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

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

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

Ans= sample={ HHH,THH,HTH,HHT,TTH,TTT,HTT,THT}

P(2H&1T)=3/8

**Q4) Two Dice are rolled, find the probability that sum 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)(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) }

Total outcomes=36

1. **Equal to 1** = 0
2. **Less than or equal to 4 =**

{(1,1)(1,2)(1,3)(2,1)(2,2)(3,1)}

6/36

1. **Sum is divisible by 2 and 3 =**

{(1,5)(2,4)(3,3),(4,2)(5,1)(6,6)}

6/36

**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: The probability of expected number of candies for a randomly selected child are

= Xi\* P (Xi)

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

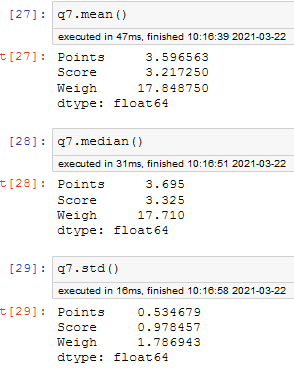
=3.09

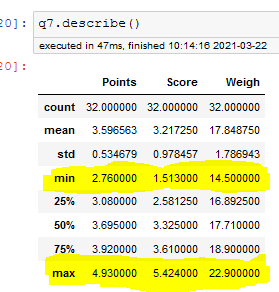
**Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset**

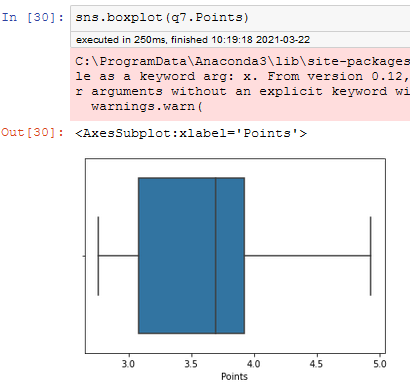
* **For Points,Score,Weigh>**

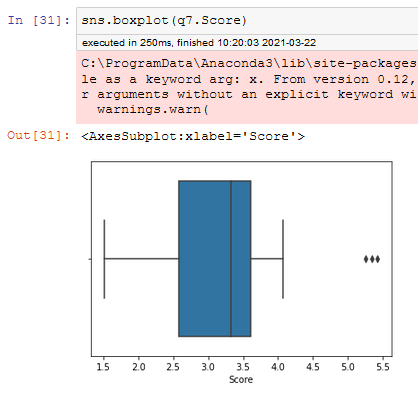
**Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.**

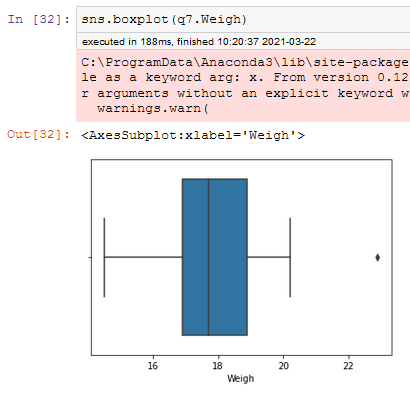
**Use Q7.csv file**









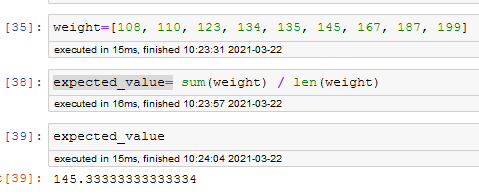


**Q8) Calculate Expected Value for the problem below**

1. **The weights (X) of patients at a clinic (in pounds), are**

**108, 110, 123, 134, 135, 145, 167, 187, 199**

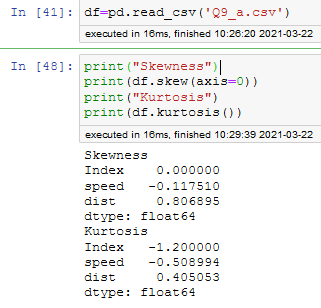
**Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?**

