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 | Ordinal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| 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 | Interval |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Interval |

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

Sol: If three coins are tossed,

Total number of possible combinations = 23 = 8

The combinations are HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.

Number of combinations that have two heads and one tail = 3, i.e.,

HHT, HTH, TTH

The probability of two heads and one tail when three coins are tossed simultaneously are

P (Two heads and One tail) = Number of desired outcomes

= ⅜ or 0.375

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

Equal to 1

Less than or equal to 4

Sum is divisible by 2 and 3

Sol: a). If two dices were rolled, then total possible cases =36 Total Favourable cases (Having sum =1) = 0 As minimum sum is 2 for outcome (1,1).

Hence, probability is 0.

b). When we roll two dice, the possibility of getting number less than or equal to 4 is (1,1), (1,2), (1, 3), (2,1), (2, 2), (3, 1).

So, the number of favorable outcomes = 6

Total number of possibilities = 36

Probability = The number of favorable outcomes / Total number of possibilities = 6/36 = 1/6.

c). When we roll two dice, the possibility of getting sum divisible by 2 and 3 is (1,5), (2,4), (3,3), (4,2), (5,1), (6,6).

So, the number of favorable outcomes = 6

Total number of possibilities = 36

Probability = The number of favorable outcomes / Total number of possibilities = 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?

Sol: Total number of balls = (2 + 3 + 2) = 7 Let S be the sample space.

Then, n(S) = Number of ways of drawing 2 balls out of 7

=7C2

=(2×1)(7×6) =21

Let E = Event of drawing 2 balls, none of which is blue.

n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls.

=5C2

=(2×1)(5×4)

=10

P(E)=n(E) / n(S) = 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

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

Q8) Calculate Expected Value for the problem below

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?

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

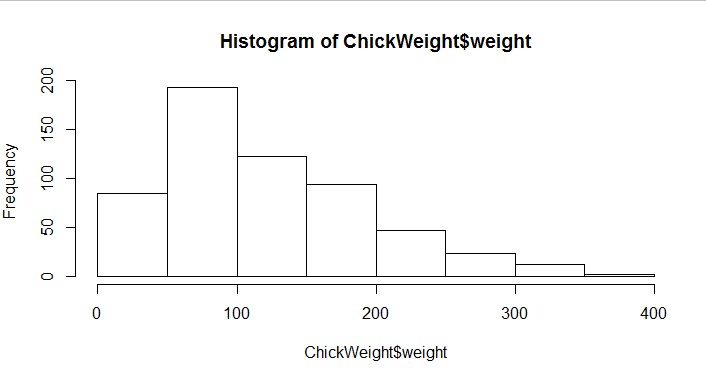
Cars speed and distance

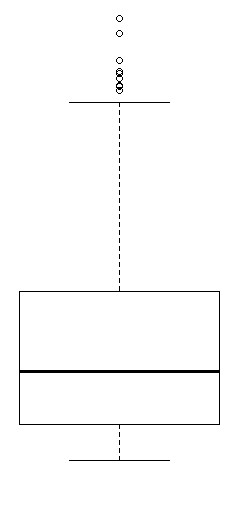
Use Q9\_a.csv

SP and Weight (WT)

Use Q9\_b.csv

Q10) Draw inferences about the following boxplot & histogram





Sol: Histogram:- The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.

Box Plot: - The boxplot has outliers on the maximum side

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 Find mean, median, variance, standard deviation.

What can we say about the student marks?

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

Sol: No skewness is present we have a perfect symmetrical distribution

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

Sol: Skewness and tail is towards Right

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

Sol: Skewness and tail is towards left

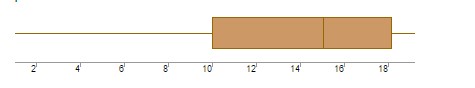
Q16) What does positive kurtosis value indicates for a data?

Sol: Positive kurtosis means the curve is more peaked and it is Leptokurtic.

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

Sol: Negative Kurtosis means the curve will be flatter and broader.

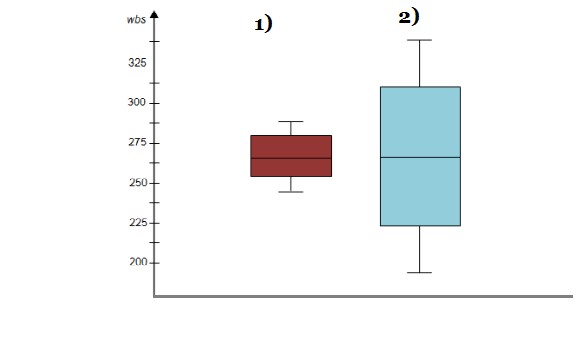
Q18) Answer the below questions using the below boxplot visualization.



1. What can we say about the distribution of the data?
2. What is nature of skewness of the data?
3. What will be the IQR of the data (approximately)?

Sol: a: The above Boxplot is not normally distributed the median is towards the higher value

1. The data is a skewed towards left. The whisker range of minimum value is greater than maximum
2. The Inter Quantile Range = Q3 Upper quartile – Q1 Lower Quartile = 8–10 =8 Q19) Comment on the below Boxplot visualizations?



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

Sol: Firstly, there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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

P(MPG>38) P(MPG<40)

c. P (20<MPG<50)

Q 21) Check whether the data follows normal distribution

Check whether the MPG of Cars follows Normal Distribution Dataset: Cars.csv

Sol: - MPG of cars follow Normal Distribution

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

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 Hint: rcode → pt(tscore,df)

df → degrees of freedom