|  |  |
| --- | --- |
| 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 | 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 | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Interval |
| 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:** Probability that two heads and one tail are obtained is 3/8.

{HHH, HTT, HHT, HTH, THT, THH, TTH, TTT}

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:** a) Sum is equal to zero. [ P (1) = 0]

b) The possibility of getting number less than or equal to 4 is 1/6.

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

The no. of favorable outcomes = 6

Total No. of possibilities = 36

Probability = 6/36 = 1/6

c) Probability of sum is divisible by 2 and 3 is 1/6.

(1,5), (3,3), (4,2), (2,4), (5,1), (6,6).

Favorable outcomes = 6

Total no. of possibilities = 36

Probability = No. of favorable outcome/No. of possible outcomes

= 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?

**ANS:** Probability that none of the balls drawn is blue = 10/21.

Total No. of balls = (2+3+2) = 7

n(S) = (7\*6)/ (2\*1) = 21

n(E) = (5\*4)/ (2\*1) = 10

p(E)= n(E)/n(S) = 10/21

Q6) Calculate the expected number of candies for a randomly selected child is

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 no. of candies for randomly selected child is 3.09.

Expected no. of candies for a randomly selected child = [(1\*0.015) + (4\*0.20) + (3\*0.65) + (5\*0.005) + (6\*0.01) + (2\*0.120)] = 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

**ANS:**  Parameters Point Score Weigh

Mean 3.62 3.22 17.85

Median 3.79 3.32 17.71

Mode 4.08 3.44 18.9

Variance 0.286 0.9574 3.19

Standard Deviation 0.535 0.9784 1.78

Range 2.17, 4.93 1.513, 5.424 14.5, 22.9

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:** Expected Value of the Weight of that patient is 145.33.

There are 9 patients at a clinic probability of selecting each patient = 1/9

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

= (1/9) (1308) = 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

**ANS:**

Car speed and distance data:

Skewness = -0.114 (car speed), 0.782 (distance)

Kurtosis = -0.51 (car speed), 0.41(distance)

SP and WT data:

Skewness = 1.61 (SP), -0.615 (weight)

Kurtosis = 2.98 (SP), 0.95 (weight)

In given data of car speed and distance the skewness of car speed is negative so, the data are negatively skewed meaning that the left tail is long.

In kurtosis case car speed is negative so, simply means that more data values are located near the mean and less data values are located on the tails as negative kurtosis is the uniform distribution, which has no peak at all and is a completely flat distribution.

In Distance data skewness value is positive so, the mean and median is greater than mode.

In kurtosis case of distance data is positive so, simply means that fewer data values are located near the mean and more data values are located on the tails.

Q10) Draw inferences about the following boxplot & histogram



**ANS:**

Here figure shows histogram of ChickWeight$weight vs frequency as well as a box plot.

Where it is positively skewed on right side i.e. mean and median of the data is greater than mode.

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?

**ANS:** Sample Mean = 200

Sample standard deviation = 30

Sample size = 2000

Using the t-distribution,

1. The 94% confidence interval is (198.73, 201.39) (t= 1.8916)

2. The 96% confidence interval is (198.61, 201.39) (t= 2.0623)

3. The 98% confidence interval is (198.43, 201.57) (t= 2.3452)

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?

**ANS:**  1) Mean = 41

Median = 40.5

Variance = 25.529

Standard Deviation = 5.05

2) Mass of students marks between 38-42. Skewness (1.52) is positive because mass of marks in left side of plot.

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

**ANS:** When mean, median of data are equal, the skewness of the distribution is zero.

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

**ANS:** When mean is greater than the median, the distribution is positively skewed.

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

**ANS:** Whenmean is less than the median, the distribution is negatively skewed.

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

**ANS:** Positive values of kurtosis indicate that distribution is peaked and possesses thick tails. An extreme positive kurtosis indicates a distribution where more of the numbers are located in the tails of the distribution instead of around the mean.

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

**ANS:** If a distribution has negative kurtosis, it is said to be platykurtic, which means that it has a flatter peak and thinner tails compared to a normal distribution. This simply means that more data values are located near the mean and less data values are located on the tails. negative kurtosis is the uniform distribution, which has no peak at all and is a completely flat distribution.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

**ANS:**

1. A negatively skewed distribution, it is a type of distribution in which more values are concentrated on the right side of the distribution graph while the left tail of the distribution graph is longer.

2. The nature of skewness of the data is negatively skewed data where Mean will be less than the Median.

3. The IQR describes the middle 50% of values when ordered from lowest to highest. In above example of data, the IOR = (-10-18).

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot2.

**ANS:**

* Boxplot 1 is positively skewed i.e. Mean and Median are greater than Mode.
* Boxplot 2 shows normal distribution that means the skewness for a normal distribution is zero.

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)

**ANS:** From Given Data,

P(MPG>38)- 1- p-norm (38,34.422,9.13144) = 0.3475908

P(MPG<40)- p-norm (40,34.422,9.13144) = 0.7293527

P (20<MPG<50)- p-norm ((50,34.422,9.13144) - (1-pnorm (20,34.422,9.13144))

= 0.01311818

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

**ANS:**

1. Here it shows that data is highly negative skewed.
2. Here it shows that AT data is bimodal and Waist is multimodal data hence, AT data is positively skewed and Waist data is fairly symmetric.

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

**ANS:**

|  |  |
| --- | --- |
| Confidence Interval | Z score |
| 90 % | 0.8416212 |
| 94 % | 1.644854 |
| 60 % | 1.880794 |

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

**ANS:**

|  |  |
| --- | --- |
| Confidence Interval | T score |
| 95 % | 2.063899 |
| 96 % | 2.171545 |
| 99 % | 2.79694 |

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:**

Population mean = 270 days

Standard deviation = 90 days

No. of items in sample = 18 bulbs

Mean of sample bulbs = 260 days

Df = 17

Using t-distribution,

The probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 0.3217. Assuming alpha = 0.05 thus, p-value > alpha i.e. The CEO’s claim is false and the average life of bulb is greater than 260 days.