Homework #5

Due: Tuesday, October 26 @ 6pm

Please remember to give R code, as well as answers, for any problems where you used R

Problem 1:

For this homework assignment, we are all going to try using R markdown, which is not only great for these homework assignments, but is a great skill for doing analyses in R. For more on R markdown, check out any of the numerous websites online, including this tutorial: https://ourcodingclub.github.io/tutorials/rmarkdown/

a. An R markdown is a special type of R script that includes both plain text and code, which makes it a perfect choice for writing up a report or a page for a lab notebook! To create an R markdown, select `File` > `New File` > `R markdown`. A popup window will ask you to name the document and choose the output type. For now, let's stick with HTML output (default) but note you could also choose to output the document as a PDF (if you first install latex to your machine). Name the document and press "OK" to create the report.
b. Take a look at the file that is generated. It is written in "markdown syntax" with normal text, *italics*, **bold**, and headers (i.e. "problem 1"). Most importantly, you also have "code chunks"(see below). Any code you type in these chunks will be evaluated and the results printed.

````
```{r}
plot(pressure)
```
````

c. If you are using RStudio 1.4 or newer, you can also view/write R Markdowns in the "visual markdown editor" (i.e. you don't really have to know markdown syntax anymore!). To switch to this mode, click the "visual markdown editor" button on the upper right of the script panel (looks like an A). This will more or less show you what the report will look like once generated.
d. Without changing anything, press the "knit" button at the top. An HTML document should be generated showing text, code, and figures.
e. Play around with adding text or code chunks in the test markdown and when you feel comfortable, create a new markdown to use for this homework assignment. **You will get full credit on this problem if your HW5 is turned in as an HTML or PDF markdown output.**
f. *Note: to make things easier on your graders (which, really is in your best interest!) consider starting each problem with a HEADER (i.e. `# Problem 1`) and maybe even using sub-headers for sub-parts of problems (i.e. `## 1a`)*

Problem 2:

Consider the following two strains of mice:

Strain1	42.6	57.89	45.22	45.4	40.01	48.64	48.42	46.86	49.47	52.14	46.11	43.53
Strain2	46.2	54.12	52.48	46.97	65.89	57.41	59.07	51.07	56.24			

a. Write down the t statistic for testing the hypothesis that the population means are equal. (Show your work, do not just simply give a number!)
b. Using the value you obtained in (a) and the `rt/qt/pt/dt` functions (not `t.test`!), test the hypothesis $H_0 : \mu_1 = \mu_2$ against the alternative $H_A : \mu_1 \neq \mu_2$. Compute the p-value and give an interpretation (in words) of the finding.
c. How would you change the above to test $H_A : \mu_1 < \mu_2$? $H_A : \mu_1 > \mu_2$?

d. Now, using R's `t.test()` function, repeat parts (b) and (c)
e. Suppose that the data above could NOT be assumed to be normal. How might you test these hypotheses? Give R code and explain what you are doing and why you chose this test.

Problem 3:

Diastolic blood pressure measurements on American men ages 18-44 years follow approximately a normal curve with $\mu = 81$ mm Hg and $\sigma = 11$ mm Hg. The distribution for women ages 18-44 is also approximately normal with the same SD but with a lower mean: $\mu = 75$ mm Hg. Suppose we are going to measure the diastolic blood pressure of n randomly selected men and n randomly selected women in the age group 18-44 years. Let E be the event that the difference between men and women will be found statistically significant by a t test. How large must n be in order to have $Pr[E] = 0.9$:

a. if we use a two-tailed test at $\alpha = 0.05$?
b. if we use a two-tailed test at $\alpha = 0.01$?
c. if we use a one-tailed test (in the correct direction) at $\alpha = 0.05$?
d. Suppose we only have a sample size of n = 45. How large must the difference between means be in order to have 80% power if we used a two-tailed test at $\alpha = 0.05$?

Problem 4:

A scientist conducted a study of how often her pet parakeet chirps. She recorded the number of distinct chirps the parakeet made in a 30-minute period, sometimes when the room was silent and sometimes when music was playing. Import the data she collected into R with the following code:

```r
url <- "https://raw.githubusercontent.com/katiesevans/IGP_biostatistics/main/data/chirps.tsv"
chirps2 <- read.delim(url)
```

a. Construct a 95% confidence interval for the mean increase in chirps (per 30 minutes) when music is playing over when music is not playing. Be sure to show your work!
b. What conditions are necessary for the confidence interval to be valid? Are those conditions satisfied?
c. Use a t test to test the null hypothesis of no difference in the number of chirps with or without music against the alternative hypothesis that there is an increase in chirps when music is playing.
d. Suppose that the data above could NOT be assumed to be normal. How might you test these hypotheses? Give R code and explain what you are doing and why you chose this test.