PBA W10 Extra Credit Exercises (II)

By: Kevin Karnadi Kirmansjah 紀維鑫 - 109006241

Import libraries

```
library(data.table)
library(knitr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:data.table':
##
##     between, first, last
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Question 1

Load the judges.csv into a data frame called judges. Create a cross-tab (of proportions, not counts) of judge gender and whether the appointing president was Republican or Democrat. Store this table as gender_rep_table. Use knitr::kable() to create a nicely formatted version of this table. In your response, answer the following questions:

1. How many judges are in this data set?
2. What proportion of the judges are men?
3. Is the party composition different for male and female judges?

Load data

```
judges = fread("http://bit.ly/3UZ5fjf")
```

```
gender_rep_table = table(judges$woman, judges$republican)/length(judges$republican)
colnames(gender_rep_table) = c("Democrat", "Republican")
rownames(gender_rep_table) = c("Man", "Woman")
```

Proportion table of judge gender and whether the appointing president was Republican or Democrat:

```
knitr::kable(gender_rep_table)
```

	Democrat	Republican
Man	0.3360656	0.5081967
Woman	0.1106557	0.0450820

1. How many judges are in this data set?

```
print(nrow(judges))
```

```
## [1] 244
```

2. What proportion of the judges are men?

```
print(sum(gender_rep_table["Man",]))
```

```
## [1] 0.8442623
```

3. Is the party composition different for male and female judges?

> Yes, they are different as can be seen from the table in 1.

Question 2

The outcome of interest is the proportion of feminist rulings on issues related to gender, progressive_vote. Create a well-formatted histogram of this variable (set bin = 10) and provide a written summary of this graph. Roughly speaking, where is the region of highest density of this variable?

```
hist(judges$progressive_vote, breaks=10, main="Histogram of 'progressive_vote'", xlab="progre
ssive_vote", ylab="Frequency")
```

Histogram of 'progressive_vote'
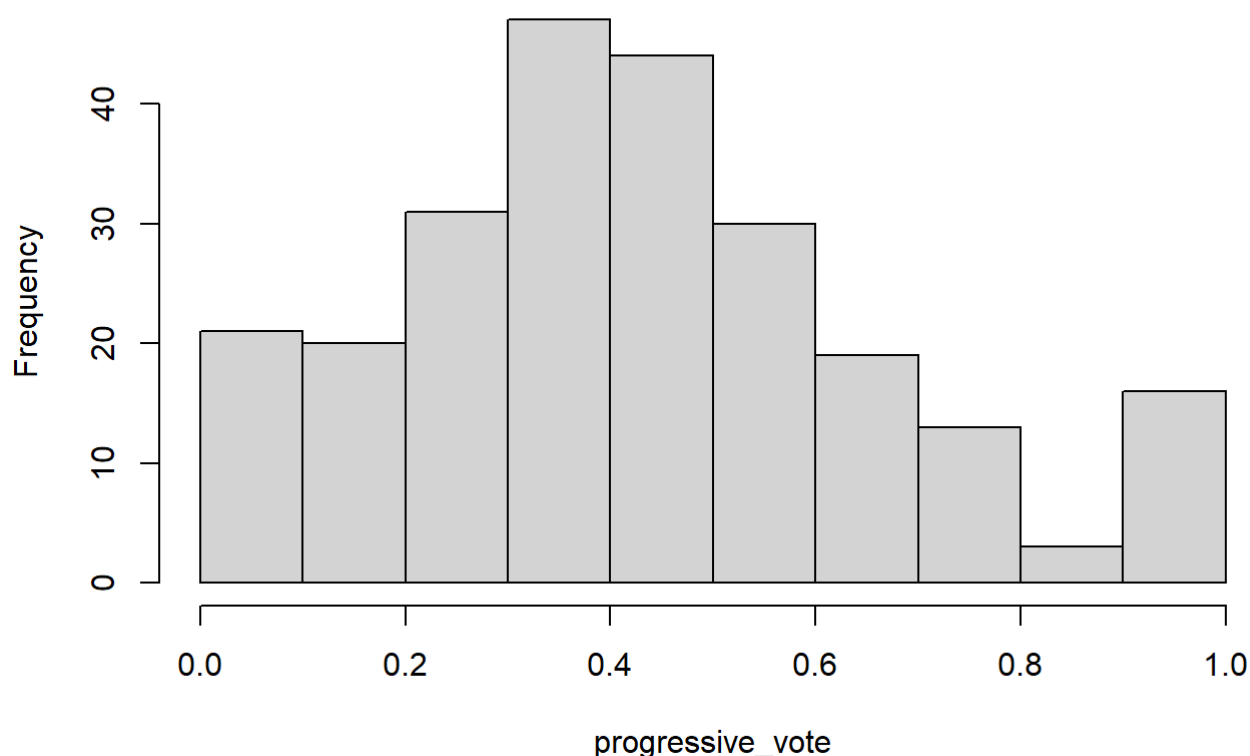
The distribution of the 'progressive_vote' variable, as can be seen from the histogram, is slightly skewed to the right. The region of the highest density of the 'progressive_vote' variable is located in the values of 0.3-0.4.

Question 3

Next, we consider differences between some groups. Create a new factor variable called judges$gender_party which indicates each of the four following groups:

- "F_Demo" for female judges appointed by Democratic presidents.

- "F_Repub" for female judges appointed by Republican presidents.

- "M_Demo" for male judges appointed by Democratic presidents.

- "M_Repub" for male judges appointed by Republican presidents.

Use tapply() to calculate the mean of progressive_vote in each of these groups and store this vector as gender_party_means.

Plot these means using a barplot then interpret the results of the analysis. Does anything stand out to you? Does it appear that party association, gender, or both contribute to progressive voting patterns? If so, how?

