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

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

## Ans: Total outcomes when three coins are tossed

## = (HHH, TTT, HTT, THT, TTH, THH, HTH, HHT)

## 

## We have to find the probability of getting Two Heads and one Tail,

## So the favourable outcomes are= THH, HTH, HHT

## Probability of getting two heads and one tail P (E)

## =Number of Favourable outcomes/Total number of outcomes

## P (E)= 3/8

## Hence when three coins are tossed, probability of getting two heads and

## one tail is 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: The Sample space S of two dice is=

{(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)

S = (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)}

n (S) =36

1. Favorable outcomes ( Having sum = 1) = 0

P (A) = Favorable outcomes / Total number of samples

P (A) = n (A) / n (S)

P (A) = 0 / 36

P (A) = 0

1. Favorable outcomes = {(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}

P (B) = Favorable outcomes / Total number of outcomes

P (B) = n (B) / n (S)

P (B) = 6 / 36

P (B) = 1 /36

So, the probability of getting sum less than or equal to 4 is 1 / 36

1. Favorable outcomes = {(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)}

P (C) = Favorable outcomes / Total number of outcomes

P (C) = n (C) / n(S)

P (C) = 6 / 36

P (C) = 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 number of balls = 2(red) + 3(green) + 2(blue)

Total number of outcomes = 7C2 = 7\*6 / 2\*1 = 42 / 2 = 21

Favorable outcomes = 5C2 = 5\*4 / 2\*1 = 20 / 2 = 10

P (A) = Favorable outcomes / Total number of outcomes

P (A) = 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: ∑ (Candies\*Probability)

= (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.

**Use Q7.csv file**

**Ans:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Mean** | **Median** | **Mode** | **Standard**  **deviation** | **Variance** | **Range** |
| **Point** | 3.596 | 3.695 | 3.07 | 0.534 | 0.2858 | 2.17 |
| **Weigh** | 3.217 | 3.325 | 17.02 | 0.978 | 0.9573 | 8.399 |
| **Score** | 17.84 | 17.10 | 3.44 | 1.786 | 3.1931 | 3.911 |

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: The weight of the patient is **145.33333333333334**

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

**Cars speed and distance**

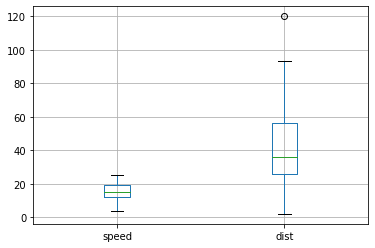
**Use Q9\_a.csv**

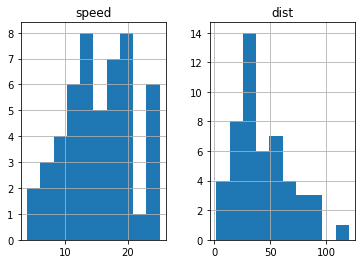
**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:**

|  |  |  |
| --- | --- | --- |
|  | **skew** | **kurtosis** |
| **Speed** | -0.117 | -0.508 |
| **distance** | 0.405 | 0.806 |

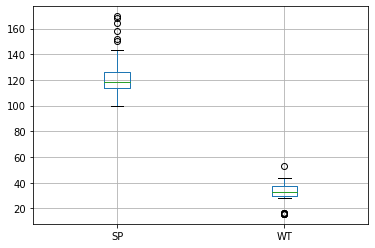


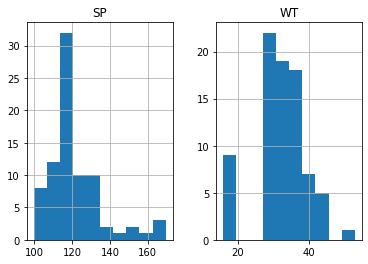


**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | **skew** | **kurtosis** |
| **speed** | 1.611 | 2.977 |
| **distance** | -0.614 | 0.950 |





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



Ans: The Histogram data has peak on right side and tail on left side, Mean > Median. So outliers have on the higher side



Ans: The box plot has outlier maximum side only.

