|  |  |
| --- | --- |
| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ordinal |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Ans=3/8=0.375

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1

Ans=0

1. Less than or equal to 4

Ans=6/36=0.16

1. Sum is divisible by 2 and 3

Ans=6/36=0.16

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=10/21=0.47

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

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= points score weigh

Mean 3.596563 3.217250 17.848750

Median 3.695 3.325 17.710

Mode 0 3.07 3.44 17.02

1 3.92 NaN 18.90

var 0.285881 0.957379 3.193166

std. dev 0.534679 0.978457 1.786943

range 2.170 3.911 8.400

average of point , score and weigh is 3.596563 , 3.217250, 17.848750 respectively. After ordering the dataset the middle value found for point, score and weigh are 3.695, 3.325,170.710respectively.

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 the patient=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans. skewness for speed= -0.12**

**skewness for distance= 0.81**

**Kurtosis for speed= -0.51**

**Kurtosis for distance= 0.41**

**Speed is left skewed. The distribution is flatter than normal distribution.**

**Distance is right skewed.**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans. skewness for SP= 1.61**

**skewness for weight= -0.61**

**Kurtosis for SP=2.98**

**Kurtosis for weight= 0.95**

**SP is right skewed. And the distribution is flatter than normal distribution.**

**Weight is right skewed.**

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



Ans= The above boxplot suggests that the distribution has lots of outliers towards upper extreme.

**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=94% confidence interval=[195.841, 204.159]

98% confidence interval=[195.061, 204.939]

96% confidence interval=[195.841, 204.159]

**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?
3. Mean=41, median=40.5, variance=25.52, standard deviation=5.05
4. Mean is greater than median . this means distribution is slightly skewed towards right.

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

Ans-Nature of skewness is symmetric

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

Ans- Nature of skewness is Right skew

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

Ans- Nature of skewness is Left skew

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

Ans-Positive kurtosis indicates that more data are crowded around the mean value. Chances of finding extreme values are much lesser

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

Ans- Negative kurtosis indicates that data are not crowded around the mean value. Chances of finding extreme value is much more

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



What can we say about the distribution of the data?

Ans= It is not a normal distribution of the data.

What is nature of skewness of the data?

Ans=It is left skewed.

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

Ans= IQR=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.

Ans= The median of two boxplot is approximately same that is 260.

The boxplots are not skewed in any direction.

Both the boxplots have not any outliers.

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)

Ans=0.348

* 1. P(MPG<40)

Ans=0.729

* 1. P(20<MPG<50)

Ans= 0.214

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans= MPG of cars follow 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

Ans= Adipose Tissue and Waist Circumference follows the normal distribution.

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

Ans- Z score of 90% confidence interval=1.64

Z score of 94% confidence interval=1.88

Z score of 60% confidence interval=0.84

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

Ans= t score of 95% confidence interval=2.064

t score of 96% confidence interval=2.064

t score of 99% confidence interval=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:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

Ans= probability for average life of no more than 260 days =0.32