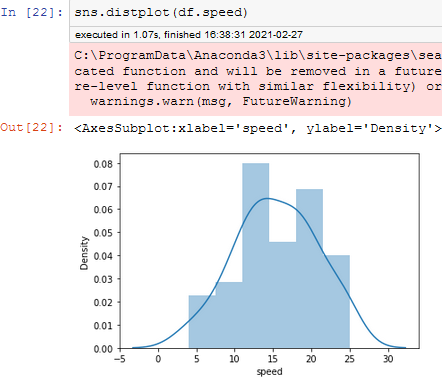
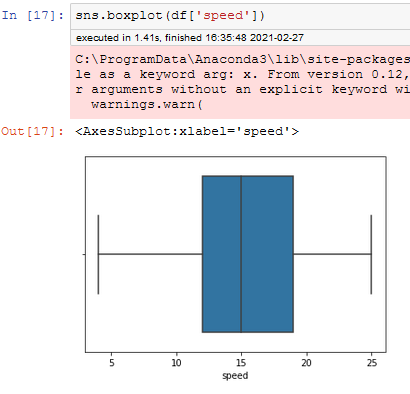


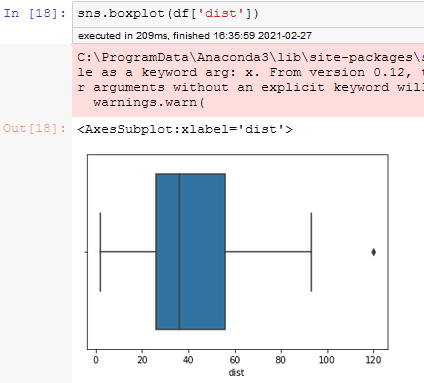
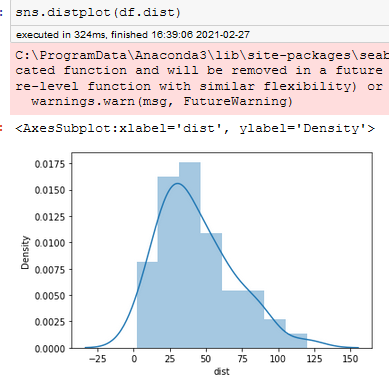
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

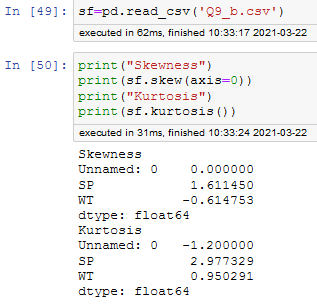


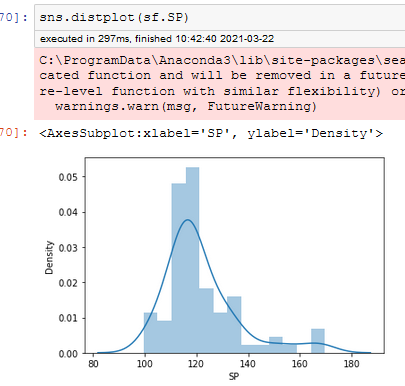
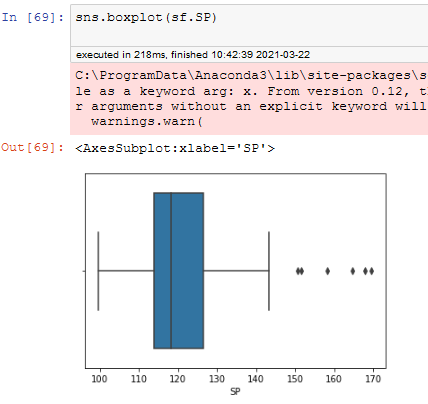


