Homework 10

Your name and student ID

Today's date

- Due date: April 28th 10pm (make sure to provide enough time for Gradescope submission to be uploaded if you include large image files).
- Late penalty: 50% late penalty if submitted within 24 hours of due date, no marks for assignments submitted thereafter.
- Submission process: Please submit a PDF of your assignment to Gradescope. You must tell Gradescope which questions are on which pages. If you can't see it properly on Gradescope, open the PDF in a PDF viewer at the same time so you can make the selections accurately, otherwise points will be deducted since this makes grading much less efficient.

Helpful hint:

- Knit your file early and often to minimize knitting errors! If you copy and paste code, you are bound to get an error that is hard to diagnose. Hand-writing 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!

- 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.

You would like to conduct a survey of highschool students to determine the proportion who are current e-cigarettes users. Before you conduct your survey, you need to determine how large of a sample size. Suppose that you would like the width of the 95% confidence interval to be 5 percentage points.

1. [1 point] Determine the most conservative sample size you would require and assign it to object p1. Recall that to do this, you need to use a p^* of 0.5.

```r
p1 <- "YOUR ANSWER HERE" #remember to remove " " if you want to store a number

check_problem1()
```

```
## [1] "Checkpoint 1 Error: p1 should be numeric"
## [1] "Checkpoint 2 Error: Wrong answer"
##
## Problem 1
## Checkpoints Passed: 0
## Checkpoints Errored: 2
## 0% passed
## --------
## Test: FAILED
```

2. [1 point] You've seen a recent publication from the Annals of Internal Medicine that estimated that 9.2% of individuals aged 18 to 24 years old are current e-cigarette users. What is the sample size estimate assuming that $p^* = 0.092$.

```r
p2 <- "YOUR ANSWER HERE"

check_problem2()
```

```
## [1] "Checkpoint 1 Error: p2 should be numeric"
## [1] "Checkpoint 2 Error: Wrong answer"
##
## Problem 2
## Checkpoints Passed: 0
## Checkpoints Errored: 2
## 0% passed
## --------
## Test: FAILED
```

3. [1 point] The recent publication referenced in the previous question only looked at adults (aged 18+), but observed that the rate of current use was inversely related to age among the population they surveyed. Because of this finding would you suppose that the sample size estimated in part (b) is too low or too high?

```r
# Uncomment one of the options

#p3 <- "too low"
#p3 <- "too high"

check_problem3()
```

```
## [1] "Checkpoint 1 Error: Did you uncomment your answer?"
## [1] "Checkpoint 2 Error: Wrong answer"
##
## Problem 3
## Checkpoints Passed: 0
## Checkpoints Errored: 2
## 0% passed
## --------
## Test: FAILED
```

Exclusive breastfeeding during the first six months of life is recommended for optimal infant growth and development. Suppose that you conducted a survey of randomly chosen women from California and found that 775 out of 5615 new mothers exclusively breast fed their infants.

[4 points]. Perform all four of the methods discussed in lecture and during lab to create a 95% confidence interval for the proportion of infants exclusively breast fed.

```
library(tidyverse)
library(tibble)
```

Store your upper and lower bounds as unrounded values (i.e. to 7 decimal places) and as a proportion bounded between [0,1]. Store your answer to p4-p7 using the following format:

```
example <- c(lowerbound, upperbound)
```

4. [1 point] Use the large sample method of constructing a 95% CI.

```
p4 <- "YOUR ANSWER HERE"

check_problem4()
```

```
## [1] "Checkpoint 1 Error: Did you store your CI as a numeric object?"
## [1] "Checkpoint 2 Error: Did you store 2 elements in your object?"
## [1] "Checkpoint 3 Error: Wrong answer"
##
## Problem 4
## Checkpoints Passed: 0
## Checkpoints Errored: 3
## 0% passed
## --------
## Test: FAILED
```