**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:

|  |  |
| --- | --- |
| **Confidence interval** | **Average weight** |
| 94% | (198.738325292158,201.261674707842) |
| 98% | (198.43943840429978,201.56056159570022) |
| 96% | (198.62230334813333,201.37769665186667) |

**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.

Ans:

|  |  |
| --- | --- |
| **Mean** | 41.0 |
| **Median** | 40.05 |
| **Standard deviation** | 4.91 |
| **Variance** | 24.11 |

1. What can we say about the student marks?

Ans: Here we didn’t have outliers and the data is slightly towards right because mean is greater than median.

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

Ans: When mean is equal median of data set, the distribution is normalized and zero skewness

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

Ans: When mean is greater than median of data set, the distribution is positively skewed and tail is expanded to right side.

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

Ans: When median is greater than mean of data set, the distribution is negatively skewed and tail is expanded to left side.

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

Ans: A positive kurtosis value indicates that the data has heavier tails and a more peaked or concentrated distribution compared to a normal distribution.

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

Ans: A negative kurtosis value indicates that the data lighter tails and a flatter or more dispersed distribution compared to a normal distribution.

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



What can we say about the distribution of the data?

Ans: The box plot is not following normal distribution

What is nature of skewness of the data?

Ans: The nature of skewness of data is left side skewed.

What will be the IQR of the data (approximately)?   
Ans: Inter Quartie Range => IQR

Q1 = 10 => Lower whisker

Q3 = 18 => Upper whisker

IQR= Q3-Q1 => 18-10

IQR = 8

Q19) Comment on the below Boxplot visualizations?



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

Ans: 1) Boxplot 1 and Boxplot 2 “Median” is equal.

2) Boxplot 1 and Boxplot 2 are “symmetric”

3) They don’t have any outliers.

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)

Ans:

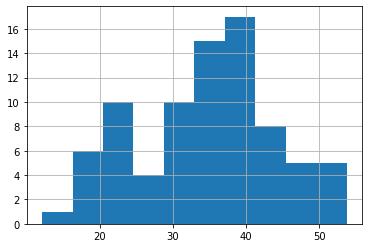
|  |  |
| --- | --- |
| **Probability** | **Probability of MPG** |
| P(MPG>38) | 0.34759392515827137 |
| P(MPG<40) | 0.7293498762151609 |
| P(20<MPG<50) | 0.8988689169682047 |

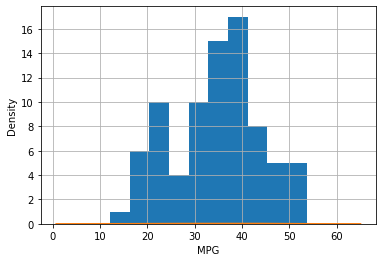
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: It is following Normal distribution.

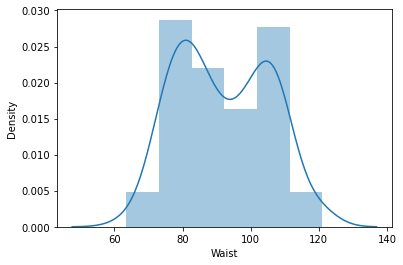


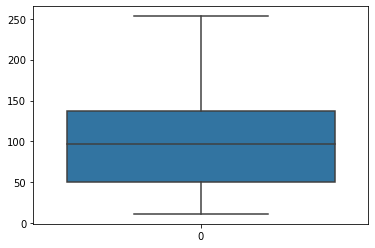


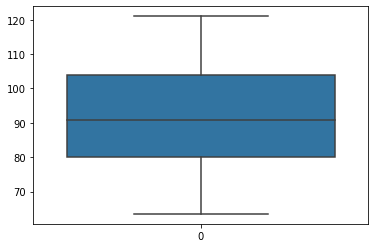
b) Check whether the Adipose Tissue(AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Ans:







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

Ans:

|  |  |
| --- | --- |
| **Confidence Interval** | **Z Score value** |
| 90% | 1.6448536269514722 |
| 94% | 1.8807936081512509 |
| 60% | 0.8416212335729143 |

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

Ans:

|  |  |
| --- | --- |
| **Confidence Interval** | **T Score value** |
| 95% | 1.973994288847133 |
| 96% | 2.1715446760080677 |
| 99% | 2.796939504772804 |

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: µ =270, =260, SD = 90, n = 18, df = n-1 = 18-1 = 17

Probability is 32%