A5 (Q3) - Ram Yoogesh Gopu (20867060)

```
knitr::opts_chunk$set(echo = TRUE,
                      warning = FALSE,
                      message = FALSE,
                      fig.align = "center",
                      fig.width = 6,
                      fig.height = 5,
                      out.height = "40%")
set.seed(12314159)
library(loon.data)
library(loon)
```

```
## Loading required package: tcltk
```

```
library(gridExtra)

codeDirectory <- "../../img"
imageDirectory <- "./img"
dataDirectory <- "./data"
path_concat <- function(path1, ..., sep="/") paste(path1, ..., sep = sep)
```

(3)

The full data set is then read in as

```
labData <- read.csv("labData.csv")
```

(A)

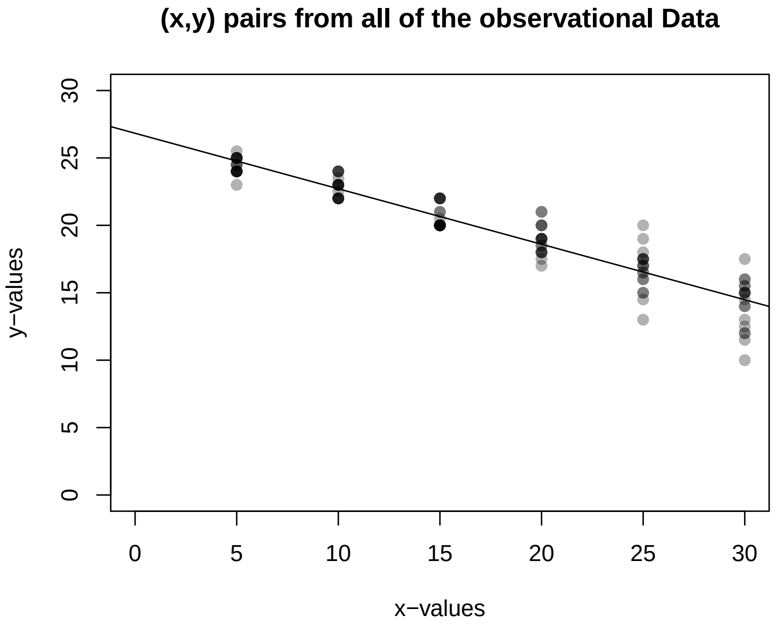
```
observational <- labData[labData$type == "observational", ]
```

(4)

(B)

```
plot(observational$x, observational$y, xlim = c(0, 30), ylim = c(0, 30), pch = 19, col = adjustcolor("bl

Bmod <- lm(y~x, observational)
abline(Bmod)
```



```
#print(Bmod$coefficients)
print("Slope Estimate is ")
```

```
## [1] "Slope Estimate is "
```

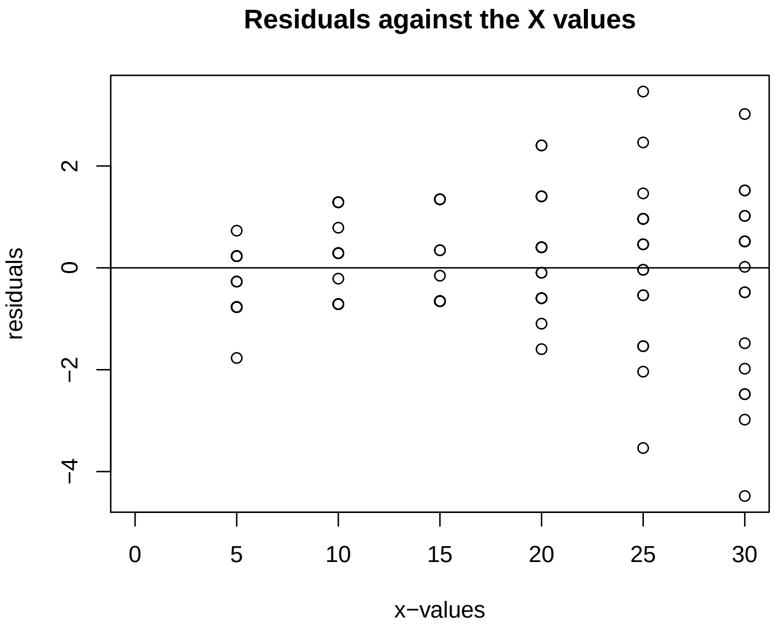
```
print(Bmod$coefficients[2])
```

```
##          x
## -0.4115873
```

