Homework 4

Answer Key

Sarah E. Grabinski

2024-11-16

Table of contents

Instructions

You work for Secretary of Transportation Pete Buttigieg, and he has asked you to prepare a year end report comparing flight delays during the Trump administration and the Biden administration. He would like to know on average what percentage of flights were delayed and how long they were delayed for during each administration. He would also know if there was a meaningful difference between the two administrations.

Your data comes from the Bureau of Transportation Statistics, run by the US Department of Transportation. It is a random sample of 800 airports from the years 2016-2024.

- Follow the previously given instructions to start a new project in RStudio for Homework 4. Go to the project folder.
- Download the files `homework4_template.qmd` and `airline_delays.xlsx` from Canvas. Move these files to your Homework 4 project folder.
- Please save your homework as `homework4_firstnamelastname.qmd`. When you render, it should produce the file `homework4_firstnamelastname.html`.
- Follow the prompts to complete the steps of each statistical analysis. Work must be shown in code chunks for partial credit.
- Add packages to the `load_packages` code chunk as you work through the analysis and need their functions. This portion has not been pre-filled for you.
- Answer in complete sentences where indicated.

Codebook

Table 1: Variables

Variable	Description	Type
admin	Administration name (Biden or Trump)	Categorical, binary
flights	Total count of flights through airport	Numeric, continuous
delayed	total count of delayed flights through airport	Numeric, continuous
avg_delay	Average delay in minutes for airport	Numeric, continuous

Packages

Load any packages required for your analysis below using the `library()` function.

```
library(readxl)
library(dplyr)
```

```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

Load Data

In the code chunk below, use the `getwd()` function to save the path to your project folder as a text string called `folder_path`.

```r
folder_path <- getwd()
```

Now, create a variable called `file_name` which sorts the name of your data file `airline_delays.xlsx`.

```r
file_name <- "airline_delays.xlsx"
```

Use the `paste()` function to connect `folder_path` to `file_name`, using the slash character "/" to separate them. Save this result as `file_path`.

```r
file_path <- paste(folder_path,
                   file_name,
                   sep = '/')
```

Now that you have made a path to your file, load it using a function that will handle Excel spreadsheets ending in `.xlsx`.

```r
delays <- read_xlsx(file_path)
```

Inspect data

In the code chunk below, use one of the functions we've learned to summarize your data set.

```r
summary(delays)
```

```
    admin              airport          airport_name          flights
 Length:800         Length:800         Length:800         Min.   :      1
 Class :character   Class :character   Class :character   1st Qu.:   3052
 Mode  :character   Mode  :character   Mode  :character   Median :  10258
                                                          Mean   :  70183
                                                          3rd Qu.:  39318
                                                          Max.   :1450670
    delayed            pct_delay          avg_delay
 Min.   :     0.0   Min.   :0.0000   Min.   :  0.000
 1st Qu.:   444.8   1st Qu.:0.1433   1st Qu.:  9.357
 Median :  1650.5   Median :0.1651   Median : 11.349
 Mean   : 12293.0   Mean   :0.1683   Mean   : 11.924
 3rd Qu.:  6903.5   3rd Qu.:0.1871   3rd Qu.: 13.366
 Max.   :252657.0   Max.   :1.0000   Max.   :222.000
```

Write 2-3 sentences describing anything you learned from this summary.

> **From this summary, you can tell which variables are numeric and which are categorical or text. You can also see the minimums and maximums, the values of the quartiles, and the mean.**

```r
Hmisc::describe(delays)
```

```
delays

 7  Variables      800  Observations
--------------------------------------------------------------------------------
admin
       n  missing distinct
     800        0        2

Value      Biden Trump
Frequency    403   397
Proportion 0.504 0.496
--------------------------------------------------------------------------------
airport
       n  missing distinct
     800        0      391

lowest : ABE ABI ABQ ABR ABY, highest: XWA YAK YKM YNG YUM
--------------------------------------------------------------------------------
airport_name
       n  missing distinct
     800        0      415

lowest : Aberdeen, SD: Aberdeen Regional                  Abilene, TX: Abilene Regional
highest: Wrangell, AK: Wrangell Airport                   Yakima, WA: Yakima Air Terminal/Mc
--------------------------------------------------------------------------------
flights
       n  missing distinct     Info     Mean      Gmd      .05      .10
     800        0      782        1    70183   113692    510.7   1530.4
     .25      .50      .75      .90      .95
  3051.8  10257.5  39317.8 185218.5 409950.7

lowest :       1       2       9      12      46
highest: 1283179 1283576 1312353 1406671 1450670
--------------------------------------------------------------------------------
```

```
delayed
       n  missing distinct     Info     Mean      Gmd      .05      .10
     800        0      735        1    12293    19991     78.9    216.9
     .25      .50      .75      .90      .95
   444.8   1650.5   6903.5  32810.4  71576.4

lowest :      0      1      2      7      9, highest: 203649 222857 230997 234577 252657
--------------------------------------------------------------------------------
pct_delay
       n  missing distinct     Info     Mean      Gmd      .05      .10
     800        0      797        1   0.1683  0.04734   0.1025   0.1203
     .25      .50      .75      .90      .95
  0.1433   0.1651   0.1871   0.2097   0.2279

lowest : 0         0.0453053 0.0648596 0.0658777 0.0662252
highest: 0.351932  0.406076  0.416667  0.439394  1
--------------------------------------------------------------------------------
avg_delay
       n  missing distinct     Info     Mean      Gmd      .05      .10
     800        0      799        1    11.92    4.702    5.959    7.371
     .25      .50      .75      .90      .95
   9.357   11.349   13.366   16.199   18.112