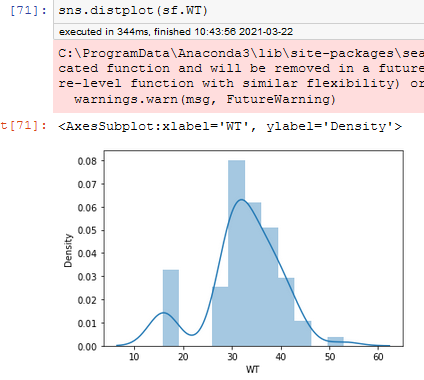
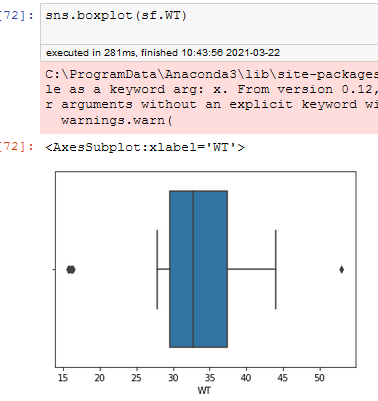
 

**SP and Weight(WT)**

**Use Q9\_b.csv**



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



Ans. Right Skewness

Due to high density on left side and tail on the right 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:

94 %=( 198.7398, 201.260131)

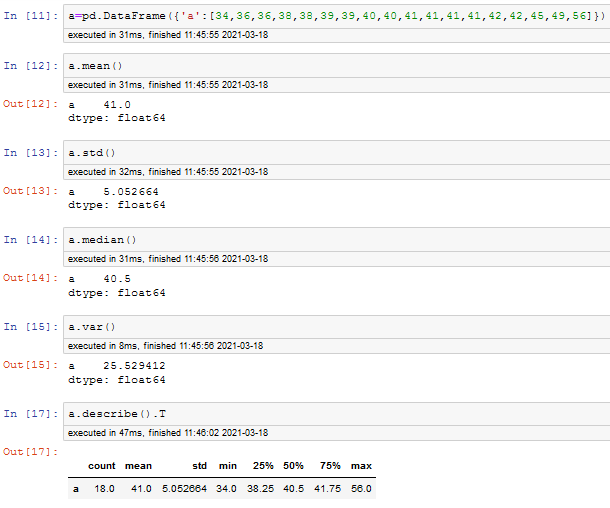
96 %=( 198.6239, 201.376011)

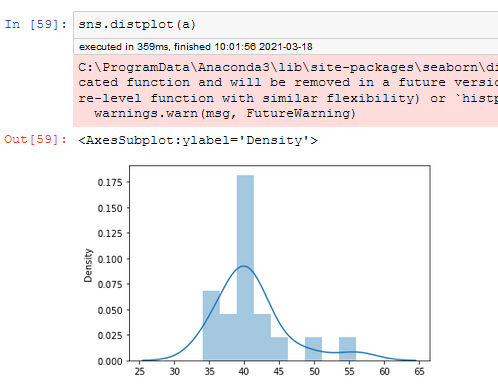
98 %=( 198.4413, 201.55865)

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

**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.



1. **What can we say about the student marks?** 

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

Ans. The distribution is symmetric and have ZERO skewness.

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

Ans. The Distribution is Right Skewed or positively Skewed .

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

Ans. The Distribution is Left Skewed or negatively Skewed .

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

Ans. The distribution has heavier tails than the normal distribution.

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

Ans. The distribution is flat and has thin tails.

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



**What can we say about the distribution of the data?**

The Distribution doesn’t follow the normal distribution a

The Distribution is Left Skewed or negatively Skewed because area between Q1 and Q2 is greater than Q3 and Q2

Mean is to be smaller than median.

