431 Lab 07

Deadline: See Course Calendar | Last Edited 2022-11-27 17:40:31

Table of contents

Deadline

Lab 07 has 8 questions, all of which you need to complete by the deadline specified on the Course Calendar.

- To receive full credit on a Lab, it must be received on Canvas no later than 59 minutes after the posted deadline. (This allows for small issues with uploading to Canvas to occur without penalty.)

Learning Objectives

1. Appropriately prepare data, check assumptions, execute appropriate modeling strategies, and interpret results

An Important Note

Your response to **every** question, whether we explicitly ask for it or not, should include a complete English sentence responding to the question. Code alone is not a sufficient response, even if the code is correct. Some responses might not need any code, but every response needs at least one complete sentence.

The Data for Lab 07

Lab 07 uses some data which we have simulated to reflect a clinical trial. In this trial, the investigators are testing out a new drug to see its effect on a subject's systolic blood pressure (SBP). For our purposes, if a subject's SBP is over 130 mm Hg they are considered to have hypertension. In this trial, the goal was to reduce the SBP from a baseline level by using a new drug (Treatment C), and comparing that to the current top-of-the-line drug (Treatment B), and the oldest drug (Treatment A). The trial focused specifically on non-Hispanic African-American women who were in long-term relationships and between the ages of 55 and 65 years old, with a minimal comorbidity profile. The outcome of interest is the post-treatment systolic blood pressure (`sbp_follow`), and we are also given the subject's age, their pre-treatment systolic blood pressure (`sbp_baseline`) and whether or not the subject's partner has hypertension.

Variable	Description
subjectid	unique subject identifier
group	treatment group where 1 = Group A, 2 = Group B, 3 = Group C
partner	whether or not the subject's partner also has hypertension
age	subject age in years at baseline

Variable	Description
sbp_baseline	subject's baseline systolic blood pressure (mm Hg)
sbp_follow	subject's follow-up systolic blood pressure (mm Hg)

The data has been made available to you in an Excel file called `lab07_trial.xls` available on our 431-data page. Note that I adjusted this file on 2022-11-01 to eliminate a column that preceded the `subjectid`. Be sure to grab the latest version of this file.

Question 1 (10 points)

Ingest the data, and then make sure that (a) group is a meaningfully leveled factor, (b) partnership status is a factor, and (c) age and the two blood pressure results are properly numeric variables, and to look for missing data. Once you've done this, run some simple and attractive, well-annotated summaries to report the number of subjects per group, as well as the partnership status, age and baseline blood pressure levels of the subjects in each group. Do the groups look comparable on these three baseline variables?

Question 2 (10 points)

Next, we'd like to know if our outcome, `sbp_follow`, seems appropriately modeled with a Normal distribution. Create a figure containing at least two nicely annotated panels (presented together using tools from the patchwork package) to assess whether a Normal model might be appropriate for this outcome. One of your panels should include a well-labeled Normal Q-Q plot. Then write a sentence of two to describe what you've done and your conclusions.

Question 3 (15 points)

Produce a figure to compare the three groups that allows you to assess the Normality and Equal Variances assumptions of an ANOVA to compare the SBP at follow-up means across the three treatment groups, ignoring all other information available in the data. What conclusions can you draw about ANOVA assumptions in this setting?

Question 4 (10 points)

Now complete the comparison of the SBP at follow-up means of the three treatment groups (A, B and C) using an analysis of variance. What conclusions do you draw, using a **90%** confidence level? Call this `model4`.

Question 5 (10 points)

Augment your results in Question 4 by incorporating baseline SBP levels into the comparison, without using an interaction term. Call this `model5`. How do your conclusions change about the effects of the various treatment groups in this revised model? Again, use a 90% confidence level.

Question 6 (10 points)

Now, create `model6` by augmenting the model you fit in Question 5 to see if `partner` may also play a meaningful predictive role in a model for our outcome, again without using an interaction term. Interpret whether the model's quality of fit has improved, and and discuss what the addition of `partner` did to your estimates about the impact of the treatment groups on the outcome.

Question 7 (15 points)

Now, to create your final model (`model7`), instead of adjusting for `partner`, take into account the subject's `age` as well as the baseline systolic blood pressure, again without using an interaction term.

Then build a comparison of the four models you've fit (in models `model4` through `model7`) in terms of the quality of fit (as measured by R-square, adjusted R-square, sigma, AIC and BIC) in the available data. What conclusions can you draw about fit quality for in comparing these four models? Does one model stand out as better or worse than the others? Why or why not?

Question 8 (20 points)

Throughout the above questions, we've been presented with a number of p values. In Chapter 10 of Spiegelhalter's *The Art of Statistics*, there is a robust discussion of p values. Write a short essay (150 words would be sufficient), which applies what you've learned from Spiegelhalter's Chapter 10 to the analyses you completed in Questions 1-7.

Session Information

Be sure to include the session information using one of the methods we have demonstrated.

Submitting your Response

Submit both your working R Markdown file and the HTML output file to Canvas in the Lab 07 section of the Assignments folder by the deadline specified in the Course Calendar.

Getting Help

You are encouraged to discuss Lab 07 with Professor Love, the teaching assistants or your colleagues, but your answer must be prepared by you alone. Don't be afraid to ask questions, using any of the methods described on our Contact Us page.

Grading

We will summarize some of the more interesting responses to Question 8 after the Lab has been graded.

- This Lab will be graded on a scale from 0-100.
- Note that the teaching assistants will review your responses to all Questions carefully to assess clarity of writing, attention to detail, and adherence to grammatical and syntax requirements. Spelling, grammar, syntax and the rest all matter for grading purposes in this and all other assignments this term.

A detailed answer sketch for this Lab will be provided on the day after the submission deadline, and a grading rubric will be provided when the grades are made available, approximately one week after the submission deadline.

Late Penalties for Lab Work

- Labs that are turned in 1-12 hours after the deadline will lose 10% of available points.
- Labs turned in more than 12 but less than 72 hours after the deadline will lose 25% of available points.
- No extensions to Lab deadlines will be permitted this semester. Labs turned in more than 72 hours after the deadline will receive no credit.
- Your lowest lab score (out of Labs 1-7) will be dropped before we calculate your lab grade.