Quiz 2.3

eCornell

9/3/2021

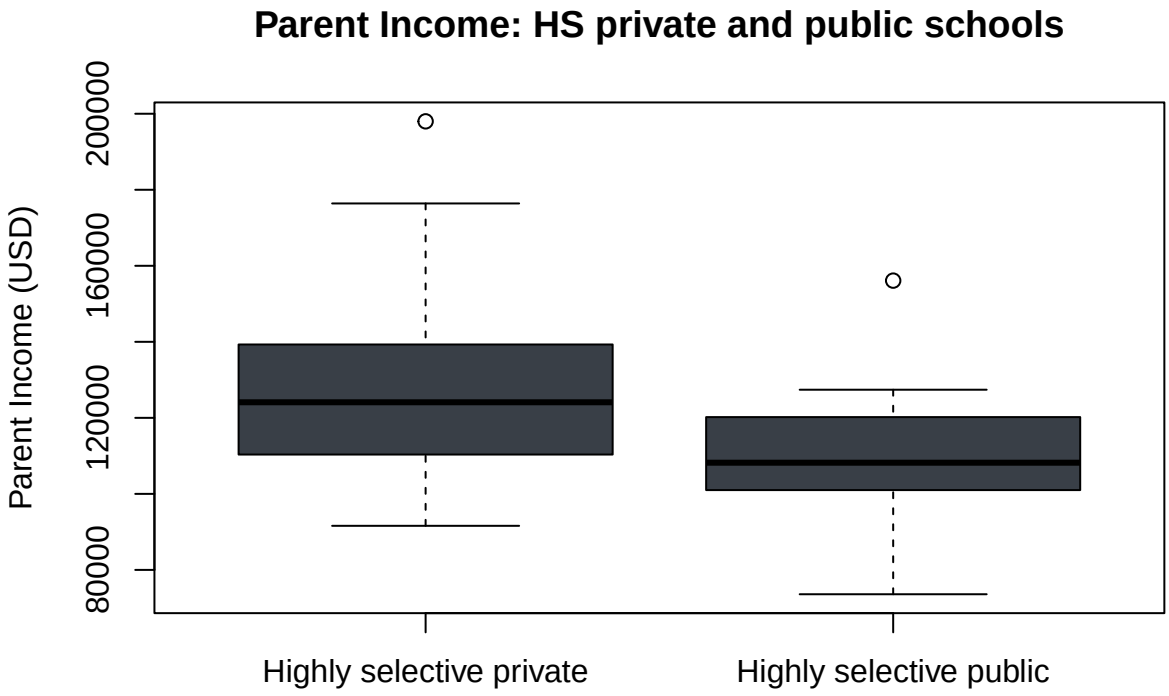
```
# eCornell Hex Codes:
crimson = '#b31b1b' #Crimson
lightGray = '#cecece' #lightGray
darkGray = '#606366'
skyBlue = '#92b2c4' #skyblue
gold = '#fbb040' #gold
ecBlack = '#393f47' #ecBlack

# Load the data.
school = read.csv('mrc_table2.csv', header = TRUE, check.names = FALSE)
school = school[,names(school) %in%
                  c('name', 'type', 'tier', 'tier_name', 'mr_kq5_pq1',
                    'par_median', 'k_median')]
names(school)[5:7] <- c("mobility_rate", "parent_income", "student_income")

## Calculate the mean parent income for students at highly selective private schools:
school = school[school$tier_name %in% c("Highly selective private", "Highly selective public"),]
par.income.prv = school$parent_income[school$tier_name == 'Highly selective private']
mean(par.income.prv)
```

```
## [1] 126631
```

