|  |  |
| --- | --- |
| 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** |

**Name : Shreyas Bendsure**

**Batch : Weekdays 1pm to pm Pune**

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 | **Discrete – Ordinal** |
| Level of Agreement | **Discrete – Interval** |
| IQ(Intelligence Scale) | **Discrete – Interval** |
| Sales Figures | **Discrete – Interval** |
| Blood Group | **Discrete – 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 – Ratio** |
| SAT Scores | **Continuous – Ratio** |
| Years of Education | **Discrete - Nominal** |

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

**Sol : P (HTH) + P(HHT) + P(THH)**

* **(1/2 \*1/2\*1/2) + (1/2 \*1/2\*1/2) + (1/2 \*1/2\*1/2)**
* **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

**Sol : a) Zero**

**b) 3/36 = 1/12**

**c) 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 : 7 C 2 = (7\*6)/(2\*1) = 21**

**5 C 2 = (5\*4)/(2\*1) = 10**

**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 : 1\*0.015 + 4\* 0.20 + 3 \* 0.65 + 5 \* 0.005 + 6 \* 0.01 +2 \* 0.120**

* **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.

**Sol :**

|  |  |  |  |
| --- | --- | --- | --- |
| **Values** | **Points** | **Score** | **Weigh** |
| **Mean** | **3.59** | **3.21** | **17.84** |
| **Median** | **3.69** | **3.32** | **17.71** |
| **Mode** | **3.92** | **3.44** | **17.02** |
| **Minimum** | **2.76** | **1.51** | **14.5** |
| **Maximum** | **4.93** | **5.42** | **22.9** |
| **Range** | **2.17** | **3.91** | **8.4** |
| **Variance** | **0.28** | **0.95** | **3.19** |
| **Std. Deviation** | **0.53** | **0.97** | **1.78** |

**Use Q7.csv file**

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 : Mean = 1308**

**Expected value -> xP = (X-x)**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Sol:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Speed** | **Distance** |
| **Skewness** | **-0.1175** | **0.8068** |
| **Kurtosis** | **-0.5089** | **0.4050** |

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Sol:**

|  |  |  |
| --- | --- | --- |
| **Name** | **SP** | **WT** |
| **Skewness** | **1.6114** | **-0.6147** |
| **Kurtosis** | **2.9773** | **0.9502** |

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



**Sol :**

**The most of the data points are concentrated in the range of 50 to 100 with frequency of 200 and the least concentrated range of weight is 400 with frequency less than 15**

**By seeing the Histogram, we can clearly see a tail towards the right side so it is right 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?

**Sol : X +/- (Z1 – Alpha . sigma/sqrt (n))**

**Degrees of freedom = 2000 – 1 = 1999**

**(1- sigma/2) = 1- 0.03 = 0.97**

**Confidence interval for 94% is 1.882**

**Confidence interval for 96% is 2.05**

**Confidence interval for 98% is 2.33**

**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 : Mean = 41**

**Median = 40**

**Variance = 24.11**

**Std. deviation = 4.910**

**The average marks scored by all the students is 41**

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

**Sol : Symmetrical**

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

**Sol : right skewed**

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

**Sol : left skewed**

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

**Sol : the data is normally distributed**

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

**Sol : the distribution of data has light tails**

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



What can we say about the distribution of the data?

**Sol : Above 15’ the data is approx. 40%**

What is nature of skewness of the data?

**Sol : Left skewed**

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

**Sol : appr0x = -8**

Q19) Comment on the below Boxplot visualizations?



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

**Sol : Mean and median of both the plots is equal so the distribution is symmetrical and the whiskers level is high in boxplot 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)
  2. P(MPG<40)

c. P (20<MPG<50)

**Sol : there are 33 data points in MPG which are greater than 38**

**And 61 data points in Mpg which are less than 40**

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

**Sol : Z score of 90% confidence interval is 1.65**

**Z score of 94% confidence interval is 1.55**

**Z score of 60% confidence interval is 0.85**

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

**Sol : T score for**

**95% = 1.96**

**96% = 2.5**

**99% 2.47**

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