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

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 | Ratio |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| 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 | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

**A)** When three coins are tossed the total number of possible outcomes are 2^3=8.

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

The number of combination which have two heads and one tail are:

HHT, HTH, TTH which three in number.

Then the probability of two heads and one tail are:

P (Two heads and One tail) = Number of desired outcomes / Total No. of outcomes

= 3/8

= 0.375

∴ P (Two heads and One tail) = 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

**A)** a) When two dice are rolled the probable outcomes will be 36.

The probability that sum is equal to 1 is Zero, because the least possible

sum will be 2 of (1,1).

b) The number of outcomes of dice rolled for sum less than or equal to 4 are 6:

(1,1), (1,2), (1,3), (2,1), (2,2), (3,1).

Probability that sum is less than or equal to 4 is

P (sum less than or equal to 4) = 6/36 =1/6 = 0.166.

c) The number of outcomes of dice rolled for sum is divisible by 2 and 3 are 6 (1,5), (2,4), (3,3), (4,2), (5,1), (6,6).

Probability that sum is divisible by 2 and 3 is

P (Sum is divisible by 2 and 3) = 6/36 =1/6 =0.166.

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)** 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

= (7x6)/(2x1) = 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 = (5x4)/(2x1) = 10

∴ P(E) = n(E)/n(S) =10/21 =0.476.

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

**A)** Expected number of candies for a randomly selected child =

1x0.015 + 4x0.20 + 3x0.65 + 5x0.005 + 6x0.01 + 2x0.120 = 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.

Weigh has more variance, hence it has more variety of data which will be used to find best suited model for the given dataset. In the given datasets no case has

Mean= Median = Mode.

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?

**A)** Expected value = ∑ (probability X 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.

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

= (1/9) (1308) = 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**

 In the given data of car speeds and distances, distances are right skewed majority of the distances are in interval of 20-40 and speeds are normally distributed majority of the speeds are in middle.

Both distance and speed have positive kurtosis.

**SP and Weight(WT)**

 In the given data of SP and WT, SP is right skewed, most of the data lies in the interval of 110-120 which are 41 in number and WT are normally distributed majority of the WTs are in middle.

Both WT and SP have Positive kurtosis.

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



**A)** For the given histogram of Chick weight of weights, the histogram is right skewed. Majority of the weights are in left side in the interval of 50-100.



For the given box plot there are outliers are present which are dominating the data which needed to be eliminated. Data 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?



|  |  |  |  |
| --- | --- | --- | --- |
|  | 94% | 98% | 96% |
| Upper | 201.04 | 201.38 | 201.17 |
| Lower | 198.96 | 198.62 | 198.83 |

**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?

**A)**  1) Mean= 41

Median=40.50

Variance= 25.52

Standard deviation = 5.052

2) Marks of students are not normally distributed. The given data has outliers. Majority of the students scored between 35-45 marks.

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

**A)** When mean and median of the data is equal it indicates that Skewness will be 0.

Curve of the data will be perfectly symmetric bell shaped.

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

**A)** When mean > median, it indicates that Skewness of the given data is positive (Right skewed). And the data will be distributed on left side.

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

**A)** When median > mean, it indicates that Skewness of the given data is negative (Left skewed). And the data will be distributed on right side.

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

**A)** Positive kurtosis value will indicate high and narrow peak on central part of data and light tail.

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

**A)** Negative kurtosis value will indicate wider peak on central part of the data and heavy tailed.

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



What can we say about the distribution of the data?

**A)** The given data indicates that it is not symmetrical and it is more concentrated towards right side.

What is nature of skewness of the data?

**A)** Given data indicates it is left skewed (Negative skewness).

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

**A)** IQR of the given data is 8 (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.

**A)**

* Data is Normally Distributed. No outliers. Center around 263.5.

Comparatively, first graph has less range.

* Data is Normally Distributed. No outliers. Center around 263.5.

Comparatively second graph has more 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)

**A)**  P(MPG>38) = 0.40740

* 1. P(MPG<40)

**A)** P(MPG<40) **=** 0.75308

c. P (20<MPG<50)

**A)**  P (20<MPG<50) = 0.85158



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



**A)** MPG is in 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

**A)** Both AT and waist doesn’t follow Normal Distribution.

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

**A)** Z-score (90%) = 1.28155

Z-score (94%) = 1.55477

Z-score (60%) = 0.25443



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

**A)**

|  |  |
| --- | --- |
| 95% | ± 2.060 |
| 96% | ± 2.167 |
| 99% | ± 2.787 |

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:

r-code 🡪 pt (t-score, df)

df 🡪 degrees of freedom

**A)** t-score = -0.4714

Degree of freedom= 17

P(t) = 0.32167

