Homework 4

The purpose of this homework is to gain a better understanding of how to run parametric hypothesis tests and to see the connections between different types of hypothesis tests. Please fill in the appropriate code and write answers to all questions in the answer sections, then submit a compiled pdf with your answers through Gradescope by 11pm on Sunday October 1st.

As always, if you need help with the homework, please attend the TA office hours which are listed on Canvas and/or ask questions on Ed Discussions. Also, if you have completed the homework, please help others out by answering questions on Ed Discussions, which will count toward your class participation grade.

Parts 1 and 2: Examining discrimination in hiring continued

On homework 3 you used a **randomization hypothesis test**, to investigate whether resumes that had names that were perceived to be White had callbacks at a higher rate than applicants that had names that were perceived to be Black.

On parts 1 and 2 of this homework, you will compare **randomization** and **parametric** hypothesis tests to investigate whether resumes that had names that are perceived to be female receive callbacks at the same rate as applicants that are perceived to be male.

To do this analysis, we will again use the data collected by Bertrand and Mullainathan in their study in the American Economic Review. If you have forgotten the design of the study, please refer back to homework 3.

Part 1: Challenge question:

All components of part 1 together comprise a "challenge problem" that you should try to figure out without getting help from the TAs.

To begin comparing randomized and parametric tests, we must first run a randomization hypothesis test that examines whether the callback rate for resumes that had stereotypical female names was the same as callback rate for resumes that had stereotypical male names. This problem is very similar to problem 1 from homework 3 so if you get stuck please refer back to homework 3. Also, since this analysis is similar to what you did on homework 3, you should complete this problem without help from the TAs (i.e., it's a "challenge problem").

In order to have more consistent answers from different students, (to make the grading easier), we have outlined the steps you should take, names of key objects you should use, and values you should print out. In particular, it is important to use the same object names discussed since we will refer to these objects in subsequent parts of the homework.

Part 1.1: State the null and alternative hypothesis (2 points)

Please state the null and alternative hypotheses in words and symbols, along with the significance level we typical use.

Note: we are testing whether there is **a difference** in the callback rate for resumes that had stereotypical female names compared to resumes that had stereotypical male names, rather than testing whether one of these rates is higher than the other. Please be sure to state your hypotheses, and run your analyses, appropriately.

In words

Null hypothesis: There is no difference in callback rates between resumes with stereotypical female names and resumes with stereotypical male names.

Alternative hypothesis: There is a difference in callback rates between resumes with stereotypical female names and resumes with stereotypical male names

In symbols

$H_0 : \pi_{female} - \pi_{male} = 0$

$H_A : \pi_{female} - \pi_{male} \neq 0$

$\alpha = 0.05$

Part 1.2: Calculate the observed statistic (2 points)

The observed statistic is the difference in the callback rates for resumes with stereotypical female names and the callback rate for resumes with stereotypical male names. Please save the difference in these callback rates to an object called `obs_stat`. Print out this rate, along with the female and male callback rates, in the R chunk below.

```
load("resume.rda")

# Callbacks filtered by female and male names
female_callback <- resume$received_callback[resume$gender == "f"]
male_callback <- resume$received_callback[resume$gender == "m"]

# Female callback rate
(female_rate <- mean(female_callback))
```

```
## [1] 0.08248799
```

```
# Male callback rate
(male_rate <- mean(male_callback))
```

```
## [1] 0.07384342
```

```
# Difference in callback rates
(obs_stat <- female_rate - male_rate)
```

```
## [1] 0.008644571
```

Part 1.3: Create the null distribution (3 points)

Please use the `rbinom()` functions to randomly create an (approximate) null distribution that has 10,000 points in it. Save this null distribution in an object called `null_dist`. Also, plot the null distribution as a histogram and put a red vertical lines at the value of the observed statistic and at the value of -1 times the observed statistic.

```
overall_rate <- mean(resume$received_callback)

size_female <- length(female_callback)
size_male <- length(male_callback)

null_female_rates <- rbinom(1e4, size_female, overall_rate) / size_female
null_male_rates <- rbinom(1e4, size_male, overall_rate) / size_male
null_dist <- null_female_rates - null_male_rates

hist(null_dist, xlab = "Callback rates",
     main = "Distribution of Callback Rates for Null Hypothesis")
abline(v = obs_stat, col = "red")
abline(v = -1 * obs_stat, col = "red")
```