(C)

(i)

```
plot(observational$x, Bmod$residuals, xlim = c(0, 30), abline(h = 0), main = "Residuals against the X va
```



(ii)

From the above residual plot, we can infer that the range of the residuals increases when the value of x increases. I would say that the measuring system is baised towards the low X values.

(D)

(i)

```
fit <- array()
estimates <- c()

for (value in 1:18){
  fit[value] <- lm(y~x, observational[observational$team == value, ])
  estimates <- c(estimates, fit[[value]][2])
}
print("Average of estimated slopes is ")
```

```
## [1] "Average of estimated slopes is "
```

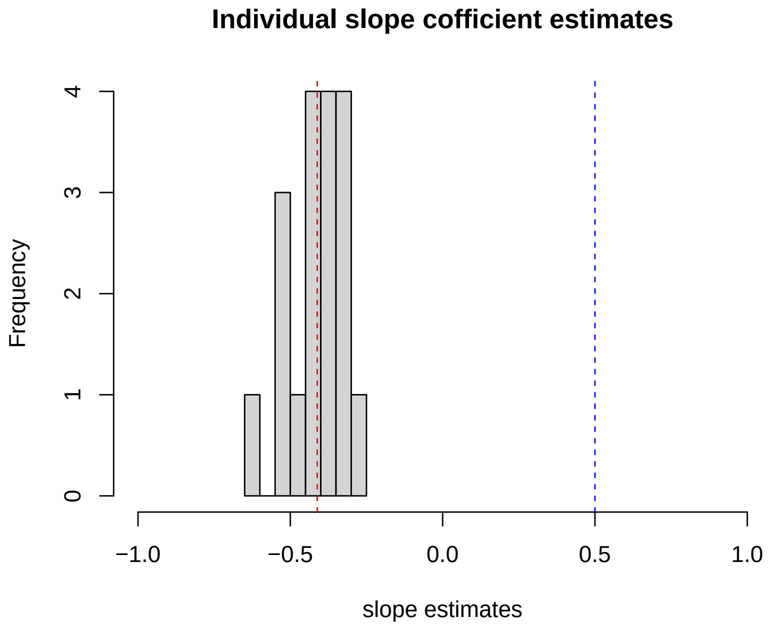
```
print(mean(estimates))
```

```
## [1] -0.4115873
```

(ii)

(iii)

```
hist(estimates, xlim = c(-1, 1), col = "lightgrey", main = "Individual slope cofficient estimates", xla
abline(v = mean(estimates), col = "red", lty = 2)
abline(v = 0.5, col = "blue", lty = 2)
```



(5)

(E)

```
print(estimates)
```

```
##          x          x          x          x          x          x          x 
## -0.3628571 -0.3828571 -0.5314286 -0.3257143 -0.3771429 -0.4000000 -0.6171429 
##          x          x          x          x          x          x          x 
## -0.3314286 -0.4028571 -0.4314286 -0.2885714 -0.5200000 -0.5000000 -0.4828571 
##          x          x          x          x 
## -0.3342857 -0.4085714 -0.3628571 -0.3485714
```

From the observational study and seeing the estimate values, it is evident that the estimates lies between -0.6171 and -0.2885. On the contrary the true value is 0.5. Hence i would conclude that the quality of slope estimates is bad.

(F)

Since Z is a lurking variable for the above problem, it is clear that it has a fixed value which fixes (hyperplane). This has an effect on the values of y. Also, the hyperplane and the height of the markers imposes a constraint on y.