Problem Set 3

PUBHLTH 490Z

Due 23 November 2021

Problem set policies. *Please provide concise, clear answers for each question. Note that only writing the result of a calculation (e.g., "SD = 3.3") without explanation is not sufficient. For problems involving* R*, include the code in your solution, along with any plots.*

Each problem set is due by 11:59 pm on the due date; please submit your problem set via Moodle as a PDF.

We encourage you to discuss problems with other students (and, of course, with the instructor and the TA), but you must write your final answer in your own words. Solutions prepared "in committee" are not acceptable. If you do collaborate with classmates on a problem, please list your collaborators on your solution.

Behavioral Risk Factor Surveillance System (BRFSS) dataset

"Do men and women think differently about their body weight?" To address this question, use the Behavioral Risk Factor Surveillance System (BRFSS) dataset; the data are in `BRFSS.Rdata`. The dataset includes information about the actual weight and desired weight (coded as `weight` and `wtdesire`, respectively) for survey participants. The variable `sex` is coded 0 for males and 1 for females. For your convenience, the variable `sex.factor` is also provided; sex is coded with two levels, "Male" and "Female".

This problem is structured in a way that reflects how statistics typically arises in a realistic scenario; start with a substantive question and associated data, then use the tools of statistics to examine what insights the data may provide. As happens in practice, the question does not specify exactly what to do with the data. There will be multiple perfectly reasonable ways to approach the problem and present conclusions. Focus on crafting a coherent solution that is clear, well-written, and supported by the data.

a) Draw a random sample of size 300 from the data, using 2017 in the `set.seed()` function. Name the sample `brfss`.

b) Fit a linear model to predict desired weight from actual weight and assess whether actual weight is a significant predictor of desired weight.

c) Fit a linear model to predict desired weight from actual weight and sex. Compare this model to the one fitted in part b); state which model is preferable, and explain your reasoning.

d) Investigate whether the association of desired weight with actual weight is different for males versus females.

   i. Create a plot that shows the regression line for desired weight versus actual weight in males and the regression line for desired weight versus actual weight in females.

ii. Create a model that includes actual weight, sex, and the interaction between actual weight and sex. Write the model equation.

iii. Write the prediction equation for males and and the prediction equation for females.

iv. Is there statistically significant evidence of an interaction between actual weight and sex? Explain your answer.