Distribution of Callback Rates for Null Hypothesis
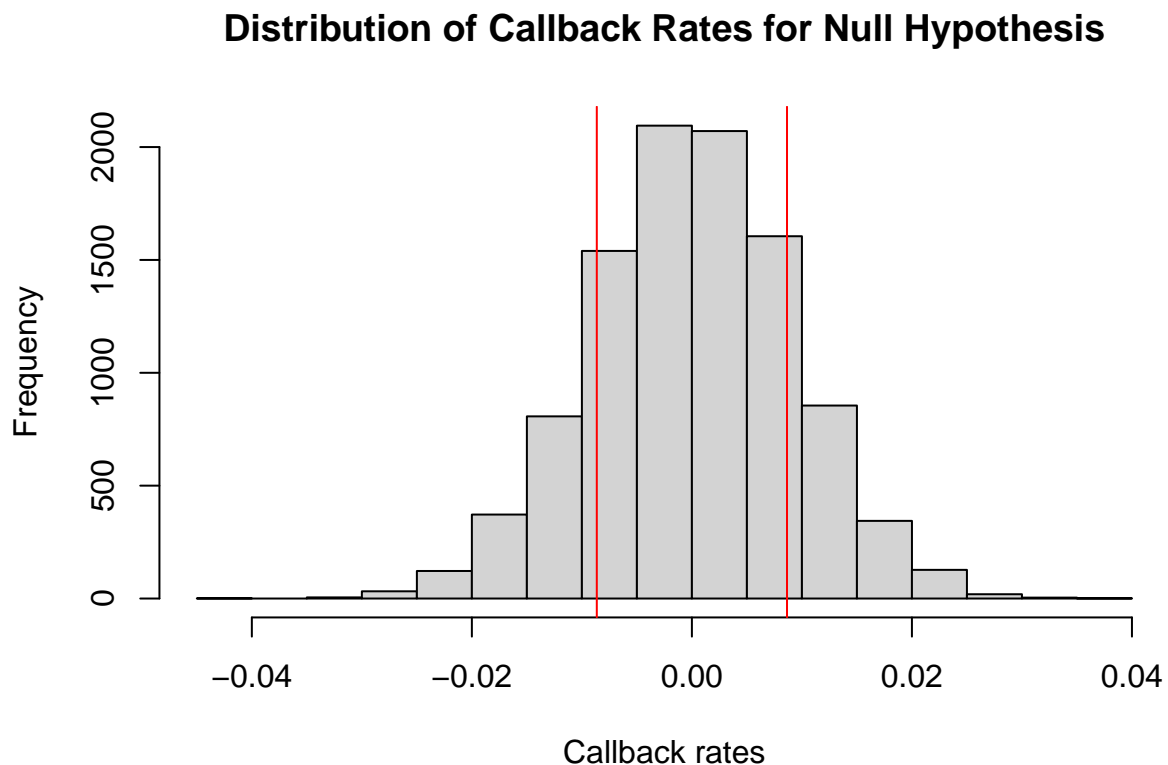
Part 1.4: Calculate the p-value (2 point)

3

Now calculate the p-value and print it below.

```
pval_right <- sum(null_dist >= obs_stat) / length(null_dist)

pval_left <- sum(null_dist <= -1 * obs_stat) / length(null_dist)

(pval <- pval_left + pval_right)
```

```
## [1] 0.3389
```

Part 1.5: Conclusion (1 point)

Based on the p-value you calculated, describe whether there is evidence to reject the null hypothesis and consequently what you would conclude.

Answer

Since the p-value 0.3389 is greater than $\alpha = 0.05$ there is not evidence to reject the null hypothesis. Thus there is not sufficient evidence to say that there is a difference in callback rates between resumes with stereotypical female names and resumes with stereotypical male names.

Part 2: Comparing randomization tests to parametric methods

Now let's continue to analyze the same questions as in part 1, by rerunning our analyses using a parametric hypothesis test. To do this we will again use the same 5 steps of hypothesis testing, but we will use a mathematical density function for the null distribution rather than creating the null distribution by generating random data.

Part 2.1: State the null and alternative hypothesis (1 point)

In order to consistently go through the steps of hypothesis testing, please start with step 1 by restating the null and alternative hypotheses in symbols here (you can skip writing them in words). Please also state again the significance level required for us to reject the null hypothesis.

$H_0 : \pi_{female} - \pi_{male} = 0$

$H_A : \pi_{female} - \pi_{male} \neq 0$

$\alpha = 0.05$

Part 2.2: Calculate the observed statistic (5 points)

We can use the same statistic as in part 1 (i.e., a difference in proportions) to run a parametric hypothesis test. The mathematical theory tells us that the difference of proportions should come from a normal distribution. As you recall, a normal distribution has two parameters μ and σ. The mathematical theory also tells us that if the null distribution is true (which we always assume when we are running a hypothesis test, and attempt to "disprove" through testing), then:

1. The value for μ is equal to the value specified in the null hypothesis
2. The standard error of our null distribution is given by the formula:

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}$$

Where: * $\hat{p}$ is the overall proportion of resumes that received callbacks regardless of whether resume had a stereotypical male or female name. * n_1: Is the number of resumes with stereotypical male names. * n_2: Is the number of resumes with stereotypical female names.

A tip: If you have trouble reading formulas in raw LaTeX notation, it may be helpful to first knit the homework and view the PDF output of the same formula to make sure you understand the equation.

Since we have already calculated the observed statistic in part 1.2 above, let's use the R chunk below to calculate the theoretical standard error value for the null distribution, save it to an object called `SE_prop`, and print out this value.

