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

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

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 – P(A) = Favorable Outcome/Total Outcome

Possible Outcomes = {HHH,HHT,HTH,HTT,THH,THT,TTH,TTT}

Favorable Outcome = {HHT,HTH,THH}= 3

Total Outcome = 8

P(A)=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

Ans. Probability – P(A) = Favorable Outcome/Total Outcome

Possible Outcomes = 36

1. When equal to 1 = 0
2. Less than or equal to 4

Possible Outcomes = 6

Total Outcome = 36

P(A)=6/36 = 1/6=0.16

1. Sum is divisible by 2 and 3

Possible Outcomes = 24

Total Outcome = 36

P(A)=24/36 = 2/3=0.66

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 -

P(A) = 5C2/7C2

= ((5\*4)/(2\*1) / (7\*6)/(2\*1)) = 10/21 = 0.467

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.

Expected number of candies = (1 \* 0.015) + (4 \* 0.20) + (3 \* 0.65) + (5 \* 0.005) + (6 \* 0.01) + (2 \* 0.120)

Expected number of candies = 0.015 + 0.80 + 1.95 + 0.025 + 0.06 + 0.24

Expected number of candies = 3.09

Therefore, the expected number of candies for a randomly selected child is 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. Python Code Attached.**

**Mean-**

Points 3.596563  
Score 3.217250  
Weigh 17.848750

**Median**-

Points 3.695  
Score 3.325  
Weigh 17.710

**Range -**

Points 4.930000-2.760000= 2.17  
Score 5.424000-1.513000= 3.911  
Weigh 22.000000-14.500000= 7.5

**Variance -**

Points 0.285881  
Score 0.957379  
 Weigh 3.193166

**Standard Deviation -**

Points 0.534679  
Score 0.978457  
Weigh 1.786943

Mode -

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Unnamed: 0** | **Points** | **Score** | **Weigh** |  |
| **0** | AMC Javelin | 3.07 | 3.44 | 17.02 |
| **1** | Cadillac Fleetwood | 3.92 | NaN | 18.90 |

**comment about the values / draw inferences -**

From the above data it is observed that points and score data are closely matched, whereas weigh data has outliers present.

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 -

Average = 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. For Speed**

|  |  |
| --- | --- |
| mean | 15.4 |
| sd | 5.234501 |
| skewness | -0.11751 |
| kurtosis | -0.50899 |

**Inference – Skewness in Negative, Which indicates that the distribution is slightly skewed to the left.**

**Kurtosis is Negative, which indicates that the distribution is slightly more peaked than a normal distribution.**

**For Distance**

|  |  |
| --- | --- |
| mean | 42.98 |
| sd | 25.51038 |
| skewness | 0.782484 |
| kurtosis | 0.405053 |

**Kurtosis is positive, which indicates that the distribution is slightly more peaked than a normal distribution. Skewness is positive, which indicates that the distribution is slightly skewed to the right.**

**Excel Attached.**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans. For SP**

|  |  |
| --- | --- |
| Skewness | 1.61145 |
| Kurtosis | 2.977329 |

**Kurtosis is positive, which indicates that the distribution is slightly more peaked than a normal distribution. Skewness is positive, which indicates that the distribution is slightly skewed to the right.**

**For Weight(WT)**

|  |  |
| --- | --- |
| Skewness | -0.61475 |
| Kurtosis | 0.950291 |

**Skewness in Negative, Which indicates that the distribution is slightly skewed to the left. Kurtosis is positive, which indicates that the distribution is slightly more peaked than a normal distribution.**

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



Ans. **Histogram** - The histogram shows that it is right skewed and has Outliers present in it.

**Box Plot –** Blot plot shows that there are outliers present in the higher side of the data Q3

**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.

For a 94% confidence interval, the z\* value is 1.88 (from a standard normal distribution table).

For a 98% confidence interval, the z\* value is 2.33.

For a 96% confidence interval, the z\* value is 1.96.

formula for the confidence interval for the population mean, given by:

CI = x̄ ± z\*(σ/√n)

where CI is the confidence interval

x̄ is the sample mean

z\* is the critical value

σ is the population standard deviation

n is the sample size.

* For 94% confidence interval:

CI = 200 ± 1.88\*(30/√2000) = (198.84, 201.16)

* For 98% confidence interval:

CI = 200 ± 2.33\*(30/√2000) = (197.38, 202.62)

* For 96% confidence interval:

CI = 200 ± 1.96\*(30/√2000) = (198.01, 201.99)

Therefore, we can say with 94% confidence that the average weight of an adult male in Mexico is between 198.84 and 201.16 pounds.

**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 – Excel file attached

|  |
| --- |
| Mean = 40.83 |
| Median = 40.5 |
| Standard Deviation = 4.42 |

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

Ans – Normally Distributed / Zero Skew.

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

Ans – Skewed to left.

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

Ans – Skewed to Right.

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

Ans – Indicates that the distribution has thicker tail than normal distribution.

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

Ans – Indicates that the distribution has tinner tail than normal distribution.

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



What can we say about the distribution of the data?

Ans: - Not normally distributed

What is nature of skewness of the data?

Ans: - Negative skewness

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

Ans: - 10 -18

Q19) Comment on the below Boxplot visualizations?



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

Q 20) Calculate probability from the given dataset for the below cases

Ans. By observing both the plots whisker’s level is high in boxplot

2.Mean and median are equal hence distribution is symmetrical.

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. a. P(MPG > 38) = 0.0003

b. P(MPG < 40) = 0.7389

c. P(20 < MPG < 50) = 0.9776

Code attached.

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans. The data follows a normal distribution

Code attached.

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

Dataset: wc-at.csv

ANS.The data does not follow a normal distribution

Code attached.

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

Ans .

1.For a 90% confidence interval, the Z-score is 1.645

2.For a 94% confidence interval, the Z-score is 1.880

3.For a 60% confidence interval, the Z-score is 0.253

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

Ans. df = 24 # (n-1)

1.For a 95% confidence interval with df=24, the t-score is 2.064

2.For a 96% confidence interval with df=24, the t-score is 2.171

3.For a 99% confidence interval with df=24, the t-score is 2.797

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 à pt(tscore,df)

df à degrees of freedom

ANS.

t = (260 - 270) / (90 / √18) = -0.94

the corresponding p-value using a t-distribution table or a t-distribution calculator with 17 degrees of freedom (18-1).

The p-value for a one-tailed test with a t-statistic of -0.94 and 17 degrees of freedom is 0.179.