Photo credit: Dave DiCello

17-803 Empirical Methods

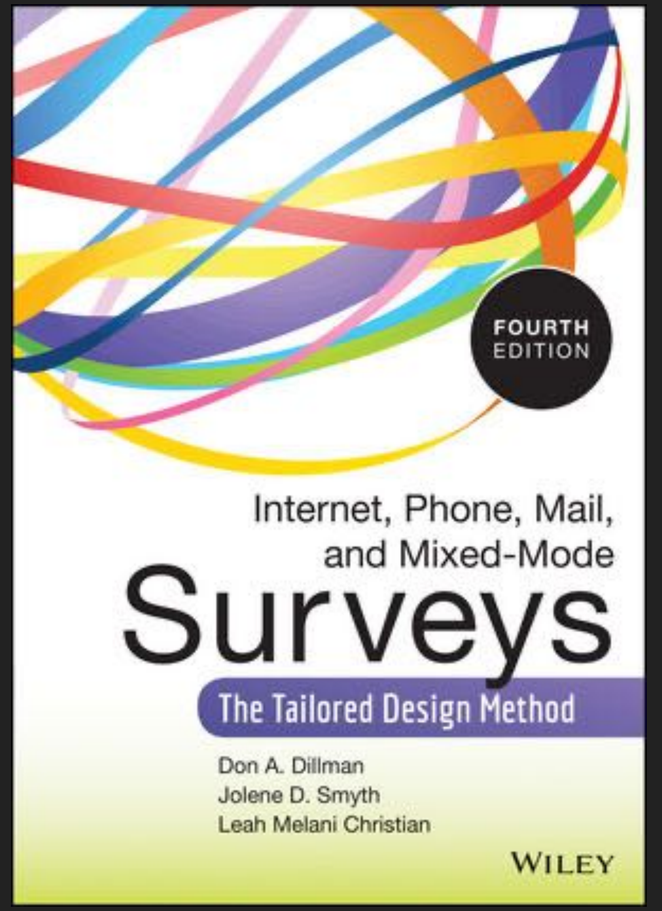
Bogdan Vasilescu, S3D

Survey Design (Part 1)

Thursday, February 22, 2024

Outline for Today

- Chapters from Dillman, D., Smyth, J. D., & Christian, L. M. (2014). Internet, Phone, Mail and Mixed-Mode Surveys: The Tailored Design Method (4th ed.). Hoboken, NJ: Wiley.
  - Ch1: Sample Surveys in our Electronic World
  - Ch2: Reducing People's Reluctance to Respond to Surveys
  - Ch4: The Fundamentals of Writing Questions
  - Ch5: How to Write Open and Closed Ended Questions
- Hof, M. (2012). Questionnaire Evaluation with Factor Analysis and Cronbach's Alpha. Student project. Seminar in Methodology and Statistics. Uni. Groningen
- Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. Tutorials in quantitative methods for psychology, 9(2), 79-94.
- Cairns, P. (2019). Doing better statistics in human-computer interaction. Cambridge University Press.
  - Ch15: What Makes a Good Likert Item?
  - Ch16: The Meaning of Factors
  - Ch17: Unreliable Reliability: The Problem of Cronbach's Alpha
  - Ch18: Tests for Questionnaires



What Is a Survey Good for?

- Prevalence (sort of)
- Relations between variables
- Differences among subpopulations

Probability sample survey strength:

Collect data from only a sample of the population but generalize results to the whole

Main goal when designing probability sample surveys:

Reduce survey error (the difference between an estimate that is produced using survey data and the true value of the variables in the population that one hopes to describe)

Four Types of Error To Minimize To Improve the Survey Estimates

- Sampling frame – how to construct?
  - Coverage error
- Draw sample – probability sampling?
  - Sampling error
- Administer survey – who responded?
  - Nonresponse error
- Questions as measures – valid and reliable?
  - Measurement error

