Week 9 Lab: Hypothesis Testing and Statistical Power

Statistics 100

March 28, 2024

Topics:

- Hypothesis testing
- Constructing null distributions
- Computing and interpreting p-values
- Power calculation

Introduction

Hypothesis testing is based around the idea of assessing how unusual an observed test statistic is under an assumption about the population parameter. If an observed test statistic is highly unlikely to have been observed given that assumption, then this represents evidence that the assumption was incorrect.

- For example, suppose that you assume a coin is fair; i.e., there are equal probabilities of seeing a head versus a tail. Then, suppose that you flip the coin 100 times and see only 5 heads. This result would be highly unlikely if the coin were actually fair; thus, we have reason to believe that the coin is biased towards tails.
- In this scenario, the competing hypotheses are the **null hypothesis** that the coin is fair versus the **alternative hypothesis** that the coin is biased. These hypotheses can be stated in terms of parameters. Let p represent the true proportion of times the coin shows heads when flipped. The null hypothesis is $H_0 : p = 0.50$ and the alternative hypothesis is $H_A : p \neq 0.50$.
- In the sample, we observed 5 heads out 100 coin tosses; i.e., $\hat{p} = 5/100 = 0.05$ is the observed **test statistic**. To understand how likely (or unlikely) it is to see this result under the assumption that the null hypothesis is true, we generate a **null distribution**.

- The **p-value** equals the probability of seeing a test statistic as or more extreme than the one observed if the null hypothesis is true.[1] A small p-value constitutes evidence against the null hypothesis.

Constructing a null distribution

A **null distribution** is a sampling distribution generated assuming that the null hypothesis is true.

A null distribution can be constructed using the `infer` package:

1. Use `specify()` to specify the variables of interest.
2. Use `hypothesize()` to specify the hypothesis.
3. Use `generate()` to generate/draw a specific number of samples.
4. Use `calculate()` to compute the sample statistic of interest within each of the generated samples.
5. Use `visualize()` to see the null distribution.

Computing a p-value

A **p-value** quantifies the likelihood of seeing a sample statistic as or more extreme than what was observed if the null hypothesis is true.

Use `get_p_value()` to compute the p-value. The `shade_p_value()` function can be used to visualize the p-value.

Statistical power

The **statistical power** of a test is the probability that the test rejects the null hypothesis H_0 when the alternative hypothesis H_A is true.

Several factors can affect the power of a test:

- As sample size increases, power increases.
- As standard deviation increases, power decreases.
- As effect size increases, power increases.

Conducting a test at a less strict significance level α also increases power.

[1] The direction of the alternative hypothesis dictates what is considered ”more extreme”.

Background information

This lab uses data from a random sample of 1,728 houses in Saratoga County, New York from 2006. The dataset `SaratogaHouses` is in the `mosaicData` package.

Some key variables:

- `price`: house sale price in US dollars
- `livingArea`: living area in square feet
- `bedrooms`: number of bedrooms
- `fireplaces`: number of fireplaces
- `bathrooms`: number of bathrooms
- `age`: age of house in years
- `waterfront`: whether the property includes waterfront
- `centralAir`: whether the house has central air
- `newConstruction`: whether the property is a new construction

For information on other variables in the dataset, run `?SaratogaHouses` to view the documentation file. The data were collected by Candice Corvetti (Williams College, class of 2007) for her senior thesis; data are from public records kept by the Saratoga Real Property Tax Service.

Practice Questions

1. Let's investigate mean house living area (`livingArea`) based on the sample of houses in the `SaratogaHouses` dataset. Suppose we are interested in investigating whether 5-bedroom houses in Saratoga County have mean living area different from 2700 sq. ft.

```r
#load packages and dataset
library(tidyverse)
library(infer)
library(mosaicData)
data("SaratogaHouses")

library(wesanderson)
wes_green <- wes_palette("Royal2")[5]
wes_green_pale <- "#9dbcac"
```

a) State the null and alternative hypotheses in terms of conjectures and in terms of parameters.

b) Compute the observed test statistic. Interpret the observed test statistic in the context of the data.

```
test_stat <- SaratogaHouses %>%
  filter(bedrooms == 5) %>%
  specify(response = livingArea) %>%
  calculate(stat = "mean")

test_stat
```

```
Response: livingArea (numeric)
# A tibble: 1 x 1
   stat
  <dbl>
1 2476.
```

c) Generate the null distribution and compute a p-value. Interpret the p-value in the context of the data.

```
# set.seed(2022)

living_areas <- SaratogaHouses %>%
  filter(bedrooms == 5) %>%
  specify(response = livingArea)
living_areas
```

```
Response: livingArea (numeric)
# A tibble: 53 x 1
   livingArea
        <dbl>
 1       1701
 2       1912
 3       2304
 4       2464
 5       2310
 6       2662
 7       2576
 8       3140
 9       2310
10       2462
# i 43 more rows
```

```r
min(living_areas)
```

```
[1] 1040
```

```r
max(living_areas)
```

```
[1] 4856
```

```r
mean(living_areas$livingArea)
```

```
[1] 2475.925
```

```r
sd(living_areas$livingArea)
```

```
[1] 695.6618
```

```r
# I actually quite don't know now how would I generate this manually,
# need to look some source code somewhere, or read more on normal distributions?
null_dist <- SaratogaHouses %>%
  filter(bedrooms == 5) %>%
  specify(response = livingArea) %>%
  hypothesize(null = "point", mu = 2700) %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "mean")

null_dist
```

```
Response: livingArea (numeric)
Null Hypothesis: point
# A tibble: 1,000 x 2
   replicate  stat
       <int> <dbl>
 1         1 2608.
 2         2 2872.
 3         3 2491.
 4         4 2686.
 5         5 2658.
 6         6 2832.
```

```
 7          7 2559.
 8          8 2920.
 9          9 2520.
10         10 2595.
# i 990 more rows
```

```
min(null_dist$stat)
```

```
[1] 2424.642
```

```
max(null_dist$stat)
```

```
[1] 3016.226
```

