|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete #As it is Countable |
| Results of rolling a dice | Discrete #As it is Countable |
| Weight of a person | Continuous #As it is Measurable |
| Weight of Gold | Continuous #As it is Measurable |
| Distance between two places | Continuous #As it is Measurable |
| Length of a leaf | Continuous #As it is Measurable |
| Dog's weight | Continuous #As it is Measurable |
| Blue Color | Nominal #As it can be Grouped |
| Number of kids | Discrete #As it is Countable |
| Number of tickets in Indian railways | Discrete #As it is Countable |
| Number of times married | Discrete #As it is Countable |
| Gender (Male or Female) | Nominal #As it can be Grouped |

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 #As it can be Grouped |
| High School Class Ranking | Ordinal #As it can be Ranked |
| Celsius Temperature | Ratio #As it is Measurable |
| Weight | Ratio #As it is Measurable |
| Hair Color | Nominal #As it can be Grouped |
| Socioeconomic Status | Ordinal #As it can be Ranked |
| Fahrenheit Temperature | Ratio #As it is Measurable |
| Height | Ratio #As it is Measurable |
| Type of living accommodation | Ordinal #As it can be Ranked |
| Level of Agreement | Ordinal #As it can be Ranked |
| IQ (Intelligence Scale) | Interval #As it can’t be Zero |
| Sales Figures | Ratio #As it is Mesrable & can’t be -ve |
| Blood Group | Nominal #As it can be Grouped |
| Time Of Day | Ordinal /Interval #As it can be Ranked |
| Time on a Clock with Hands | Ratio #As it is Measurable |
| Number of Children | Ratio #As it is Measurable |
| Religious Preference | Nominal #As it can be Grouped |
| Barometer Pressure | Ratio #As it is Measurable |
| SAT Scores | Ordinal #As it can be Ranked |
| Years of Education | Interval #As it can’t be Zero |

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

Ans

Events May happen: {HHH, HTT, HHT, HTH, TTT, THH, TTH, THT} =23 = 6 Two heads and one tail events: {HHT, HTH, THH} = 3

Probability of two heads and one tail is = Favorable Events/Total events = 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

Total Events

{(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)} = 62 = 36

1. Equal to 1: 0% Probability (minimum value can be zero)
2. Less than or equal to 4: 6/36 = 1/6
3. Sum is divisible by 2 and 3:

{(1,5),(2,4),(3,3),(4,3),(5,1),(6,6)}=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?

Ans:

Total balls = 2+3+2 = 7 Balls drawn at random = 2

Combinations formula = (n!/(n-r)!\*r!)

Total No.of ways 2 balls can be drawn from seven = 7C2 = (7!/(7-2)!\*2!) = 21 ways

No.of ways of drawing 2 balls so that none is blue = 5C2 = (5!/(5-2)!\*2!) =10 ways

Probability of drawing 2 balls such that none is blue = 5C2 / 7C2 = 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

Ans:

Expected Value Formula =  ∑ ( probability  \* Value )

Expected val = 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

= 3.09

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:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Mean** | **Median** | **Mode** | **Variance** | **Std. Dev** | **Range** |
| **Points** | 3.59 | 3.69 | 3.92 | 0.28 | 0.53 | 2.17 |
| **Score** | 3.21 | 3.32 | 3.44 | 0.95 | 0.97 | 3.91 |
| **Weight** | 17.84 | 17.71 | 17.02 | 3.19 | 1.78 | 8.39 |

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 Formula = ∑ (probability \* Value), Total Patients=9

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

**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:**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **Speed** | -0.113 | -0.577 |
| **Distance** | 0.782 | 0.248 |

* **”Speed” is Left Skewed & “Distance” is right Skewed**
* **“Speed” & “Distance” are < 3 so it’s Platykurtic Kurtosis**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **SP** | 1.581 | 2.723 |
| **WT** | -0.603 | 0.819 |

* **“SP” is Right Skewed & “WT” is Left Skewed**
* **“SP” & “WT” are < 3 so it’s Platykurtic Kurtosis**

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



* The data is Right Skewed
* Majority of Chick weighs in between 50-100



* Data has Outliers
* It is Positively Skewed

**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:

|  |  |  |  |
| --- | --- | --- | --- |
|  | 94% | 98% | 96% |
| Upper | 201.26 | 201.56 | 201.37 |
| Lower | 198.73 | 198.43 | 198.62 |

**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.

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.5 |
| Variance | 25.52 |
| Deviation | 5.05 |

1. What can we say about the student marks?

* Avg marks of students =41
* Majority of the students scored between 35 – 45 Marks

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

Ans: Skewness = 0. Perfectly symmetric bell shaped curve

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

Ans: Skewness = Positive (Right Skewed)

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

Ans: Skewness = Negative (Left Skewed)

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

Ans: \* Positive values of kurtosis indicate that distribution is peaked (or)

\* High peak at center of data

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

Ans: \* Negative kurtosis value is less peaked compared to normal distribution.

\* Wider peak in central part of data

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



What can we say about the distribution of the data?

Ans: It is not a Normal Distribution

What is nature of skewness of the data?

Ans: It is Left Skewed

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

Ans: The Interquartile range = Q3 - Q1 = 18 – 10 = 8

Q19) Comment on the below Boxplot visualization



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

Ans: \* Data is Normally Distributed

\* No outliers

\* Center lies around 262

\* The range of Fig 1 is less compared to Fig 2

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)

Ans: 0.4074074074074074

* 1. P(MPG<40)

Ans: 0.7530864197530864

c. P (20<MPG<50)

Ans: 0.8518518518518519

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: It follows 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

Ans: Adipose Tissue and Waist Circumference follows Normal Distribution

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

Ans: Z score for 90% confidence interval: 1.6448536269514722

Z score for 94% confidence interval: 1.8807936081512509

Z score for 60% confidence interval: 0.8416212335729143

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

Ans: T score for 95% confidence interval 2.06389856

T score for 96% confidence interval 2.17154468

T score for 99% confidence interval 2.7969395

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: population mean(mu)= 270, sample mean(x bar)=260, SD=90,

Sample size (n)=18, Degree of freedom (df) = n-1 = 18-1 = 17

T-test = (xbar – mu) / (SD/sqrt(n)) = -0.4714045207910317

Probability = 0.32167253567098364