|  |  |
| --- | --- |
| 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 | Nominal |
| Level of Agreement | Ordinal |
| 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 | Ordinal |
| 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 :** 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):** 0

**Ans b):** 1/6

**Ans c):** 5/36

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:** 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.

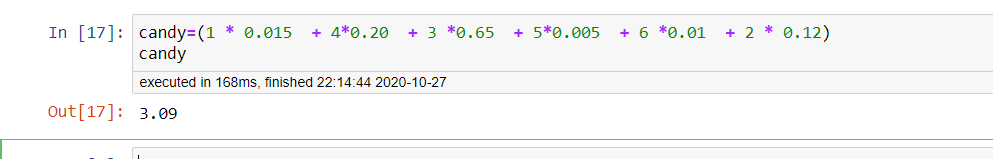
Child B – probability of having 4 candies = 0.20

**Ans:**

Expected number of candies for a randomly selected child

=  1 \* 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

* 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:**

**For Points :**

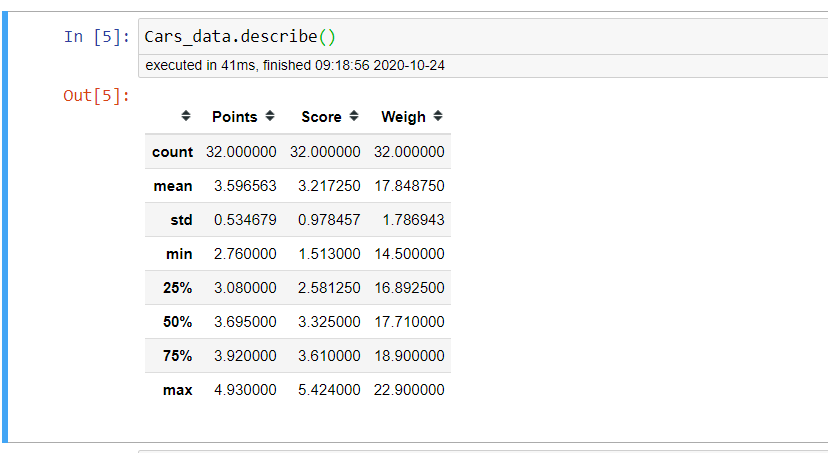
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mean | Median | Mode | Variance | Standard Deviation | Range (Max-Min) |
| 3.596563 | **3.6950000000000003** | 3.07 | **0.27694755859375003** | 0.534679 | 2.17 |

