Assignment 1

1) A 2012 Pew Research survey asked 2,373 randomly sampled registered voters their political affiliation (Republican, Democrat, or Independent) and whether or not they identify as swing voters. 35% of respondents identified as Independent, 23% identified as swing voters, and 11% identified as both.

a) Are being Independent and being a swing voter disjoint, i.e. mutually exclusive?

No, being Independent and being a swing voter are not disjoint because both these events can happen at the same time.

b) What percent of voters are Independent but not swing

P(I) = 35%
P(I and S) = 11%
P(S) = 23%
P(I and S') = 35 -11 = **24%**

Hence, 24% of voters are independent but not swing.

c) What percent of voters are Independent or swing voters?

P(I) = 35%
P(S) = 23%
P(I and S) = 11%
P(I or S) = P(I) + P(S) – P(I and S)
= 35 + 23 – 11
= **47%**

Hence, 47% of voters are independent or swing voters.

d) What percent of voters are neither Independent nor swing voters?

P(I or S)' = 100 – 47 = **53%**

Hence, 53% of voters are neither independent nor swing.

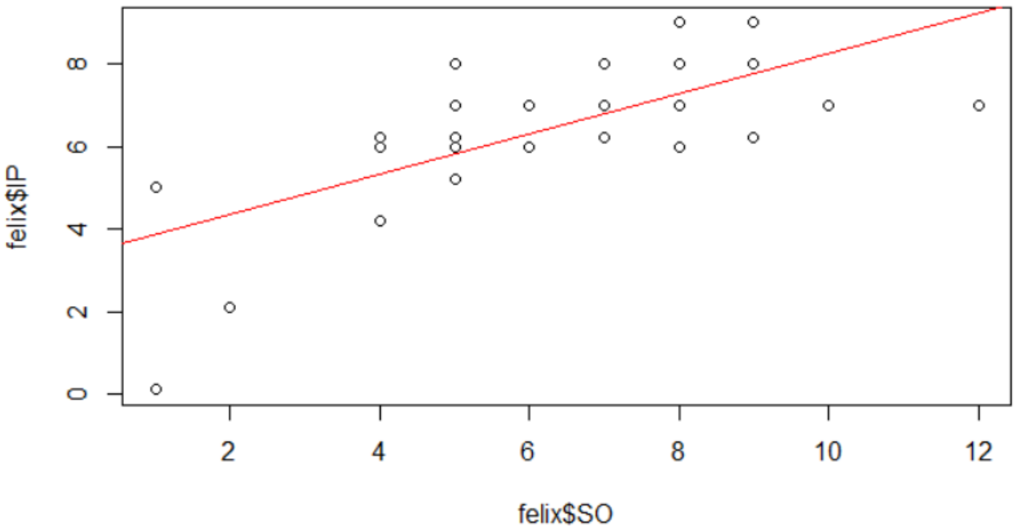
e) Is the event that someone is a swing voter independent of the event that someone is a political Independent?

No, they are not independent. For two events to be independent P(A and B) = P(A)P(B)
P(I and S) = 0.11 = 11%
P(I)P(S) = 0.35*0.23 = 0.0805 =8.05%
Since P(I and S) is not equal to P(I)P(S), the two events are not independent.

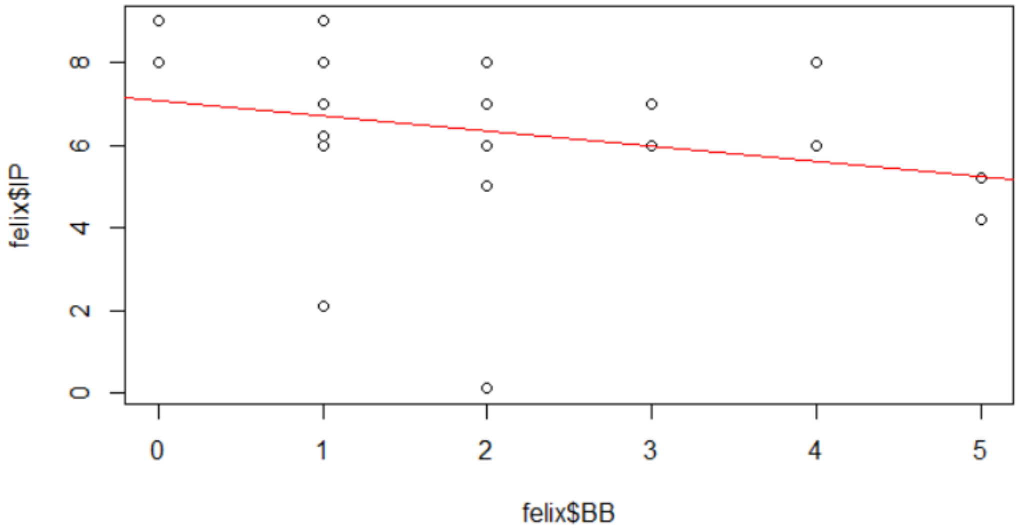
2) To execute the code, I have attached the R file.

```
setwd("D:/UW/Quarter 2/INFX 573/Assignment")
```

