ISYS3401
IT Evaluation

Week 10: Measurement of Constructs
Scale Reliability

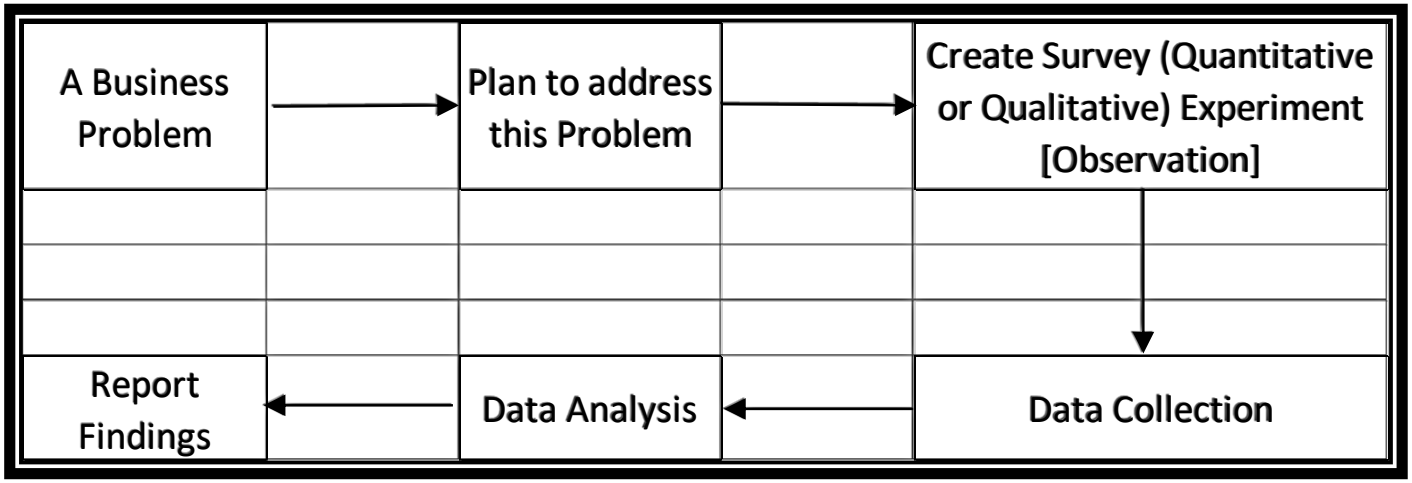
Vincent.Pang@sydney.edu.au

THE UNIVERSITY OF SYDNEY

Agenda

- Recap on what we have done so far…
- Extra Information On Measurement
- Scale Reliability
  - Excel
  - SPSS
- Final Model (Using Regression)

So what have we done so far…



Now look at the photo you took in Week 1

So what have we done so far…

Model and Questions

1. Learn about TAM model, and how it can be used to evaluate a system
2. Examine some of questions used in TAM model (plus reading of Chapter 6)
3. Learn about Qualitative Research – how not to ask a question, and the use of NVivo to analyse text
4. Understand the needs of Pilot Testing

Data Analysis

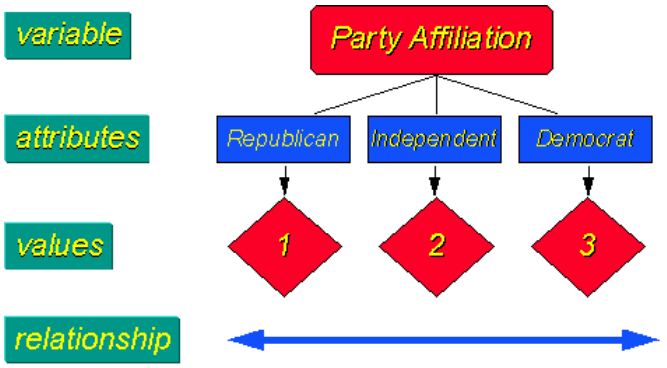
1. Assuming you have collected data, you use Factor Analysis to check the constructs/factors using
2. Check Convergent (>=0.60) and Discriminant Validity (<=0.30)
3. Internal Consistency Reliability (Today)
4. Create your Final Model (next week)

