

Question 1 (25 points)

A private 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 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)
```

Find a 90% confidence interval for the student IQ in the school assuming the population of IQ from which our random sample has been selected is normally distributed.

```
1 sd(y) #13.09
2 qnorm(.95) #1.64
3 mean(y) #98.44
4
5 me <- 1.64*(13.09/sqrt(10))
6 me #6.78
7
8 mean(y)-me #91.65
9 mean(y)+me #105.22
```

A 90 Percent Confidence Interval = (91.65, 105.22)

Question 2 (25 points)

A private school counselor was curious whether the average of IQ of the students in her school is higher than the average IQ score 100 among all the schools in the country. She took a random sample of 25 students' IQ scores. The following is the data set:

```
1 #####
```

```
1 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)
```

Conduct a test with 0.05 significance level assuming the population of IQ from which our random sample has been selected is normally distributed.

```
1 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)
```

data fits the assumptions of random sampling, quantitative data, and normal distribution. state hypotheses: $H_0 = 100$, H_A does not = 100 calculate the test statistic:

```
1 mean(y) #sample mean = 98.44
```

```

1 sd(y) #sd of sample = 13.09
2 13.09/sqrt(25) #standard deviation of sampling distribution = 2.618
3
4 (98.44-100)/2.618 #-0.5958747, df=24

```

find P value:

```

1 2*pt(abs(-0.59), df=24, lower.tail=F)

```

P value = 0.56. Since $p > 0.05$, we conclude that the evidence supports the null hypothesis. The average IQ of the students in her school is not higher than the average IQ score 100 among all the schools in the country.

Question 3 (50 points)

```

1 expenditure.txt
2 expenditure <- read.table("expenditure.txt", header=T)

```

Researchers are curious about what affects the education expenditure on public education. The following is available variables in a data set about the education expenditure.

State	50 states in US
Y	per capita expenditure on public education
X1	per capita personal income
X2	Number of residents per thousand under 18 years of age
X3	Number of people per thousand residing in urban areas
Region	1=Northeast, 2= North Central, 3= South, 4=West

- Please plot the relationships among Y, X1, X2, and X3? What are the correlations among them? Describe the graph and the relationships among them.

Plot Y

```

1 hist(expenditure$Y, main="Per Capita Expenditure on Public Education",
      xlab="Y", ylab="Frequency")

```

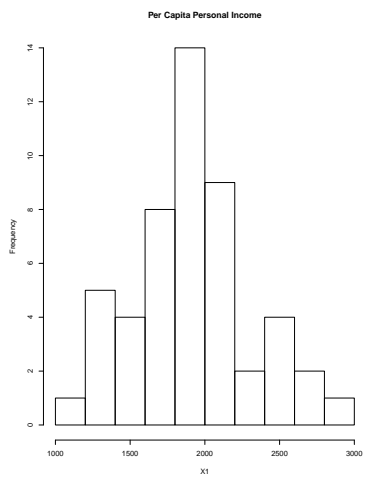
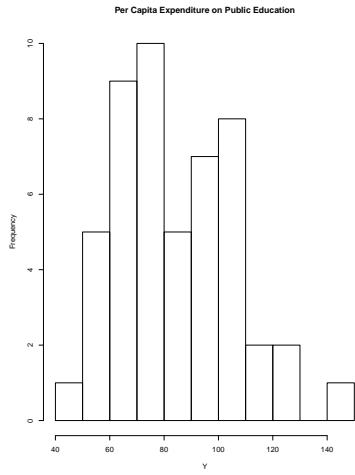
The data for Y presents a bimodal histogram, with two spikes. The first at around 60 and the second at 100.

Plot X1

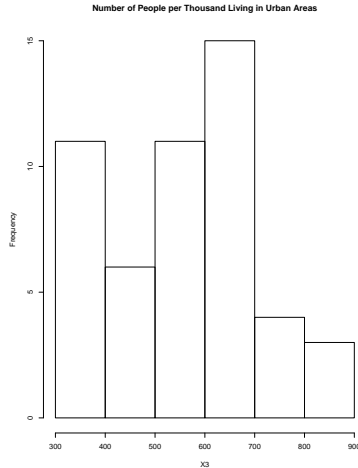
```

1 hist(expenditure$X1, main="Per Capita Personal Income", xlab="X1", ylab="
      Frequency")

```



X2.pdf



X1 displays a relatively normal distribution, with its peak per capita personal income at about 2000.

Plot X2

```
1 hist(expenditure$X2, main="Number of Residents per Thousand Under 18  
   Years", xlab="X2", ylab="Frequency")
```

X2 has the highest frequency at about 400, with a largely right-skewed distribution. Thus, there is a higher frequency of residents per thousand under 18 between the values of 300 and 400.

Plot X3

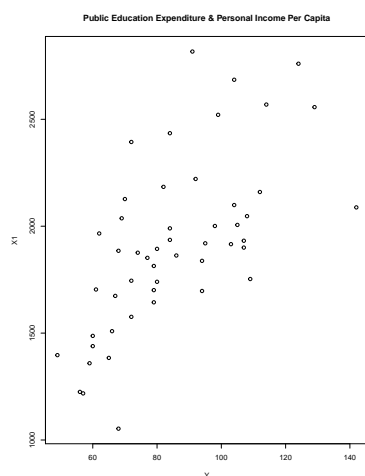
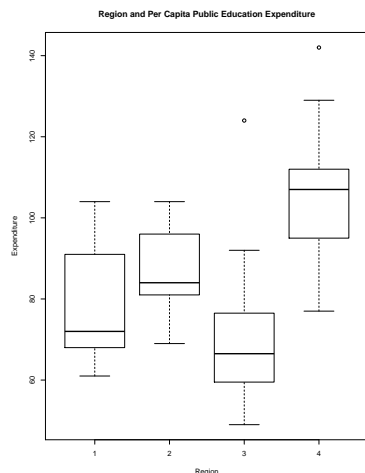
```
1 hist(expenditure$X3, main="Number of People per Thousand Living in Urban  
   Areas", xlab="X3", ylab="Frequency")
```

X3 has the highest frequency at its mean, which is around 600. It is slightly skewed right, less significantly than X2.

- Please plot the relationship between Y and *Region*? On average, which region does have the highest per capita expenditure on public education?

```
1 boxplot(Y~Region, data=expenditure, main="Region and Per Capita Public  
   Education Expenditure", xlab="Region", ylab="Expenditure")
```

Region 4 has, on average, the highest per capita expenditure on public education.



- 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 plot(expenditure$Y, expenditure$X1, main = "Public Education Expenditure
    & Personal Income Per Capita", xlab="Y", ylab="X1")
```

There is a positive association between Public Education Expenditure and Personal income per capita; as income increases, so too does spend on public education.

```
1 plot(expenditure$Y, expenditure$X1, col=as.integer(expenditure$Region),
    pch=as.integer(expenditure$Region), main = "Public Education
    Expenditure & Personal Income Per Capita by Region", xlab="Y", ylab="
    X1")
```

