Psychometric Instrument Development

Lecture 6

Survey Research & Design in Psychology

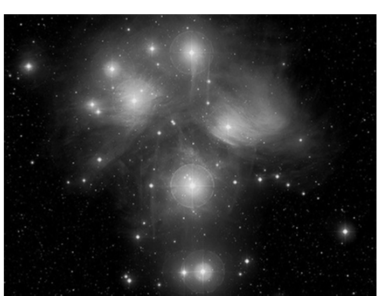
James Neill, 2018

Overview

1 Recap: Exploratory factor analysis
2 Concepts & their measurement
3 Measurement error
4 Psychometrics
5 Reliability & validity
6 Composite scores
7 Writing up instrument development

2

Recap: Exploratory Factor Analysis



What is factor analysis?

- Factor analysis is:
  - a *family* of multivariate correlational methods used to identify clusters of covariance (called factors)
- Two main purposes:
  - Theoretical (PAF)
  - Data reduction (PC)
- Two main types (extraction methods):
  - Exploratory factor analysis (EFA)
  - Confirmatory factor analysis (CFA)

4

EFA steps

1 Test assumptions
- Sample size
  - 5+ cases x no. of variables (min.)
  - 20+ cases x no. of variables (ideal)
  - Another guideline: $N > 200$
- Outliers & linearity
- Factorability - Use any of:
  - Correlation matrix: Some > .3?
  - Anti-image correlation matrix diags > .5
  - Measures of Sampling Adequacy:
    - KMO > ~ .5 to 6
    - Bartlett's sig?

5

EFA steps

2 Select type of analysis
- Extraction
  - Principal Components (PC)
  - Principal Axis Factoring (PAF)
- Rotation
  - Orthogonal (Varimax)
  - Oblique (Oblimin)

6

EFA steps

3. Determine no. of factors
   - Theory?
   - Kaiser's criterion?
   - Eigen Values and Scree plot?
   - % variance explained?
   - Interpretability of weakest factor?

7

EFA steps

4. Select items
   - Use factor loadings to help identify which items belong in which factor
   - Drop items one at a time if they don't belong to any factor e.g., consider any items for which
     - primary (highest) loading is low? (< .5 ?)
     - cross- (other) loading(s) are high? (> .3 ?)
     - item wording doesn't match the meaning of the factor

8

EFA steps

5 Name and describe factors
6 Examine correlations amongst factors
7 Analyse internal reliability
8 Compute composite scores
9 Check factor structure across sub-groups

Covered in this lecture

9

EFA example 4: University student motivation

10

Example EFA: University student motivation

- 271 UC students responded to 24 student motivation statements in 2008
- 8-point Likert scale (False to True)
- For example:
  "I study at university … "
  - to enhance my job prospects.
  - because other people have told me I should.
- EFA PC Oblimin revealed 5 factors

11

Example EFA: Pattern matrix

	Component				
	1	2	3	4	5
motiv15	.964				
motiv20	.914				
motiv25	.779				
motiv10	.750				
motiv05	.713				
motiv09		.955			
motiv14		.922			
motiv24		.912			
motiv04		.885			
motiv19		.765			
motiv07			-.906		
motiv22			-.884		
motiv17			-.883		
motiv01			-.876		
motiv12			-.734		
motiv03			-.725		
motiv13				.925	
motiv23				.862	
motiv18				.847	
motiv11					.817
motiv21					.767
motiv02					.740
motiv16			-.248		.664
motiv06					.628

This is a pattern matrix showing factor loadings. It shows a simple factor structure.

Primary loadings for each item are above .5

Cross-loadings are all below .3.

