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

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

Answer-P(Two heads and one tail)=3/8=0.375=37.5%

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

Answer- (a) O

(b) 6/36

(c) 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?

Answer –10/21

Total balls = 7 balls

No. of ways of drawing 2 balls at random out of 7 balls ,n(S) = 7C2= 21

E event is drawing 2 balls such that balls are not blue

n (E) =No. of ways drawing balls out of 5 =5C2= 10

P(E)=n(E)/n(S)=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

Answer- E(A)=0.015

E(B)=0.8

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** | **Score** | **Weigh** |
| --- | --- | --- |
| **mean** | 3.596563 | 3.217250 | 17.848750 |
| **median** | 3.695000 | 3.325000 | 17.710000 |
| **var** | 0.285881 | 0.957379 | 3.193166 |
| **std** | 0.534679 | 0.978457 | 1.786943 |

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- 1/9\*1308=145.332

The expected value is 145.332

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

**Cars speed and distance**

**Use Q9\_a.csv**

Answer- Skewness for speed is -0.11751 which is negative skewed

Skewness for distance is 0.806 which is positive skewed

Kurtosis for speed is -0.5089(Platykurtic distribution) indicates the data is light –tailed relative to the normal distribution or the curve is flatter than the normal curve(lack of outliers).

Kurtosis for distance is 0.4050 (Leptokurtic distribution) indicates the data is heavy-tailed relative to ND and it has outliers

**SP and Weight(WT)**

**Use Q9\_b.csv**

Answer - Skewness for SP is 1.61145 which is Positive skewed

Skewness for WT is -0.61475 which is negative skewed

Kurtosis for SP is 2.977 which is high indicates that it has more outliers and the data is heavy –tailed relative to ND

Kurtosis for WT is 0.950 indicates distribution is peaked.

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



Answer- The data is positively skewed .Meaning –There are outliers

The no. of chicks whose weight is 100 are more



Answer -The top whisker is much longer than the bottom whisker

**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 90%=stats.t.interval(0.94,1999,200,0.67)=(198.7-201.26)

For 98%=stats.t.interval(0.98,1999,200,0.67)=(198.44-201.55)

For 96%=stats.t.interval(0.96,1999,200,0.67)=(198.62-201.3)

**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-25.3,standard deviation-5.053

1. What can we say about the student marks?

Answer - Most of the students have scored 40.There is variation of scores of 25 students.

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

Answer-The data has the perfectly symmetrical distribution.

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

Answer-The skewness is positive

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

Answer-The skewness is negative

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

Answer- The positive kurtosis value indicates that the distribution has higher peak and taller tails than a normal distribution.

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

Answer- The negative kurtosis value indicates that the distribution has lighter tails than the normal distribution

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



What can we say about the distribution of the data?

Answer - The median is 15.5 approx

Min=-2 max=20

The distribution is not normal

The 50% of the data ranges from 10-18

What is nature of skewness of the data?

Answer-Its 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?



Answer - The median for both the box plots is same i.e, 262.5

The distribution is normal

(50% of the data ) IQR for box plot 1 : 250-275

IQR for box plot 2 : 225-300

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

Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

* 1. P(MPG>38)

Answer-0.3476

* 1. P(MPG<40)

Answer-0.729

* 1. P (20<MPG<50)

Answer-0.956

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer- The graph is not approximately bell shaped and symmetric about the mean.Its’s bimodal.There are outliers .

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

Dataset: wc-at.csv

Answer- There is positive linear correlation in the beginning but then there r clusters and few outliers .

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

Answer=The Z scores of 90% confidence interval =1.645

The Z scores of 94% confidence interval =1.8807

The Z scores of 60% confidence in terval =0.85

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

Answer The T score of 95% confidence interval= 2.064

The T score of 96% confidence interval=2.085

The T score of 99% confidence interval =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 🡪 pt(tscore,df)

df 🡪 degrees of freedom

Answer : t=-0.4714

Stats.t.cdf(-4714,17)=32167%