|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | DISCRETE DATA |
| Results of rolling a dice | DISCRETE DATA |
| Weight of a person | CONTINUOUS DATA |
| Weight of Gold | CONTINUOUS DATA |
| Distance between two places | CONTINUOUS DATA |
| Length of a leaf | CONTINUOUS DATA |
| Dog's weight | CONTINUOUS DATA |
| Blue Color | QUALITATIVE DATA |
| Number of kids | DISCRETE DATA |
| Number of tickets in Indian railways | DISCRETE DATA |
| Number of times married | DISCRETE DATA |
| Gender (Male or Female) | QUALITATIVE DATA |

**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/INTERVAL |
| Level of Agreement | RATIO |
| IQ(Intelligence Scale) | INTERVAL |
| Sales Figures | RATIO |
| Blood Group | NOMINAL |
| Time Of Day | ORDINAL |
| Time on a Clock with Hands | INTERVAL |
| Number of Children | RATIO |
| 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?**

**ANS:** When 3 coins are tossed the total number of possible outcomes are 23 = 8.

The total outcomes: TTT, HHH, HTT, HHT, HTH, TTH, THT, THH.

The number of combinations having 2 heads and 1 tail are 3: HHT, HTH, THH

The probability of having 2 heads and 1 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 total possible outcomes 62 = 36.

(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)

(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)

a) sum is equal to 1 = not possible, so probability = 0

b) less than or equal to 4

The favorable outcomes are: (1,1) (1,2) (1,3), (2,1), (2,2) (3,1) = 6

P = 6/36 = 1/6

c) sum divisible by 2 and 3

The favorable outcomes are: (1,5) (2,4) (3,3) (4,2) (5,1) (6,6) = 6

P = 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?**

**ANS:** Total number of balls in a bag = 7

Total Outcomes = 7C2 = 7! / (2! 5!) = 7 x 6 / 2 = 21

Therefore, 21 total outcomes

Favorable outcomes = 5C2 = 5! / (2! 3!) = 5 x 4 / 2 = 10

Hence, the probability is 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

= 1 x 0.015 + 4 x 0.20 + 3 x 0.65 + 5 x 0.005 + 6 x 0.01 + 2 x 0.120

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

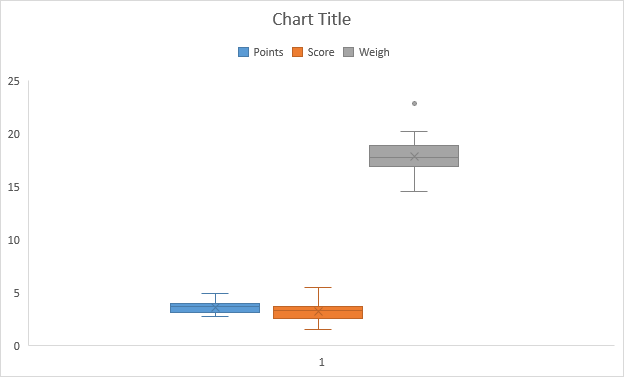
= 3.09

**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:** 



**Inference:** The points and score is negatively skewed whereas weigh is slightly positively skewed and has a data point outside the range.

**Q8) Calculate Expected Value for the problem below.**

**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 = ∑ (Probability x Value)

As there are 9 patients, the probability of selecting each patient is 1/9.

Hence,

Assuming X patient is chosen at random,

= (1/9) x (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9) x (1308)

= 145.34

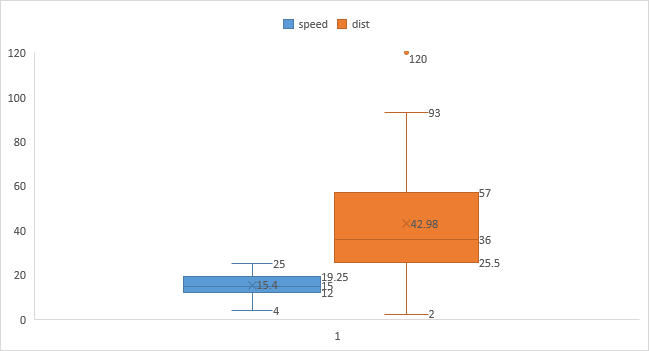
Hence, the age of X patient chosen at random would be 145.34

