|  |  |
| --- | --- |
| 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/Categorical |

Q1) Identify the Data type for the Following:

Q2) Identify the Data types, which were among the following

|  |  |
| --- | --- |
| 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) | Ordinal |
| 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 | Ratio |
| SAT Scores | Interval |
| Years of Education | Interval |

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

Ans:

The probability of 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:

a) The probability of sum is Equal to 1 is 0

b) The probability of sum is Less than or equal to 4 is 6/36 or 1/6

c) The probability of sum is divisible by 2 and 3 is 6/36 or 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:

The probability that none of the balls drawn blue is 10/21 or 5/7

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:

The Expected number of candies for a randomly selected child is **3.09**

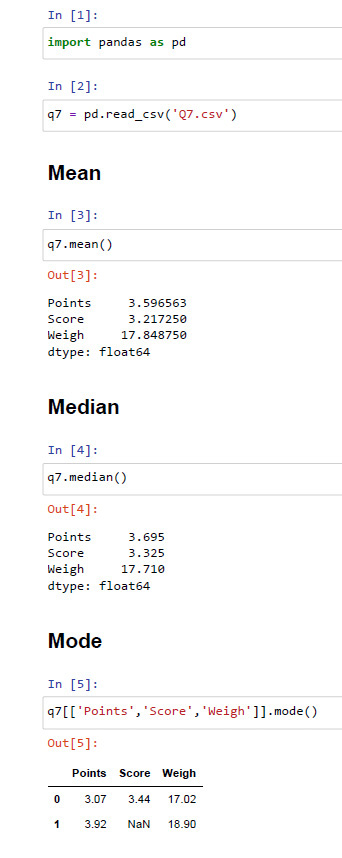
Explanation:

= 1\*0.015 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.01 + 2\*0.12

= 3.09

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

**Use Q7.csv file**

**ANS:**

* Mean:

Points = 3.596563

Score = 3.217250

Weigh = 17.848750

* Median:

Points = 3.695

Score = 3.325

Weigh = 17.710

* Mode:

Points = 3.07

Score = 3.44

Weigh = 17.02

Variance:

Points = 0.28

Score = 0.95

Weigh = 3.19

Standard Deviation:

Points = 0.53

Score = 0.97

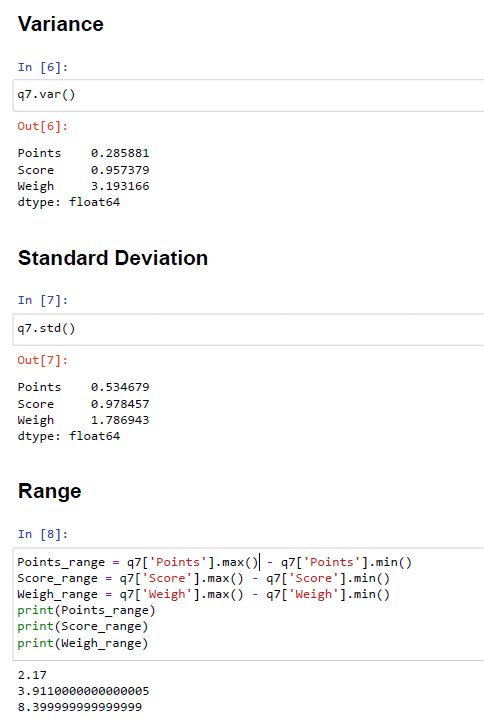
Weigh = 1.78

Range:

Points = 2.17

Score = 3.91

Weigh = 8.39

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 when one of the patients is chosen at random is 145.33(in pounds).