5. [1 point] Use the Clopper Pearson (Exact) method of constructing a 95% CI.

```
p5 <- "YOUR ANSWER HERE"

check_problem5()
```

```
## [1] "Checkpoint 1 Error: Did you store your CI as a numeric object?"
## [1] "Checkpoint 2 Error: Did you store 2 elements in your object?"
## [1] "Checkpoint 3 Error: Wrong answer"
##
## Problem 5
## Checkpoints Passed: 0
## Checkpoints Errored: 3
## 0% passed
## --------
## Test: FAILED
```

6. [1 point] Use the Wilson Score method of constructing a 95% CI with a continuity correction.

```r
p6 <- "YOUR ANSWER HERE"

check_problem6()
```

```
## [1] "Checkpoint 1 Error: Did you store your CI as a numeric object?"
## [1] "Checkpoint 2 Error: Did you store 2 elements in your object?"
## [1] "Checkpoint 3 Error: Wrong answer"
##
## Problem 6
## Checkpoints Passed: 0
## Checkpoints Errored: 3
## 0% passed
## --------
## Test: FAILED
```

7. [1 point] Use the Plus Four method of constructing a 95% CI.

```r
p7 <- "YOUR ANSWER HERE"

check_problem7()
```

```
## [1] "Checkpoint 1 Error: Did you store your CI as a numeric object?"
## [1] "Checkpoint 2 Error: Did you store 2 elements in your object?"
## [1] "Checkpoint 3 Error: Wrong answer"
##
## Problem 7
## Checkpoints Passed: 0
## Checkpoints Errored: 3
## 0% passed
## --------
## Test: FAILED
```

8) [2 points] Create a plot comparing the confidence intervals, *with the scale changed to 100* (i.e. 0.50 becomes 50). If you are stuck, refer back to the example code presented in lab 10.

```
# Create a tibble using the code template below. Make sure you change the scale of the CI bounds and es
breastfeed_CIs <- tibble(method   = c("large sample", "Exact", "Wilson", "Plus 4"),
                  lower_CI = c(0.0          , 0.0    ,  0.0    ,  0.0),
                  upper_CI = c(0.0          , 0.0    ,  0.0    ,  0.0),
                  estimate = c(0.0          , 0.0    ,  0.0    ,  0.0)
                  )
p8 <- "YOUR ANSWER HERE"

check_problem8()
```

```
## [1] "Checkpoint 1 Error: You did not define a ggplot."
## [1] "Checkpoint 2 Error: Did not use breastfeed_CIs as the data"
## [1] "Checkpoint 3 Error: Did you specify the data correctly?"
## [1] "Checkpoint 4 Error: Did you specify the data correctly?"
## [1] "Checkpoint 5 Error: Did you plot the point estimates?"
## [1] "Checkpoint 6 Error: Did you plot the CI segments?"
##
## Problem 8
## Checkpoints Passed: 0
## Checkpoints Errored: 6
## 0% passed
## --------
## Test: FAILED
```

9. [1 point] Do the methods produce confidence intervals that are basically the same or very different? Why?

<YOUR ANSWER HERE (remove the brackets).>

10) [1 point] Suppose that in 2010, the rate of exclusive breastfeeding in California was known to be 18.6%. Based on the 95% CIs calculated in questions 4-7, is there evidence against the null hypothesis that the underlying rate is equal to 18.6% in favor of the alternative that the rate is different from 18.6%?

<YOUR ANSWER HERE (remove the brackets).>

[4 points] To confirm your answer to question 9, perform a two-sided hypothesis test and interpret the p-value.

11) [1 point] State the null and alternative hypotheses:

<YOUR ANSWER HERE (remove the brackets).>

12) [1 point] Calculate the test statistic:

```r
p12 <- "YOUR ANSWER HERE"

check_problem12()
```

```
## [1] "Checkpoint 1 Error: p12 should be numeric"
## [1] "Checkpoint 2 Error: Wrong answer"
##
## Problem 12
## Checkpoints Passed: 0
## Checkpoints Errored: 2
## 0% passed
## --------
## Test: FAILED
```