Side note: An alternative way to run this analysis would be to use a z-statistic statistic which is the difference in proportions divided by the SE. You likely have seen this type of analysis in the Introductory Statistics class you took (a test that uses this statistic goes by the name of a z-test). This type of z-statistic comes from a standard normal distribution; i.e., a normal distribution with a mean of 0 and a standard deviation of 1. The reason we are not doing this here is because we want to compare our parametric null distribution to the randomization null distribution we created in part 1.4. Another alternative would have been to use a z-statistic in part 1 of this homework by calculating the SE using the formula above and then using it to convert our difference of proportion statistics to a z-statistic by dividing by the SE.

```
(SE_prop <- sqrt(((overall_rate * (1 - overall_rate)) / size_male) +
        ((overall_rate * (1 - overall_rate)) / size_female)))
```

```
## [1] 0.009252382
```

Part 2.3: Create/visualize the null distribution (5 points)

Now let's visualize the parametric null distribution's density function and compare it to the randomization null distribution we created in part 1.3. Please complete the following steps to do this:

1. Create a sequence of x-values that cover the domain of where the density function is above ~0. Hint: the `seq()` function could be useful here.

2. Create a sequence of y-values that correspond to the density values for the x-values you created in step 1. Hint: recall density functions start with the letter `d` and think through what the density function is, and what the parameter values should be based on the information given in part 2.2.

3. Plot a histogram of the null distribution as you did in part 1.3

4. Use the `points()` function to plot the density curve that you created in steps 1 and 2 above. Hint: the `points()` function works the same as the `plot()` function except that it plots on top of an existing plot rather than creating a new plot.

5. In the answer section below, write 1-2 sentences stating whether the parametric density curve we are using for the null distribution appears similar to the null distribution we created by generating random statistics.

5

```
xvals <- seq(-.04, .04, .001)
yvals <- dnorm(xvals, sd = SE_prop)

hist(null_dist, freq = FALSE, xlab = "Null callback rates", main = "Null Distribution for Callback Rates

points(xvals, yvals, type = "l", col = "blue")
```

Null Distribution for Callback Rates
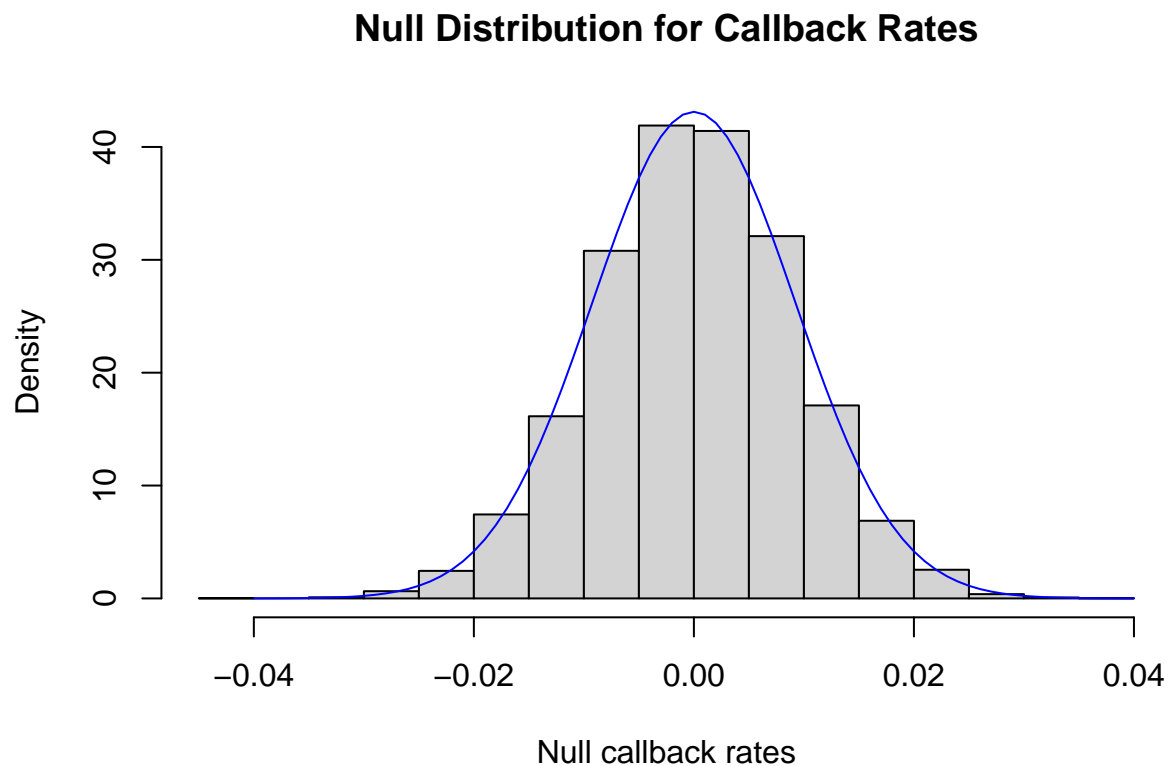
Answer:

The parametric density curve does appear similar to the null distribution created by generating random statistics.

Part 2.4: Calculate the parametric p-value (5 points)

Now calculate the p-value using the appropriate parametric density curve and print out this p-value. In the answer section describe whether this p-value is similar to the p-value you got in part 1.4.

