

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	Discrete or Categorical
Number of kids	Discrete
Number of tickets in Indian railways	Discrete
Number of times married	Discrete
Gender (Male or Female)	Discrete or Categorical

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

Nominal, Ordinal, Interval, Ratio.

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

Barometer Pressure	Ratio
SAT Scores	Interval
Years of Education	Ordinal

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

Ans: - $2^3 = \text{Total 8 Possibilities.}$

Therefore, probability of 2 Heads and One Tails is $\frac{6}{8}$ or $\frac{3}{4}$.

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) Since two dice are rolled simultaneously there is not possibility for sum equals to 1. Therefore, **zero probability. 0**

b) Possibilities for less than or equal to 4 are (1,1), (1,2), (2,1), (3,1), (1,3), (2,2).

Therefore, **possibility is** $\frac{6}{36}$ or $\frac{1}{6}$.

c) Possibilities of getting sum is divisible of 2 and 3 are (1,5), (5,1), (2,4), (4,2),

(3,3), (6,6). Therefore, **possibility is** $\frac{6}{36}$ or $\frac{1}{6}$.

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: - 2 Red, 3 Green and 2 Blue.

$${}^5C_2 / {}^7C_2 = \frac{10}{21}$$

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

Ans: -

	mean	median	mode	variance	std	range
Points	3.5965	3.695	3.07	0.2858	0.5346	2.170
Score	3.2172	3.325	3.44	0.9573	0.9784	3.910
Weigh	17.8487	17.710	17.02	3.1930	1.7869	8.399

By analyzing the Car data set we could say that. Average overall score of the cars are 3.2172 and most repeated score is 3.44.

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?

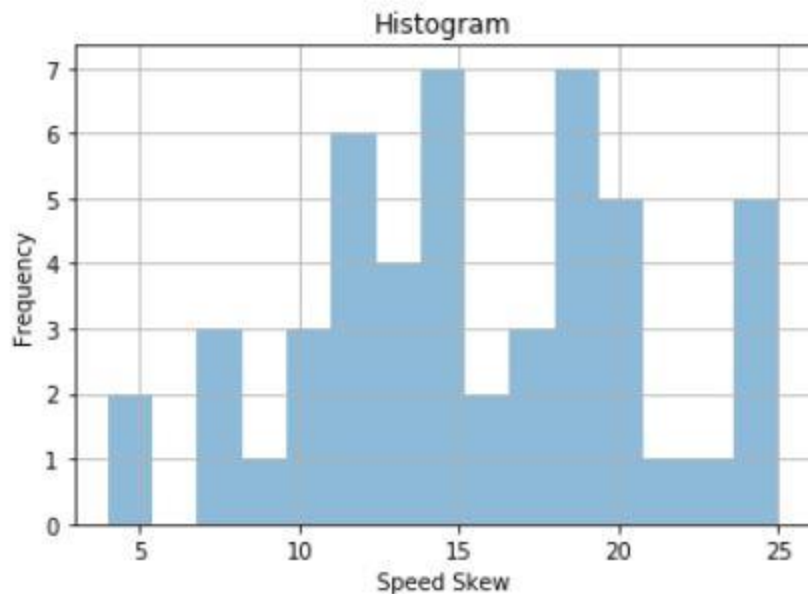
Ans: - **Estimated Value is 145.333**

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

Cars speed and distance

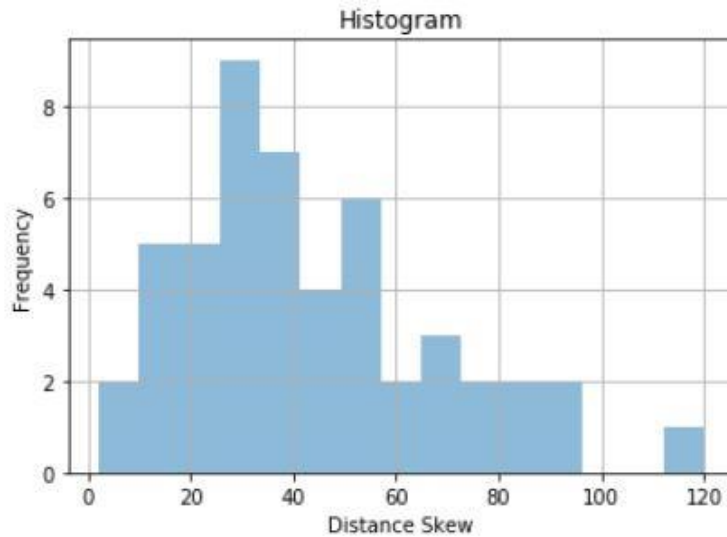
Car Speed Skewness

Ans: -



- Speed skewness Value -0.11750

Since the above speed histogram is left tailed. We could say that it's a negative skew. The average speed for all the cars is 15.4.



