PS01 Response

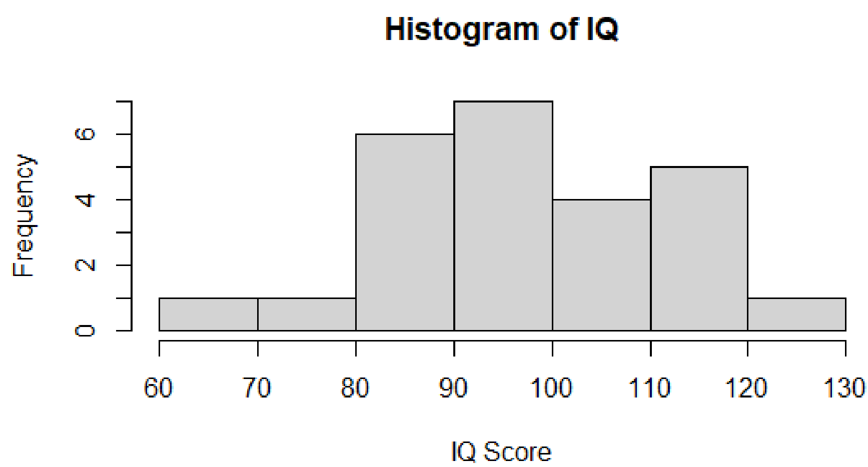
Applied Stats/Quant Methods 1

Zach Keller

Question 1

First, the data was explored. Both histograms and QQ plots were created to determine whether to use T scores or Z scores by investigating whether the data is normal

Figure 1: Histogram of IQ Scores.



As shown above, the data does not look to be normally distributed. Therefore, it was decided that T Scores were more suitable to use. More specifically, a one-tailed one-sampled t-test was used, as we are only testing in one direction. In order to construct a confidence interval, the following code was used:

```
stderror <- sd(y)/sqrt(length(y))
t_score <- qt(.05, df = length(y)-1, lower.tail = FALSE)
CI_lower_t <- mean(y) - (stderror * t_score)
CI_upper_t <- mean(y) + (stderror * t_score)
```

Figure 2: QQ Plot of IQ Scores.

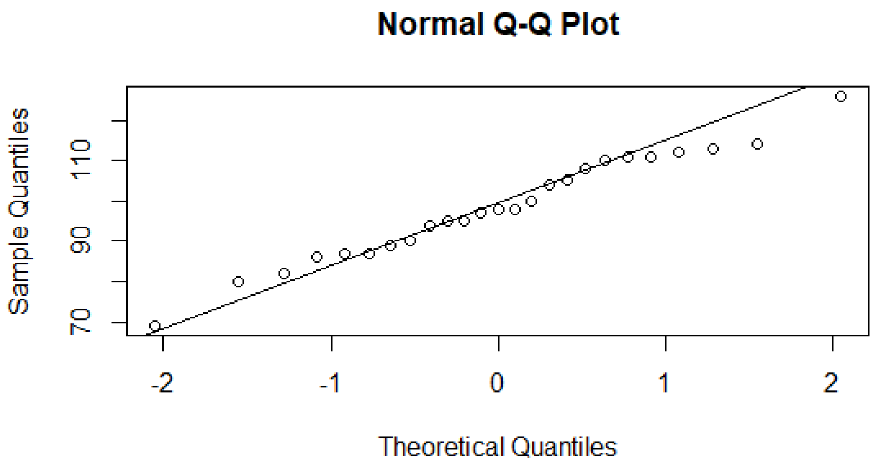
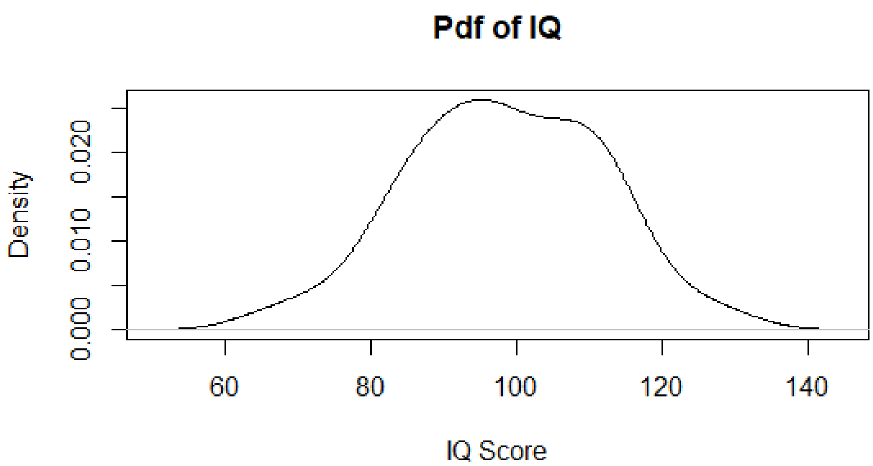


Figure 3: Plot Density Plot of IQ Scores.



