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

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

Answer: Total no. of chances = 2^3 = 8

Possible outcomes=[HHH,HHT,HTH,THH,HTT,TTH,THT,TTT]

Probability of 2 heads and one tail = 3/8 = 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:

a) 0

b) Possible outcomes = [(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)]

Total no. of outcomes = 36

Probability [sum less than or equal to 4] = 6/36 = 1/6 = 0.167

1. Possible outcomes = [ (1,5), (2,4),(3,3),(4,2),(5,1),(6,6)]

Total no. of outcomes = 36

Probability [sum is divisible by 2 and 3] = 6/36 = 0.167

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 no. of balls = 2+3+2 = 7

Total no. of outcomes = 7C2 = (7\*6)/2 = 21

Expected no. of outcomes = 5C2 = 5!/2!\*3! = 10

Probability that none of the balls drawn is blue = 5C2 / 7C2 = 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:

Sum of probabilities = 0.015 + 0.20 + 0.65 + 0.005 + 0.01 + 0.120 = 1

Expected no.of candies for a randomly selected child = 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

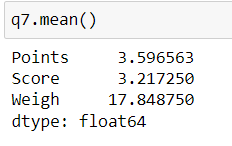
* For Points,Score,Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

Answer:

1. Mean =



1. Median =

Text, table

Description automatically generated

1. Mode =

Points : It is multimoded – 3.07 & 3.92

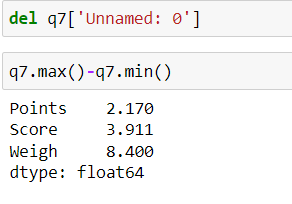
Score : 3.44

Weigh : It his multimoded –

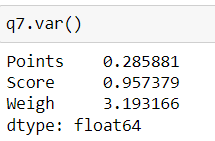
Table

Description automatically generated

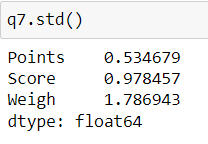
1. Range =

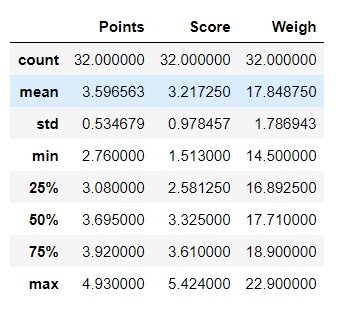


1. Variance =



1. Standard Deviation =





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:

**Probability Distribution Table :**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| E(x) | 108 | 110 | 123 | 134 | 135 | 145 | 167 | 187 | 199 |
| P(x) | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 |

Expected value of the weight of a randomly chosen patient = 108/9 + 110/9 + 123/9 + 134/9 + 135/9 + 145/9 + 167/9 + 187/9 + 199/9 = 1/9\*(1308) = 145.34

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

**Cars speed and distance**

**Use Q9\_a.csv**

Answer:

1. Skewness of speed = -0.114

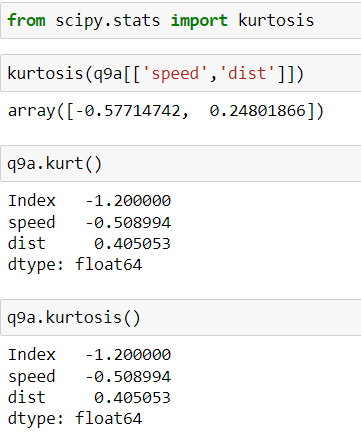
Skewness of disp = 0.782

Graphical user interface, text, application, email

Description automatically generated

1. Kurtosis of speed = -0.577

Kurtosis of dist = 0.25



Inference :

* Speed is negative skewness value hence it is Left / Negatively skewed.
* Distance is positive skewness value hence it is Right / Positively skewed.
* Speed has negative kurtosis hence it is flatter than a normal curve with the same mean and standard deviation.
* Distance has kurtosis positive value near to zero. So it has less outliers.

Chart, histogram

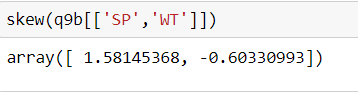
Description automatically generated

**SP and Weight(WT)**

**Use Q9\_b.csv**

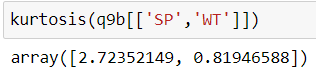
1. Skewness of SP = 1.581

Skewness of Weight = -0.603

****

1. Kurtosis of SP = 2.723

Kurtosis of Weight = 0.82

****

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



Answer:

Inference of Histogram : This graph is positively/right skewed. It has its mean>median>mode, hence majority of chick weight is less than mean value.



Inference of Boxplot: This boxplot is positively 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?

