|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Numeric |
| Results of rolling a dice | Numeric |
| Weight of a person | Numeric |
| Weight of Gold | Numeric |
| Distance between two places | Numeric |
| Length of a leaf | Numeric |
| Dog's weight | Numeric |
| Blue Color | Categorical |
| Number of kids | Numeric |
| Number of tickets in Indian railways | Numeric |
| Number of times married | Numeric |
| Gender (Male or Female) | Categorical |

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

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

**Ans**: Sample space = {HHH,HHT,HTH,THH,TTH,THT,HTT,TTT}

Number of favorable events = 3

Total number of events = 8

P (Getting two heads and one tail) = 3/8

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:**

1. Sample space = {(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) }

Number of favorable events = 0

Total number of events = 36

P (Sum equal to 1) = 0/36

= 0

1. Number of favorable events = 6

P (Sum is less than or equal to 4) = 6/36

= 1/6

1. Number of favorable events = 6

P (Sum is divisible by 2 and 3) = 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:** Total number of balls = 7

Total number of events = 7C2

= 7!/(5!\*2!)

= 21

E: Balls drawn are not blue otherwise ball drawn are either red or green

Number of events that balls drawn are either red or green = 5C2

= 5!/(3!\*2!)

= 10

P (E) = 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

**Ans:** X: Number of candies for children

E(x) = Σ x\*P(x)

= 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**: The mean value for Point, Score and Weigh are 3.596563, 3.217250 and 17.848750 respectively.

The median values for Point, Score and Weigh are 3.6950, 3.3250 and 17.7100respectively.

The mode values for Point are 3.07 and 3.92.

The mode value for Score is 3.44.

The mode values for Weigh are 17.02 and 18.90.

The variance for Points, Score, Weigh are 0.2859, 3.1932, 0.9574 respectively.

The standard deviation for Points, Score, Weigh are 0.5347, 0.9785, 1.7869 respectively.

The range for Points, Score, Weigh are 2.17, 3.9110, 8.4 respectively.

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:**  X: Weights of the patients

E(x) = Mean value x

= (1/n) Σ x

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

= 145.33

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data Cars speed and distance**

**Use Q9\_a.csv**

**Ans:** Speed is negatively skewed and is mesokurtic. Distance is positively skewed and is leptokurtic.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:** SP is positively skewed and leptokurtic. Weight is negatively skewed and leptokurtic.

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



**Ans:** The histogram is positively skewed or right skewed, since the tail is to the right side.

The boxplot is negatively skewed or left skewed, since area of the first quartile region is higher than that of third. The data has 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?

**Ans:** Confidence interval for confidence level 94% is (143.5762, 256.4238).

Confidence interval for confidence level 98% is (130.2096, 269.7904).

Confidence interval for confidence level 96% is (138.3875, 261.6125).

**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 = 25.50

Standard deviation = 5.05

2) Mark of students are normally distributed with mean = 41 and

variance = 24.11

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

**Ans:** The distribution is not skewed, that is symmetric if mean = median = mode.

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

**Ans:** Right skewed or positively skewed.

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

**Ans:** Left skewed or negatively skewed.

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

**Ans:** Positive kurtosis indicates that the data is leptokurtic.

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

**Ans:** Negative kurtosis indicates that the data is mesokurtic.

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:** The data is negatively skewed. There are more data points on the left side of the data.   
The IQR of the data will be approximately 8.

Q19) Comment on the below Boxplot visualizations?



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

**Ans:** Both data has the same median value. The data shown in Boxplot 1 has less inter-quartile range compared to the data shown in Boxplot 2. Both data are symmetrically distributed. There are no outliers in both data.

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)
  3. P (20<MPG<50)

**Ans:** a. 0.3476

b. 0.7293

c. 0.8989

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:** a) MPG of cars follows normal distribution.

b) Adipose Tissues from the data set wc-at.csv does not follows normal distribution. But Waist circumference follows normal distribution.

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

**Ans:** The Z values for confidence intervals 90%,94%,60% are 1.2816, 1.5548, 0.2533 respectively.

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

**Ans:** The corresponding t scores are 1.7108, 1.8280, 2.4921 respectively.

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:**  P [sample mean<=260] = 0.5