PS09

Aditya Sanjay Mhaske

2023-03-27

Q1) i)

Null hypothesis H_0 : <= 0.02 H_1 : > 0.02

```
library(infer)
```

```
f = data.frame(frc = c(rep("diseased", 40), rep("non-diseased", 960)))
```

citation: “032323_lab_infer.R”

```
null_sim_p <- f |>
  specify(response = frc, success = "diseased") |>
  hypothesize(null = "point", p = 0.02) |>
  generate(reps = 10000, type = "draw") |>
  calculate(stat = "prop")

p_st <- f |>
  observe(response = frc, success = "diseased", stat = "prop")
p_st
```

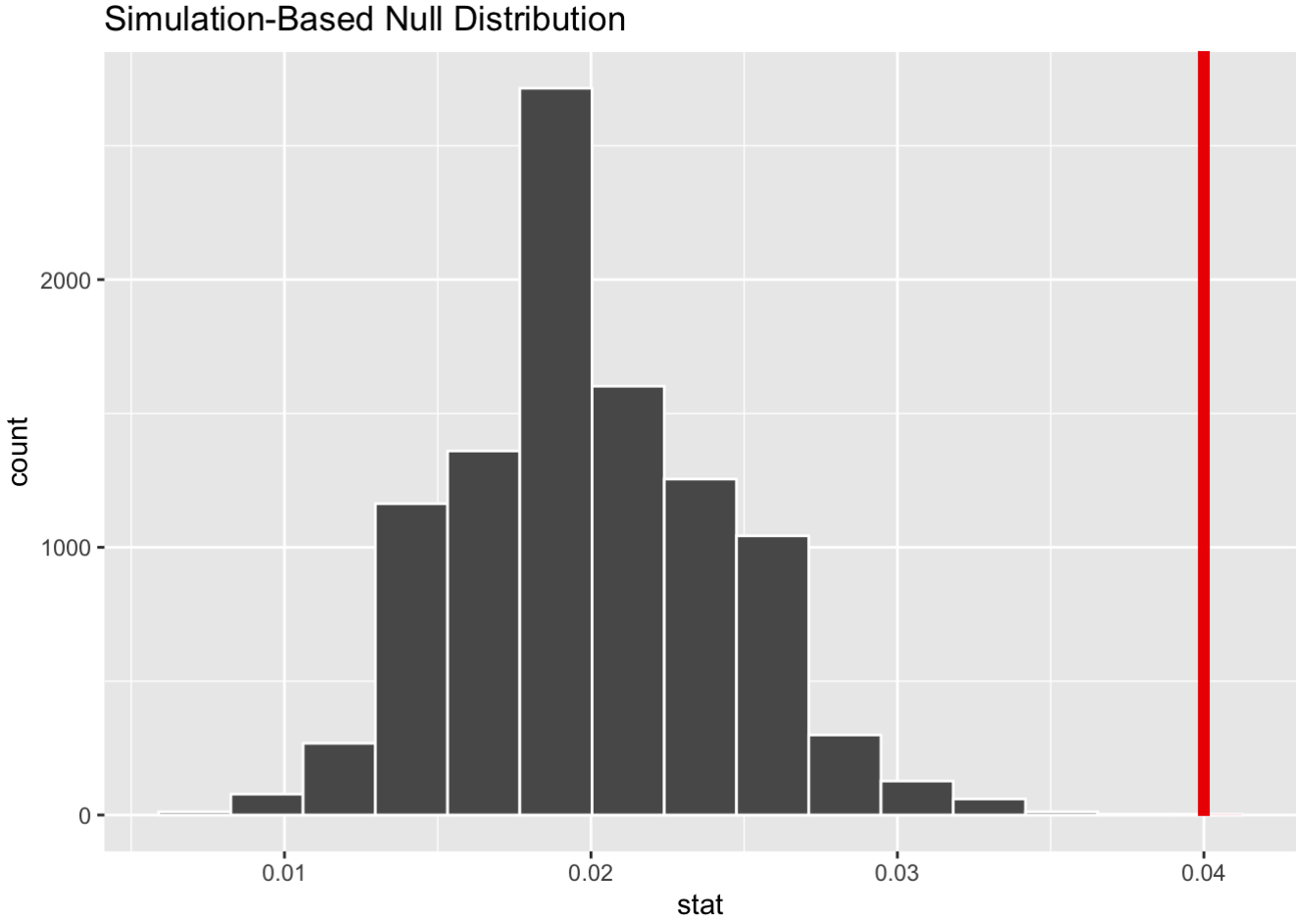
```
## Response: frc (factor)
## # A tibble: 1 × 1
##    stat
##   <dbl>
## 1  0.04
```

```
null_sim_p |>
  get_p_value(obs_stat = p_st, direction = "right")
```

```
## 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.
```

