MGT256_LAB1_Monika

Monika Baloda

2022-10-05

tinytex::install_tinytex()

```
library("readxl")
setwd("F:/UCR/Academics/Fall2022/MGT256_Business_Analytics/Homeworks/Lab/Lab1")
```

Question-1

We are installing a library package epiDisplay to make it easy to find frequency table.

```
health=read_excel("jaggia_ba_1e_ch03_Health.xlsx")
#install.packages('epiDisplay')
library(epiDisplay)
```

```
## Loading required package: foreign
```

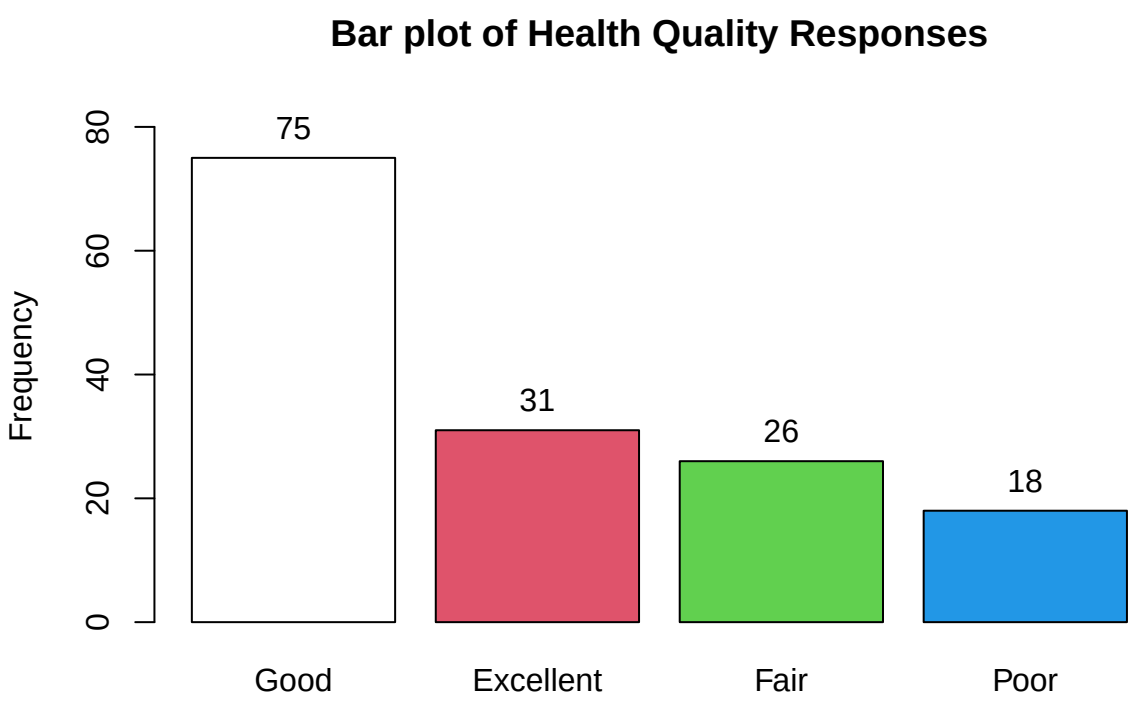
```
## Loading required package: survival
```

```
## Loading required package: MASS
```

```
## Loading required package: nnet
```

Part a) ***Answer***: The most common response is 'Good'. It is the 50 percent of the total. We show the table below:

```
tab1(health$Quality, sort.group = "decreasing", cum.percent = TRUE,
      xlab="Health Quality Response", ylab="Frequency",
              main="Bar plot of Health Quality Responses")
```




```
## health$Quality :
##           Frequency Percent Cum. percent
## Good             75    50.0         50.0
## Excellent        31    20.7         70.7
## Fair             26    17.3         88.0
## Poor             18    12.0        100.0
##   Total         150   100.0        100.0
```

Part b)

