Survey Design

By: Kellen Whetstone, Joanne Nie, Leslie Medina, Jim Liu, Roger Gao, Owen Sun

Survey Design Objective

Survey design is used to properly collect information within a survey.

It is typically used in the following scenarios:

1) Collecting customer feedback and satisfaction
2) Collecting employee feedback and satisfaction
3) Research studies

Survey design can be used in a survey to ensure it is valid and reliable

It ensures that the survey correctly measures what it is intended to measure or collect.

Key words: valid and reliable

Mathematical basis: Reliability

About Cronbach’s alpha:

- It is one of the most popular measures of internal consistency (reliability)
- It measures the degree of association among a group of items
  - Their correlation
- The Cronbach's alpha reliability coefficient tends to range between 0 and 1
  - However, there is no real lower limit
  - The closer Cronbach’s alpha is to 1, the larger the internal consistency of the items
- When is Cronbach’s alpha typically used ?
  - It is commonly used in surveys that contain Likert questions.

Note: Cronbach’s alpha is not a measure of validity

Mathematical basis: Reliability

Cronbach's Alpha:

- Among items: compares the amount of shared variances, or covariances and gives the overall variance
- If reliable:
  - "There should be a great deal of covariance among the items relative to the variance"
- Deletion of items
  - If Cronbach's alpha increases after an item is deleted from a group, then the item does not belong in the survey

- Alpha value closer to 1: items are highly correlated (higher precision/consistency)
- Alpha value closer to 0: weak correlation, lower internal consistency

Formula

$$\alpha = \frac{N * \bar{c}}{\bar{v} + (N - 1) * \bar{c}}$$

- N = number of items
- $\bar{c}$ = mean covariance between items.
- $\bar{v}$ = mean item variance.

Evaluating Cronbach's Alpha

- 00 to . 69 = Poor.
- 70 to . 79 = Fair.
- 80 to . 89 = Good.
- 90 to . 99 = Excellent/Strong.

Mathematical basis: Reliability

Split-Half Reliability:

- Split the test into 2 halves and calculate the score of each half
  - Correlate score of both halves
    - Similar to Cronbach's alpha: if correlation is high, indicates reliability

Using Pearson Correlation:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

x_i = values of the x-variable in a sample

$\bar{x}$ = mean of the values of the x-variable

y_i = values of the y-variable in a sample

$\bar{y}$ = mean of the values of the y-variable

- Note: Cronbach's alpha is equivalent to "taking the average of all possible split-half reliabilities"

Different ways to split the data into 2 halves :

1) First half vs last half
2) Odd-numbered items vs even-numbered items
3) Random Split
4) Parallel items

How to design a good survey

1. Define your objectives: Clearly identify the purpose of your survey and the specific information you want to gather. This will guide the design process and help you create focused and relevant questions.
2. Keep it concise: Aim for a reasonable survey length to maintain respondent engagement. Long surveys can lead to fatigue and lower response rates. Be mindful of the time it takes to complete the survey and prioritize essential questions.
3. Use clear and precise language: Write questions that are easy to understand and unambiguous. Avoid jargon or technical terms that respondents may not be familiar with. Keep sentences short and use simple language whenever possible.

How to design a good survey (Continued)

4. Avoid leading or biased questions: Ensure your questions are neutral and unbiased. Avoid wording that may influence or sway respondents' answers. For example, instead of asking, "Don't you agree that...?" use a more neutral phrasing such as, "What is your opinion on...?"

5. Use appropriate response scales: Choose response scales that are suitable for the type of data you wish to collect. Common options include Likert scales (e.g., strongly agree, agree, neutral, disagree, strongly disagree) and numerical scales (e.g., on a scale of 1 to 10).

6. Ensure construct validity: Construct validity refers to the degree to which a survey measures what it intends to measure. To enhance construct validity, consider the following:

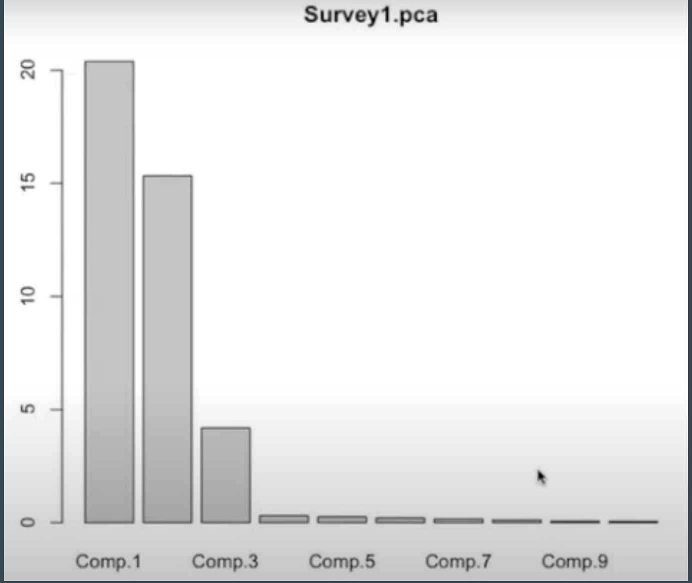
   a. Clearly define the constructs or variables you are measuring.
   b. Use established scales or questions that have been previously validated, if available.
   c. Seek expert feedback to ensure the survey accurately captures the desired constructs.
   d. Conduct statistical analysis to assess the relationship between different survey items and the overall construct.