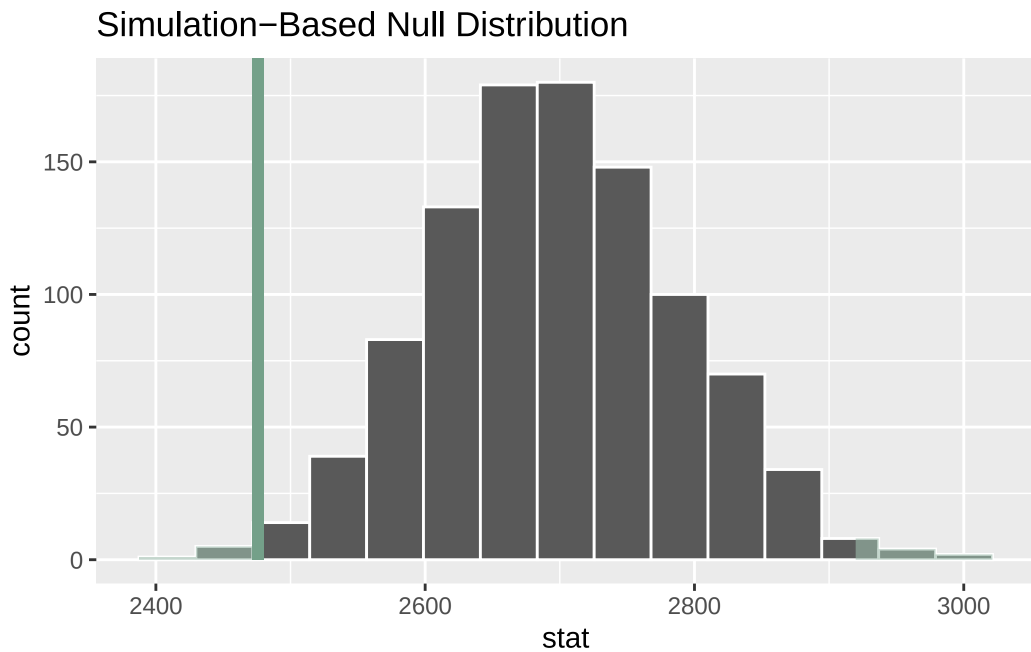
```
mean(null_dist$stat)
```

```
[1] 2697.258
```

```
sd(null_dist$stat)
```

```
[1] 91.72741
```

```
null_dist %>%
  visualise() +
  shade_p_value(test_stat, direction = "two.sided",
                col = wes_green, fill = wes_green_pale)
```




```
p_value <- null_dist %>%
  get_p_value(obs_stat = test_stat,
              direction = "two-sided")
p_value
```

```
# A tibble: 1 x 1
  p_value
    <dbl>
1   0.014
```

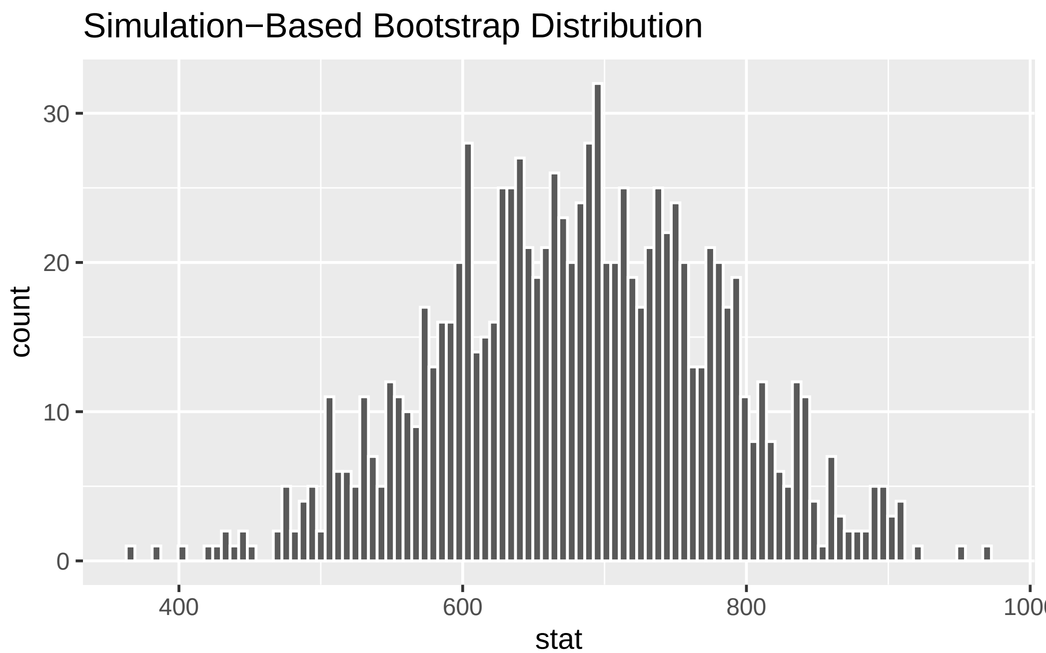
The two-sided p-value is 0.014. If the mean living area for all 5-bedroom homes in Saratoga was 2700 sqft, there would only be a 0.014 probability that the observed sample had a mean living area smaller than 2476 or larger than 2924 sqft. Estimating the significance level at $\alpha = 0.05$, this is evidence to reject H_0 and suggest that the mean living area for all 5-bedroomers in Saratoga is *different* than 2700 sqft.

```
# Personal tests

# SD of the area in the sample
SaratogaHouses %>%
  filter(bedrooms == 5) %>%
  specify(response = livingArea) %>%
  calculate(stat = "sd")
```

```
Response: livingArea (numeric)
# A tibble: 1 x 1
   stat
  <dbl>
1  696.
```

```r
# SDs of the null hypotheses
SaratogaHouses %>%
  filter(bedrooms == 5) %>%
  specify(response = livingArea) %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "sd") %>%
  visualize(bins = 100)
```




```r
# My takeaway is that the null hypotheses
# are built using normal distributions that have
# the same SD as the original sample.
```

d) Suppose the alternative hypothesis had been $H_A : \mu < 2700$ sq. ft. Would you expect this p-value to be smaller or larger than the p-value from part c)? Explain your reasoning.

