

Hypothesis Testing

Olivia Bryant

2023-10-26

Hypothesis Testing

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr  1.0.1
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.3.0      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(MASS)

##
## Attaching package: 'MASS'
##
## The following object is masked from 'package:dplyr':
##
##      select

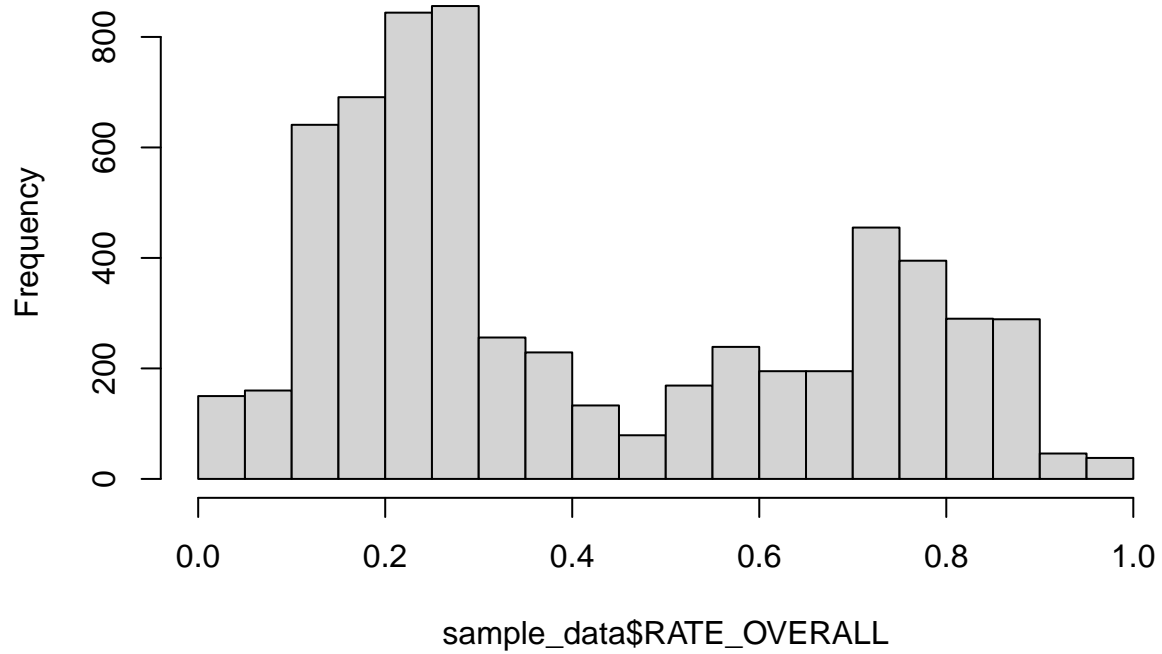
sample_data <- read.csv("choropleth_sample_data.csv")
head(sample_data)