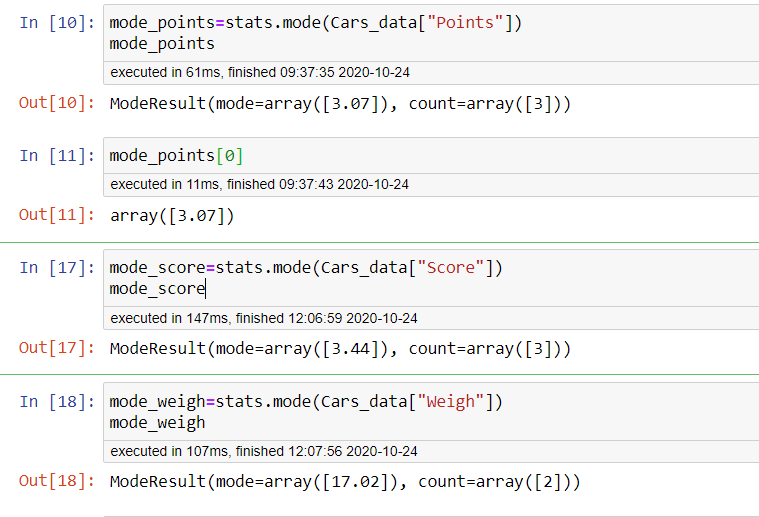
**For Score :**

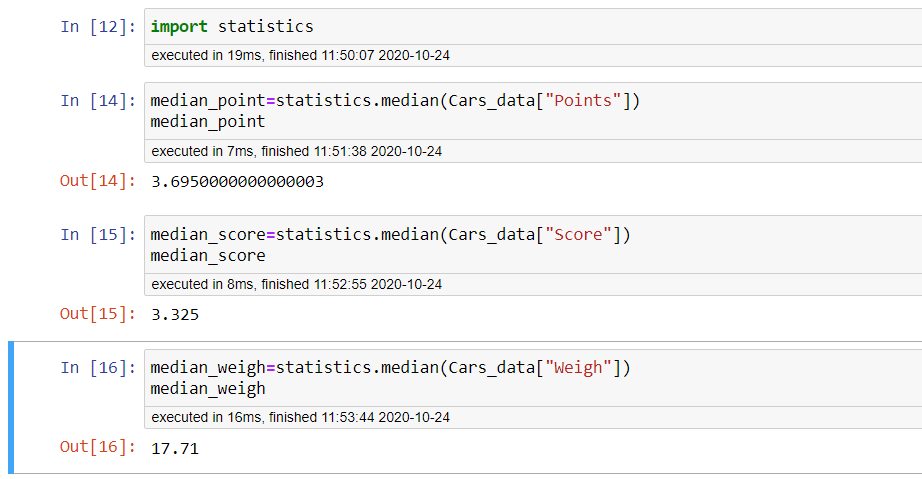
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mean | Median | Mode | Variance | Standard Deviation | Range (Max-Min) |
| 3.217250 | **3.325** | 3.44 | **0.927460875** | 0.978457 | 3.911 |

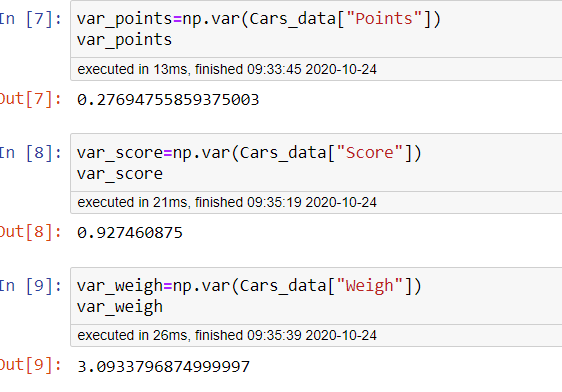
**For Weigh :**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mean | Median | Mode | Variance | Standard Deviation | Range (Max-Min) |
| 17.848750 | **17.71** | 17.02 | **3.0933796874999997** | 1.786943 | 8.4 |



****

****

****

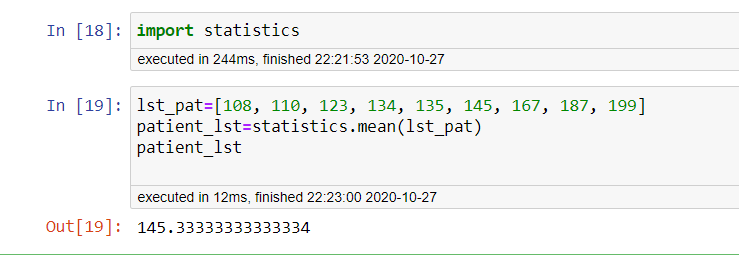
Q8) Calculate Expected Value for the problem below

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?

**Ans:** 145.33



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

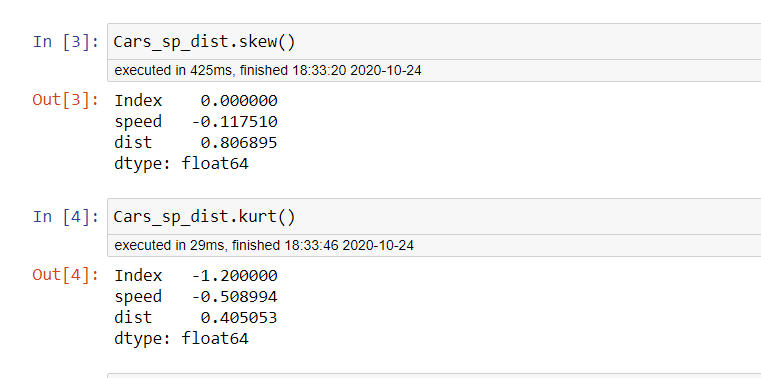
**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:**

