Lab 10 Assignment

Dr. Purna Gamage

11/1/2020

Problem 1: Independence of Religion and View on Death Penalty.

- Religion is not obvious and we cannot easily sample from different religious groups. Therefore we do not consider this a test about homogeneity of populations.
- Permutation test
- Try χ^2 test

Examine relation between "religion" and "views of the death penalty" using Chi square test. Use gss2002 data.

a. Create a two-way table and examine the data.

b. Delete all rows with NAs for the two attributes. then create a new two way table with cleaned data.

c. Write a function to calculate chi square statistic. Then calculate the chi square statistic for the table you created above.

d. Write a function to make random permutation, make table, and compute test statistic. Try it out for the cleaned data.

e. Simulate null distribution (10000 simulations uding the permutation function you used in part d). Plot the histogram of the null distribution and draw the red line where you observed the chi-square test statistic.

f. find the p-value. What is your Conclusion?

The observed value is somewhere to the right but not too far. In fact 9% of null distribution values are larger. do not contadict the null hypothesis that views of the death penalty are independent of releigion.

f. Redo with chi squared test (use chisq.test()). Do you get the same conclusion?

Problem 2:

load the Verizon Data and examine the data (we have done this in last week lab)

```
Verizon <- read.csv("Verizon.csv")

head(Verizon)
```

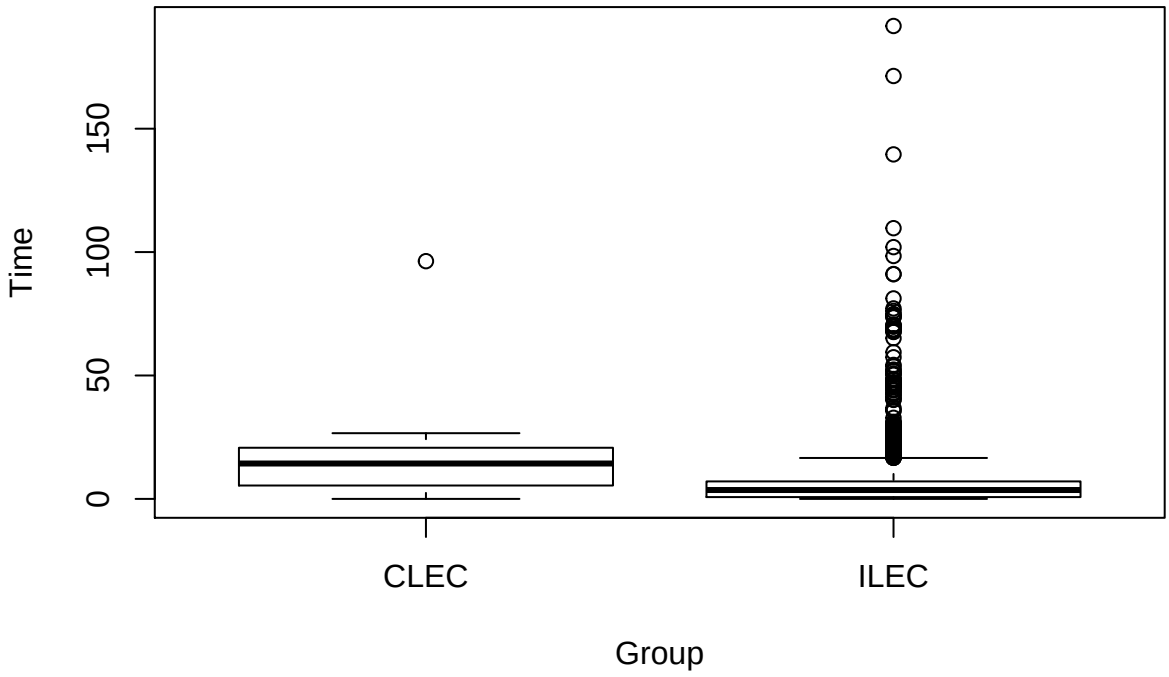
```
##    Time Group
## 1 17.50  ILEC
## 2  2.40  ILEC
## 3  0.00  ILEC
## 4  0.65  ILEC
## 5 22.23  ILEC
## 6  1.20  ILEC
```

