IST772 Problem Set 4

Shivani Sanjay Mahaddalkar

The homework for week 4 is based is based on exercises 7-10 on page 66, but with changes as noted in the text in this notebook (i.e., follow the problems as given in this document and not the textbook).

Attribution statement: 2. I did this homework with help from the book and the professor and these Internet sources: a. https://stackoverflow.com/questions/12384071/how-to-coerce-a-list-object-to-type-double

Chapter 4, Exercise 7

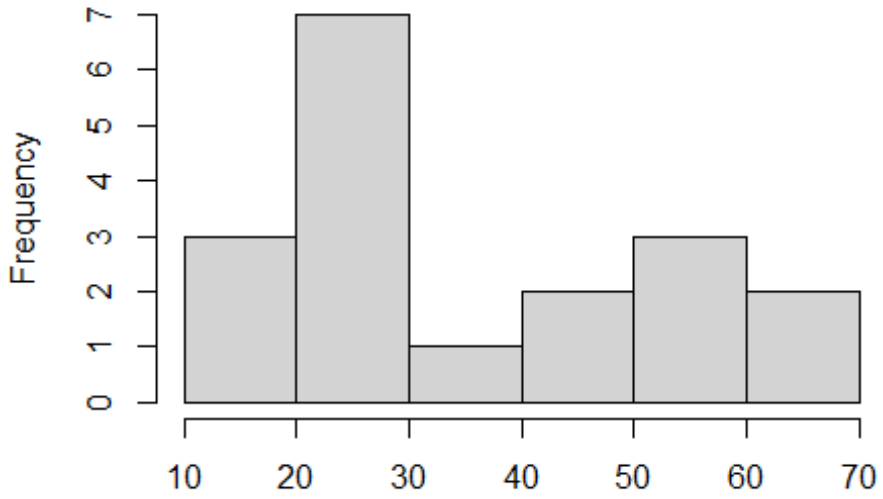
The built-in warpbreaks data set contains data for the number of warp breaks per loom with different amounts of tension (we will not consider the variable for the type of wool). The tensions are labelled "L", "M" or "H" for low, medium and high tension. Run the summary() command on warpbreaks and explain the output. Create a histogram of the breaks for low tension (1 pt). As a reminder about R syntax, here is one way that you can access the low tension data:

```
summary(warpbreaks)
#The summary command shows that there are 3 variables in the data. "Breaks"
is the number of Breaks in yarn, wool is the type of wool (A or B) #and
tension is the level of tension(Low, medium or High)

hist(warpbreaks$breaks[warpbreaks$tension=="L"]) #Creating a histogram of the
breaks for low tension.
```

Using the dplyr package, you can instead write:

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

hist(as.integer(unlist(warpbreaks %>% filter(tension == "L") %>%
select(breaks))))
```

nteger(unlist(warpbreaks %>% filter(tension == "L") '

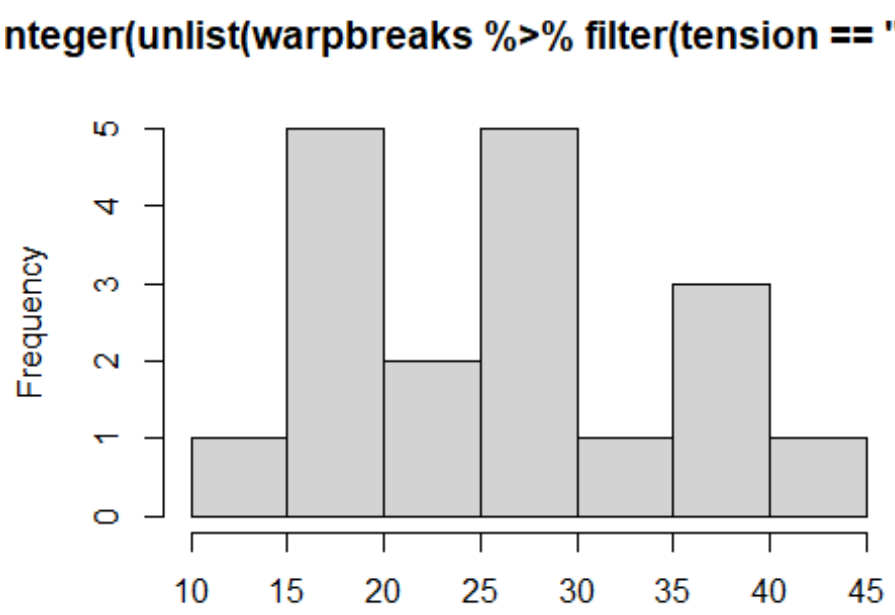


s.integer(unlist(warpbreaks %>% filter(tension == "L") %>% select(bre

(Note that a select function is defined in multiple packages, so if you want to be sure you're using the one from the dplyr library, call dplyr::select.)

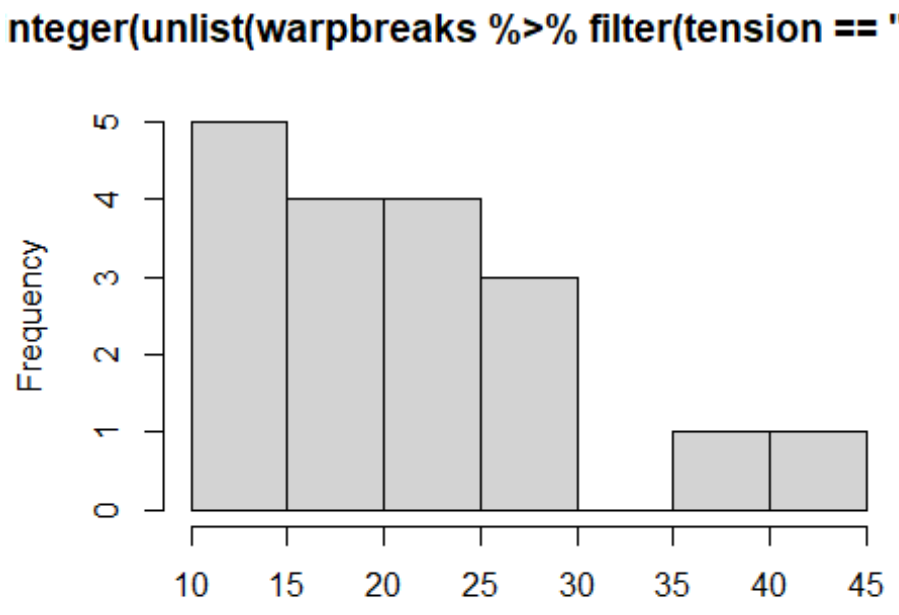
Also create histograms of the breaks for medium and high tensions. What can you say about the differences in the effects of tension by looking at the histograms? (1 pt)

```
hist(as.integer(unlist(warpbreaks %>% filter(tension == "M") %>%
select(breaks))))
```



s.integer(unlist(warpbreaks %>% filter(tension == "M") %>% select(bre

