Q1) Identify the Data type for the Following:

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

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

Nominal, Ordinal, Interval, Ratio.

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

Barometer Pressure	Ratio scale data
SAT Scores	Interval scale data
Years of Education	Ratio scale data

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

A3) 3 coins are tossed(HHH,HHT,HTT,HTH,TTT,TTH,THH,THT)= 8 Possibilities Probability for 2 heads & one tail= (HHT,HTH,THH)=3÷8 = 0.375

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

A4) a) Equal to 1= 0 % Probability

  b) Less than or equal to 4= 6÷36= 1÷6= 0.167

  c) Sum is divisible by 2 and 3= 8÷36= 0.222

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?

A5) (Total balls= 2red+3green+2blue=7balls) drawn 2 balls by blue color, therefore excluding 2 balls having already blue in color then remained 5 balls(2red+3green) so probability that none of the balls drawn is blue= 5C2÷7C2

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

A6) Expected no. of candies= 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.

A7) **Using Q7.csv file (Tool - *R studio*)**

S.no	Mean	Median	Mode	Variance	S.D	Range
Points	3.596563	3.695	3.07 & 3.92	0.2858814	0.5346787	2.76 & 4.93
Score	3.21725	3.325	3.44	0.957379	0.9784574	1.513 & 5.424
Weigh	17.84875	17.71	17.02 & 18.90	3.193166	1.786943	14.5 & 22.9

Inferences: The mode of Points and Weigh Columns Are Bi-modal Series.

From the Range column in the above table, we can observe the extreme values in the given data set.

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?

A8) Expected value of the weight of that patient= 145.333

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

Car's speed and distance

A9) Using Q9_a.csv (Tool – *R Studio*)

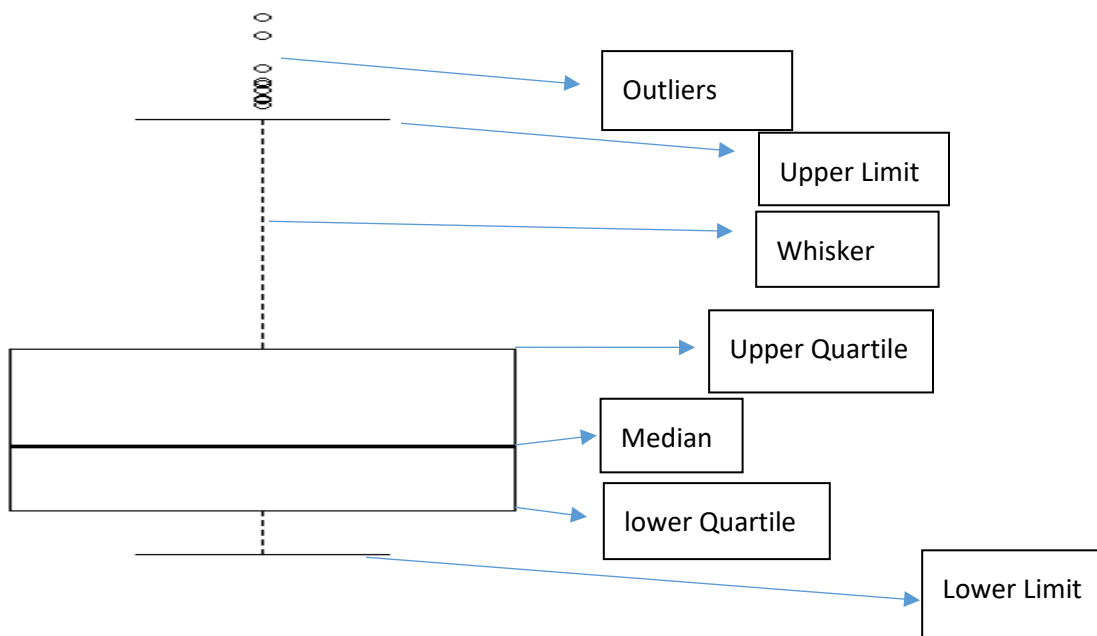| S.No	Skewness	Kurtosis
Speed	-0.1105533	2.422853
Distance	0.7591268	3.248019

Inferences:

1. Skewness of Speed is negative in nature, which lets us know that the data is left skewed and also the median > mean.
2. Skewness of Distance is positive in nature, which lets us know that the data is right skewed and also the mean > median.
3. The Kurtosis values of both the data of Speed and Distance are positive in nature, which means the curve from the given data is sharp in nature and the values in the extremes have thin tails.

