PSC 103B

Homework 3

Spring 2024

**Instructions**

Please use R/RStudio to complete the following questions. You will submit your answers on Canvas as a PDF file, which can just be this file saved as a 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 Monday, May 15th**.

1. **Conceptual Understanding**
2. \*(1pt) In your own words, explain what the group and residual sum of squares (SS) from an ANOVA table represent.
3. \*(1pt) Why, when the null hypothesis is true, is the expected value of the F-statistic 1? And why, when the null hypothesis is false, is the expected value of the F-statistic greater than 1. (Hint: talk about what the numerator and denominator in the F-ratio represent).
4. You have been asked to serve as a statistical consultant for several proposed projects. For each of the following, indicate what statistical test you would use. Briefly explain why you chose that test, and state the H0 and H1 hypothesis. (Hint: It may help to write-out whether each variable is continuous or categorical, and, if categorical, how many levels that variable has).
   1. (\*1pt) The President of UC Davis wants to see whether anything can be done about student drinking. She intends to have a random sample of 250 students fill out a questionnaire reporting on how much alcohol they drank in the last month. The students will then see a film on alcohol abuse. A month later, they will again report on how much alcohol they drank in the last month.
   2. (\*1pt) An office manager wants to test whether productivity per worker (measured on a scale that runs from 0 to 100) is the same regardless of how crowded an office is. The manager plans to draw independent samples (each of size 60) from Office A (no crowding), Office B (some crowding), and Office C (severe crowding) and compare their productivity.
   3. An insurance company sells three kinds of life insurance: Whole Life Insurance, Universal Life insurance, and Term Life Insurance. An analyst for the company wants to know whether there is any association between region of the country (North, East, South, and West) and the type of life insurance the company’s customers buy. A random sample of 500 customers from the West, 700 customers from the North, 300 customers from the East, and 600 customers from the South will be drawn.
   4. (\*1pt) A researcher believes that racism (measured on a scale that ranges from 0, not racist, to 25, extremely racist) varies by geographic region (North, South, East, or West) and Religion (Catholic or NonCatholic). For each combination of region and religion, 200 individuals are interviewed.
5. **ANOVA (Include your code!)**

A school district wants to know about the effects of using positive, negative, or no reinforcement to help children learn math. To examine this, they randomly assigned each student to one of three possible learning programs: positive, in which the children received positive reinforcement for getting correct answers on their practice problems; negative, in which the children received negative reinforcement for their incorrect responses on practice problems; or control, in which students were neither positively or negatively reinforced for their responses on practice problems. At the end of the learning program, students took a final math exam and their scores were recorded.

Use the data set (OneWayAnovaData.csv) from the study to answer the following questions. The outcome variable is labeled “score”, these values indicate the student’s score on the final math exam after the learning program. The predictor variable is “condition”, which indicates the student’s learning program (“Positive” = positive reinforcement; “Negative” = negative reinforcement; “Control” = neither positive or negative reinforcement).

1. State the null hypothesis and alternative hypothesis (use , , and to denote the means for each condition) for a one-way ANOVA testing the effect of reinforcement.
2. Conduct a one-way ANOVA to determine whether assigned condition had an effect on math scores. Can you reject the H0?
3. Run a post-hoc comparison to test if the reinforcement conditions (positive and negative) are significantly different from the control condition (use the data and ANOVA results from the previous question).
   1. (\*1pt) State the H0 for the contrast and write out the equation for the weighted sum of group means (ψ). Make it clear what comparison weights you are assigning to each mean.
   2. (\*0.5pt) Calculate the F-statistic for the contrast. Report and show your work for all steps.
   3. (\*0.5pt) Can you reject the H0? Make sure you correct for multiple testing. Interpret your decision (reject or retain H0) in the context of your research question.
4. (\*1 pt) Write a paragraph reporting your results from the one-way ANOVA and the post-hoc as you would in a research paper, including all relevant detail from the tests.
5. **Factorial ANOVA (Include your code!)**

A clinical researcher is interested in whether two different medications (Anxifree and Joyzepam) are effective at treating anxiety and increasing patients’ mood, compared to a placebo. They suspect that the impact of the medication would also depend on whether the patient was in therapy or not while taking the medication. To test this, they recruited patients who were and were not in therapy, and randomly assigned them to take 1 of the 3 medications (Anxifree, Joyzepam, and a placebo) for 1 month.

Use the dataset (FactorialAnova\_Mood.csv) to answer the following questions. The outcome variable is called “mood.gain”, which indicates the patient’s improvement in mood after 1 month (so higher values indicate more improvement). The two grouping variables are “drug”, which indicates which medication the patient was assigned to take, and “therapy”, which indicates whether the patient was enrolled in therapy or not.

1. (\*1 pt) State the null hypothesis and alternative hypothesis.
2. Conduct a factorial ANOVA to answer the research question. Make sure to include an interaction term for the two grouping variables!
3. Compute Bayes factors for the factorial ANOVA you conducted. Assuming equal prior probabilities, compare the full model with all the main effects and the interaction to reduced models, where a single effect is omitted at a time (hint: this is exactly how we did in lab).
4. (\*1pt) Write a paragraph reporting the results from the previous analyses. Include the test statistic, degrees of freedom, and p-value for each of the main effects and the interaction. If there is a main effect of drug type, don’t worry about conducting and reporting post-hoc tests, just report the information mentioned above. If there is a main effect of therapy, report which group experienced greater mood gain.