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

```
visualize(null_sim_p, method = "simulation") +
  shade_p_value(obs_stat = p_st, direction = "right")
```



ii)

```
n = 100
p = 0.04
mu = 0.02
x_bar = 0.04



v = p * (1 - p)
v
```

```
## [1] 0.0384
```

```
s = sqrt(v)
s
```

```
## [1] 0.1959592
```

```
z = (x_bar - mu)/(s/sqrt(n))
z
```

```
## [1] 1.020621
```

```
1 - pnorm(z)
```

```
## [1] 0.1537171
```

iii)

```
f = data.frame(frc = c(rep("diseased", 4), rep("non-diseased", 96)))
```

reference - “032323_lab_infer.R”

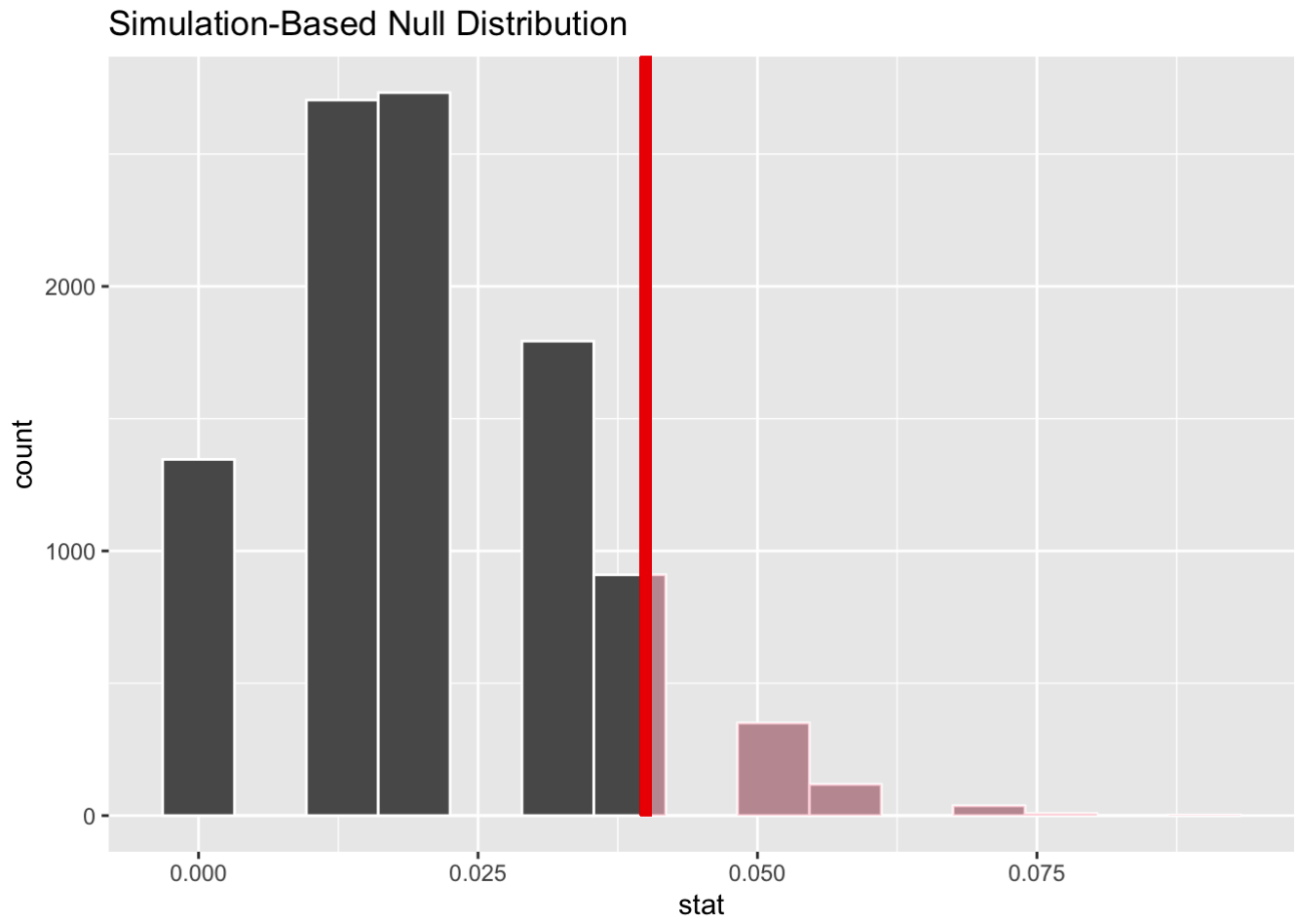
```
null_sim_p <- f |>
  specify(response = frc, success = "diseased") |>
  hypothesize(null = "point", p = 0.02) |>
  generate(reps = 10000, type = "draw") |>
  calculate(stat = "prop")

p_st <- f |>
  observe(response = frc, success = "diseased", stat = "prop")
p_st
```

```
## Response: frc (factor)
## # A tibble: 1 × 1
##    stat
##   <dbl>
## 1  0.04
```

```
null_sim_p |>
  get_p_value(obs_stat = p_st, direction = "right")
```

```
## # A tibble: 1 × 1
##   p_value
##     <dbl>
## 1   0.142
```

```
visualize(null_sim_p, method = "simulation") +
  shade_p_value(obs_stat = p_st, direction = "right")
```



