All students are requested to complete this activity and submit before 05.02.2021
**************************Assignment No: 02 Descriptive Statistics**************************

Name: Vaibhav Jumde

Date: 5 Feb.2021

Following are the published weights (in pounds) of all of the team members of the San Francisco 49ers from a previous year (Source: San Jose Mercury News).

177; 205; 210; 210; 232; 205; 185; 185; 178; 210; 206; 212; 184; 174; 185; 242; 188; 212; 215; 247; 241; 223; 220; 260; 245; 259; 278; 270; 280; 295; 275; 285; 290; 272; 273; 280; 285; 286; 200; 215; 185; 230; 250; 241; 190; 260; 250; 302; 265; 290; 276; 228; 265

1. Organize the data from smallest to largest value.

```
 >x=c(177,205,210,210,232,205,185,185,178,210,206,212,184,174,185,242,188,212,215,247,24
1
,223,220,260,245,259,278,270,280,295,275,285,290,272,273,280,285,286,200,215,185,230,250
, 241,190,260,250,302,265,290,276,228,265)

> length(x)
[1] 53

> sort(x)
 [1] 174 177 178 184 185 185 185 185 188 190 200 205 205 206 210 210 210 212 212
215 [21] 215 220 223 228 230 232 241 241 242 245 247 250 250 259 260 260 265 265
270 272 [41] 273 275 276 278 280 280 285 285 286 290 290 295 302
```

2. Find the median.

```
> median(x)
[1] 241
```

3. Find the first quartile.

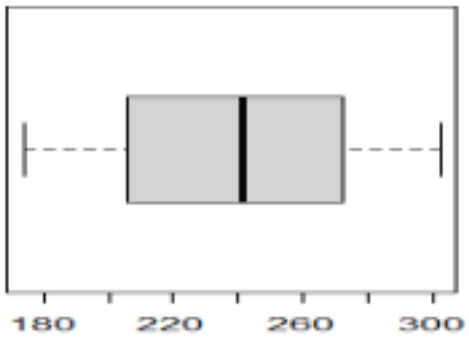
```
> q1=quantile(x,0.25)
> q1
25%
206
```

4. Find the third quartile.

```
> q3=quantile(x,0.75)
> q3
75%
272
```

5. Construct a box plot of the data.

```
> boxplot(x,horizontal = T)
```



6. The middle 50% of the weights are from 206 to 272.

7. If our population were all professional football players, would the above data be a sample of weights or the population of weights? Why?

The above data would be a sample of weights because they represent a subset of the population of all football players.
It represents a sample. Supposing Football Team "San Francisco 49ers" is all professional, they can be considered a sample from all professional football players.

8. If our population were the San Francisco 49ers, would the above data be a sample of weights or the population of weights? Why? i. Assume the population was the San Francisco 49ers.

The data would be a population of weights because they represent all of the players in the Football Team "San Francisco 49ers".

Find

(i) The population mean, μ

```
> mean(x)
[1] 236.3396
```

(ii) The population standard deviation, σ

```
> v=var(x)
> v
[1] 1433.152
> sqrt(v)
[1] 37.85699
```

(iii) The weight that is 2 standard deviations below the mean.

$$\bar{x} - 2s = 236.3396 - (2 \times 37.85699) = 160.63$$

9. When Steve Young, quarterback, played football, he weighed 205 pounds. How many standard deviations above or below the mean was he?

$$\bar{x} - \text{Young} = 205$$

$$\therefore z = \frac{\bar{x} - 205}{s}$$

$$= \frac{236.3396 - 205}{37.85699} = 0.8278$$

He was 0.8278 standard deviations below the mean.

10. That same year, the average weight for the Dallas Cowboys was 240.08 pounds with a standard deviation of 44.38 pounds. Emmit Smith weighed in at 209 pounds. With respect to his team, who was lighter, Smith or Young? How did you determine your answer?

$$\bar{x} - \text{Young} = \frac{x - \mu}{\sigma}$$

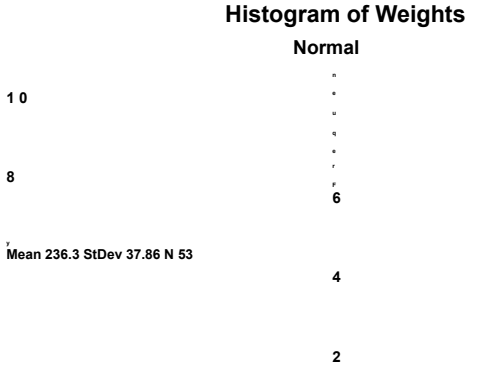
$$= \frac{205 - 236.3396}{37.85699} = -0.8278$$

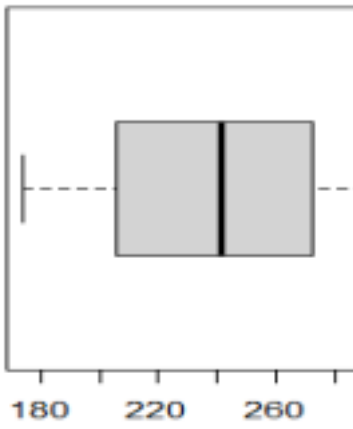
$$\bar{x} - \text{Smith} = \frac{x - \mu}{\sigma}$$

$$= \frac{209 - 240.08}{44.38} = -0.7003$$

Since, z-score of Young is less than z-score of Smith; therefore Steve **Young is lighter than Emmit Smith**.

11. Based on the shape of the data, what is the most appropriate measure of center for this data: mean, median, or mode? Explain.





From Histogram and Box-plot we can see that the data is approximately **Bell-shaped** (that is data is normally distributed). Therefore, the most appropriate measure of center for this data will be **mean**.

12. Are there any outliers in the data? Use an appropriate numerical test involving the IQR to identify outliers, if any, and clearly state your conclusion.

Here, for our data set we have,

First quartile (Q1)=206 and Third quartile(Q3)=272

Inter Quartile Range (IQR) = Q3-Q1 = 272-206 = 266

Now,

Lower Fence = Q1-(1.5*IQR) = 206-(1.5*66) = 107

Upper Fence = Q3+(1.5*IQR) = 272+(1.5*66) = 371

Any observation (Data point) outside of the interval (107,371) will be considered as an outlier.

Since, all the data values are within this interval, **there are no outliers in the data set**.

13. Are any data values further away than 2 standard deviations from the mean? Clearly state your conclusion and show numerical work to justify your answer.

$\bar{x} - 2s = 236.3396 - (2 \times 37.85699) = 160.66$

$\bar{x} + 2s = 236.3396 + (2 \times 37.85699) = 312.05$

No data values further away than 2 standard deviations from the mean. All data values are within 2 standard deviations from the mean.