Problem Set 10 (Answer Key)

Joe Ornstein

1. Load the data

```
library(tidyverse)
```

```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.3.6     v purrr   0.3.4
v tibble  3.1.8     v dplyr   1.0.9
v tidyr   1.2.0     v stringr 1.4.0
v readr   2.1.2     v forcats 0.5.1
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```
library(here)
```

```
here() starts at C:/Users/jo22058/Documents/intro-political-methodology
```

```
load( here('data/ces-2020/cleaned-CES.RData') )
```

2.

```
# recode as 0-1 variable
ces <- ces |>
  mutate(china_tariffs = as.numeric(china_tariffs == 'Support'))

ces |>
  summarize(pct_support = mean(china_tariffs))
```

```
# A tibble: 1 x 1
  pct_support
        <dbl>
1          NA
```

```
# drop out the people who didn't respond to that question
ces <- ces |>
  filter(!is.na(china_tariffs))

ces |>
  summarize(pct_support = mean(china_tariffs))
```

```
# A tibble: 1 x 1
  pct_support
        <dbl>
1       0.592
```

3. What's the difference in tariff support for men and women?

```
ces |>
  group_by(gender) |>
  summarize(pct_support = mean(china_tariffs))
```

```
# A tibble: 2 x 2
  gender pct_support
  <chr>        <dbl>
1 Female       0.591
2 Male         0.594
```

4. Write a function that samples 100 respondents and computes the difference in means.

```
diff_in_means <- function(sample_size = 100){

  small_ces <- ces |>
    slice_sample(n = sample_size) |>
    mutate(male = as.numeric(gender == 'Male'))


  linear_model <- lm(china_tariffs ~ male,
                     data = small_ces)
```

```
  return(linear_model$coefficients['male'])

}

diff_in_means()
```

```
     male
-0.1103896
```

```
diff_in_means()
```

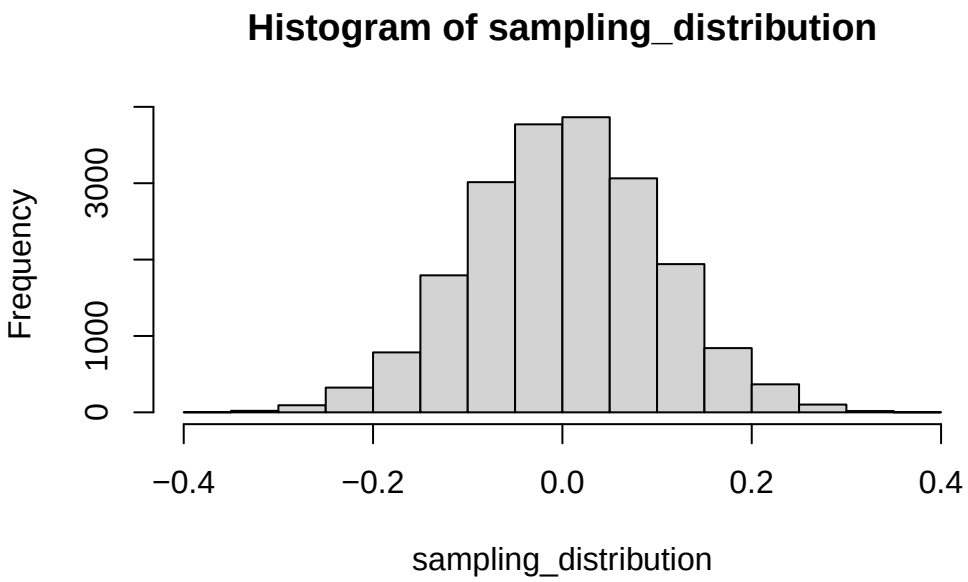
```
     male
0.08484848
```

5. Generate sampling distribution

```
sampling_distribution <- replicate(20000,
                                   diff_in_means())
```

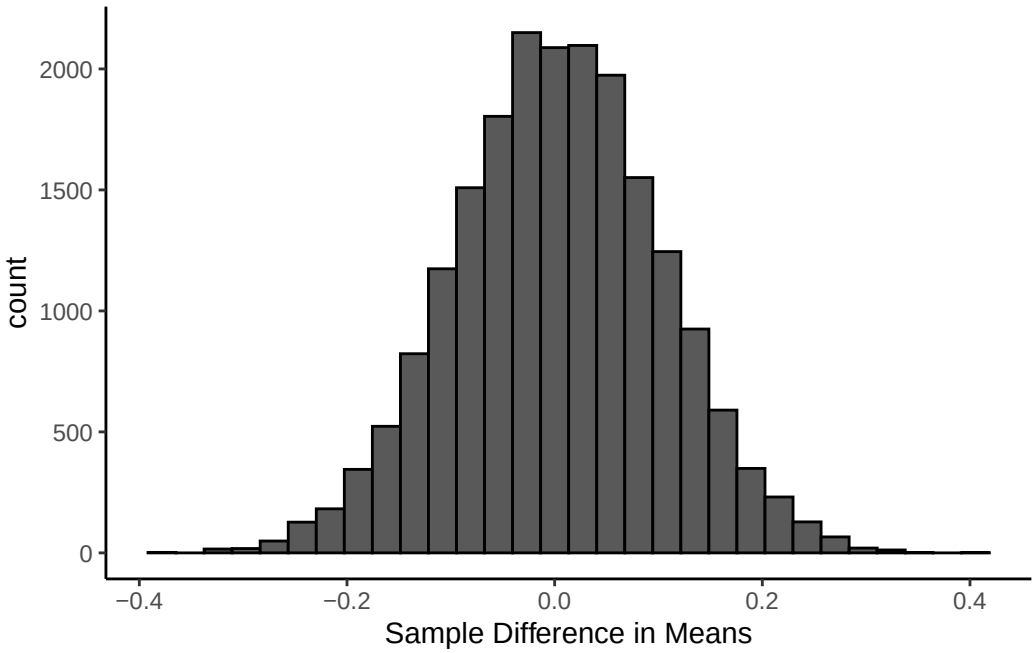
6. Plot it

```
hist(sampling_distribution)
```




```
ggplot(mapping = aes(x=sampling_distribution)) +
  geom_histogram(color = 'black') +
  theme_classic() +
  labs(x = 'Sample Difference in Means')
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



What's the expected value?

```
mean(sampling_distribution)
```

```
[1] 0.002899952
```

Standard error?

```
sd(sampling_distribution)
```

```
[1] 0.09922471
```