Problem Set 1

Claire Mooney

Due: October 1, 2021

A school counselor was curious about the average of IQ of the students in her school and took a random sample of 25 students' IQ scores. The following is the data set:

1 Question 1: Education

1. Find a 90% confidence interval for the average student IQ in the school.

I began by inputting the values of the data set.
y ¡-105,69,86,100,82,111,104,110,87,108,87,90,94,113,112,98,80,97,95,111,114, 89,94,126,98)
t.test(y,conf.level=0.9)
One Sample t-test

data: y = 37.593, df = 24, p-value ¡ 2.2e-16

alternative hypothesis:
true mean is not equal to 0

90 percent confidence interval:
93.95993 102.92007
sample estimates: mean of x is 98.44
2. One cannot reject the null hypothesis is the school mu is lower in comparison to other countries around the country

The alternative hypothesis is that the school's average IQ is greater than 100

t.test(y,mu=100,alternative="greater")

I ran a test test using our sample y and I set the mu to the population average 100. The teacher wants to find out if her students IQS are greater so we insert "alternative=greater"

Running a one sided t-test with our alternative hypothesis yields a p value of 0.7215

This test is not significant with alpha =o.o5 therefore p is greater than alpha

Question 2: Political Economy

Researchers are curious about what affects the amount of money communities spend on addressing homelessness. The following variables constitute our data set about social welfare expenditures in the USA.

`State`	*50 states in US*
`Y`	*per capita expenditure on shelters/housing assistance in state*
`X1`	*per capita personal income in state*
`X2`	*Number of residents per 100,000 that are "financially insecure" in state*
`X3`	*Number of people per thousand residing in urban areas in state*
`Region`	*1=Northeast, 2= North Central, 3= South, 4=West*

Explore the `expenditure` data set and import data into `R`.

- Please plot the relationships among *Y*, *X1*, *X2*, and *X3*? What are the correlations among them (you just need to describe the graph and the relationships among them)?

- I began by bringing in my raw data commanding expenditure to read the raw data.
  I then created a string of expenditure and created lines within the string.
  In order plot the relationship amoung Y, X1,X2 and X3 I began by examining historgrams separately of each.
  Finally in order to see all the data I entered the code:
  plot(expenditure,ylim=range(expenditure(dollar)Y,expenditure(dollar)X1,expenditurw(dollar)X2,ex
  main = "Expenditure of States in US")
  This promoted a series of graphs. It offered a variety of data some that didn't correlate at all. Some that offered strong correlation and suggested a significant relationship. There was both positive and negative correlation. There was a variety of strength of relationship between the data.

  Please plot the relationship between *Y* and *Region*? On average, which region has the highest per capita expenditure on housing assistance?

- I installged ggplot2.
  install.packages("ggplot2")
  library(ggplot2)

  expenditure(dollar)Region = as.factor(expenditure(dolla)Region) library(ggplot2)

  data=as.data.frame(expenditure[c(2,6)])
  data(dollar)Region = as.factor(data(dollar)Region)
  mode(data(dollar)Region)

  This then changes expenditure to a dataframe.
  As I learnt R will stores regions as intergers unless you make them factors.
  This has created a data frame out of the two variables we are looking at.
  In this case we are using Y(2) and Region (6) based on their position on the list.

  Mode(data(dollar)region)was to check if the region had changed to a factor

  ggplot(aes(y = Y, x = Region, fill=Region), data = data)+ geomboxplot()+ggtitle("Box plots of Expenditure by Region")

  I had originally tested this mode but had not realised that ggplot2 needed to be installed. Following guidance I installed ggplot2.

  I then created a box plot graph comparing expenditure and region Having analysed the data it appears region 4 the 'West'has the highest per capita expenditure on housing assistance. Therefore 4 the 'west' has on average the highest per capita expenditure on housing assistance.

  Please plot the relationship between Y and $X1$? Describe this graph and the relationship. Reproduce the above graph including one more variable *Region* and display different regions with different types of symbols and colors.

  I repeated the same code as above to explore the relationship between y and x1.
  I then replaced where value x1 was with Region and executed the code.

  This brought up a graph with a weak positive correlation. The data being quite dispersed may even be suggestive there is very little correlation. This suggests there is not a very strong connection between the two.

This meant that X1 was now available to code alongside my already established relationship between y and x1

After this is I then entered the following code:

data=as.data.frame(expenditure[c(2,3,6)]) data(dollar)Region = as.factor(data(dollar)Region) mode(data(dollar)Region)

This created a data frame out of the three instead of two variables.
In this case we are using Y(2) and X1(3) and Region (6) and based on their position on the list

I then ran this line of code:
ggplot(aes(y = Y, x = X1, fill= Region), data = data)+ geomboxplot()+ggtitle("Box plots of Expenditure per capita")

This then brought up a graph that showed a positive correlation between the data. The data had a slightly stronger relationship then the previous data but there was no especially strong relationship between the data.