SP and Weight (WT)

Using Q9_b.csv

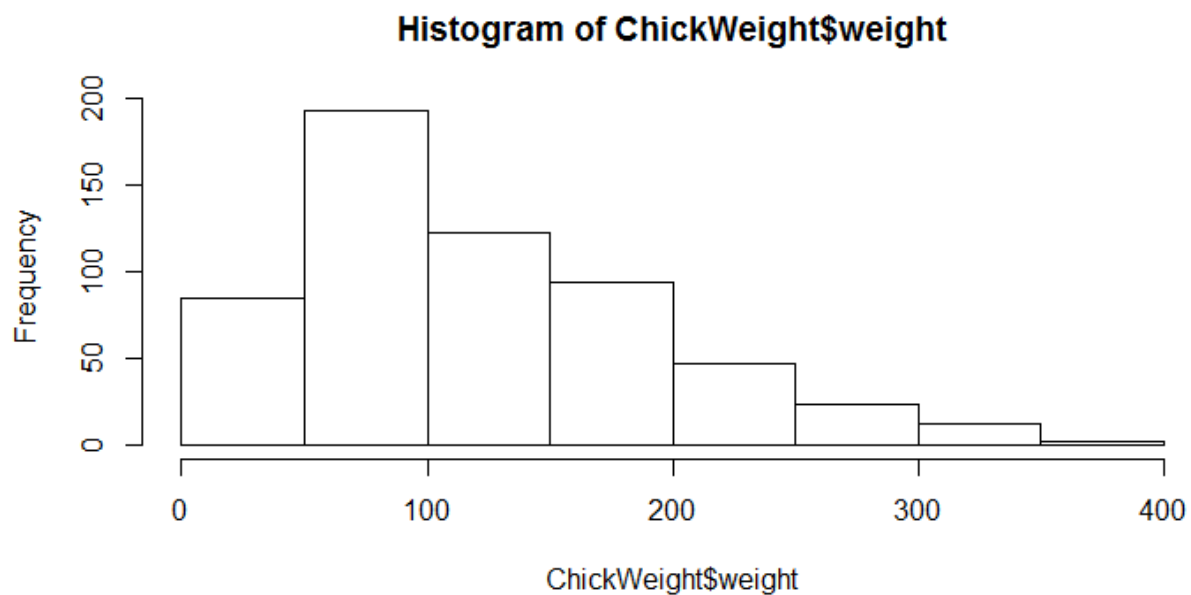| S.No	Skewness	Kurtosis
SP	1.552258	5.723521
WT	-0.5921721	3.819466

Inferences:

1. Skewness of WT is negative in nature, which lets us know that the data is left skewed and also the median > mean.
2. Skewness of SP is positive in nature, which lets us know that the data is right skewed and also the mean > median.
3. The Kurtosis values of both the data of SP and WT are positive in nature, which means the curve from the given data is sharp in nature and the values in the extremes have thin tails.

Q10) Draw inferences about the following boxplot & histogram

A10) The skewness for the below histogram is right skewed in nature which depicts that the mean > median.

The peak of the below histogram has a sharp therefore the kurtosis is positive.

Histogram of ChickWeight$weight

Outliers

Upper Limit

Whisker

Upper Quartile

Median

lower Quartile

Lower Limit

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?

A11) Class interval for 94% = 143.57619175546247, 256.42380824453755

Class interval for 98% = 130.2095637787748, 269.7904362212252

Class interval for 96% = 138.38753268104531, 261.61246731895466

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.

Sol: Mean = sum of all obs./ total obs.= 738/18=41

Median = 40+41/2 = 81/2 = 40.5

S.no	Deviation (obs. – mean)	Deviation^2
1	-6.5	42.25
2	-4.5	20.25
3	-4.5	20.25
4	-2.5	6.25
5	-2.5	6.25
6	-1.5	2.25
7	-1.5	2.25
8	-0.5	.25
9	-0.5	.25
10	0.5	.25
11	0.5	.25
12	0.5	.25
13	0.5	.25
14	2.5	6.25
15	2.5	6.25

16	5.5	30.25
17	9.5	90.25
18	16.5	272.25
19	total	506.5

Variance= 506.5/18 = 28.13

SD = √28.13 = 5.303

2) What can we say about the student marks?

Ans) Since mean is greater than median the skewness of the data is positive.

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

A13) Normal distribution curve is Symmetrical in nature.

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

A14) The normal distribution curve is right skewed.

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

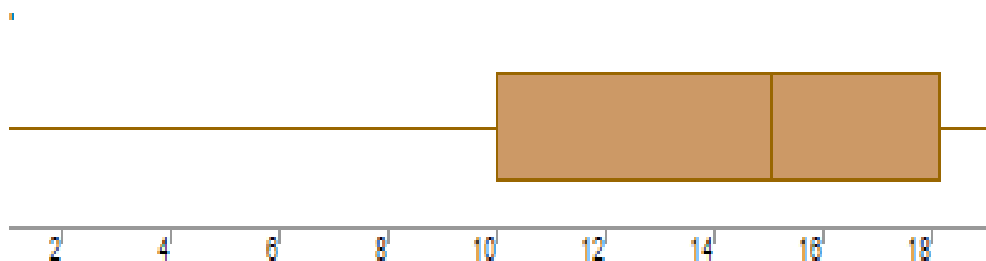A15) The normal distribution curve is left skewed.

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

A16) The peak of the normal distribution curve is sharp.

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

A17) The peak of the normal distribution curve is flat.

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

What can we say about the distribution of the data?

Ans)1. The IQR lies mostly between 10 & 18.

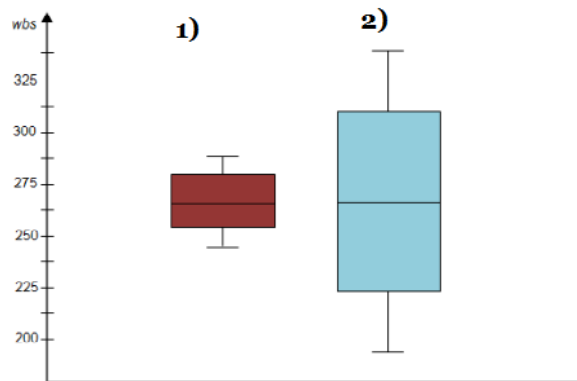2. Median is close to Upper quartile.

What is nature of skewness of the data?

Ans) Median is close to Upper Quartile therefore the data is negatively skewed.

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

Ans) Q3-Q1 = 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.

A19) The median of both the boxplots are on the 50% margin of the IQR therefore there is no skewness and the data is normally ditributed.

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)

b. P(MPG<40)

c. P(20<MPG<50)

A20) a) P(MPG>38) = 0.347

b) P(MPG<40) = 0.729

c) P(20<MPG<50) = 0.898

Q 21) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans) No, the data of MPG is not normally distributed because the mean, median and mode are not equal and the skewness is also not equal to 0.

Mean =34.42, Median=35.15, Mode=29.629 & skewness = -0.17.

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

Dataset: wc-at.csv

Ans) No, either of the data in the respective columns are normally distributed because the mean, meduan and mode are not equal and the skewnwss of the data is also not equal to 0.

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

A 22) Confidence Level Z-Score 90% = 1.644

Confidence Level Z-Score 94% = 1.750

Confidence Level Z-Score 60% = 0.841

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

A 23) Confidence Level T-Score 95% = 2.063

Confidence Level T-Score 96% = 2.171

Confidence Level T-Score 99% = 2.796

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

A 24) T-Score = 260-270/(90/18^0.5) = -0.47

P(X<260) = Stats.t.cdf(-0.47,17) = 0.32