|  |  |
| --- | --- |
| 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) | Nominal |

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

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

Solution: P(X=2)= Probability of occurrence of 2 heads and 1 tail

= P(HHT) + P(HTH) + P(THH)

Where H= Heads

P= Tails

The Probability of occurrence of an event when a coil is tossed is ½

So Therfeore,

= [1÷2×1÷2×1÷2]+ [1÷2×1÷2×1÷2]+ [1÷2×1÷2×1÷2]

= 3÷8

= 0.375

Therefore the probability that two heads and one tail are obtained is 0.375

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

Solution: Total Possible outcome when two dice are rolled is 62=36

1. Sum equal to 0 is not possible as the sum always exceeds 0 when two dice are rolled.

Probability(Sum equal to 1) = 0÷36

= 0

Therefore the probability of sum equal to 1 when two dice are rolled is 0.

1. The outcomes are (1,1), (1,2), (1,3), (3,1), (2,2), (3,1)

The number of outcomes are 6

Probability (Sum is less than or equal to 4) = 6÷36

= 1÷ 6

= 0.1667

Therefore the probability of sum is less than or equal to 4 when two dice are rolled is 0.1667

1. The outcomes are (1,5), (2,4), (3,3), (4,2), (5,1)

The number of outcomes are 5

Probability (Sum is divisible by 2 and 3) = 5÷36

= 0.1388

Therefore the probability of sum is divisible by 2 and 3 when two dice are rolled is 0.1388

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?

Solution: The bag contains 2 red, 3 green and 2 blue balls.

Total number of balls = 2 + 3 + 2 = 7

Two balls can be drawn in 7C2 ways = 21 ways

Number of ways of drawing 2 balls such that none is blue

Number of ways of drawing 2 balls from 2 red and 3 green balls = 5C2 = 10 ways

Probability of drawing 2 balls such that none is blue = 10÷21

= 0.476

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

Solution: 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 × 0.12

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

= 3.090

= 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**

**Solution:**

The mean, median, mode, variance, standard deviation, range as displayed accordingly in the pdf named Question 7.pdf

Points: Mean=3.596563

Median=3.695

Mode= As the values are numerical it would be meaningless to calculate the mode and hence the python console returns that Not a number (NaN)

Variance=0.285881

Standard Deviation =0.534679

Range =2.170

Score : Mean= 3.217250

Median= 3.325

Mode= As the values are numerical it would be meaningless to calculate the mode and hence the python console returns that Not a number (NaN)

Variance= 0.957379

Standard Deviation = 0.978457

Range = 3.911

Weigh: Mean= 17.848750

Median= 17.710

Mode= As the values are numerical it would be meaningless to calculate the mode and hence the python console returns that Not a number (NaN)

Variance= 3.193166

Standard Deviation = 1.786943

Range = 8.400

Inference:  
The most appeared car brand is Volvo 142E and the least appeared one is AMC Javelin

**\*Pdf is attached**

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?

Solution:

The Expected value can be found by calculating the mean of the data given

Mean = (108 + 110 + 123 + 134 + 135 + 145, 167 + 187 +199) ÷ 9

= 145.33

**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**

Solution: The Skewness of the speed is -0.117510, that is the data is negatively skewed

The Skewness of the distance is 0.806895, and the data is positively

skewed.

The kurtosis value of speed is -0.508994 that is the curve is platykurtic.

The kurtosis value of distance is 0.405053 that is the curve is Leptokurtic

The code has been attached in pdf format

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





**Solution:**

Histogram :

The values are concentrated in the range 50-100 with the frequency 200

The least range of weight is around 400 with a frequency around 0-10

We can notice that the long tail is towards the right so it is right skewed.

Box Plot:

Median is less than mean. Right skewed and we have outlier on the upper side of boxplot and there is 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?

**Solution :**

Avg. weight of Adult in Mexico with 94% CI is

(198.738325292158, 201.261674707842)

Avg. weight of Adult in Mexico with 98% CI is

(198.43943840429978, 201.56056159570022)

Avg. weight of Adult in Mexico with 96% CI is

(198.62230334813333, 201.37769665186667)

The code has been attached in pdf format

**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?

Solution:

1)Mean=41

Median=40.5

Variance=25.5294

Standard Deviation=5.05266

2) Most of the students are score are in the range of 40-45.

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

Solution: The skewness will be symmetrical. That is the data is normally distributed.

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

Solution: The skewness is positive. Most of the data will be lying on the left side of the plot. Mean always tends to go towards the most skewed part.

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

Solution: The skewness will be negative. Most of the data will be lying on the right side of the plot.

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

Solution: Positive kurtosis value indicates that the distribution has heavier tails than the normal distribution.

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

Solution: 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?

Solution: The data is not equally distributed across the plane. There might be outliers influencing the data. Median of the data is approximately 14.

25% of the data lies between 0-10

50% of the data lies between 10-18

25% of the data lies after 18-20 Approximately

What is nature of skewness of the data?

Solution: The data will be left skewed since whisker length on the upper quadrant is higher than the data on the lower quadrant. Median will be greater than the mean since data is left skewed.

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

Solution:

Q1= 10

Q2=14.7

Q3=18

IQR=Q3-Q1=8

Therefore, the IQR is 8

Q19) Comment on the below Boxplot visualizations?



Solution:

In the box plots, plot 2) is highly distributed across the plane and plot 1) is slightly less distributed.

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

Solution:

First there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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)

Solution:

The probability for MPG > 38 is 0.34759

The probability for MPG < 40 is 0.72934

The probability for 20< MPG > 50 is 1.2430968797327491e-05

The code has been attached in pdf format

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Solution: After the analysis of the data we can conclude that the data is fairly

distributed as the skewness value is -0.177 and the kurtosis value is -0.611

The code has been attached in pdf format

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

Dataset: wc-at.csv

Solution:

Adipose Tissue (AT): From the distribution plot, Mean> Median, right whisker is larger than left whisker, data is positively skewed.

Waist: Mean> Median, both the whisker is of same length, median is slightly shifted towards left. Data is fairly symmetrically distributed.

The code has been attached in pdf format

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

Solution:

Z-score of 90% confidence interval is 1.6448

Z-score of 94% confidence interval is 1.880

Z-score of 60% confidence interval is 0.8416

The code has been attached in pdf format

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

Solution:

t scores of 95% confidence interval for sample size of 25 is 2.0638

t scores of 96% confidence interval for sample size of 25 is 2.1715

t scores of 99% confidence interval for sample size of 25 is 2.7969

The code has been attached in pdf format

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

Solution: The probability is 0.32

The code has been attached in pdf format