|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | discrete |
| Results of rolling a dice | discrete |
| Weight of a person | continous |
| Weight of Gold | continous |
| Distance between two places | continous |
| Length of a leaf | continous |
| Dog's weight | continous |
| 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 | nomial |
| High School Class Ranking | Ordinal |
| Celsius Temperature | interval |
| Weight | ratio |
| Hair Color | nomial |
| Socioeconomic Status | ordinal |
| Fahrenheit Temperature | interval |
| Height | ratio |
| Type of living accommodation | nominal |
| Level of Agreement | ordinal |
| IQ(Intelligence Scale) | interval |
| Sales Figures | ratio |
| Blood Group | nomial |
| Time Of Day | ratio |
| Time on a Clock with Hands | ratio |
| Number of Children | ratio |
| Religious Preference | ordinal |
| Barometer Pressure | ratio |
| SAT Scores | ratio |
| Years of Education | ordinal |

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

Answer. 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

Answer. No .of possible outcomes that their sum is equal to 1 = 0 Therefore, that probability = 0

1. Less than or equal to 4

Answer. No.of possible outcomes that their sum is less than or equal to 4=6 i.e.,(1,1),(1,2),(1,3),(2,1),(2,2),(3,1) Therefore, probability=6/36=⅙

1. Sum is divisible by 2 and 3

Answer. No.of possible outcomes that their sum is divisible by 2 and 3=6 i.e.,(1,5),(2,4),(3,3),(4,2),(5,1),(6,6) Therefore, probability = 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?

Answer. No.of red balls = 2 No.of green balls = 3 No.of blue balls = 2 Probability of two balls are drawn at random is red=(2/7)\*(2/2)=2/7 Probability of two balls are drawn at random is green = (2/7)\*(2/3) = 4/21 Probability of two balls are drawn at random is not blue = 2/7+4/21=1

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

Answer. Expected number of candies for a randomly selected child is = 1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120 = 0.015+0.80+1.95+0.025+0.06+0.240=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**

**Answer.** Points: Mean = 3.596563 Median = 3.695 Mode = 3.92 Variance = 0.285881 Standard deviation = 0.534679 Range = 2.76-4.93 From the above the data we can see that Mean < Median < Mode then it shows that it has negatively skewness

Score: Mean = 3.21725 Median = 3.325 Mode = 3.44 Variance = 0.957379 Standard deviation = 0.978457 Range = 1.513-5.424 From the above the data we can see that Mean < Median < Mode then it shows that it has negatively skewness

Weigh: Mean = 17.84875 Median = 17.71 Mode = 17.02 Variance = 3.193166 Standard deviation = 1.786943 Range = 14.5-22.9 From the above the data we can see that Mode < Median < Mean then it shows that it has positively skewness

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?

Answer. Given weights of patients at a clinic(in pounds)= 108,110,123,134,135,145,167,187,199 If one of the patients is choosen at random, Then the Expected Value at the weight of that patient =Average of weights of patients = (108+110+123+134+135+145+167+187+199)/9 =145.333

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Answer.**

cars Speed:

Skewness = -0.11751

Kurtosis = -0.50899

Distance:

Skewness = 0.806895

Kurtosis = 0.405053

**SP and Weight(WT)**

**Use Q9\_b.csv**

SP:

Skewness = 1.61145

Kurtosis = 2.977329

Weight:

Skewness = -0.61475

Kurtosis = 0.950291

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



Answer.This is right side skewed histogram and its skewness is also positive



A) This is a 6 point chart and it is going to give the information about: Q1, Median value(Q2), Q3, lower limit, Upper Limit and 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?

Answer.For the 94% confidence interval:

CI=(198.23,201.76)

For the 98% confidence interval:

CI=(197.20,202.80)

For the 96% confidence interval:

CI=(198.58,201.42)

**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.

Answer. Mean = 41

Median = 40.5

Variance = 25.52941

Standard deviation = 5.052664

1. What can we say about the student marks?

Answer. Mean is greater than Median It is right skewed and contains no outlier

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

Answer. The nature of skewness is Perfectly skewness

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

Answer. The nature of skewness is Positively skewness

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

Answer. The nature of skewness is Negatively skewness

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

Answer. It indicates that a distribution is peaked and possess thick tails

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

Answer. It indicates that a distribution is flat and possess thin tails

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



What can we say about the distribution of the data?

Answer. We can say that there is a big difference between upper quartile and upper extreme compared to the lower quartile and lower extreme

What is nature of skewness of the data?

Answer.Left skewed

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

Answer.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.

Answer. Medium is same for both graph, there is difference between upper limit and lower limit (IQR1-IQR2) and there is no outliers in both the boxplot.

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)

Answer. P(MPG>38)=0.3475939251582705

* 1. P(MPG<40)

Answer.P(MPG<40)=0.7293462508918711

* 1. P (20<MPG<50)

Answer. P(20<MPG<50)=0.898868916740733

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer. Yes, it follows 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

Answer. Yes, both follows normal distribution.

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

Answer.from scipy.stats import norm

#Z-score of 90% confidence interval

norm.ppf(0.95)

1.6448536269514722

#Z-score of 94% confidence interval

norm.ppf(0.97)

1.8807936081512509

#Z-score of 60% confidence interval

norm.ppf(0.80)

0.8416212335729143

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

Answer. t scores of 95% confidence interval for sample size of 25

t.ppf(0.975, 24) #df= n-1=24

2.0638985616280205

t scores of 96% confidence interval for sample size of 25

t.ppf(0.98, 24) #df= n-1=24

2.1715446760080677

t scores of 99% confidence interval for sample size of 25

t.ppf(0.995, 24) #df= n-1=24

2.796939504772804

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

Answer. t-statistics for the data is given as follows:

