PSC 103B

Homework 2

Spring 2024

**Instructions**

Please use R/RStudio to complete the following questions. You will submit your answers on Canvas as either a Word Document or PDF. You will be required to **include the code you used to generate your answer for each question as well as the final answer and/or relevant output** for each question.If you choose to submit only the graded questions, make sure you include relevant code and output that might be in previous questions. You can copy and paste the code you used as well as the output, or take screenshots of it and paste them into your document, or any combination that works for you. Please include all the relevant work and information for each question under that questions’ prompt. If you want to make the output more readable, you can format it using a monospaced font like Courier or consolas.

We encourage you to do every question, but only those marked with an asterisk will be graded.

You may get help from your classmates while working on the assignment, but you must do all the work yourself. You may consult each other, but everything you turn in must be your own code and words. Please make good use of the Slack discussion channel if you get stuck. Posting there can help other people who also have that question! Remember that you are welcome to ask questions about both graded and ungraded questions – we really want you to complete the whole thing, because it’ll help you learn.

Please submit your answers on Canvas by **10am on May 2nd.**

1. **Conceptual Understanding**
2. \*(1pt) Provide a definition of statistical power in the context of an independent samples t-test.
3. \*(1pt) What 3 factors affect power? What is the factor that is most commonly used to increase power? In other words, if someone wants to increase the power of their study, which of these factors would they likely modify?
4. Describe in your own words how t-distributions are used to test null hypotheses.
5. Are there any circumstances under which sample size affects effect size? Why?
6. Compare the t-distribution to the chi-square distribution. How are they similar and how are they different?
7. **Computing Power for an Independent Samples T-Test (Don’t forget to include your code!)**
8. \*(1pt) How many participants do you need to detect a mean difference of *d* = 1.0, with 95% power, and an alpha level of .05?
9. How much power do you have to detect a mean difference of *d* = .1, with a sample size of 1,000 per group, and an alpha level of .05
10. \*(1pt) What effect size would I need to be able to have 90% power to reject the null hypothesis, given a sample size of 100 per group, at an alpha level of .05?
11. **T-Tests**
12. Imagine that you have developed two new breeding programs for the common housefly. In the first sample, you have specially bred flies to have longer than average lifespans (*N* = 10). In addition to your mutant flies, you have another sample of flies you have bred to have much shorter life spans (*N* = 10). You want to see if there is a mean difference in the lifespans of flies in your two breeding programs. Use the data below to answer the following questions:

Lifespan of Mutant flies (days) = {27, 25, 20, 25, 23, 21, 27, 25, 25, 22}

Lifespan of Short-lived flies (days) = {24, 23, 19, 21, 22, 20, 25, 27, 21, 22}

1. Use the code below to input the data on fly life spans into R. Check to see that the information on mutant and short-lived flies are in separate columns.

flies <- data.frame(mutant = c(27, 25, 20, 25, 23, 21, 27, 25, 25, 22), shortlived = c(24, 23, 19, 21, 22, 20, 25, 27, 21, 22))

1. State your null and alternative hypotheses. Do you have a directional or non-directional hypothesis?
2. \*(1pt) What test would you use to statistically test your hypothesis? Justify your answer.
3. \*(0.5pt) Run the test you selected in the previous question (c). Use an alpha level of .05. Include both your code and the test output in your answer
4. What can you conclude? Report your results as you would in a research paper, (i.e., all relevant information in sentence form).
5. It just so happens that, after you ran your analysis, a research assistant told you some new details about these fly experiments. It turns out that, to control for environment, pairs of flies were assigned to live in the same cages for the duration of their lives. That is Fly 1 of the Mutant Flies and Short-lived flies lived in the same enclosure, as did Fly 2 of the Mutant Flies and Short-lived Flies, and so on. This suggests that each pair of flies had some common factors and that the samples are not actually independent. Use this information, along with the same data from Question 11, to answer this next set of questions:
   1. \*(1pt) State your null and alternative hypotheses. Do you have directional or non-directional hypotheses?
   2. What test would you use to statistically test your hypothesis? Justify your answer.
   3. \*(0.5pt) Run the test you selected in the previous question (b). Use an alpha level of .05. Include both your code and the test output in your answer
   4. \*(1pt) What can you conclude? Report your results as you would in a research paper, (i.e., all relevant information in sentence form).
6. **Chi Square Tests**

For the next few questions, we will use a real dataset about cereals! Download the cereal\_103B.csv dataset from Canvas and load it into R using read\_csv. You can read more about the dataset here: <https://www.kaggle.com/crawford/80-cereals/data>. We’ve modified it to include a few more variables!

1. The variables *SodiumLevel* and *CaloriesLevel* were created based on the variables *sodium* and *calories* respectively in the original dataset. They classify cereals into High vs. Low sodium, and High vs. Low calories. Test the independence of these two variables (SodiumLevel and CaloriesLevel)
   1. Use the function table() to see the number of cereals in each of the 4 categories: low calories and low sodium, high calories and low sodium, etc. Do you think these variables are independent?
   2. \*(1pt)What is H0? What is H1?
   3. Run the chi square test of independence. Report both your code and the test output. Report a p-value and a Bayes factor.
   4. \*(1pt) Report your results as you would in a research paper, (i.e., all relevant information in sentence form) and interpret these results (you can interpret either the p-value or the Bayes factor). What can we conclude about the levels of sodium and calories of these cereals, in general?