How to use EDA in Survey Design factor analysis

```
> head(Survey1)
  Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10
1  3  2  3  2  5  2  2  5  2   6
2  7  7  7  7  0  5  4  3  5   4
3  4  4  4  3  4  4  4  4  4   3
4  5  5  5  5  2  5  5  2  6   2
5  7  6  6  6  1  3  2  5  3   2
6  2  3  2  2  5  0  1  7  1   5
```

pca<-princomp((Survey1))

plot(pca)



```r
Survey1.fa1 <- factanal(Survey1, factors=3, rotation="varimax")
Survey1.fa1
```

```
Call:
factanal(x = Survey1, factors = 3, rotation = "varimax")

Uniquenesses:
   Q1    Q2    Q3    Q4    Q5    Q6    Q7    Q8    Q9   Q10
0.053 0.046 0.051 0.051 0.006 0.006 0.045 0.011 0.046 0.980
```

```
Loadings:
    Factor1 Factor2 Factor3
Q1   0.971
Q2   0.973
Q3   0.972
Q4   0.972
Q5  -0.996
Q6            0.995
Q7            0.948   0.228
Q8           -0.975  -0.184
Q9            0.958   0.189
Q10                   0.131

               Factor1 Factor2 Factor3
SS loadings      4.785   3.771   0.149
Proportion Var   0.479   0.377   0.015
Cumulative Var   0.479   0.856   0.871

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 14.14 on 18 degrees of freedom.
The p-value is 0.72
```

Cronbach's Alpha

```r
library(psych)

df <- read.csv("Cronbachs_alpha.csv")
a <- alpha(df)
a$total[1]
```

```
##   raw_alpha
##   0.7853028
```

Method: Analyze the effect of Cronbach's alpha when dropping elements of the survey. If an omitted question substantially increase Cronbach's alpha, we should remove it.

Here, our baseline alpha = 0.79, which increases to 0.92 when omitting Item 4.

```r
kable(a$alpha.drop)
```

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha se	var.r	med.r
Item.1	0.5994987	0.5231156	0.6194132	0.2677469	1.096944	0.0748448	0.2039250	0.0168981
Item.2	0.6062792	0.5344361	0.6287445	0.2767482	1.147933	0.0752110	0.2014913	0.0384436
Item.3	0.6259956	0.5505051	0.6474650	0.2898937	1.224719	0.0692574	0.2063986	0.0384436
Item.4	0.9216739	0.9227700	0.8888359	0.7993087	11.948332	0.0172939	0.0001769	0.7945225

Cronbach's Alpha (Continued)

```r
df2 <- df[,1:3]
b <- alpha(df2)
b$total[1]
```

```
##  raw_alpha
##  0.9216739
```

```r
kable(b$alpha.drop)
```

	raw_alpha
Item.1	0.8785626
Item.2	0.8846082
Item.3	0.8967848

Now we delete item 4 and rerun the analysis to verify that no more items should be dropped.

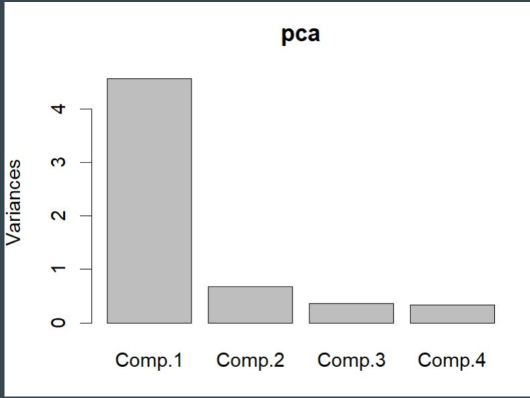
Alpha	Guideline
0-.69	Poor
.70-.79	Fair
.80-.89	Good
.90-.99	Excellent

