

NAME : SACHIN

Q1) Identify the Data type for the Following:

Activity	Data Type
Number of beatings from Wife	Discrete
Results of rolling a dice	Discrete
Weight of a person	Continuous
Weight of Gold	Continuous
Distance between two places	Continuous
Length of a leaf	Continuous
Dog's weight	Continuous
Blue Color	Qualitative
Number of kids	Discrete
Number of tickets in Indian railways	Discrete
Number of times married	Discrete
Gender (Male or Female)	Qualitative

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

Data	Data Type
Gender	Nominal
High School Class Ranking	Ratio
Celsius Temperature	Interval
Weight	Ratio
Hair Color	Nominal
Socioeconomic Status	Ordinal
Fahrenheit Temperature	Interval
Height	Ratio
Type of living accommodation	Normal
Level of Agreement	Ratio
IQ(Intelligence Scale)	Ratio
Sales Figures	Ratio
Blood Group	Ordinal
Time Of Day	Interval
Time on a Clock with Hands	Interval
Number of Children	Ratio
Religious Preference	Ordinal

Barometer Pressure	Interval
SAT Scores	Ratio
Years of Education	Ordinal

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

Ans: $\frac{3}{8}$

Q4) Two Dice are rolled, find the probability that sum is

- a) Equal to 1
- b) Less than or equal to 4
- c) Sum is divisible by 2 and 3

Ans: a) 0

b) $\frac{1}{6}$

c) $\frac{1}{6}$

Q5) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

Ans: $\frac{10}{21}$

[5 non balls, 1st draw $\frac{5}{7}$ and 2nd draw $\frac{4}{6}$ so by multiplying we get $\frac{10}{21}$, since it is continuous]

Q6) Calculate the Expected number of candies for a randomly selected child

Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

CHILD	Candies count	Probability
A	1	0.015
B	4	0.20
C	3	0.65
D	5	0.005
E	6	0.01
F	2	0.120

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

Ans: Expected number of candies for a randomly selected child

$$=1*0.015+4*0.20+3*0.65+5*0.005+6*0.01+2*0.12$$

$$=3.09$$

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

- For Points,Score,Weigh>
Find Mean, Median, Mode, Variance, Standard Deviation, and Range
and also Comment about the values/ Draw some inferences.

Use Q7.csv file

Ans:(this question answer I have executed in anaconda navigator)

```
# mean
cars.mean()
```

```
Points      3.596563
Score       3.217250
Weigh       17.848750
```

```
# Median
cars.median()
```

```
Points      3.695
Score       3.325
Weigh       17.710
```

```
# Mode
cars.Points.mode()
```

```

0    3.07
1    3.92
cars.Score.mode()
0    3.44
cars.Weigh.mode()
0    17.02
1    18.90
# Variance
cars.var()

Points    0.285881
Score     0.957379
Weigh     3.193166
# Standard Deviation
cars.std()

Points    0.534679
Score     0.978457
Weigh     1.786943
# Range
cars.describe()

```

```
In [ ]: # Range
```

```
In [12]: cars.describe()
```

```
Out[12]:
```

	Points	Score	Weigh
count	32.000000	32.000000	32.000000
mean	3.596563	3.217250	17.848750
std	0.534679	0.978457	1.786943
min	2.760000	1.513000	14.500000
25%	3.080000	2.581250	16.892500
50%	3.695000	3.325000	17.710000
75%	3.920000	3.610000	18.900000
max	4.930000	5.424000	22.900000

```
Points_Range=cars.Points.max()-cars.Points.min()
```

```
Points_Range
```

```
2.17
```

```
Score_Range=cars.Score.max()-cars.Score.min()
```