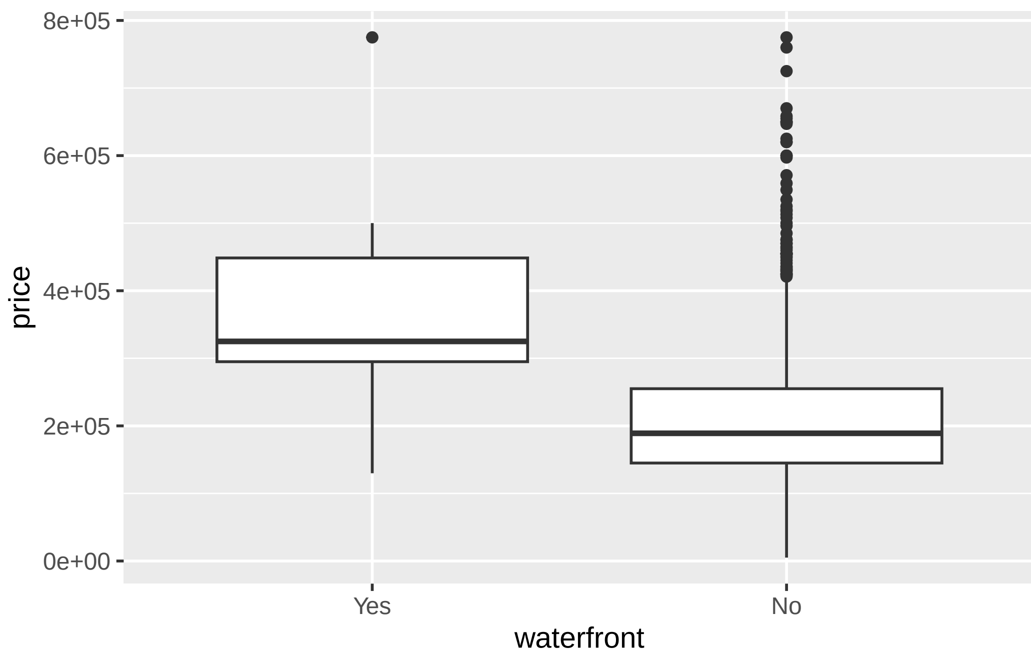
SOLUTION: This -value should be smaller than the -value from part c) since for a one-sided alternative, only the extreme values in one tail constitute evidence against 0 and this alternative uses the left tail area. This -value represents the probability of observing a sample mean living area of 2,476 sq. ft. or smaller if the population mean is actually 2,700 sq. ft.

ME: Is that right? Because when we do two-sided, don't we take have the area? But maybe if we come from the direction of estimation, it doesn't matter?

e) Suppose the alternative hypothesis had been $H_A : \mu > 2700$ sq. ft. Why would it make sense to expect this p-value to be larger than 0.50? Explain your reasoning.

2. Do these data suggest that houses which include waterfront are on average more expensive than those that do not? Conduct a hypothesis test and summarize the findings.

```
ggplot(SaratogaHouses, aes(x = waterfront,
                           y = price,)) +
  geom_boxplot()
```




```
# The mean diff suggests that waterfront houses are more expensive
test_stat <- SaratogaHouses %>%
  drop_na(waterfront) %>%
  specify(price ~ waterfront) %>%
  calculate(stat = "diff in means", order = c("Yes", "No"))
test_stat
```

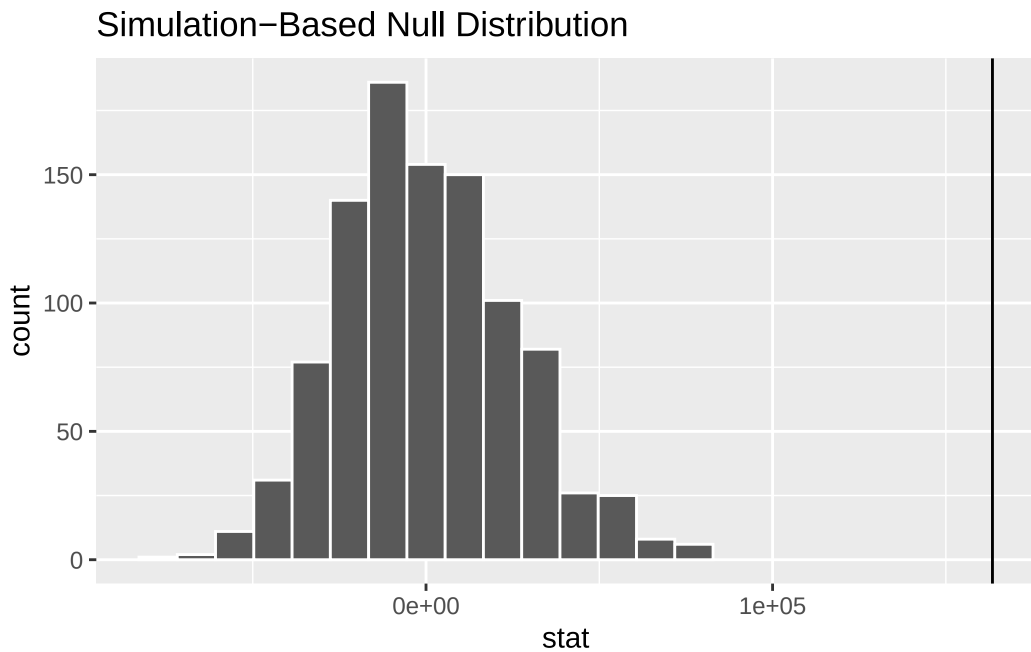
```
Response: price (numeric)
Explanatory: waterfront (factor)
# A tibble: 1 x 1
     stat
    <dbl>
1 163444.
```

```
# Generate null distribution
null_dist <- SaratogaHouses %>%
  drop_na(waterfront) %>%
  specify(price ~ waterfront) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 1000, type = "permute") %>%
  calculate(stat = "diff in means",
            order = c("Yes", "No"))
null_dist
```

```
Response: price (numeric)
Explanatory: waterfront (factor)
Null Hypothesis: independence
# A tibble: 1,000 x 2
   replicate    stat
       <int>   <dbl>
 1         1   7319.
 2         2  33276.
 3         3   2179.
 4         4  -8907.
 5         5  30824.
 6         6  32665.
 7         7 -19990.
 8         8 -27189.
 9         9   -563.
10        10  16697.
# i 990 more rows
```

```
# glimpse(null_dist)
# summary(null_dist)
# summary(SaratogaHouses)

visualise(null_dist) +
  geom_vline(xintercept = test_stat$stat)
```




```r
p_value <- null_dist %>%
  get_p_value(obs_stat = test_stat,
              direction = "greater")
```

```
Warning: Please be cautious in reporting a p-value of 0. This result is an
approximation based on the number of `reps` chosen in the `generate()` step.
See `?get_p_value()` for more information.
```

```r
p_value
```

```
# A tibble: 1 x 1
  p_value
    <dbl>
1       0
```