Answer: We note that most of the patient at this medical practice are in good or better quality health. From the table in part-a), 88 percent people are not in 'poor' quality health. Therefore one can say that overall health of patients is good.

```
barplot(table(health$Quality),ylim=c(0, 80), xlab="Health Quality Response", ylab="Frequency",
              main="Bar plot of Health Quality Responses")
```

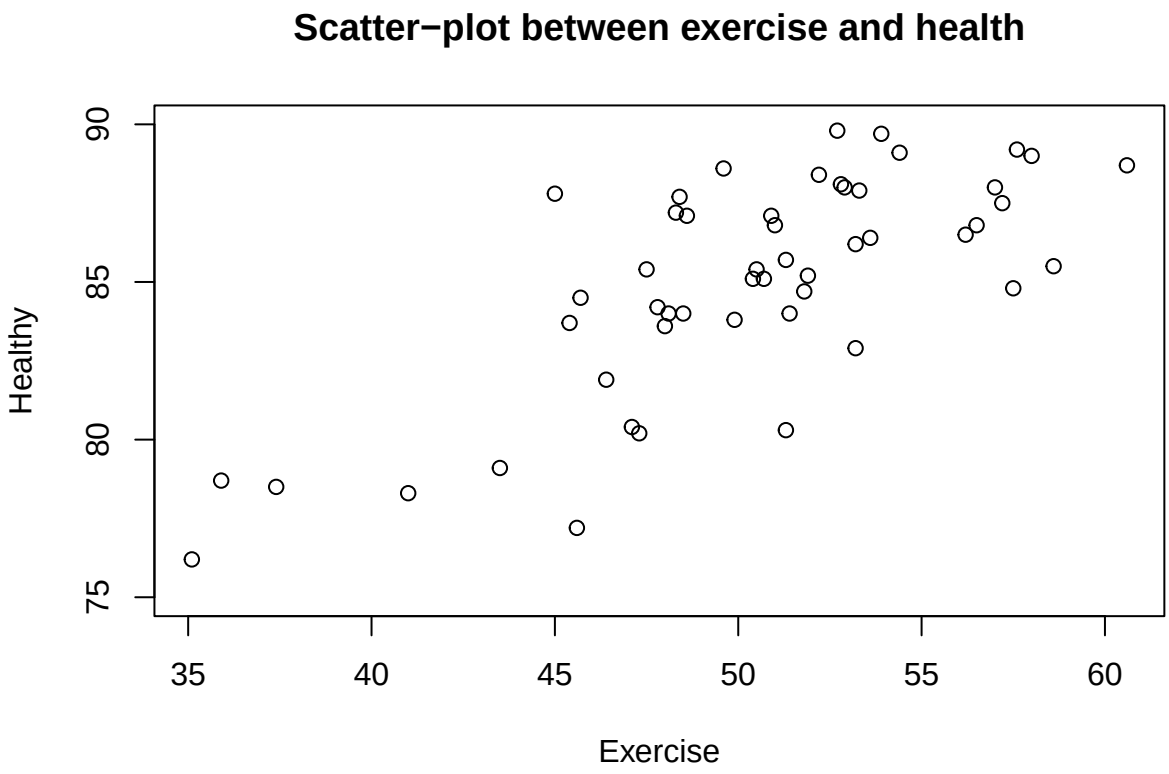
Question-2

```
healthy_liv=read_excel("jaggia_ba_1e_ch03_Healthy_living.xlsx")
```

a) Constructing a scatterplot of Health against Exercise:

Answer: We plot the aforementioned relationship below. We find that there appears to be positive relationship between the two variables. Further we set appropriate limit for variable on x axis using its maximum and minimum values. On y axis, we set the limits manually.