```
hist(as.integer(unlist(warpbreaks %>% filter(tension == "H") %>%
select(breaks))))
```



s.integer(unlist(warpbreaks %>% filter(tension == "H") %>% select(bre

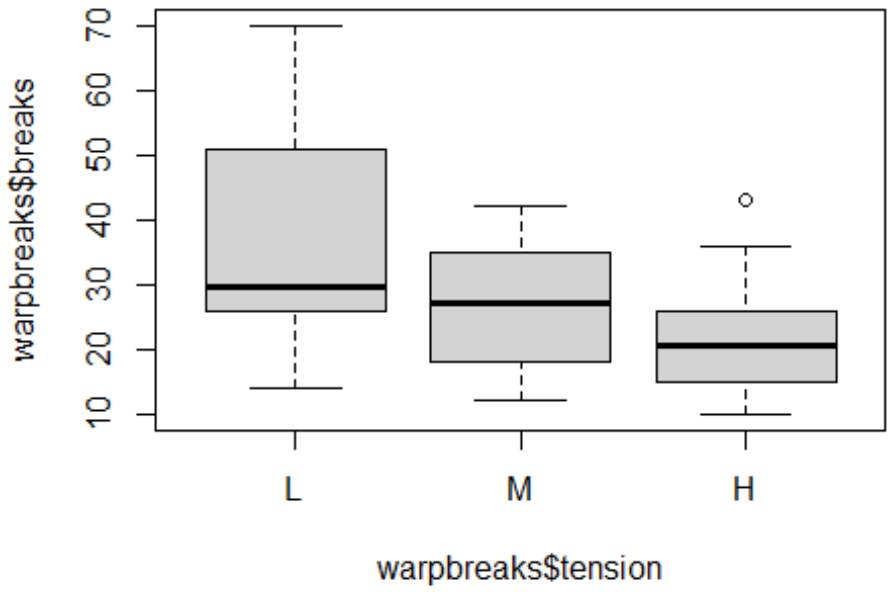
Looking at the three histograms, it can be noticed that the most number of breaks for the H level of

tension happen within the 10-15 region followed by 15-25. For M, the most number of breaks happen between 15-20 and 25-30. For L, the most number of breaks happen between 20-30. The scale for L type of tension goes upto 70 number of breaks, whereas for the other two it goes upto 45. It is visually seen that the L level of tension shows higher frequency at a higher number of breaks. The H level of tension has lower range for the highest number of breaks.

Chapter 4, Exercise 8

Create a boxplot (or violin plot) of the breaks data, using the model "breaks ~ tension". (1 pt) What can you say about the differences in the tensions by looking at the boxplots for the different tensions? (1 pt)

```
boxplot(warpbreaks$breaks~warpbreaks$tension)
```



As it can be seen by the boxplot, it shows that L has a higher median for the the number of breaks followed by M, followed by H. The M and H yarns have a lower upper bound for the number of breaks as compared to L. H has one outlier which is beyond the upper whisker of the barplot. H has lesser number of breaks of the three and, if the quality is judged by having lesser number of breaks, the L yarn is a better fit.

Chapter 4, Exercise 9

Run a t-test to compare the means of high and medium tension in the warpbreaks data. (1 pt) Report and interpret the confidence interval. (1 pt) Make sure to include a carefully worded statement about what the confidence interval implies with respect to the population mean difference between the high and medium tensions. (1 pt)

```
warpbreaksH <- warpbreaks %>% filter(tension == "H") %>% select(breaks)
warpbreaksM <- warpbreaks %>% filter(tension == "M") %>% select(breaks)
warpbreaksL <- warpbreaks %>% filter(tension == "L") %>% select(breaks)
t.test(warpbreaksH, warpbreaksM)
```

```
##
##  Welch Two Sample t-test
##
## data:  warpbreaksH and warpbreaksM
## t = -1.6199, df = 33.74, p-value = 0.1146
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.648042   1.203597
## sample estimates:
## mean of x mean of y
##  21.66667  26.38889
```

The confidence interval is (-10.648, 1.2036) The confidence interval here implies nothing with certainty. If this experiment is run enough number of times the population difference between the mean difference will lie in the confidence intervals 95% of the time. So it suggests that there is a high probability(95%) that the population mean difference lies within the confidence interval, however, there is no certainty attached to it.

Chapter 4, Exercise 10

Run a t-test to compare the means of low and medium tension in the warpbreaks data. (1 pt) Report and interpret the confidence interval. (1 pt + 1 pt for statement about means)

```
t.test(warpbreaksL, warpbreaksM)
```

```
##
##  Welch Two Sample t-test
##
## data:  warpbreaksL and warpbreaksM
## t = 2.256, df = 26.554, p-value = 0.03252
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   0.8976796 19.1023204
## sample estimates:
## mean of x mean of y
##  36.38889  26.38889
```

The confidence interval is (-0.8977, 19.1023204) The confidence interval here implies nothing with certainty. If this experiment is run enough number of times the population difference between the mean difference will lie in the confidence intervals 95% of the time. So it suggests that there is a high probability(95%) that the population mean difference lies within the confidence interval, however, there is no certainty attached to it.