Examine the data

```
table(Verizon$Group)
```

```
##
## CLEC ILEC
##   23 1664
```

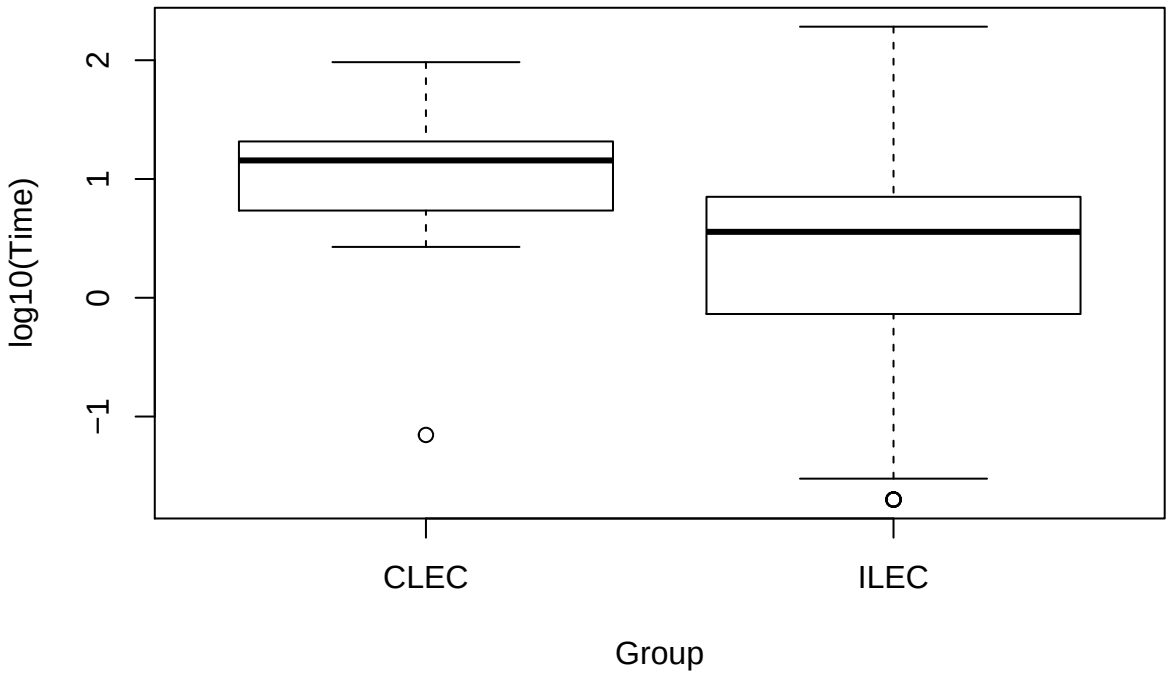
```
boxplot(Time ~ Group, data = Verizon)
```




```
boxplot(log10(Time) ~ Group, data = Verizon)
```

```
## Warning in bplt(at[i], wid = width[i], stats = z$stats[, i], out = z$out[z$group
## == : Outlier (-Inf) in boxplot 1 is not drawn
```

```
## Warning in bplt(at[i], wid = width[i], stats = z$stats[, i], out = z$out[z$group
## == : Outlier (-Inf) in boxplot 2 is not drawn
```



Means by customer group

```
aggregate(Time ~ Group, data = Verizon, FUN = mean)
```

```
##   Group      Time
## 1  CLEC 16.509130
## 2  ILEC  8.411611
```

a. Make a function to compute the difference of two means. Using that function,calculate the *observed difference of means*.

(Hint: Assume the data frame has the structure of the Verizon data and use the aggregate() function to calculate the means.)

b. Make a perumtation of the labels and return the difference of means of the permuted data. (use the function that you creatred in part b)

c. Repeat this many times(1000). Make histogram of the differences of means. Draw vertical line where the *observed difference of means* (calculated this in part a) is.

d. Compute p-value

e. Conclusion?

f. (Do not submit but try) Can you Repeat this with difference of medians instead of means? Conclusions?