|  |  |  |
| --- | --- | --- |
|  | **Speed** | **Inferences** |
| **Skewness** | **-** 0.117510 | Left skewed |
| **Kurtosis** | - 0.508994 | Flat peak |

|  |  |  |
| --- | --- | --- |
|  | **Distance** | **Inferences** |
| **Skewness** | 0.806895 | Right Skewed |
| **Kurtosis** | 0.405053 | Sharp Peak |

****

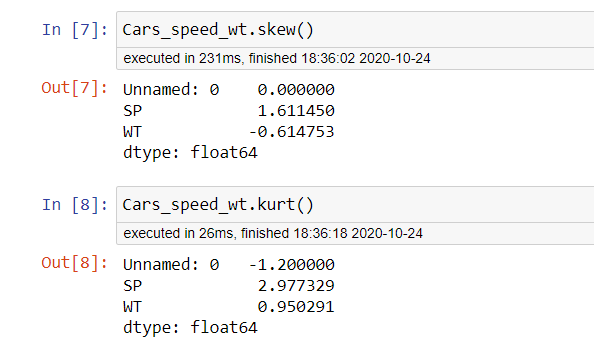
**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:**

|  |  |  |
| --- | --- | --- |
|  | **SP** | **Inferences** |
| **Skewness** | 1.611450 | Right skewed |
| **Kurtosis** | 2.977329 | Sharp Peak |

|  |  |  |
| --- | --- | --- |
|  | **Weight** | **Inferences** |
| **Skewness** | -0.614753 | Left Skewed |
| **Kurtosis** | 0.950291 | Sharp Peak |

****

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



**Ans:** It is Right Skewed as skewness is positive

Outliers

Min

Q3

Q2

Q1

Max



Box plot is left skewed with outliers present above.

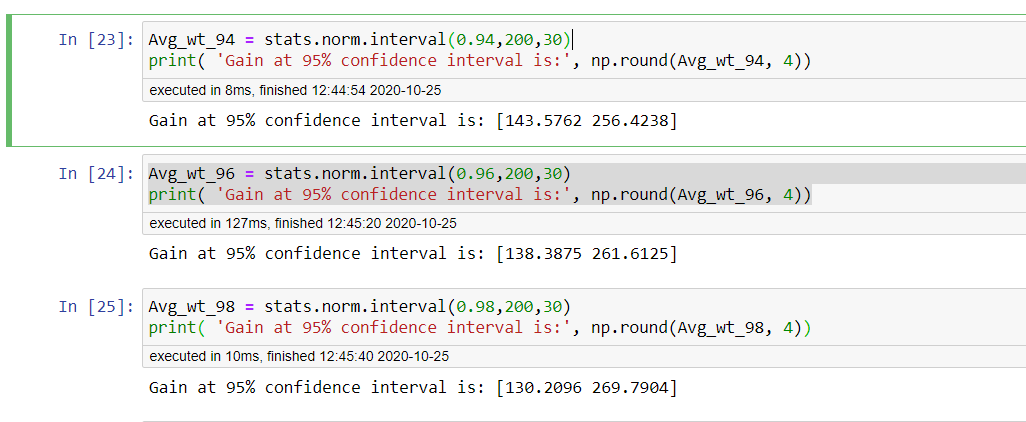
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

**Ans:**

Confidence interval at 94% = **[143.5762 256.4238]**

Confidence interval at 98% = **[138.3875 261.6125]**

Confidence interval at 96% = **[130.2096 269.7904]**



**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?

**Ans:1)**

|  |  |  |  |
| --- | --- | --- | --- |
| Mean | Median | Variance | Std Dev |
| 41 | 40.5 | 24.1111 | 5.05 |

**Ans:2)**

Inference: The student average mark is 41.