##   PRACTICE_CODE RATE_OVERALL HOMELESS_GP
## 1      G82107      0.999           1
## 2      G83026      0.998           0
## 3      H82012      0.998           1
## 4      A82646      0.997           0
## 5      H81044      0.997           1
## 6      H81113      0.996           0
```

Visualise the Data

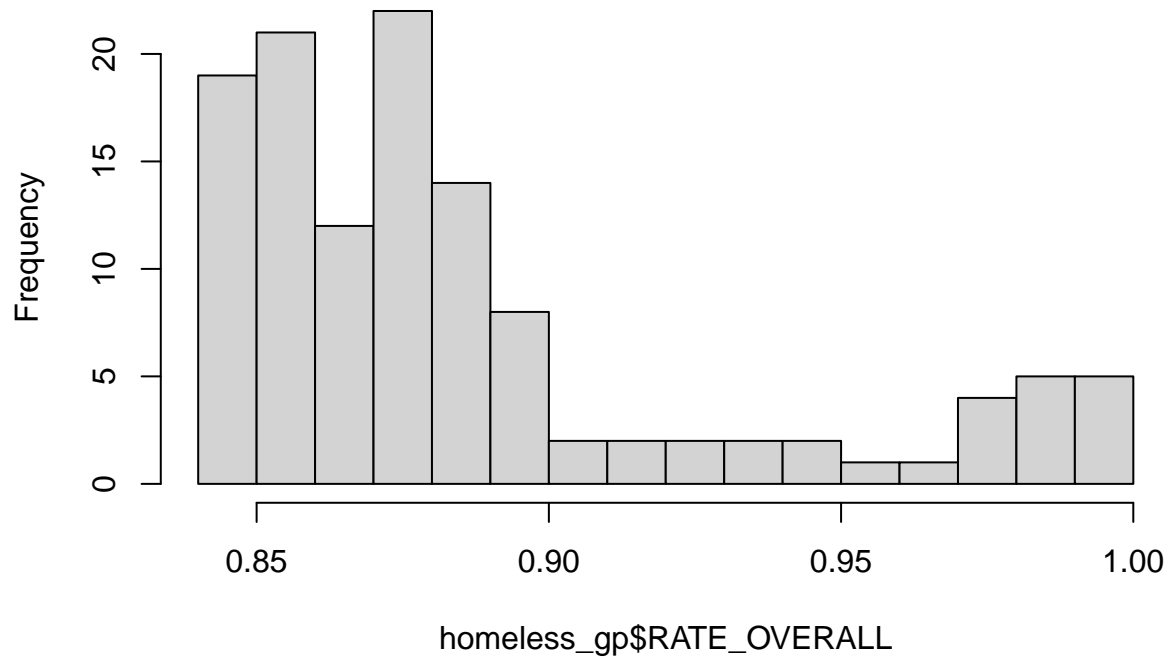
```
# visualise prescribing rates for all GPS
hist(sample_data$RATE_OVERALL, breaks = 15)
```

Histogram of sample_data\$RATE_OVERALL



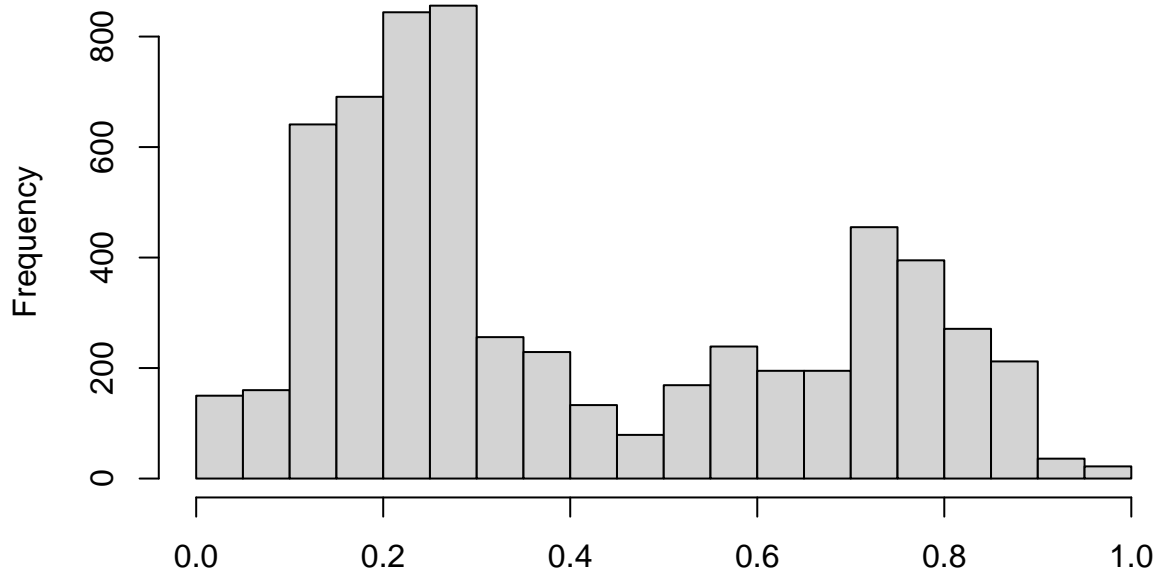
```
# visualise the distribution for homeless GPs  
homeless_gp <- sample_data %>%  
  filter(HOMELESS_GP == 1)  
hist(homeless_gp$RATE_OVERALL, breaks=20)
```

Histogram of homeless_gp\$RATE_OVERALL



```
# visualise the distribution for non-homeless GPs
nonhomeless_gp <- sample_data %>%
  filter(HOMELESS_GP == 0)
hist(nonhomeless_gp$RATE_OVERALL, breaks=20)
```

Histogram of nonhomeless_gp\$RATE_OVERALL



nonhomeless_gp\$RATE_OVERALL

```
# create box plots for the two groups
boxplot(RATE_OVERALL ~ HOMELESS_GP, data = sample_data,
  col = c("#FFE0B2", "#F57C00"), names=c("Non-Homeless GP", "Homeless GPs"))
```



###

T-Test

```
t.test(RATE_OVERALL ~ HOMELESS_GP, data = sample_data, var.equal = FALSE)
```

```
##  
## Welch Two Sample t-test  
##  
## data: RATE_OVERALL by HOMELESS_GP  
## t = -95.237, df = 343.44, p-value < 2.2e-16  
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0  
## 95 percent confidence interval:  
## -0.4977563 -0.4776124  
## sample estimates:  
## mean in group 0 mean in group 1  
## 0.3995206 0.8872049
```

Wilcoxon Rank Sum Test

```
wilcox.test(RATE_OVERALL ~ HOMELESS_GP, data = sample_data)
```

```
##  
## Wilcoxon rank sum test with continuity correction  
##  
## data: RATE_OVERALL by HOMELESS_GP  
## W = 20056, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0
```