The null distribution de-correlates waterfront var from price by combining all possible values from the samples, i.e. any observed house (WF or not) can have any observed price. The mean of the null dist is centers around 0, which makes sense if there is no correlation. The test stat (diff in means) is at ~160k. The p-value of this test stat is 0. This suggests that, if H_0 were true, there would be a 0 probability of the diff in mean prices between WF and no WF being equal or larger than ~160k. This gives us evidence to reject H_0 and accept the hypothesis (at alpha = 0.05) that WF houses are on average more expensive than not, from the sample.

SOLUTION: The null hypothesis is that there is no difference in the average house price of houses which include waterfront versus those which do not, 0 − = 0. The alternative hypothesis is that the average house price of houses which include waterfront is greater than that of those which do not, 0 − > 0. Let = 0.05. The observed test statistic is $163,443.70. The sample mean house price of houses with waterfront is $163,443.70 higher than the sample mean house price of houses without waterfront. The -value is practically 0 (< 0.001). It would be practically impossible to see such a large difference in sample mean house price (or larger) if the mean house prices were actually the same between houses which include waterfront versus those that do not. These data represent sufficient evidence to reject the null hypothesis at significance level = 0.05; there is strong evidence that the mean house price of houses which include waterfront is greater than the mean house price of houses that do not include waterfront.

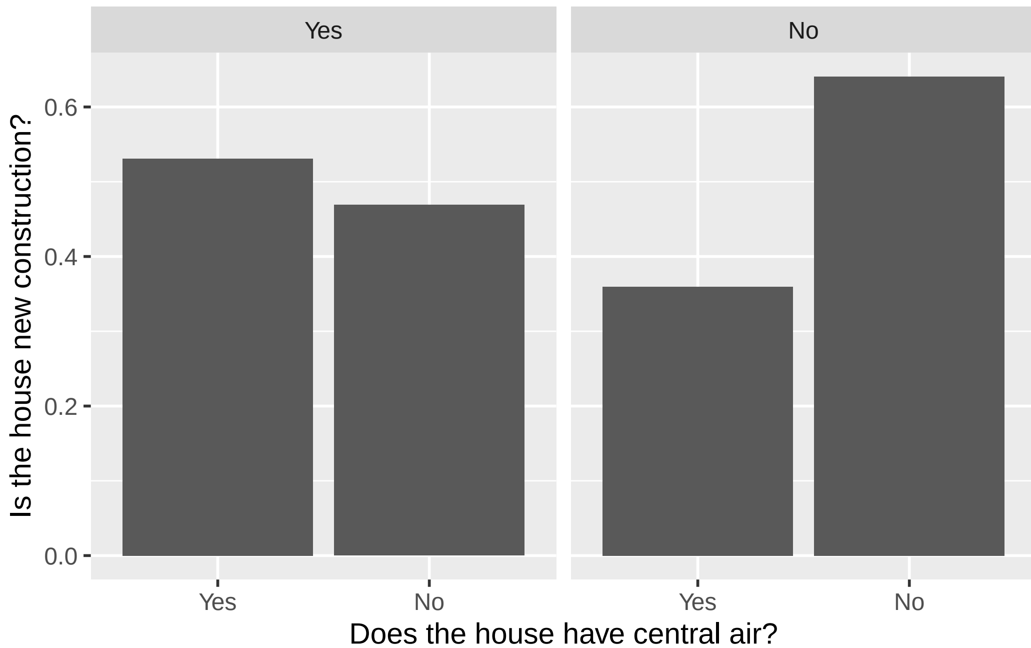
3. Do these data suggest that houses that are a new construction are more likely to have central air than houses which are not? Conduct a hypothesis test and summarize your findings.

The null distribution de-correlates new construction var from AC by combining all possible values from the samples, i.e. any observed house (NC or not) can have any AC value. The mean of the null dist is centers around 0, which makes sense if there is no correlation. The test stat (diff in ratios) is at ~0.17: Among newly constructed houses, about 17% more have central air than among houses that are not newly constructed. The p-value of this test stat is .004. This suggests that, if H_0 were true, there would be a 0.004 probability of the diff in AC ratios between NC and no NC being equal or larger than .17 . This gives us evidence to reject H_0 and accept the hypothesis (at alpha = 0.05) that NC houses are more likely to have AC, from the sample.

```r
# set.seed(2022)

ggplot(SaratogaHouses, aes(x = centralAir)) +
  geom_bar(aes(y = ..prop.., group = 1),
           stat = "count") +
  facet_wrap(~ newConstruction) +
  labs(x = "Does the house have central air?",
       y = "Is the house new construction?")
```

```
Warning: The dot-dot notation (`..prop..`) was deprecated in ggplot2 3.4.0.
i Please use `after_stat(prop)` instead.
```




```
# Compute observed test statistic
test_stat <- SaratogaHouses %>%
  drop_na(newConstruction, centralAir) %>%
  specify(centralAir ~ newConstruction, success = "Yes") %>%
  calculate(stat ="diff in props",
            order = c("Yes", "No"))
test_stat
```

```
Response: centralAir (factor)
Explanatory: newConstruction (factor)
# A tibble: 1 x 1
   stat
  <dbl>
1 0.171
```

```
# Generate null distribution
null_dist <- SaratogaHouses %>%
  drop_na(newConstruction, centralAir) %>%
  specify(centralAir ~ newConstruction, success = "Yes") %>%
  hypothesize(null = "independence") %>%
  generate(reps = 1000, type = "permute") %>%
  calculate(stat ="diff in props",
```

```
             order = c("Yes", "No"))
null_dist
```

```
Response: centralAir (factor)
Explanatory: newConstruction (factor)
Null Hypothesis: independence
# A tibble: 1,000 x 2
   replicate    stat
       <int>   <dbl>
 1         1  0.0419
 2         2  0.0289
 3         3 -0.0229
 4         4 -0.0876
 5         5 -0.0876
 6         6  0.0160
 7         7  0.133
 8         8 -0.0876
 9         9  0.0678
10        10 -0.0229
# i 990 more rows
```

```
mean(null_dist$stat)
```