12

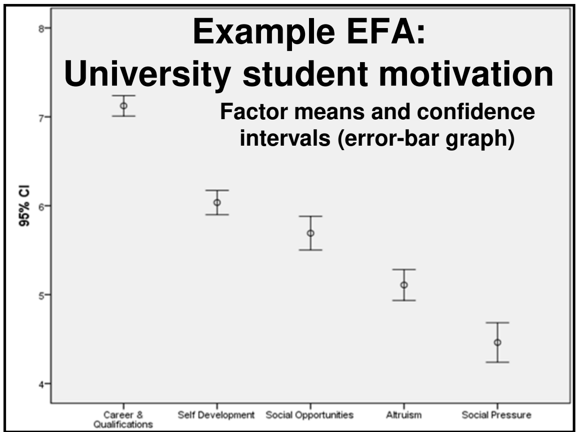
Example EFA: University student motivation

1. Career & Qualifications
   (6 items; α = .92)
2. Self Development
   (5 items; α = .81)
3. Social Opportunities
   (3 items; α = .90)
4. Altruism
   (5 items; α = .90)
5. Social Pressure
   (5 items; α = .94)

13

Example EFA: University student motivation

Factor means and confidence intervals (error-bar graph)



Example EFA: University student motivation

Factor correlations

Motivation	Self Development	Social Enjoyment	Altruism	Social Pressure
Career & Qualifications	**.26**	**.25**	**.24**	.06
Self Development		**.33**	**.55**	-.18
Social Enjoyment			**.26**	**.33**
Altruism				.11

15

Psychometric Instrument Development

Image source: http://commons.wikimedia.org/wiki/File:Soft_ruler.jpg, CC-by-SA 3.0

Readings: Psychometrics

1 Bryman & Cramer (1997).
Concepts and their measurement. [UCLearn Reading List]
2 DeCoster, J. (2000).
Scale construction notes. [Online]
3 Howitt & Cramer (2005).
Reliability and validity: Evaluating the value of tests and measures. [Textbook/UCLearn Reading List]
4 Howitt & Cramer (2014).
Ch 37: Reliability in scales and measurement: Consistency and measurement. [Textbook/UCLearn Reading List]
5 Wikiversity.
Composite scores. [Online]
Measurement error. [Online]
Reliability and validity. [Online]

18

Concepts and their measurement

Operationalising fuzzy concepts

Image source: https://www.flickr.com/photos/lenore-m/441579944/in/set-72157600039854497

Concepts and their measurement

Bryman & Cramer (1997)

Concepts

- express common elements in the world (to which we give a name)
- form a linchpin in the process of social research

20

Concepts and their measurement

Bryman & Cramer (1997)

Hypotheses

- specify expected relations between concepts

21

Concepts and their measurement

Bryman & Cramer (1997)

Operationalisation

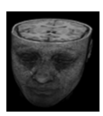
- A concept needs to be ***operationally defined*** in order to be systematically researched.
- "An ***operational definition*** specifies the procedures (operations) that will permit differences between individuals in respect of the concept(s) concerned to be precisely specified ..."

22

Concepts and their measurement

Bryman & Cramer (1997)

"... What we are in reality talking about here is ***measurement**,* that is, the assignment of numbers to the units of analysis - be they people, organizations, or nations - to which a concept refers."



23

Operationalisation

- The act of making a ***fuzzy concept*** measurable.
- Social science often uses ***multi-item measures*** to assess related but distinct aspects of a fuzzy concept.

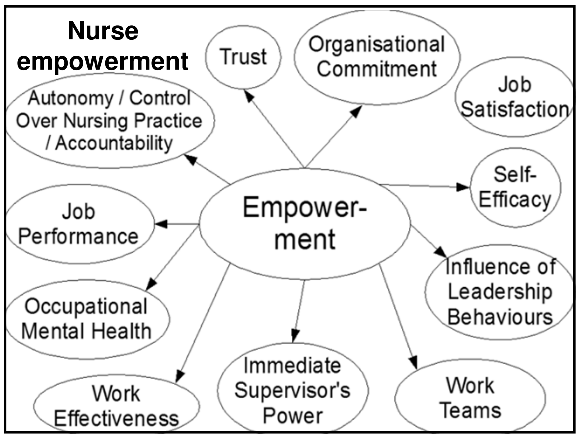
Operationalisation steps

1 Brainstorm indicators of a concept
2 Define the concept
3 Draft measurement items
4 Pre-test and pilot test
5 Examine psychometric properties
 - how precise are the measures?
6 Redraft/refine and re-test

25

Operationalisation
Example (Brainstorming indicators)

