

Problem Set 1

Molly Marino

Due: October 9, 2025

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 23:59 on Thursday October 9, 2025. No late assignments will be accepted.

Question 1: 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 average_student_IQ <- 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)
```

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

```
1 average_student_IQ <- 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)  
2 #I renamed "y" as an object to average_student_IQ because it better  
  represents the data set.  
3 sample.mean <- mean(average_student_IQ)
```

```

4 # I found the mean of the sample because it is used as a point estimate
  for the unknown population mean.
5 sample.mean
6 #This step shows the mean in the console. The mean is 98.44.
7 sample.n <-length(average_student_IQ)
8 # I did this step to calculate my sample size. The sample.n is 25L
9 sample.sd <-sd(average_student_IQ)
10 # The next step of calculating the confidence interval is to find the
    standard deviation or spread of the data. The sample.sd is 13.09.
11 sample.se <-sample.sd/sqrt(sample.n)
12 # I did this step to calculate the standard error and to see how much the
    mean will fluctuate between the samples.
13 sample.se
14 # This steps shows the standard error in my console. The sample.se is
    2.618.
15 df <- sample.n-1
16 # I did this step to calculate the shape of the distribution. The df is
    24.
17 t_critical <- qt(0.95,df=sample.n-1)
18 # This step is using a t distribution because the population size is less
    than 30. I adjusted for 90 percent confidence to find the t score.
    The t_critical value is 1.71.
19 margin_error <- t_critical * sample.se
20 # I did this step to find the discrepancy between the sample and the t
    value.The margin_error is 4.48.
21 lower_bound <-sample.mean-margin_error
22 #The lowest possible value within the 90 percent confidence level for the
    population parameter. The lower_bound is 93.95.
23 upper_bound <-sample.mean +margin_error
24 # The highest possible value within the 90 percent confidence level for
    the population parameter. The upper bound is 102.92.
25 print(c(lower_bound, upper_bound))
26 # This shows the 90 percent confidence intervals of the average student
    IQ in the school for upper and lower values. Meaning, if sampling was
    conducted numerous times 90 percent of the student's IQ scores would
    fall between 93.95-102.92

```

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 $\alpha = 0.05$.

```

1 # H0:  =100
2 # Ha:  >100
3 mu0 <- 100
4 # mu0 is a way to store the value null hypothesis mean under a value name
  .
5 t_value <- (sample.mean - mu0) / sample.se
6 # The t value is used because the sample is less than 30 and it shows how
    far the sample mean is away from the null hypothesis. The t_value is
    -0.595.

```