Hint: a function that starts with the letter **p** will be useful here.

```
(pval_right <- pnorm(SE_prop, sd = SE_prop, lower.tail = FALSE))
```

```
## [1] 0.1586553
```

6

```r
(pval_left <- pnorm(-1 * SE_prop, sd = SE_prop))
```

```
## [1] 0.1586553
```

```r
(pval <- pval_right + pval_left)
```

```
## [1] 0.3173105
```

Answer

This p-value (0.3173105) is similar to the p-value I got in part 1.4 (0.3389).

Part 2.5: Conclusions (2 points)

Based on the p-value you calculated in part 2.4, describe whether there is evidence to reject the null hypothesis and consequently what you would conclude. Did the randomization test lead to different results from the parametric test?

Answer

Since the p-value is greater than $\alpha = 0.05$ there is not enough evidence to reject the null hypothesis. Thus there is not sufficient evidence to say that there is a difference in callback rates between resumes with stereotypical female names and resumes with stereotypical male names.

The randomization test did not lead to different results than the parametric test.

Part 3: Analyzing data using t-tests

The term "Freshman 15" refers to the belief that college students frequently gain weight during their freshman year (i.e., it is believed that students gain 15 lbs). To test whether students do in fact gain weight during their first year at college, David Levitsky, a Professor of Nutrition at Cornell College, recruited students from an introductory health course. The students were weighed during the first week of the semester, then again 12 weeks later.

The data from 68 students in Professor Levitsky's class is loaded below and contains the following variables:

- `Subject`: An ID number for each participant who volunteered their weight.
- `Initial.Weight`: The weight of each participant at the start of the semester.
- `Terminal.Weight` The weight of each participant at the end of the semester.

Let's run a few t-tests to see if students do indeed gain any weight (on average) over the course of the semester. Though the term is called "freshman 15," our analysis will investigate if students gain any weight at all.

```r
freshman <- read.table("freshman-15.txt", header = TRUE)
```

Part 3.1 (3 points):

Let's start by running an "independent samples" t-test where we treat the initial weight and final weight as independent samples. Although the final weight and initial weight are not independent, because they are measurements of the same subjects across different points in time, we will first assume they are independent here for practice running independent t-tests and observing the outcome of the "independent samples" t-test.

Please start with step 1 by stating the null and alternative hypothesis in words and using our commonly used symbols.

Answer

Null hypothesis: The average final weight is not greater than the average initial weight.

Alternative hypothesis: The average final weight is greater than the average initial weight.

$H_0 : \mu_{final} \leq \mu_{initial}$

$H_A : \mu_{final} > \mu_{initial}$

Part 3.2 (6 points):

Now let's visualize the data and calculate the statistic of interest. Please use an appropriate plot to compare the students' weights at the start and end of the semester. Also calculate and store the value of the statistic in an object called `t_stat`, and report its value. Recall for Welch's t-test, the t-statistic is defined as:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n1} + \frac{s_2^2}{n2}}}$$

where: - $\bar{x}_1$ and $\bar{x}_2$ are the means of the two groups. - s_1^2 and s_2^2 are the variances of the two groups. - n_1 and n_2 are the samples sizes of the two groups.

Based on the plot, does it seem that students gain weight?

Note: as before, if you have trouble reading the above equation, please knit this document to a pdf, and you should then be able to zoom in on the equation (and also see the class slides).

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n1} + \frac{s_2^2}{n2}}}$$

```
initial <- freshman$Initial.Weight
final <- freshman$Terminal.Weight

boxplot(initial, final, names = c("Initial Weight (lbs)", "Final Weight (lbs)"), ylab = "Pounds", main =
```

Student Weights
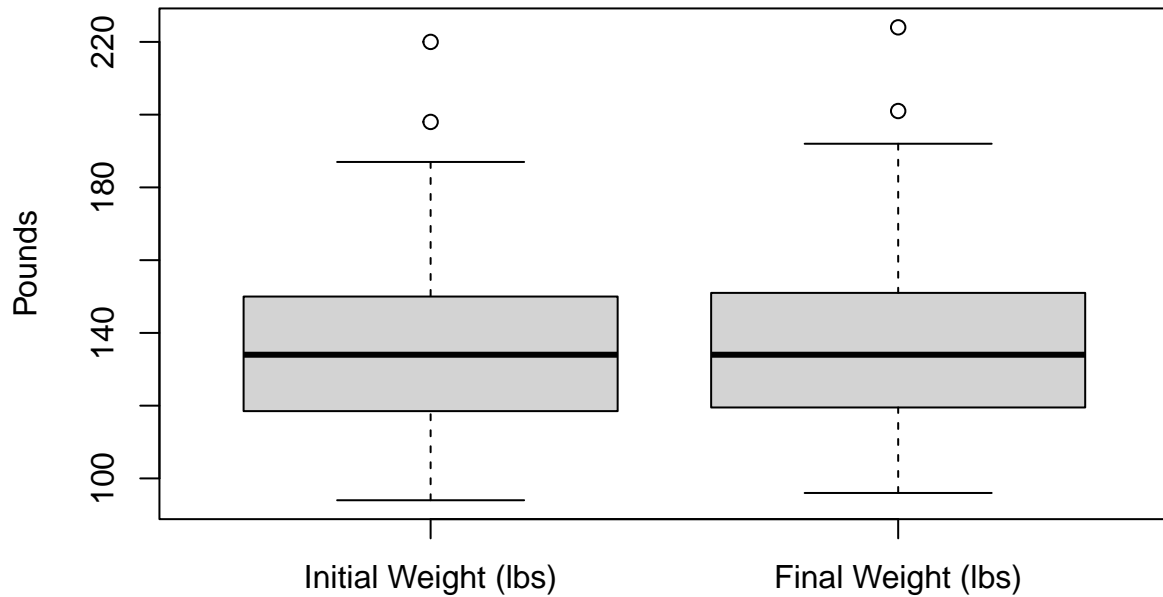

```r
avg_initial <- mean(initial)
avg_final <- mean(final)

var_initial <- var(initial)
var_final <- var(final)

size_initial <- length(initial)
size_final <- length(final)

(t_stat <- (avg_initial - avg_final) / sqrt(((var_initial^2) / size_initial) + ((var_final)^2 / size_fir
```

```
## [1] -0.01858423
```