From this code, it was calculated that the confidence interval was that the mean was between 93.96 and 102.92. This was confirmed when running the data through the t-test function

```
One Sample t-test

data:  y
t = 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
98.44
```

Now for the hypothesis test. The first step involves making assumptions about your data. As shown from the data visualisations in the last part, the data is not normally distributed and has a small sample size. Step two involves creating your null and alternative hypotheses. In this case, the hypotheses would be as follows:

```
# H0: sample mean <= 100
# H1: sample mean > 100
```

Step three is the test statistic, which was obtained from the the following code, which produced the results below:

```
> t.test(y, conf.level = 0.95, alternative = "greater", mu = 100)

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
```

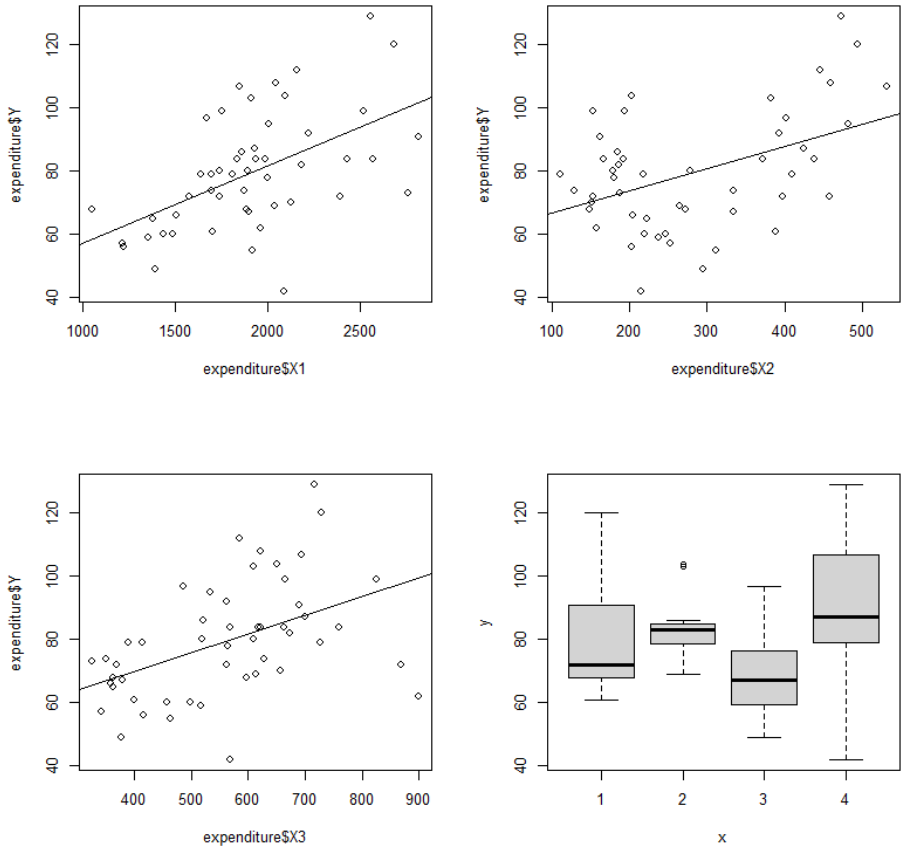
This p value is greater than .05, meaning we fail to reject the null hypothesis. This implies that the sample mean is either less than or equal to 100, the population mean for IQ Scores.

Question 2

Figure 4 contains three scatterplots which show the relationships between Y and the different X variables.

All of these scatterplots display a positive relationship, with the sample slope being positive on all tables. However, these relationships would be considered moderate, with the above data having correlations of $r = .53$, $r = .44$, and $r = .46$ respectively, which would be considered moderate correlations. This was done by calculating the square root of r from

Figure 4: Required Graphs.



Tables 1-3 shown below.

Also, the boxplot in Figure 4 displays that Region 4 has the highest mean spending compared to the other regions.

As previously shown in Table 1, Y and X1 have a moderate positive relationship, with the estimated slope being 0.025. The R squared value is 0.283, meaning that the regression model can only account for 28.3 percent of the variability in the data.

The graph in Figure 5 displays the expenditure depending on the region, with each region being represented by a different colour, as shown by the key to the right of the graph.

Table 1

	Dependent variable:
	Y
X1	0.025***
	(0.006)
Constant	32.546***
	(11.034)
Observations	50
R^2	0.283
Adjusted R^2	0.268
Residual Std. Error	15.836 (df = 48)
F Statistic	18.920*** (df = 1; 48)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2

	Dependent variable:
	Y
X2	0.070***
	(0.020)
Constant	59.761***
	(6.164)
Observations	50
R^2	0.201
Adjusted R^2	0.184
Residual Std. Error	16.714 (df = 48)
F Statistic	12.072*** (df = 1; 48)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3

	Dependent variable:
	Y
X3	0.059***
	(0.016)
Constant	46.306***
	(9.461)
Observations	50
R^2	0.215
Adjusted R^2	0.199
Residual Std. Error	16.567 (df = 48)
F Statistic	13.146*** (df = 1; 48)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 5: Scatterplot of Y against X1