|  |  |
| --- | --- |
| 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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| 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 | Ordinal |

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

**Sol**: - The combinations are: -

H H H

H H T

H T H

T H H

T T H

T H T

H T T

T T T

When three coins are tossed the total number of possible combinations are = 8.

H H T

H T H

T T H

The number of combinations which have two heads and one tail are = 3

*The Probability of getting two heads and one tails in the toss of three coins simultaneously is =* **3/8** or **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

**Sol**: - The combinations are: -

The set of possible outcomes when we roll a die are {1, 2, 3, 4, 5, and 6}

(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)

1. **Equal to 1: -**

The minimum sum is 1+1 = 2. It’s impossible to get a sum of 1

Total favorable outcomes to get a sum is 1 when 2 dice are rolled = 0

1. **Less than or equal to 4: -**

When we roll two dice, the possibility of getting number less than or equal 4 is

(1, 1), (2, 1), (1, 3), (2, 1), (2, 2), (3, 1)

The number of favourable outcomes = 6

Total number of possibilities = 36

Probability = the number of favourable outcomes / Total number of possibilities = **6/36** => **1/6**.

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

When we roll two dice, the possibility of getting number Sum is divisible by 2 and 3 is

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

The number of favourable outcomes = 6

Total number of possibilities = 36

Probability = the number of favourable outcomes / Total number of possibilities = **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?

**Sol**: - Total number of balls = 7

The number of favourable outcomes = 10

Total number of possibilities = 21

Probability = the number of favourable outcomes / Total number of possibilities = **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

**Sol**: - Expected number of candies for a randomly selected child

= 1 \* 0.015 + 4 \* 0.20 + 3 \* 0.65 + 5 \* 0.005 + 6 \* 0.01 + 2 \* 0.12

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

= 3.090

= 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**

**Sol**: -

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weigh** |
| **Mean** | 3.596563 | 3.21725 | 17.84875 |
| **Median** | 3.695 | 3.325 | 17.71 |
| **Mode** | 3.07, 3.92 | 3.44 | 17.02, 18.90 |
| **Variance** | 0.2858814 | 0.957379 | 3.193166 |
| **Standard Deviation** | 0.5346787 | 0.9784574 | 1.786943 |
| **Range** | 2.76 - 4.93 | 1.513 - 5.424 | 14.5 - 22.9 |

Points have Positive skewness of **0.2788734**, and it has **0 outliers**, as in scatter plot Points value increasing the Score value is decreasing and plots are going towards down. Score Positive skewness of **0.4437855** and has **two outliers** on higher side. Weigh has Positive skewness of **0.3870456** and has **one outlier** on higher side.

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?

**Sol**: - There are 9 patients

Probability of selecting each patient = 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 =* **145.33**

**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**

**Sol**: -

1. Cars speed skewness has negative tail with Left skewness of **-0.1139548**, kurtosis of **2.422853** and Cars speed has no Outliers.

Distance skewness has positive tail with Right skewness of **0.7824835** and kurtosis of **3.248019** and distance have **one outlier** on higher side

1. Skewness of SP has positive tail with Right skewness **1.581454**, kurtosis of **5.723521** and SP has high outliers on higher side.

Skewness of WT has negative tail with Left skewness **-0.6033099**, kurtosis of **3.819466** and WT has outliers on both lower and higher side, lower side has more outliers.

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



**Sol**: - Above Histogram has more frequency at range between 50 to 100 and has Right tail with Positive skewness.  Mean > Median. We have outliers on the higher side.



**Sol**: - The boxplot has outliers on the higher side and mean is near to the lower extreme. Lower whisker is smaller than higher whisker.

**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 weight 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?

**Sol**: - The **94%** confidence interval is **198.73, 201.27**.

The **98%** confidence interval is **198.43, 201.57**.

The **96%** confidence interval is **198.61, 201.39**.

**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?

**Sol**: - (1)

Mean - 41

Median - 40.5

Variance - 25.5294

Standard deviation – 5.052

1. We have two outliers 49 and 56 and the data is skewed towards right because mean is greater than median.

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

**Sol**: - The mean is equal to the median, then the distribution is Symmetric and the distribution has **Zero skewness.**

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

**Sol**: - The mean of **positively skewed** data will be greater than the median and tail is towards **Right**.

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

**Sol**: - The median of **negatively skewed** data will be greater than the mean and tail is towards **Left**.

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

**Sol**: - Positive values of kurtosis indicate that distribution is **peaked and possesses thick tails**.

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

**Sol**: - 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?

What is nature of skewness of the data?

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

**Sol**: -

1. Boxplot is not normally distributed the median is towards the higher value.
2. The data is a skewed towards left. The whisker range of minimum value is greater than maximum.
3. The Inter Quantile Range = Upper quartile (Q3) – Lower Quartile (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.

**Sol**: - There are no outliers. Both the box plot shares the same median that is approximately in a range of 275 to 250 and 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)

**Sol**: - (a) P (MPG > 38) = **0.348**

(b) P (MPG < 40) = **0.729**

(c) P (20 < MPG < 50) = **0.013000**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Sol**: - MPG of cars follow 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

**Sol**: -Adipose Tissue (AT) follows Normal Distribution and Waist Circumference (Waist) does not follows Normal Distribution.

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

**Sol**: -The Z score at **90%** confidence interval is **1.645**

The Z score at **94%** confidence interval is **1.555**

The Z score at **60%** confidence interval is **0.253**

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

**Sol**: - The t scores of **95%** confidence interval is **2.0638985616280205**

The t scores of **96%** confidence interval is **2.1715446760080677**

The t scores of **96%** confidence interval is **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

**Sol**: - The Probability is **0.32167411684460556** or **32%**