PCA and Parallel Analysis

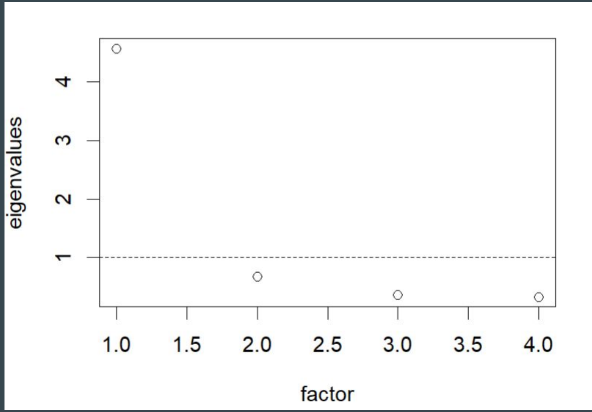
```
> pca <- princomp(df)
> pca
Call:
princomp(x = df)

Standard deviations:
   Comp.1    Comp.2    Comp.3    Comp.4 
2.1376151 0.8213641 0.6048979 0.5737841 

 4  variables and  60 observations.
```



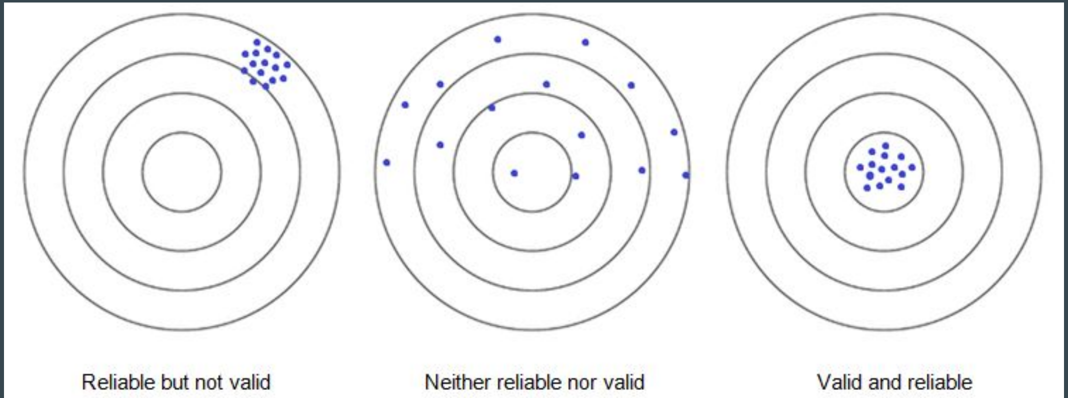
- Parallel analysis: there is only one point with eigenvalue greater than one, so we go with only one factor
- Just the first factor will gives us 76.9% of variance explained



How to assess validity/reliability?

Ways to measure validity:

- Content validity → Assesses whether a test represents all aspects of the construct
- Construct validity → Refers to if the scale measures the construct adequately
- Face validity → Does the test appear to measure what it is supposed to measure?
- Criterion-related validity → Measure of quality of measurement methods



Measures of Reliability

- Cronbach's alpha (already discussed): Measure of internal consistency or how closely related a set of items are as a group
- Split-half reliability: Internal consistency approach to quantify the reliability of a test
- Test-retest reliability: Conducts test given twice over a given period → time interval between two administrations should be long enough
- Parallel-forms reliability: Two similar surveys made to assess same construct. Tests administered to same group of individuals and coefficient of correlation between scores is used to measure reliability
- Internal consistency reliability: Measures consistency of individual items of a test
- Inter-rater reliability: Measures feedback of someone assessing the test given. Test is reliable if scores are consistent between multiple people scoring test

$$\alpha = \frac{N * \bar{c}}{\bar{v} + (N-1) * \bar{c}}$$

$$r_{spearman\ Brown} = \frac{2\, r_{XY}}{(1 + r_{XY})}$$

	Item.1	Item.2	Item.3	Item.4
Item.1	1.00000000	0.81433945	0.794522453	0.038443572
Item.2	0.81433945	1.00000000	0.789064222	0.016898051
Item.3	0.79452245	0.78906422	1.000000000	-0.002721536
Item.4	0.03844357	0.01689805	-0.002721536	1.000000000

Why measure reliability/limitations

- One of the easiest ways to quantify the quality of the test
- May be an oversimplification, but the convenience makes reliability indices very popular
- Many tests require conditions to be met to obtain relevant results:
  - Ex: Test-retest reliability requires the time interval between administration of tests to be far large enough that participants do not recall the information

- Big picture: Reliability and validity tests do not tell full story → often conducted on a specific sample or time period
- Also limitations with statistical tests:
  - Ex: Cronbach's alpha assumes unidimensionality and homogeneity of construct