Explanation:

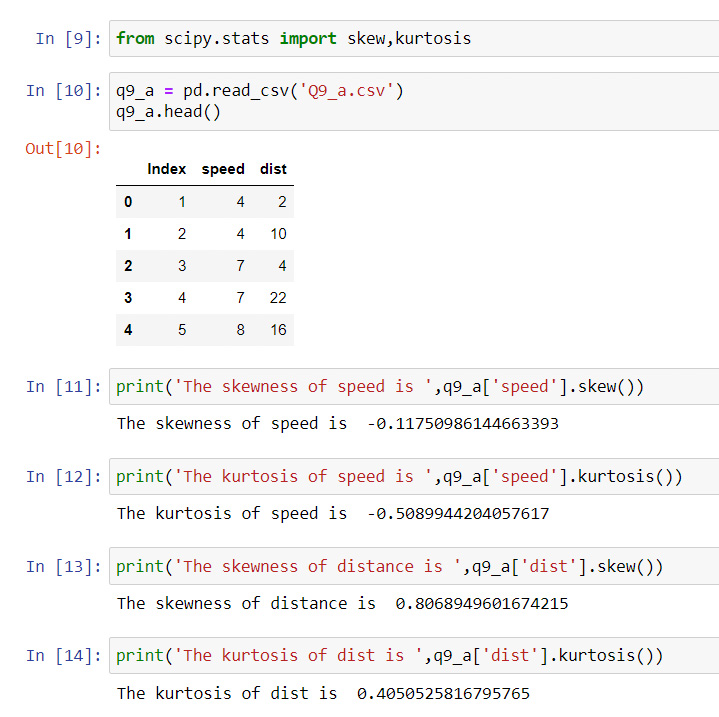
No. of data given here is 9, then probability of choosing one patient is 1/9, so

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

=145.33

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

Cars speed and distance



Speed:

Skewness = -0.117

Kurtosis = -0.508

Distance:

Skewness = 0.806

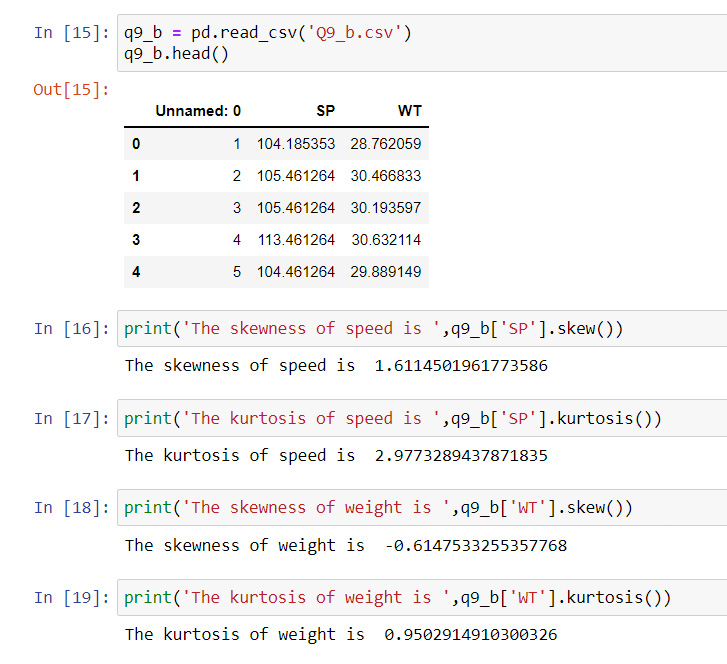
Kurtosis = 0.405

**Inference:**

1. The skewness of speed is negative, it signifies that the data are negatively skewed or left skew
2. Mean<median<mode
3. The kurtosis of speed is Negative, it signifies that the curve is flat with thin tails and data is platykurtic data.
4. The skewness of distance is positive, it signifies that the data are positively skewed or right skew
5. Mean>median>mode
6. The kurtosis of distance is positive, it signifies that the curve is peaked with thick tails and data is called as leptokurtic data.

SP and Weight (WT)

Use Q9\_b.csv



Speed:

Skewness = 1.611

Kurtosis = 2.977

Weight:

Skewness = -0.614

Kurtosis = 0.950

**Inference**:

1. The skewness of SP is Positive, it signifies that the data are positively skewed or right skewed
2. Mean>median>mode
3. The kurtosis of SP is Positive, it signifies that the curve is peaked with thick tails and data is Leptokurtic data.
4. The skewness of Weight is negative, it signifies that the data are negatively skewed or left skewed
5. Mean<median<mode
6. The kurtosis of Weight is positive, it signifies that the curve is peaked with thick tails and data is called as leptokurtic data.

