Best Practices in Quantitative Methods

Resources guide

Jason W. Osborne, Ph.D.

September 01, 2016

In this document I have collected as many resources as I could. Some of the promised resources disappeared from the web before I could do this, and some authors are not responding to requests any more. My apologies for this.

Chapter 3: Inter-rater reliability (data not available currently; am attempting to track down author)

Chapter 4: link to MINISTEPS installer

Chapter 5: link to download WINSTEPS trial software MINISTEPS (WINSTEPS.com does not seem to exist any more…. Apologies). Here is another option: http://assess.com/xcart/home.php

- **This is a dissertation (included in this website with permission of the author) that used a Rasch Partial Credit model: Fidelity of implementation to instructional strategies as a moderator of curriculum unit effectiveness in a large-scale middle school science quasi-experiment** by *O'Donnell, Carol Lynn*, Ed.D., The George Washington University, 2007, 219 pages; AAT 3276564
- Here is a link to Dr. O'Connell's WINSTEPS control file if you want to replicate some of her analyses... Thanks to her for sharing!!!

Chapter 6: Since this chapter was written I released two books on the topic.

- One focuses on EFA using SPSS with its own resources and exercises. You can find them HERE.
- One focuses on EFA using SAS with its own resources and exercises. They will be on my web site soon.

Chapter 7: To make calculations of p(rep) simpler, Dr. Killeen has provided this Excel spreadsheet.

Chapter 8: no resources provided by authors

Chapter 9: no resources provided by authors

Chapter 10: no resources provided by authors

Chapter 11: Data sets referred to in the chapter:

1. The Lalonde/Dehejia and Wahba dataset used to illustrate methods

2. The "MatchIt" R library used to do the matching in the paper: http://gking.harvard.edu/matchit/. There are also some other references for software packages at the end of the chapter, and you could include their weblinks as well (they should be in the text or in the reference list). (If that link is broken, here is a hosted copy)

Links provided by the authors: (from Dr. Stuart's page on the topic)

- Propensity score matching in SPSS http://arxiv.org/ftp/arxiv/papers/1201/1201.6385.pdf
  - o Thoemmes, F. (2012). Propensity score matching in SPSS. http://sourceforge.net/projects/psmspss/files/
  - o Nearest neighbor propensity score matching with various options (with/without replacement, calipers, k to 1, etc.)
  - o Detailed balance statistics and graphs
  - o Actually calls MatchIt using a point and click interface
- Matching software for R (http://www.r-project.org)
  - o MatchIt: http://gking.harvard.edu/matchit
  - o Matching: http://sekhon.berkeley.edu/matching
  - o twang: https://cran.r-project.org/web/packages/twang/twang.pdf
- Matching software for Stata:
  - o PSMATCH2: http://repec.org/bocode/p/psmatch2.html
  - o discussion on STATA blog: http://blog.stata.com/2015/08/24/introduction-to-treatment-effects-in-stata-part-2/
  - o another discussion: https://kb.iu.edu/d/avll
- Matching software for SAS:
  - o 1:N propensity score matching
  - o SAS usage note: http://support.sas.com/kb/30/971.html
  - o Local and global optimal propensity score matching

    - Coca-Perraillon, M. (2007). Local and global optimal propensity score matching. In SAS Global Forum 2007. Paper 185-2007. Available here.
    - Variety of matching methods. No built in diagnostics. Assumes propensity score already estimated.
  - o Greedy matching (1:1 nearest neighbor)
    - Parsons, L. S. (2001). Reducing bias in a propensity score matched-pair sample using greedy matching techniques. In SAS SUGI 26, Paper 214-26. Available here.
    - Parsons, L.S. (2005). Using SAS software to perform a case-control match on propensity score in an observational study. In SAS SUGI 30, Paper 225-25. Available here.
    - Kosanke, J., and Bergstralh, E. (2004). gmatch: Match 1 or more controls to cases using the GREEDY algorithm. http://www.mayo.edu/research/departments-divisions/department-health-sciences-research/division-biomedical-statistics-informatics/software/locally-written-sas-macros
  - o 1:1 Mahalanbois matching within propensity score calipers
    - Feng, W.W., Jun, Y., and Xu, R. (2005). A method/macro based on propensity score and Mahalanobis distance to reduce bias in treatment comparison in observational study. www.lexjansen.com/pharmasug/2006/publichealthresearch/pr05.pdf
  - o Weighting
    - Leslie, S. and Thiebaud, P. (2006). Using propensity scores to adjust for treatment selection bias. http://www.lexjansen.com/wuss/2006/Analytics/ANL-Leslie.pdf
  - o Variable ratio matching, optimal matching algorithm
    - Kosanke, J., and Bergstralh, E. (2004). Match cases to controls using variable optimal matching. http://www.mayo.edu/research/departments-