Nurse empowerment



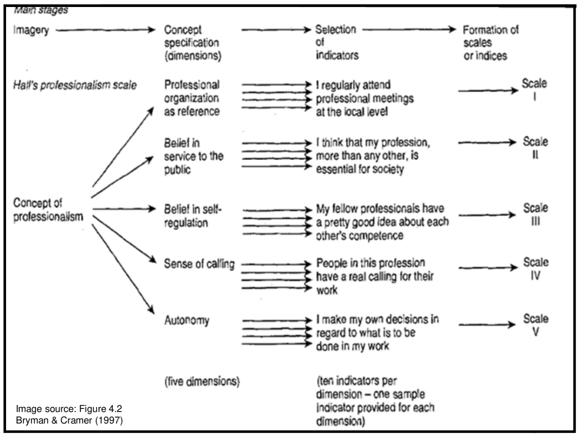


Image source: Figure 4.2
Bryman & Cramer (1997)

Measurement error

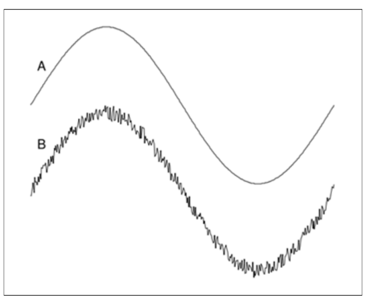


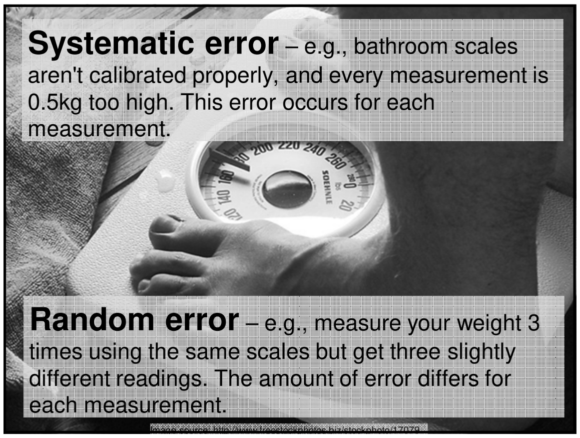
Image source: http://commons.wikimedia.org/wiki/File:Noise_effect.svg

Measurement error

Measurement error is statistical deviation from the **true value** caused by the measurement procedure.

- **Observed score** =
  true score +/- measurement error
  - **Measurement error** =
    systematic error +/- random error
    - Systematic error =
      sampling error +/- non-sampling error

30



Sources of systematic error

Non-sampling

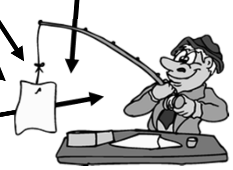
Test reliability & validity
(e.g., unreliable or invalid tests)

Respondent bias
(e.g., social desirability)

Researcher bias
(e.g., researcher favours a hypothesis)

Paradigm
(e.g., focus on positivism)

Sampling
(non-representative sample)



32

Measurement precision & noise

- The lower the measurement precision, the more participants are needed to make up for the "noise" in the measurements.
- Even with a larger sample, noisy data can be hard to interpret.
- Especially when testing and assessing individual clients, special care is needed when interpreting results of noisy tests.

http://www.sportsci.org/resource/stats/precision.htm

33

Minimising measurement error

- Standardise administration conditions with clear instructions and questions
- Minimise potential demand characteristics (e.g., train interviewers)
- Use multiple indicators for fuzzy constructs

34

Minimising measurement error

- Obtain a representative sample:
  - Use probability-sampling, if possible
  - For non-probability sampling, use strategies to minimise selection bias
- Maximise response rate:
  - Pre-survey contact
  - Minimise length / time / hassle
  - Rewards / incentives
  - Coloured paper
  - Call backs / reminders

35

Minimising measurement error

- Ensure administrative accuracy:
  - Set up efficient coding, with well-labelled variables
  - Check data (double-check at least a portion of the data)

36

Psychometrics

Image source: http://commons.wikimedia.org/wiki/File:Information_icon4.svg

Psychometrics: Goal

To validly measure differences between individuals and groups in psychosocial qualities such as attitudes and personality.

