

STAT 614 - HW 4

Due: Thursday, October 22, 2020 in Blackboard by 11:59pm.

Instructions: Please type your solutions and upload the document as a pdf file in Blackboard. There is only one file to submit for this assignment. As part of this assignment, please take the completely anonymous Midterm Course Evaluation under the Survey tab in Blackboard.

Notes:

- For this HW you will need some concepts from chapter 3 on checking assumptions and transformations and chapter 4 on nonparametric methods.
- You will also be revisiting the “big ideas” around confidence intervals and hypothesis tests.

The food-frequency questionnaire (FFQ) is an instrument often used in dietary epidemiology to assess consumption of specific foods. A person is asked to write down the number of servings per day typically eaten in the past year of over 100 individual food items. A food-composition table is then used to compute nutrient intakes (protein, fat, etc.) based on aggregating responses for individual foods. The FFQ is inexpensive to administer but is considered less accurate than the diet record (DR) (the gold standard of dietary epidemiology). For the DR, a participant writes down the amount of each specific food eaten over the past week in a food diary and a nutritionist, using a special computer program, computes nutrient intakes from the food diaries. This is a much more expensive method of dietary recording. To validate the FFQ, 173 nurses participating in the Nurses’ Health Study completed 4 weeks of diet recording about equally spaced over a 12-month period and an FFQ at the end of diet recording. Data are in Blackboard in the file valid.txt.

Consider the data on total alcohol consumption for both the DR and FFQ, `alco_dr` and `alco_ffq`, respectively. You are to assess whether the two methods, diet record and the food-frequency questionnaire, are comparable for total alcohol consumption. In particular, is there evidence that FFQ *underestimates* total alcohol consumption, in general? Estimate by how much the FFQ generally underestimates total alcohol consumption.

1. Explain why the initial model needed to address these research goals is a matched-pairs t-procedure.
2. Use both the model notation we developed in class and a brief written description of the model (you may also use pictures) to illustrate the model. (Be careful! The matched-pairs procedure works on the *difference* in the two measures on each individual. Start with $y = \text{alco_dr} - \text{alco_ffq}$ and describe the model for y !)
3. What are the model assumptions?
4. Which of the model assumptions are not met? Give and refer to specific output.
5. Consider a square root transformation of the alcohol data: $\text{salcoDR} = \sqrt{\text{alco_dr}}$ and $\text{salcoFFQ} = \sqrt{\text{alco_ffq}}$. Are the model assumptions met for the transformed data? Give and refer to specific output.
6. Conduct the appropriate test on the square root transformed data and interpret the results. Be sure to address the research questions stated above.
7. Consider a nonparametric method for addressing the research questions. What null and alternative hypotheses are addressed by the appropriate nonparametric method? Carry out and interpret the results of the nonparametric method. Include and interpret the confidence interval estimate.
8. Which of the results in (6) or (7) do you prefer to use to draw conclusions for this study and why?