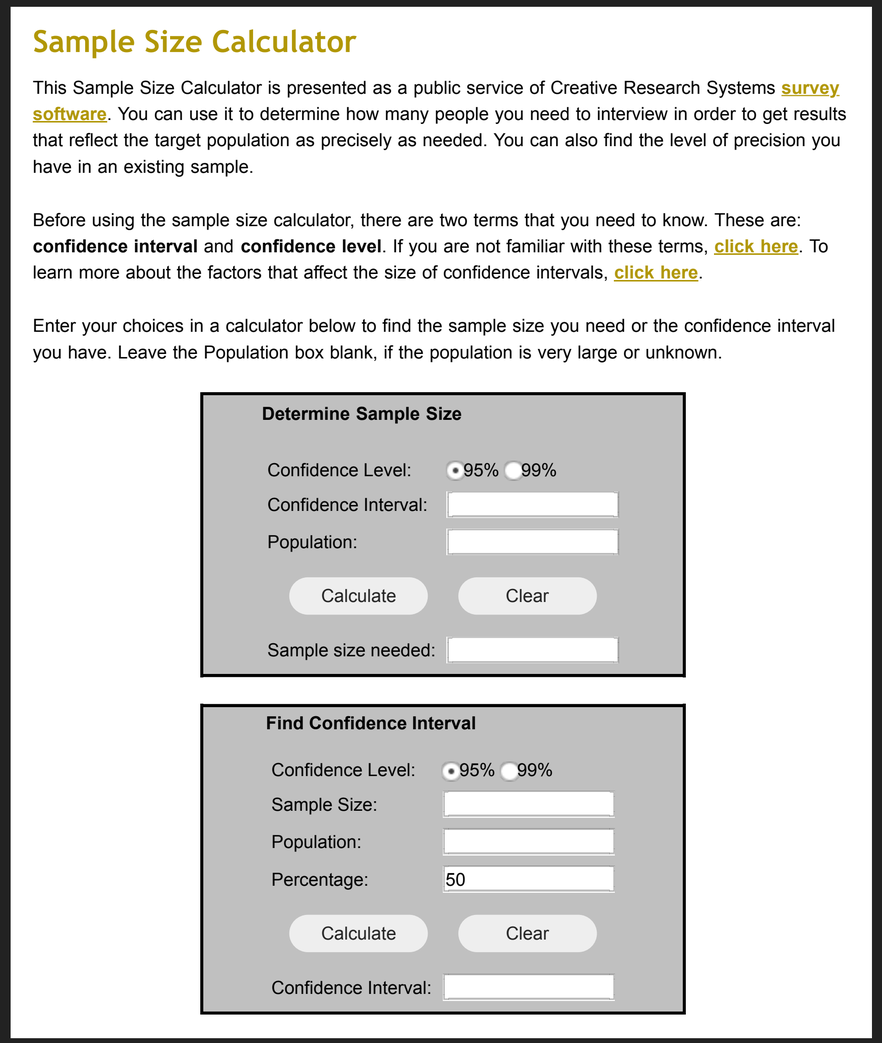
Coverage Error Example

- Landline random digit dial telephone survey
  - People who have landlines are quite different from those who do not on a number of important characteristics (e.g., higher socioeconomic status).
- Context matters
  - Internet survey among US population vs CMU students

"Butler c. 1922" CC-BY-SA-4.0 Wikipedia

Sampling Error

- Surprising how few people you would need to survey to obtain estimates with acceptable levels of precision!
- CMU: 14k students
  - 95% confidence +/- 2% margin of error:
    - Random sample of 2050 students
- US population: 328 million
  - 95% confidence +/- 2% margin of error:
    - Random sample of 2384 people
  - 95% confidence +/- 10% margin of error:
    - Random sample of 96 people



Sample Size Calculator

This Sample Size Calculator is presented as a public service of Creative Research Systems survey software. You can use it to determine how many people you need to interview in order to get results that reflect the target population as precisely as needed. You can also find the level of precision you have in an existing sample.

Before using the sample size calculator, there are two terms that you need to know. These are: **confidence interval** and **confidence level**. If you are not familiar with these terms, click here. To learn more about the factors that affect the size of confidence intervals, click here.

Enter your choices in a calculator below to find the sample size you need or the confidence interval you have. Leave the Population box blank, if the population is very large or unknown.

Determine Sample Size

Confidence Level: 95% 99%
Confidence Interval:
Population:
Calculate Clear
Sample size needed:

Find Confidence Interval

Confidence Level: 95% 99%
Sample Size:
Population:
Percentage: 50
Calculate Clear
Confidence Interval:

https://www.surveysystem.com/sscalc.htm

Aside: Confidence Level, Confidence Interval

- Confidence interval (aka “margin of error”):
  - Plus-or-minus figure usually reported
  - Lower is better
  - Example:
    - if confidence interval is 4 and 47% percent of sample answer X
    - then you can be "sure" that among the entire relevant population between 43% (47-4) and 51% (47+4) would have answered X.
- Confidence level (how sure can you be):
  - Represents how often the true percentage of the population who would answer X lies within the confidence interval.
  - Higher is better
  - Common: 95% or 99%