38

Psychometrics: Tasks

- Develop approaches and procedures (theory and practice) for measuring psychological phenomena
- Design and test psychological measurement instrumentation (e.g., examine and improve reliability and validity of psychological tests)

39

Psychometrics: In demand

"Psychometrics, one of the most obscure, esoteric and cerebral professions in America, is now also one of the hottest."
- As test-taking grows, test-makers grow rarer, David M. Herszenhor, May 5, 2006, New York Times

Psychometricians are in demand due to increased testing of educational and psychological capacity and performance.

40

Psychometrics: Methods

- Factor analysis
  - Exploratory
  - Confirmatory
- Classical test theory
  - Reliability
  - Validity

41

Reliability and Validity

Image source: http://www.flickr.com/photos/psd/17433783/in/photostream

Reliability and validity
(Howitt & Cramer, 2005)

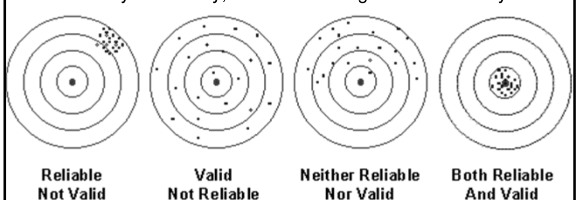
Reliability and validity ("classical test theory") are ways of evaluating the accuracy of psychological tests and measures.

- Reliability is about consistency of
  - items within the measure
  - the measure over time
- Validity is about whether the measure actually measures what it is intended to measure.

43

Reliability vs. validity

In classical test theory, reliability is generally thought to be necessary for validity, but it does not guarantee validity.



In practice, a test of a relatively changeable psychological construct such as suicide ideation, may be valid (i.e., accurate), but not particularly reliable over time (because suicide ideation is likely to fluctuate).

Reliability vs. validity

Reliability
- A car which starts every time is **reliable**.
- A car which only starts sometimes is **unreliable**.

Validity
- A car which always reaches the desired destination is **valid**.
- A car which misses the desired destination is **not valid**.

Image source: https://commons.wikimedia.org/wiki/File:Aiga_carrental_cropped.svg

Reliability and validity
(Howitt & Cramer, 2005)

- Reliability and validity are not inherent characteristics of measures. They are affected by the context and purpose of the measurement → a measure that is valid for one purpose may not be valid for another purpose.

46

Reliability

Reproducibility of a measurement

Reliability: Types

- **Internal consistency**: Correlation among multiple items in a factor
  - Cronbach's Alpha (α)
- **Test-retest reliability**: Correlation between test at one time and another
  - Product-moment correlation (*r*)
- **Inter-rater reliability**: Correlation between one observer and another:
  - Kappa

48

Reliability: Rule of thumb

< .6 = Unreliable
.6 = OK
.7 = Good
.8 = Very good, strong
.9 = Excellent
> .95 = may be overly reliable or redundant – this is subjective and depends on the nature what is being measured

49

Reliability: Rule of thumb

Table 7 Fabrigar et al (1999).

Table 7 Fabrigar et al. (1999)

Variable	*Journal of Personality and Social Psychology* *N*	%	*Journal of Applied Psychology* *N*	%
Average reliability of variables				
Less than .60	3	1.9	2	3.4
.60–.69	6	3.8	5	8.6
.70–.79	33	20.8	9	15.5
.80–.89	33	20.8	11	19.0
.90–1.00	14	8.8	9	15.5
Unknown	70	44.0	22	37.9

Rule of thumb - reliability coefficients should be over .70, up to approx. .95

Internal consistency
(or internal reliability)

Internal consistency refers to:
- How well multiple items combine as a measure of a single concept
- The extent to which responses to multiple items are consistent with one another

Internal consistency can measured by:
- Split-half reliability
- Odd-even reliability
- Cronbach's Alpha (α)

51

Internal consistency
(recoding)

If dealing with a mixture of positively and negatively scored items, remember to recode so that all items are measured i the same direction.

