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

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

**ANSWER**

**When 3 coins are tossed the total no. of possible combinations are**

**23 = 8.**

**Combinations are {HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}.**

**So, {HHT, HTH, TTH} are the 3 combinations having 2 heads & 1 tail.**

**Therefore, the probability = 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

**ANSWER**

**Sample Space (S) = {(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 = **0**
2. Less than or equal to 4 = **6/36 = 1/6 = 16.66**
3. Sum is divisible by 2 and 3 **= 6/36 = 1/6 = 16.66**

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?

**ANSWER Total Number of balls = (2+3+2) = 7**

**n(S) The number of ways of drawing 2 balls out of 7**

**= 7C2**

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

**= 21**

**n(E) Event of drawing 2 balls, none is blue (2+3)**

**= 5C2**

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

**= 10**

**Therefore, Probability**

**= n(E) / n(S)**

**= 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

**ANSWER 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**

**= 3.09**

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**

**ANSWER**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **POINTS** | **SCORE** | **WEIGHT** |
| **MEAN** | **3.596** | **3.217** | **17.848** |
| **MEDIAN** | **3.695** | **3.325** | **17.71** |
| **MODE** | **3.92** | **3.44** | **17.02** |
| **VARIANCE** | **0.285** | **0.957** | **3.193** |
| **STD. DEVIATION** | **0.534** | **0.978** | **1.786** |
| **MAX** | **4.93** | **5.424** | **22.9** |
| **MIN** | **2.76** | **1.513** | **14.5** |
| **RANGE** | **2.17** | **3.911** | **8.4** |

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?

**ANSWER**

**Expected Value**

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

**= 1308 / 9**

**= 145.3333**

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

**Car’s speed and distance**

**Use Q9\_a.csv ANSWER**

|  |  |  |
| --- | --- | --- |
|  | SPEED | DIST |
| SKEWNESS | **-0.117** | **0.806** |
| KURTOSIS | **-0.508** | **0.405** |

**SP and Weight (WT) Use Q9 b.csv ANSWER**

|  |  |  |
| --- | --- | --- |
|  | **SP** | **WT** |
| **SKEWNESS** | **1.611** | **-0.614** |
| **KURTOSIS** | **2.977** | **0.950** |

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



**ANSWER**

**The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.**



**ANSWER**

**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?

**ANSWER**

**For 94% confidence interval Range is = 198.73 – 201.26**

**For 98% confidence interval range is = 198.43 – 201.56**

**For 96% confidence interval range is = 198.62 – 201.37**

**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?

**ANSWER**

**#Mean = 41**

**#Median = 40.5**

**#Variance = 25.52**

**#Standard deviation = 5.05**

**MEAN > MEDIAN SO DATA IS SLIGHTLY SKEWED TOWARDS RIGHT.**

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

**ANSWER**

**NO SKEWNESS IS PRESENT WE HAVE A PERFECT SYMMETRICAL DISTRIBUTION.**

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

**ANSWER**

**SKEWNESS AND TAIL IS TOWARDS RIGHT.**

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

**ANSWER**

**SKEWNESS AND TAIL IS TOWARDS LEFT.**

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

**ANSWER**

**POSITIVE KURTOSIS MEANS THE CURVE IS MORE PEAKED & IT IS LEPTOKURTIC.**

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

**ANSWER**

**NEGATIVE KURTOSIS MEANS THE CURVE WILL BE FLATTER & BROADER.**

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)?

What can we say about the distribution of the data?

**ANSWER THE BOXPLOT IS NOT NORMALLY DISTRIBUTED AS THE MEDIAN IS TOWARDS THE HIGHER VALUE**.

What is nature of skewness of the data?

**ANSWER** **THE DATA IS SKEWED TOWARDS LEFT AS THE WHISKER RANGE OF MINIMUM VALUE IS GREATER THAN MAXIMUM.**

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

**ANSWER** **THE INTER QUARTILE RANGE = Q3 – 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.

**ANSWER**

**#1 NO OUTLIERS.**

**#2 THE BOXPLOT SHARES THE SAME MEDIAN WHICH RANGES BETWEEN 275-250.**

**#3 THEY ARE NORMALLY DISTRIBUTED WITH 0 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)

**ANSWER**

* 1. P(MPG>38) **= 0.348**
  2. P(MPG<40) **= 0.729**
  3. P(MPG>20) **= 0.943**
  4. P(MPG<50) **= 0.956**
  5. P (20<MPG<50) **= 0.013000000000000012**

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

**ANSWER(a)** **FOLLOWS NORMAL DISTRIBUTION.**

**ANSWER(b)** **DOES NOT** **FOLLOWS NORMAL DISTRIBUTION.**

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

**ANSWER**

|  |  |
| --- | --- |
| CONFIDENCE INTERVAL | Z-SCORES |
| **90%** | **1.644** |
| **94%** | **1.880** |
| **60%** | **0.841** |

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

**ANSWER**

|  |  |
| --- | --- |
| CONFIDENCE INTERVAL | T-SCORES |
| **95%** | **2.063** |
| **96%** | **2.171** |
| **99%** | **2.796** |

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

**ANSWER**

# Population mean = 270 days

# Sample mean = 260 days

# Sample Std Dev = 90days

# Sample(n) = 18 bulbs

# Degree of freedom = n - 1 = 18-1 = 17

# T-SCORE = -0.471

# 0.32%