|  |  |
| --- | --- |
| 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 | 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Interval |

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

A3) Here,

Total outcomes = HHH, HHT, HTH, THH, HTT, THT, TTH & TTT.

It has 3 outcomes containing 2 heads one tail = HHT, HTH & THH.

P(A) = 3/8

=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.

A4)

a) The probability of sum is Equal to 1 is 0.

b) The probability of a sum Less than or equal to 4 is 1/6.

c) The probability of sum divisible by 2 and 3 is 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) Here,

C(7,2) = 7!/(2!\*5!) = 21

Now, we will count the number of ways to draw 2 balls from the 5 non-blue balls, and we will divide that by the total number of ways.

C(5, 2) = 5! / (2! \* 3!) = 10

So, the probability is,

P(a)= 10 / 21

=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

A6) The expected candies are **3.09.**

=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 Comment about the values/ Draw some inferences.

**Use Q7.csv file**

**Ans**

* Mean, Median, Mode

**A screenshot of a computer

Description automatically generated with medium confidence**

* **Mode:**

Points = 3.07

Score = 3.44

Weigh = 17.02

* **Median**:

Points = 3.69

Score = 3.32

Weigh = 17.71

* **Mean:**

Points = 3.59

Score = 3.21

Weigh = 17.84

**Graphical user interface, text, application

Description automatically generatedText

Description automatically generated**

* **Range:**

Points = 2.17

Score = 3.91

Weigh = 8.39

* **Standard Deviation:**

Points = 0.53

Score = 0.97

Weigh = 1.78

* **Variance:**

Points = 0.28

Score = 0.95

Weigh = 3.19

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:**

The weight of the patient when one of the patients is chosen at random is 145.33 pounds.

Explanation:

Here, n is 9, then probability of choosing one patient is 1/9.

so,

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

p(X)=1/9\*(1308)

**p(X)=145.33**

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

**Cars speed and distance**

**Ans:**

**Text

Description automatically generated**

**Inference:**

1. The skewness of speed is negative, it signifies the data is negatively skewed.
2. Mean<median<mode.
3. The kurtosis of speed is negative, it signifies that the curve is flat with thin tails and data is platykurtic data.
4. The skewness of distance is positive, it signifies the data is positively skewed
5. Mean>median>mode.
6. The kurtosis of distance is positive, it signifies that the curve is peaked with thick tails and data is called as leptokurtic data.

**SP and Weight (WT)**

**Text

Description automatically generated**

**Inference**:

1. The skewness of SP is Positive, it signifies the data are positively skewed.
2. Mean>median>mode.
3. The kurtosis of SP is positive, it signifies that the curve is peaked with thick tail and data is leptokurtic data.
4. The skewness of weight is negative, it signifies the data is negatively skewed
5. Mean<median<mode.
6. The kurtosis of Weight is positive, it signifies that the curve is peaked with thick tails and data is called as leptokurtic data.

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

**Ans:**



**Inference:**

* The above histogram shows that the values are positively skewed or right skewed since a greater number of data is present on the left side.
* The skewness values are greater than zero.
* Thin tail present towards the right side.
* Mean>Median>Mode.



**Inference:**

* The boxplot shows that greater density of data is present towards the left side.
* The data is positively skewed.
* The data contains outliers in +ve direction, the dots represent the presence of outliers.

**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:**

Text

Description automatically generated

**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.

**Ans:**

Graphical user interface, application

Description automatically generated

**Mean = 41**

**Median = 40.5**

**Mode = 41**

**Variance = 25.52**

**Std Dev = 5.05**

2)What can we say about the student marks?

**Ans:**

* The average mark of students in test is 41.
* Most of the students mark between 35 to 45.
* The maximum and minimum marks are 56 and 34.

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

**Ans:** When mean & median of data are equal then there will be no skewness.

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

**Ans:** When mean>median then it’s nature will be Positive skewness.

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

**Ans:** When median > mean than Negative skewness.

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

**Ans:** The positive kurtosis (>3) indicate the distribution is peaked and has thick tails.

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

**Ans:** The negative kurtosis (<3) indicates the distribution is flat and has thin tails.

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



What can we say about the distribution of data?

**Ans:**

* Most of the values are on the left. So, it is left skewed.
* The distribution is not Normal.
* The median value for this is 15.
* The lower and upper quartiles are 10 & 18.

What is the nature of skewness of the data?

**Ans:** It is left skewed because most of the values are on the left side.

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

**Ans:** IQR = QR3-QR1

=18-10

IQR=10

Q19) Comment on the below Boxplot visualizations?



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

**Ans:** The given box plots are Normally distributed they have no outlier & both of them have median around 262.5.

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

**Ans:**

* 1. P(MPG>38) = 0.34
  2. P(MPG<40) = 0.27
  3. P (20<MPG<50) = 1.24

Graphical user interface, text, application

Description automatically generated

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Chart, histogram

Description automatically generated

Here, the Skewness of MPG is -0.17. So, it follows Normal Distribution.

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

Graphical user interface, application

Description automatically generated

Here, Skewness of Waist is 0.13. Hence it follows normal distribution.

Chart, histogram

Description automatically generated

Here, Skewness of AT is 0.57. Hence, it follows normal distribution.

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

**Ans:**

Graphical user interface, application

Description automatically generated

Z90 = 1.64

Z94 = 1.88

Z60 = 0.84

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

**Ans:**

Graphical user interface, text, application

Description automatically generated

t95= 2.06

t96 = 2.17

t99 = 2.79

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

Text

Description automatically generated