52

Internal consistency: Split-half reliability

- Sum the scores for the first half (e.g., 1, 2, 3) of the items.
- Sum the scores for the second half (e.g., 4, 5, 6) of the items.
- Compute a correlation between the sums of the two halves.

53

Internal consistency: Odd-even reliability

- Sum the scores for odd items (e.g.,1, 3, 5)
- Sum the scores for even items (e.g., 2, 4, 6)
- Compute a correlation between the sums of the two halves.

54

Internal consistency: Cronbach's alpha (α)

- Averages all possible split-half reliability coefficients.
- Akin to a single score which represents the degree of intercorrelation amongst the items.
- Most commonly used indicator of internal reliability.

55

How many items per factor?

- More items → greater reliability
  (The more items, the more "rounded" the measure)
- Minimum items to create a factor is 1.
- No maximum. Law of diminishing returns = each additional item will add less and less to the reliability.
- Typically ~ 3 to 10 items per factor are used.
- Final decision is subjective and depends on research context

56

Internal reliability example

Student-rated quality of maths teaching

- 10-item scale measuring students' assessment of the educational quality of their maths classes
- 4-point Likert scale ranging from: strongly disagree to strongly agree

57

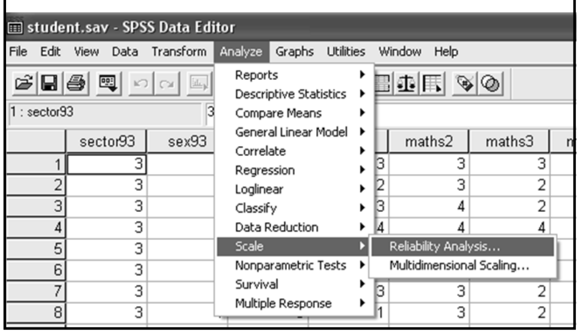
Internal reliability example
Quality of mathematics teaching

1. My maths teacher is friendly and cares about me.
2. The work we do in our maths class is well organised.
3. My maths teacher expects high standards of work from everyone.
4. My maths teacher helps me to learn.
5. I enjoy the work I do in maths classes.

+ 5 more

58

Internal reliability example
Quality of maths teaching



Internal reliability example
SPSS: Corrected Item-total correlation

Item-total correlations should be > ~.5

Reliability Statistics

Cronbach's Alpha	N of Items
.885	10

A measure for examining the relationship between individual items and the total scale, this is the correlation between the given item and the item sum if the given item is not included in the scale. Smaller values indicate the given item is not well correlated with the others.

Item-To

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
school1	41.15	98.608	.438	.888
school2	40.04	91.500	.648	.872

Internal reliability example

SPSS: Cronbach's α

Reliability Statistics

Cronbach's Alpha	N of Items
.885	1[illegible]

If "Cronbach's α if item deleted" is higher than the α, consider removing item.

A measure for examining the relationship between individual items and the total scale, this is the value of Cronbach's Alpha for the remaining items if the given item is not included in the scale.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
school1	41.15	98.608	.438	.888
school2	40.04	91.500	.648	.872

Internal reliability example

SPSS: Reliability output

Item-total Statistics

Item Deleted	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
MATHS1	25.2749	25.5752	.6614	.8629
MATHS2	25.0333	26.5322	.6235	.8661
MATHS3	**25.0192**	**30.5174**	**.0996**	**.9021**
MATHS4	24.9786	25.8671	.7255	.8589
MATHS5	25.4664	25.6455	.6707	.8622
MATHS6	25.0813	24.9830	.7114	.8587
MATHS7	25.0909	26.4215	.6208	.8662
MATHS8	25.8699	25.7345	.6513	.8637
MATHS9	25.0340	26.1201	[illegible]	.8623
MATHS10	25.4642	25.7578	.6495	.8638

N of Cases = 1353.0

Alpha = .8790

Remove this item.

Maths3 does not correlate well with the other items and the Cronbach's alpha would increase without this item.

