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

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

Answer:0.375

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

1. Equal to 1 The Answer:0(Not Possible)
2. Less than or equal to 4

Answer: 6/36=0.166

1. Sum is divisible by 2 and 3: 6/36=0.166

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: The probability is 10/21=0.47

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 for the above question: the expected value is 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.



**Answer:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **points** | **Score** | **Weigh** |
| **mean** | **3.596** | **3.217** | **17.84** |
| **median** | **3.695** | **3.325** | **17.710** |
| **mode** | **3.07** | **3.44** | **17.02** |
| **variance** | **0.285** | **0.9573** | **3.193** |
| **Standard deviation** | **0.534** | **0.9784** | **1.7869** |
| **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 for above question:145.333

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

**Cars speed and distance**

****

**Answer:**

**Skewness for Speed and Distance:**

**for speed =-0.1175 skewness value is negative so it is left skewed and data concerned on right side.**

**for distance=0.8068 skewness value is positive so it is right skewed and data concerned left side.**

**Kurtosis for Speed and Distance:**

**For speed =-0.5089 kurtosis value is negative which means it has flatter peaks and thinner tails**

**For distance=0.4050 kurtosis value is positive which means it has sharper peaks and heavier tails**

**SP and Weight(WT)**



**Skewness for SP and Weight:**

**For SP: 1.611 value is positive so it is right skewed with data concerned left side**

**For WT: -0.6147 value is negative so it is left skewed with data concerned on right side**

**Kurtosis for SP and Weight:**

**For SP:2.977 value is positive means sharper peaks and heavier tails**

**For WT:0.950 value is positive means sharper peaks and heavier tails**

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



The nature of histogram

Answer: It is a right skewed histogram with data concentrated on left side



Answer: The above histogram has lot of outliers towards upper extreme

**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=198.74-201.26

For 98 % confidence interval =198.44-201.56

For 96 % confidence interval =198.62-201.38

**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.

Answer: mean=41, median=40, variance:25.52, standard deviation:5.052

1. What can we say about the student marks?

Answer: mean>median so it means that data is skewed towards right

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

Answer: skewness=0 skewness is symmetrical

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

Answer: the nature is right skewed

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

Answer: the nature is left skewed

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

Answer: It indicates that a distribution is peaked and possess thick tails

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

Answer: It indicates that the distribution is flatter and has thick tails

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



What can we say about the distribution of the data?

Answer: It is not a Normal Distribution

It has median value of 15

It has lower quartile value of 10

It has higher quartile value of 18

What is nature of skewness of the data?

Answer: The nature is left skewed

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

Answer: 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:

Mean and Median are equal hence distribution is symmetrical

The median of both the boxplots are 260 approximately and no outliers are present

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

Answer:

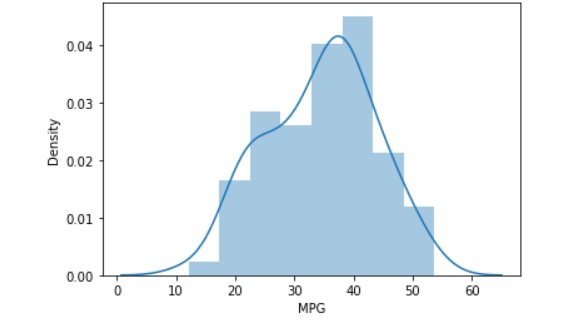
* 1. P(MPG>38) =0.34
  2. P(MPG<40) =0.72

c. P (20<MPG<50) =1.243

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

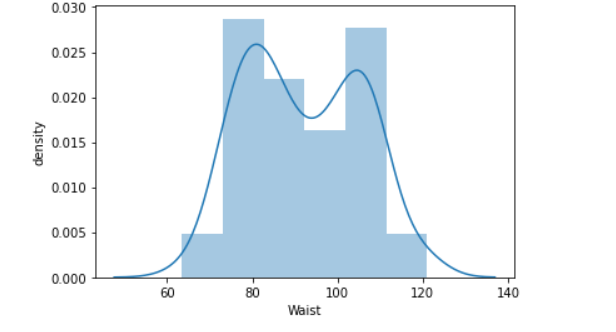
Dataset: Cars.csv



Answer: The Data Does Not Follow Normal Distribution as indicated by the above plot

1. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv



Answer: The Data does not follow normal distribution as indicated by the above plot

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

Answer:

Z Score of 90 % confidence Interval is 1.65

Z Score of 94% confidence Interval is 1.88

Z Score of 60% confidence Interval is 0.84

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

Answer:

For 95%=2.063899

For 96%= 2.171545

For 99%=2.7969

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: Required probability:0.32=32%