lowest : 0       1.39735 2.38593 2.46189 3.02167
highest: 26.2338 26.4537 44.4394 59.5    222
--------------------------------------------------------------------------------
```

Write 2-3 sentences describing anything you learned from this summary.

Like with the `summary()` function from base R, you can tell which variables are numeric and which are categorical or text. You can also see the minimums and maximums, the values of the quartiles, and the mean.
An important feature of this description is that it includes the total number of observations, the number of missing observations, and the number of distinct values in each variable. You also get the values of the 5th, 10th, 90th, and 95th percentiles, which can help give an idea of skew in data. You also get the lowest and highest 5 values, which can highlight outliers.

Proportions

Research Question: On average, how frequently were flights delayed during 2016-2020 and 2021-2024, and were the rates different between the 2 administrations?

1. Calculate your sample statistics from your data: the proportions of flights delayed during the Trump ($\hat{p}_{\text{Trump}}$) and Biden administrations ($\hat{p}_{\text{Biden}}$).
2. Check the assumptions for inference and hypothesis testing using the Central Limit Theorem for proportions.
3. Infer the average proportion of flights in the US were delayed during the Trump and Biden administrations from your sample statistics.
4. Test the hypothesis that the proportion of flights delayed during the Trump administration is different from the proportion of flights delayed during the Biden administration.

Sample Statistics

Use the `summarize()` function from the `dplyr` package to find the sample size (n), the total number of flights (`flights`), and the total number of delayed flights (`delayed`) for each administration (`admin`).

```
proportion_statistics <- delays |>
  summarize(n = n(),
            flights = sum(flights),
            delayed = sum(delayed),
            proportion = sum(delayed) / sum(flights),
            .by = 'admin')

proportion_statistics
```

```
# A tibble: 2 x 5
  admin     n  flights delayed proportion
  <chr> <int>    <dbl>   <dbl>      <dbl>
1 Trump   397 25611227 4295997      0.168
2 Biden   403 30534936 5538365      0.181
```

Calculate your point estimate of the proportion of flights that were delayed during the Trump administration $\hat{p}_{\text{Trump}}$ and save it as the variable `phat_trump`.

```r
phat_trump <- 4503558/26613488

phat_trump
```

```
[1] 0.1692209
```

Calculate your point estimate of the proportion of flights that were delayed during the Biden administration $\hat{p}_{\text{Biden}}$ and save it as the variable `phat_biden`.

```r
phat_biden <- 5497718/30326333

phat_biden
```

```
[1] 0.1812853
```

You will need sample sizes later in the analysis. Save the sample size for airports from the Trump administration as the variable `n_trump`.

```r
n_trump <- 403
```

Save the sample size for airports from the Biden administration as the variable `n_biden`.

```r
n_biden <- 397
```

Assumptions

Reliability

When the data you observe in your sample is very close to the "ground truth" or what you would expect to see under perfect conditions in that sample, the data and its sample statistics are considered to be *reliable*. You don't expect $\hat{p}_{\text{observed}}$ to be very different from $\hat{p}_{\text{expected}}$, so $\hat{p}_{\text{observed}}$ is a reliable/accurate estimator for your sample.

- You don't believe there is much bias or measurement error in the data
- Your sample does not have a lot of missing data
- You expect your sample statistic to be close to the "true" sample population parameter

$$\hat{p}_{\text{sample}} \approx p_{\text{sample}}$$

Write 1-2 sentences reflecting on whether the data from your sample could be considered reliable.

> **The proper timing of flights is a safety issue, so there is not likely to be much error in the recording of delay times. In addition, there is not likely to be a lot of missing data. Finally, the delay time is recorded in seconds, but a flight is considered late if it is delayed over 15 minutes. The time is precisely measured, so there are unlikely to be many flights misclassified as late due to poor measurement. Therefore, it's reasonable to assume that this data is reliable.**

Validity

When the data you observe in your sample is very close to the "ground truth" or what you would expect to see under perfect conditions in any sample of size n from your study population, the data and sample statistics are considered to be *valid*. You don't expect the sample statistic $\hat{p}$ to be very different from the population parameter p, so $\hat{p}$ is a valid/accurate estimator for your study population.

- You don't believe there is much bias or measurement error in the data
- You have a large, representative sample (or you have all the data available)
- Your sample does not have a lot of missing data
- You expect your sample statistics to be close to the "true" study population parameter

$$\hat{p}_{\text{sample}} \approx p$$

Since I have access to the full data, I can tell you the "true" population parameters.

$$p_{\text{Trump}} = 0.169$$
$$p_{\text{Biden}} = 0.181$$

Compare these population parameters to your sample statistics and write 1-2 sentences reflecting on whether or not your data is valid.

> **If you calculated your sample statistics correctly, these should be approximately equal to your $\hat{p}$ values `phat_trump` and `phat_biden`. When the sample**

statistic approximately equals the population parameter, your data is valid. Sample statistics calculated from your data are valid estimators of the study population parameters because they are good approximations of the study population.
In addition, this was a random sample of airports. Data from a random sample is likely to be representative of the population, which contributes to the data's validity.
Finally, you have large sample sizes with $n_{\text{Trump}} = 403$ (`n_trump`) and $n_{\text{Biden}} = 397$ (`n_biden`), which tend to produce valid estimates. Assuming the sample is representative (e.g. random), as the sample size increases ($n \to \infty$), the sample statistic $\hat{p}$ will be closer and closer to the "true" population parameter p ($\hat{p} \to p$).

Sample Size

For the Central Limit Theorem to apply for a proportion, you need at least 10 successes and 10 failures in your sample such that…

$$
\begin{aligned}
n &\geq 20 \\
np &\geq 10 \\
n(1-p) &\geq 10
\end{aligned}
$$

In the code chunk below, check the assumption that, on average, airports had at least 10 delayed flights during the Trump administration.

