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

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

**Answer –**

When three coins are tossed the total number of possible combinations are HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.

The number of combinations which have two heads and one tail are:

HHT, HTH, TTH which makes them 3 in number.

Therefore the Probability of getting two heads and one tails in the toss of three coins simultaneously is defined as:

P(Two heads and One Tail) = Number of desired outcomes

= 3 = 0.375

The Probability of getting two heads and one tails in the toss of three coins simultaneously is 3/8 or 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**

**Answer –**

a) There is no outcomes which corresponds sum is equal to one. i.e.

0/36. Probability is 0.

b) (1,3) (2,2) (3,1) = 3 outcomes, 3/36 i.e. 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?**

**Answer –**

Given : Bag contains 2R, 3G and 2B Balls.

Total number of ball = 7

Number of ways picking up two balls out of 7 = 7x6/2x1 = 21

Therefore, number of picking up two balls out of (2+3) balls = 5x4/2x1 = 10

Number of ways picking up none of the ball is blue = number of picking up two balls out of (2+3) balls/ Number of ways picking up two balls out of 7

=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 –

Expected number of candies for a randomly selected child

= 1x0.015 + 4x0.20 + 3x0.65 + 5x0.005 + 6x0.01 + 2x0.120

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

= 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= “numeric”,

Variance= 0.2858814

, Standard deviation= 0.5346787.

Score:

Mean= 3.21725, Median= 3.325, Mode= “numeric”,

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= “numeric”,

Variance= 3.193166, Standard deviation= 1.786943

***Attached Question No 7.ipynb***

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?

The weights (X) of patients at a clinic (in pounds), are 108, 110, 123, 134, 135, 145, 167, 187, 199

one of the patients is chosen at random.

**To Find :** Expected Value

**Solution:**

Expected Value  =  ∑ ( probability  \* Value )

 ∑ P(x).E(x)

there are 9 patients

Probability of selecting each patient = 1/9

Ex  108, 110, 123, 134, 135, 145, 167, 187, 199

P(x)  1/9  1/9   1/9  1/9   1/9   1/9   1/9   1/9  1/9

Expected Value  =  (1/9)(108) + (1/9)110  + (1/9)123 + (1/9)134 + (1/9)135 + (1/9)145 + (1/9(167) + (1/9)187 + (1/9)199

= (1/9) ( 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9)  (  1308)

= 145.33

Expected Value of the Weight of that patient = 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**

**Answer –**

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.

***Attached Question 9.ipynb***

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



**Answer –**

The majority of the data points are distributed between 50 and 100 with a frequency of 200. The lowest weight range is 400, which is somewhere between 0 and 10. Consequently, the aforementioned distribution's expected value is 75. Skewness: Because of the large tail to the right, it is strongly skewed to the right.

**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 –**

Using the t-distribution, it is found that:

The 94% confidence interval is (198.73, 201.27).

The 96% confidence interval is (198.61, 201.39).

The 98% confidence interval is (198.43, 201.57).

***Attached Question 11.ipynb***

**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?

**Answer –**

Mean = 41.0

Median = 40.5

Variance = 25.52

Standard Deviation = 5.52

***Attached Question No. 12.ipynb***

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

The skewness might be positive, negative, or zero depending on the mean, median, and mode. If the mean, median, and mode are all equal, then the skewness is 0. The mean and median are equal and the distribution has no skewness if the distribution is symmetric.

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

To determine if your distribution is positively or negatively skewed, use the mean, median and mode. Positive skewness occurs when the mean exceeds the median in the distribution. The distribution is said to be negatively skewed if the mean is lower than the median.

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

When the median of a distribution is greater than the mean, negative skewness results. When the median is less than the mean, the distribution is said to be positively skewed.

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

Positive Kurtosis suggests that the likelihood of outliers is higher. Positive values of kurtosis suggest a peaked distribution with thick tails. A distribution where more of the numbers are distributed away from the mean and in the tails is said to have an extreme positive kurtosis.

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

In comparison to the normal distribution, the data distribution features flatter peaks and lighter tails. A negative kurtosis means that your distribution is flatter than a normal curve with the same mean and standard deviation.

The easiest way to visualise this is to plot a histogram with a fitted normal curve.

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



What can we say about the distribution of the data?

Assuming that the above data represents age of different students.

We can say that that 50% students are less than 10 year of age and the remaining students are more than 10 years of age. Also, from the above data we can conclude that approximately 40% of the students are more than 15 year of age.

What is nature of skewness of the data?

By looking at the above data we can conclude that the data is left skewed and median is greater than the mean.

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

By looking at the data, Inter-quartile range (approx.). = -8

Q19) Comment on the below Boxplot visualizations?



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

**Answer -** By comparing the two plots, we can see that the boxplot 2 distribution is symmetrical because the mean and median are equal.

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)
  3. P (20<MPG<50)

***Attached Question No. 20.ipynb***

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer (a)- From plot and values we can say that data is fairly symmetrical, i.e fairly normally distributed.

***Attached Question 21 - A.ipynb***

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

Dataset: wc-at.csv

Answer - mean> median, both the whisker are of same length, median is slightly shifted towards left. Data is fairly symmetrically distributed.

***Attached Question No. 21-B.ipynb***

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

Answer – Z score at 90% confidence interval - 1.6448536269514722

Z score at 94% confidence interval - 1.8807936081512509

Z score at 60% confidence interval - 0.8416212335729143

***Attached Question No. 22.ipynb***

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

**Answer -** Formula A = (1 + CL)/2 where A= area under the normal distribution curve, CL = confidence level

for 95% A = (1+0.95)/2 = 0.975. df = n-1 = 25-1=24

for 96% A = (1+0.96)/2 = 0.98. df = n-1 = 25-1=24

for 99% A = (1+0.99)/2 = 0.995. df = n-1 = 25-1=24

***Attached Question No. 23.ipynb***

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

***Attached Question No. 24.ipynb***