**Basic Statistics Level - 1**

Q1) Identify the Data type for the following :

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

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

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

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

1. Equal to 1

1. Less than or equal to 4:

1. Sum is divisible by 2 and 3:

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?

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)

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset.For Points,Score,Weight find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ draw some inferences.

Use Q7.csv file.

Q8) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

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

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

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

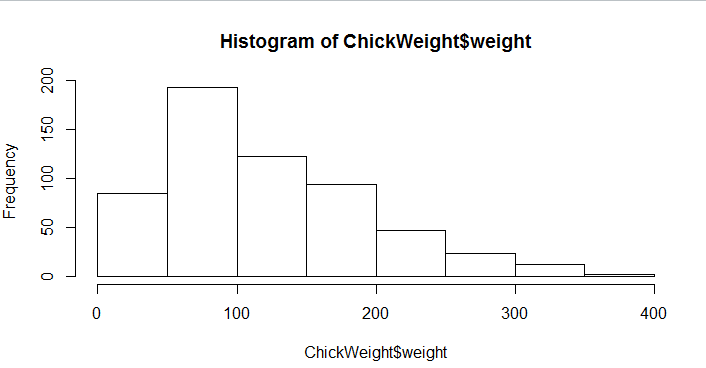
1] Cars speed and distance

Use Q9\_a.csv

2] SP and Weight(WT)

Use Q9\_b.csv

Q10) Draw inferences about the following boxplot & histogram.





1]

Inference :

| ChickWeight$Weight | Frequency |
| --- | --- |
| 0-50 | 90 |
| 50-100 | 190 |
| 100-150 | 120 |
| 150-200 | 95 |
| 200-250 | 45 |
| 250-300 | 25 |
| 300-350 | 15 |
| 350-400 | 5 |

The histogram is skewed to the right.

2]

In the given boxplot the maximum value is much higher than the other quartiles and minimum value. The outliers are more than the maximum value of the box plot.

The 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 ?

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?

From the given data we can infer that the student scores are skewed to the left or positively skewed.

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

The nature of skewness is zero when mean median mode of data are equal.

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

The nature of skewness is positively skewed when mean > median.

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

The nature of skewness is negatively skewed when mean > median.

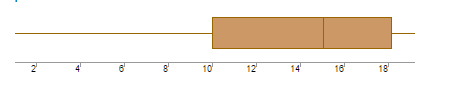
Q16) What does positive kurtosis value indicate for a data ?

Positive values of kurtosis indicate that distribution is peaked and possesses thick tails. An extreme positive kurtosis indicates a distribution where more of the numbers are located in the tails of the distribution instead of around the mean.Leptokurtic distributions have positive kurtosis values.

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

Negative excess values of kurtosis (<3) indicate that a distribution is flat and has thin tails. Platykurtic distributions have negative kurtosis values.

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

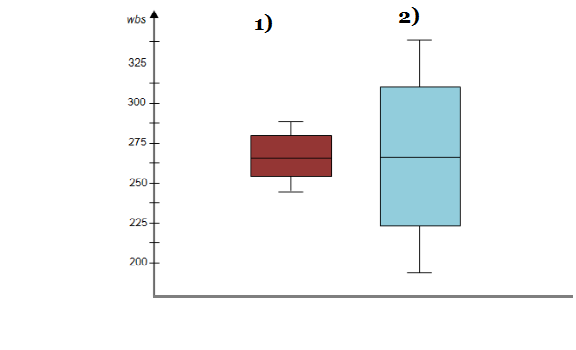


1. What can we say about the distribution of the data?
2. What is the nature of skewness of the data?
3. What will be the IQR of the data (approximately)?
4. Minimum = 0 , Maximum = 20 , Q1 = 10 , Q3 = 18 , Median = 15.6

The data is unevenly distributed.

1. The data is negatively skewed as the minimum of the data is much lesser than the maximum and the other quartiles.
2. IQR = Q3-Q1 = 18-10 = 8

Q19) Comment on the below Box Plot visualizations?



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

Inference :

1. Minimum = 243.75

Q1 = 256.25

Median = 262.5

Q3 = 281.25

Maximum = 287.5

1. Minimum = 193.75

Q1 = 225

Median = 262.5

Q3 = 312.5

Maximum = 337.5

Both the boxplot visualizations have the same median of 262.5 and are not skewed or having outliers.However Boxplot 1 has lesser IQR and Whiskers than Boxplot 2 indicating lesser range of the data.

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

* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval,60% confidence interval for sample size of 25.

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

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.