```r
n_trump * phat_trump
```

```
[1] 68.19602
```

On average, each airport during the Trump administration had 68 delayed flights per year. This is larger than 10.

In the code chunk below, check the assumption that, on average, airports had at least 10 flights that were not delayed during the Trump administration.

```r
n_trump * (1 - phat_trump)
```

```
[1] 334.804
```

On average, each airport during the Trump administration had 334 on time flights per year. This is larger than 10.

In the code chunk below, check the assumption that, on average, airports had at least 10 delayed flights during the Biden administration.

```
n_biden * phat_biden
```

```
[1] 71.97026
```

On average, each airport during the Biden administration had 72 delayed flights per year. This is larger than 10.

In the code chunk below, check the assumption that, on average, airports had at least 10 flights that were not delayed during the Biden administration.

```
n_biden * (1 - phat_biden)
```

```
[1] 325.0297
```

On average, each airport during the Biden administration had 325 on time flights per year. This is larger than 10.

Inference

Once you have checked your assumptions, you can use the Central Limit Theorem to infer the sampling distribution of $\hat{p}$ in the study population.

The Central Limit Theorem states that as the sample size n increases ($n \to \infty$), the distribution of the sample statistic $\hat{p}$ approximates a normal distribution with mean p and standard deviation SE or the standard error of a proportion.

$$\hat{p} \sim \mathrm{N}\left(p, \mathrm{SE}\right)$$

The standard deviation for this sampling distribution, the standard error SE of a single sample proportion $\hat{p}$, is calculated as follows.

$$\mathrm{SE} = \sqrt{\frac{p(1-p)}{n}}$$

However, the population parameter p, the "true" proportion of flights that are delayed, is "unknowable" unless you have all the data possible. You have a random sample, so you can not directly observed the sampling distribution of $\hat{p}$.

However, when the data in your sample is valid, $\hat{p} \approx p$ and can be used to infer the sampling distribution of $\hat{p}$.

$$\hat{p} \sim \mathrm{N}\left(\hat{p}, \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

Work through the steps below to infer the sampling distributions for $\hat{p}_{\text{Trump}}$ and $\hat{p}_{\text{Biden}}$ as $\hat{p} \sim \mathrm{N}(\hat{p}, \mathrm{SE})$.

You calculated your point estimates for $\hat{p}_{\text{Trump}}$ and $\hat{p}_{\text{Biden}}$ in the last section. Now, you will calculate the standard errors $\mathrm{SE}_{\text{Trump}}$ and $\mathrm{SE}_{\text{Biden}}$.

Use these sampling distributions to calculate 95% confidence intervals for the average proportion of flights that were delayed more than 15 minutes during each administration. Because these are proportions, use the Z or standard normal distribution for inference.

$$\text{point estimate} \pm Z^* \times SE$$

The critical value Z^* corresponds to the Z-score for the probability $\alpha/2$ or $1-\alpha/2$. You can find α from the confidence level.

$$\alpha = 1 - \text{confidence}$$

Standard Errors

Use the formula for the standard error of a single proportion to calculate $\mathrm{SE}_{\text{Trump}}$ below and save it as the variable `se_phat_trump`.

```r
se_phat_trump <- sqrt((phat_trump*(1-phat_trump))/n_trump)

se_phat_trump
```

```
[1] 0.01867744
```

Use the formula for the standard error of a single proportion to calculate SE_{Biden} below and save it as the variable `se_phat_biden`.

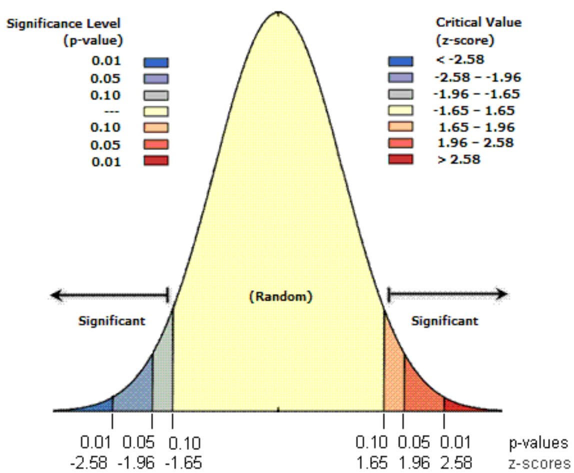
```
se_phat_biden <- sqrt((phat_biden*(1-phat_biden))/n_biden)

se_phat_biden
```

```
[1] 0.01933536
```

Critical Value

The *critical value* Z^* for a 95% confidence interval from the Z distribution indicates how many standard errors away from the mean proportion (p) that 95% of the data can be in this distribution. In other words, 95% of the data in this distribution falls between $-Z^*$ and Z^*.



Calculate α in the code chunk below for a 95% confidence interval.

```
alpha <- 1 - 0.95
```

Use the `qnorm()` function to find the Z-score which corresponds to the probability $\alpha/2$ or $1-\alpha/2$. Save this value as `z_star`.

```
z_star <- qnorm(alpha/2)

z_star
```

```
[1] -1.959964
```

```r
z_star <- qnorm(1-alpha/2)

z_star
```

```
[1] 1.959964
```

Confidence Interval

A confidence interval for a single proportion using the Z-distribution is calculated as point estimate $\pm Z^* \times SE$.

Trump

Use the code chunk below to calculate the lower boundary of the 95% confidence interval for average proportion of flights delayed during the Trump administration.