Q10) Draw inferences about the following boxplot & histogram



**Inference:**

* The above histogram shows that the values are positively skewed or right skewed since more no of data present in left side
* The skewness values is greater than zero
* Thin tail present towards the right side
* Mean>median>mode

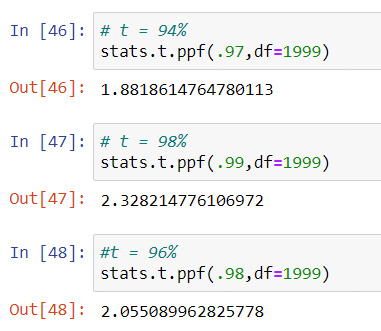


**Inference:**

* The boxplot shows that greater density of data present towards left side
* The data are positively skewed or right skewed
* The data contains outliers in positive direction, the dots represents the presence of outliers

**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?

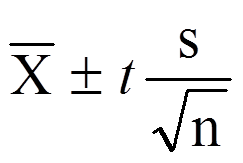
**ANS:**

****

t94 = 1.88

t98 = 2.33

t96 = 2.05

**Confidence Interval =**

94% confidence interval:

= 200-1.882\*(30/√2000) to 200+1.882\*(30/√2000)

=198.73 to 201.26

98% confidence interval:

= 200-2.328\*(30/√2000) to 200+2.328\*(30/√2000)

=198.43 to 201.56

96% confidence interval:

= 200-2.055\*(30/√2000) to 200+2.055\*(30/√2000)

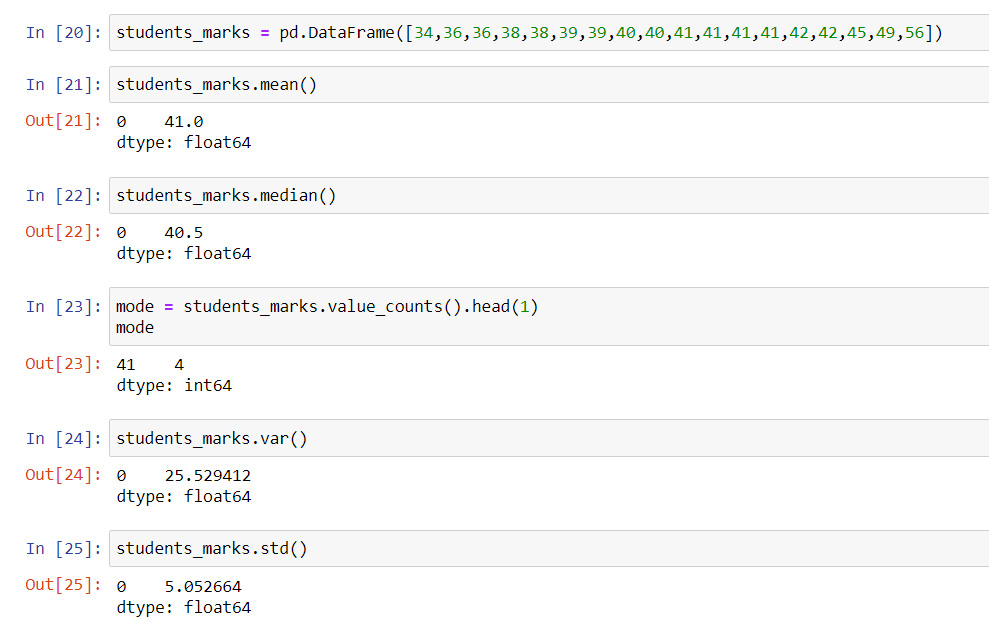
=198.62 to 201.378

**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

Median = 40.5

Mode = 41

Variance = 25.52

Std deviation = 5.05

1. What can we say about the student marks?

**ANS:**

* The average mark of students in test is 41
* The maximum and minimum marks are 56 and 34
* Most of the students mark between 35 to 45

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

ANS:

No Skewness

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

ANS:

Positive Skewness

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

ANS:

Negative Skewness

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

ANS:

In Positive Kurtosis (>3), the distribution is peaked and has thick tails (i.e. most of the values in distribution located in tails rather than around the mean).