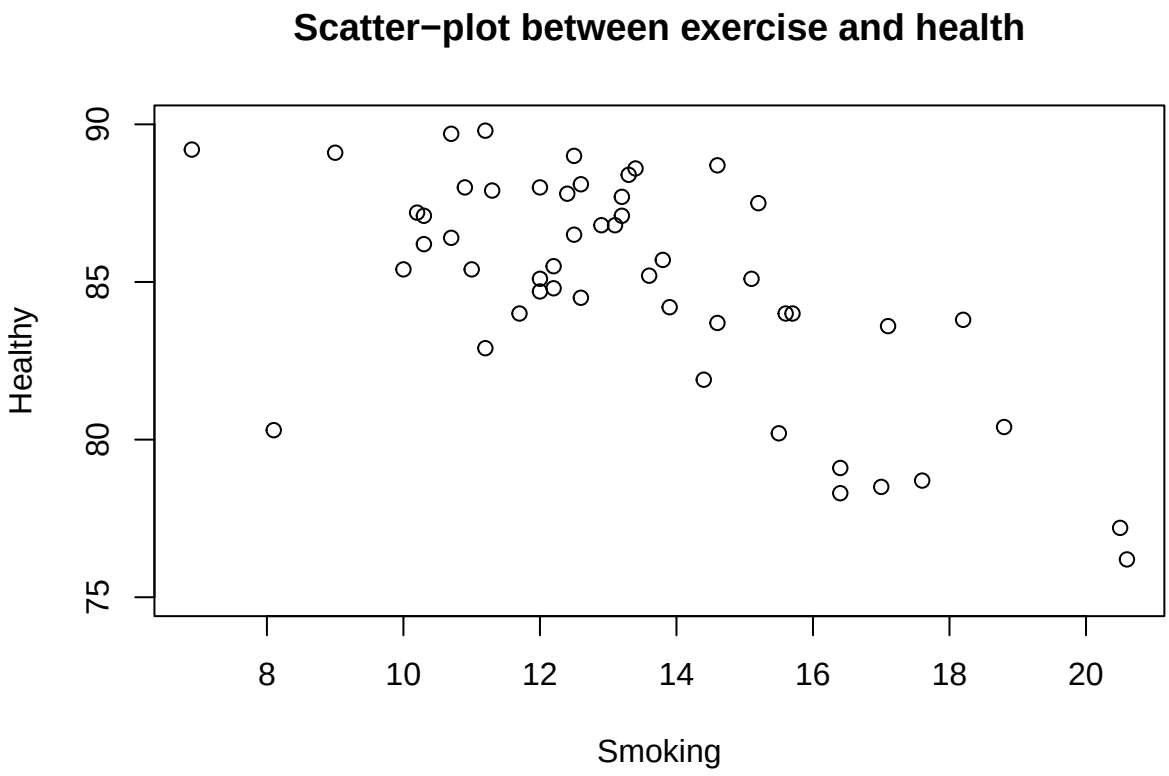
```
plot(healthy_liv$Exercise,healthy_liv$Healthy, xlab="Exercise", ylab="Healthy",
     main="Scatter-plot between exercise and health", ylim=c(75, 90),
     xlim=c(min(healthy_liv$Exercise),max(healthy_liv$Exercise)))
```



b) Constructing a scatterplot of Health against Smoking

Answer: We plot the aforementioned relationship below. We find that there appears to be a negative relationship between the two variables. Further we set appropriate limit for variable on x axis using its maximum and minimum values. On y axis, we set the limits manually.

```
plot(healthy_liv$Smoking,healthy_liv$Healthy, xlab="Smoking", ylab="Healthy",
     main="Scatter-plot between exercise and health", ylim=c(75, 90),
     xlim=c(min(healthy_liv$Smoking),max(healthy_liv$Smoking)))
```



Question-3

```
prime=read_excel("jaggia_ba_1e_ch03_Prime.xlsx")
```

a) mean and the median of annual expenditures

Answer: The mean of amazon prime expenditure is 1306.94 and median is 1287.5

```
mean(prime$Expenditures)
```

```
## [1] 1306.94
```

```
median(prime$Expenditures)
```

```
## [1] 1287.5
```

b) first quartile and the third quartile for annual expenditures.