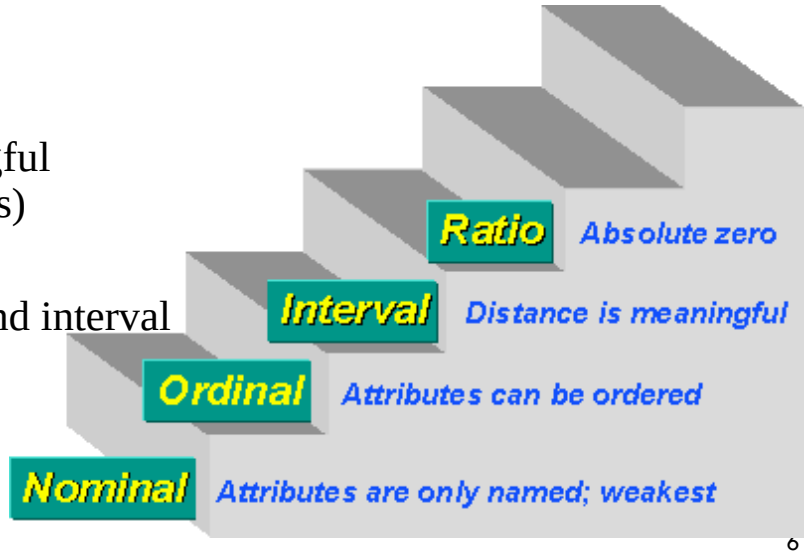
Extra Information on Measurements

THE UNIVERSITY OF SYDNEY

Rating Scales (Level of Measurement)

- The relationship among the values that are assigned to the attributes of a variable
  - The value that an indicator can take
- General Rating Scales
  1. Nominal
     - Numerical values just "name" the attributes uniquely
     - Attributes cannot be rank-ordered
  2. Ordinal
     - Attributes can be rank-ordered
     - Distance between attributes does not have any meaning
  3. Interval
     - Attributes can be rank-ordered
     - Distance between attributes is meaningful (equidistant between adjacent attributes)
  4. Ratio
     - All the qualities of nominal, ordinal, and interval scales
     - "True zero" point (absolute zero)
     - Ratio (or fraction) is meaningful





Specific Rating Scales

1. **Binary:** "Yes" or No", or "True" or "False"

2. **Likert:** A **Likert scale** is a psychometric scale commonly involved in research that employs questionnaires. It is the most widely used approach to scaling responses in survey research, such that the term (or more accurately the **Likert-type scale**) is often used interchangeably with *rating scale*, even though the two are not synonymous.

3. **Guttman:** A set of items if they can be ordered in a reproducible hierarchy. For example, a series of items could be (1) "I am willing to be near ice cream"; (2) "I am willing to smell ice cream"; (3) "I am willing to eat ice cream"; and (4) "I love to eat ice cream"

4. **Semantic Differential** is a scale used for measuring the meaning of things and concepts. There are two aspects of meaning: **denotative and connotative.** The semantic differential measures connotative meaning.
   - *Denotation* - what a name or concept refers to (*denote* - to mark out plainly, to indicate)
   - *Connotation* - the suggestive significance of a word, apart from its explicit and recognized meaning

1. Specific Rating Scales: Binary Scale

- Nominal scale consisting of binary items that assume one of two possible values (e.g., yes or no, true or false, etc.)

Have you ever written a letter to a public official	Yes	No
Have you ever signed a political petition	Yes	No
Have you ever donated money to a political cause	Yes	No
Have you ever donated money to a candidate running for public office	Yes	No
Have you ever written a political letter to the editor of a newspaper or magazine	Yes	No
Have you ever persuaded someone to change his/her voting plans	Yes	No

Table 6.2. A six-item binary scale for measuring political activism

2. Specific Rating Scales: Likert Scale

- Includes simply-worded statements to which respondents can indicate their extent of agreement or disagreement on a multiple-point scale ranging from "strongly disagree" to "strongly agree"
- Number of values
  - Odd number allows for a "neutral anchor"
  - Even number forces respondents to agree or disagree
- Ordinal scale in theory (anchors not necessarily equidistant), treated like interval scale in practice (anchors are assumed equidistant)

	Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
I feel good about my job	1	2	3	4	5
I get along well with others at work	1	2	3	4	5
I'm proud of my relationship with my supervisor at work	1	2	3	4	5
I can tell that other people at work are glad to have me there	1	2	3	4	5
I can tell that my coworkers respect me	1	2	3	4	5
I feel like I make a useful contribution at work	1	2	3	4	5

Table 6.3. A six-item Likert scale for measuring employment self-esteem

3. Specific Rating Scales: Guttman Scale

- **A composite scale uses a series of items arranged in increasing order of intensity** of the construct of interest, from **least intense to most intense.**

How will you rate your opinions on the following statements about immigrants?		
Do you mind immigrants being citizens of your country	Yes	No
Do you mind immigrants living in your own neighborhood	Yes	No
Would you mind living next door to an immigrant	Yes	No
Would you mind having an immigrant as your close friend	Yes	No
Would you mind if someone in your family married an immigrant	Yes	No

Table 6.5. A five-item Guttman scale for measuring attitude toward immigrants

4. Specific Rating Scales: Semantic Differential Scale

- A composite (multi-item) scale where respondents are asked to indicate their opinions or feelings toward a single statement using different pairs of adjectives framed as polar opposites

How would you rate your opinions on national health insurance?						
	Very much	Somewhat	Neither	Somewhat	Very much	
Good	□	□	□	□	□	Bad
Useful	□	□	□	□	□	Useless
Caring	□	□	□	□	□	Uncaring
Interesting	□	□	□	□	□	Boring

Table 6.4. A semantic differential scale for measuring attitude toward national health insurance

Topic: Scale Reliability

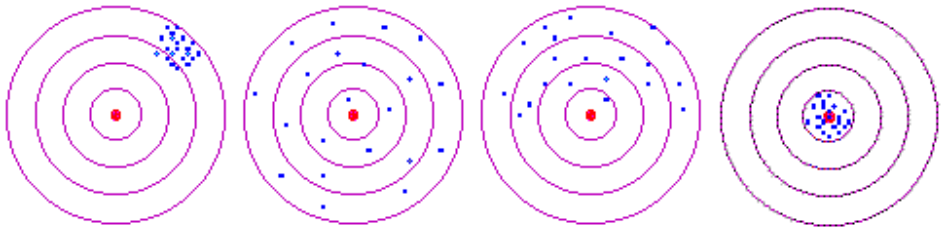
Reference:

http://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=1002&context=oa_textbooks **(Chapter 6 and 7)**

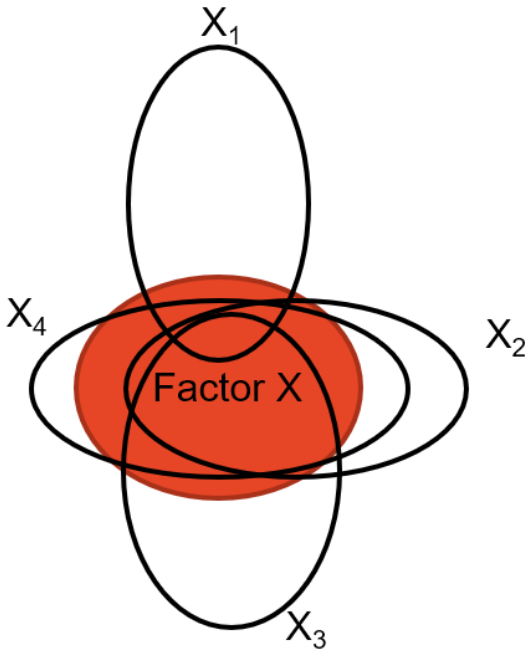
THE UNIVERSITY OF SYDNEY

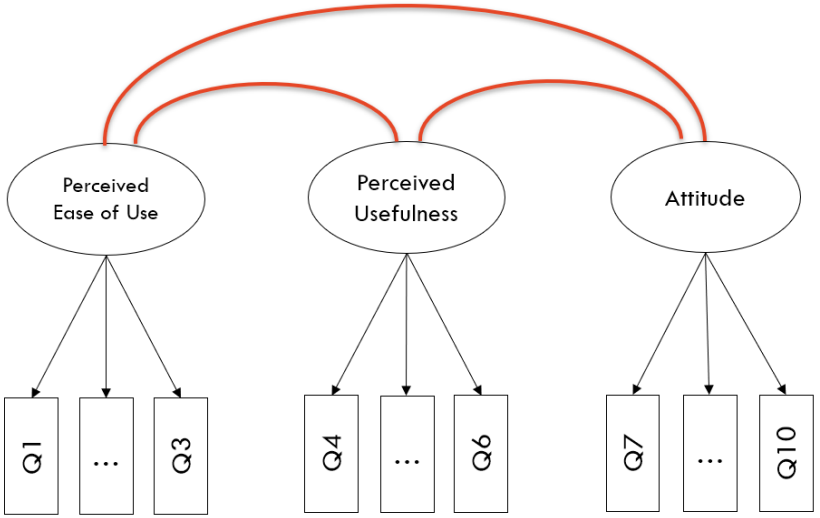
Recap: Psychometric Properties of Measurement Scales

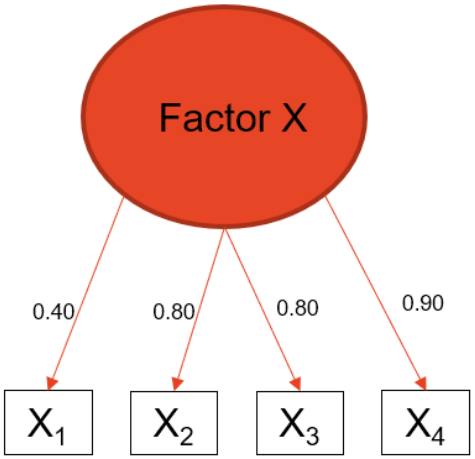
- Reliability (correlation & variance!!)
  - The degree to which the measure of a construct is consistent or dependable
  - Sources of unreliable measurements
    - Observer's (researcher's) subjectivity
    - Imprecise or ambiguous questions
    - Questions about issues that respondents are not very familiar about or care about
- Construct Validity
  - The extent to which a measure adequately represents the underlying construct that it is supposed to measure
  - Validity of the measurement procedures
  - Distinct from the validity of hypotheses testing procedures
    - Internal validity
    - External validity



Recall: Use Factor analysis to heck if the variables belong to the same Factor/Construct.







Recap: Convergent and Discriminant Validity

- Convergent validity
  - Comparing the observed values of one indicator of one construct with that of other indicators of the same construct
  - Demonstrating similarity (or high correlation) between values of these indicators
  - **Same-factor loadings ≥ 0.60**
- Discriminant validity
  - Demonstrating that indicators of one construct are dissimilar from (i.e., have low correlation with) other constructs
  - **Cross-factor loadings ≤ 0.30**

	KL1	KL2	KL3	PF1	PF2	PF3
KL1	1.00	0.83	0.79	0.23	0.21	0.19
KL2		1.00	0.75	0.11	0.20	0.03
KL3			1.00	0.03	-0.11	0.17
PF1				1.00	0.84	0.91
PF2					1.00	0.77
PF3						1.00

High correlations between items of the same construct (convergent validity)

Low correlations between items of different constructs (discriminant validity)

Rotated Component Matrix[a]

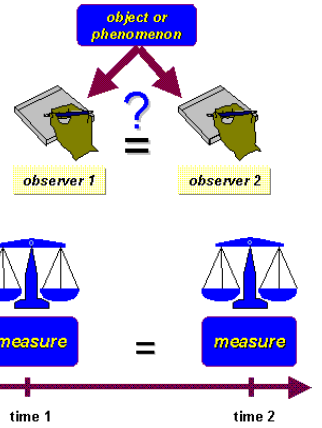
	Component		
	1	2	3
PU1		.938	
PU2		.913	
PU3		.896	
PEOU1	.964		
PEOU2	.970		
PEOU3	.968		
PEOU4	.720		.389
ATT1			.849
ATT2			.926
ATT3			.882