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

ANS:

In Negative Kurtosis (<3), the distribution is flat and has thin tails(i.e. it has fewer values in its shorter tails)

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



What can we say about the distribution of the data?

ANS:

* The distribution of this data is not Normally distributed
* Most of the values in left so it is left skewed
* The median value of the distribution is 15
* The Lower and Upper quartile are 10 and 18

What is nature of skewness of the data?

ANS:

Most of the values are skewed towards left side so the nature of the skewness is left skewed

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

IQR = QR3-QR1

=18-10

IQR=10  
  
Q19) Comment on the below Boxplot visualizations?



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

**ANS:**

The given box plots are Normally distributed, both have median around 262.5 and they have no outlier.

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

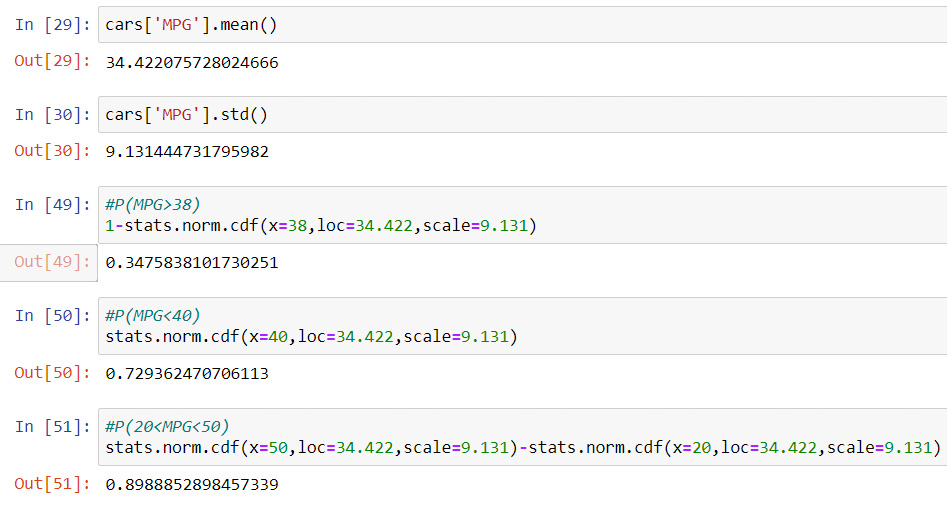
Data \_set: Cars.csv

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:**

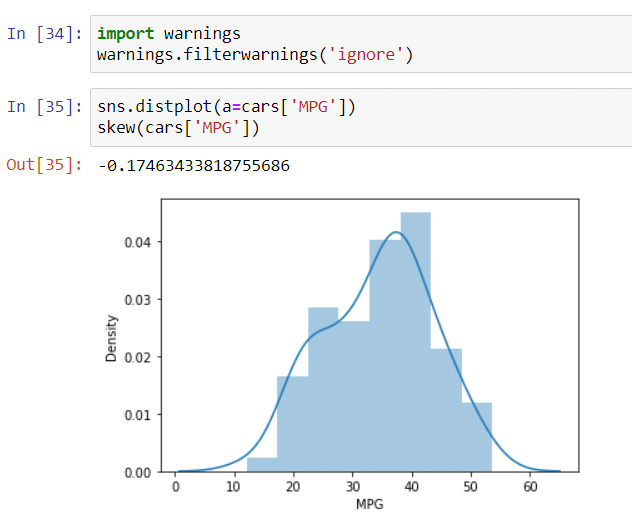
* + P(MPG>38) = 0.34
  + P(MPG<40) = 0.72
  + P (20<MPG<50) = 0.89