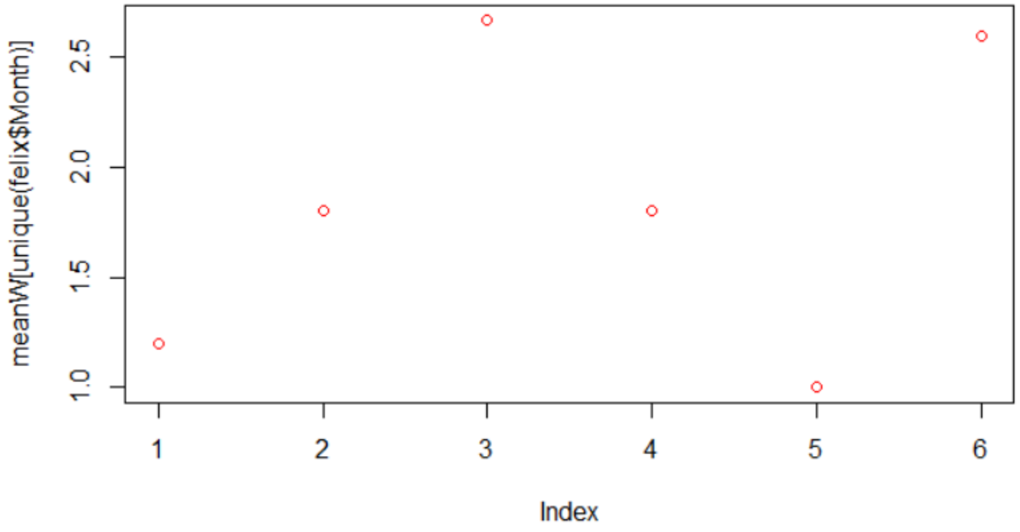
```
felix <- read.csv("FelixHernandez2015.csv")
str(felix)
felix[1:5,]
# 2(a) starts
nrow(felix[felix$W==1,])
# 2(a) ends: 18

# 2(b) starts
rMode <- function(x){
  ux <- unique(x)
  ux[which.max(tabulate(match(x, ux)))]
}

mean(felix$SO)
median(felix$SO)
rMode(felix$SO)
# mean: 6.16129, median: 6, mode: 5
# 2(b) ends

# 2(c) starts
plot(felix$IP ~ felix$SO)
model <- lm(felix$IP ~ felix$SO)
abline(model, col =2 )
```



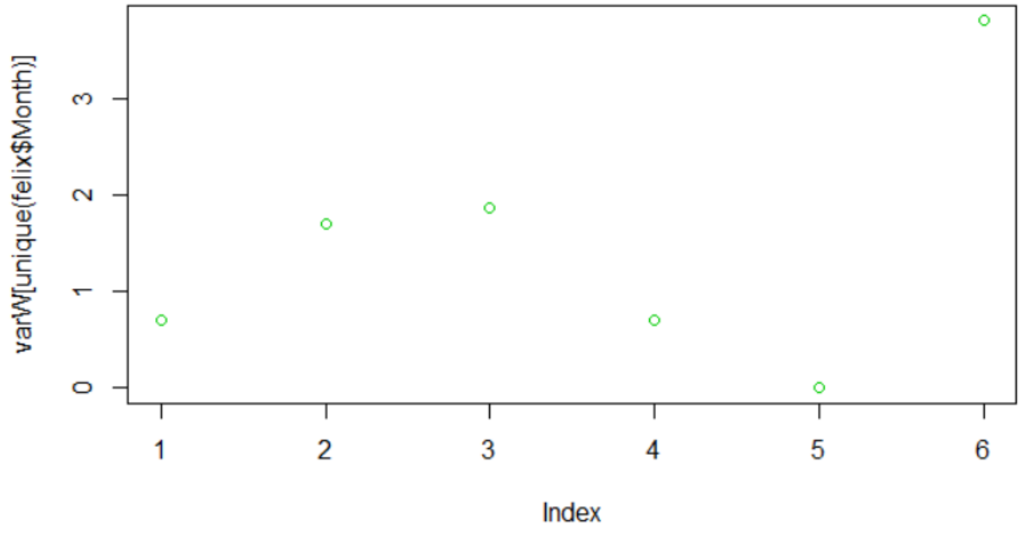
```
plot(felix$IP ~ felix$BB)
model1 <- lm(felix$IP ~ felix$BB)
abline(model1, col =2 )
```



```
# As number of innings pitched increase, strike outs also increase.
# As number of innings pitched increase, base on balls decrease.
# 2(c) ends

# 2(d) starts
corIpSo <- cor(felix$IP, felix$SO)
# correlation between IP and SO: 0.6816081
corIpBb <- cor(felix$IP, felix$BB)
# correlation between IP and BB: -0.2638496
# Yes, these align with the plots. Correlation between IP and SO is positive as graph linearly
#increases.
# Correlation between IP and BB is negative as graph linearly decreases.
# 2(d) ends

#2(e) starts
meanW <- by(felix$BB,felix$Month,mean)
plot(meanW[felix$Month], col=2)
```



```
varW <- by(felix$BB,felix$Month,var)
plot(varW[felix$Month], col=3)
```



```
cor(meanW, varW)
# Since corelation between mean and variance is positive, variance increases with mean.
#It means that the walks are very much spread out from the mean for months - Apr, May and
#Jun. For the next two months,
# walks are again close to the mean.

#2(e) ends
```

```
#2(f) starts
table(felix$away, felix$W)
# 7 wins away from home and 11 wins at home. Therefore, there are more wins at home.
#2(f) ends

#2(g) starts
randy <- read.csv("RandyJohnson1995.csv")
str(randy)
sum(randy$SO) > sum(felix$SO)
# Randy outperformed Felix since number of strikeouts of Randy are more.
#2(g) ends

# Question 3
# 3(c) starts
curve(dnorm, from = -5, to=5)
abline(v=0.714, col="red")
abline(v=0.52, col="blue")
text(1.285714+1, 0.3, "Verbal: 0.714",col="red")
text(0.5215124-1.5, 0.1, "Quantitative: 0.52", col="blue")
# 3(c) ends

# 3(d) starts
pnorm(0.714)
# 0.762
pnorm(0.52)
# 0.698
#3(d) ends
```

