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

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

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

🡪When three coins are tossed the total number of possible combinations are = 8

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

The number of combinations which have two heads and one tails are: HHT,HTH,TTH.

The Probability of getting two heads and one tails in the toss of three coins P(Two heads and One tail)

= Number of desired outcomes

= 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

🡪Total possible outcomes = = 36

1. Favorable outcome ( sum equal to 1 ) = 0 (i.e. not possible that sum always exceed to 1)

Required probability

= 0/36

= 0

1. When we roll two dice,the possibility of getting number 4 is (1,3), (2,2),(3,1),(1,1),(2,1),(1,2).

The number of favorable outcomes = 6

Probability = The number of favorable outcomes / Total number of possible outcomes = 6/36 = 1/6 = 0.17

1. Favorable Outcomes = sum is divisible by 2 and 3 = (1,5),(2,4),(3,3),(4,2),(5,1),(6,6)

Number of favorable outcomes = 6

Probability = number of favorable outcomes / number of possible outcomes = 6/36 = 1/6 = 0.17

Thus the probability that sum is divisible by 2 and 3 is 0.17

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?

🡪Total number of balls = 2+3+2 = 7

Number of ways of drawing 2 balls out of 7 = 7C­2

(7 \* 6) / (2 \* 1) = 42/2 = 21

Number of balls other than blue = 5

Number of ways of drawing 2 balls out of 5 = 5C2

(5 \* 4) / (2 \* 1) = 20/2 = 10

Probability = 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 = 0.20

🡪 Expected number of candies for a randomly selected child

= 1\*0.15 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.01 + 2\*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**

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?

🡪Expected Value = ∑ (Probability \* Value)

∑ P(x).E(x)

There are 9 patients

Probability of selecting each patient = 1/9

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| E(x) | 108 | 110 | 123 | 134 | 135 | 145 | 167 | 187 | 199 |
| P(x) | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 |

= (1/9)(108) + (1/9)(110) + (1/9)(123) + (1/9)(134) + (1/9)(135) + (1/9)(145) + (1/9)(167) + (1/9)(187) + (1/9)(199)

= 145.33

Expected Value of the Weight of that patient = 145.33

**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 🡪 The histograms peak has right skew and tail is on right. Mean > Median. We

have outliers on the higher side.

Ans 🡪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**

1. Find mean, median, variance, standard deviation.
2. What can we say about the student marks?

🡪1) Mean = ∑(Xi)/n

= 34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56 / 18

= 738 / 18

= 41

Median = 40.5

Variance = ∑(Xi - )2/n-1

= 49+25+25+9+9+4+4+1+1+1+1+16+64+225/17

= 434/17

= 25.52

Standard Deviation = 5.05

2) We don’t have outliers and the data is slightly skewed towards right

because mean is greater than median.

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

🡪If the distribution is symmetric, then the mean is equal to the median, and

the distribution has zero skewness.

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

🡪If the mean is greater than median, the distributed is positively skewed.

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

🡪If the median is greater than the mean, the distributed is negatively skewed.

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

🡪Positive value of kurtosis indicate that distribution is peaked and possesses thick tails.

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

🡪 distribution with a negative kurtosis value indicates that the distribution

has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

🡪The above boxplot is not normally distributed the median is towards the higher value.

What is nature of skewness of the data?

🡪The data is a skewed towards left. The whisker range of minimum value is

greater than maximum.

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

🡪The Inter Quartile 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.

🡪First there are no outliers. Second both the box plot shares the same median that is approximately in range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum 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

* 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

🡪MPG of cars 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

🡪Adipose Tissue (AT) and Waist does not follows Normal Distribution.

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