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

**Ans:** Skewness will be =0 i.e, data is symmetrical

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

**Ans:** Skewness will be Positive

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

**Ans:** Skewness will be Negative

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

**Ans:** Positive value of kurtosis indicates sharp peak

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

**Ans:** Negative value of kurtosis indicates flat peak

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

**Ans:**

**Distribution of data:** Skewed

**Skewness:** Left skewed

**IQR :** Q3=18 and Q1=10

**IQR=** 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.

**Ans:**

**Boxplot1:** Min=240, Max=280, Median= 262.5

Variations is less. Most of the data is in between 252 to 277 range

Distribution of the data is peak symmetric.

**Boxplot2:** Min=190, Max=350, Median= 262.5

Variations is more. Most of the data is in between 225 to 312 range

Distribution of the data is Uniformly 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.

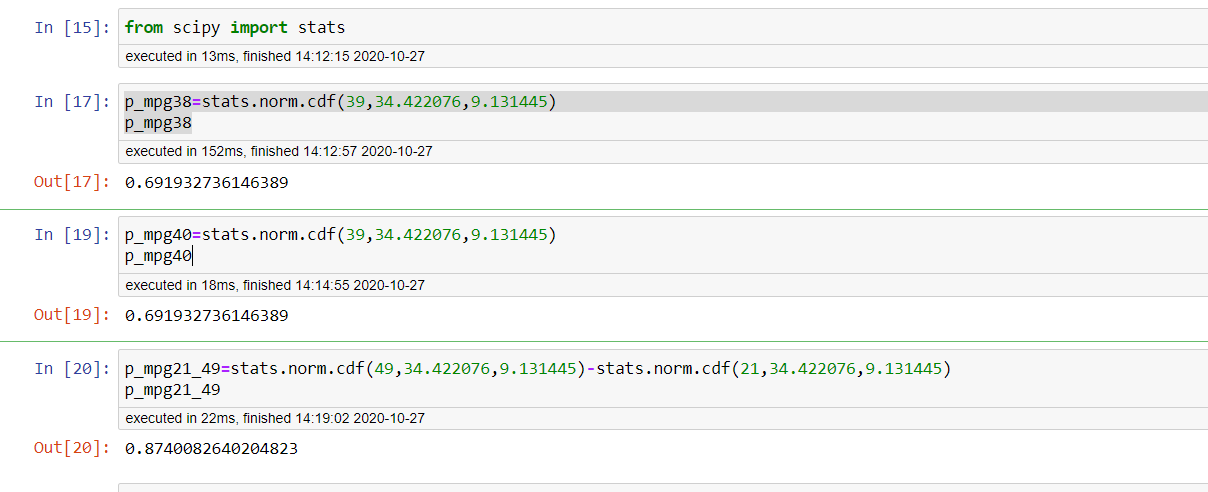
MPG <- Cars$MPG

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

c. P (20<MPG<50)

**Ans:**

|  |  |  |
| --- | --- | --- |
| P(MPG>38) | P(MPG<40) | P (20<MPG<50) |
| 0.691932736146389 | 0.691932736146389 | 0.8740082640204823 |



Q 21) Check whether the data follows normal distribution

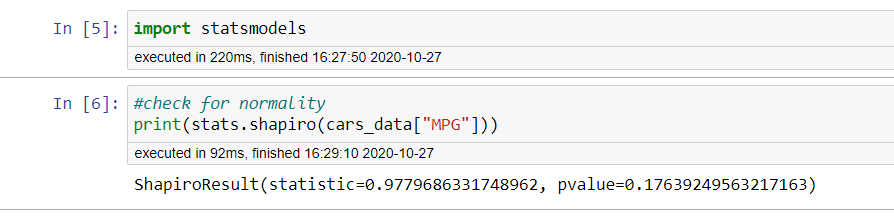
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans:** shapiro test used for checking the normality