Q2

$H_0 : \mu = 0$ $H_1 : \mu! = 0$

```
x_bar = -0.1833
s = 5.18633

#t value
z = (-0.1833 - 0) / (5.18633 / sqrt(60))
z
```

```
## [1] -0.273765
```

```
#p value
p_val <- 2 * pt(abs(z), 59)
p_val
```

```
## [1] 1.214779
```

the p-value is greater than alpha value 0.05 as well as 0.25. Thus, we can't reject the null hypothesis.

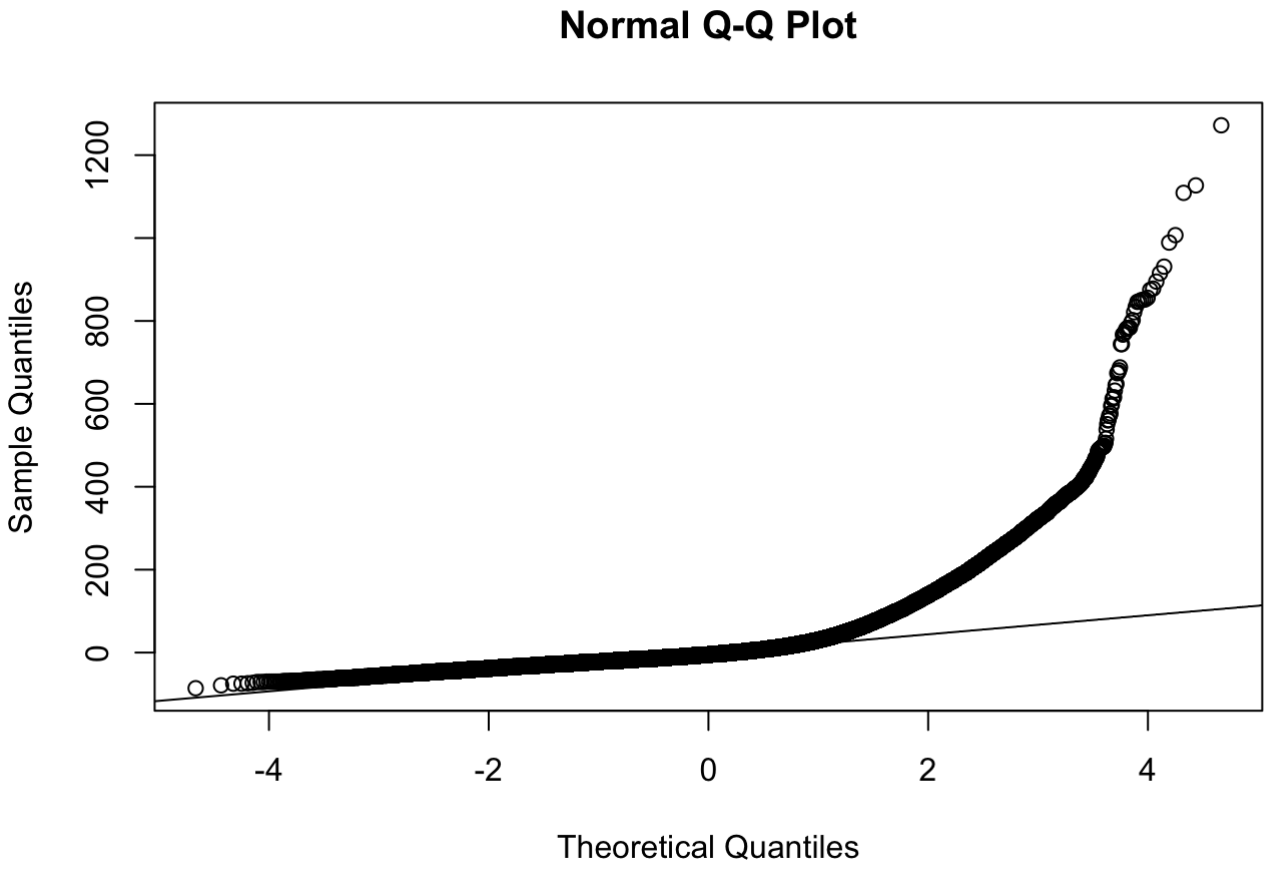
Q3

a)

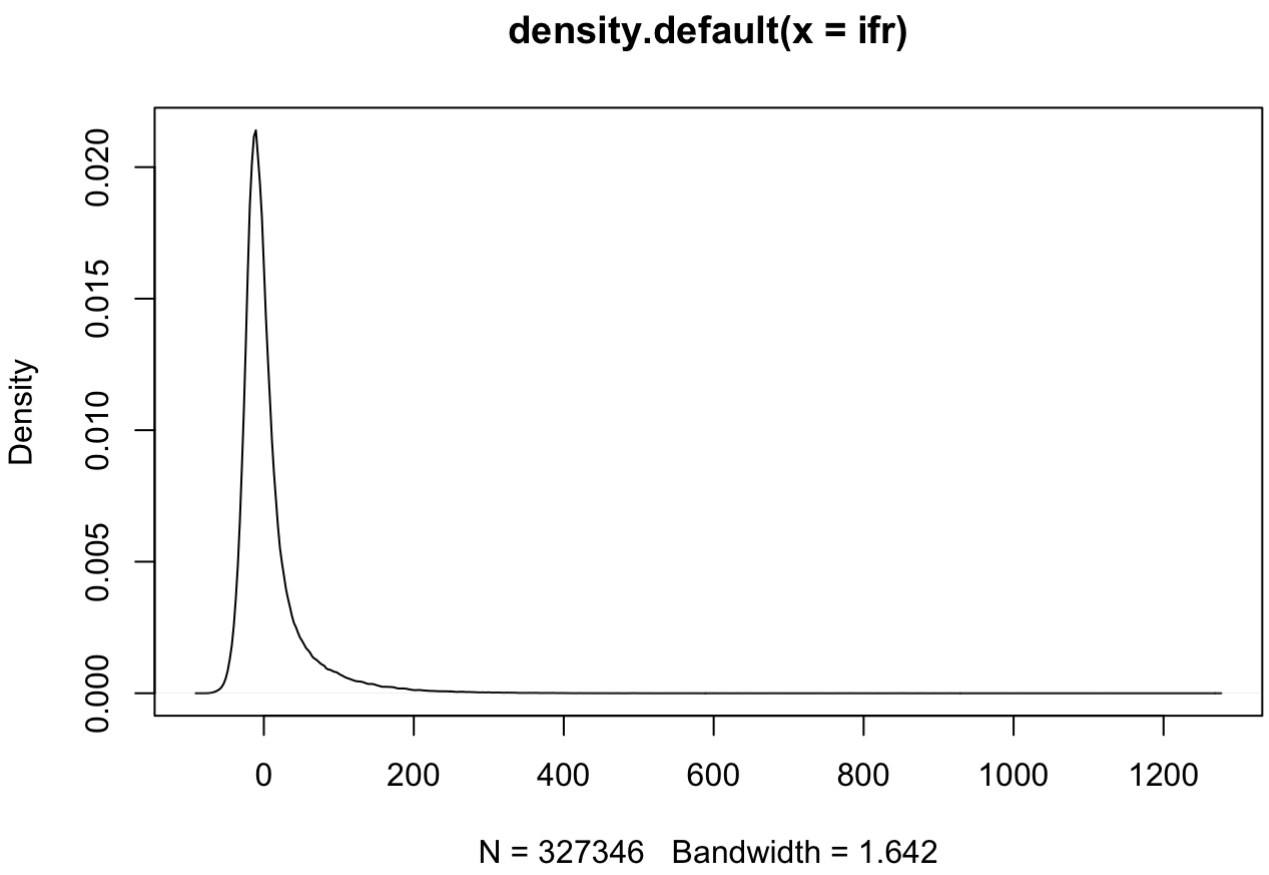
```
library(nycflights13)

ifr = na.omit(flights$arr_delay)
```

