

iiiiii HEAD ===== llllll af2a91eef73f36234332980a3a4c406b0f477d9e

# Problem Set 1

Applied Stats/Quant Methods 1

Due: October 3, 2022

## Instructions

- Please show your work! You may lose points by simply writing in the answer. If the problem requires you to execute commands in **R**, please include the code you used to get your answers. Please also include the **.R** file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask.
- Your homework should be submitted electronically on GitHub.
- This problem set is due before class on Monday October 3, 2022. No late assignments will be accepted.
- Total available points for this homework is 80.

iiiiii HEAD

## Question 1 (50 points): Education

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

```
1 # Problem 1
2 #####
3
4 #load data
```

```

5 y <- c(105, 69, 86, 100, 82, 111, 104, 110, 87, 108, 87, 90, 94, 113, 112, 98,
      80, 97, 95, 111, 114, 89, 95, 126, 98)
6
7 #calculate mean
8 ybar <- mean(y)
9
10 #visualize y
11 hist(y, xlab = "IQ scores")
12
13 # get confidence intervals
14 CI_lower <- qnorm(0.05,
15                   mean = mean(y),
16                   sd = (sd(y)/sqrt(length(y))) # the equation for the standard
      error of the mean
17 )
18
19 CI_upper <- qnorm(0.95,
20                   mean = mean(y),
21                   sd = (sd(y)/sqrt(length(y)))
22 )

```

Lower : 94.13283 Upper : 102.7472

```

1 matrix(c(CI_lower, CI_upper), ncol = 2,
2         dimnames = list("", c("Lower", "Upper")))

```

2.618575

```

1 #get confidence intervals using t distribution

```

1.710882

```

1 se

```

93.95993

```

1 t_score

```

102.9201

1. Next, the school counselor was curious whether the average student IQ in her school is higher than the average IQ score (100) among all the schools in the country.

Using the same sample, conduct the appropriate hypothesis test with  $\alpha = 0.05$ .

Hypothesis testing

Step 1 : I assuming that this is a normal distribution

Step 2: The null hypothesis is that ybar is less than or equal to 100  
The alternative hypothesis is that ybar is greater than 100

Step 3: Calculating the t-statistic

```
1  
2 # Step 3:
```

-0.5957439

```
1 # t.score = ybar - mu/ se  
2 ts <- (ybar - 100)/se  
3 ts  
4  
5 #check t.score result with R t.test()  
6 t.test(y,
```

One Sample t-testdata: yt = -0.59574, df = 24, p-value = 0.7215  
Alternative hypothesis: true  
Mean is greater than 100  
95 percent confidence interval: 93.95993  
Infsample estimates:mean of x 98.44

## Question 2 (50 points): 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.

```
1 mean = mean(y) ,
```

- 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)?

Figure 1:  $Y$  and  $X1$  plot

The relationship between these variables is broadly linear however there are a few outliers with a cluster at the bottom of the plot

Figure 2:  $Y$  and  $X2$  plot

The relationship between these variables is broadly linear however there are a few outliers

Figure 3:

There is a somewhat linear relationship between these variables

Figure 4:  $X1$  and  $X2$  plot

There is not a linear relationship between these variables

Figure 5:

There is a linear relationship between these variables

Figure 6:

There is no relationship between these variables

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

Figure 7:

Region 4 (West) spends the most on average per capita on shelters/ 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.

Figure 8:  $Y$  and  $X1$  plot with *Region* variable

=====

## Question 1 (40 points): Education

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 # Problem 1

1. Find a 90% confidence interval for the average student IQ in the school.  
 11111111 af2a91eef73f36234332980a3a4c406b0f477d9e
2. Next, the school counselor was curious whether the average student IQ in her school is higher than the average IQ score (100) among all the schools in the country.

Using the same sample, conduct the appropriate hypothesis test with  $\alpha = 0.05$ .

## Question 2 (40 points): 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.

```
1 mean = mean(y) ,
```

- 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)?
- Please plot the relationship between  $Y$  and  $Region$ ? On average, which region has 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.