- Speed skewness Value - 0.8068

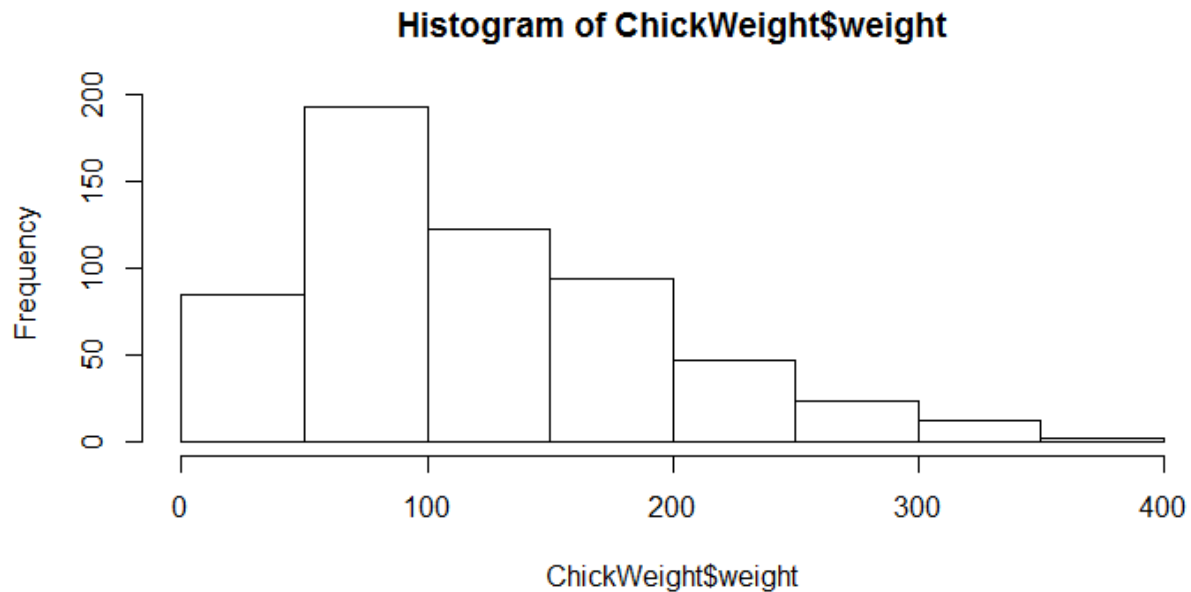
Since the above speed histogram is right tailed. We could say that it's a "positive skew". The average distance covered by all the cars are 42.98.

- Speed Kurtosis Value is: -0.508994
- Distance Kurtosis Value is: 0.405053

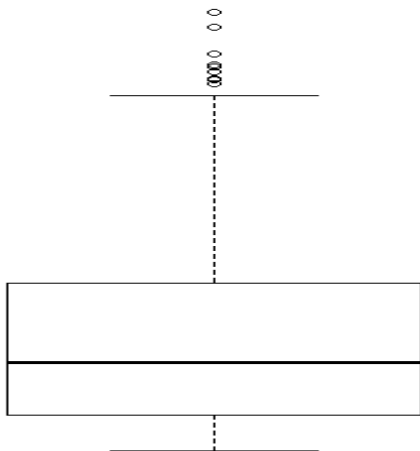
SP and Weight (WT)

- SP Skewness value is 1.611450.
It is Positive Skewed.
- WT Skewness value is -0.614753.
It is Negative Skewed.
- SP Kurtosis value is 2.977329.
- WT Kurtosis value is -0.614753.