b)

```
qqnorm(ifr)
qqline(ifr)
```




```
plot(density(ifr))
```



The data does not seem to be drawn from a Normal Distribution.

c)

```
set.seed(520)

nc = sample(ifr, 100, F)
```

H_0 : >= 10
H_1 : < 10

```
nc_bar = mean(nc)
nc_s = sqrt(var(nc)/100)
```

test statistic

```
t_str = (nc_bar - 10)/nc_s
t_str
```

```
## [1] -2.474439
```

p - value

```
pt(t_str, 100 - 1)
```

```
## [1] 0.007522487
```

p - value = 0.0075. = Alternate hypothesis is true. Therefore, average arrival time delay > 10 mins.

d)

```
dly = nc == 0

dly_bar = mean(dly)
```

test statistic:

```
t_st = (dly_bar - 0.5) / sqrt(0.5*(1 - 0.5)/100)
t_st
```

```
## [1] -9.2
```

p-value:

```
1 - pnorm(t_st)
```

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

Q4

```
# a)
```

Citation:“032323_lab_infer.R”

```
null_sim_t <- flights |>
  specify(response = arr_delay) |>
  hypothesize(null = "point", mu = 10) |>
  generate(reps = 100, type = "bootstrap") |>
  calculate(stat = "t")
```

```
## Warning: Removed 9430 rows containing missing values.
```

test statistic

```
t_bar <- flights |>
  observe(response = arr_delay, null = "point", mu = 10, stat = "t")
```

```
## Warning: Removed 9430 rows containing missing values.
```

```
t_bar
```

```
## Response: arr_delay (numeric)
## Null Hypothesis: point
## # A tibble: 1 × 1
##    stat
##   <dbl>
## 1 -39.8
```

P-value

```
null_sim_t |>
  get_p_value(obs_stat = t_bar, direction = "two.sided")
```

```
## 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.
```

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

```
# b)
```

```
null_sim_t <- flights |>
  specify(response = arr_delay) |>
  hypothesize(null = "point", mu = 0.5) |>
  generate(reps = 100, type = "bootstrap") |>
  calculate(stat = "t")
```

```
## Warning: Removed 9430 rows containing missing values.
```

test statistic

```
t_bar <- flights |>
  observe(response = arr_delay, null = "point", mu = 0.5, stat = "t")
```

```
## Warning: Removed 9430 rows containing missing values.
```

```
t_bar
```

```
## Response: arr_delay (numeric)
## Null Hypothesis: point
## # A tibble: 1 × 1
##    stat
##   <dbl>
## 1  82.0
```

P-value

```
null_sim_t |>
  get_p_value(obs_stat = t_bar, direction = "two.sided")
```

```
## 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.
```

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