Chapter 12: no resources provided by authors

Chapter 13: no resources provided by authors

Chapter 14: since release of this book, I wrote an entire book on data cleaning and updated some information on outliers. You can find links to many exercises and data sets on my web page. Below are several data files in SPSS format:

ANOVA
Sample size=52
- outliers in one cell
- outliers in both cells

Samples size=416
- outlliers in one cell
- outliers in both cells

You can replicate some of the analyses presented in the chapter. In each data file you have two cells (mean1, SD1, mean2 SD2) with a *t* value for the difference between the two. Means 1 and 2 and SD 1 and 2 are means and standard deviations for the cells before outliers were removed. Means 3/4 and SD 3/4 and *t* 2 are means, standard deviations, and t scores after outliers are removed from the data. We provide data for small samples (N=52) and large samples (N=416) which have outliers in the same proportion. Analyze the differences before/after outliers are removed, and compare what happens when you have outliers in one or both cells, and you will immediately see that accuracy gets better after outliers are removed.

Chapter 15: no resources provided by authors

Chapter 16: no resources provided by authors

Chapter 17: Note that since the publication of this book, I have addressed application of bootstrap resampling in several places as a methodology for easily estimating confidence intervals for effect sizes. A recent paper by Erin Banjanovic and myself covers the topic succinctly.

Regarding data from the chapter: The data presented in Table 1 and below electronically represent a random sample of real data provided by roughly 500,000 library users at over 700 libraries from around the world with respect to the perceived quality of academic library services (cf. Thompson, Cook & Kyrillidou, 2005, 2006).
Included in the data are scores on the LibQUAL+® total scale, and three subscales:
- Affect of Service, involving how librarians are perceived as interacting with users;
- Information Control, involving users' perceptions of whether users can find information how they want it, when they want it; and
- Library as Place, which involves the physical library facility and the role of the physical library as a symbol.

Also included are scores on perceived outcome impacts of library use, and generic user-reported satisfaction with library service quality. Finally, user group (1 = undergraduate student, 2 = graduate student, 3 = faculty) and gender (0 = female, 1 = male) are reported.. This data from table 17.1 is presented in Excel format and can be imported into most statistical software packages for analyses as indicated in the Chapter.

Chapter 18: Dr. Wilcox has kindly provided an updated paper on the topic:

Wilcox, R. R. & Keselman, H. J. (2012). Modern regression methods that can substantially increase power and provide a more accurate understanding of associations. European Journal of Personality, 26, 165--174. DOI: 10.1002/per.860

He also wrote: "Also, a lot has happened during the last 10 years. In case it helps, here is a list of books that includes recent advances:

- Wilcox, R. R. (2017). Understanding and Applying Basic Statistical Methods Using R. New York: Wiley ISBN: 978-1-119-06139-7
- Wilcox, R. R. (2017). Introduction to Robust Estimation and Hypothesis Testing 4th Edition. New York: Elsevier.
- Wilcox, R. R. (to appear). Modern Statistics for the Social and Behavioral Sciences: A Practical Introduction 2nd Ed. New York: Chapman & Hall/CRC press ISBN: 978-1-4987-9678-1 Cat \#: K30405".