```

7 t_value
8 # This shows the t value in the console
9 p_value <- pt(abs(t_value), df, lower.tail = FALSE)
10 # the p value is used to compare against the a=0.05 to determine if we
    can reject the sample data based on the null hypothesis. I did tail=
    false because I am only interested in scores above 100 in the upper
    tail of the data.
11 p_value
12 # The p value is 0.27 and this means that is is greater than 0.05. The
    null hypothesis cannot be rejected.
13 alpha <- 0.05
14 if(p_value < alpha){
15   print("Reject H0: Average student IQ is greater than 100")
16 } else {
17   print("Do not reject H0: Not enough evidence that average student IQ is
        greater than 100")
18 }
19 # The null hypothesis cannot be rejected because the p value of 0.27 is
    greater than the alpha value of 0.05. This means that there is not
    enough evidence that the average student IQ score in the teacher's
    class is higher than the national average.

```

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.

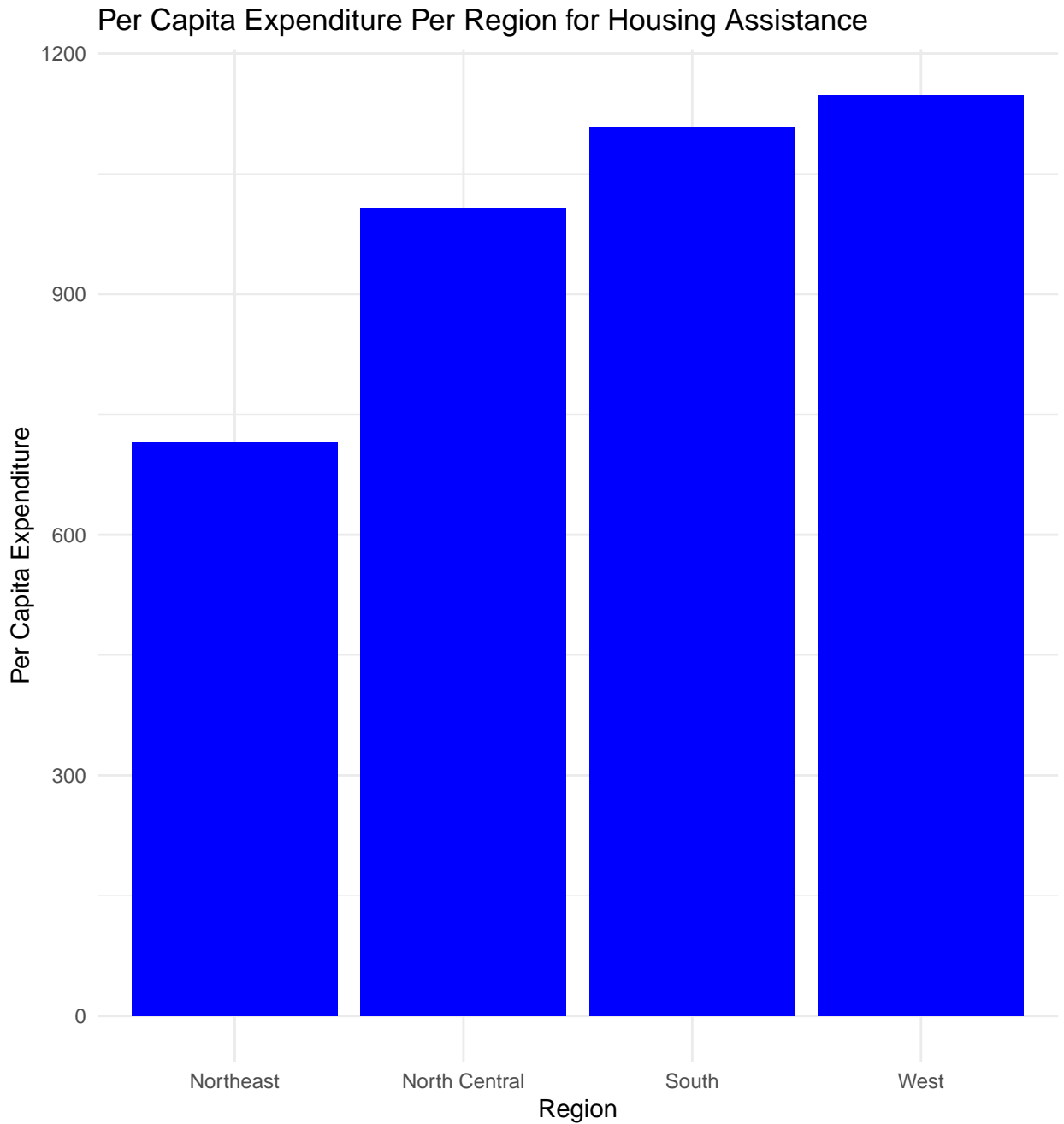
- Please plot the relationships among Y , $X1$, $X2$, and $X3$? What are the correlations among them (you just need to describe the graph and the relationships among them)?
- Please plot the relationship between Y and $Region$? On average, which region has the highest per capita expenditure on housing assistance?

```
1 ggsave("C:/Users/molly/OneDrive/Documents/GitHub/StatsI_2025/problemSets/
  PS01/my_answers/Per_Capita_Expenditure.pdf")
2 # I needed to create a barplot because one of the variables is
  categorical and one is numerical.
3 # The region that has the highest per capita expenditure on housing is
  the West.
4
5
6 expenditure$Per_Capita_Personal_Income_State <-expenditure$X1
7 Per_Capita_Personal_Income_State <-expenditure$X1
8 # I created a new variable name to name X1 to, "Per Capita Personal
  Income State."
9 pdf("C:/Users/molly/OneDrive/Documents/GitHub/StatsI_2025/problemSets/
  PS01/my_answers/Per_Capita_Housing_Assistance.pdf")
10 plot(Per_Capita_Personal_Income_State,Per_Capita_Expenditure,main="Per
  Capita Housing Assistance Vs.Personal Income",xlab="Per Capita Income
  In State",ylab="Per Capita Expenditure On Housing In State",
11       pch=19, frame=FALSE)
12 dev.off()
13 # I needed to do this part because this is not ggplot and has to be saved
  a special way.
14 # The relationship that is displayed by the graph is that as per capita
  income in the state goes up, the per capita expenditure on housing
```