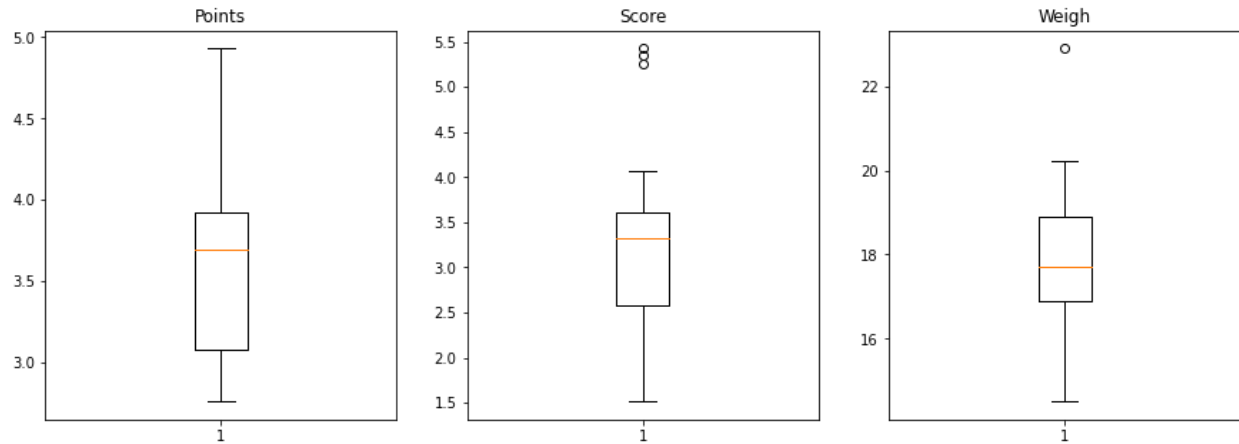
```
Score_Range
```

```
3.9109999999999996
```

```
Weigh_Range=cars.Weigh.max()-cars.Weigh.min()
```

```
Weigh_Range
```

```
8.399999999999999
```



Q8) Calculate Expected Value for the problem below

a) The weights (X) of patients at a clinic (in pounds), are
108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

Ans: Expected Value = $\sum (\text{probability} * \text{Value})$

$$\sum P(x).E(x)$$

there are 9 patients

Probability of selecting each patient = $1/9$

Ex 108, 110, 123, 134, 135, 145, 167, 187, 199

P(x) $1/9$ $1/9$ $1/9$ $1/9$ $1/9$ $1/9$ $1/9$ $1/9$ $1/9$

Expected Value = $(1/9)(108) + (1/9)110 + (1/9)123 + (1/9)134 + (1/9)135 + (1/9)145 + (1/9)(167) + (1/9)187 + (1/9)199$

= $(1/9) (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)$

= $(1/9) (1308)$

= 145.33

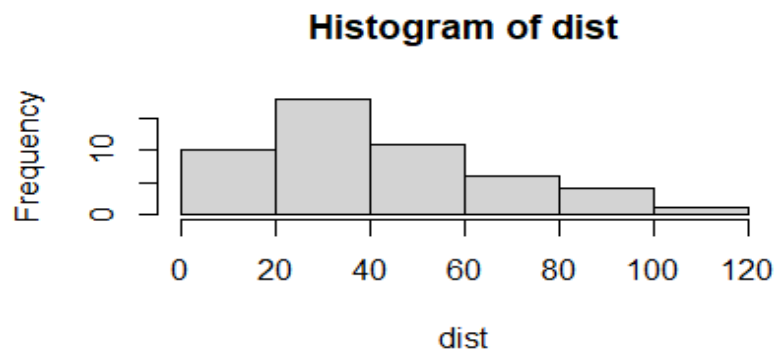
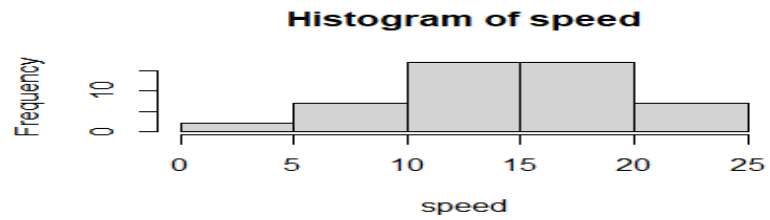
Expected Value of the Weight of that patient = 145.33

