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

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

**ANS : 3/8 ( HHH,HHT,HTH,HTT,THH,THT,TTH,TTT** **)**

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1 = **0**
2. Less than or equal to 4 = **1/6**
3. Sum is divisible by 2 and 3 = **2 / 3**

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: 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 number of candies for a randomly selected child is 3.090

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.

**ANSWER : MEAN AND MEDIAN**

|  |
| --- |
| Here is the dataset Q7    1) Mean(Points)=3.59 1) Median(Points)==3.69    2) Mean(Score)==3.21 2) Median(Score)===3.32  3) Mean(Weigh)==17.84 3) Median(Weigh)===17.71 |
|  |
| |  | | --- | |  | |
| |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | **VARIANCE, STANDARD DEVIATION AND RANGE**   |  | | --- | | Var (Points) = 0.285  Sd (Points) = 0.534  Range (Points) = 2.76 - 4.93  Var (Score) = 0.95  Sd (Score) = 0.978  Range (Score) = 1.53 - 5.42  Var (Weigh) = 3.19  Sd (Weigh) = 1.78  Range (Weigh) = 14.5 - 22.9  **MODE**  Mode (Points) = 4.93  Mode (Score) = 5.42  Mode(Weigh) = 22.9 | |  | |  | | |  | | --- | |  | |  | | | |

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 of the weight of that patient is 145.333

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

**Cars speed and distance**

**SP and Weight(WT)**

**ANS - 9A**

|  |
| --- |
| Skewness – Speed = -0.11395  Kurtosis -- Speed = 2.42285  Skewness -- Dist = 0.78248  Kurtosis -- Dist = 3.2480  **ANS - 9B**  Kurtosis -- SP = 5.7235  Skewness --SP = 1.5814  Kurtosis -- WT = 3.8194  Skewness – WT =-0.603309 |
|  |
| |  | | --- | |  | |

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

**Answer-** Here “ChickWeight” of the data is distributed more in left side. We can say it resembles Gaussian Distribution & it is positively skewed. And there is a peak in a distribution so it is positive kurtosis.



**Answer -** Here the data is not symmetric, the data distributed wider in in the upper quartile (i. e 75%) and the longer part of the data lies in right(upper) so the data is skewed right. It consists of outliers too. We can see the outliers as the points lying above 100th Quartile

**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 :**  Z (94%) = 198.74 - 201.25 (ME=1.26)

Z (96%) = 198.62 - 201.37 (ME=1.38)

Z (98%) = 198.44 - 201.55 (ME = 1.56)

**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?

**ANS:**

**Mean = 41**

**Median = 40.5**

**Variance = 24.1**

**S.D = 4.910**

Data is symmetric as mean is almost equal to median & the average marks obtained by students = 40

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

ANS : Symmetric.

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

ANS : Positively Skewed.

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

ANS : Negatively skewed.

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

ANS : Distribution has sharper peak and wide tail.

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

ANS: Distribution has Flatter peak and lighter tail.

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



What can we say about the distribution of the data?

**ANS** =The data distribution is wider in the first 25% quartile.

What is nature of skewness of the data?

**ANS** =Skewed left.

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

ANS= **IQR=8**

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect 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) = **0.3475**
  2. P(MPG<40) = **0.7294**

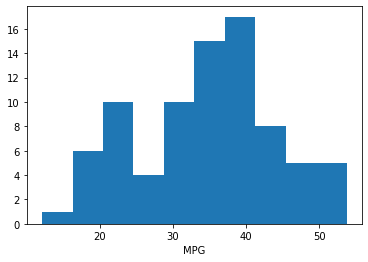
c. P (20<MPG<50) = **0.899**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

**ANS: Data follows normal distribution.**

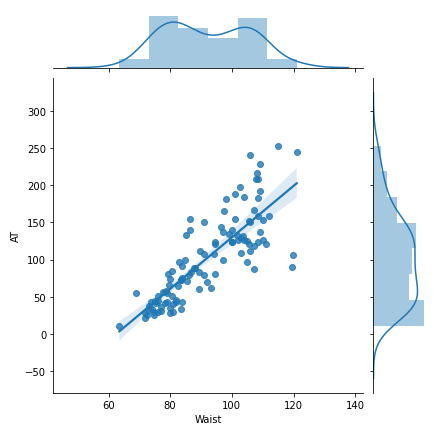
Dataset: Cars.csv



|  |
| --- |
|  |
|  |
| |  | | --- | |  | |

b)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

**ANS:**

1)Z ( 90% )=**1.64**

2)Z ( 94% )=**1.880**

3)Z ( 60% )=**0.841**

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

**ANS:**

1)qt(95%)=**2.06**

2)qt(96%)=**2.17**

3)qt(99%)=**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

**ANS: Required probability is 0.32=32%.**