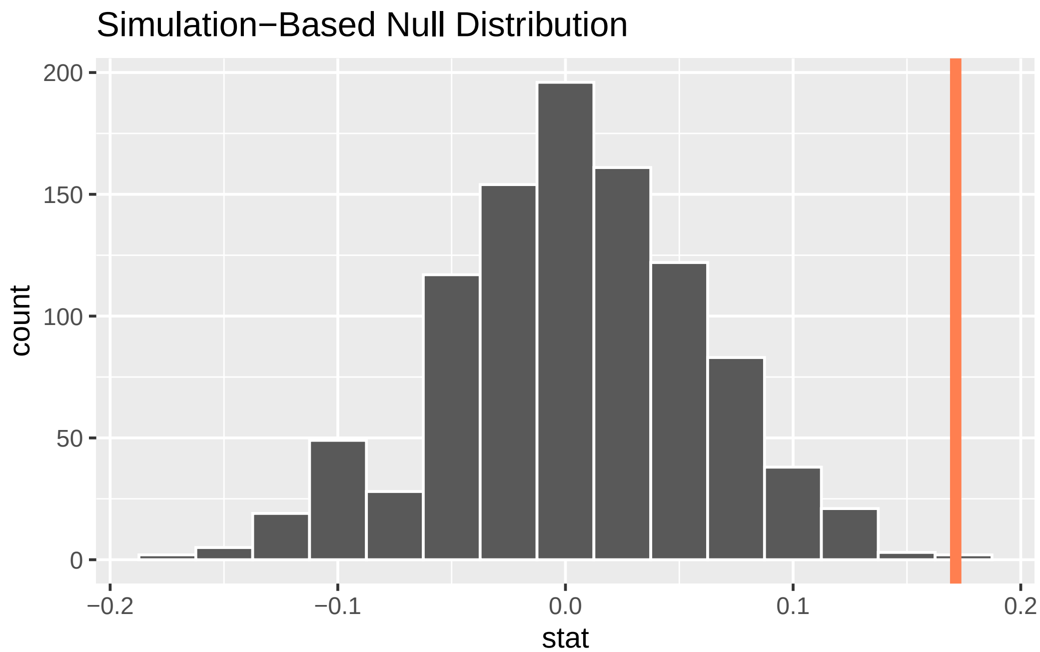
```
[1] 0.00115766
```

```
sd(null_dist$stat)
```

```
[1] 0.05648538
```

```
# Graph null distribution with test statistic
visualize(null_dist) +
  geom_vline(xintercept = test_stat$stat,
             color = 'coral', size = 2)
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```




```
# Compute p-value
p_value <- null_dist %>%
  get_p_value(obs_stat = test_stat,
              direction = "greater")
p_value
```

```
# A tibble: 1 x 1
  p_value
    <dbl>
1   0.002
```

4. Suppose that the city government would like to assess whether there is broad support for launching a comprehensive affordable housing program. They plan to proceed with launching the program if there is evidence that over 70% of residents are in favor of the program.

a) If they survey a random sample of 100 residents and actually 75% of all residents are in favor of the program, what is the power for a test of the one-sided alternative $H_A : p > 0.70$ conducted at the $\alpha = 0.10$ significance level? Do these results suggest that a larger sample size is advisable? Justify your reasoning.

Here, $H_0 : p = 0.70$. So, we can construct a null distribution centered around 0.70 and with sample size 100. This distribution shows the probability of approval values. For a significance level $\alpha = 0.10$, the critical value is 0.76. The alt hypothesis is $H_a : p > 0.70$, we can build an alt dist around 0.75 with 100 samples. After computing the power, we get a 37.9% probability that a sample of 100 residents will be over 70% in favor of the program, with a significance level of 0.10, and assuming that the reality is 75%. This is not a big power, probably wouldn't go for the experiment.

SOLUTION: The power of the test is only 0.379; i.e., there is only a 37.9% chance of rejecting the null correctly. Since the power is so low, this indicates that a larger sample size should be collected.

```
set.seed(2023)

# Construct data frame of sample results with 100 values
n <- 100

# # The data frame has a dist as expected...
# dat <- data.frame(favor = c(rep("Yes", 0.75*n),
#                             rep("No", 0.25*n)))

# Does creating the sample data with the right ratio even matter in this example?
# Since the null and alt hypothesis already have a set target stat value,
# the distributions are created using such ratios, and ignoring the 75% proportions!
# Technically, we can get away with any response ratio in the fake sample data?
# (so long as the sample count is the same, this will affect the SDs)
dat <- data.frame(favor = c(rep("Yes", 0.5*n), rep("No", 0.5*n)))


# Generate a null distribution matching this ratio
# H0 is p == 0.70
null_dist <- dat %>%
  specify(response = favor, success = "Yes") %>%
  hypothesise(null = "point", p = 0.70) %>%
  generate(reps = 1000, type = "draw") %>%
  calculate(stat = "prop")

null_dist
```

```
Response: favor (factor)
Null Hypothesis: point
# A tibble: 1,000 x 2
   replicate  stat
       <int> <dbl>
 1         1  0.67
 2         2  0.71
```

```
 3          3  0.64
 4          4  0.68
 5          5  0.81
 6          6  0.73
 7          7  0.73
 8          8  0.74
 9          9  0.77
10         10  0.73
# i 990 more rows
```