Nonresponse Error

- Do the characteristics of respondents differ from those who chose not to respond in a way that is relevant to the study results?
- Higher response rates reduce the likelihood of nonresponse error.
  - But: nonresponse error may occur in surveys with higher as well as lower response rates (Groves & Peytcheva, 2008)

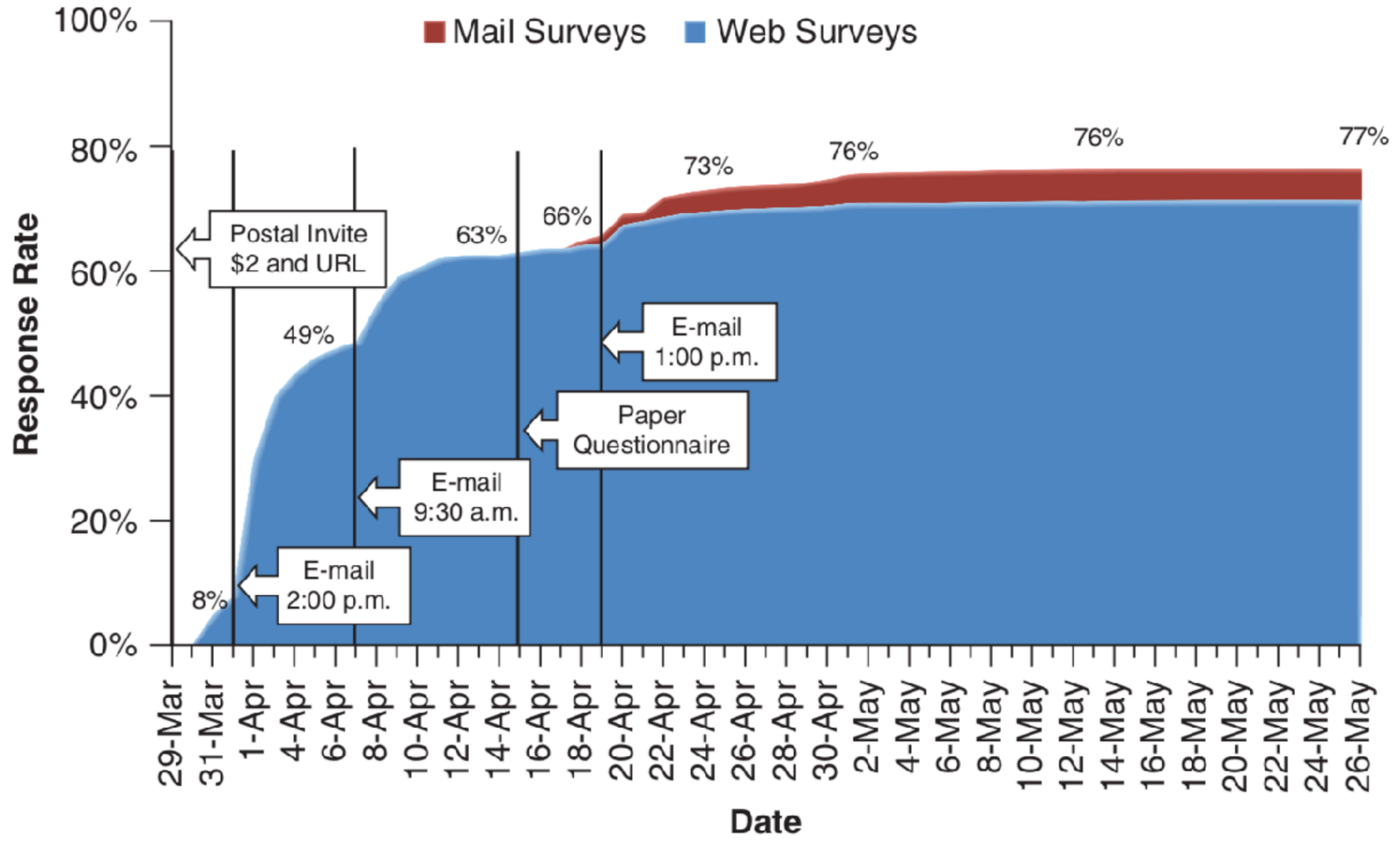
Measurement Error

- Asking good questions:
  - Does question adequately measure the idea or concept of interest?
  - “Construct validity”
- Example:
  - household income for the previous year as a measure of household wealth.
  - annual income decreases after retirement, but wealth may not.

Getting people to respond:

Example survey of 600 PhD students at Washington State U.
about their dissertation work and graduate training

FIGURE 2.2 Cumulative response rate by day and mode for the 2013 WSU Doctoral Student Experience Survey, showing contribution of each contact to final response rate.



