431 Lab 07

Deadline: Tuesday 2023-11-28 at Noon

431 Staff

2023-10-15

Table of contents

0.1 Deadline

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

0.2 Learning Objectives

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

0.3 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
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.

1 Question 1 (5 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?

2 Question 2 (5 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.

3 Question 3 (5 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?

4 Question 4 (5 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`.

5 Question 5 (5 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.

6 Question 6 (5 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.

7 Question 7 (10 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?

8 Question 8 (10 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.

9 Session Information

Be sure to include the session information, in a section of its own, using one of the methods we have demonstrated.

10 What Happens After Submission?

Submit **both** your Quarto file **and** the HTML output file to the Lab 07 section in the Assignments folder in Canvas by the deadline specified in the Course Calendar. We will need both the Quarto and HTML file submitted before we can grade your work.

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 0-50 points.
- 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 the Lab will be provided 48 hours after the submission deadline, and a grading rubric will be provided when the grades are made available, approximately one week after the submission deadline.