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

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

**Ans:** Trails = (2)^3 = 8

No. of (Two heads and one tail) = {(H,H,T),(H,T,H),(T,H,H)} = 3

P(Tow heads & one tail) = 3/8 = 0.375

% of P(Tow heads & one tail) = 37.50%

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:**



a) P(Equal to 1) = 0/36 = 0

b) P(Less than or equal to 4) = 6/36 = 1/6 = 1.6667

% of P(Less than or equal to 4) = 16.67%

c)P(Sum is divisible by 2 and 3) = 6/36 = 1/6 = 1.6667

% of P(Sum is divisible by 2 and 3) = 16.67%

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) = 7

Let N be the Total sample size.

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

= 7C2 = (7\*6) / (2\*1) = 21

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

n(E) = Number of away of drawing 2 balls out (2 + 3) balls

= 5C2

= (5\*4) / (2\*1) = 20

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 Value 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)

Expected Value of candies for a randomly selected child = 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

**Expected Value of candies for a randomly selected child = 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 : 1) Points:**

Mean = 3.5965

Median = 3.695

Mode = 3.07, 3.92 (Binomial Mode)

Variance = 0.2858

Standard Deviation = 2.17

Range = 2.17

**2) Score:**

Mean = 3.2172

Median = 3.325

Mode = 3.44

Variance = 0.9573

Standard Deviation = 0.9784

Range = 3.911

**3) Weigh:**

Mean = 17.8487 Standard Deviation = 1.7869

Median = 17.71 Range = 8.4

Mode = 17.02, 18.9 (Binomial Mode)

Variance = 3.1931

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

**Expected Value = 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**

**Ans: a) Car’s speed and distance**

Skewness: Kurtosis:

Speed: -0.11751 Speed: -0.5089

Distance: 0.8068 Distance: 0.40505

**b) SP and Weight(WT)**

Skewness: Kurtosis:

SP: 1.6114 SP: 2.9773

WT: -0.6147 WT: 0.9502

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



**Ans:** The most of the data points are connected in the range 50-100 with frequency 200 and least range of weight is 400 somewhere around 0-10. So, the expected value the above distribution is 75. Skewness – Right Skewed.

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?

**Ans:**

1. For 94%:

t\_value = 1.881

Confidence Interval = Upper Limit: 198.7376

Lower Limit: 201.2623

1. For 96%:

t\_value = 2.055

Confidence Interval = Upper Limit: 198.6214

Lower Limit: 201.3785

1. For 98%:

t\_value = 2.3282

Confidence Interval = Upper Limit: 198.4381

Lower Limit: 201.5618

**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:** 1) Mean: 41

Median: 40.5

Variance: 52.5294

Standard Variance: 5.05266

2) we can see that most of student’s marks lies between range 35 to 45.

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

**Ans:** Symmetrical

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

**Ans:** Right Skewed(Positively Skewed)

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

**Ans:** Left Skewed(Negatively Skewed)

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

**Ans:** It is a normal distribution and kurtosis value = 0

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

**Ans:** The distribution of the data has lighter tails and a 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?



What is nature of skewness of the data?

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

**Ans:**

1. Here, Median = 15, Mean = 14.

Consider, above boxplot shows ages of children in a family.

In this family, more than 50% children’s age is above 10 yrs and remaining are less. Children who’s age is above 12 approx. 22%

1. It is a Left skewed (Negatively skewed). Median is greater than mean.
2. Inter Quartile Range(IQR) = 3rd Quartile(Q3) – 1st Quartile(Q1) = 14 – 6 = 8

Q19) Comment on the below Boxplot visualizations?



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

**Ans:** Mean and median are equal so, this is symmetrical distribution. After, observation we can see that, whisker’s level is high in boxplot 2 as compare to boxplot 1.

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:** Probability of MPG of Cars for different cases are as follows:

1. P(MPG>38) : 0.348
2. P(MPG<40) : 0.7293
3. P(20<MPG<50) : 0.8989

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) Yes. MPG of Cars follows Normal Distribution.

1. Adipose Tissues(AT) and Waist from wc-at data set doesn’t follows Normal Distribution.

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

**Ans:** z scores of different confidence intervals are as follows,

1. z score of 90% confidence interval : 1.645
2. z score of 94% confidence interval : 1.881
3. z score of 60% confidence interval : 0.842

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

**Ans:** t scores of different confidence intervals for sample size of 25 are as follows,

1. t scores of 95% confidence interval : 2.064
2. t scores of 96% confidence interval : 2.172
3. t scores of 99% confidence interval : 2.797

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:**

Null Hypothesis: Average life of bulb is greater than or equal to 260 days.

Alternative Hypothesis: Average life of bulb is smaller than 260 days.

Population mean(pop\_mean) = 260

Sample Mean(sample\_mean) = 270

Sample size(n) = 18

Sample standard deviation(sample\_stdev) = 90

df = n – 1 = 18 – 1 =17

t-value = (sample\_mean – pop\_mean)/(sample\_stdev/sqrt(n))

= (270-260)/(90/sqrt(18))

‘t-value’ = 0.4714

p-value = 1 – stats.t.cdf(t,df)

= 1 – stats.t.cdf(0.4714,17)

‘p-value’ = 0.32167

The probability that 18 randomly selected bulbs would have an average life of no more than 260 days, if the CEO’s claim were true is approx. 0.3212 or 32%