62

Internal reliability example

SPSS: Reliability output

Item-total Statistics

Item Deleted	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
MATHS1	22.2694	24.0699	.6821	.8907
MATHS2	22.0280	25.2710	.6078	.8961
MATHS4	21.9727	24.4372	.7365	.8871
MATHS5	22.4605	24.2235	.6801	.8909
MATHS6	22.0753	23.5423	.7255	.8873
MATHS7	22.0849	25.0777	.6166	.8955
MATHS8	22.8642	24.3449	.6562	.8927
MATHS9	22.0280	24.5812	.7015	.8895
MATHS10	22.4590	24.3859	.6524	.8930

N of Cases = 1355.0 N of Items = 9

Alpha improves

Alpha = .9024

63

Table * . Definitions of the Life Effectiveness Questionnaire dimensions, with Internal Consistency and Test-Retest Correlations

LEQ 8-factor model	Description	3 items per scale	
		Test-Retest *r*	Alpha
Achievement Motivation	Motivation to achieve excellence and put the required effort into action to attain it.	.68	.87
Active Initiative *	Initiating action in new situations.	.73	.81
Emotional Control	Maintaining emotional control when faced with potentially stressful situations.	.75	.87
Intellectual Flexibility	Adapting thinking and accommodating new information from changing conditions and different perspectives.	.60	.78
Self Confidence *	Confidence in abilities and the success of actions.	.73	.84
Social Competence	Ability in and success of social interactions.	.75	.86
Task Leadership	Ability to lead other people effectively when a task needs to be done and productivity is the primary requirement.	.81	.82
Time Management	Makes optimum use of time.	.75	.84
Total	Effective in generic life skills.	.72	.84
N		.67	.93

Validity

Validity is the extent to which an instrument actually measures what it purports to measure.

Validity = does the test measure what its meant to measure?

Validity

- Validity is multifaceted and includes:
  - Comparing wording of the items with theory and expert opinion
  - Examining correlations with similar and dissimilar measures
  - Testing how well the measure predicts the future

66

Validity: Types

- Face validity
- Content validity
- Criterion validity
  - Concurrent validity
  - Predictive validity
- Construct validity
  - Convergent validity
  - Discriminant validity

67

Face validity
(low-level of importance overall)

- **Asks**:
  "At face-value, do the questions appear to measure what the test purports to measure?"
- **Important for**:
  Respondent buy-in
- **How assessed**:
  Read the test items

68

Content validity
(next level of importance)

- **Asks**:
  "Are questions measuring the complete construct?"
- **Important for**:
  Ensuring holistic assessment
- **How assessed**:
  Diverse item generation (lit. review, theory, interviews, expert review)

69

Criterion validity
(high importance)

- **Asks**:
  "Can a test score predict real world outcomes?"
- **Important for**:
  Test relevance and usefulness
- **How assessed**:
  Concurrent validity: Correlate test scores with recognised external criteria such as performance appraisal scores
  Predictive validity: Correlate test scores with future outcome e.g., offender risk rating with recidivism

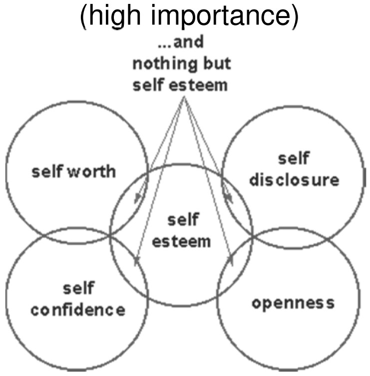
70

Construct validity
(high importance)

- **Asks**:
  Does the test assess the construct it purports to?
  ("the truth, the whole truth and nothing but the truth")
- **Important for**:
  Making inferences from operationalisations to theoretical constructs
- **How assessed**:
  - **Theoretical** (is the theory about the construct valid?)
  - **Statistical**
    Convergent – correlation with similar measures
    Discriminant – not correlated with other constructs

71

Construct validity
(high importance)



Composite Scores

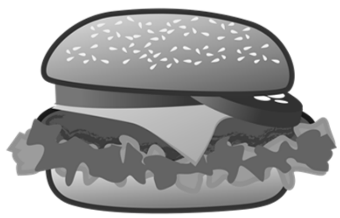


Image source: https://commons.wikimedia.org/wiki/File:PEO-hamburger.svg

Composite scores

Combine item-scores into an overall factor score which represents individual differences for the target construct.
The new composite score can then be used for:

- Descriptive statistics and histograms
- Correlations
- As IVs and/or DVs in inferential analyses such as MLR and ANOVA

74

Composite scores

Ways of creating composite scores:

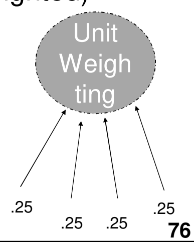
- Unit weighting
- Regression weighting

75

Unit weighting

Average (or total) of item scores within a factor.
(each variable is equally weighted)

$X = \text{mean}(y_1 \ldots y_p)$



76

Composite scores: Missing data

To maximise the sample size, consider computing composite scores in a way that allows for some missing data.

77

Composite scores: Missing data

SPSS syntax:

```
Compute X = mean (v1, v2, v3, v4, v5, v6)
Compute X = mean.4 (v1, v2, v3, v4, v5, v6)
```

Specifies a min. # of items. If the min. isn't available, the composite score will be missing.

In this example, X will be computed for a case when the case has responses to at least 4 of the 6 items.

How many items can be missed? Depends on overall reliability. A rule of thumb:

- Allow 1 missing per 4 to 5 items
- Allow 2 missing per 6 to 8 items
- Allow 3+ missing per 9+ items

A researcher may decide to be more or less conservative depending on the factors' reliability, sample size, and the nature of the study.

78

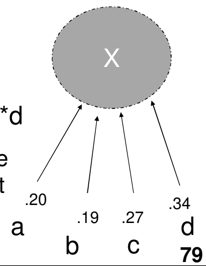
Regression weighting

Factor score regression weighting

The contribution of each item to the composite score is weighted to reflect responses to some items more than other items.

X = .20*a + .19*b + .27*c + .34*d

This is arguably more valid, but the advantage may be marginal, and it makes factor scores between studies more difficult to compare.

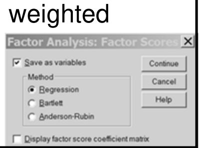


79

Regression weighting

Two calculation methods:

- Manual (use **Compute** – New variable name = MEAN.*(list of variable names separated by commas) - Unit weighted
- Automatic (use **Factor Analysis** – Factor Scores – Save as variables - Regression) - Regression weighted



Regression weighting

Variable view: of variables auto-calculated through SPSS factor analysis

64	FAC1_1	Numeric	11	5	REGR factor score 1 for analysis 1	N
65	FAC2_1	Numeric	11	5	REGR factor score 2 for analysis 1	N
66	FAC3_1	Numeric	11	5	REGR factor score 3 for analysis 1	N
67	FAC4_1	Numeric	11	5	REGR factor score 4 for analysis 1	N
68	FAC5_1	Numeric	11	5	REGR factor score 5 for analysis 1	N
69	FAC6_1	Numeric	11	5	REGR factor score 6 for analysis 1	N
70	FAC7_1	Numeric	11	5	REGR factor score 7 for analysis 1	N
71	FAC8_1	Numeric	11	5	REGR factor score 8 for analysis 1	N
72	FAC9_1	Numeric	11	5	REGR factor score 9 for analysis 1	N

Data view: Data are standardised, centred around 0

FAC1_1	FAC2_1	FAC3_1	FAC4_1	FAC5_1	FAC6_1	FAC7_1	FAC8_1	FAC9_1
.46	.41	-4.41	-1.29	.93	.26	-2.63	.99	-1.21
-1.34	-1.90	3.17	-1.06	-.10	1.95	-1.39	.66	-.08
-.36	-.02	1.61	-1.27	-2.05	-1.77	-.74	.72	1.00
.51	-.09	.11	.56	1.05	-.72	-.93	1.06	-.17
.30	-.54	-.14	2.65	-.54	.11	1.82	.53	1.23
-.01	1.18	.56	-.26	1.35	-1.36	-.58	-1.06	-.63
-1.91	-1.74	1.73	-.36	-2.47	1.34	.37	.86	-.38

Writing up instrument development

Writing up instrument development

- Introduction
  - Review previous literature about the construct's underlying factors – consider both theory and research
  - Generate a research question e.g., "What are the underlying factors of X?".
  - Could also make a hypothesis about the number of factors and what they will represent.