Truer words have rarely been spoken. Thanks to a scholar and pioneer in the field of robust estimation for sharing his

Software. A practical issue is being able to apply the modern methods, summarized here, with easy-to-use software. All of the methods that were described are readily applied using functions written in R or S-PLUS. These software packages are nearly identical and provide a powerful and flexible approach to data analysis that, in general, makes modern methods readily accessible. R is free and can be downloaded from cran.r-project.org. The R functions for applying the methods covered here are available in two files: Rallfunv1.v4 and Rallfunv2.v4.
With version 2.2.0 of R, store these files in C:\Program Files\R\R-2.2.0\.
(So as new versions come out, 2.2.0 will be replaced by the new version number.) Once R is started, the command
    source("Rallfunv1.v4")
will incorporate all of the functions in the file Rallfunv1.v4 into your version of R. (The packages listed on the R web site is another excellent source of functions aimed at applying recently developed methods.) When using S-PLUS, download the files allfunv1.v4 and allfunv2.v4 instead and again use the source command. Some of the functions relevant to the methods covered here are described in Wilcox (2003, 2005). But some of the newer functions are not covered in any book, so for convenience, the names of these functions are given here.

Exercises and other ways to investigate the issues presented in this chapter: (download data)
There are four parts covering basics about modern methods, methods for comparing groups, methods for studying associations, and some descriptions and illustrations of my R and S+ functions. Some of the data sets are listed here. The first two deal with measures of depression among Palestinian youths who have or have not had a family member killed or wounded by an Israeli: depression1 depression2 *(these data sets have been lost)*. The file schiz contains measures of skin resistance stemming from four groups of individuals having to do with schizophrenia. The file read contains data on measures related to predicting reading ability in children. The files pygc and pyge deal with what is called Pygmalion in the classroom and provide an interesting ANCOVA example. The data are described in my 2003 and 2005 books in the ANCOVA section.

Chapter 19: Use the following data sets, provided by the chapter author, to reinforce your understanding of the chapter by working through the examples.

Lady Tasting Tea (click here to download data)
Lady tasting tea
In 1920, R. A. Fisher shared a story with his colleagues about how he resolved a statistical question in an innovative way. Once, Fisher met a lady who insisted that her tongue was sensitive enough to detect a subtle difference between a cup of tea with the milk being poured first and a cup of tea with the milk being added later. Fisher was skeptical and he presented eight cups of tea to this lady. Four of these eight cups were "milk-first" and four others were "tea-first." All cups were arranged in a random order yet the lady correctly identified six out of the eight cups (Salsburg, 2001). The test results are summarized in Table 1 (in the book).
Did the woman really have a super-sensitive tongue? This question can be reformulated as a statistical problem with the following two hypotheses:
- Null Hypothesis: The order in which milk or tea is poured in a cup and the lady's detection of the order are independent.
- Alternate Hypothesis: The lady can correctly tell the order in which milk or tea is poured in a cup.

Law School (click here to download data)

Bootstrapping
In classical procedures, parameter estimation requires certain parametric assumptions, but bootstrapping replaces the unknown population distribution with known empirical distributions, which are also called bootstrap distributions. The bootstrap methods began to attract more and more attention after Diaconis and Efron (1983) published an essay explaining bootstrapping using layman's terms in *Scientific American*. As discussed in this chapter's introduction, the beauty of resampling is its conceptual clarity. Resampling is highly accessible to many researchers whose primary concern is the content area of psychology or biology rather than mathematics. More importantly, as many permutation tests are built upon Fisher's counterfactual reasoning, basic bootstrapping principles also pave the way to advanced bootstrapping.

In the following example, let us revisit the simple, yet intellectually powerful example depicted by Diaconis and Efron in *Scientific American*. Please notice that in the following the bootstrap method will be illustrated with the use of Splus (Insightful, 2004), which was not available at the time of Diaconis and Efron's writing. Henceforth, the demonstration below is slightly different from that in the *Scientific American* essay.

In their article, Diaconis and Efron asked the readers to consider a group of 15 law schools, for which the academic achievements of each freshman are measured in terms of the average undergraduate GPA and the average score on the Law School Admission Test (LSAT) (Table 4, in the book). This small dataset indicates that the correlation between GPA and LSAT score is .776. Given Table 4, how confidently can the researcher assert that there is a positive correlation between GPA and LSAT in the law student population? Diaconis and Efron proposed the following strategy:
1. The original sample is duplicated one billion times. As a result, we have 15 billion observations instead of 15. This expanded sample is treated as a *virtual population* or a *proxy population*.
2. Samples are drawn from this virtual population to verify the estimators. Unlike permutation methods in which observations are *resampled without replacement*, the bootstrap employs *resampling with replacement*.