```

    also goes up. This is a positive correlation.
15 colors<-c("Northeast"="blue", "Northcentral"="green", "South"="red", "West"
    ="yellow")
16 symbols<-c("Northeast"=16,"Northcentral"=15,"South"=14,"West"=17)
17 # I am assigning colors and symbols to the different regions.
18 region_colors<-colors[Region]
19 # I am assigning a variable called region_colors to represent the colors
    for the region variable.
20 region_symbols<-symbols[Region]
21 # I am assigning a variable called region_symbols to represent symbols
    for the region variable.
22 pdf("C:/Users/molly/OneDrive/Documents/GitHub/StatsI_2025/problemSets/
    PS01/my_answers/Per_Capita_Housing_Assistance_Colors.pdf")
23 plot(Per_Capita_Personal_Income_State, Per_Capita_Expenditure, main="Per
    Capita Housing Assistance Vs. Personal Income", xlab="Per Capita Income
    In State", ylab="Per Capita Expenditure On Housing In State",
24      pch=region_symbols, col=region_colors, frame=FALSE)

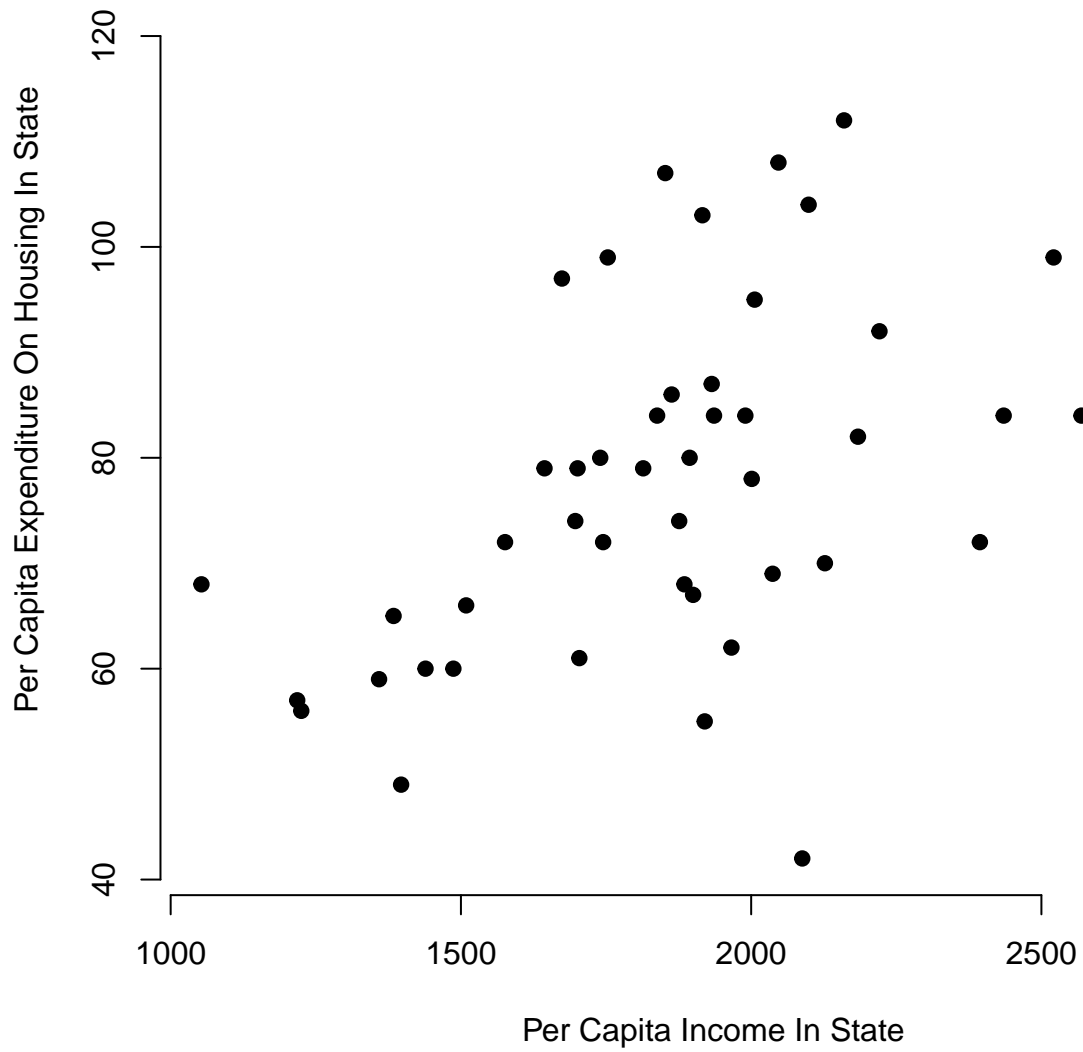
```



- 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.

```
1 dev.off()  
2 #I did this because this is not ggplot and has to be saved a special way.
```

Per Capita Housing Assistance Vs. Personal Income



Per Capita Housing Assistance Vs. Personal Income