Answer:

Total Population (N) = 30,00,000

No. of samples (n) = 2000

Sample mean = 200 pounds

Standard deviation from mean of the random sample = 30 pounds

Since we don’t have population parameters, we can go with t-test

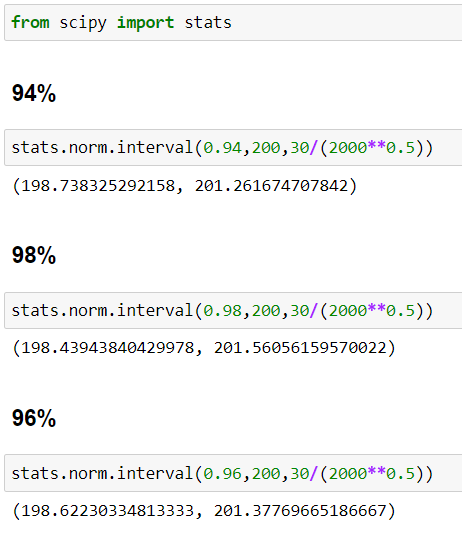
The **interval** is:



94% : 200 +/- (1.881 x 0.671) = [198.74 201.261]

98% : 200 +/- (2.33 x 0.671) = [198.437 201.563]

96% : 200 +/- (2.054 x 0.671) = [198.617 201.383]



**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?

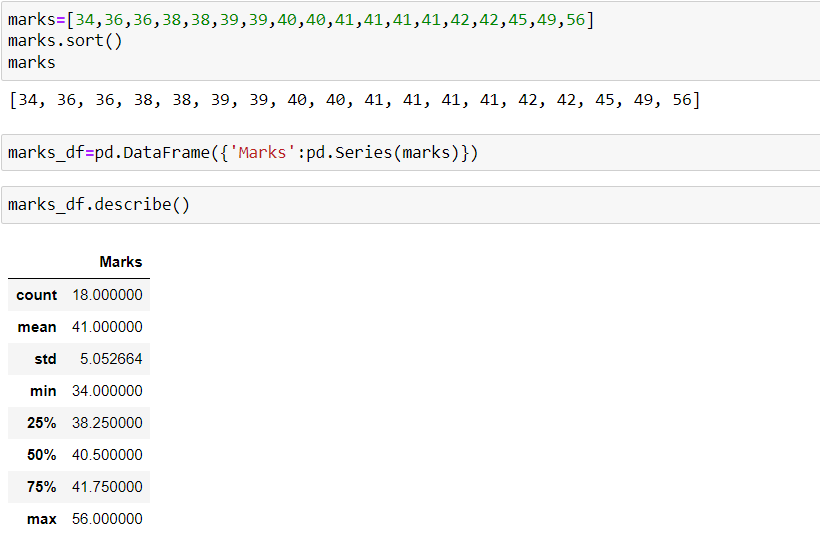
Answer:

1. Mean = 41

Median = 40.5

Standard deviation = 5.052664

Variance = 25.52942



Graphical user interface, text, application

Description automatically generated

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

Answer: Skewness will be zero when mean=median. It is the case of normal distribution.

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

Answer: Nature will be Right/Positively skewed.

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

Answer: Nature will be Left/Negatively skewed.

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

Answer: Positive kurtosis value greater than 3 means the data has many outliers and the distribution will have a steep curve. Its called Leptokurtic.

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

Answer: Negative kurtosis value means the data has very less outliers and the distribution will have a flatter curve. It is called Platykurtic.

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



What can we say about the distribution of the data?

Answer: The data are not normally distributed

What is nature of skewness of the data?

Answer: Negatively/left skewed nature where mean<median

What will be the IQR of the data (approximately)?   
  
  
Answer: 18 - 10 = 8 (approx).

Q19) Comment on the below Boxplot visualizations?



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

Answer: The box plot 1 designed with range = 3 , The second one range is = 1.5

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:

Graphical user interface, text, application, email

Description automatically generated

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Chart, histogram

Description automatically generated

The MPG data is not normally distributed.

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

Dataset: wc-at.csv

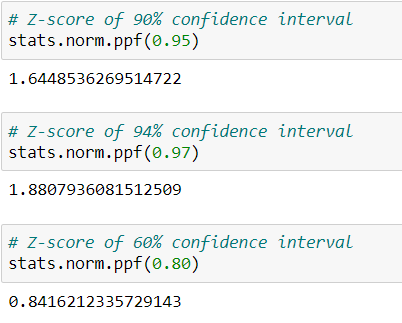
Chart, line chart

Description automatically generated

The Waist circumference and Adipose tissue does not follow normal distribution.

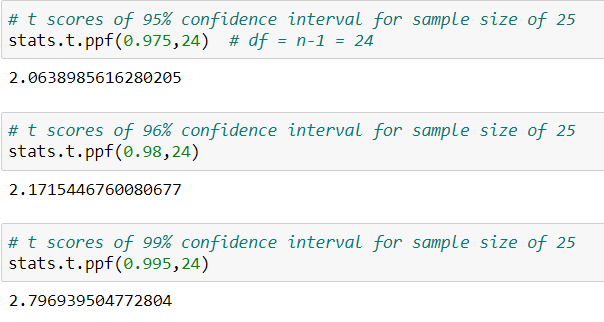
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

Answer:



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

Answer:



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:

