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

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

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

Sol- P= {HHH, HHT, HTH, HTT,THH,THT,TTH,TTT}

=3/8

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

Sol- a) There is no outcomes whose sum is equal to one i.e Probability, 0/36= 0

b) (1, 3) (2, 2) (3, 1)= 3/36= 1/12

c) 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?

Sol- Total no. of balls= (2R+3G+2B)= 7 balls

n (s) = two balls drawn at random= 7C2= 21

Where, event of 2 balls, none of which is blue means 2R+3G=5

n(E)= 5C2= 10

P(E)= 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:-

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

=**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.

**Use Q7.csv file**

**Sol- Points**: Mean =3.596563, Median= 3.695, Mode= “3.07 3.92”,

Variance= 0.2858814, Standard deviation= 0.5346787.

**Score:** Mean= 3.21725, Median= 3.325, Mode= 3.44,

Variance= 0.957379, Standard deviation= 0.9784574

Note: Mean value are closer for both ‘Point’ and ‘Score’.

**Weight:** Mean= 17.84875, Median= 17.71, Mode= “17.02, 18.90”,

Variance= 3.193166, Standard deviation= 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?

Ans- since, no probability is given so expected value is also known as weighed mean, Mue= 1308/9= 145.333

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

Cars speed and distance

Use Q9\_a.csv

SP and Weight(WT)

Use Q9\_b.csv

**Ans- 1)** Skewness for speed= -0.1139548, skewness value is negative so it is left skewed. Since magnitude is slightly greater than 0 it is slightly left skewed

And for distance= 0.7824835, right skewed (Positive) slight magnitude to right.

Kurtosis for speed= 2.422853 Kurtosis for distance= 3.248019

**2)** Skewness for speed= 1.581454, skewness value is positive so it is a right skewed. And for weight = -0.6033099 and it’s a negative skewed.

Kurtosis for speed= 5.723521 positive, Kurtosis for weight= 3.819466 positive

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



Sol- The most of the data points are concentrated in the range 50-100

And the least range of the weight is 400 somewhere around 0-10.

Skewness- It is noticeable a long tail towards a right so it is heavily right skewed.



Sol- Mean is greater than the Median so it is a right skewed and we have a outlier on the upper side of the box plot and there is a less data points between Q1 and bottom point.

**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?

Sol- N (sample size)=2000 >30 so it is a normal, mean weight as the sample statistic is 200. Confidence level= 94%

Standard error= 30/ (square root of 2000)= 0.670.

Find, alpha (a)= 1-(conf.level/100)=0.06

Critical probability= P\*=1-alpha/2= 0.97

Degree of freedom(df)=(n-1)= (2000-1)=1999

To find the critical value is the t score while using df 1999 and the critical probability 0.97 using t chart is 1.88.

Now, margin of error (ME)= critical value \* standard error of the mean=1.88\*0.670= 1.2596

200+-1.2596

**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?

Sol- 1) Mean= 738/18= 41

2) Median = 40.5

3) Variance= 22.1667

4) Standard deviation = 4.7081

Mean &gt; Median, This implies that the distribution is slightly skewed

towards right. No outliers are present.

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

Sol- Symmetrical

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

Sol- Right skewed

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

Sol- Left Skewed

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

Sol- The data is normally distributed and kurtosis value is 0.

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

Sol- The distribution of the data has lighter tails and flatter peaks than the normal distribution.

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



\*What can we say about the distribution of the data?

Sol- Let’s assume above box plot is about age’s of the students in a school.

50% of the people are above 10 yrs old and remaining are less.

And students whose age is above 15 are approx 40%.

\*What is nature of skewness of the data?

Sol- Left skewed, median is greater than mean.

\*What will be the IQR (**Interquartile** **Range**) of the data (approximately)?

Sol- Approximately= -8

Q19) Comment on the below Boxplot visualizations?



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

Sol- 1) median of the two boxes are similar 260, outliers doesn’t exist and its not skewed.

By observing both the plots whisker’s level is high in box plot 2, mean and

Median are equal hence distribution is symmetrical.

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)
  2. P(MPG<40)

c. P (20<MPG<50)

Sol- By using filter command and installing the dplyr package into the ‘R’.

a) 0.6524908

b) 0.7294261

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans- Yes it is follows normal distribution based on the qqnorm and qqline and the boxplot calculation.

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

Dataset: wc-at.csv

Ans- Waist circumference and Adipose tissue follows the normal distribution all the points in a straight line.

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

Sol- Z score of 90% confidence interval is 1.65

Z score of 94% confidence interval is 1.55

Z score of 60% confidence interval is 0.85

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

Sol- For 95%= 1.96

For 96%= 2.5

For 99% = 2.47

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?

Ans- rcode 🡪 pt(tscore,df) = 0.3218

T= X-mu/(S/square root of N)