|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Descrete |
| Results of rolling a dice | Descrete |
| 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 | Descrete |
| Number of kids | Descrete |
| Number of tickets in Indian railways | Descrete |
| Number of times married | Descrete |
| Gender (Male or Female) | Descrete |

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

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

3/8

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

1. Equal to 1

0

1. Less than or equal to 4

1/6

1. Sum is divisible by 2and 3

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 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-> 3.090

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.

ANS->

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Mode | Variance | Std deviation | Range |
| Points | 3.5965 | 3.6950 | 3.07, 3.92 | 0.2858 | 0.5346 | 2.17 |
| Score | 3.2172 | 3.325 | 3.44 | 0.9573 | 0.9784 | 3.910 |
| Weigh | 17.8487 | 17.71 | 17.02 , 18.90 | 3.1931 | 1.7868 | 8.399 |

Inferences->

Here we can easily see that the variance in points and score is not too high but in weight it is about 3 so the points and score of a particular individual is close to the mean of the data set but the weight of a particular individuals are not close enough to each other.

And also by range we get to know that the difference between maximum and minimum points, score and weight are 2.17,3.91,8.399 respectively.

**Use Q7.csv file**

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->

Here the dataset is small so the expected value of the weight of that patient will be equal to the mean of the dataset which is 153.333.

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans->**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **Speed** | **-0.1175** | **-0.5089** |
| **Distance** | **0.8068** | **0.40505** |

By calculating the skewness of speed we came to know that it is left skewed which means that the data is spread from the mean towards right.

For distance skewness is positive so the data is spread towards left from the actual mean.

For kurtosis->

Kurtosis tells about the peakedness in the data .

For positive kurtosis it tells that the data has a more sharp peak.

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **SP** | **1.61145** | **2.9773** |
| **WT** | **-0.6147** | **0.9502** |

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



By observing the histogram we can say that it is positively skewed. So the mean weight of chicks is around 100 as most of the data is concentrated here.

By observing the boxplot we can say that there are 7-8 outliers whose weight can affect the mean of the total weights in the dataset.



**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 94% C.I. =(143.57619175546247, 256.42380824453755)

98% C.I. =(130.1535847418068, 269.8464152581932)

96% C.I. = (138.34732124381935, 261.65267875618065)

**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?

Mean : 41.0

Median : 40.5

Variance : 25.529412

Std Deviation: 5.052664

About the student marks-> as the standard deviation in not too high for the marks we can say that the group contains the students which are average in studies .

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

Ans> Then the skewness is equal to zero.

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

Ans> Then it is Right Skew or positive skew

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

Ans> Then it is Left Skew or negative skew

Q16) What does positive kurtosis value indicates for adata ?

Ans>A positive kurtosis indicates that the distribution has heavier tails than normal distribution.

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

Ans> A negative kurtosis indicates that the distribution has lighter tails than normal distribution.

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



What can we say about the distribution of the data?

ANS-> Here the 50% of the data is distributed in the range of 10-18.

What is nature of skewness of the data?

ANS-> As we can see that the data is much more distributed in the left of median so it is left skewed

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

ANS-> IQR = Q3-Q1 = 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-> both of the boxplot have the same median i.e. 262.5.. for boxplot 1 50% of the data lies in the range of 275-250 whereas in boxplot 2 50% of the data lies in the range of 325-225.

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

Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

* 1. P(MPG>38) = 0.3475939404153007
  2. P(MPG<40) = 0.7293498604157946

c. P (20<MPG<50) = 0.8988689076273199

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 shows Normal Distribution to some of the extent and can be considered as 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->no Adipose Tissue and Waist Circumference (Waist) both does not follow Normal Distribution.

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

ANS -> Z score of  
 90 % C.I. = [-1.65 , 1.65]

94 % C.I. = [-1.89 , 1.89]

60 % C.I. = [-0.85 , 0.85]

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

ANS -> 95 % C.I. = [-2.492 , 2.492]

96 % C.I. = [-2.492 , 2.492]

99% C.I. = [ -2.064, 2.064]

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 -> n=18, mean=270 , sample mean =260 , sample =90

t score = -0.47

prob = 0.32