PS Response

Applied Stats/Quant Methods 1

Zhexuan Yin

Instructions

This is my responses in ***R*** *and* ***Latex****.*

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:

```
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)
```

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

   90% of observations lie within +/-1.645

   The **approximate** solution
   Lower bound, 90 confidence level

```
lower_90 = mean(y) - (1.645*sd(y)/sqrt(length(y)))
```

   Upper bound, 90 confidence level

```
upper_90 = mean(y) + (1.645*sd(y)/sqrt(length(y)))
```

   The **precise** solution, using normal distribution
   Lower bound, 90 confidence level

```r
lower_90_n <- qnorm(0.05,
                    mean = mean(y),
                    sd = (sd(y)/sqrt(y)))
```

Upper bound, 90 confidence level

```r
upper_90_n <- qnorm(0.95,
                    mean = mean(y),
                    sd = (sd(y)/sqrt(length(y))))
```

90% confidence interval for the average student IQ: 94.13244 to 102.7476

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 $\alpha = 0.05$.

```r
t_test_result<-t.test(y, mu=100, alternative = "greater")
t_test_result
```

   One Sample t-test
   data: y
   t = -0.59574, df = 24, p-value = 0.7215
   alternative hypothesis: true mean is greater than 100
   95 percent confidence interval:
   93.95993 Inf
   sample estimates:
   mean of x
   98.44

   Test Statistic: -0.5957439
   P-value: 0.7215383

   So according to the sample estimates, the average student IQ in her school only has 98.44, isn't higher than the average IQ score (100) among all the schools in the country.

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`.

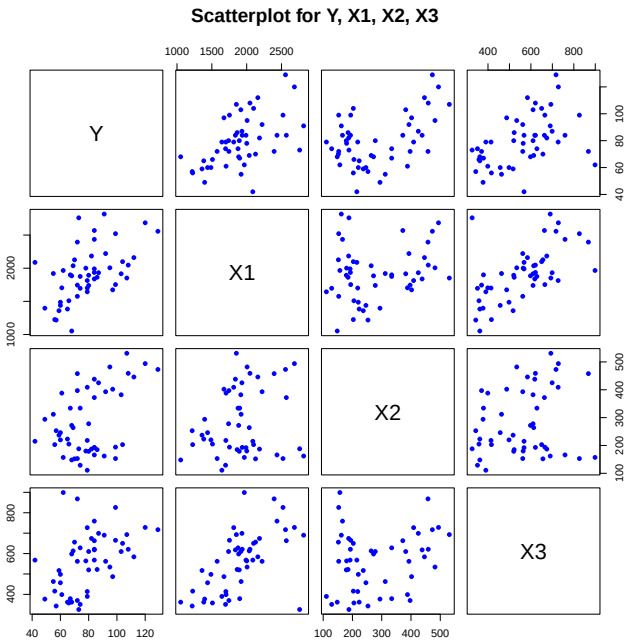
```
lower_90 = mean(y) - (1.645*sd(y)/sqrt(length(y)))
```

- 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)?

  Plot the relationships among Y, X1, X2, and X3

```
pairs(expenditure[, c("Y", "X1", "X2", "X3")],
      main = "Scatterplot for Y, X1, X2, X3",
      pch = 16,
      col = "blue")
```

  Scatterplot matrix for Y, X1, X2, and X3



Calculate matrix of the correlations

```
cor(expenditure[, c("Y", "X1", "X2", "X3")])
```

```
           Y        X1        X2        X3
Y  1.0000000 0.5317212 0.4482876 0.4636787
X1 0.5317212 1.0000000 0.2056101 0.5952504
X2 0.4482876 0.2056101 1.0000000 0.2210149
X3 0.4636787 0.5952504 0.2210149 1.0000000
```

Y and X1 X2 X3, X1 and X3 also have a moderately positive correlation, suggesting that X1 X2 X3 increases as Y, and X1 increases as X3. But the correlation between X1 and X2, X2 and X3, is only 0.2, which shows that X1 is not a strong predictor of X2, and X2 is not a strong predictor of X3.

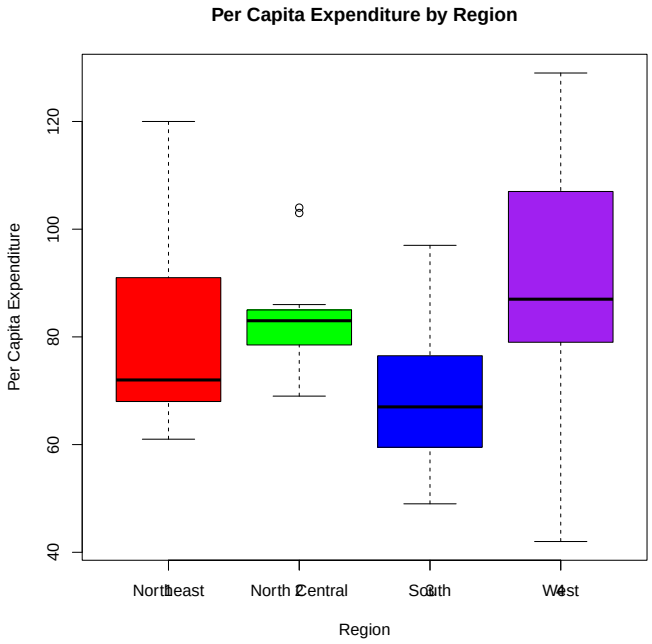
- Please plot the relationship between Y and *Region*? On average, which region has the highest per capita expenditure on housing assistance?