Q9) Calculate Skewness, Kurtosis & draw inferences on the following data

Cars speed and distance

Use Q9_a.csv

Ans:



skewness	-0.11751	0.806895
kurtosis	-0.50899	0.405053

SP and Weight(WT)

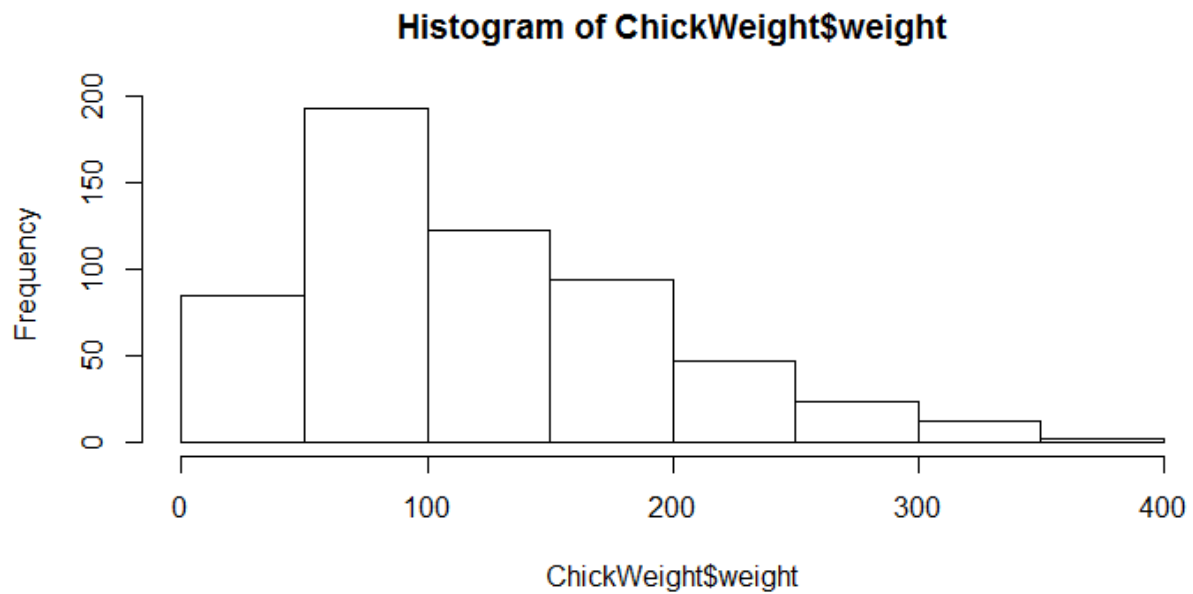
Use Q9_b.csv

Ans:

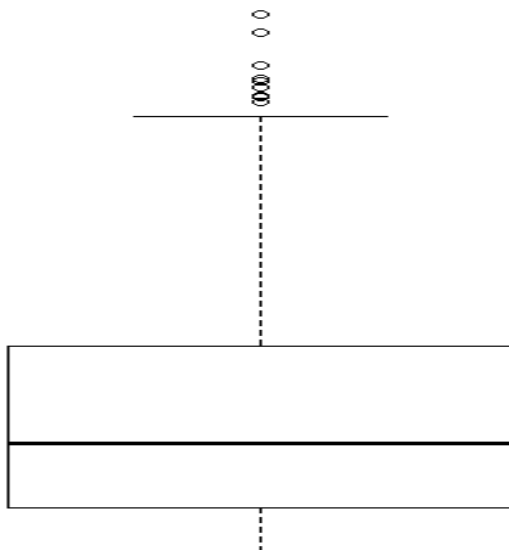
SP	WT
104.1854	28.76206
105.4613	30.46683
105.4613	30.1936
113.4613	30.63211
104.4613	29.88915
113.1854	29.59177
105.4613	30.30848
102.5985	15.84776
102.5985	16.35948
115.6452	30.92015
111.1854	29.36334
117.5985	15.75353
122.1051	32.81359
111.1854	29.37844
108.1854	29.34728
111.1854	29.60453
114.3693	29.53578
117.5985	16.19412
114.3693	29.92939
118.4729	33.51697
119.1051	32.32465
110.8408	34.90821
120.289	32.67583
113.8291	31.83712
119.1854	28.78173
114.5985	16.04317
120.7605	38.06282
119.1051	32.83507
99.56491	34.48321
121.8408	35.54936
113.4846	37.04235
112.289	33.23436
119.9211	31.38004
121.3926	37.57329

