Applied Statistical Analysis I: Problem Set One

Sophie McDonnell

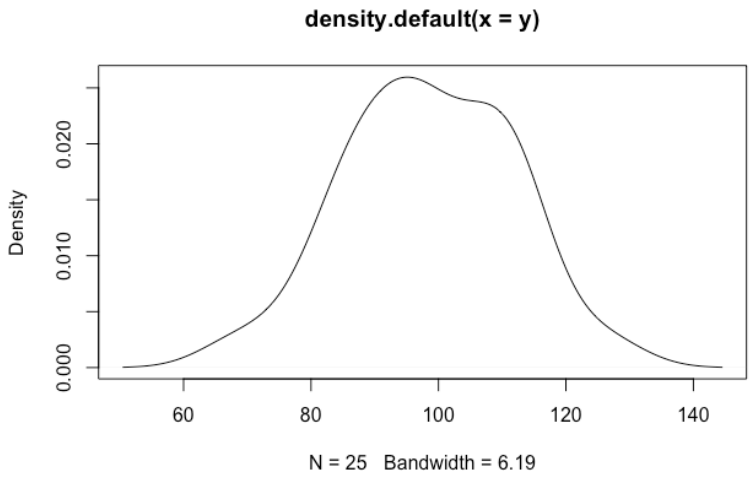
15328561

Question One:

Part 1:

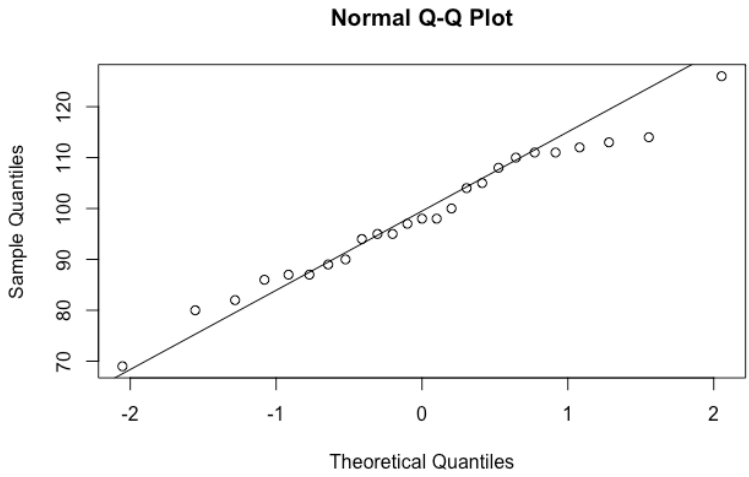
Code:

```
plot(density(y))
qqnorm(y)
qqline(y, distribution = qnorm)
CI_lower <- qnorm(0.05,
        mean = mean(y),
        sd = (sd(y)/sqrt(length(y))))
CI_upper <- qnorm(0.95,
        mean = mean(y),
        sd = (sd(y)/sqrt(length(y))))
matrix(c(CI_lower, CI_upper), ncol = 2,dimnames = list("",c("Lower", "Upper")))
var.test(y, ratio = 1, alternative = "two.sided", conf.level = 0.95)
```



Confidence interval: Lower: 94.13283 Upper: 102.7472

Difference between Average IQ and National Average IQ are not statistically significant.



Question two:

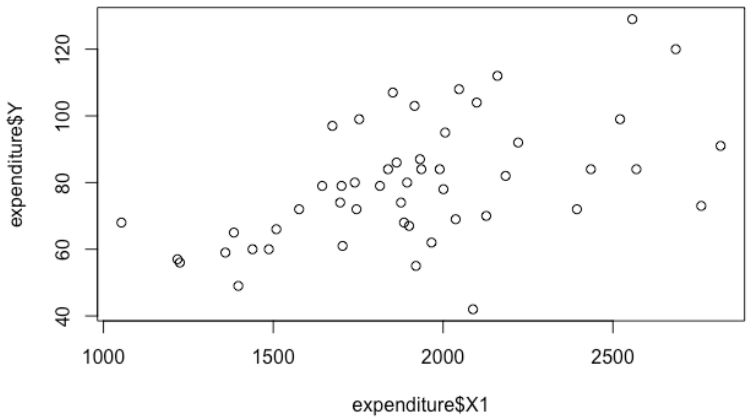
Part 1:

Code:

```
expenditure <- read.table("~/Documents/GitHub/StatsI_Fall2022/datasets/expenditure.txt", header=T)

plot(expenditure$X1, expenditure$Y)

plot(expenditure$X2, expenditure$Y)

plot(expenditure$X3, expenditure$Y)
```

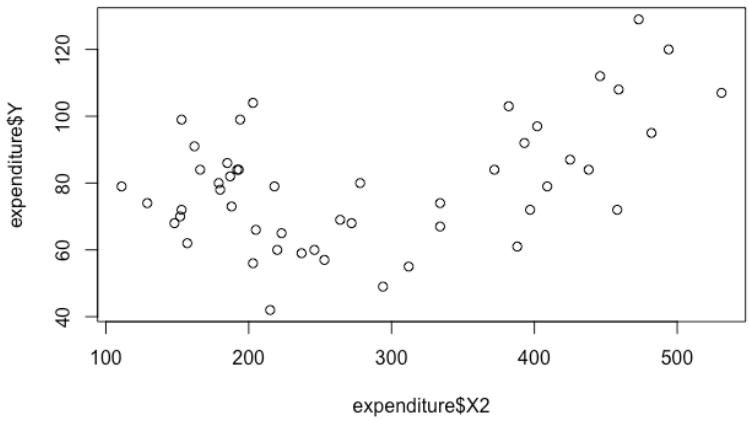
X1 and Y:

Per capita expenditure on shelters/housing assistance in state (Y) and per capita personal income in state (X1) has a positive linear moderate correlation.



