|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Numerical (Discrete) |
| Results of rolling a dice | Numerical (Discrete) |
| Weight of a person | Numerical (Continuous) |
| Weight of Gold | Numerical (Continuous) |
| Distance between two places | Numerical (Continuous) |
| Length of a leaf | Numerical (Continuous) |
| Dog's weight | Numerical (Continuous) |
| Blue Color | Categorical (Nominal) |
| Number of kids | Numerical (Discrete) |
| Number of tickets in Indian railways | Numerical (Discrete) |
| Number of times married | Numerical (Discrete) |
| Gender (Male or Female) | Categorical (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 | Continuous |
| Weight | Continuous |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Continuous |
| Height | Continuous |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Continuous |
| Blood Group | Nominal |
| Time Of Day | Continuous |
| Time on a Clock with Hands | Continuous |
| Number of Children | Discrete |
| Religious Preference | Nominal |
| Barometer Pressure | Continuous |
| SAT Scores | Continuous |
| Years of Education | Discrete |

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

A3) Output for 3 tossed coins is

(H,H,H)(H,H,T)(H,T,H)(T,H,H)(T,T,T)(T,T,H)(T,H,T)(H,T,T,)

So probability for 2 Heads & 1 Tail

(H,H,T), (H,T,H), (T,H,H) **P(S) = 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

A4) (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)

Aa) Since sum of any two number of two dice will not give 1 so P(S) = 0

Ab) (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)

So P(S) = 6/36 = 1/6

Ac) (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)

So P(S) = 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?

A5) No blue ball = 5C2 = (5x4x3x2x1)/(2x1)(5-2)! = 10

2 Balls out = 7C2 = (7x6x5x4x3x2x1)/(2x1)(7-2)! = 21

So 10/21 ans.

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

A6) (1\*0.015)+(4\*0.2)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.12) = 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**

**A7) See attachment**

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 =(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

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

**Cars speed and distance**

**Use Q9\_a.csv**

A9a 1) Speed is negatively skewed that means more data points are at the left side

2) There is a possibility of presence of outlier for dist data as shown.

SP and Weight (WT)

**Use Q9\_b.csv**

A9b 1) SP is normally distributed as its kurtosis value is very close to 3 and WT tend to produce fewer and less extreme outliers than the normal distribution.

1. Both the data of SP & WT contains outliers.

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



1. Most chicks lies in weight ranges of 50-100.
2. Data is spread mostly to the right side
3. Outliers are present in the data.



**Ans:** 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?

A11) s=2000

P = 3000000

X = 200 pounds

SD =30 pounds

Z=(x-mu)/sd

Confidence interval 94% = 200 + 0.94 (30 / sqrt(2000)) = 200.6306

200 - 0.94 (30 / sqrt(2000)) = 199.3694

[199.3694,200.6306]

Confidence interval 98% = 200 + 0.98 (30 / sqrt(2000)) = 200.6574

200 - 0.98 (30 / sqrt(2000)) = 199.3426

[199.3426,200.6574]

Confidence interval 96% = 200 + 0.96 (30 / sqrt(2000)) = 200.644

200 - 0.96 (30 / sqrt(2000)) = 199.356

[199.356,200.644]

**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.

|  |  |  |
| --- | --- | --- |
| **X** | **X-x** | **sq(X-x)** |
| 34 | -7 | 49 |
| 36 | -5 | 25 |
| 36 | -5 | 25 |
| 38 | -3 | 9 |
| 38 | -3 | 9 |
| 39 | -2 | 4 |
| 39 | -2 | 4 |
| 40 | -1 | 1 |
| 40 | -1 | 1 |
| 41 | 0 | 0 |
| 41 | 0 | 0 |
| 41 | 0 | 0 |
| 41 | 0 | 0 |
| 42 | 1 | 1 |
| 42 | 1 | 1 |
| 45 | 4 | 16 |
| 49 | 8 | 64 |
| 56 | 15 | 225 |
|  | Total | 434 |
|  |  |  |
| **Mean (x)** | Sum(X)/n | 41 |
| **Variance (V)** | Sum (Sq(X-x))/n-1 | 25.529 |
| **Standard deviation** | sqrt(V) | 5.053 |

Median = 18/2 = 9th i.e. 40

1. What can we say about the student marks?

Most of the students got average marks between 34 and 45, which shows that the test was difficult and only 2 students have understood the concept compared to others.

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

**Ans:** Symmetrical distribution is seen when mean and median are equal.

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 means that it has a sharper peak and heavier tails compared to a normal distribution. This simply means that fewer data values are located near the mean and more data values are located on the tails.

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

A distribution with a 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?

**Ans:** The above Box plot is not normally distributed the median is towards the higher value.

What is nature of skewness of the data?

**Ans:** The data is skewed towards left. The whisker range of maximum value is greater than maximum.

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

**Ans:** The Inter Quartile range = Upper Quartile - Lower Quartile

= 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:**

1. There are no outliers in both the boxplot visualizations.
2. Both the box plot shares the same median.
3. Median for both is in between a range of 250 to 275 and 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)

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

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

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

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