```r
phat_trump - z_star * se_phat_trump
```

```
[1] 0.1326138
```

Use the code chunk below to calculate the upper boundary of the 95% confidence interval for average number of flights delayed during the Trump administration.

```r
phat_trump + z_star * se_phat_trump
```

```
[1] 0.205828
```

Interpret this confidence interval using a complete sentence.

> **With 95% confidence, 13.26% to 20.58% of flights were delayed at airports during the Trump administration.**

Biden

Use the code chunk below to calculate the lower boundary of the 95% confidence interval for average proportion of flights delayed during the Biden administration.

```
phat_biden - z_star * se_phat_biden
```

```
[1] 0.1433887
```

Use the code chunk below to calculate the upper boundary of the 95% confidence interval for average number of flights delayed during the Biden administration.

```
phat_biden + z_star * se_phat_biden
```

```
[1] 0.2191819
```

Interpret this confidence interval using a complete sentence.

> **With 95% confidence, 14.34% to 21.92% of flights were delayed at airports during the Trump administration.**

Hypothesis Test

Now, you need to test the hypothesis that there's a difference between the two proportions. Testing for the difference between 2 proportions is called a *2-sample proportion test*.

Please do a 2-sided hypothesis test with a significance level of $\alpha = 0.05$.

1. Define your null (H_0) and alternate (H_A) hypotheses regarding the difference between the 2 proportions $\hat{p}_{\text{Biden}} - \hat{p}_{\text{Trump}}$.
2. Calculate a point estimate for the difference between the 2 proportions $\hat{p}_{\text{Biden}} - \hat{p}_{\text{Trump}}$ from your sample.
3. Calculate the standard error SE for the difference between 2 proportions and infer the sampling distribution of the $\hat{p}_{\text{Biden}} - \hat{p}_{\text{Trump}}$ under your null hypothesis.
4. Construct a confidence interval under the Z distribution for the difference between the 2 proportions $\hat{p}_{\text{Biden}} - \hat{p}_{\text{Trump}}$ for a significance threshold of $\alpha = 0.05$.
5. Calculate the test statistic for your observed data $\hat{p}_{\text{Biden}} - \hat{p}_{\text{Trump}}$ under your null hypothesis.
6. Find the p-value, or the probability of observing a difference between proportions $\hat{p}_{\text{Biden}} - \hat{p}_{\text{Trump}}$ as extreme or more extreme than the one you found under the null hypothesis.
7. Compare your p-value to your significance level α and choose whether or not to reject your null hypothesis H_0.

Hypothesis Statements

H_0

State your null hypothesis below in 1 sentence.

Here, you needed to return to the research question motivating your research.

> *Research Question: On average, how frequently were flights delayed during 2016-2020 and 2021-2024, and were the rates different between the 2 administrations?*

The null hypothesis is a statement of skepticism regarding your research question. If your research question has you investigating whether or not there IS a difference, then your null hypothesis should be that there is NOT a difference.
The null hypothesis is that the difference between the proportion of flights delayed at airports during the Trump administration and the proportion of flights delayed at airports during the Biden administration is 0.
$H_0 : p_{\text{Biden}} - p_{\text{Trump}} = 0$

H_A

State your alternative hypothesis below in 1 sentence.

The alternate hypothesis is a statement regarding the scenario proposed by your research question. If your research question has you investigating whether or not there IS a difference, then your alternate hypothesis should be that there IS a difference.
Additionally, you were asked to do a 2-sided hypothesis test, which means you do not assume in advance the direction of the difference. We don't know whether airports under Biden or Trump were expected to have more delays, so we allow for the possibility of both by saying that the difference is not 0. If we were doing a 1-sided hypothesis test, assuming in advance that airports had more delayed flights under one or the other, we would say that the difference is greater or less than zero.
The alternate hypothesis is that the difference between the proportion of flights delayed at airports during the Trump administration and the proportion of flights delayed at airports during the Biden administration is not 0.
$H_A : p_{\text{Biden}} - p_{\text{Trump}} \neq 0$

Point Estimate

In the code chunk below, use your variables `phat_biden` and `phat_trump` to find $\hat{p}_{\text{Biden}} - \hat{p}_{\text{Trump}}$. Save the result as the variable `phat_diff`.

```
phat_diff <- phat_biden - phat_trump
```

Standard Error

Use the formula for the standard error of the difference between 2 proportions to calculate $\text{SE}_{\hat{p}_{\text{Biden}} - \hat{p}_{\text{Trump}}}$ below and save it as the variable `se_phat_diff`.

Choose from the formulas below based on your null hypothesis.

$$\text{SE}_{\text{diff}} = \begin{cases} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}, & \text{when } H_0 : \hat{p}_1 - \hat{p}_2 \neq 0 \\ \sqrt{\frac{p_{\text{pool}}(1-p_{\text{pool}})}{n_1} + \frac{p_{\text{pool}}(1-p_{\text{pool}})}{n_2}}, & \text{when } H_0 : \hat{p}_1 - \hat{p}_2 = 0 \end{cases}$$

Based on your research question, your null hypothesis should have been $H_0 : p_1 - p_2 = 0$. You should have chosen the bottom equation for the standard error of the difference between 2 proportions and completed the steps below.

If your null hypothesis is $H_0 : \hat{p}_1 - \hat{p}_2 \neq 0$, you will also need to calculate $\hat{p}_{\text{pool}}$. That's because this hypothesis states that the two samples come from the same population. Therefore, we only need 1 parameter $\hat{p}_{\text{pool}}$ for the population.

$$\hat{p}_{\text{pool}} = \frac{\text{count(delays)}_{\text{Trump}} + \text{count(delays)}_{\text{Biden}}}{\text{count(flights)}_{\text{Trump}} + \text{count(flights)}_{\text{Biden}}}$$

Use the `summarize()` function from the `dplyr` package to find the sample size (n), the total number of flights (`flights`), and the total number of delayed flights (`delayed`), but do not include a `.by` parameter to separate the results by administration.