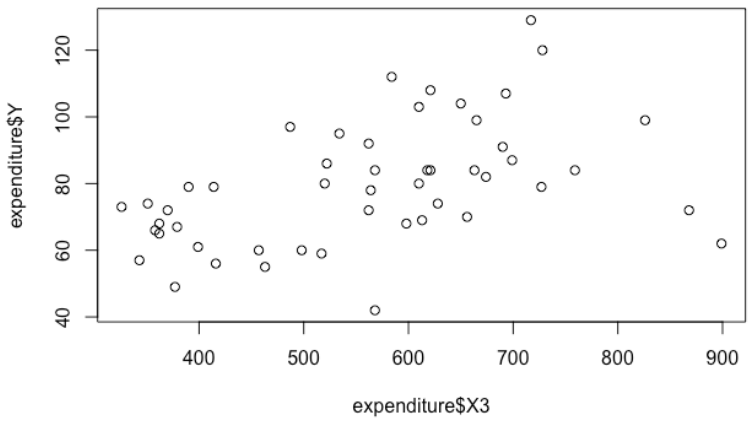
X2 and Y:

Per capita expenditure on shelters/housing assistance in state (Y) and the number of residents per 100,00 that are financially insecure in state (X2) has a positive, non-linear relationship (potentially quadratic).



X3 and Y:

Per capita expenditure on shelters/housing assistance in state (Y) and the number of residents per thousand residing in urban areas has a positive, linear, moderate correlation.

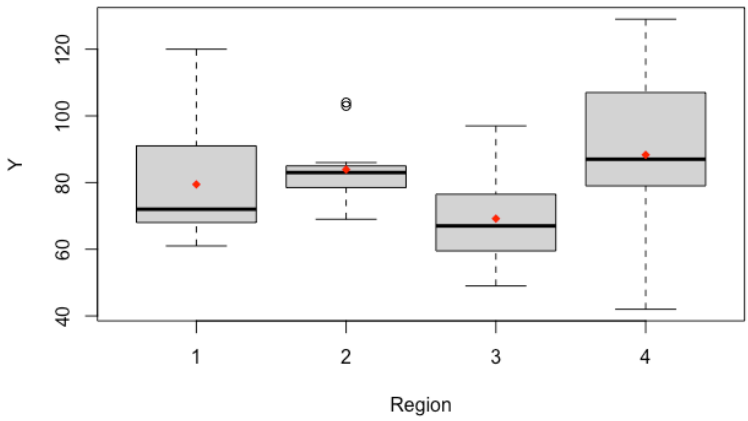


Part Two:

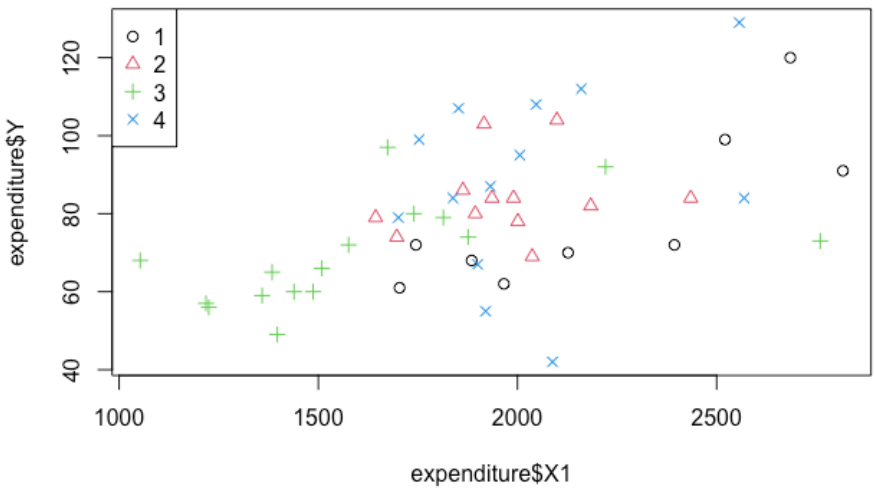
Code:

```
boxplot(Y~Region, data=expenditure)
means <- tapply(expenditure$Y, expenditure$Region, mean)
points(means, col="red", pch=18)
means
```

On average, Region 4 has the highest per capita expenditure on housing assistance, with about 88.3 $ per capita.



Part Three:

Code:

```
plot(expenditure$X1, expenditure$Y, col=expenditure$Region, pch=expenditure$Region)
legend("topleft", legend=1:4, col=1:4, pch=1:4)
```