Statistics for Engineers
Assignment 2
Fatal Force

Natalya Pya Arnqvist, Konrad Abramowicz, Viktor Mostberg, Hugo Englund

October 9, 2020

General Information

In this lab assignment you are supposed to extract a sample of data, visualize it and perform some statistical inference. The lab consists of one task and is to be solved in groups. You are supposed to write a lab report as described on the lecture. It should include:

- Task description

- Method section including:

    - Necessary notation, definitions and assumptions
    - Connection between your task and underlying theory
    - Calculations that lead to the solutions

- Results & Conclusions

    IMPORTANT INFORMATION:

- The lab report shall be submitted in **pdf** format trough Cambro.

- The *r*-files with the code should be submitted together with the report.

Manual for report writing

General idea

The manual is thought to be a guideline for you while writing the report. It is of the utmost importance that you follow the guideline in order to create a complete report. The manual is divided into separate **sections** and in each section there are a certain amount of questions as well as instructions of which you will have to answer as well as follow.

Further, the assignment consists of the two following steps:

1. Answer the questions in the task by following each step in the manual. Then, save the manual **with** your calculations, answers and discussions in a separate pdf-file.

2. Create a copy of the file above. Now, remove all the guidelines from the copied file, but **keep** the sections and your answers. Then, write a report by converting your answers under each section to a fluent and readable text. Lastly, add a front page to your report containing a title, course name, group members (name & e-mail), group number and date. Save the report as a pdf-file.

What is to be handed in?

To accomplish the assignment, you have to hand in:

- The manual with your answers as a pdf-file.

- The report as a pdf-file.

- The commented r-code used to solve the task as a r-file.

If anything is excluded, your report will not be corrected and therefore automatically graded **U**.

Check the Statistics

The Dataset

The Washington Post's database contains records of every fatal shooting in the United States by a police officer in the line of duty since January 1, 2015.

In 2015, The Post began tracking more than a dozen details about each killing - including the race of the deceased, the circumstances of the shooting, whether the person was armed and whether the person was experiencing a mental-health crisis - by culling local news reports, law enforcement websites and social media, and by monitoring independent databases such as Killed by Police and Fatal Encounters. The Post conducted additional reporting in many cases.

The Post is documenting only those shootings in which a police officer, in the line of duty, shoots and kills a civilian — the circumstances that most closely parallel the 2014 killing of Michael Brown in Ferguson, Mo., which began the protest movement culminating in Black Lives Matter and an increased focus on police accountability nationwide. The Post is not tracking deaths of people in police custody, fatal shootings by off-duty officers or non-shooting deaths.

Read more about the dataset and its variables **here**.

The Task

The population data related to the task is found in the file `shootings.csv`. You are to use the data uploaded on Cambro and **not** download the data from webpage. Import it to R and add to workspace as variable `shootings`. Assure that all variables are treated as they should be (numeric as numeric, factors as factors).

Generate your sample   Calculate the sum of the digits of your birth date (e.g. 981224) of all members in the group, and call this variable r. Use r to initialize random number generator, and generate the sample dataset for your group, by writing:

```
set.seed(r)
data=shootings[sample(4895, 100, replace = TRUE),]
data=droplevels(data)
```

The sample now contains information about 100 individuals. In most of this lab you will only work with the sample, and at the end you will see if the results correspond to what you see in population of all 4895 observations. Let's analyze!

Summary & Visualization   Perform a summary on your sample. Also visualize the distributions of age, fleeing and signs of mental illness in your sample. Comment on the results.

Question 1   What is the average age of the victims in your sample? Using your sample, construct a relevant 95% set for the population mean, in order to determine the lower bound for the mean age along the total population of victims. Provide a probabilistic interpretation of this confidence interval. Can we use it to say something about the age of victims?

Question 2   Based on your sample, what can you say about the proportion of the unarmed victims in the population. Construct a relevant 95% confidence interval.

Question 3   As an argument in discussion, one can claim that the mental illness is the reason for the shooting. Using the observed sample, can we say that more than 20% of the victims have signs of mental illness? Perform a relevant hypothesis test to verify the statement. Use significance level 5%.

Question 4   Intuitively, one would argue that fleeing from the police officer would increase the likelihood to be shot. Based on your sample, perform a hypothesis test to determine if this is the case or not. Use significance level 5%.

Question 5   Now, return back to analysis of the full population and verify if the conclusions you made based on your sample are correct or not. Why or why not?