**What is nature of skewness of the data?**

The Distribution is Left Skewed or negatively Skewed

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

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.**

* Median for both are same
* Data is normally distributed
* 2nd data will have larger standard deviation
* 2nd data contain more values than 1 data

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

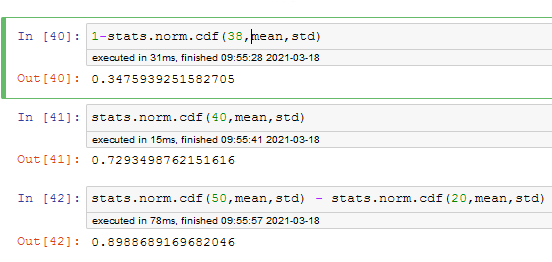
**Data \_set: Cars.csv**

**Calculate the probability of MPG of Cars for the below cases.**

**MPG <- Cars$MPG**

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

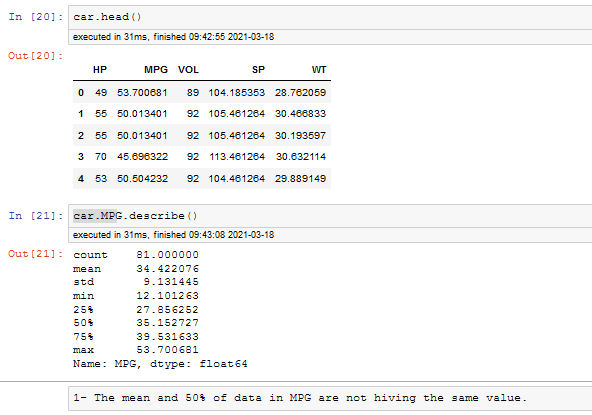
**c. P (20<MPG<50)**

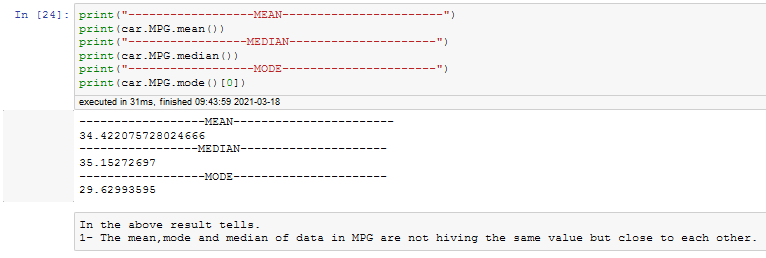


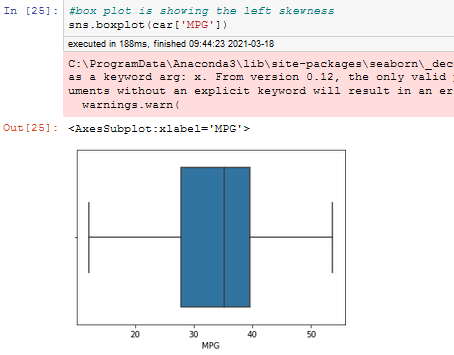
**Q 21) Check whether the data follows normal distribution**

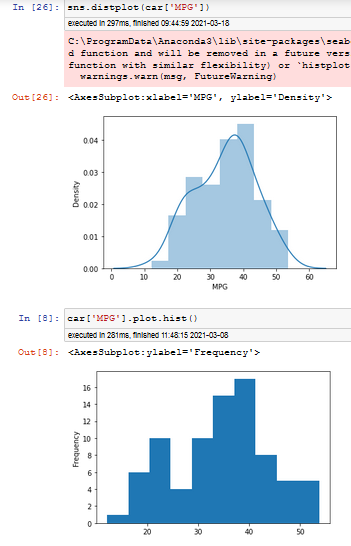
1. **Check whether the MPG of Cars follows Normal Distribution**