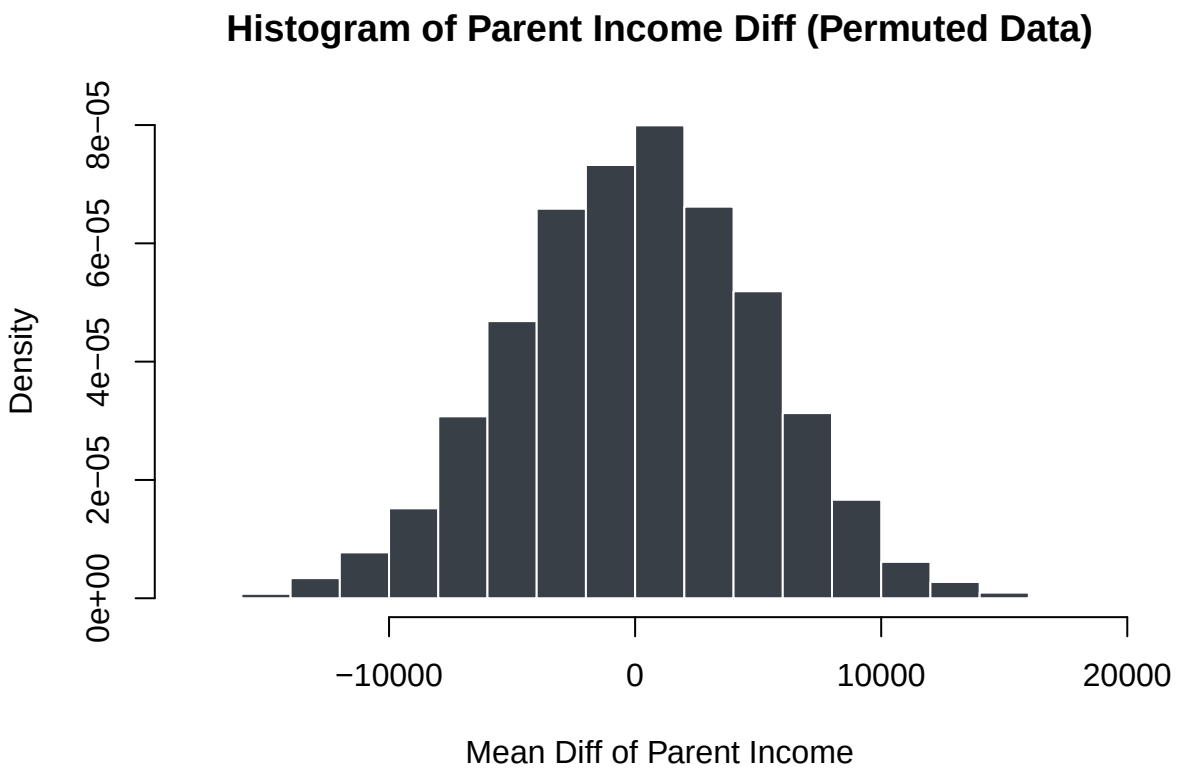
```
## Calculate the mean parent income for students at highly selective public schools.
par.income.pub = school$parent_income[school$tier_name == 'Highly selective public']
mean(par.income.pub)
```

```
## [1] 109753.8
```

```
# Create the boxplot:
boxplot(par.income.prv, par.income.pub, names = c('Highly selective private', 'Highly selective public'
        main = 'Parent Income: HS private and public schools', col = ecBlack)
```




```r
# Create a permutation distribution:
set.seed(1)
P = 10000
store_mean_diff = rep(0, P)

for (n in 1:P){
  school.perm = school
  school.perm$parent_income = sample(school.perm$parent_income, replace = FALSE)
  school.perm.prv = school.perm$parent_income[school.perm$tier_name == 'Highly selective private']
  school.perm.pub = school.perm$parent_income[school.perm$tier_name == 'Highly selective public']

  store_mean_diff[n] = mean(school.perm.prv) - mean(school.perm.pub)
}

# Plot the observed sample statistic on the histogram:

hist(store_mean_diff, breaks = 20, freq = FALSE, col = ecBlack, border = 'white',
     xlab = 'Mean Diff of Parent Income',
     main = 'Histogram of Parent Income Diff (Permuted Data)')
```




```
# Observed statisic
obs_stat = mean(par.income.prv)-mean(par.income.pub)
obs_stat
```

```
## [1] 16877.14
```

```
# Calculate p-value
mean(abs(store_mean_diff) > abs(obs_stat))
```

```
## [1] 2e-04
```