
  - Rotate factors

Stage 1: How Many Factors to Extract?

- We want to distinguish between *real* factors and *junk* factors
  - Don't want the junk
- Two (main) ways to do it
  - Both involve eigenvalues

Eigenvalues

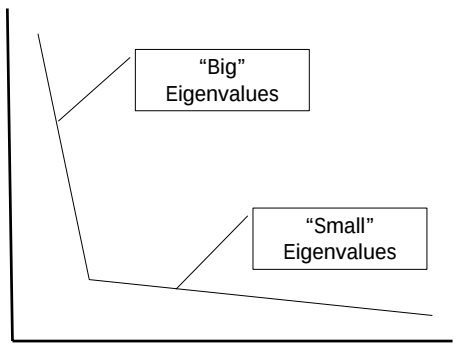
- Fairly complicated and mathematical to understand where they come from
  - Easy to see how they are used
- Each factor has an eigenvalue associated with it
  - Bigger the eigenvalue, the more variance the factor accounts for
- Total of all eigenvalues = number of items
  - In GHQ example, 12 items, so sum of eigenvalues = 12

Using Eigenvalues (1)

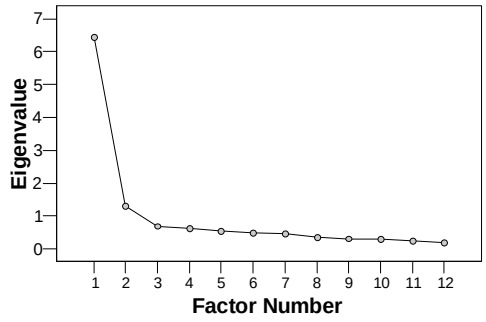
- The Kaiser criterion
  - Select items with eigenvalues greater than 1
    - Bigger than average
- From I-GHQ analysis
- **6.443 1.316** .696 .635 .554 .489 .467 .352 .317 .297 .247 .186
  - Two eigenvalues greater than 1
    - Two factors

Using Eigenvalues (2)

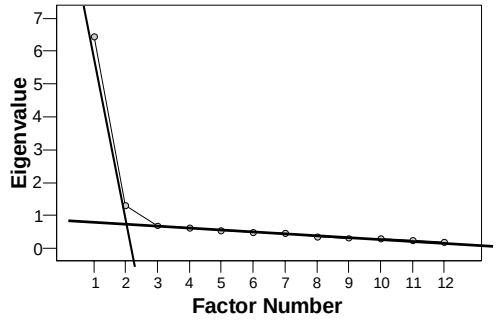
- Catell's Scree Plot
  - Plot the eigenvalues against factor numbers
    - See if there is a difference between the big ones and the small ones
    - Like a scree, at the side of a mountain



Our Scree Plot



Our Scree Plot



Proportion of Variance

- Total amount of variance in the correlation matrix ≈ information
  - How much information have we retained?
- Our eigenvalues = 6.4 + 1.3 = 7.7
  - 7.7 / 12 = 64.6 ≈ 65%
- We have a 2 factor solution
  - 6 times simpler
  - Retains 2/3 of information

More on Extraction

- Requires some (semi-) arbitrary decisions
  - Sometimes interpretability also used as a criterion
- Some other statistics:
  - Kaiser Meyer Olkin Measure of Sampling Adequacy (KMO-MSA) – measure of data quality
  - Bartlett's test of sphericity: not useful

Stage 2: Rotation

- Rotation increases interpretability
- Only 1 decision
  - Orthogonal rotation or oblique rotation
- Different types, most common:
  - Varimax: orthogonal
  - Direct oblimin or promax: oblique

Orthogonal versus Oblique

- Orthogonal:
  - Keeps factors uncorrelated
- Oblique:
  - Allows factors to correlate
- Orthogonal (particularly varimax)
  - Much easier to interpret
- Oblique
  - Probably closer to the truth

Reliability

- Reliability means two things:
  - Internal consistency
  - Temporal stability
- Internal consistency
  - Closely related to factor analysis
  - Measured using Coefficient (Cronbach's) Alpha

Cronbach's Alpha

- Alpha estimates
  - The correlation between the measured variable (the sum of the scores) and
  - The latent variable
- Higher alpha (0.7, 0.8) more acceptable reliability

Confirmatory Factor Analysis

- We have been discussing *exploratory* factor analysis (EFA)
- *Confirmatory* factor analysis (CFA) also exists
- Exploratory
  - Explore the data, see what we have
- Confirmatory
  - Confirm a pre-determined structure
- CFA: More complex