```
proportion_statistics2 <- delays |>
  summarize(n = n(),
            flights = sum(flights),
            delayed = sum(delayed),
            proportion = sum(delayed) / sum(flights))

proportion_statistics2
```

```
# A tibble: 1 x 4
      n  flights delayed proportion
  <int>    <dbl>   <dbl>      <dbl>
1   800 56146163 9834362      0.175
```

If you need to calculate $\hat{p}_{\text{pool}}$, use the code chunk below to save the result as the variable `phat_pool`.

```
phat_pool <- 10001276/56939821
```

Use the code chunk below to find the standard error of the difference $\text{SE}_{\hat{p}_{\text{Biden}}-\hat{p}_{\text{Trump}}}$ under your null hypothesis. Save the result as `se_phat_diff`.

```
se_phat_diff <- sqrt((phat_pool*(1-phat_pool))/n_trump + (phat_pool*(1-phat_pool))/n_biden)

se_phat_diff
```

```
[1] 0.02690752
```

Critical Score

The difference of 2 proportions uses the Z distribution, so the critical score for this confidence interval is Z^*. Find the Z-score for the probability $\alpha/2$ or $1-\alpha/2$ using the `qnorm()` function.

```
z_star <- qnorm(alpha/2)

z_star
```

```
[1] -1.959964
```

```
z_star <- qnorm(1-alpha/2)

z_star
```

```
[1] 1.959964
```

Confidence Interval

As with the 1-sample proportions, the confidence interval for the difference between 2 proportions is calculated as point estimate $\pm Z^* \times SE$.

Use the code chunk below to calculate the lower boundary of your confidence interval for average difference in the proportion of flights delayed during the Trump administration and proportion of flights delayed during the Biden administration.

```
phat_diff - z_star * se_phat_diff
```

```
[1] -0.04067336
```

Use the code chunk below to calculate the upper boundary of your confidence interval for average difference in the proportion of flights delayed during the Trump administration and proportion of flights delayed during the Biden administration.

```
phat_diff + z_star * se_phat_diff
```

```
[1] 0.06480217
```

Interpret this confidence interval in 1 sentence.

With 95% confidence, the difference between the proportion of flights delayed more than 15 minutes at airports during the Biden administration and the proportion of flights delayed more than 15 minutes at airports during the Trump administration was -4.1% to 6.5%.

Test Statistic

Calculate the test statistic Z under your null distribution for your observed value $\mathrm{SE}_{\hat{p}_{\mathrm{Biden}}-\hat{p}_{\mathrm{Trump}}}$. Save the value as `z_diff`.

$$Z = \begin{cases} \frac{(\hat{p}_1-\hat{p}_2)-\mu}{\mathrm{SE}}, & \text{when } H_0: \hat{p}_1 - \hat{p}_2 \neq 0 \\ \frac{\hat{p}_1-\hat{p}_2}{\mathrm{SE}}, & \text{when } H_0: \hat{p}_1 - \hat{p}_2 = 0 \end{cases}$$

Based on your research question, your null hypothesis should have been $H_0 : p_1 - p_2 = 0$. You should have chosen the bottom equation for the test statistic calculation.

```r
z_diff <- phat_diff/se_phat_diff

z_diff
```

```
[1] 0.4483655
```

Interpret the test statistic in 1 sentence.

> **In our sample, we observed that 1.2% more flights were delayed more than 15 minutes at airports during the Biden administration than during the Trump administration, which is 0.45 standard errors above the null hypothesis that there was no difference in the proportion of flights delayed more than 15 minutes at airports during the two administrations.**

P-Value

Find the p-value for your test statistic `z_diff` using the function `pnorm()`. Use a 2-sided hypothesis test.

```r
pnorm(z_diff, lower.tail = F) * 2
```

```
[1] 0.6538894
```

Decision

Given the p-value for your observed data under the null hypothesis and a significance threshold of $\alpha = 0.05$, would you reject the null hypothesis? Why or why not?

> **The p-value of 0.654 is greater than our significance threshold of 0.05, so you should not have rejected the null hypothesis with this data.**
> **If the null hypothesis were true, the difference between the proportion of flights delayed at airports during the Trump and Biden administrations is 0. Assuming the null hypothesis is true, the probability of taking two samples of size $n_{\text{Trump}} = 403$ and $n_{\text{Biden}} = 397$ from this study population and seeing a probability difference more extreme than the one we got in our sample (greater than 1.2% or less than -1.2%) is 65%.**
> **In other words, if there were no real difference and you repeated this study over and over, you would get $\hat{p}_{\text{Biden}} - \hat{p}_{\text{Trump}} > 0.012$ or $\hat{p}_{\text{Biden}} - \hat{p}_{\text{Trump}} < -0.012$ about 65% of the time. To reject a null hypothesis, we need our observed**

data to be UNLIKELY under the null hypothesis, or a small p-value. When the observed data is unlikely under the null hypothesis, we can reject that theory of the world and accept the alternate hypothesis. Here, however, our data is consistent with the null hypothesis, so we fail to reject it.
This study did not detect a difference between the proportion of flights delayed more than 15 minutes at airports during the Trump and Biden administrations.

Means

Research Question: On average, how many minutes were flights delayed during 2016-2020 and 2021-2024, and were the times different between the 2 administrations?

1. Calculate your sample statistics from your data: the length of delays during the Trump ($\bar{x}_{\text{Trump}}$) and Biden administrations ($\bar{x}_{\text{Biden}}$).
2. Check the sample size assumption.
3. Infer the average delay length in the US during the Trump and Biden administrations from your sample statistics.
4. Test the hypothesis that the average delay during the Trump administration is different from the average delay during the Biden administration.

Sample Statistics

Use the `summarize()` function from the `dplyr` package to find the sample size (n), the mean (`mean`) of the average delay times in minutes (`avg_delay`), and standard deviation (`sd`) of the average delay times in minutes.