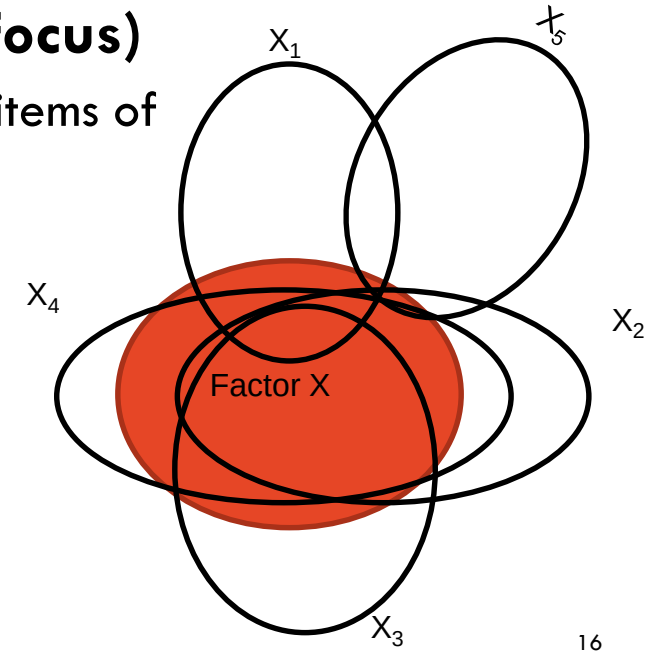
Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

Reliability

- Inter-Rater Reliability
  - A measure of consistency between two or more independent rater (observers) of the same construct
- Test-Retest Reliability
  - A measure of consistency between two measurements (tests) of the same construct administered to the same sample at two different points in time
- **Internal Consistency Reliability (main focus)**
  - A measure of consistency between different items of the same construct
  - Methods
    1. Average inter-item correlation
    2. Average item-to-total correlation
    3. Split-half reliability
    4. Cronbach's alpha (α)





Internal Consistency Reliability

1. After Factor Analysis (last week), we got the constructs/factors and the variables associated with each construct.
2. We now want to check if the variables for a selected construct is reliable, i.e. we are performing **internal consistency reliability** test. For instance, you can either take a sample of your data (sample) or you can use the whole set of data (population)

Demonstration for Consistency Reliability Test

Response	Q1	Q2	Q3	Q4	Q1+Q2	Q3+Q4	Q1toQ4
1	1	4	2	2	5	4	9
2	2	4	5	4	6	9	15
3	3	5	4	4	8	8	16
4	4	5	4	3	9	7	16
5	5	3	5	5	8	10	18
Variance	2.5	0.7	1.5	1.3	2.7	5.3	11.7
cov	4.75	-0.2	3.75	3.4	4.55	7.15	11.7
cor	0.8783	-0.0699	0.8951	0.8718			

	Q1	*Q2*	*Q3*	*Q4*	*Q1+Q2*	*Q3+Q4*	*Q1toQ4*
Q1	1						
Q2	- 0.1890	1					
Q3	0.6455	-0.2440	1				
Q4	0.6934	-0.4193	0.8951	1			
Q1&2	0.8660	0.3273	0.4969	0.4537	1		
Q3&4	0.6868	-0.3375	0.9753	0.9715	0.4890	1	
Q1to4	0.8783	-0.0699	0.8951	0.8718	0.8095	0.9080	1

Average inter-item correlation:	0.230	
Average item-to-total correlation:	0.644	
Split-Half Correlation (Q1+Q2) vs (Q3+Q4):	0.489	
Split-Half Reliability (Q1+Q2) vs (Q3+Q4):	0.657	
Alpha (Unstandardised) (Q1toQ4):	0.650	
Standardised Cronbach's alpha (Q1toQ4):	0.545	Poor

Formulae

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N} = \frac{\sum X^2}{N} - \mu^2$$

$$r = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}}$$

$$r_{xx} = \frac{2r_{hh}}{1 + r_{hh}}$$

$$\alpha = \frac{K}{K - 1}\left(1 - \frac{\sum_{i=1}^{K} \sigma_{Yi}^2}{\sigma_X^2}\right)$$