  Plot the relationship between Y and Region
  Box plot

```
boxplot(Y ~ Region,
        data = expenditure,
```

```
        main = "Per Capita Expenditure by Region",
        xlab = "Region",
        ylab = "Per Capita Expenditure",
```

```
axis(1, at=1:4, labels = c("Northeast", "North Central", "South", "West")
    )
```



Calculate the mean expenditure for each region

```
tapply(expenditure$Y, expenditure$Region, mean)
```

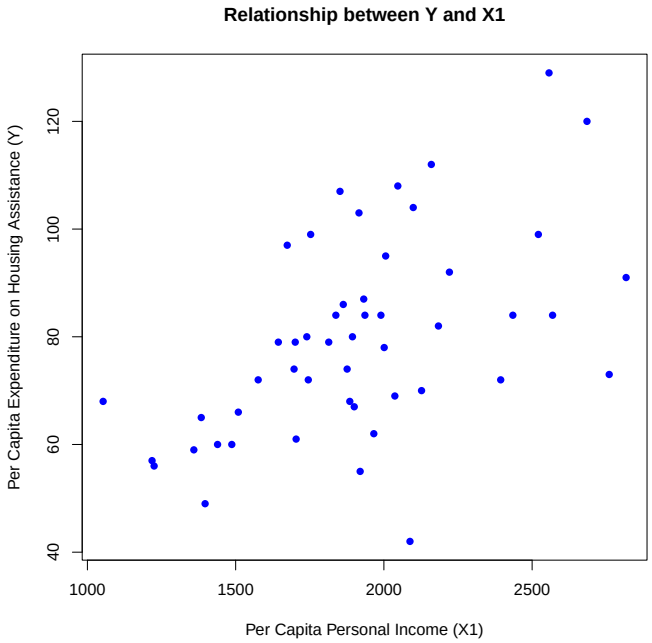
```
1        2        3        4
79.44444 83.91667 69.18750 88.30769
```

West 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.

Plot the relationship between Y and X1
Scatter plot for Y and X1

```r
plot(expenditure$X1, expenditure$Y,
     main = "Relationship between Y and X1",
     xlab = "Per Capita Personal Income (X1)",
     ylab = "Per Capita Expenditure on Housing Assistance (Y)",
     pch = 16, col = "blue")
```

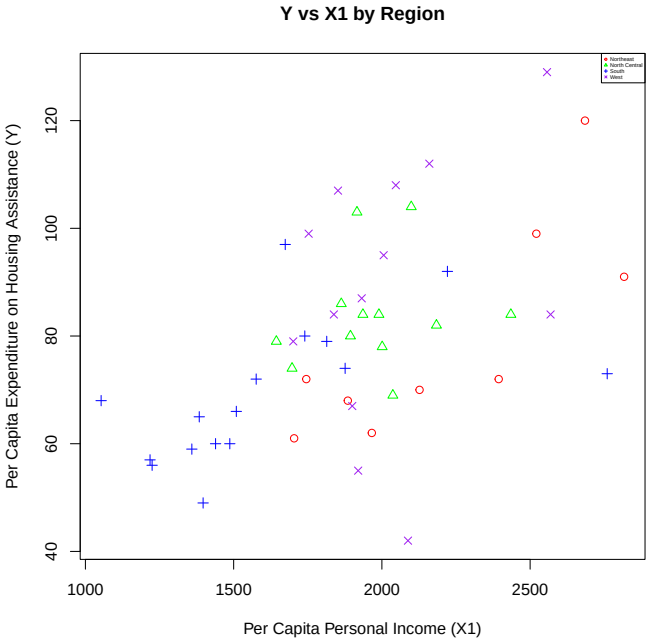


According to the graph, we can see X1 increases as Y. Y and X1 have the relationship of a positive correlation.

Scatter plot for Y and X1, colored by Region

```r
plot(expenditure$X1, expenditure$Y,
     main = "Y vs X1 by Region",
     xlab = "Per Capita Personal Income (X1)",
     ylab = "Per Capita Expenditure on Housing Assistance (Y)",
     pch = as.numeric(expenditure$Region),
     col = c("red", "green", "blue", "purple")[expenditure$Region])
```

```r
legend("topright",
       legend = c("Northeast", "North Central", "South", "West"),
       col = c("red", "green", "blue", "purple"),
       pch = 1:4,
       cex = 0.3)
```


Y vs X1 by Region
Northeast
North Central
South
West
Per Capita Expenditure on Housing Assistance (Y)
40
60
80
100
120
1000
1500
2000
2500
Per Capita Personal Income (X1)