Homework 5

Your name and student ID

Today's date

- Due date: Tuesday, March 3 10:00pm.
- Late penalty: 50% late penalty if submitted within 24 hours of due date, no marks for assignments submitted thereafter.
- This assignment is marked out of 29. Marks are indicated for each question.

Remember: This homework does not involve autograder checking!

- Submission process: Follow the submission instructions on the final page. Make sure you do not remove any `\newpage` tags or rename this file, as this will break the submission.

Helpful hints:

- Every function you need to use was taught during lecture! So you may need to revisit the lecture code to help you along by opening the relevant files on Datahub. Alternatively, you may wish to view the code in the condensed PDFs posted on the course website. Good luck!
- Knit your file early and often to minimize knitting errors! If you copy and paste code for the slides, you are bound to get an error that is hard to diagnose. Typing out the code is the way to smooth knitting! We recommend knitting your file each time after you write a few sentences/add a new code chunk, so you can detect the source of knitting errors more easily. This will save you and the GSIs from frustration! **You must knit right before submitting**
- If your code runs off the page of the knitted PDF then you will LOSE POINTS! To avoid this, have a look at your knitted PDF and ensure all the code fits in the file. When it doesn't, go back to your .Rmd file and add spaces (new lines) using the return or enter key so that the code runs onto the next line.

[7 points] Part 1: Pregnancy Length Probabilities

An average pregnancy for humans lasts 266 days, with a standard deviation of 16 days. Assume that human pregnancies are Normally distributed.

1. [3 marks] Approximately what proportion of births are expected to occur on or before 298 days? To aid your answer, hand-draw (or use any software) to sketch a Normal curve, and shade in the area under the Normal density curve the question represents. Add dashed lines at the mean +/- 1SD, 2SD and 3SD. Then calculate the proportion asked about in the first sentence. You shouldn't need to use R to perform any calculations for this question. Report the proportion to one decimal place.

(Use the code chunk below to include an image file of your drawing. To do so you need to delete the hashtask, upload the image to Datahub into the `src` directory and replace the file name with your file name. JPG or PNG will both work.)

```
#Take off the '#' in the following code and replace the code with you file name
#knitr::include_graphics("src/Your-file-name.JPG")
```

2. [1 mark] Check your answer from part a) using R code. Store your answer and the absolute difference between your answer and the exact probability to the vector `p2`

```
p2 <- "Your answer here"
p2
```

```
## [1] "Your answer here"
```

3. [3 marks] What is the range, in days, that the middle 50% of pregnancies last? To aid your answer, hand-draw (or use any software) to sketch a Normal curve, and shade in the area under the Normal density curve that the middle represents. Then, use R to calculate the requested range. Round the lower and upper bound of the range each to two decimal places.

(Use the code chunk below to include an image file of your drawing. To do so you need to delete the hashtag, upload the image to Datahub into the `src` directory and replace the file name with your file name. JPG or PNG will both work.)

```
#Take off the '#' in the following code and replace the code with you file name
#knitr::include_graphics("src/Your-file-name.JPG")

#Your code for range here
```

<TODO: YOUR ANSWER HERE>

[7 points] Part 2: Assessing Normality and Interpreting QQ Plots

The number of trees for nine plots of land of 0.1 hectare each have been recorded. They are: 18, 4, 22, 15, 18, 19, 22, 12, 12. Are these data Normally distributed?

4. [3 marks] Make a Normal quantile plot for these data using R. Remember, to make a ggplot of these data, you need to first input the data as a vector and then convert that vector to a data frame. Example code has been provided to you to get you started. After making the plot, assess whether the data appear to approximately follow a Normal distribution.

```
library(tidyverse)
# example code
counts <- c(1, 2, 3)
tree_data <- data.frame(counts)

# your code here
```

<TODO: YOUR ANSWER HERE>

5. [4 marks] Read this blog post by Sean Kross (up to and including the Takeaways), no need to read the Updates, or to understand the code Sean is using. It is different from the code we've been learning in class. Pay most attention to the presentation of the Quantile-Quantile plots for all the distributions he presents. Important note: Sean is plotting "Q-Q" plots and we've been plotting Normal quantile plots. Q-Q plots are a little different, but the same takeaways apply, meaning that if you understand how to interpret Q-Q plots, you can also apply those interpretations to Normal quantile plots.