```
judges = mutate(judges,
            gender_party = ifelse(woman == 1 & republican == 0, "F_Demo",
                            ifelse(woman == 1 & republican == 1, "F_Repub",
                            ifelse(woman == 0 & republican == 0, "M_Demo",
                            ifelse(woman == 0 & republican == 1, "M_Repub",
                "no")))))
judges = judges[, gender_party:=as.factor(gender_party)]
head(judges)
```

```
##                       name child circuit girls progressive_vote race religion
## 1:     Alarcon, Arthur L.     3       9     1        0.0000000     3        4
## 2:        Aldisert, Ruggero     3       3     1        0.6666667     1        4
## 3:          Aldrich, Bailey     2       1     0        0.3333333     1        1
## 4: Alito, Samuel A., Jr.      2       3     1        0.5000000     1        4
## 5:      Altimari, Frank X.    4       2     1        0.5000000     1        4
## 6:          Ambro, Thomas    NA       3    NA        0.0000000    NA       NA
##    republican sons woman yearb gender_party
## 1:          0    2     0  1925       M_Demo
## 2:          0    2     0  1919       M_Demo
## 3:          1    2     0  1907      M_Repub
## 4:          1    1     0  1950      M_Repub
## 5:          1    3     0  1928      M_Repub
## 6:          0   NA     0    NA       M_Demo
```
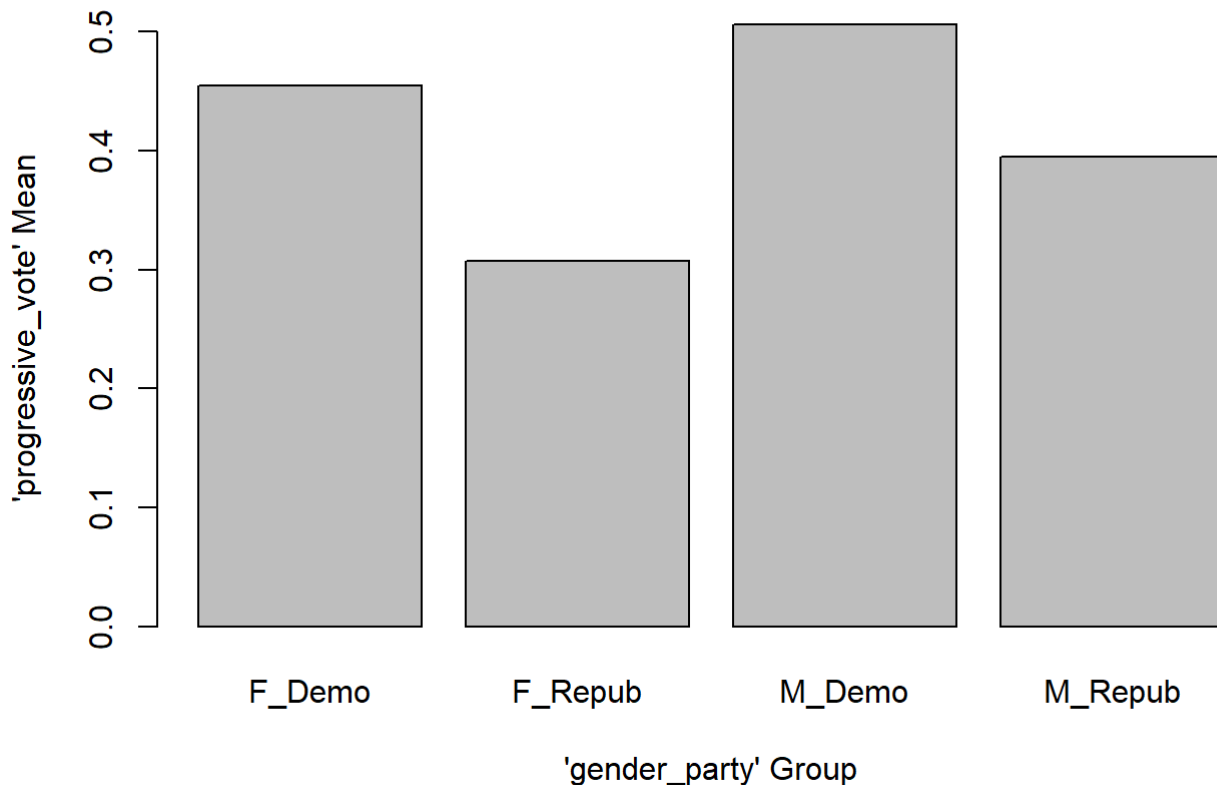
```
gender_party_means = tapply(judges$progressive_vote, judges$gender_party, mean)
gender_party_means
```

```
##    F_Demo    F_Repub    M_Demo    M_Repub
## 0.4547162 0.3069867 0.5062359 0.3952614
```

```
barplot(gender_party_means, main="Bar Plot of 'progressive_vote' Mean Values", xlab="'gender_
party' Group", ylab="'progressive_vote' Mean")
```

Bar Plot of 'progressive_vote' Mean Values