Source: Adapted from *Determining Whether Research Is Interdisciplinary: An Analysis of New Indicators* (Technical Report 13-049), by M. M. Millar, 2013, Pullman: Washington State University, Social and Economic Sciences Research Center.

Initial commu-nication

Jane Doe
123 Cougar Road
Pullman, WA 99163

Dear Jane,

I am writing to ask for your help with an important survey we are conducting of WSU doctoral students. I understand that you have successfully completed your preliminary examinations and are now at the stage of needing to complete a dissertation.

My colleague, Morgan Millar, and I have been working with the National Science Foundation to better understand how the needs of doctoral training in the U.S. are changing. We are hoping you could spend a few minutes sharing some of your experiences in your doctoral program. In particular, we are interested in factors that may affect bringing the dissertation writing process to a successful conclusion.

To this end, we would greatly appreciate if you would answer a few questions for us. To do so, simply go to this website: www.opinion.wsu.edu/phdexperience

In order to begin the survey, you will need to enter this access code: «RESPID».

We think it should only take about ten minutes to complete the questionnaire.

The survey is confidential. Your individual answers will not be linked with your name or department in any reports of the data. Your participation is voluntary and if you come to any question you prefer not to answer, you are welcomed to skip it and go on to the next. Should you have any questions or comments, please contact me (dillman@wsu.edu) or Thom Allen, the study director (509) 335-1722 or ted@wsu.edu.

We very much appreciate your help with this study, and a small token of appreciation is enclosed with this letter as way of saying thank you.

Many Thanks,

Don A. Dillman
Regents Professor and Deputy Director

Reminder

Dear Jane,

Earlier this week we sent you a letter asking for your help with an important survey. We are conducting this study of WSU doctoral students to learn more about the processes they go through to complete their dissertations and finish their degrees.

I am following up with this e-mail to provide you with an electronic link to the survey website. I hope this link makes it easier for you to respond. It should only take a few minutes to complete the questionnaire.

Simply click on this link and you will automatically be logged into the survey:

http://www.opinion.wsu.edu/phdexperience

And enter your personal Access Code in the space provided: **<<RESPID>>**

The results of this study will help us better understand the needs and experiences of students as they work on their dissertation research. Your participation is very important, and we appreciate you considering our request.

Sincerely,

Don A. Dillman
Regents' Professor and Deputy Director
Social and Economic Sciences Research Center

The Social Reward Value of Modest Incentives With the Request (Avdeyeva & Matland, 2013)

- Control group – no incentive
- A second group – 50 rubles (~$1.65) with survey request
- A third group – promised 300 rubles if questionnaire was returned
- Response rates:
  - no incentive – 10%
  - 50 rubles pre pay – 37%
  - 300 rubles post pay – 24%
  - Combined pre and post pay – 48%

Using social exchange concepts to motivate potential respondents:

People are more likely to comply with a request from someone else if they believe and trust that the rewards for complying with that request will eventually exceed the costs of complying

Increasing the Benefits of Survey Participation

- Make it seem valuable, important
  - How will results be useful?
  - Ask for help/advice
  - Make it interesting
  - Gamification
  - Scarcity – only some people get asked
  - Others have responded
  - Pay (a little bit) forward

Decreasing the Costs of Participation

- Make it easy
  - Keep it short!
  - Keep it simple and clear and convenient
  - Giving people a choice of modes may reduce response rates
- Be likable
  - Show respect
  - Don’t use subordinate language

Establishing Trust

- Worry about malware, fake surveys
- Ways to assess your trustworthiness (contact info)
- Sponsorship
- Assure confidentiality
- Minimize requests to obtain personal information

FIGURE 2.3 Summary of ways to increase benefits, reduce costs, and establish trust.