Look at the charts entitled "Skewed right" and "Skewed left" and the Quantile-Quantile plots beside them. Why does the Quantile-Quantile plot for the skewed right plot curve upwards and to the right (i.e., above the line), while the Quantile-Quantile plot for the skewed left plot curve downwards and to the left?

<TODO: YOUR ANSWER HERE>

[15 points] Part 3: Conducting a general anxiety disorder study

Suppose that a new treatment for general anxiety disorder has undergone safety and efficacy trials and based on these data 30% of patients with general anxiety disorder are expected to benefit from the new treatment. You are conducting a follow-up study and so far have enrolled 8 participants with general anxiety disorder into your study. These patients do not know each other and represent individuals who responded to a call for study participants that they saw on a flyer on campus.

6. [2 marks] Let X represent the number of patients that you have enrolled who benefit from the treatment. Does X meet the assumptions of a Binomial distribution? Thoroughly explain why or why not.

<TODO: YOUR ANSWER HERE>

7. [1 mark] Using one of the distributions learnt in class that X meets the assumptions of, calculate by hand the probability that exactly 5 participants will benefit. Show your work.

<TODO: YOUR ANSWER HERE>

8. [1 mark] Confirm your previous calculation using an R function and store your answer to p8.

```
p8 <- "Your answer here"
p8
```

```
## [1] "Your answer here"
```

9. [2 marks] Calculate by hand the probability that 6 or more participants will benefit. Show your work.
<TODO: YOUR ANSWER HERE>

10. [1 mark] Confirm your previous calculation using code that depends on `pbinom()`. Store your answer to p10.

```
p10 <- "Your answer here"
p10
```

```
## [1] "Your answer here"
```

11. [1 mark] Re-confirm your previous calculation, this time using code that depends on `dbinom()`. Store your answer to p11.

```
p11 <- "your answer here"
p11
```

```
## [1] "your answer here"
```

12. [2 marks] Interpret the binomial coefficient, $\binom{8}{7}$, in the context of this study. Write out all the possible combinations to achieve $\binom{8}{7}$.

<TODO: YOUR ANSWER HERE>

13. [4 marks] Calculate the number of patients you would expect to benefit from the treatment and the standard deviation. Write a sentence to interpret the meaning of the mean. If the mean is not a whole number, what whole number is most probable?

<TODO: YOUR ANSWER HERE>

14. [1 mark] Should you apply a Normal approximation to these data using the μ and σ you calculated in the last question? Why or why not?

<TODO: YOUR ANSWER HERE>

Check your score

Submission

For assignments in this class, you'll be submitting using the **Terminal** tab in the pane below. In order for the submission to work properly, make sure that:

1. Any image files you add that are needed to knit the file are in the `src` folder and file paths are specified accordingly.
2. You **have not changed the file name** of the assignment.
3. The file is saved (the file name in the tab should be **black**, not red with an asterisk).
4. The file knits properly.

Once you have checked these items, you can proceed to submit your assignment.

1. Click on the **Terminal** tab in the pane below.
2. Copy-paste the following line of code into the terminal and press enter.

cd; cd ph142-sp20/hw/hw05; python3 turn_in.py

3. Follow the prompts to enter your Gradescope username and password. When entering your password, you won't see anything come up on the screen–don't worry! This is just for security purposes–just keep typing and hit enter.
4. If the submission is successful, you should see "Submission successful!" appear as output.
5. If the submission fails, try to diagnose the issue using the error messages–if you have problems, post on Piazza.

The late policy will be strictly enforced, **no matter the reason**, including submission issues, so be sure to submit early enough to have time to diagnose issues if problems arise.