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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| 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 0
2. Less than or equal to 4
3. 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?

A)

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)

|  |  |  |
| --- | --- | --- |
| 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

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.

1. Mean for Points is 3.596563

Median for Points is 3.695

Mode for Points is 3.07 & 3.92

Variance for Points is 0.2858814

Standard Deviation for Points is 0.5346787

Range for Points is 2.76 & 4.93

Mean for Score is 3.21725

Median for Score is 3.325

Mode for Score is 3.44

Variance for Score is 0.957379

Standard Deviation for Score is0.9784574

Range for Score is 1.513 & 5.424

Mean for Weight is 17.84875

Median for Weight is 17.71

Mode for Weight is 17.02 & 18.90

Variance for Weight is 3.193166

Standard Deviation for Weight is1.786943

Range for Weight is 14.5 & 22.9

From above details we can say that weight has more information compare to Points & Score data by observing the variance and standard deviation of weight.



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?

The Expected Value of the Weight of that is 145 (Aproxx.)

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

**Cars speed and distance**



Skewness of speed is -0.1139548 which is Left skewness

Skewness of dist is 0.7824835 which is Right skewness

Kurtosis of speed is 2.422853

Kurtosis of dist is 3.248019

**SP and Weight(WT)**



Skewness of SP is 1.581454 which is Right skewness

Skewness of WT is -0.6033099 which is Left skewness

Kurtosis of SP is 5.723521

Kurtosis of WT is 3.819466

**Q10) Draw inferences about the following boxplot & histogram**



From the above Histogram Plot we can say that the above plot is Right skewed as most of the observation/data is on the left side of histogram plot and most of the observation are below 200.



From the above Boxplot we can say that 75 Percentile of the data/observations are present/situated near to the Lower Limit of the Boxplot and few observations/data are situated above the Upper Limit of the Boxplot are formed as 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?

Confidence Interval for 94% is (143.54415570565965, 256.45584429434035)

96% is (138.34730111522666, 261.6526988847733)

98% is (130.15355671679083, 269.84644328320917)

**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 of the above data is 41

Median of the above data is 40.5

Variance of the above data is 25.52941

Standard Deviation of the above data is 5.052664

1. What can we say about the student marks?

From the details of Answer 1 in question 12 we can say that mean is greater than median. So the data is slightly Right Skewed data.

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

When mean & median are equal then the data is perfectly symmetric and skewness is 0(zero).

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

When mean is greater than median then it is a Right Skewed Data.

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

When mean is lessthan median then it is a Left Skewed Data.

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

Positive Kurtosis Value indicates Sharper Peak and Broad/Fatter Tails for a data.

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

Negative Kurtosis Value indicates Flatter/Wider Peak and Thin Tails for a data.

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



Q) What can we say about the distribution of the data?

A) From above Boxplot we can say that most of the data is present between 10 and 15.

Q) What is nature of skewness of the data?

1. We can say it is Left Skewed Data.

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

A) The IQR is the width of the Box in Boxplot, IQR = Q3 – Q1 = 18-10 = 8  
  
Q19) Comment on the below Boxplot visualizations?



1. By observing Boxplot of 1 & 2 we can say both data of Boxplots are normally distributed and both Boxplot median is 260 (approximately)

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

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)

* 1. P(MPG>38) –> 0.3475941
  2. P(MPG<40) –> 0.7293497

c. P(20<MPG<50) –> 0.01311637

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

1. The Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at does not follows Normal Distribution

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

A) Z Scores of 90% Confidence Interval is 1.6448536269514722

94% Confidence Interval is 1.8807936081512509

60% Confidence Interval is 0.8416212335729143

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

A) T Scores of 95% Confidence Interval is 2.0638985616280205

96% Confidence Interval is 2.1715446760080677

99% Confidence Interval is 2.796939504772804

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)Probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 0.32167253567098364