pvalue=0.17639249563217163

as pvalue> 0.05, so MPG is 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

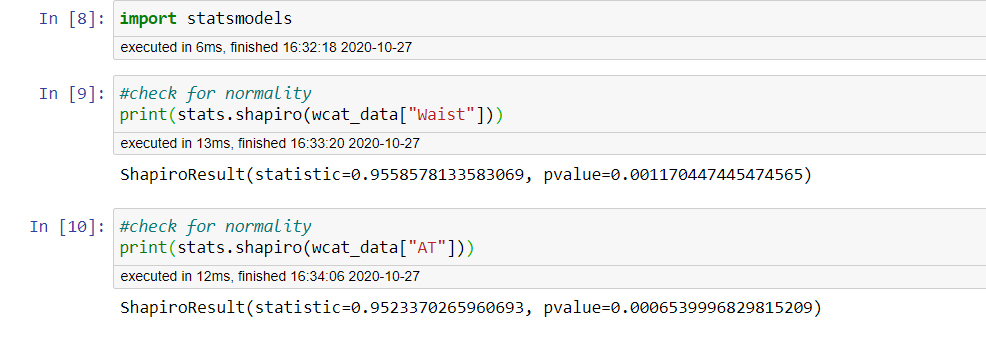
**Ans:** shapiro test used for checking the normality

Pvalue for AT= 0.0006539996829815209

as pvalue<0.05, so AT does not follows normal distribution

Pvalue for Waist=0.001170447445474565

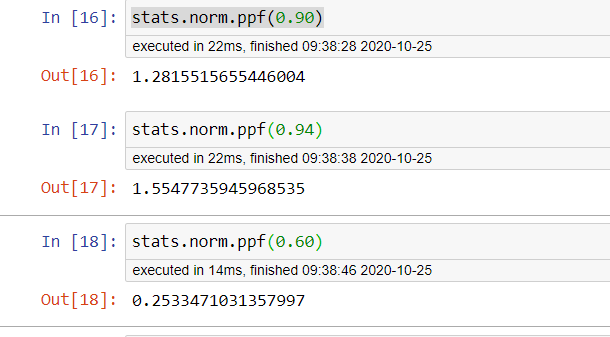
as pvalue<0.05, so Waist does not follows normal distribution



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

**Ans:**

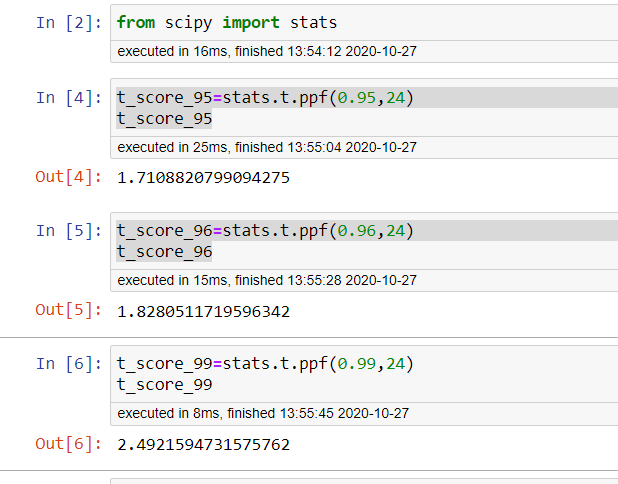
|  |  |  |
| --- | --- | --- |
| **Z score at 90%** | **Z score at 94%** | **Z score at 60%** |
| 1.2815515655446004 | 1.5547735945968535 | 0.2533471031357997 |

****

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

**Ans:**

|  |  |  |
| --- | --- | --- |
| **t score at 95%** | **t score at 96%** | **t score at 99%** |
| 1.7108820799094275 | 1.8280511719596342 | 2.4921594731575762 |



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

**Ans:**

P(x=18)= 34.975%