Q10) Draw inferences about the following boxplot & histogram



Ans: - From the above histogram, we could say that it's a "right tailed" or "Positive skewed histogram". Majority of data are concentrated around 100 – 200's. And there is only few chick's whose weight are above 300.



Ans:- From the above box plot we could say that it's a right skewed or positive skewed data. And there are several outliers in the data set

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: -

Confidence Interval of 94% is **[198.740 – 201.26]**

Confidence Interval of 98% is **[198.44 – 201.56]**

Confidence Interval of 96% is **[198.62 – 201.38]**

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.

Mean = 41.3125

Median = 41

Variance = 25.529

St. Deviation = 5.05

2) What can we say about the student marks?

There are 2 outliers in the data set which is 49 and 56.

And Mean > Median which means distribution is right skewed or positive skewed.

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

Ans: -There will be no skewness. Data will be Normally Distributed.

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

Ans: - Right Skewed or Positive Skew

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

Ans: - Left Skewed or Negative Skew

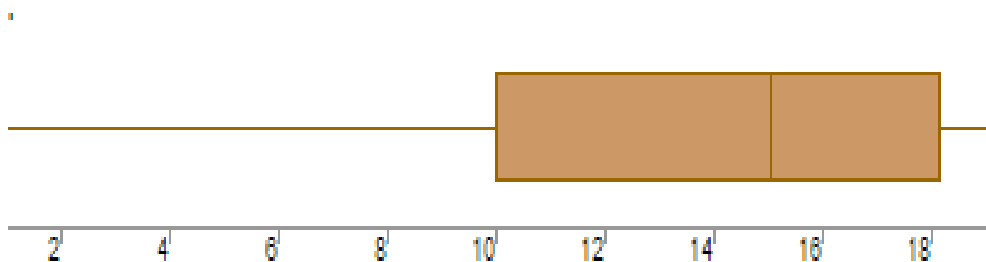
Q16) What does positive kurtosis value indicates for a data?

Ans: - Positive Kurtosis indicated more Peak in the distribution.

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

Ans: - Negative Kurtosis indicated Low Peak or no Peak at all.

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



What can we say about the distribution of the data?

Ans: - The distribution doesn't look symmetric. Therefore, it's not a normal Distribution.

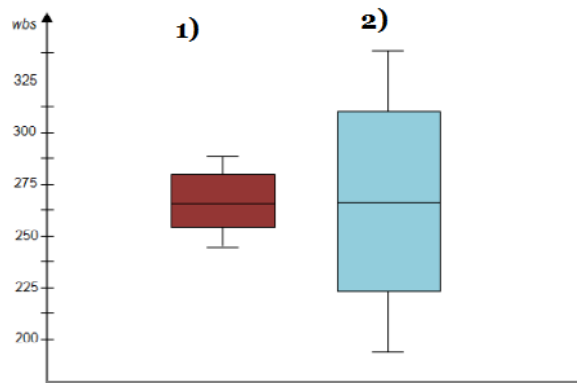
What is nature of skewness of the data?

Ans: - Left Skewed or Negative skew.

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

Ans: - $IQR = Q1 - Q3$. $18 - 10 = 8$

Q19) Comment on the below Boxplot visualizations?



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

Ans: - We could saw that both are Normal Distribution. The Median of both box-plot are same which is approximately around 265.

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. $P(\text{MPG} > 38)$
- b. $P(\text{MPG} < 40)$
- c. $P(20 < \text{MPG} < 50)$