Answer: The first quartile is 1118.75, this means that if we arrange all the customers in ascending order and if number of customers is 100 then the value of expenditure of 25th customer is 1118.75. Similarly, the third quartile is 1574.25 which is the expenditure of 75th customer in order. Further if there are suppose 1000 customers then 1st quartile is the expenditure of 250th person in order i.e. 25% of the people spend less than this person's expendtire.

```
first_quartile=summary(prime$Expenditures)[2]
third_quartile=summary(prime$Expenditures)[5]
print(first_quartile)
```

```
## 1st Qu.
## 1118.75
```

```
print(third_quartile)
```

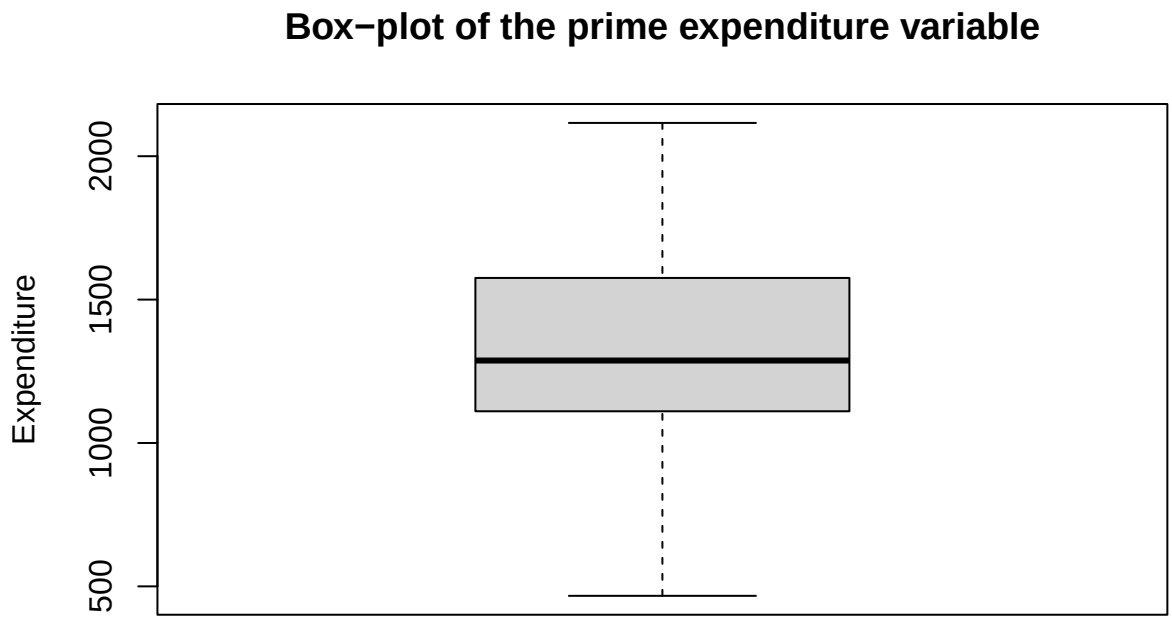
```
## 3rd Qu.
## 1574.25
```

Question-4 We are still using 'prime' dataset in this question, therefore we do not load it again.

a) boxplot of expenditure variable

Answer We make the boxplot below. The box plot does not suggest the existence of outliers.

```
boxplot(prime$Expenditures, ylab="Expenditure",
        main="Box-plot of the prime expenditure variable")
```



b) z-score ***Answer*** We first construct z-score and then see if the datapoints are outside the interval [-3,3]. Outlier means either greater than mean+3xSD or less than mean-3xSD. But in Z score mean=0, SD=1. So, if a point outside [-3,3] interval i.e. either >3 or <-3, then we call it a outlier. We do not find any outlier as minimum and maximum value are in [-3,3] interval.

```r
x_bar=mean(prime$Expenditures)
mean=rep(x_bar, length=length(prime$Expenditures))
sigma=sd(prime$Expenditures)
sd=rep(sigma, length=length(prime$Expenditures))
prime$z=(prime$Expenditures)
prime$z=prime$z-mean
prime$z=prime$z/sd

summary(prime$z)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -2.49881 -0.55986 -0.05783  0.00000  0.79524  2.40695
```