83

Writing up instrument development

- Method
  - Materials – summarise the design and development of the measures and the expected factor structure
    e.g., present a table of the expected factors and their operational definitions.

84

Writing up instrument development

- Results
  - Factor analysis
    - Assumption testing
    - Extraction method & rotation
    - # of factors, with names and definitions
    - # of items removed and rationale
    - Item factor loadings & communalities
    - Factor correlations
  - Reliability for each factor
  - Composite scores for each factor
  - Correlations between factors

85

Writing up instrument development

- Discussion
  - Theoretical underpinning – Was it supported by the data? What adaptations should be made to the theory?
  - Quality / usefulness of measure – Provide an objective, critical assessment, reflecting the measures' strengths and weaknesses
  - Recommendations for further improvement
- Writing up a factor analysis
  - Download examples: http://goo.gl/fD2qby

86

Summary

87

Summary: Psychometrics

1. Science of psychological measurement
2. Goal: Validly measure individual psychosocial differences
3. Design and test psychological measures e.g., using
   1. Factor analysis
   2. Reliability and validity

88

Summary: Concepts & their measurement

1. Concepts name common elements
2. Hypotheses identify relations between concepts
3. Brainstorm indicators of a concept
4. Define the concept
5. Draft measurement items
6. Pre-test and pilot test
7. Examine psychometric properties
8. Redraft/refine and re-test

89

Summary: Measurement error

1. Deviation of measure from true score
2. Sources:
   1. Non-sampling (e.g., paradigm, respondent bias, researcher bias)
   2. Sampling (e.g., non-representativeness)
3. How to minimise:
   1. Well-designed measures
   2. Representative sampling
   3. Reduce demand effects
   4. Maximise response rate
   5. Ensure administrative accuracy

90

Summary: Reliability

1 Consistency or reproducibility
2 Types
  1 Internal consistency
  2 Test-retest reliability
3 Rule of thumb
  1 > .6 OK
  2 > .8 Very good
4 Internal consistency
  1 Split-half
  2 Odd-even
  3 Cronbach's Alpha

91

Summary: Validity

1 Extent to which a measure measures what it is intended to measure
2 Multifaceted
  1 Compare with theory and expert opinion
  2 Correlations with similar and dissimilar measures
  3 Predicts future

92

Summary: Composite scores

Ways of creating composite (factor) scores:

1. **Unit weighting**
   1. Total of items or
   2. Average of items (recommended for lab report)
2. **Regression weighting**
   1. Each item is weighted by its importance to measuring the underlying factor (based on regression weights)

93

Summary: Writing up instrument development

1. **Introduction**
   1. Review constructs & previous structures
   2. Generate research question or hypothesis
2. **Method**
   1. Explain measures and their development
3. **Results**
   1. Factor analysis
   2. Reliability of factors
   3. Descriptive statistics for composite scores
   4. Correlations between factors
4. **Discussion**
   1. Theory? / Measure? / Recommendations?

94

References

1 Allen, P. & Bennett, K. (2008). *Reliability analysis* (Ch 15) in SPSS for the health & behavioural sciences (pp. 205-218). South Melbourne, Victoria, Australia: Thomson.

2 Bryman, A. & Cramer, D. (1997). Concepts and their measurement (Ch. 4). In *Quantitative data analysis with SPSS for Windows: A guide for social scientists* (pp. 53-68). Routledge. UCLearn Reading List.

3 DeCoster, J. (2000). *Scale construction notes*.

4 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.

5 Fowler, F. (2002). Designing questions to be good measures. In *Survey research methods* (3rd ed.)(pp. 76-103). Thousand Oaks, CA: Sage. UCLearn Reading List.

6 Howitt, D. & Cramer, D. (2005). Reliability and validity: Evaluating the value of tests and measures (Ch. 13). In *Introduction to research methods in psychology* (pp. 218-231). Harlow, Essex: Pearson. UCLearn Reading List.

95

Next lecture

Multiple linear regression I

- Correlation (Review)
- Simple linear regression
- Multiple linear regression

96