```r
mean(null_dist$stat)
```

```
[1] 0.702
```

```r
sd(null_dist$stat)
```

```
[1] 0.04710646
```

```r
# Where is the critical value for alpha = 0.10?
alpha_x <- quantile(null_dist$stat, 0.90)
alpha_x
```

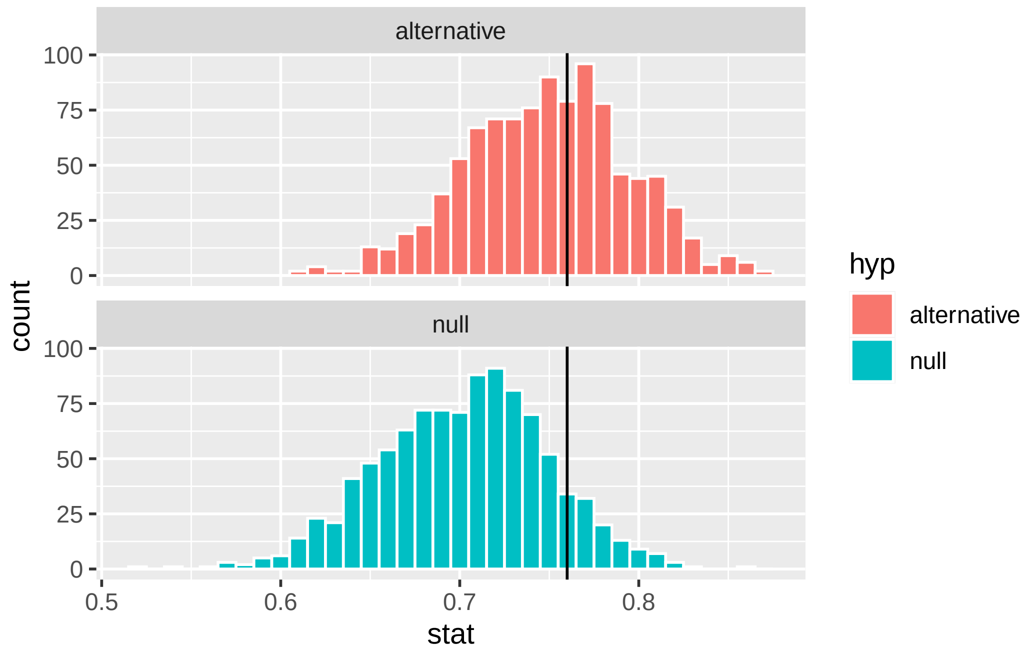
```
 90%
0.76
```

```r
# Generate alternative distribution
# Ha is p != 0.70
# We use the assumption that we will get a p of 0.75 in a sample
alt_dist <- dat %>%
  specify(response = favor, success = "Yes") %>%
  hypothesise(null = "point", p = 0.75) %>%
  generate(reps = 1000, type = "draw") %>%
  calculate(stat = "prop")
alt_dist
```

```
Response: favor (factor)
Null Hypothesis: point
# A tibble: 1,000 x 2
   replicate  stat
```

```
       <int> <dbl>
 1         1  0.82
 2         2  0.77
 3         3  0.75
 4         4  0.79
 5         5  0.73
 6         6  0.74
 7         7  0.78
 8         8  0.71
 9         9  0.71
10        10  0.74
# i 990 more rows
```

```
# Visualize piled distributions
dist_combo <- rbind(null_dist %>% mutate(hyp = "null"),
                    alt_dist %>% mutate(hyp = "alternative"))
ggplot(dist_combo, aes(x = stat, fill = hyp)) +
  geom_histogram(color = "white", binwidth = 0.01) +
  facet_wrap(~ hyp, nrow = 2) +
  geom_vline(xintercept = alpha_x)
```




```
power <- alt_dist %>%
  summarize(power = mean(stat >= alpha_x))
power
```

```
# A tibble: 1 x 1
  power
  <dbl>
1 0.458
```

```r
alt_dist$stat > alpha_x
```

```
  [1]  TRUE  TRUE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
 [13] FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
 [25] FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE
 [37]  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE
 [49]  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [61] FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
 [73] FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
 [85]  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE
 [97] FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE
[109] FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
[121] FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE
[133] FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE
[145] FALSE FALSE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE
[157] FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE
[169] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE
[181] FALSE FALSE  TRUE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
[193] FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE
[205] FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE
[217] FALSE  TRUE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE  TRUE FALSE
[229] FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE
[241] FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE
[253] FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE  TRUE FALSE
[265] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE
[277] FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[289] FALSE FALSE  TRUE  TRUE  TRUE FALSE  TRUE FALSE FALSE  TRUE  TRUE FALSE
[301] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE
[313] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE
[325] FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE
[337] FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
[349]  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE
[361]  TRUE  TRUE FALSE  TRUE  TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE
[373] FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE
[385]  TRUE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE
[397]  TRUE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE  TRUE  TRUE
[409]  TRUE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE
```

```
[421]  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE
[433] FALSE FALSE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE
[445] FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE  TRUE FALSE  TRUE FALSE FALSE
[457]  TRUE  TRUE FALSE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE
[469]  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[481] FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE
[493] FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE
[505] FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE
[517]  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE
[529]  TRUE  TRUE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE
[541] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE
[553] FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE FALSE FALSE
[565] FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE FALSE
[577] FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE
[589] FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE
[601]  TRUE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
[613]  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE
[625]  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE
[637] FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE
[649] FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE
[661] FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
[673] FALSE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE
[685] FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE FALSE
[697]  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE
[709] FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE  TRUE
[721] FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE
[733]  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE
[745]  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE FALSE
[757]  TRUE FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE
[769] FALSE FALSE FALSE  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE FALSE  TRUE
[781] FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE
[793] FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE
[805] FALSE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE
[817] FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE
[829] FALSE FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE
[841] FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE
[853] FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE
[865] FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE
[877] FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
[889] FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE  TRUE  TRUE
[901] FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE
[913] FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE  TRUE
[925] FALSE  TRUE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
```

```
 [937] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
 [949]  TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE
 [961] FALSE  TRUE FALSE FALSE  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE FALSE
 [973]  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE
 [985] FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE  TRUE
 [997] FALSE  TRUE FALSE  TRUE
```

b) What is the power if they increase the sample size to $n = 500$ and conduct a two-sided test with $H_A : p \neq 0.70$ at the $\alpha = 0.10$ significance level?

I don't know that a two-sided test here makes a lot of sense…

```
set.seed(2023)

# Construct data frame of sample results with 100 values
n <- 500
dat <- data.frame(favor = c(rep("Yes", 0.5*n), rep("No", 0.5*n)))

# Generate a null distribution matching this ratio
# H0 is p == 0.70
null_dist <- dat %>%
  specify(response = favor, success = "Yes") %>%
  hypothesise(null = "point", p = 0.70) %>%
  generate(reps = 1000, type = "draw") %>%
  calculate(stat = "prop")

