**ASSIGNMENT-1**

**Q1)** Identify the Data type for the Following:

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

**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 | Nominal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ (Intelligence Scale) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Nominal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| 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.** Three Coins are tossed i.e.,

Sample Space (S) = {HHH, TTT, HTT, THT, TTH, THH, HTH, HHT}

i.e., n(S) = 8

We have to find two heads and one tail,

So, Favourable Outcomes will be = {THH, HTH, HHT}

i.e., n(A) = 3

Hence required Probability is,

Probability = Number of favourable outcomes / number of possible outcomes or sample space

Hence, when three coins are tossed, the probability of two heads and one tail is 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.** Two dice are rolled i.e.,

Sample Space (S) = **{**(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)

(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)

(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)

(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)

(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)

(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)**}**

i.e., n(S) = 6\*6

=36

1. **Equal to 1**

The sum is equal to 1 is zero (0) because, the above table starts with (1,1)….likewise, other than in the dice we are not having zero.

1. **Less than or Equal to 4**

The sum is equal to 4 is possible i.e.,

The favourable outcomes are = {(1,3),(2,2),(3,1)}

i.e., n(b) = 3

Hence the required probability,

1. **Sum is divisible by 2 and 3**

The sum is divisible by 2 and 3 is possible and is less than 13. Here the total favourable outcomes will come i.e., n(c) = 36

Hence the required probability is,

**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 number of balls = (2+3+2)

Let S be the sample space

Then, n(S) = Number of ways of drawing 2 balls out of 7

= n!/[r!(n-r)!]

=

= 21

Now,

Let E = Event of drawing 2 balls, none of which is blue.

Since, n(E) = Number of ways of drawing 2 balls out of (2+3) balls.

=

=(2\*1)(5\*1)

=10

Hence,

P(E) = 0.47

**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 is

E(X) = \*P(X)

= 1\*0.015 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.01 + 2\*0.120

= 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.** See Jupyter file.

**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.** a) The weights (X) of patients at a clinic (in pounds) are

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

Expected Value [E(X)] = = \*P(X)

Since,

Probability of 108 patients is 1/9

Similarly, for all patients given here probability is 1/9

Now,

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

= 145.33

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

1. **Car’s speed and distance**

**Use Q9\_a.csv**

1. **SP and Weight (WT)**

**Use Q9\_b.csv**

**Ans.** See Jupyter File

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



**Ans.**

* The histograms peak has right skew and tail is on right.
* Mean>Median.
* We have outliers on the higher side.
* The boxplot has outliers on the maximum 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.** See Jupyter 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.**

a)

Mean = 41,

Median = 40.5

Variance = 25.52

Standard Deviation = 0.05

b) We don’t have outliers and the data is slightly skewed towards right because mean is greater than median.

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

**Ans.** No skewness is present we have a perfect symmetrical distribution.

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

**Ans.** Skewness and tail is towards right.

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

**Ans.** Skewness and tail is towards left.

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

**Ans.** Positive Kurtosis value indicates that the curve is more peaked and it is Leptokurtic.

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

**Ans.** Negative Kurtosis value indicates that the curve is flatter and broad and it is Platykurtic.

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



1. What can we say about the distribution of the data?
2. What is nature of skewness of the data?
3. What will be the IQR of the data (approximately)?

**Ans.**

a) The above boxplot is not normally distributed. The median is towards the higher value.

b) The data is skewed towards left. The whisker range of minimum value is greater than maximum.

c) The Inter Quartile Range = Q3 – Q1

= 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.**

* First there are no outliers.
* Second, both the boxplot shares the same median that is approximately in a range between 275 to 250 an 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)
  3. P (20<MPG<50)

**Ans. See Jupyter file.**

**Q 21)** Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

**Ans.**

a) MPG of cars follows normal distribution

b) Adipose Tissue (AT) and Waist Circumference (Waist) does not follow normal distribution.

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

**Ans.** Check Jupyter File

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

**Ans.** See Jupyter File.

**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.** See Jupyter File.

**Submitted by: - Somya Agrawal**