Survey Examples

- Ecosystem survey

http://cmu.ca1.qualtrics.com/jfe/form/SV_d4M66VwPlZYd5kh

  - Results: http://breakingapis.org/survey/

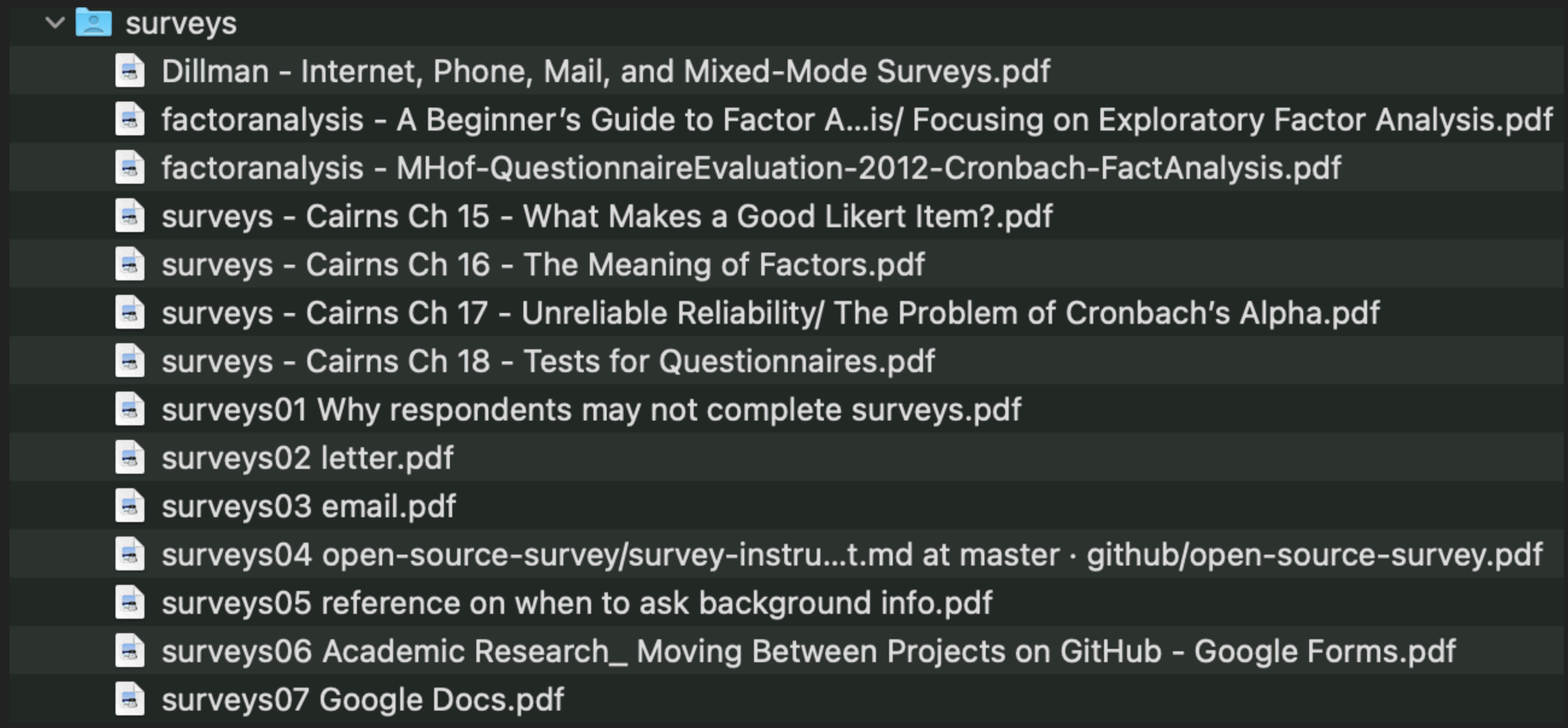
- GitHub open source survey

https://github.com/github/open-source-survey

  - Results: http://opensourcesurvey.org/2017/

Readings

- https://drive.google.com/drive/folders/1ISOQlbw-cRmT47_itpIJkTZua_IASNgB?usp=sharing



Credits

- Graphics:
  - Dave DiCello photography (cover)
- Content:
  - Chapters from Dillman, D., Smyth, J. D., & Christian, L. M. (2014). Internet, Phone, Mail and Mixed-Mode Surveys: The Tailored Design Method (4th ed.). Hoboken, NJ: Wiley.
    - Ch1: Sample Surveys in our Electronic World
    - Ch2: Reducing People's Reluctance to Respond to Surveys
    - Ch4: The Fundamentals of Writing Questions
    - Ch5: How to Write Open and Closed Ended Questions
  - Hof, M. (2012). Questionnaire Evaluation with Factor Analysis and Cronbach's Alpha. Student project. Seminar in Methodology and Statistics. University of Groningen
  - Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. Tutorials in quantitative methods for psychology, 9(2), 79-94.
  - Cairns, P. (2019). Doing better statistics in human-computer interaction. Cambridge University Press.
    - Ch15: What Makes a Good Likert Item?
    - Ch16: The Meaning of Factors
    - Ch17: Unreliable Reliability: The Problem of Cronbach's Alpha
    - Ch18: Tests for Questionnaires