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

**Car’s speed and distance**

**Use Q9\_a.csv**

 **ANS:**

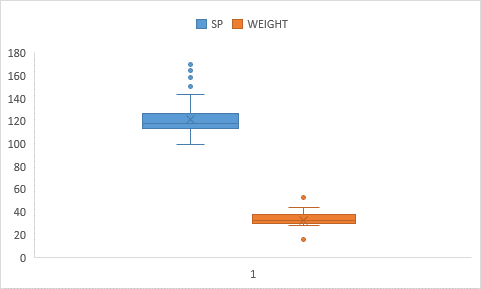


**Inference:** It indicates that the given data is skewed. Distance and speed of a car are positively skewed.

**SP and Weight**

**Use Q9\_b.csv**

**ANS:** 



**Inference:** The SP and Weight are positively skewed.

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



**ANS:** Chick weight data is positively skewed.

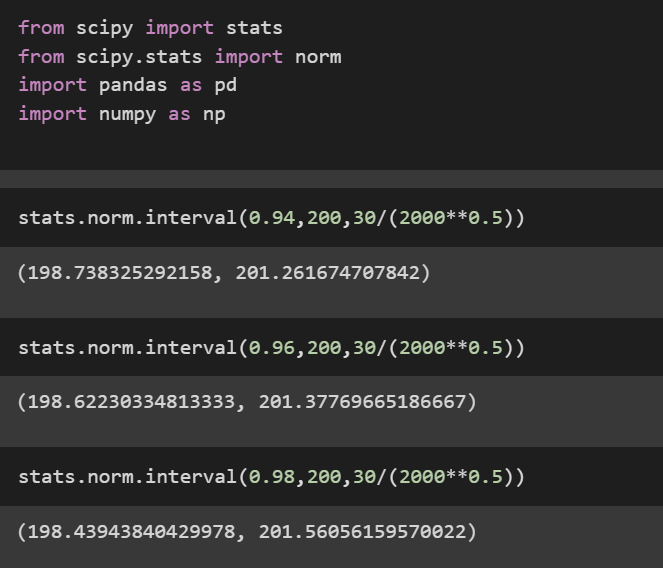
More than 50% of the Chick Weight is between 50 to 150.

As Chick Weight increases, the frequency decreases.

**ANS:** The data is positively skewed and there are data points(outliers) outside the range on the upper side.

**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:**



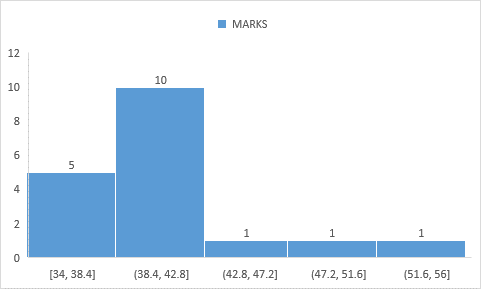
**Q12) Below are the scores obtained by a student in tests:**

**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:** 

1. What can we say about the student marks?

**ANS:** 

There are outliers, and the data is negatively skewed towards the right as mean < median.

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

**ANS:** It is Normal Distribution and skewness = 0.

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

**ANS:** It indicates that the data is positively skewed having the bending towards the right side.

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

**ANS:** It indicates that the data is negatively skewed having the bending towards the left side.

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

**ANS:** It indicates that the data has high peak, more values close to mean and thick at tails.

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

**ANS:** It indicates that the data has flat peak, more dispersed values with lighter tails.

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



1. What can we say about the distribution of the data?

**ANS:** The data is not normally distributed as the median is towards the lower side.

1. What is nature of skewness of the data?

**ANS:** It is positive skewed. 10 -18 = 8

1. What will be the IQR of the data (approximately)?   
   **ANS:** 10-18 = 8

**Q19) Comment on the below Boxplot visualizations?**



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

**ANS:** There are no outliers for both the boxplots. Also, both the boxplots share the same median ranging between 250 and 275. The first one is slightly positively skewed and the second is normally distributed.