$$\alpha_{standardized} = \frac{K\bar{r}}{1 + (K - 1)\bar{r}}$$

Average Inter-Item Correlation

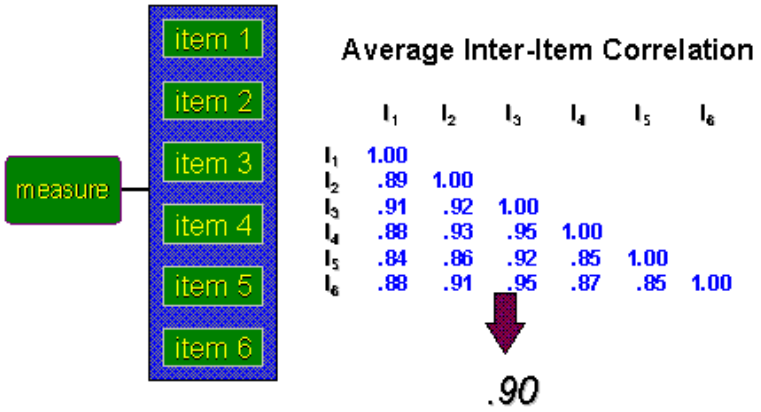
1. Compute the correlation between each pair of items

$$r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2][N\Sigma y^2 - (\Sigma y)^2]}}$$

Where:

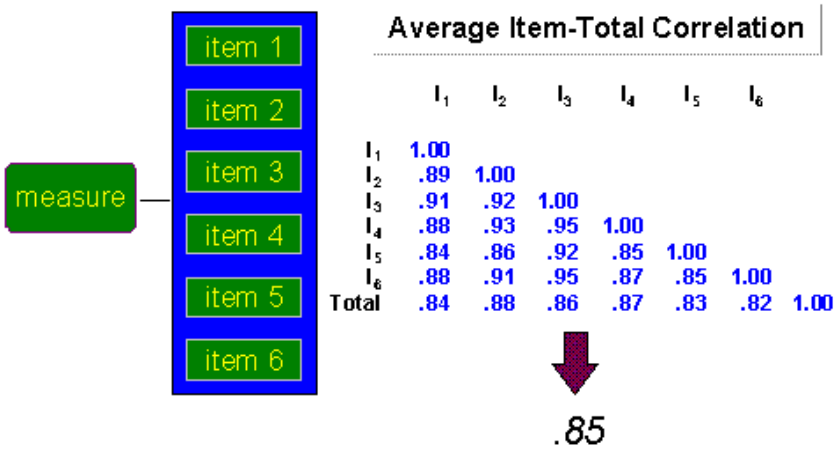
N	=	number of pairs of scores
Σxy	=	sum of the products of paired scores
Σx	=	sum of x scores
Σy	=	sum of y scores
Σx^2	=	sum of squared x scores
Σy^2	=	sum of squared y scores

2. Computer the average of all the correlations



Average Item-to-Total Correlaton

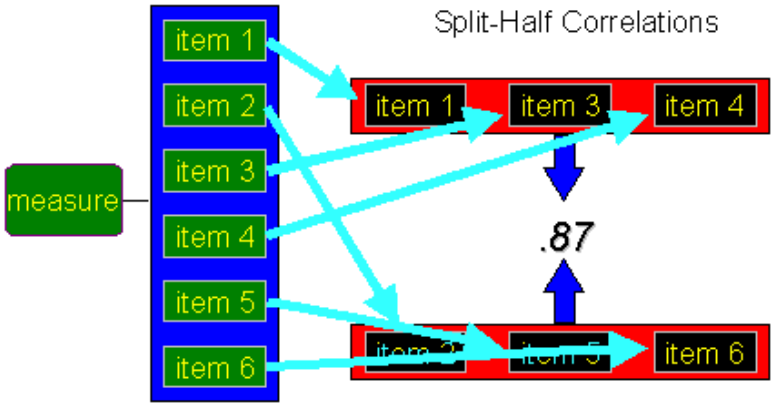
1. Compute the total score for the items
2. Compute the correlation between each item and the total



3. Compute the average of the item-to-total correlations

Split-Half Reliability – Spearman-Brown Formula

1. Randomly divide all items into two halves
2. Compute the total score for each half
3. Compute the correlation (r_{hh}) between the two halves