skewness	1.61145	-0.61475
kurtosis	2.977329	0.950291

Q10) Draw inferences about the following boxplot & histogram



Ans: Right side skewed or positively skewed.



Ans: the interface for this box plot is positively skewed.

Q11) Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

Ans: Sample mean = 200

Sample SD = 30

n = 2000

➤ **Avg. weight of Adult in Mexico with 94% CI**
stats.norm.interval(0.94,200,30/(20000.5)**

(198.738325292158, 201.261674707842)

➤ *Avg. weight of Adult in Mexico with 98% CI*
*stats.norm.interval(0.98,200,30/(2000**0.5))*

(198.43943840429978, 201.56056159570022)

➤ *Avg. weight of Adult in Mexico with 96% CI*
*stats.norm.interval(0.96,200,30/(2000**0.5))*

(198.62230334813333, 201.37769665186667)

Q12) Below are the scores obtained by a student in tests

34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56

- 1) Find mean, median, variance, standard deviation.
- 2) What can we say about the student marks?

Ans: 1) mean=sum of the terms/number of terms

=738/18

=41

▪ **Median = 1**

- Variance = 22.705
- Standard deviation = 4.764

2) There are 2 Outliers in Student's marks: 49 and 56

Q13) What is the nature of skewness when mean, median of data are equal?

Ans: Normalized skewness.

Q14) What is the nature of skewness when mean > median ?

Ans: Right skewed.

Q15) What is the nature of skewness when median > mean?

Ans: Left skewed.

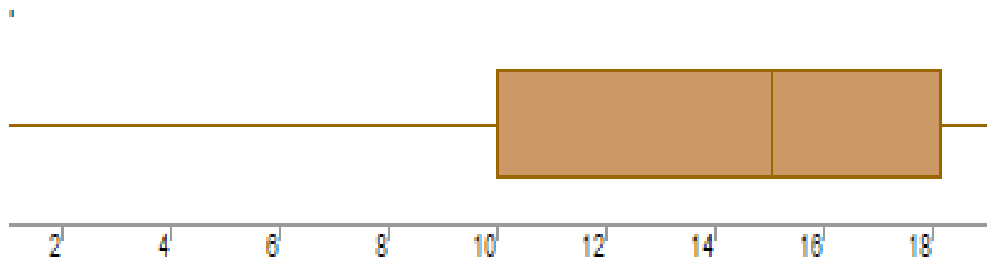
Q16) What does positive kurtosis value indicates for a data ?

Ans: Sharp peak in the plot. less gap between tails to x-axis

Q17) What does negative kurtosis value indicates for a data?

Ans: Border peak under the curve and more gap between tails and x-axis.

Q18) Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

Ans: The data is distributed in De-assigned format

What is nature of skewness of the data?