Answer

Based on the plot, it does not seem that the students gain weight.

Part 3.3 (6 points):

To run Welch's t-test, the following conditions need to be met:

1. The data come from normal populations, or, as a rule of thumb, the sample sizes for the two groups are greater than 30.

2. Like all hypothesis tests we run, we assume that data points within each group are independent (we will ignore any dependence between the two samples for this analysis).

If these conditions are met, the t-statistic should come from a null t-distribution with (conservative) degrees of freedom of the minimum of n_1 - 1 and n_2 - 1.

Please plot this null t-distribution below, and draw a red line at the observed t-statistic value you calculated in part 3.2. Based on the plot you created, does it appear that the results will be statistically significant?

```r
degree_free <- min(length(initial), length(final)) - 1

x_vals <- seq(-5, 5, .01)
y_vals <- dt(x_vals, degree_free)

plot(x_vals, y_vals, type = "l", col = "blue", main = "Null t-distribution")
abline(v = abs(t_stat), col = "red")
```
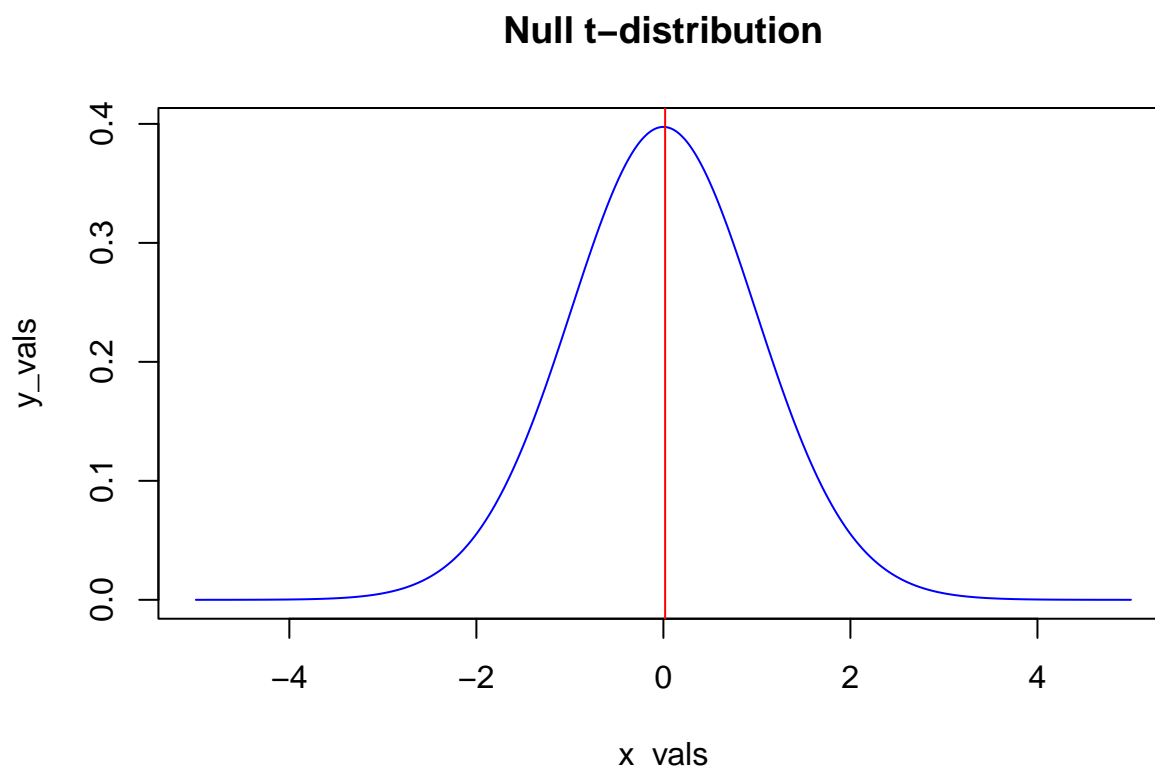

Answer

It does not appear that the results will be statistically significant.