4. Step up the split-half correlation using the Spearman-Brown formula:

$$r_{xx} = \frac{2r_{hh}}{1 + r_{hh}}$$

Split-Half Reliability – Spearman-Brown Formula

Example

- A construct (e.g., Satisfaction) is measured using 10 items (survey questions) from 15 respondents
- Calculate the split-half reliability using the Spearman-Brown Formula
  1. Randomly divide the 10 items into two halves (e.g., Q1-Q5 and Q6-Q10)
  2. Compute the total score for each half (H1 and H2)
  3. Compute the correlation (r_{hh}) between the two halves (0.47581)
  4. Step up the split-half correlation using the SB formula:

$$r_{xx} = \frac{2(0.475481)}{1+0.475481} = 0.645$$

	A	B	C	D	E	F	G	H	I	J	K
3		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
4	1	3	2	4	1	4	5	1	4	3	2
5	2	5	4	6	3	5	1	2	5	5	5
6	3	7	3	5	5	6	4	3	3	6	5
7	4	2	4	3	3	2	2	2	6	6	3
8	5	6	5	6	5	4	4	4	6	7	5
9	6	1	2	3	4	2	5	2	3	5	2
10	7	4	3	5	3	5	2	1	2	4	3
11	8	2	4	4	3	7	2	2	4	6	5
12	9	6	5	5	5	1	6	5	2	3	2
13	10	5	4	6	4	3	5	3	1	7	2
14	11	4	3	5	2	5	3	2	3	4	1
15	12	4	2	3	2	5	3	4	2	3	4
16	13	5	5	6	2	3	3	2	4	4	5
17	14	6	5	5	4	2	2	1	4	6	5
18	15	4	3	5	5	3	1	3	2	3	4

	A	B	C	D	E	F	G	H
22		H1	H2					
23	1	14	15		Correlation coefficient			0.475481
24	2	23	18		Spearman-Brown correction			0.64451
25	3	26	21					
26	4	14	19					
27	5	26	26					
28	6	12	17					
29	7	20	12					
30	8	20	19					
31	9	22	18					
32	10	22	18					
33	11	19	13					
34	12	16	16					
35	13	21	18					
36	14	22	18					
37	15	20	13					

Cronbach's Alpha (Coefficient Alpha) (α)

- Alpha (Unstandardized)

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{i=1}^{K}\sigma_{Yi}^2}{\sigma_X^2}\right)$$

K: number of items

σ_{Yi}^2: variance of item i

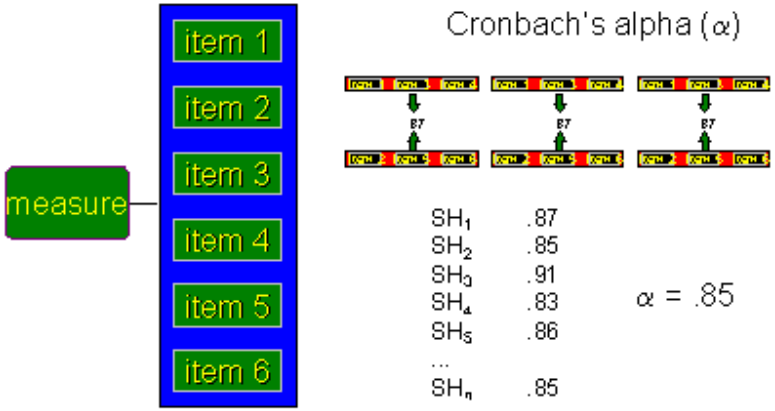
σ_X^2: variance of total score

- Standardized Alpha
  - When item variances are all equal
  - Mathematically equivalent to the average of all possible split-half correlations

$$\alpha_{standardized} = \frac{K\bar{r}}{1 + (K-1)\bar{r}}$$

$\bar{r}$: average correlation among the k items

Cronbach's alpha	Internal consistency
α ≥ 0.9	Excellent
0.9 > α ≥ 0.8	Good
0.8 > α ≥ 0.7	Acceptable
0.7 > α ≥ 0.6	Questionable
0.6 > α ≥ 0.5	Poor
0.5 > α	Unacceptable



Cronbach's Alpha (Coefficient Alpha) (α)

Example

- A construct (e.g., Satisfaction) is measured using 10 items (survey questions) from 15 respondents
- Calculate the Cronbach's alpha

$$\alpha = \frac{10}{10-1}\left(1 - \frac{2.72889 + 1.17333 + \cdots + 1.9822}{40.6933}\right)$$
$$= \frac{10}{10-1}\left(1 - \frac{19.0222}{40.6933}\right)$$
$$= 0.592$$