13) [1 point] Calculate the p-value:

```r
p13 <- "YOUR ANSWER HERE"

check_problem13()
```

```
## [1] "Checkpoint 1 Error: p13 should be numeric"
## [1] "Checkpoint 2 Error: Wrong answer"
##
## Problem 13
## Checkpoints Passed: 0
## Checkpoints Errored: 2
## 0% passed
## --------
## Test: FAILED
```

14) [1 point] Interpret the p-value:

<YOUR ANSWER HERE (remove the brackets).>

[2 points total]. The quadrivalent HPV vaccine was introduced to Canada in 2007, and was given to girls in Ontario, Canada who were enrolled in grade 8 (13-14 year olds). Before 2007, no girls recieved the vaccine, while in the 4 years after it was introduced nearly 40% of girls recieved the vaccine each year. One concern that some people had was that the vaccine itself would increase promiscuity if the girls felt a false sense of protection, which could thereby increase the prevalence of other sexually transmitted infections (STIs) among vaccinated girls. This paper examines this question using an advanced method called the "regression discontinuity" design which harnesses the abrupt change in vaccination status across cohorts of girls to estimate the causal effect of vaccination against HPV on the occurrence of other STIs.

Read only the abstract of the paper, and don't worry about the details because these are advanced methods. Note that the term "RD" is the difference in risk of STIs between girls exposed and unexposed to HPV vaccination. We can therefore think of this risk difference as the difference between two proportions.

15) [1 point] Interpret this result from the abstract: We identified 15 441 (5.9%) cases of pregnancy and sexually transmitted infection and found no evidence that vaccination increased the risk of this composite outcome: RD per 1000 girls -0.61 (95% confidence interval [CI] -10.71 to 9.49). **In particular, what does -0.61 estimate?**

<YOUR ANSWER HERE (remove the brackets).>

16) [1 point] The 95% confidence interval includes 0. What can you conclude about the p-value for a two-sided test of the difference between vaccinated and unvaccinated girls and their risk of sexually transmitted diseases?

<YOUR ANSWER HERE (remove the brackets).>

Question 4 [7 points total]. An allergy to peanuts is increasingly common in Western countries. A randomized controlled trial enrolled infants with a diagnosed peanut sensitivity. Infants were randomized to either avoid peanuts or to consume them regularly until they reached age 5. At the end of the study, 18 out of the 51 randomized to avoid peanuts were tested to be allergic to peanuts. Only 5 out of the 47 randomized to consuming them regularly were tested to be allergic to peanuts.

17) [1 point] Estimate the difference between the two proportions.

```r
p17 <- "YOUR ANSWER HERE"

check_problem17()
```

```
## [1] "Checkpoint 1 Error: p17 should be numeric"
## [1] "Checkpoint 2 Error: Wrong answer"
##
## Problem 17
## Checkpoints Passed: 0
## Checkpoints Errored: 2
## 0% passed
## --------
## Test: FAILED
```

18) [1 point] Use the plus four method to find a 99% confidence interval for the difference between the two groups. Store the upper and lower bounds into an object called p16.

```r
p18 <- "YOUR ANSWER HERE"

check_problem18()
```

```
## [1] "Checkpoint 1 Error: Did you store your CI as a numeric object?"
## [1] "Checkpoint 2 Error: Did you store 2 elements in your object?"
## [1] "Checkpoint 3 Error: Wrong answer"
##
## Problem 18
## Checkpoints Passed: 0
## Checkpoints Errored: 3
## 0% passed
## --------
## Test: FAILED
```

19) [1 point] Why would it have been inappropriate to use the large sample method to create a 99% CI?

<YOUR ANSWER HERE (remove the brackets).>

[4 points] Perform a two-sided hypothesis test for the difference between the groups. Start by stating the null and alternative hypotheses, then calculate the test statistic, the p-value, and conclude with your interpretation of the p-value.