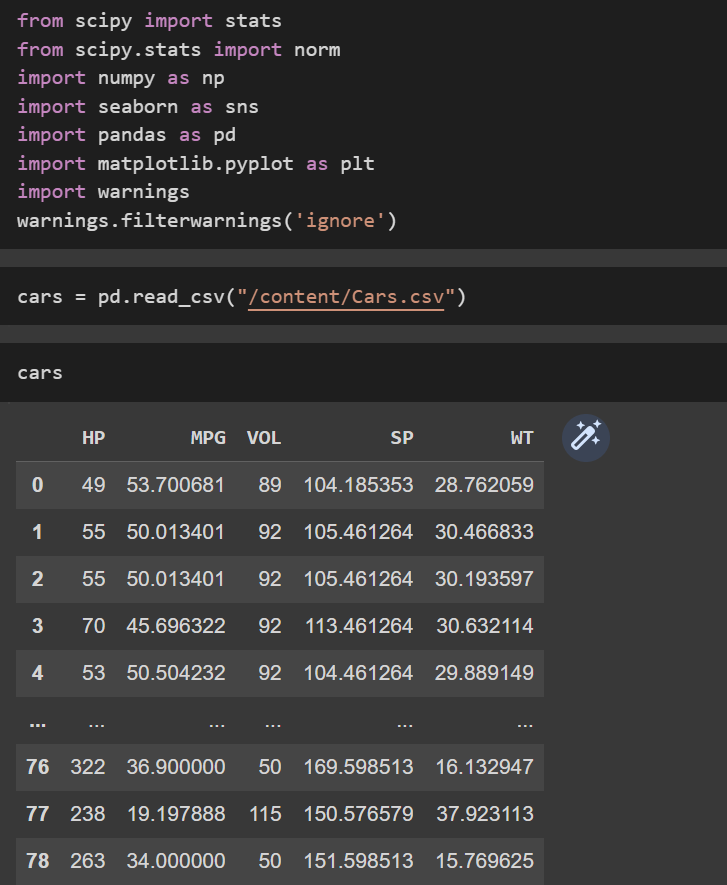
**Q 20) Calculate probability from the given dataset for the below cases:**