```
mean_statistics <- delays |>
  summarize(n = n(),
            mean = mean(avg_delay),
            sd = sd(avg_delay),
            .by = 'admin')

mean_statistics
```

```
# A tibble: 2 x 4
  admin     n  mean    sd
  <chr> <int> <dbl> <dbl>
1 Trump   397  11.3  4.69
2 Biden   403  12.6 11.1
```

Save your point estimate of $\bar{x}_{\text{Trump}}$ rounded to 2 decimal places as the variable `mean_trump`.

```
mean_trump <- 11.27

mean_trump
```

```
[1] 11.27
```

Save your point estimate of $\bar{x}_{\text{Biden}}$ rounded to 2 decimal places as the variable `mean_biden`.

```
mean_biden <- 12.56

mean_biden
```

```
[1] 12.56
```

Save your point estimate of s_{Trump} rounded to 2 decimal places as the variable `sd_trump`.

```
sd_trump <- 4.69

sd_trump
```

```
[1] 4.69
```

Save your point estimate of s_{Biden} rounded to 2 decimal places as the variable `sd_biden`.

```
sd_biden <- 11.09

sd_biden
```

```
[1] 11.09
```

Assumptions

For the Central Limit Theorem to apply for means, there must be at least $n \geq 30$ observations.

In the code chunk below, check the assumption that you have at least 30 observations from the Trump administration.

```
n_trump
```

```
[1] 403
```

The sample statistic $\bar{x}_{\text{Trump}}$ is based on data from 403 airports under the Trump administration, which is more than 30.

In the code chunk below, check the assumption that you have at least 30 observations from the Biden administration.

```
n_biden
```

```
[1] 397
```

The sample statistic $\bar{x}_{\text{Biden}}$ is based on data from 397 airports under the Trump administration, which is more than 30.

Inference

Once you have checked your assumptions, you can use the Central Limit Theorem to infer the sampling distribution of $\bar{x}$ in the study population.

The Central Limit Theorem states that as the sample size n increases ($n \to \infty$), the distribution of the sample statistic $\bar{x}$ approximates a normal distribution with mean μ and standard deviation SE or the standard error of a mean

$$\bar{x} \sim \mathrm{N}\left(\mu, \mathrm{SE}\right)$$

The standard deviation for this sampling distribution, the standard error SE of a single sample mean $\bar{x}$, is calculated as follows.

$$\text{SE} = \frac{\sigma}{\sqrt{n}}$$

However, the population parameter s, the "true" standard deviation, is "unknowable" unless you have all the data possible. You have a random sample, so you can not directly observed σ.

However, when the data in your sample is valid, $\bar{x} \approx \mu$ and can be used to infer the sampling distribution of $\bar{x}$. We also use s as an estimate for σ, but we will need to account for the additional uncertainty of using 2 estimated statistics in our distribution.

$$\bar{x} \sim \text{N}\left(\bar{x}, \frac{s}{\sqrt{n}}\right)$$

Work through the steps below to infer the sampling distributions for $\bar{x}_{\text{Trump}}$ and $\bar{x}_{\text{Biden}}$ as $\bar{x} \sim \text{N}\left(\bar{x}, \text{SE}\right)$.

You calculated your point estimates for $\bar{x}_{\text{Trump}}$ and $\bar{x}_{\text{Biden}}$ in the last section. Now, you will calculate the standard errors SE_{Trump} and SE_{Biden}.

Use these sampling distributions to calculate 95% confidence intervals for the average delay during each administration.

Because we do not know σ and are approximating it with the sample statistic s, we need to account for the additional uncertainty in our estimates. We use the T or Student's t distribution for inference when we don't know σ.

$$\text{point estimate} \pm T^*_{\text{df}} \times SE$$

The critical value T^* corresponds to the T-score for the probability $\alpha/2$ or $1 - \alpha/2$. You can find α from the confidence level. The T-distribution has an additional parameter called degrees of freedom..

$$\text{df} = n - 1$$

Standard Errors

Use the formula for the standard error of a single mean to calculate SE_{Trump} below and save it as the variable `se_mean_trump`.

```
se_mean_trump <- sd_trump / sqrt(n_trump)

se_mean_trump
```

```
[1] 0.2336255
```

Use the formula for the standard error of a single proportion to calculate $\mathrm{SE}_{\mathrm{Biden}}$ below and save it as the variable `se_mean_biden`.

```
se_mean_biden <- sd_biden / sqrt(n_biden)

se_mean_biden
```

```
[1] 0.5565911
```

Critical Value

The *critical value* T^* for a 95% confidence interval from the T distribution indicates how many standard errors away from the mean delay (μ) that 95% of the data can be in this distribution. In other words, 95% of the data in this distribution falls between $-T^*$ and T^*.

Calculate α in the code chunk below for a 95% confidence interval.