# Generate alternative distribution
# Ha is p != 0.70
# We use the assumption that we will get a p of 0.75 in a sample
alt_dist <- dat %>%
  specify(response = favor, success = "Yes") %>%
  hypothesise(null = "point", p = 0.75) %>%
  generate(reps = 1000, type = "draw") %>%
  calculate(stat = "prop")
mean(alt_dist$stat)
```

```
[1] 0.749822
```

```
sd(alt_dist$stat)
```

```
[1] 0.01981608
```

```r
# Where are the two-sided critical values for alpha = 0.10?
crit.val.lt <- quantile(null_dist$stat, 0.05)
crit.val.ut <- quantile(null_dist$stat, 0.95)  # note how alpha is split in half for two-side
crit.val.lt
```

```
   5%
0.664
```

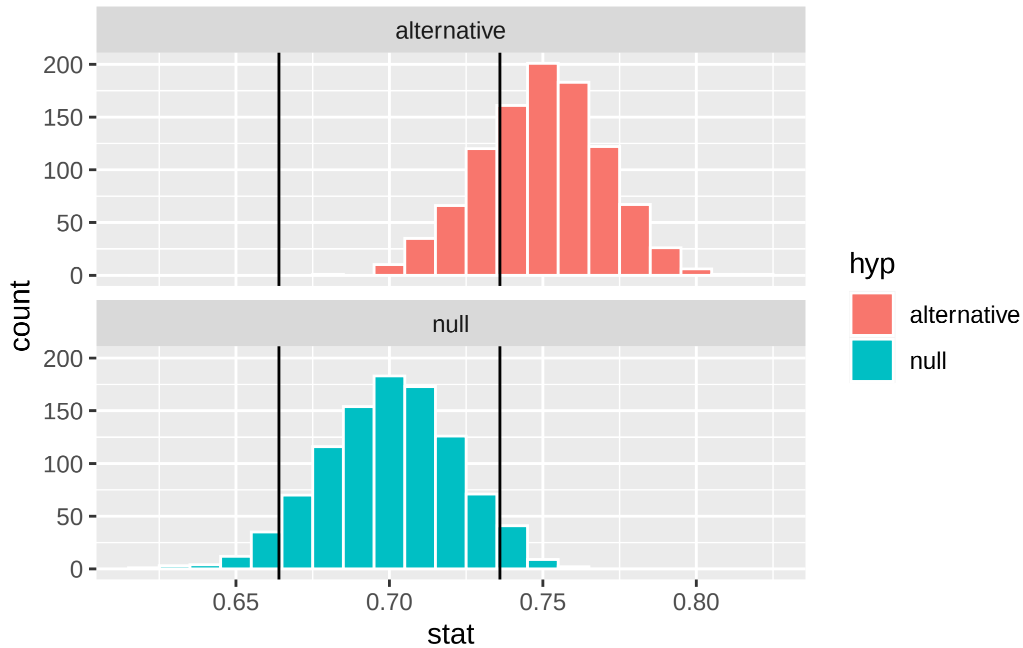
```r
crit.val.ut
```

```
  95%
0.736
```

```r
# Compute power
power <- alt_dist %>%
  summarize(power = mean(stat <= crit.val.lt) +
              mean(stat >= crit.val.ut))
power
```

```
# A tibble: 1 x 1
  power
  <dbl>
1 0.768
```

```r
# Visualize piled distributions
dist_combo <- rbind(null_dist %>% mutate(hyp = "null"),
               alt_dist %>% mutate(hyp = "alternative"))
ggplot(dist_combo, aes(x = stat, fill = hyp)) +
  geom_histogram(color = "white", binwidth = 0.01) +
  facet_wrap(~ hyp, nrow = 2) +
  geom_vline(xintercept = crit.val.lt) +
  geom_vline(xintercept = crit.val.ut)
```




```
threshold <- 3250
target_value <- 135
```

JLX TESTS: I am going to test here the question I tested in my JS:

Q: given the Saratoga housing data, test if the average price per square foot is below 135, for houses above 3250 sf. Use a significance level of 0.05.

```
large_houses <- SaratogaHouses %>%
  filter(livingArea > threshold) %>%
  mutate(price_per_sqft = price / livingArea)
large_houses
```

```
    price lotSize age landValue livingArea pctCollege bedrooms fireplaces
1  382500    4.08  13     75500       4534         64        6          2
2  625000    0.45  14    119500       5228         64        4          4
3  415000    0.58   9     86400       3358         64        4          1
4  412500    0.60  13     88000       3896         64        5          2
5  435000    1.00  25     25000       4211         57        5          2
6  310000    0.17 169    220000       3347         57        6          2
7  325000    0.23   5     73800       3313         57        4          1
8  496000    0.34   3     82400       3467         57        4          1
9  620000    1.06  14    125100       4856         57        5          2
```

10	500075	0.91	0	239300	3400	57	3	0
11	649000	1.04	10	192900	4128	57	3	2
12	449000	1.00	20	124800	3457	57	3	2
13	597185	1.07	0	193200	4210	57	4	1
14	535000	1.00	14	192500	3254	57	4	1
15	405000	0.61	6	23900	3296	63	4	1
16	420000	1.14	6	82100	3279	63	4	1
17	355465	0.35	0	233000	3328	63	4	1
18	460000	0.47	1	14100	3336	63	4	1
19	313635	1.93	16	131500	3824	63	3	2
20	355840	0.87	0	108900	3259	40	4	1
21	775000	0.48	31	72600	3968	62	5	4
22	650000	0.34	3	82400	3770	62	4	1
23	403040	0.49	0	233000	3320	64	4	1
24	317105	0.44	1	108900	3285	40	4	1
25	319000	0.47	1	108900	3285	40	4	1
26	314000	0.53	1	108900	3344	40	4	1
27	300000	0.56	34	30800	3604	64	6	1
28	508000	0.17	1	116700	3511	64	4	1
29	469900	1.40	2	74800	3422	62	4	1
30	422680	0.16	176	46200	4486	51	6	1

	bathrooms	rooms	heating	fuel	sewer	waterfront
1	2.5	12	hot air	oil	septic	No
2	4.0	12	hot air	gas	public/commercial	No
3	3.5	12	hot air	gas	public/commercial	No
4	4.5	12	hot water/steam	gas	public/commercial	No
5	3.5	12	hot water/steam	gas	septic	No
6	2.5	12	hot water/steam	gas	public/commercial	No
7	2.5	12	hot air	gas	public/commercial	No
8	2.5	11	hot air	gas	public/commercial	No
9	4.0	12	hot air	oil	septic	No
10	3.0	12	hot air	gas	public/commercial	Yes
11	3.5	12	hot air	gas	septic	No
12	2.5	12	electric	electric	septic	No
13	3.5	12	hot air	gas	septic	No
14	2.5	12	hot air	gas	septic	No
15	2.5	12	hot air	gas	public/commercial	No
16	3.0	11	hot air	gas	public/commercial	No
17	2.5	12	hot air	gas	public/commercial	No
18	3.5	12	hot air	gas	public/commercial	No
19	4.0	12	hot air	gas	septic	No
20	2.5	10	hot air	gas	public/commercial	No
21	3.5	12	hot air	gas	public/commercial	No

