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

NAME – SIDDHI DESHMANE

Q1) Identify the Data type for the Following:

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

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

ANS- Sample Space ={HHH,HHT,HTH,THH,HTT,TTH,THT,TTT} ; n(S) = 8

A ={HHT,HTH,THH}; n(A) = 3

P(A) = n(A)/n(S)

= 3/8

= 0.375

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

1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2 and 3

ANS- Sample Space S = {(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)}

n(S) = 36

1. EVENT A is when the sum is 1 which is impossible

n(A) = 0

P(A) = 0

1. EVENT B is when sum is Less than or equal to 4

B = {(1,3),(3,1),(2,2)} ; n(B) =3

P(B) = 1/12

=0.0833

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 – Sample space n(S) = 21

Event A being no blue ball in two balls

n(A) = 10

P(A) = n(A) /n(S)

= 10/21

=0.4761

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

ANS - Expected number of candies for a randomly selected child

=  1 \* 0.015  + 4\*0.20  + 3 \*0.65  + 5\*0.005  + 6 \*0.01  + 2 \* 0.12

= 0.015 + 0.8  + 1.95 + 0.025 + 0.06 + 0.24

=       3.090

=  3.09

THIS CAN BE CONSIDERED AS 3

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

1. 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- The Expected Value of the Weight of a patient chosen at random is: 145.33 pounds

**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**



ANS – Here we can observe that

1. the distribution is positively skewed
2. mass of the distribution is concentrated between 1-350 which can be considered as the range of the distribution
3. As the given data is displayed as histogram we can conclude that the data is continuous
4. The data has some outliers above 350

**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.

ANS Mean: 40.47

Median: 40.5

Variance: 49.67

Standard Deviation: 7.05

1. What can we say about the student marks?

The mean and median are quite close, this shows that the data is roughly symmetrically distributed around the central value of approximately 40. Also the SD and variance are relatively high indicates the spreading of data on large scale

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

ANS – The data has zero skewness in that case . we can imply that the given data is forming a relatively symmetric figure on the graph and is neither positively nor negatively skewed

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

ANS – The data is positively skewed as mean is greater than median

The data is right skewed

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

ANS – The data is negatively skewed as mean is smaller than median

The data is left skewed

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

ANS - Positive kurtosis indicates leptokurtic distribution.

This includes sharper peak and longer tails

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

ANS - Negative kurtosis indicates platykurtic distribution.

This includes broader peak and shorter tails

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



What can we say about the distribution of the data?

ANS - The above Boxplot is not normally distributed the median is towards the higher value

What is nature of skewness of the data?

ANS - The data is positively skewed as mean is greater than median

The data is right skewed

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

Ans: The Inter Quantile Range = Q3 Upper quartile – Q1 Lower Quartile

= 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.

ANS – 1. there are no outliers.

2. both the box plot shares the same median that is approximately in a range between 275 to 250

3. 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)= 0.3466923853688819

P(MPG<40) = 0.7306083416219191

P (20<MPG<50) = 0.01297

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

ANS – The MPG follows normal distribution

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

Dataset: wc-at.csv

ANS – AT and Waist follow normal distribution

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

ANS -Z score for 90% Confidence Interval = -1.6449

Z score for 94% Confidence Interval = -1.8808

Z score for 60% Confidence Interval = -0.8416

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

ANS -T score for 95% Confidence Interval = -2.0639

T score for 94% Confidence Inteval = -1.974

T score for 95% Confidence Interval = -2.7969

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

ANS - t-score: -0.4714

Degrees of Freedom: 17

Probability: 0.3217