```
alpha <- 1 - 0.95
```

Use the `qt()` function to find the T-score which corresponds to the probability $\alpha/2$ or $1-\alpha/2$ for the Trump administration. The `qt()` function takes the additional parameter $\mathrm{df} = n - 1$. Save this value as `t_star_trump`.

```
t_star_trump <- qt(alpha/2,
                   df = n_trump-1)

t_star_trump
```

```
[1] -1.965883
```

```
t_star_trump <- qt(1-alpha/2,
                   df = n_trump-1)

t_star_trump
```

```
[1] 1.965883
```

Use the `qt()` function to find the T-score which corresponds to the probability $\alpha/2$ or $1-\alpha/2$ for the Biden administration. Save this value as `t_star_biden`.

```
t_star_biden <- qt(alpha/2,
                   df = n_biden-1)

t_star_biden
```

```
[1] -1.965973
```

```
t_star_biden <- qt(1-alpha/2,
                   df = n_biden-1)

t_star_biden
```

```
[1] 1.965973
```

Compare your T^*'s to your Z^* from your proportions. Comment in 1 sentence on why they are similar or different.

> **The values from the Z distribution are smaller because it is a narrower distribution. 95% of the data is closer to the mean of the distribution, so the critical values will be closer to 0. The T distribution is wider to account for the uncertainty of using an estimated standard deviation to calculate the standard error. Therefore, you need to be further from 0 to capture 95% of the data.**

Confidence Interval

A confidence interval for a single mean using the T-distribution with degrees of freedom $n-1$ is calculated as point estimate $\pm\, T^*_{\text{df}} \times SE$.

Trump

Use the code chunk below to calculate the lower boundary of the 95% confidence interval for average delay during the Trump administration.

```
mean_trump - t_star_trump * se_mean_trump
```

```
[1] 10.81072
```

Use the code chunk below to calculate the upper boundary of the 95% confidence interval for average delay during the Trump administration.

```r
mean_trump + t_star_trump * se_mean_trump
```

```
[1] 11.72928
```

Interpret this confidence interval using a complete sentence.

> **With 95% confidence, flights were delayed 10.81-11.73 minutes on average at airports during the Trump administration.**

Biden

Use the code chunk below to calculate the lower boundary of the 95% confidence interval for average delay during the Biden administration.

```r
mean_biden - t_star_biden * se_mean_biden
```

```
[1] 11.46576
```

Use the code chunk below to calculate the upper boundary of the 95% confidence interval for average delay during the Biden administration.

```r
mean_biden + t_star_biden * se_mean_biden
```

```
[1] 13.65424
```

Interpret this confidence interval using a complete sentence.

> **With 95% confidence, flights were delayed 11.47-13.65 minutes on average at airports during the Biden administration.**

Hypothesis Test

Now, you need to test the hypothesis that there's a difference between the two means. Testing for the difference between 2 proportions is called a *2-sample t* test.

Please do a 2-sided hypothesis test with a significance level of $\alpha = 0.05$.

1. Define your null (H_0) and alternate (H_A) hypotheses regarding the difference between the 2 proportions $\bar{x}_{\text{Biden}} - \bar{x}_{\text{Trump}}$.
2. Calculate a point estimate for the difference between the 2 proportions $\bar{x}_{\text{Biden}} - \bar{x}_{\text{Trump}}$ from your sample.
3. Calculate the standard error SE for the difference between 2 proportions and infer the sampling distribution of the $\bar{x}_{\text{Biden}} - \bar{x}_{\text{Trump}}$ under your null hypothesis.
4. Construct a confidence interval under the T distribution for the difference between the 2 proportions $\bar{x}_{\text{Biden}} - \bar{x}_{\text{Trump}}$ for a significance threshold of $\alpha = 0.05$.
5. Calculate the test statistic for your observed data $\bar{x}_{\text{Biden}} - \bar{x}_{\text{Trump}}$ under your null hypothesis.
6. Find the p-value, or the probability of observing a difference between means $\bar{x}_{\text{Biden}} - \bar{x}_{\text{Trump}}$ as extreme or more extreme than the one you found under the null hypothesis.
7. Compare your p-value to your significance level α and choose whether or not to reject your null hypothesis H_0.

Hypothesis Statements

H_0

State your null hypothesis below in 1 sentence.

Here, you needed to return to the research question motivating your research.

> *Research Question: On average, how many minutes were flights delayed during 2016-2020 and 2021-2024, and were the times different between the 2 administrations?*

The null hypothesis is a statement of skepticism regarding your research question. If your research question has you investigating whether or not there IS a difference, then your null hypothesis should be that there is NOT a difference.
The null hypothesis is that the difference between the average delay time at airports during the Trump administration and the average delay time at airports during the Biden administration is 0.

$H_0: \mu_{\text{Biden}} - \mu_{\text{Trump}} = 0$

H_A

State your alternative hypothesis below in 1 sentence.

The alternate hypothesis is a statement regarding the scenario proposed by your research question. If your research question has you investigating whether or not there IS a difference, then your alternate hypothesis should be that there IS a difference.
Additionally, you were asked to do a 2-sided hypothesis test, which means you do not assume in advance the direction of the difference. We don't know whether airports under Biden or Trump were expected to have longer delays, so we allow for the possibility of both by saying that the difference is not 0. If we were doing a 1-sided hypothesis test, assuming in advance that airports had longer delays under one or the other, we would say that the difference is greater or less than zero.
The alternate hypothesis is that the difference between the average delay time at airports during the Trump administration and the average delay time at airports during the Biden administration is not 0.
$H_A: \mu_{\text{Biden}} - \mu_{\text{Trump}} \neq 0$

Point Estimate

In the code chunk below, use your variables `mean_biden` and `mean_trump` to find $\bar{x}_{\text{Biden}} - \bar{x}_{\text{Trump}}$. Save the result as the variable `mean_diff`.