3. *Bias* is checked by comparing the statistic of the original sample against that of the empirical distribution. The bias estimated by the bootstrap method is the mean of the empirical distribution minus the statistic for the original sample.

It is highly advisable for readers to walk through the process using the full version or the trial version of Splus, as explained below:

1. Like RSE to Excel, the Resample Library is an external add-in module to Splus. After downloading the Resample Library and opening Splus, select *Load Library* from *File* and choose "resample."

2. Enter the law school data into a dataset.

3. Select *Correlation/resample* from *Statistics/Data Summary*.

4. Go to the tab *Bootstrap*. Check the boxes *Perform Bootstrap, Both Distribution and QQ, Percentiles, BCa Confidence Interval.* Set *Number of Resamples* to 1000. Then click OK.

During the bootstrapping process, the computer randomly selects 15 pairs of scores 1,000 times. At the end, these 1,000 resamplings generate an empirical distribution as shown in Figure 5. As you would expect, sometimes the resample yields a low correlation coefficient. In some extreme cases the correlation is close to zero, however, most of the time it returns a high correlation. The mean of these correlation coefficients is depicted as a dotted line, which almost overlaps the original observed correlation coefficient, 0.776, shown as a solid line.

Chapter 20: Play with prediction equations

Download the data set I used for the chapter (based on NELS '88 data from the National Center for Educational Statistics).

In this dataset are several variables:

- GPA_12- grade point average at 12th grade
- weighting- if you wanted to weight the variables to correct for sampling in the original data, weight by this variable
- Race1- 0= white, 1=African-American
- part- participation in extracurricular activities (more is better)
- bypared
    - 1=<high school education
    - 2=High school graduate
    - 3=some college
    - 4=Master's degree or equivalent
    - 5= Ph.D., MD, or other terminal degree

Step #1: Create a prediction equation based on the entire data set. This is your reference, or "population" prediction equation.

Step #2: Sample subgroups randomly or purposefully, analyze, and compare to reference, population equation. Use very small samples (N=20-50) and large samples (N=400-1000). Do several of each and see how the regression coefficients vary wildly around the reference equation.

Step #3. Practice cross validation and double cross validation. Calculate shrinkage (both described in the chapter). Which has more shrinkage, the smaller samples or larger ones?

Chapter 21: Since the release of this book, I released a book on Regression and Linear Modeling with an updated chapter on Poisson regression, and another on ordinal regression, and others (e.g., multinomial logit modeling). You can find information about it on my web page, along with extensive exercises. For resources from the original book chapter:

Example #1: Voter turnout data (from the chapter). Click here to download.

Example #2: <u>Nonmelonoma skin cancer</u>   Data on a rare event, with a low probability of occurrence. The Poisson distribution here is used to approximate the binomial. Both nominal (city) and interval/ratio (age) variables are used as predictors. An offset is also used.

Chapter 22: <u>detailed information and data sets here</u>

Chapter 23: From the author: **The links that I reference are**:

- <u>www.uvm.edu/~dhowell/AnovaChapter/FoaDoubled.dat</u>
- <u>www.uvm.edu/~dhowell/AnovaChapter/JSLdep.sav</u>
- <u>www.uvm.edu/~dhowell/AnovaChapter/AirportModified.dat</u>

Chapter 24: In the years since this book was published, I authored <u>a book on logistic regression</u> with extensive data sets and exercises, as well as a book on regression and linear modeling that covers many other regression models.

Original chapter resources: <u>overview and links to data</u> (<u>hosted link to data</u>)

Chapter 25: In the years since this book was published, I authored <u>a book on logistic regression</u> with extensive data sets and exercises, as well as a book on regression and linear modeling that covers many other regression models. You may find this resource interesting: <u>an overview of the simple math behind logistic regression.</u>

Chapter 26: In the years since this book was published, I authored <u>a book on logistic regression</u> with extensive data sets and exercises, as well as a book on regression and linear modeling that covers many other regression models.