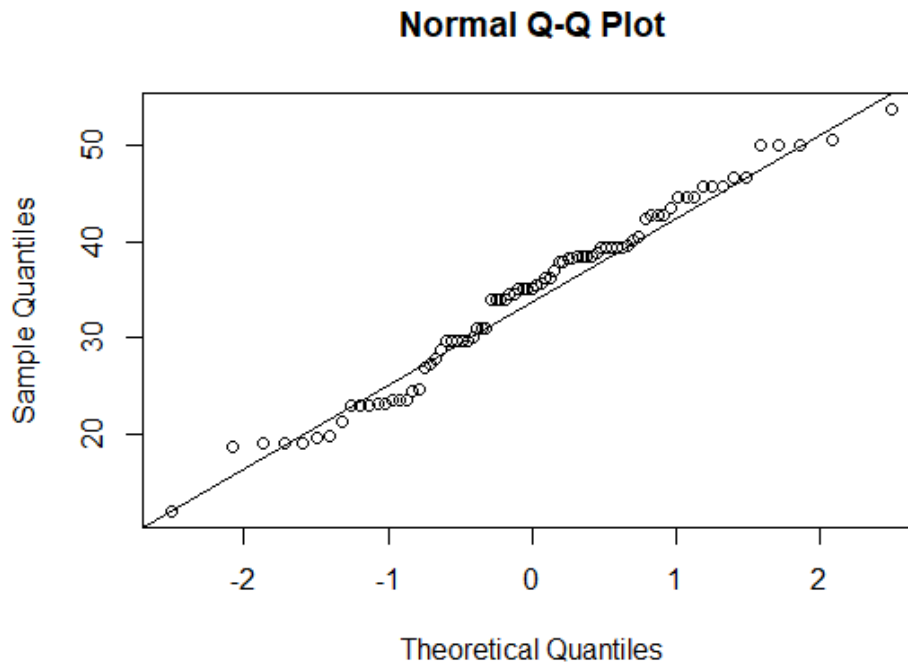
Ans: -

- a) $P(\text{MPG} > 38)$ $1 - 0.65 = 0.35$
- b) $P(\text{MPG} < 40)$ 0.7294571
- c) $P(20 < \text{MPG} < 50)$ $0.956039 - 0.05712119 = 0.8989178$

Q 21) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



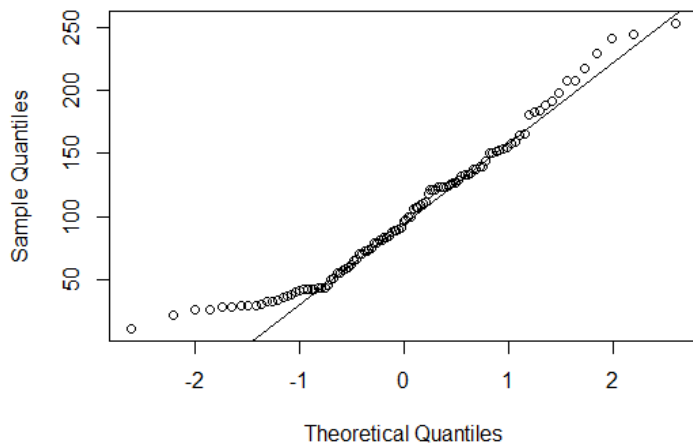
Ans: - Yes, the MPG of Cars follow Normal Distribution.

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

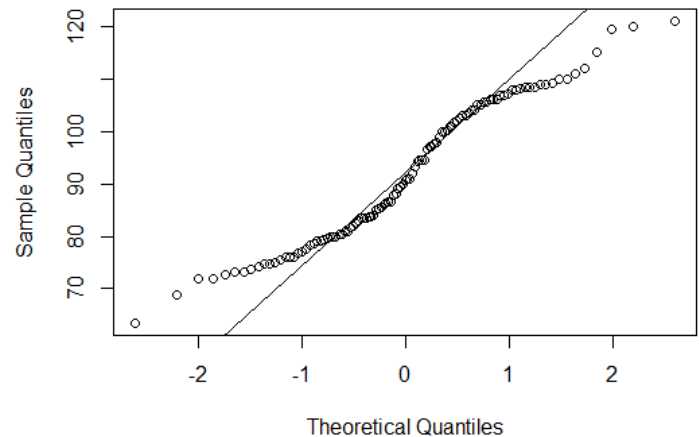
Dataset: wc-at.csv

Ans:- Both Adipose Tissue and Waist follows normal distribution.

Normal Q-Q Plot



Normal Q-Q Plot



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

Ans: - 90% -> **1.644854**
94% -> **1.880794**
60% -> **0.8416212**

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

Ans: - 95% -> **2.063899**
96% -> **2.171545**
99% -> **2.79694**

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 → `pt(tscore,df)`

df → degrees of freedom

Ans:- $\mu=270$, $\bar{x}=260$, $SD=90$, $n=18$, $df=18-1=17$

$$T\text{-Score} = \frac{260-270}{90\sqrt{18}} = \frac{-10}{21.23} = -0.47$$

Probability is 32%