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

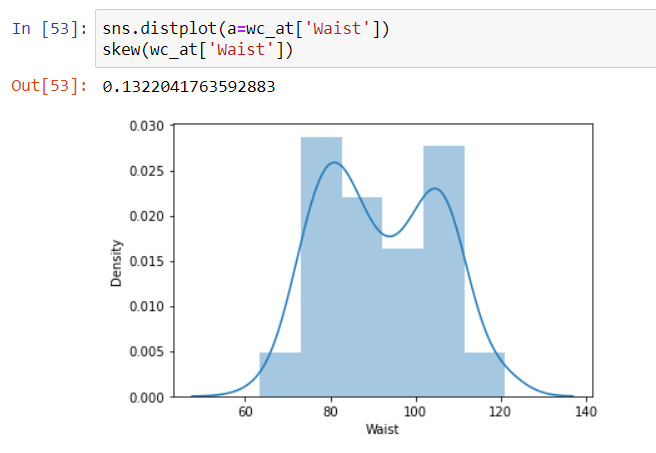
**ANS:**

**** Skewness around -0.5 to 0.5 is acceptable, here skewness of MPG is -0.17 hence it follows normal distribution.

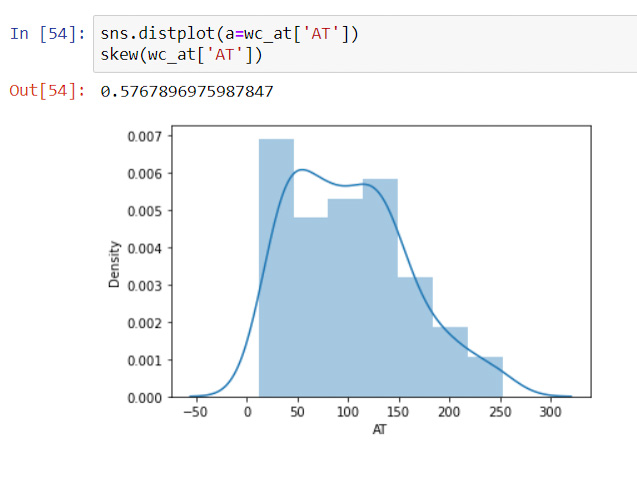
1. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

**ANS**:

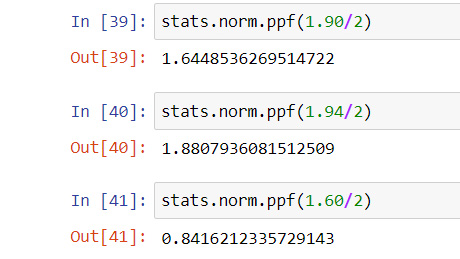


Skewness around -0.5 to 0.5 is acceptable, here skewness of Waist is 0.13 hence it follows normal distribution



Skewness around -0.5 to 0.5 is acceptable, here skewness of AT is 0.57 hence it follows normal distribution

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

**ANS:**

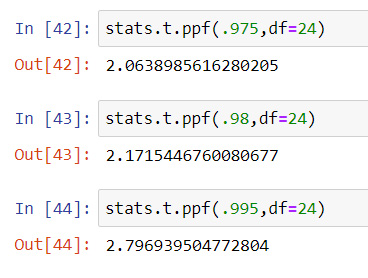
Z90 = 1.645

Z94 = 1.88

Z60 = 0.842

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

**ANS:**

t95= 2.064

t96 = 2.172

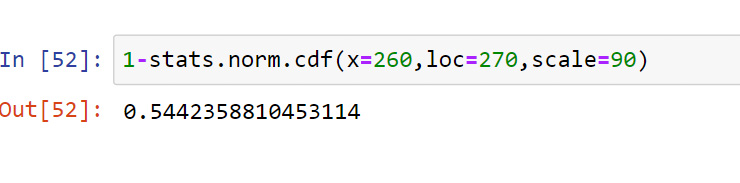
t99 = 2.797

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:**

The probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 54.4%