|  |  |
| --- | --- |
| 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 | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical |

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 | Nominal |
| Time Of Day | Ordinal |
| 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?

**Ans: 3/8**

Total possible outcome 2^3 =8

Required outcomes = 3

Then Probability of 2 HH & 1 T = 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

**Ans:**

Total no. of outcomes for this pair of dice is 36

a)Favorable outcomes Equal to 1 is **ZERO**  Because it should be 1+1

b) Favorable outcomes for <= 4 are [1,1],[1,2],[1,3], [2,2], [2,1] , [3,1] = 6

6/36= **1/6**

c) Favorable outcomes for Sum is divisible by 2 and 3

Favorable outcomes = (1 , 5) , (3 , 3) , (4 , 2) , (5 , 1) , (6 , 6),(6,12),(12,6) = 7

So **7/36.**

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?

**Ans:**

Total balls = 2+3+2 = 7 & now 2 balls can be drawn out of 7 is

7C2 = 7x6/ 2x1 = 21 similarly now 2 balls can be drawn but null of the blue ball is

5C2 = 5x4 /2x1 = 10 hence Probability that none of the balls drawn is blue is

**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:**

To calculate for Expected number of candies for a randomly selected child- multiply by candies count x probability & then apply sum function

(1⋅0.015)+(4⋅0.20)+(3⋅0.65)+(5⋅0.005)+(6⋅0.01)+(2⋅0.120)

Ans = **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**

**Ans:-**

('Mean: Points', 3.5965625) ('Score', 3.21725) ('Weigh', 17.84875)

('Median: Points', 3.6950000000000003) ('Score', 3.325) ('Weigh', 17.71)

('Mode: Points', 3.92) ('Score', 3.44) ('Weigh', 17.02)

('variance: Points', 0.2858813508064516) ('Score', 0.9573789677419355) ('Weigh', 3.1931661290322575)

('std\_deviation: Points', 0.5346787360709715) ('Score', 0.9784574429896967) ('Weigh', 1.786943236096843)

('range\_value: Points', 2.17) ('Score', 3.9110000000000005) ('Weigh', 8.399999999999999)

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-**

(108+110+123+134+135+145+167+187+199)/9=145.33

**hence 145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans-** Skewness: [-0.11395477 0.78248352]

Kurtosis: [-0.57714742 0.24801866]

**I**nferences –

1. **Skewness:** Skewness measures the asymmetry of the distribution. A skewness value of 0 indicates a symmetrical distribution where the tails on both sides of the mean are balanced. Positive skewness indicates that the tail on the right side of the distribution is longer or fatter than the left side, while negative skewness indicates the opposite—i.e., the tail on the left side of the distribution is longer or fatter than the right side. Skewness gives insight into the direction and degree of asymmetry in the distribution.

**So speed becomes left side skew & dist becomes right side skew.**

1. **Kurtosis:** Kurtosis measures the peakedness or flatness of a distribution. A kurtosis value of 3 is considered normal (mesokurtic), indicating a distribution with similar tails and peak compared to a normal distribution. Values greater than 3 indicate leptokurtic distributions, which have heavier tails and a sharper peak compared to the normal distribution, making it more peaked. Values less than 3 indicate platykurtic distributions, which have lighter tails and a flatter peak compared to the normal distribution, making it less peaked.

**Both becomes platykurtic distributions.**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans-**

Skewness: [ 1.58145368 -0.60330993]

Kurtosis: [2.72352149 0.81946588]

**SP becomes right side skew & WT left side skew**

**Both becomes platykurtic distributions.**

Code is written in attached .ipynb file

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



**Ans-**

* Chick Weight data is right skewed
* Max. chick weight is between 50-100
* 50% of the chick weight is between 50-150
* 100 chick weight observed at maximum time 200 times of frequency



**Ans-**

* Data is right skewed.
* Outlier are at upper side.

**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?

**Ans-**

94% Confidence Interval: (198.74, 201.26)

98% Confidence Interval: (198.44, 201.56)

96% Confidence Interval: (198.62, 201.38)

Code is written in attached .ipynb file

**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?

**Ans-**

Based on these statistics, we can infer the following about the student's marks:

* **Mean:** The average score obtained by the student is 41.
* **Median:** The middle score (or average of the two middle scores) is 40.5. This indicates that half of the scores are below 40.5 and half are above 40.5.
* **Variance:** The variance measures the spread of the scores around the mean. A higher variance indicates that the scores are more spread out from the mean. In this case, the variance is 24.11.
* **Standard Deviation:** The standard deviation is a measure of the dispersion or spread of the scores around the mean. A higher standard deviation indicates more variability in the scores. Here, the standard deviation is approximately 4.91.

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

**Ans-**

if the mean and median are equal, it means that there is no skewness in the data or the skewness is very close to zero. Skewness measures the asymmetry of the distribution. When the mean and median are equal, it indicates that the distribution is perfectly symmetrical, and there is an absence of skewness.

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

**Ans-**

When the mean is greater than the median, it indicates that the distribution is right-skewed or positively skewed. The tail of the distribution extends towards the right (larger values).

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

**Ans-**

When the median is greater than the mean, it indicates that the distribution is left-skewed or negatively skewed.The tail of the distribution extends towards the left (smaller values).

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

**Ans-**

A positive kurtosis value indicates that a distribution has heavier tails and a sharper peak compared to a normal distribution.**Positive Kurtosis (>3):** This is known as leptokurtic distribution. It means that the tails of the distribution are longer and fatter than the tails of a normal distribution, and the peak of the distribution is higher and sharper. In other words, the data have more extreme outliers or unusually high values, resulting in a higher frequency of observations in the tails.

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

**Ans-**

A negative kurtosis value indicates that a distribution has lighter tails and a flatter peak compared to a normal distribution.**Negative Kurtosis (<3):** This is known as platykurtic distribution. It means that the tails of the distribution are shorter and thinner than the tails of a normal distribution, and the peak of the distribution is lower and flatter. In other words, the data have fewer extreme outliers or unusually high values, resulting in a lower frequency of observations in the tails.

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



What can we say about the distribution of the data?

**Ans- This data is not normally distributed.**

What is nature of skewness of the data?

**Ans- Data has negative skewness.**

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

**Ans- 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.

**Ans-**

* There are no outliers.
* Both box plot shares the same median approximately in a range 275 to 250.
* They are normally distributed with zero.

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)

**Ans –**

Probability of MPG > 38: 0.4074074074074074

Probability of MPG < 40: 0.7530864197530864

Probability of 20 < MPG < 50: 0.8518518518518519

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans- MPG of Cars 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

**Ans – AT & Waist does not follow s Normal distribution.**

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

**Ans-**

90.0% Confidence Interval Z-score: 1.6448536269514722

94.0% Confidence Interval Z-score: 1.8807936081512509

60.0% Confidence Interval Z-score: 0.8416212335729143

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

**Ans-**

95% confidence interval t-score : 2.0638985616280205

96% confidence interval t-score : 2.1715446760080677

99% confidence interval t-score : 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

**Ans- 0.32167411684460556 = 32.16%**

***\*\*\*END of the Assignment 1\*\*\****