431 Lab 04

Deadline: Tuesday 2023-10-17 at Noon

431 Staff

2023-08-10

Table of contents

0.1 Deadline

Lab 04 has 6 questions, all of which you need to complete by the deadline specified on the Course Calendar.

0.2 Getting Help

You are welcome to discuss Lab 04 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.

0.3 Learning Objectives

1. Critically evaluate a news story on scientific literature, incorporating the principles of probability and odds
2. Given data, work through an analysis of variance including visualizing the data, building the ANOVA model, and using the Bonferroni-Holm approach to assess pairwise comparisons of means.

0.4 There is no template for Lab 04

We encourage you to use a similar approach to Lab 04 that you used in Labs 02 and 03. We have not provided a template for Lab 04, but you can adapt one of the ones we've provided previously, if you like. Or you can use an approach that you think works well. Be sure to use a new section for each question on the Lab, and do not hide your code. (If you want to use code-folding, set it to show.)

0.5 Submitting the Lab

Submit **both** your revised Quarto file **and** the HTML output file to the Lab 04 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.

Again, we encourage you in the strongest possible terms to **ask questions**, using any of the approaches described on our Contact Us page.

0.6 Our Labs Page

Visit our 431-Labs page for more information, including:

- how to get help with the Lab,
- our late policy,
- details on answer sketches and grading rubrics
- how your lab will be graded, and
- how we handle requests to regrade a Lab.

Part A: News Story and Research Article (Questions 1-4)

Find a headline from the internet related to health or medicine that describes the findings of a study published on January 1, 2017 or later. Then find the study being referred to in PUBMED. Use the formula for updating your opinions about health news developed in this article by Jeff Leek, along with the abstract and full contents of the published study to complete Questions 1-4. While it won't be necessary to prepare any R code to respond to Questions 1-4, we think it will be good practice for you to prepare your response in Quarto anyway.

1 Question 1 (3 points)

a. Specify the URL where we can see the headline and news story describing the findings of the study. Feel free to use `bit.ly` or a related tool online to produce a shortened URL for this purpose. Specify the reference completely, including the names of the author(s) of the news story, and its full title, and source.

b. Specify a URL where we can see at least the abstract of the complete study. Again, shortened URLs are fine. Give the complete reference to the study, as well, including the authors, full title, journal name and so forth.

2 Question 2 (5 points)

Describe, in a few sentences, your original opinion (gut feeling) related to the conclusions of the study as summarized in the headline and news article, first in terms of a probability statement, and then calculate the appropriate odds, remembering to convert statements about probabilities to statements about odds. Provide some motivation for your internal prior probability, describing your relevant personal experiences or other factors that drove your gut feeling.

2.1 Note for Question 2

If X is an event, and `Pr(X)` is the probability that X occurs, and `odds(X)` are the odds that X occurs, then

`Pr(X) = odds(x) / (1 + odds(x))`

and

`odds(X) = Pr(X) / (1 - Pr(X))`.

3 Question 3 (12 points)

Evaluate the study in terms of the six specifications proposed by Jeff Leek in this article at FiveThirtyEight when evaluating study support. Be sure to specify your conclusion about **each** of the six specifications, and provide direct quotes and summarize the evidence from the abstract or paper to address the issues raised and justify your conclusions. We want to see a clear, motivated conclusion about each of the six specifications, as well as direct quotes and evidence summaries to address the issues raised and justify conclusions. We suggest you use a different subheading for each of the six specifications so it's easy for us to see your conclusions in each case.

4 Question 4 (8 points)

Incorporate the study support assessment into a Bayes' Rule calculation to obtain the final odds you should now be willing to give to the headline, and specify this value in terms of a probability statement, as well. Then react to the final conclusion specified by this approach in a few sentences. How does your subjective posterior probability that the headline is true match up with the formula's conclusions? Do you feel that the formulaic approach has yielded an appropriate conclusion for you in this case? Why or why not?

Part B: Palmer Penguins, again (Questions 5-6)

5 Question 5 (12 points)

Start with the 342 penguins in the `penguins` tibble contained in the `palmerpenguins` package in R, who have **complete** data on both `island` and `body_mass_g`. Recall that we used the `palmerpenguins` data in a previous Lab.

Now, create two different attractive and thoughtfully labeled plots (as described below, including an appropriate title for each) that show the relationship between the penguin's home island and their body mass (in g) for those 342 penguins.

- The first of these plots should be a combination violin and box plot.
- The second of these plots should be a faceted set of Normal Q-Q plots.

No additional text is required besides the two well-labeled plots for this Question.

6 Question 6 (10 points)

Build a linear model to examine the relationship between body mass in grams (the outcome of interest) and island (our predictor.) Include an appropriate ANOVA table, and use a Bonferroni-Holm approach to complete the testing of available pairwise comparisons of means. Interpret your results briefly, and in addition to summarizing your model's findings, you also need to comment on the plausibility of the Normality assumption for ANOVA based on your figures in Question 5.

7 Session information

At the end of your Quarto file, please include a new code chunk to provide the **session information**, as you have done in previous Labs.

8 What Happens After Submission?

We will summarize some of the more interesting responses to Questions 1-4 after the Lab has been graded.

- This Lab will be graded on a scale from 0-50.
- 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 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.