3) Sophia who took the Graduate Record Examination (GRE) scored 156 on the Verbal Reasoning section and 157 on the Quantitative Reasoning section. The mean score for Verbal Reasoning section for all test takers was 151 with a standard deviation of 7, and the mean score for the Quantitative Reasoning was 153 with a standard deviation of 7.67. Suppose that both distributions are nearly normal.

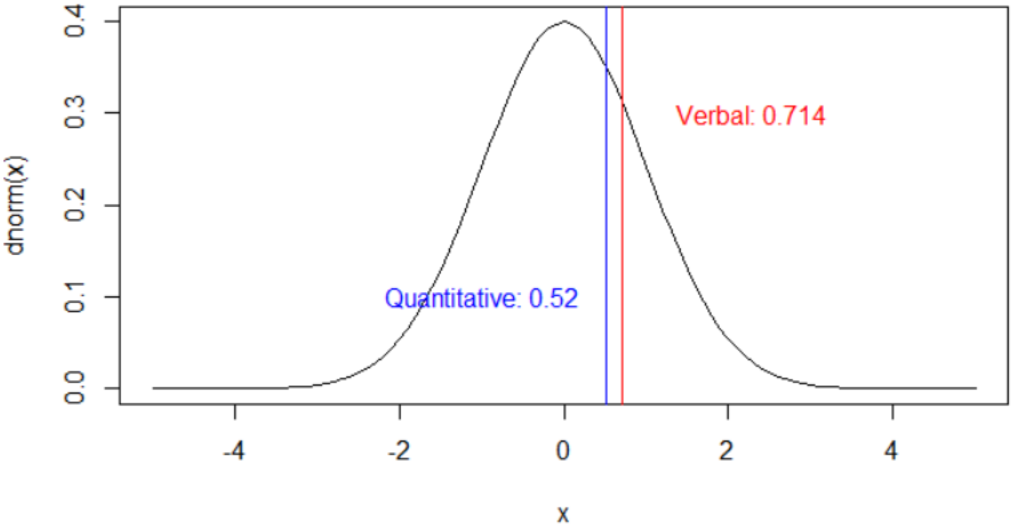
a) What is Sophia’s Z-score on the Verbal Reasoning section? On the Quantitative Reasoning section?

V = 156; Q = 157, mean(V) = 151; std(V) = 7; mean(Q) = 153, std(Q) = 7.67

Z(V) = (156 – 151) / 7 = 5/7 = **0.714**

Z(Q) = (157 – 153) / 7.67 = **0.52**

b) Draw a standard normal distribution curve and mark these two Z-scores.



c) Relative to others, which section did she do better on?

She performed better on **verbal** as she got a better percentile in verbal. We are comparing relative values and not absolute values.

d) Find her percentile scores for the two exams.

Percentile of verbal = **0.762 or 76.2%**
Percentile of quant = **0.698 or 69.8%**

e) What percent of the test takers did better than her on the Verbal Reasoning section? On the Quantitative Reasoning section?

Percent of test takers who did better than her on the Verbal Reasoning = (100 – 76.2) %
= **23.8%**

Percent of test takers who did better than her on the Quantitative Reasoning = (100 – 69.8) %
= **30.2%**

f) Explain why simply comparing her raw scores from the two sections would lead to the incorrect conclusion that she did better on the Quantitative Reasoning section (2-3 sentences).

Comparing her raw scores from two sections would lead to incorrect conclusion because we are concerned about her performance relative to others and not her absolute scores. For this reason, we check her percentile scores and not percentage scores.