Ans: Left side skewed

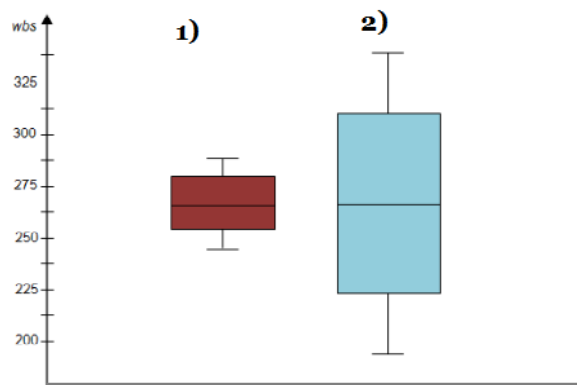
What will be the IQR of the data (approximately)?

Ans: $Q3 - Q1$

$$= 18 - 10$$

$= 8$ is IQR

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

Ans: The box plot 1 designed with range = 3 , The second one range is = 1.5

Q 20) Calculate probability from the given dataset for the below cases

Data _set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

`MPG <- Cars$MPG`

a. $P(MPG > 38)$

Ans: $1 - \text{pnorm}(38, 34.422, 9.13144) = 0.3475908$

b. $P(MPG < 40)$

Ans: $\text{pnorm}(40, 34.422, 9.13144) = 0.7293527$

c. $P(20 < MPG < 50)$

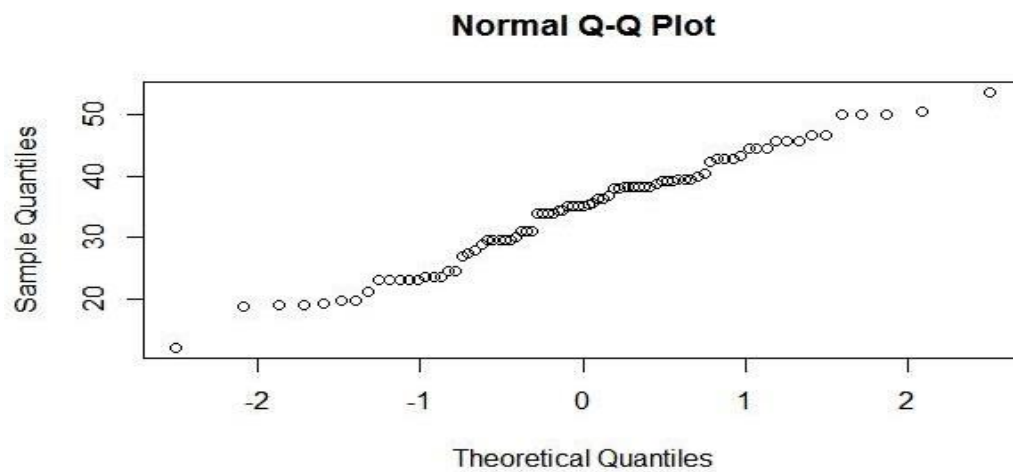
Ans: $\text{pnorm}(50, 34.422, 9.13144) - (1 - \text{pnorm}(20, 34.422, 9.13144)) =$
0.01311818

Q 21) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:



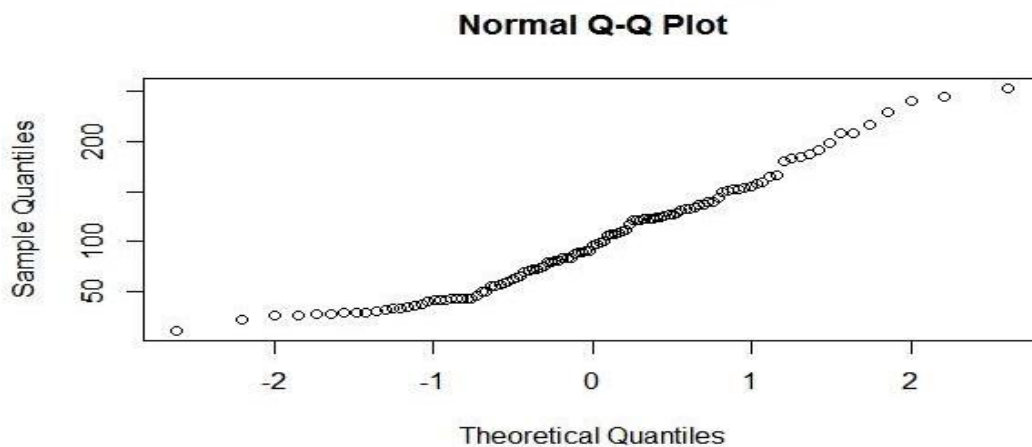
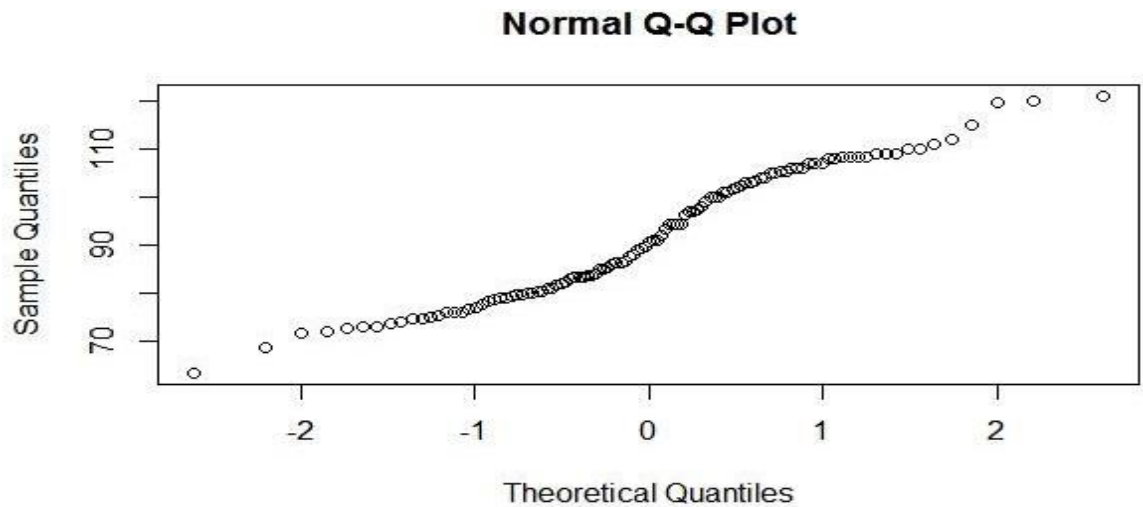
-distributed normally

b) Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Ans:

Adipose tissue (AT) normal distribution



- Waist Circumference(Waist) normal distributed

Q 22) Calculate the Z scores of 90% confidence interval, 94% confidence interval, 60% confidence interval

Ans: 90%= 1.645

94%= 1.880

60% = 0.253

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

Ans: To compute the 95% confidence interval, start by computing the mean and standard error: $M = (2 + 3 + 5 + 6 + 9)/5 = 5$. $\sigma_M = 1.118$. $Z_{.95}$ can be found using the normal distribution calculator and specifying that the shaded area is 0.95 and indicating that you want the area to be between the cutoff points

Confidence Level z

0.90 1.645

0.92 1.75

0.95 1.96

0.96 2.05

With a 90 percent confidence interval, you have a 10 percent chance of being wrong. A 99 percent confidence interval would be wider than a 95 percent confidence interval (for example, plus or minus 4.5 percent instead of 3.5 percent).

Q 24) A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode \rightarrow pt(tscore,df)

df \rightarrow degrees of freedom

Ans: t - statistics for the data is given as follows:

$$t = \frac{x - \mu}{\frac{s}{\sqrt{n}}}$$

x = mean of the sample of bulbs = 260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

$$t = \frac{260 - 270}{\frac{90}{\sqrt{18}}}$$

$$t = \frac{-10}{\frac{90}{3\sqrt{2}}}$$

$$t = \frac{-10}{\frac{30}{\sqrt{2}}}$$

$$t = \frac{-1 \times \sqrt{2}}{3}$$

$$t = -0.471$$