**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)**
  3. **P (20<MPG<50)**

**ANS: **

**Graphical user interface, chart, line chart

Description automatically generated**

**Graphical user interface, text

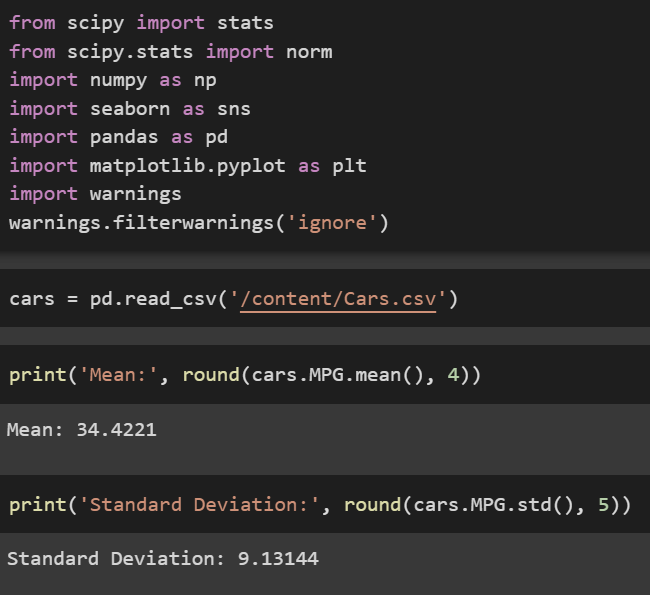
Description automatically generated**

**Q 21) Check whether the data follows normal distribution**

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**ANS:**



Chart, histogram

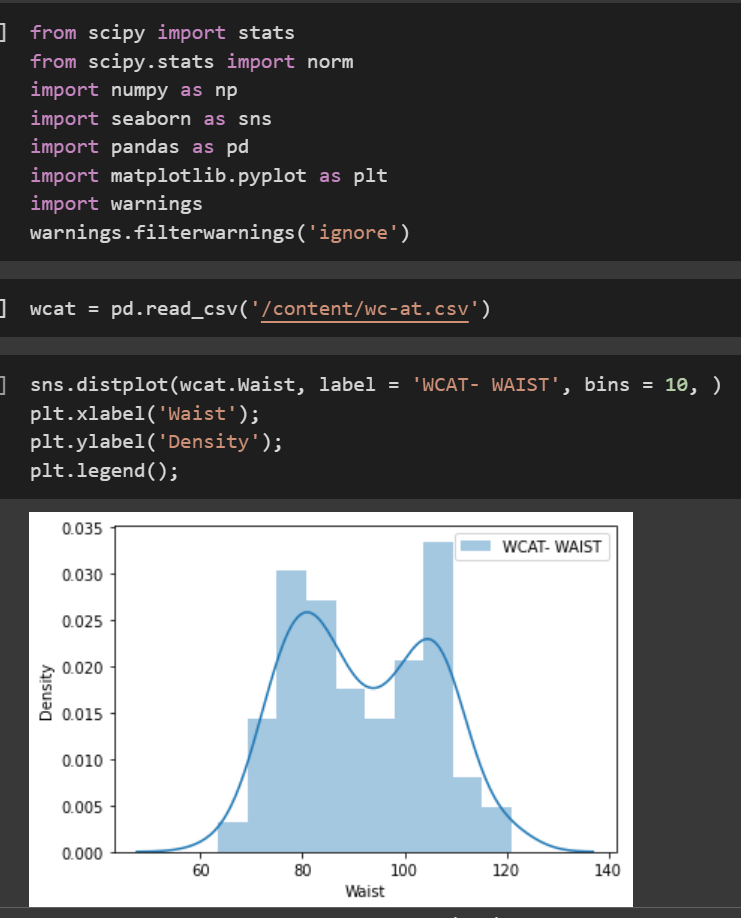
Description automatically generated

Hence, the MPG of Cars follow the Normal Distribution

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

Dataset: wc-at.csv

**ANS:**

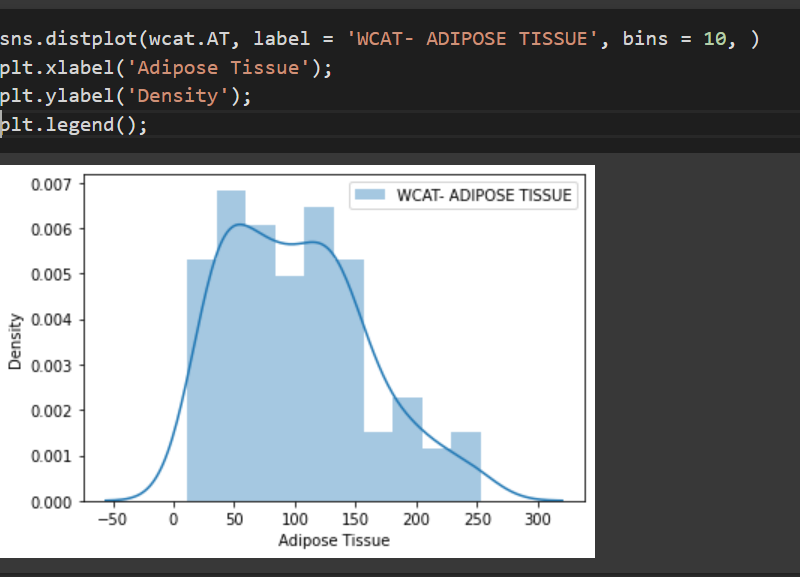


**WAIST CIRCUMFERENCE**

Chart, histogram

Description automatically generated

**ADIPOSE TISSUE**



Hence, the waist and adipose tissue doesn’t follow the normal distribution.

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

**ANS:** Text

Description automatically generated

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

**ANS:**

A screenshot of a computer

Description automatically generated with medium confidence

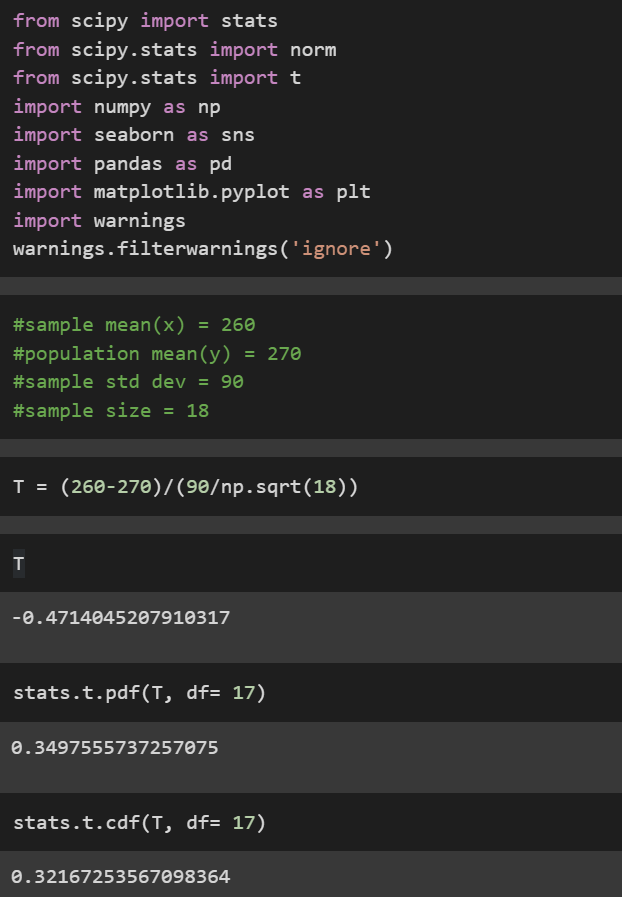
**Q 24) A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days.**

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

**ANS:**

ANS: 

Hence, 32%