Assignment - 01

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

Q2) Identify the Data types, which were among the following
Nominal, Ordinal, Interval, Ratio.

Data	Data Type
Gender	Nominal
High School Class Ranking	Interval
Celsius Temperature	Interval
Weight	Interval
Hair Color	Nominal
Socioeconomic Status	Ordinal
Fahrenheit Temperature	Interval

Height	Ratio
Type of living accommodation	Ordinal
Level of Agreement	Ordinal
IQ(Intelligence Scale)	Ratio
Sales Figures	Ratio
Blood Group	Nominal
Time Of Day	Interval
Time on a Clock with Hands	Interval
Number of Children	Nominal
Religious Preference	Nominal
Barometer Pressure	Interval
SAT Scores	Interval
Years of Education	Ratio

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

Ans: Total number of events= {hhh, hht, htt, ttt, tth, thh, hth, tht} =8

Interested events=3

Probability=3/8.

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

a) Equal to 1

Ans: Total number of outcomes when two dice are rolled=6*6=36.

(1, 1)(1, 2)(1, 3)(1, 4)(1, 5)(1, 6)
(2, 1)(2, 2)(2, 3)(2, 4)(2, 5)(2, 6)
(3, 1)(3, 2)(3, 3)(3, 4)(3, 5)(3, 6)
(4, 1)(4, 2)(4, 3)(4, 4)(4, 5)(4, 6)
(5, 1)(5, 2)(5, 3)(5, 4)(5, 5)(5, 6)

(6, 1)(6, 2)(6, 3)(6, 4)(6, 5)(6, 6)

Equal to 1 = **0% probability**

b) Less than or equal to 4

Ans: Less than or equal to 4= 6/36 = **1/6**

c) Sum is divisible by 2 and 3

Ans: sum is divisible by 2 and 3

{	2	3	4	5	6	7
	3	4	5	6	7	8
	4	5	6	7	8	9
	5	6	7	8	9	10
	6	7	8	9	10	11
	7	8	9	10	11	12 }

Probability=6/36= 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: total number of events= ===21

Interested events= ==10

Probability that none of the balls is blue =10/21=0.47

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-Generalised 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 = E(x)

= 1*0.015+4*0.20+3*0.65+5*0.005+6*0.01+2*0.120

= **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: Done in R file.

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: Done in R file.

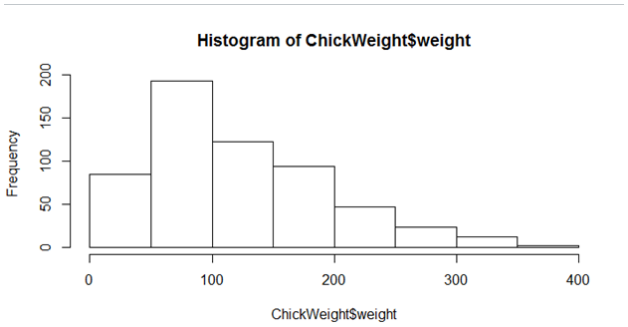
Q9) Calculate Skewness, Kurtosis & draw inferences on the following data

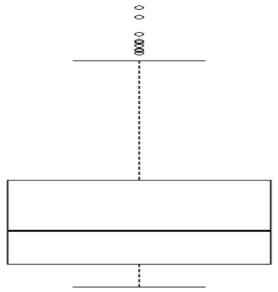
Cars speed and distance Use Q9_a.csv

SP and Weight(WT) Use Q9_b.csv

Ans: Done in R file.

Q10) Draw inferences about the following boxplot & histogram





Ans:

Histogram	Box Plot
➔ This data is positively skewed. ➔ Many chickens are having weights in between 50-100 gm (unit) ➔ Very rare no of chickens are having weight more than 300gm ➔ After a certain point i.e. (100gm) the frequency of the chicken decreases with increase in their weight.	➔ This boxplot represents the data is positively skewed ➔ The data contains outliers.

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: Done in R file.

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) Done in R file.
2) Mean > Median, This implies that the distribution is slightly skewed towards the right. Marks are not normally distributed. Person with mark 56 can be the outlier in our data. Many of the students are having 41 marks i.e. the modal value of our data.

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

Ans: Skewness doesn't exist in those cases i.e. skewness = 0. The curve is called a perfectly symmetric bell shaped curve.

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

Ans: Right skewed(tail on the right side).

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

Ans: Left Skewed(tail on the left side).

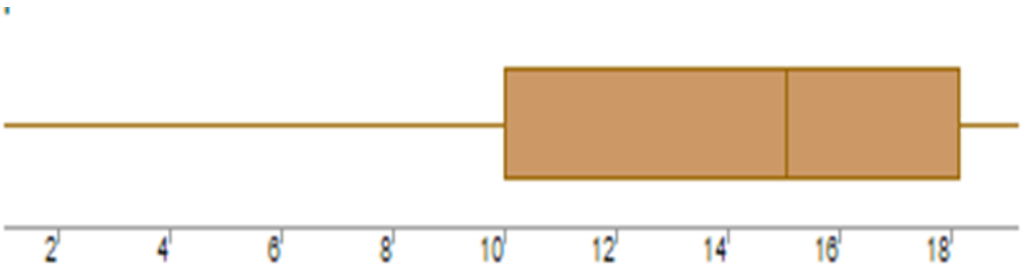
Q16) What does positive kurtosis value indicate for a data ?

Ans: sharp peak and less variation.

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

Ans: Broad peak and more variation.

Q18) Answer the below questions using the below box plot visualisation.



What can we say about the distribution of the data?

Ans: The data is not symmetric. Mass of the data is concentrated towards the right side. It is not a Normal Distribution

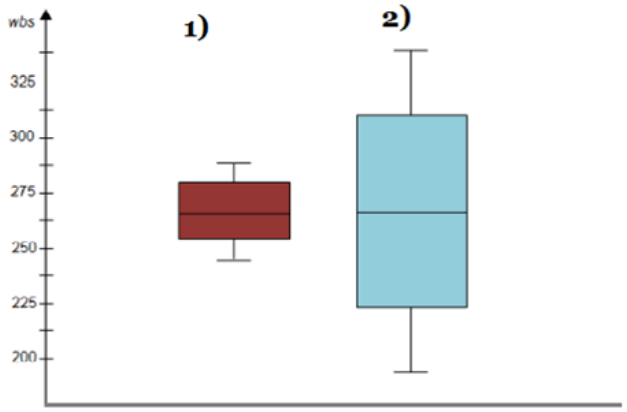
What is the nature of skewness of the data?

Ans: The Data is negatively skewed. It is left skewed.

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

Ans: InterQuartile Range =Upper Quartile- Lower Quartile => 18-10=8

Q19) Comment on the below Box Plot visualisations?



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

Ans: 1) The median of the two boxplots are approximately 260.

2) The boxplots are not skewed in +ve or –ve direction.

3) Outliers don't exist in both of the boxplots.

4) Comparing Boxplot_1 (Red) has less variability, less variation, less standard deviation, less range, less Inter-quartile-range value as compared to Boxplot_2(blue)

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)

Ans: Done in R file.

Q 21) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

Ans: Done in R file.

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

Ans: Done in R file.

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

Ans: Done in R file.

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 à pt(tscore,df) df à degrees of freedom

Ans: Done in R file.