

P8130 Fall 2019: Biostatistical Methods I

Homework 1

Due, Sept 20th @ 5:00pm

P8130 Guidelines for Submitting Homework

Your homework should be submitted only through CourseWorks. No email submissions!

All derivations, graphs, output and interpretations to each section of the problem(s) must be included in the PDF (not the code), otherwise it will not be graded.

Only 1 PDF file should be submitted. When derivations were required and handwriting was allowed, scan the derivations and merge ALL PDF files (<http://www.pdfmerge.com/>) into a single one.

We are encouraged to use R for calculations, but you still have to show the mathematical formulae. Also, make sure to also submit your commented code as a separate R/RMD file.

DO NOT FORGET:

You are encouraged to collectively look for answers, explain things to each other, and use questions to test each other knowledge.

But

Do NOT hand out answers to someone who has not done any work. Everyone ought to have ideas about the possible answers or at least some thoughts about how to probe the problem further. Write your own solutions!

Problem 1 (30p)

A study was conducted over a six-month period at local ambulatory virology clinic. The goal was to test the effect of a structured exercise program for overweight/obese low-income virally suppressed HIV positive subjects on different parameters, such as metabolic, immunologic, and well-being parameters. A number of individuals agreed to participate in the intervention group and others (remaining candidates in the pool) were selected as non-participating controls, matched by gender.

The investigator is asking you to help with the following:

- a) Create a table summarizing the subjects' demographics and medical comorbidities by intervention group (1) vs control (0). Based on the variable type, use descriptive statistics (N, mean \pm SD, median (IQR), n(%)) for the following variables: Age, Gender (1: male, 2: female), Race/Ethnicity (1 and 3: African American, 2: Hispanic, 4: Caucasian, 5 and 6: Other), Depression (1: yes, 0: no), Smoking (1: yes, 0: no), Hypertension (HTN) (1: yes, 0: no), Type 2 Diabetes (T2DM) (1: yes, 0: no) (10p)

Create this table in R using functions of your choice. For this part you are not supposed to use Word and add the values manually – code will be checked!

From R you can then export the table or if you're familiar, feel free to use R Markdown to generate it as 'html' or 'pdf' file. The variables names should be clear to a general audience reading this report (no acronyms). Below is an example of how this table should look – it does not have to be identical in style, but definitely contain the requested information (your variables/categories and descriptive values might also differ).

Table 1: Demographics and co-morbidities

	Intervention N = XX	Control N = XX
Age		
Mean \pm SD	53.58 \pm 9.58	51.5 \pm 10.81
Median (IQR)	55.5 (47 – 59.5)	51 (44.5 – 60.5)
Gender, n(%)		
Male	16 (44.44%)	16 (44.44%)
Race, n(%)		
African American	30 (83.33%)	21 (58.33%)
Hispanic	5 (13.89%)	14 (38.89%)
Caucasian	1 (2.78%)	1 (2.78%)
HTN, n(%)		
Yes	22 (61.11%)	19 (52.78%)

- b) You will now investigate changes in the following metabolic parameters: BMI, Glucose, Systolic and Diastolic Blood Pressure, HDL and LDL Cholesterol (15p)

- i) Use descriptive statistics to summarize these variables by intervention group (vs control), pre/baseline - post (6 months) measurements. Add the absolute changes as well.

This table has a more complicated layout, so you can use Word to edit the values. If you feel adventurous, please try R, but it's optional. Example table is shown below:

Table 2: Pre/Post changes in metabolic parameters

	Intervention N = XX		Control N = XX	
	Baseline	6 month	Baseline	6 month
BMI	31.91 ± 6.58 29.25 (27.35-34.7)	31.21 ± 6.13 29.15 (26.8-33.05)	34.23 ± 6.16 33.4 (29.5-37.65)	34.51 ± 5.97 33.05 (30.35-37.8)
Δ	-0.76 ± 1.44		0.28 ± 0.97	

- ii) Use only BMI and LDL variables to create two separate graphs. In each graph, generate side-by-side boxplots by intervention group and pre/baseline – post measurement. Make sure you label your figures appropriately.
- iii) Succinctly discuss your findings/trends obtained in parts i) & ii).
- c) What do you think of the study design and what are some potential issues that might affect its validity/generalizability? (5p)

Problem 2 (10p)

Down syndrome (DS) is a chromosomal condition that occurs in about one in 1000 pregnancies. The most accurate test for DS in wide use requires amniocentesis, which unfortunately carries a risk of miscarriage (about one in 200). A lower-risk test in common use is called the triple test; it screens for levels of three hormones in maternal blood at around 16 weeks of pregnancy. However, the triple test does not always correctly identify a fetus with DS, and sometimes it incorrectly identifies a fetus with a normal set of chromosomes as DS. Under normal conditions, the detection rate of the triple test (i.e., the probability that a fetus with DS will be correctly scored as having DS) is 0.60. The false-positive rate is 0.05.

If the test on a randomly chosen fetus gives a positive result (i.e., it indicates that the fetus has DS), what is the probability that this fetus actually has DS? What is this name of this specific probability value? Also, comment on the magnitude of the result obtained.

Problem 3 (20p)

Find a story in the media (newspaper, online) that discusses the results of a piece of research. If the same story was reported in several media outlets, comment on the differences re: how the facts were presented. Every (decent) reported story should have references of the original

research study. After reading the original reference(s), comment on the aspects of study design, i.e., randomized vs observational study, sample selection and sample size, possible bias, interpretation of the results. If you feel confident, you can also comment on the analysis, but it is not required for this assignment. Decide how seriously you should take the results of the research.

Your summary, including your recommendations should not exceed one page. Note: Link to the media post and journal article(s) should be listed as references.