Part 3.4 (3 points):

Now calculate the p-value (hypothesis test step 4). Report below whether the results are statistically significant (hypothesis tests step 5). Also report what you would conclude.

10

```
pvalue_left <- pt(t_stat, degree_free, lower.tail = TRUE)
pvalue_right <- pt(abs(t_stat), degree_free, lower.tail = FALSE)
(pvalue <- pvalue_left + pvalue_right)
```

```
## [1] 0.985228
```

Answer

The results are not statistically significant. Thus there is not sufficient evidence to reject the null hypothesis
and conclude that students gain weight over the course of the semester.

Part 3.5 (3 points):

R has a built in function to do a t-test called `t.test(sample1, sample2)`. Please print out the results from
running this function and report:

1. Do you get the same t-statistic value as you computed above in part 3.2?
2. Do you get the same p-value as you computed above in part 3.4?
3. If you get any different results, what could be causing the difference? Hint: take a look at the
   documentation for the function by running `?t.test` in the R console.

```
t.test(initial, final)
```

```
##
##   Welch Two Sample t-test
##
## data:  initial and final
## t = -0.45517, df = 133.99, p-value = 0.6497
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.218939   6.395409
## sample estimates:
## mean of x mean of y
##   136.0735  137.9853
```

Answer

1. I did not get the same t-statistic.

2. I did not get the same p-value.

3.

Part 4: Paired t-test

As you likely noticed in part 3, the "Freshman 15" data discussed in part 3 is based on repeated measures from the same participants. We can leverage this fact to run a paired samples t-test which is much more powerful (i.e., it has a much higher ability to reject the null hypothesis when it is false). The reason the paired samples t-test is so much more powerful is because there is a lot of variability between different people in the study, so if we can factor that out and only focus on the weight gained by each person, we will have a much better chance of rejecting the null hypothesis if it is false.

The paired samples t-test assesses whether the difference between two measurements on the same observational units is 0 on average (it can also be used to test if the difference is another value apart from 0, but that is less common). To run this test, we first calculate the difference score for each participant between the two measurements. We then calculate the follow t-statistic:

$$t = \frac{\bar{x}_d}{\frac{s_d}{\sqrt{n}}}$$

where: - $\bar{x}_d$ is the mean of the difference scores - s_d is the standard deviation of the difference scores - n is the number of samples (where each sample represents one person and consists of two values – their initial weight and final weight)

Provided that the difference scores are approximately normal (or n is large), and that, as always, the data points are independent, then the t-statistic should come from a t-distribution with n - 1 degrees of freedom.

In this exercise, please use the 5 hypothesis steps to run a paired t-test on the freshman 15 data. As part of your analysis, be sure to print out value of your t-statistic and the p-value you calculate. Also, please solve this problem without using the `t.test()` function (i.e., calculate the t-statistic, use the `pt()` function etc.), however you can use the `t.test()` to check that you have the right answer.

Part 4.1: State the null and alternative hypothesis using the appropriate symbols (3 points)

In words

Null hypothesis: The average student does not gain weight over the course of the semester.

Alternative hypothesis: The average student gains weight over the course of the semester.

In symbols

$H_0 : \mu_d = 0$

$H_A : \mu_d > 0$

Part 4.2: Create an appropriate plot of the data and calculate the statistic of interest (5 points)

You have a few choices of relevant plots although please show one that gives insight into whether the null or alternative hypothesis is true. Be sure to also print out the value of your t-statistic.