The means of the judges appointed by Democratic presidents are higher than the judges appointed by Republican presidents, regardless of their genders. Also, when comparing within the same party, the means of the Male judges are higher than the Female judges.

Question 4

What is the difference in the proportion of pro-feminist decisions between judges who have at least one daughter and those who do not have any? To compute this difference, first create a variable called judges$any_girls that is 1 when the judge has at least 1 girl and 0 otherwise. Then, create a subset of the data called parents that contains judges that have at least one child. Create an object called ATE that is the difference in means of progressive_vote between judges that have at least one girl versus those that have no girls among those judges with any children.

Can we interpret the result causally (i.e., can we safely say that the difference, if there is any, is caused by the judges having a daughter)? Why might we worry about interpreting the result causally, considering number of children as a possible confounder?

```
judges = mutate(judges, any_girls = ifelse(girls == 0 | is.na(girls), 0, 1))
parents = judges[child > 0,]
parents_with_girls = parents[any_girls == 1,]
parents_without_girls = parents[any_girls == 0,]
ATE = mean(parents_with_girls$progressive_vote) - mean(parents_without_girls$progressive_vot
e)
print(ATE)
```

```
## [1] 0.05416231
```

Ans: No, we can't interpret the result from the difference in means of progressive_vote between judges that have at least one girl versus those that have no girls among those judges with any children of pro causally, because as we can see from the difference in mean, it means that there exists a correlation, but from a correlation, we cannot determine whether they have a causation or not, so we cannot say that we can interpret the result causally.