Midterm – Stats for IR

Put your name here

Due 4/23/2020

Instructions:

1. This is a 10-days take-home exam.
2. This is an individual exam. You should not collaborate.
3. I will hold office hours Thursday, during class time.
4. In the office hours, you may not ask questions specifically about the midterm. Please rephrase the issues in terms of similar problems.
5. If it does not knit, the grade is **zero**.
6. Good luck!

Question 1

Thomas Fujiwara wrote the following tweet to analyze the COVID-19 evolution. You should replicate the same graph, not for the Brazilian States, but countries. After this, you should tell us about the COVID-19 trends in the world. In which countries the COVID is going down? In which countries the COVID is going up? Do two separate analysis for the:

1. Evolution of infections
2. Evolution of casualties

Hint: recycle the code you wrote in the last problem set.

Question 2

Let the following vector:

```
ages <- rbinom(100, 50, .2)+18
```

Compute:

1. The mean of the vector.
2. The median of the vector.
3. The quartiles.
4. The 1-percentile and the 99-percentile.
5. Make a histogram.
6. Make a box-plot.

Question 3

In the following dataset, we have an experiment where researchers send resumes for jobs, randomizing the names of the applicants. In the US, the names of applicants are correlated with race, and this allows firms to select racially based on the person's names.

The paper that this problem came from is the following: Marianne Bertrand and Sendhil Mullainathan (2004) "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination." American Economic Review, vol. 94, no. 4, pp. 991–1013. (https://www.aeaweb.org/articles?id=10.1257/0002828042002561)

The dataset is the following:

```r
resume <- read.csv('https://raw.githubusercontent.com/umbertomig/intro-prob-stat-FGV/master/datasets/res
head(resume)
```

```
##   firstname    sex  race call
## 1   Allison female white    0
## 2   Kristen female white    0
## 3   Lakisha female black    0
## 4   Latonya female black    0
## 5    Carrie female white    0
## 6       Jay   male white    0
```

If has the following variables:

- **firstname:** First name in the resume
- **sex:** Gender of the applicant
- **race:** Race of the applicant
- **call:** Whether a callback was made (call=1) or not (call=0)

Answer the following questions:

1. What is the overall callback rate?
2. What is the black callback rate?
3. What is the white callback rate?
4. What is the Black-Female callback rate?
5. What is the White-Female callback rate?
6. What is the callback rate by the first name? Which first name receives the highest and the lowest callback rates?
7. Make a mosaicplot of the callback rate by race.
8. Make a mosaicplot of the callback rate by gender.

Question 4

1. What is the fundamental problem of the causal inference? Provide an example.
2. One researcher wants to study tax conservatism among the rich. She selects lottery ticket winners to run her study, and she claims that the biggest the ticket the person won, the more the person is against income taxes. Should we believe this study? What are the strengths of this study? What are the weaknesses of this study?
3. Suppose you want to replicate the study in **question 3** in Brazil. What would you do? What are the design choices that you would make? What are the potential problems?