$$t = \frac{\bar{x}_d}{\frac{s_d}{\sqrt{n}}}$$

```
difference_score <- final - initial

x_d <- mean(difference_score)
s_d <- sd(difference_score)
n <- length(difference_score)

(tstat <- x_d / (s_d / sqrt(n)))
```
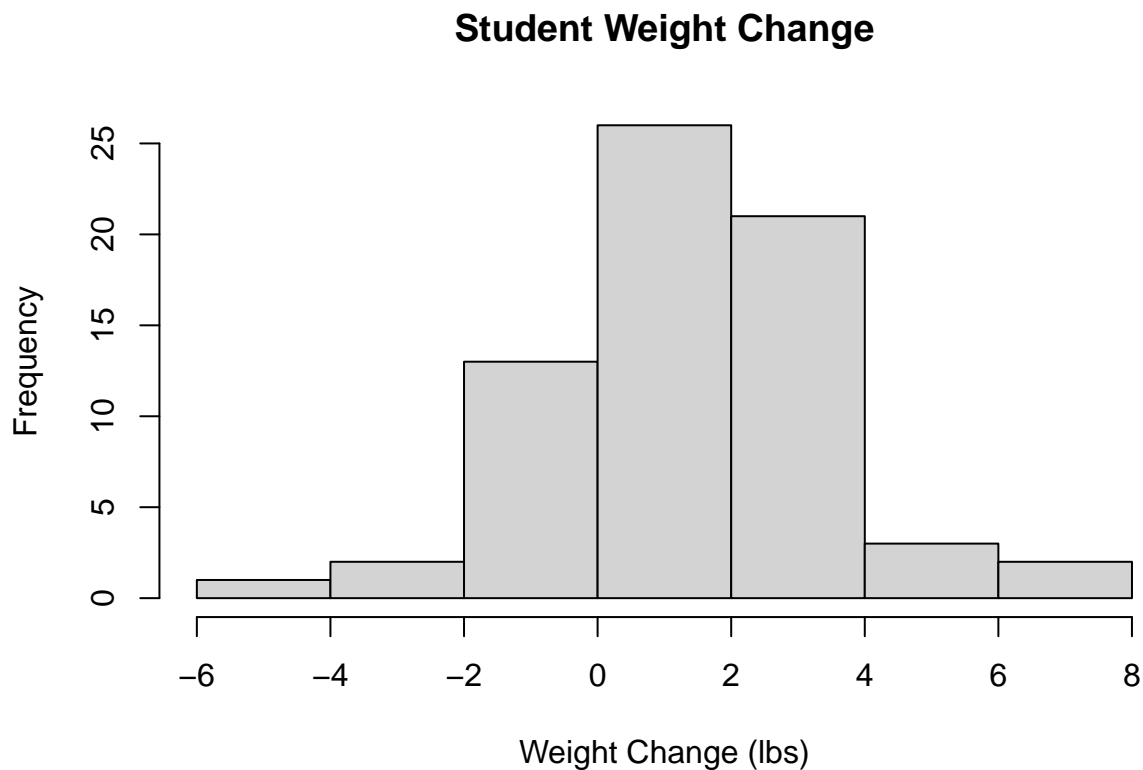
```
## [1] 7.407437
```

```
hist(difference_score, xlab = "Weight Change (lbs)",
            main = "Student Weight Change")
```

Student Weight Change

Part 4.3: Plot the null distribution with the observed statistic as a red line (4 points)

```
null_weight_data <- difference_score - mean(difference_score)

null_distribution <- NULL
for (i in 1:1e4){
  resample_data <- sample(null_weight_data, replace = TRUE)

  numerator_resample <- mean(resample_data)
```
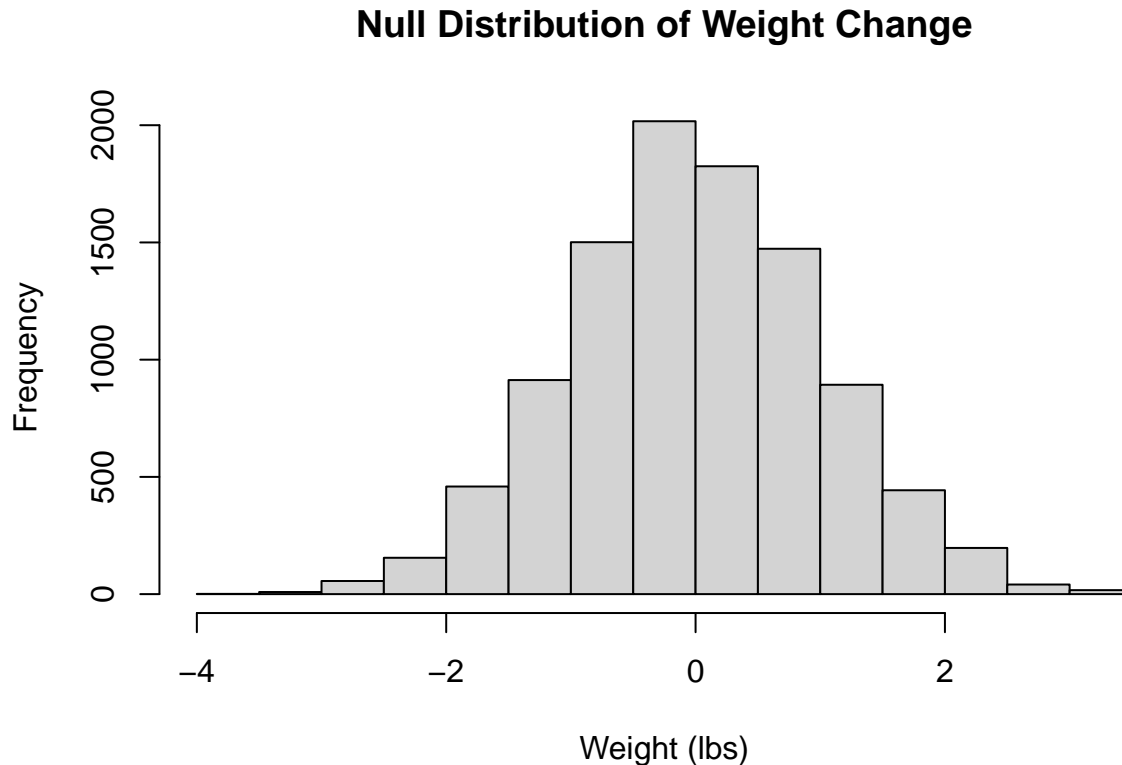
13

```
  denominator_resample <- sd(resample_data) / sqrt(length(resample_data))

  null_distribution[i] <- numerator_resample / denominator_resample
}

hist(null_distribution, xlab = "Weight (lbs)", main = "Null Distribution of Weight Change")

abline(v = tstat, col = "red")
```