```
22       2.5    12         hot air      gas public/commercial         No
23       2.5    12         hot air      gas public/commercial         No
24       2.5    10         hot air      gas public/commercial         No
25       3.0    12         hot air      gas public/commercial         No
26       3.5    11         hot air      gas public/commercial         No
27       3.5    12 hot water/steam      gas public/commercial         No
28       2.5    12         hot air      gas public/commercial         No
29       4.0    12         hot air      gas public/commercial         No
30       4.0    12 hot water/steam      gas public/commercial         No
   newConstruction centralAir price_per_sqft
1               No        Yes       84.36259
2               No        Yes      119.54858
3               No        Yes      123.58547
4               No        Yes      105.87782
5               No         No      103.30088
6               No         No       92.62026
7               No        Yes       98.09840
8               No        Yes      143.06317
9               No        Yes      127.67710
10              No        Yes      147.08088
11              No        Yes      157.21899
12              No        Yes      129.88140
13              No        Yes      141.84917
14              No        Yes      164.41303
15              No        Yes      122.87621
16              No        Yes      128.08783
17             Yes        Yes      106.81040
18              No        Yes      137.88969
19              No        Yes       82.01752
20             Yes         No      109.18687
21              No        Yes      195.31250
22              No        Yes      172.41379
23             Yes        Yes      121.39759
24              No         No       96.53120
25             Yes         No       97.10807
26             Yes         No       93.89952
27              No         No       83.24084
28              No         No      144.68812
29             Yes        Yes      137.31736
30              No         No       94.22202
```

```r
test_stat <- large_houses %>%
  specify(response = price_per_sqft) %>%
  calculate(stat = "mean")
test_stat
```

```
Response: price_per_sqft (numeric)
# A tibble: 1 x 1
   stat
  <dbl>
1  122.
```

```r
null_dist <- large_houses %>%
  specify(response = price_per_sqft) %>%
  hypothesize(null = "point", mu = target_value) %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "mean")
null_dist
```

```
Response: price_per_sqft (numeric)
Null Hypothesis: point
# A tibble: 1,000 x 2
   replicate  stat
       <int> <dbl>
 1         1  144.
 2         2  136.
 3         3  128.
 4         4  140.
 5         5  141.
 6         6  138.
 7         7  141.
 8         8  140.
 9         9  143.
10        10  135.
# i 990 more rows
```

```r
min(null_dist$stat)
```

```
[1] 118.1952
```

```r
max(null_dist$stat)
```

```
[1] 150.847
```

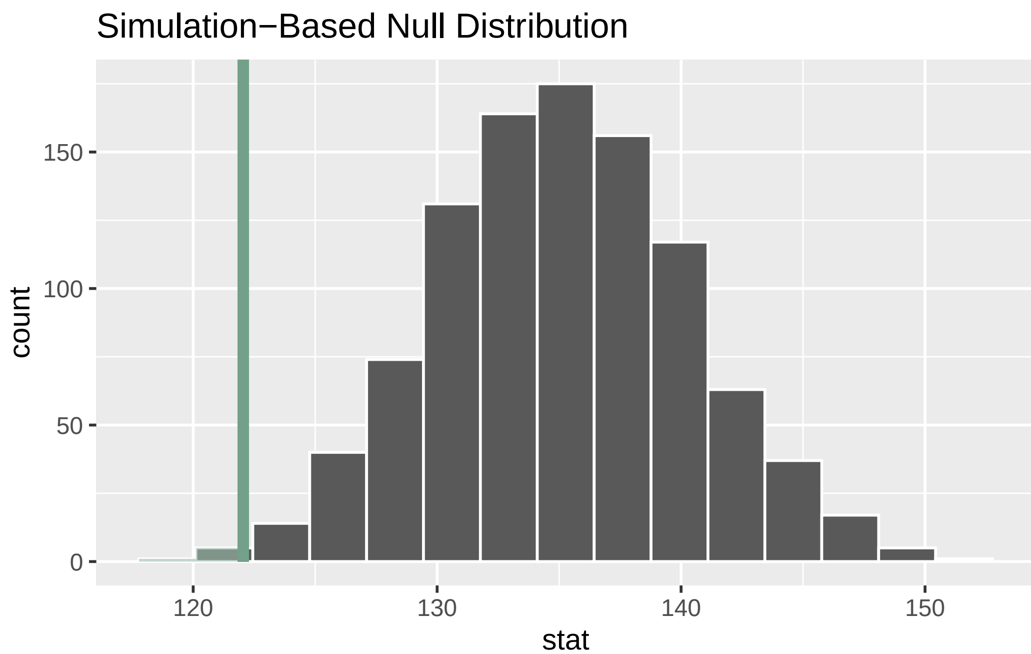
```r
mean(null_dist$stat)
```

```
[1] 135.1352
```

```r
sd(null_dist$stat)
```

```
[1] 5.191333
```

```r
null_dist %>%
  visualise() +
  shade_p_value(test_stat, direction = "less",
                col = wes_green, fill = wes_green_pale)
```




```r
p_value <- null_dist %>%
  get_p_value(obs_stat = test_stat,
              direction = "less")
p_value
```

```
# A tibble: 1 x 1
  p_value
    <dbl>
1   0.003
```