- Comment on the reliability of the scale given the Cronbach's alpha
  - Given the alpha value of 0.592, which is between 0.5 and 0.6, the reliability of the scale is considered poor
- Suppose the average correlation among the 10 items is 0.14783. Calculate the standardized alpha

$$\alpha_{standardized} = \frac{10(0.14783)}{1+(10-1)(0.14783)} = 0.634$$

	A	B	C	D	E	F	G	H	I	J	K	L
1	Cronbach's Alpha											
2												
3		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total
4	1	3	2	4	1	4	5	1	4	3	2	29
5	2	5	4	6	3	5	1	2	5	5	5	41
6	3	7	3	5	5	6	4	3	3	6	5	47
7	4	2	4	3	3	2	2	2	6	6	3	33
8	5	6	5	6	5	4	4	4	6	7	5	52
9	6	1	2	3	4	2	5	2	3	5	2	29
10	7	4	3	5	3	5	2	1	2	4	3	32
11	8	2	4	4	3	7	2	2	4	6	5	39
12	9	6	5	5	5	1	6	5	2	3	2	40
13	10	5	4	6	4	3	5	3	1	7	2	40
14	11	4	3	5	2	5	3	2	3	4	1	32
15	12	4	2	3	2	5	3	4	2	3	4	32
16	13	5	5	6	2	3	3	2	4	4	5	39
17	14	6	5	5	4	2	2	1	4	6	5	40
18	15	4	3	5	5	3	1	3	2	3	4	33
19	Total	64	54	71	51	57	48	37	51	72	53	558
20	Var	2.72889	1.17333	1.12889	1.57333	2.69333	2.29333	1.31556	2.10667	2.02667	1.98222	19.0222
21												
22	k	10										
23	Σvar	19.0222										
24	var	40.6933										
25	α	0.59172										

Class Activities on Reliability

Download Week 10 Lecture Reliability.xlsx from Canvas

Try “TamData Sample” yourself.

In Excel…

Response	Q1	Q2	Q3	Q4	Q1+Q2	Q3+Q4	Q1toQ4
1	1	4	2	2	5	4	9
2	2	4	5	4	6	9	15
3	3	5	4	4	8	8	16
4	4	5	4	3	9	7	16
5	5	3	5	5	8	10	18
Variance	2.5	0.7	1.5	1.3	2.7	5.3	11.7
cov	4.75	-0.2	3.75	3.4	4.55	7.15	11.7
cor	0.8783	-0.0699	0.8951	0.8718			
	Q1	*Q2*	*Q3*	*Q4*	*Q1+Q2*	*Q3+Q4*	*Q1toQ4*
Q1	1						
Q2	- 0.1890	1					
Q3	0.6455	-0.2440	1				
Q4	0.6934	-0.4193	0.8951	1			
Q1&2	0.8660	0.3273	0.4969	0.4537	1		
Q3&4	0.6868	-0.3375	0.9753	0.9715	0.4890	1	
Q1to4	0.8783	-0.0699	0.8951	0.8718	0.8095	0.9080	1
Average inter-item correlation:	0.230						
Average item-to-total correlation:	0.644						
Split-Half Correlation (Q1+Q2) vs (Q3+Q4):	0.489						
Split-Half Reliability (Q1+Q2) vs (Q3+Q4):	0.657						
Alpha (Unstandardised) (Q1toQ4):	0.650						
Standardised Cronbach's alpha (Q1toQ4):	0.545	Poor					

SPSS – under Analyze > Scale > Reliability Analysis

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.650	.545	4

Inter-Item Correlation Matrix

	Q1	Q2	Q3	Q4
Q1	1.000	-.189	.645	.693
Q2	-.189	1.000	-.244	-.419
Q3	.645	-.244	1.000	.895
Q4	.693	-.419	.895	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	3.700	3.000	4.200	1.200	1.400	.280	4
Item Variances	1.500	.700	2.500	1.800	3.571	.560	4
Inter-Item Covariances	.475	-.400	1.250	1.650	-3.125	.658	4
Inter-Item Correlations	.230	-.419	.895	1.314	-2.135	.300	4