Null Distribution of Weight Change

Par 4.4: Calculate the p-value (3 points)

```
(p_value <- 2 * pt(-tstat, df = (length(difference_score) - 1)))
```

```
[1] 0.0000000002812735
```

```
t.test(difference_score)
```

```
##
##   One Sample t-test
##
## data:  difference_score
## t = 7.4074, df = 67, p-value = 0.0000000002813
```

```
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.396621 2.426909
## sample estimates:
## mean of x
##  1.911765
```

Part 4.5: Report the conclusions (2 point)

Based on the p-value there is sufficient evidence to reject the null hypothesis and conclude that on average the students did gain weight over the course of the semester.
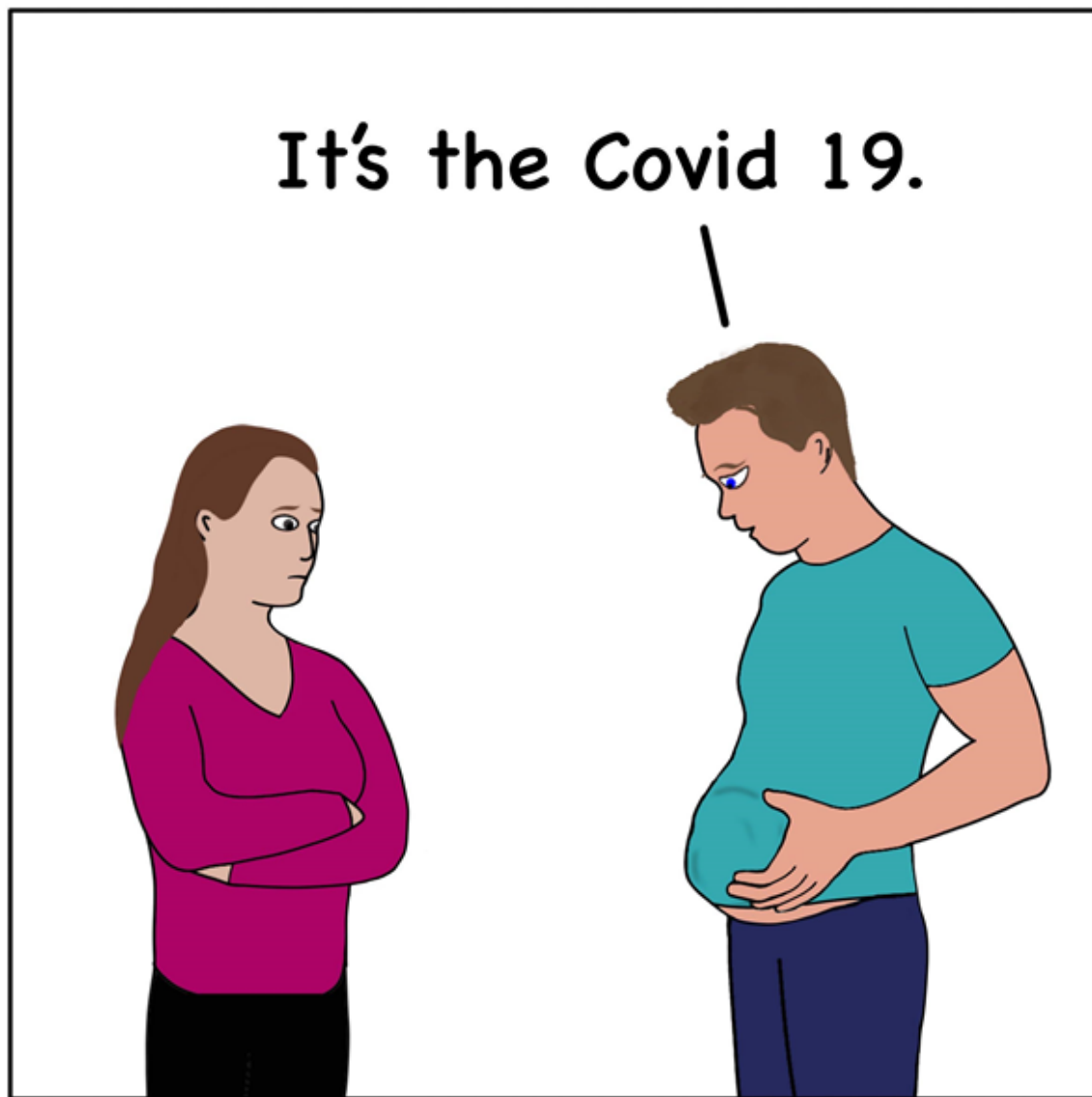
www.abovethefraycomic.com Above the fray ©'20

Reflection (2 points)

Please reflect on how the homework went by going to Canvas, going to the Quizzes link, and clicking on Reflection on homework 4.