```
mean_diff <- mean_biden - mean_trump

mean_diff
```

```
[1] 1.29
```

Standard Error

Use the formula for the standard error of the difference between 2 proportions to calculate $\text{SE}_{\bar{x}_{\text{Biden}} - \bar{x}_{\text{Trump}}}$ below and save it as the variable `se_mean_diff`.

$$\text{SE} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Use the code chunk below to find the standard error of the difference $\text{SE}_{\bar{x}_{\text{Biden}} - \bar{x}_{\text{Trump}}}$ under your null hypothesis. Save the result as `se_mean_diff`.

```
se_mean_diff <- sqrt((sd_trump^2)/n_trump + (sd_biden^2)/n_biden)

se_mean_diff
```

```
[1] 0.6036345
```

Critical Score

The difference of 2 means uses the T distribution with the degrees of freedom $\text{df} = min(n_1, n_2) - 1$, so the critical score for this confidence interval is T^*_{df}.

Use the code chunk below to find the degrees of freedom for the difference in means using the `min()` function. Save the result as `dof`.

```
dof <- min(n_trump, n_biden)-1
```

Find the T-score for the probability $\alpha/2$ or $1-\alpha/2$ using the `qt()` function and the parameter `df` set to the degrees of freedom you just calculated. Save the value as `t_star_diff`

```
t_star_diff <- qt(alpha/2,
                  df = dof)

t_star_diff
```

```
[1] -1.965973
```

```
t_star_diff <- qt(1-alpha/2,
                  df = dof)

t_star_diff
```

```
[1] 1.965973
```

Confidence Interval

As with the 1-sample means, the confidence interval for the difference between 2 means is calculated as point estimate $\pm T^*_{\text{df}} \times SE$.

Use the code chunk below to calculate the lower boundary of your confidence interval for average difference in the proportion of flights delayed during the Trump administration and proportion of flights delayed during the Biden administration.

```
mean_diff - t_star_diff * se_mean_diff
```

```
[1] 0.1032711
```

Use the code chunk below to calculate the upper boundary of your confidence interval for average difference in the proportion of flights delayed during the Trump administration and proportion of flights delayed during the Biden administration.

```
mean_diff + t_star_diff * se_mean_diff
```

```
[1] 2.476729
```

Interpret this confidence interval in 1 sentence.

> **With 95% confidence, flights were delayed 0.10 to 2.48 minutes longer at airports during the Biden administration than during the Trump administration.**

Test Statistic

Calculate the test statistic T under your null distribution for your observed value $\text{SE}_{\bar{x}_{\text{Biden}} - \bar{x}_{\text{Trump}}}$. Save the value as `t_mean_diff`.

$$Z = \begin{cases} \frac{(\bar{x}_1 - \bar{x}_2) - \mu}{\text{SE}}, & \text{when } H_0 : \bar{x}_1 - \bar{x}_2 \neq 0 \\ \frac{\bar{x}_1 - \bar{x}_2}{\text{SE}}, & \text{when } H_0 : \bar{x}_1 - \bar{x}_2 = 0 \end{cases}$$

```
t_mean_diff <- (mean_biden - mean_trump)/se_mean_diff

t_mean_diff
```

```
[1] 2.137055
```

Interpret the test statistic in 1 sentence.

> **In our sample, we observed that flights were delayed 1.29 minutes longer at airports during the Biden administration than during the Trump administration, which is 2.14 standard errors above the null hypothesis that there was no difference in the average flight delay at airports during the 2 administrations.**

P-Value

Find the p-value for your test statistic `t_mean_diff` using the function `pt()` and your degrees of freedom `dof`. Use a 2-sided hypothesis test.

```r
pt(t_mean_diff,
   df = dof,
   lower.tail = F) * 2
```

```
[1] 0.03320565
```

Decision

Given the p-value for your observed data under the null hypothesis and a significance threshold of $\alpha = 0.05$, would you reject the null hypothesis? Why or why not?

> **The p-value of 0.033 is less than our significance threshold of 0.05, so you should have rejected the null hypothesis with this data.**
> **If the null hypothesis were true, the difference between the mean delay time at airports during the Trump and Biden administrations is 0. Assuming the null hypothesis is true, the probability of taking two samples of size $n_{\text{Trump}} = 403$ and $n_{\text{Biden}} = 397$ from this study population and seeing a difference in means more extreme than the one we got in our sample (greater than 1.29 minutes or less than -1.29 minutes) is 3.3%.**
> **In other words, if there were no real difference and you repeated this study over and over, you would get $\bar{x}_{\text{Biden}} - \bar{x}_{\text{Trump}} > 1.29$ or $\bar{x}_{\text{Biden}} - \bar{x}_{\text{Trump}} < -1.29$ about 3.3% of the time. To reject a null hypothesis, we need our observed data to be UNLIKELY under the null hypothesis, or a small p-value. When the observed data is unlikely under the null hypothesis, we can reject that theory of the world and accept the alternate hypothesis. Here, we have said that it is evidence against the null hypothesis if the probability of our data under**

the null hypothesis is less than 5%, therefore, we reject the null hypothesis and accept the alternate hypothesis.
Flights were delayed longer during the Biden administration than during the Trump administration.