20) [1 point] State the null and alternative hypotheses:

<YOUR ANSWER HERE (remove the brackets).>

21) [1 point] Calculate the test statistic:

```r
p21 <- "YOUR ANSWER HERE"

check_problem21()
```

```
## [1] "Checkpoint 1 Error: p21 should be numeric"
## [1] "Checkpoint 2 Error: Wrong answer"
##
## Problem 21
## Checkpoints Passed: 0
## Checkpoints Errored: 2
## 0% passed
## --------
## Test: FAILED
```

22) [1 point] Calculate the p-value:

```r
p22 <- "YOUR ANSWER HERE"

check_problem22()
```

```
## [1] "Checkpoint 1 Error: p22 should be numeric"
## [1] "Checkpoint 2 Error: Wrong answer"
##
## Problem 22
## Checkpoints Passed: 0
## Checkpoints Errored: 2
## 0% passed
## --------
## Test: FAILED
```

23) [1 point] Interpret the p-value:

<YOUR ANSWER HERE (remove the brackets).>

24) [1 point] Suppose you were testing the hypotheses $H_0 : \mu_d = 0$ and $H_a : \mu_d \neq 0$ in a paired design and obtain a p-value of 0.21. Which one of the following could be a possible 95% confidence interval for μ_d?

a. -2.30 to -0.70
b. -1.20 to 0.90
c. 1.50 to 3.80
d. 4.50 to 6.90

```r
# Uncomment one answer only
# p24 <- "a"
# p24 <- "b"
# p24 <- "c"
# p24 <- "d"

check_problem24()
```

```
## [1] "Checkpoint 1 Error: Did you uncomment your answer?"
## [1] "Checkpoint 2 Error: Wrong answer"
##
## Problem 24
## Checkpoints Passed: 0
## Checkpoints Errored: 2
## 0% passed
## --------
## Test: FAILED
```

25) [1 point] Suppose you were testing the hypotheses $H_0 : \mu_d = 0$ and $H_a : \mu_d \neq 0$ in a paired design and obtain a p-value of 0.02. Also suppose you computed confidence intervals for μ_d. Based on the p-value which of the following are true?

a. Both a 95% CI and a 99% CI will contain 0.
b. A 95% CI will contain 0, but a 99% CI will not.
c. A 95% CI will not contain 0, but a 99% CI will.
d. Neither a 95% CI nor a 99% CI interval will contain 0.

```r
# Uncomment one answer only

# p25 <- "a"
# p25 <- "b"
# p25 <- "c"
# p25 <- "d"

check_problem25()
```

```
## [1] "Checkpoint 1 Error: Did you uncomment your answer?"
## [1] "Checkpoint 2 Error: Wrong answer"
##
## Problem 25
## Checkpoints Passed: 0
## Checkpoints Errored: 2
## 0% passed
## --------
## Test: FAILED
```

Check your score

Click on the middle icon on the top right of this code chunk (with the downwards gray arrow and green bar) to run all your code in order. Then, run this chunk to check your score.

```
# Just run this chunk.
total_score()
```

```
##                       Test Points_Possible          Type
## Problem 1           FAILED               1    autograded
## Problem 2           FAILED               1    autograded
## Problem 3           FAILED               1    autograded
## Problem 4           FAILED               1    autograded
## Problem 5           FAILED               1    autograded
## Problem 6           FAILED               1    autograded
## Problem 7           FAILED               1    autograded
## Problem 8           FAILED               2    autograded
## Problem 9  NOT YET GRADED               1 free-response
## Problem 10 NOT YET GRADED               1 free-response
## Problem 11 NOT YET GRADED               1 free-response
## Problem 12          FAILED               1    autograded
## Problem 13          FAILED               1    autograded
## Problem 14 NOT YET GRADED               1 free-response
## Problem 15 NOT YET GRADED               1 free-response
## Problem 16 NOT YET GRADED               1 free-response
## Problem 17          FAILED               1    autograded
## Problem 18          FAILED               1    autograded
## Problem 19 NOT YET GRADED               1 free-response
## Problem 20 NOT YET GRADED               1 free-response
## Problem 21          FAILED               1    autograded
## Problem 22          FAILED               1    autograded
## Problem 23 NOT YET GRADED               1 free-response
## Problem 24          FAILED               1    autograded
## Problem 25          FAILED               1    autograded
```

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/hw10; 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.