|  |  |
| --- | --- |
| 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 | Categorical/Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical/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 | ratio |
| 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 | ratio |
| Number of Children | ratio |
| Religious Preference | nominal |
| Barometer Pressure | nominal |
| SAT Scores | ordinal |
| Years of Education | ratio |

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

Solution: Total possible events: 8, no. of events occurred = 3

Probability = 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

Solution:

1. 0
2. 1/6
3. 5/36

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?

Solution: 5c2/ 7c2 = 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

Solution:

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

=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**

**Points:**

· Mean =3.59

· Median=3.695

· Mode=3.92

· Variance=0.2809

· Standard Deviation=0.53

· Range=2.17

**Score:**

· Mean =3.217

· Median=3.325

· Mode=3.44

· Variance=0.956

· Standard Deviation=0.978

· Range=3.911

**Weigh:**

· Mean =17.84

· Median=17.71

· Mode=17.02

· Variance=3.189

· Standard Deviation=1.786

· Range=8.4

* Mean<median<mode in case of Point and Score and hence these two are Negatively skewed.
* Mean>Median>mode in case of Weigh and hence it is positively skewed

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?

Solution:

Total number of patients = 9

Probability of choosing one patient = 1/9

Expected value of weight = 108/9+110/9+123/9+134/9+135/9+145/9+167/9+187/9+199/9

= 145.3

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

Skewness: speed = -0.1139

Dist = 0.78248

Kurtosis: speed = 2.422

Dist = 3.248

* “dist” is positively skewed where as “speed” is negatively skewed. dist has distribution of data concentrated on the left whereas speed has distribution on the right. As seen in the graph
* Both dist and speed has positive Kurtosis

**SP and Weight(WT)**

**Use Q9\_b.csv**

Skewness: SP= 1.5814

WT= 0.78248

Kurtosis: SP = -0.6033

WT = 3.81946

* “SP” is positively skewed where as “WT” is negatively skewed. SP has distribution of data concentrated on the left whereas WT has distribution on the right.
* Both WT and SP has positive Kurtosis

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



**Histogram:**



a) Majority of the Chicks has weight in range 50 – 100, followed by 100 -150 and 150 – 200

b) The data is positively Skewed

c) histogram mean >median>mode.

Boxplot:

a) Data has outliers

b) 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.

Mean : 41

Median : 40.50

Variance :25.53

Std Deviation: 5.05

1. What can we say about the student marks?

Data has outlier

Majority of the students scored between 35 – 45 Marks

Distribution of marks is positively skewed.

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

Skewness = 0. Perfectly symmetric bell shaped curve

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

if mean is greater than median it indicates Tail is on the right side of the distribution and it is positively skewed.

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

If Median is greater than mean it indicates tail is on left side of the distribution and it is negatively skewed.

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

Positive kurtosis indicates distribution contains peak value and has thick tails.

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

Negative kurtosis indicates distribution is flat and has thin tails.

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



What can we say about the distribution of the data?

The data is not symmetric. Data is more concentrated towards right side

What is nature of skewness of the data?

Skewness = Negative

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

Data is Normally Distributed. No Outliers. Center around 262.5. first graph has less range, 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) : 0.3474
  2. P(MPG<40): 0.2705

c. P (20<MPG<50) : 0.8989

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Chart, line chart, histogram

Description automatically generated

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

Dataset: wc-at.csv

Chart, histogram

Description automatically generated

Chart, line chart

Description automatically generated

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

* 90% confidence interval in decimal is 0.90

Calculating alpha = (1+0.90)/2= 0.95

Z score: 1.64

* 94% confidence interval in decimal is 0.94

Alpha =(1+0.94)/2=0.97

Z score: 1.88

* 60% confidence interval in decimal is 0.60

Alpha = (1+0.60)/2 = 0.8

Z score: 0.84

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

Confidence interval of 95%: t= 2.063899

• Confidence interval of 96%: t=2.171545

• Confidence interval of 99%: t=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

Population mean=270

Sample Mean = 260

Sample size = 18

Standard Deviation of Sample = 90

t-score = 260−270÷90/(√18) =- 0.471

Degree of freedom = (18-1)=17

P(t) = 0.322