<u>Original book chapter resources.</u>

Chapter 27: *Data.*

The data used in this illustrative analysis were from the Early Childhood Longitudinal Study (ECLS), an ongoing study by the U.S. Department of Education, National Center for Education Statistics that focuses on children's early school experiences beginning with kindergarten (Tourangeau, Nord, Lê, Pollack, & Atkins-Burnett, 2006). The study follows a nationally representative sample of children from kindergarten through fifth grade. The sample reflected all children from various racial and language background. Sampling for the ECLS was based on a dual frame, multi-stage sampling design, with 100 primary sampling units (PSU). For simplicity, only the data collected during 2004 from the fifth graders was in this paper. The sample size in the current analysis was 1,342 children, which included 650 males and 692 females. Among the total analysis sample of children, 797 were White, 126 were Black, 230 were Hispanic, 141 were Asian, and 48 were multiracial.

Measures. In the present analysis, four measures were used as independent variables. They were:

Self-Description Questionnaire—Math Self-Concept (Marsh, 1990). This measure assesses how children think and feel about themselves in terms of math competence. This scale includes eight items on math grades, the difficulty of math work, and interest in and enjoyment of math, with the score scale ranged from 1 to 4. The analysis used the average score of each participant.

Academic Rating Scale-Math. This is the teacher's rating of children's academic performance in math. Teachers were asked to rate each child's proficiency in the following areas: number concepts, measurement, operation, geometry, math strategies, and beginning algebraic thinking, with the score scale ranged from 1 to 5. The analysis used the average score of each participant.

Social Rating Scale-Approach to Learning. This is the teacher's judgment of children's social competence. The approach to learning scale measures behaviors that affect the ease with which children can benefit from the learning environment. It includes six items that rate the child's attentiveness, task persistence, eagerness to learn, learning independence, flexibility, organization, and following classroom rules, with the score scale ranged from 1 to 4. The analysis used the average score of each participant.

Social Rating Scale-Self-Control. It has four items that rate the child's ability to control behavior by respecting the property rights of others, controlling temper, accepting peer ideas for group activities, and responding appropriately to peer pressure, with the score scale ranged from 1 to 4. The analysis used the average score of each participant.

In all above measures, the scores were coded positively, with high scores indicating higher self-concept, and higher teacher rating on academic and social competence. The reported reliability for these independent variables ranged from .79 to .92 (Tourangeau et al., 2006).

The dependent variable used was a composite math proficiency probability score that was computed as an average across nine math skill levels: count/number, relative size, ordinality/sequence, add/subtract, multiple/divide, place value, rate and measurement, fractions, and area/volume. The probability scores were from 0.00 to 1.00, with a larger probability score indicating an overall higher achievement across these math skill levels.

In addition, children's gender and race were included as covariates. They were used to increase the classification accuracy of individuals into each latent class. In this chapter, children's race was represented in five categories: White, Black, Hispanic, Asian (which includes Pacific Islanders and American Indians), and multiracial.

Chapter 28:

Description of variables CONTAINED IN THIS DATA SET

partic	Participant's number
salary	Participants annual salary in $10
school	Number of years (& months) the participant attended school. The mean was 14.28, and it has already been mean centered.
iq	Participant's Intelligence Quotient measured during this study.
sex	Participant's sex. As the data is simulated, labeled non-descriptively as Sex1 and Sex2.
motiv	Participant's average score on a set of instruments designed to measure achievement motivation.
schmot	*motiv* ×*school*

Chapter 29: I revised this chapter several times and it is now part of the Regression and Linear Modeling book, which has exercises and data sets.

Chapter 30: Use the following data sets, provided by the chapter author, to reinforce your understanding of the chapter by working through the examples.

- [Data set #1:growth data](#)
- [Data set #2: teacher data](#)

Chapter 31: [web page for further exploration](#) with [links to data](#)

Chapter 32: [data summary](#) from chapter

Chapter 33: no resources provided

Chapter 34: no resources provided