P8122 Homework 2
Due: 09/30/2024 at 5:00pm

Instructions

- Upload a single pdf file for your homework along with your R Script on Canvas.

- The R script should be self-contained, so someone else should be able to run it as is and get your results. The script should also be well-commented, so it is clear which code goes with which question.

- You may discuss these problems with each other verbally, but must write up the answers on your own, and may not share or show your answers to anyone else.

- Short and clear answers, please.

- No late homework are allowed.

Exercises

Consider a study on the effect that light at night has on weight gain and other variables in mice. The goal of this study was to determine whether light at night may play a causal role in the obesity epidemic. The data comprises $n = 30$ mice randomized to three different treatment groups. All mice spent 16 hours in light, and the explanatory variable was the level of light during the remaining 8 hours. Some of the mice were randomized to darkness during those 8 hours (as is typical for regular mice), some were randomized to a dim light (equivalent to a TV on in the room for humans), and the remaining mice were exposed to bright light for all 24 hours. Mice are nocturnal, and typically most of their activity and eating happen at night. The hypothesis of this study was that having a light on at night may alter mouse eating habits and/or metabolism, and so increase body mass. The data are available on canvas. There are some missing values because one mouse died, and one mouse did not receive the full glucose injection for it's glucose tolerance test.

Table 1: Data Description

Variable Name	Variable Description
Light	Treatment group (dark, dim, bright)
BMGain	Change in body mass from the beginning of the experiment to the end (week 8)
Consumption	Average daily consumption (in grams)
Corticosterone	Blood corticosterene level (a measure of stress)
DayPct	Percentage of calories consumed during the day
GlucoseInt	Glucose intolerant at end of study? (Yes or No)
GTT15	Glucose level in the blood 15 minutes after a glucose injection
GTT120	Glucose level in the blood 120 minutes after a glucose injection
Activity	A measure of physical activity level

1. (5 points) We are interested in the causal effect of light at night on weight gain. Plot the outcome by treatment group.

2. (5 points) Here we will compare the mice exposed to darkness to the mice exposed to bright light overnight (once you have the code it is easy to rerun the analysis for the dim light group, if you are interested). Subset the data to only consider these two groups.

3. (15 points) Set up the data such that everything you will need has generic names (such as Y_{obs} or whatever you want to call them). Everything specific to the context of your data (variable names, sample sizes) should only be in your R Script here. Everything else should be generic so you can copy/paste it for later use. What quantities will you need to evaluate the causal effect of light at night on weight gain?

4. (10 points) Suppose we want the statistic to be the difference in means between the two treatment groups. Calculate T_{obs}.

5. (10 points) How many different possibilities are there for A? Enumerate all of these possibilities in a matrix. (Hint: it's probably easiest to first install the `ri` or `perm` package, have a look at the function `chooseMatrix` in R, it may come in handy.)

6. (15 points) State the sharp null hypothesis of no difference. Calculate the test statistic under one of these possibilities for A (the first one), under the sharp null hypothesis.

7. (10 points) Generate the exact randomization distribution for T, under the sharp null hypothesis of no difference.

8. (10 points) Plot this distribution, and mark the observed test statistic.

9. (10 points) Calculate the exact p-value, based on this distribution.

10. (10 points) What do you conclude?