M358K - Homework 2

posted on: September 24th, 2018

due: October 8th, 2018

Note. To each of the 'is ... statistically sigifnicant' question below, please set up a hypothesis test. Write down

- H_0 and H_A
- What test you use in R: name of test, one-sided or two-sided
- Your chosen significance level
- Report the p-value you obtained
- Give a conclusion

Number of questions in this homework: 3.
Maximal points possible: 6 from writeup, 2 from code, 4 from presentation. This gives a total of **10 points**.

Question 1. hsb2 dataset

Consider the hsb2 dataset. This dataset is included in the `openintro` package. You can load this dataset in R with the following commands:

```
install.packages("openintro")
library(openintro)
data(hsb2)
```

Now there will be a dataframe named `hsb2` in your R workspace. You can do usual commands, such as

```
head(hsb2)
```

For information on the variables, type in R

```
?hsb2
```

1. Describe in graphs and numbers the distribution of math scores between male and female

2. Is there a significant difference in the median score between these two groups? Use a permutation test to find out. Remember to set seed so that the grader can reproduce your result. `?set.seed`

3. Is there a significant difference between male and female in the proportion of those who math score is 65 or more? Use a test of your choice.

4. Does your analysis disprove or support the claim that "top math students tend to be male"?

Question 2. Murders dataset

The dataset `murders` contains the victim name, age, and location of every murder recorded in the Greater London area by the Metropolitan Police from January 1, 2006 to September 7, 2011.

This dataset is included in the `OIdata` package. You can load this dataset in R with the following commands:

```
install.packages("OIdata")
library(OIdata)
data(murders)
```

Now there will be a dataframe named `murders` in your R workspace. You can do usual commands, such as

```
head(murders)
```

For information on the variables, type in R

```
?murders
```

This documentation has helpful examples, including a code on how you can visualize the murder on the London map.

Our goal is to answer the question: do all boroughs have the same murder rate, or are there some "bad neighborhoods"?

1. Produce a map to visualize the murders by borough.

2. Produce a table that counts the number of murders by borough.

3. Is the count itself meaningful? What other statistics are we missing to compute the murder rate? Go online to find them and compute the murder rate by borough.

Question 3. Murders, continued

1. Is there a significant difference in the rate of murders between boroughs? Answer this question by performing a permutation test.

Hint: you will need to write a piece of R code to randomly reassign the location of the murders in the dataset. Be VERY clear on what test statistic you are using.

Bonus. A question for you that I don't know the answer to: how to obtain a similar dataset for Austin?