**Dataset: Cars.csv**



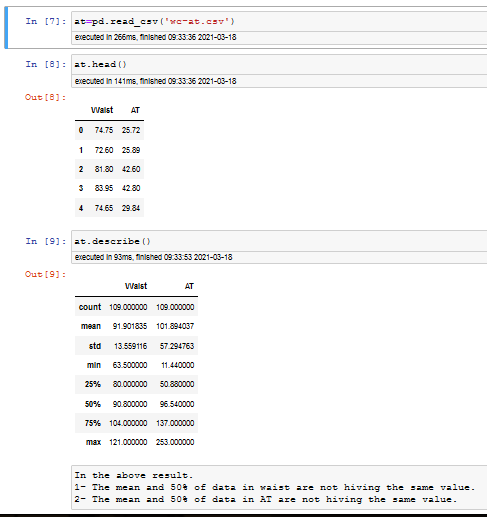


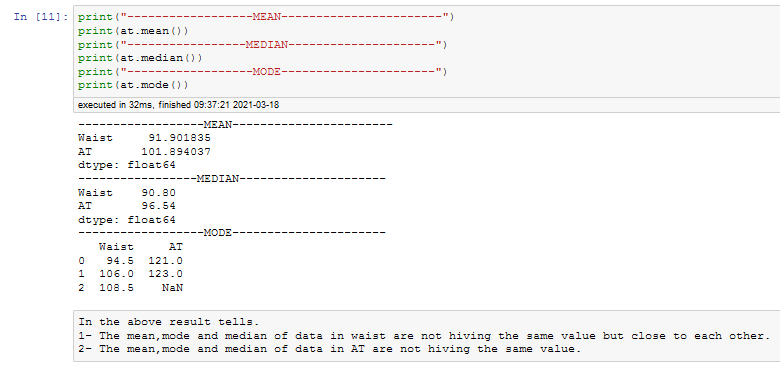


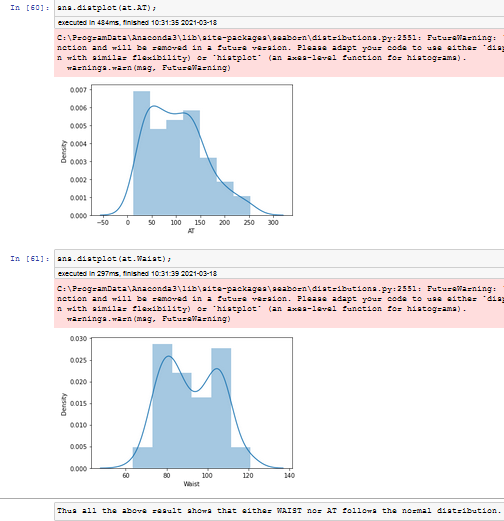


1. **Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution**

**Dataset: wc-at.csv**





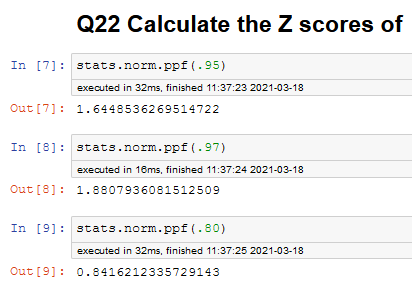


**Q 22) Calculate the Z scores of**

**90% confidence interval, =**

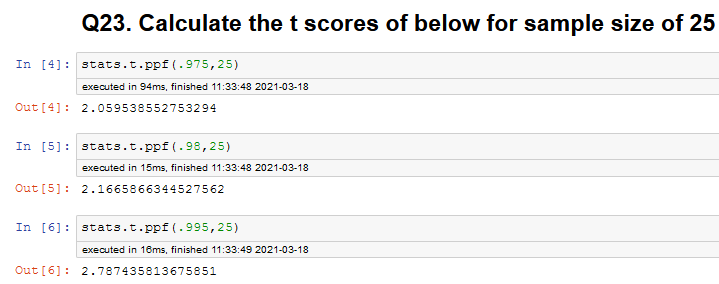
**94% confidence interval, =**

**60% confidence interval =**



**Q 23) Calculate the t scores of below for sample size of 25**

**-95% confidence interval, 96% confidence interval, 99% confidence interval**



**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 n=18 Xbar=260 S=90

We will use T-distribution

* n<30
* ơ is Unknown

t=( Xbar - µ ) / ( s / √n )

t= ( 260 -270 )/ ( 90 / √18 )

t=-0.471

