Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| 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 |

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 | Discrete-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 | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

* **When the three coins are tossed the total number of possible are: 3/8**

**These are (HHH, HHT, HTH, TTT, TTH, THT, THH, HTT)**

**The number of possibilities which have two heads and one tail are:**

**(HHT, THH, HTH)**

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1

* **0**

**(Not possible that sum is equal to 1, because of in other dice we are not having zero)**

1. Less than or equal to 4

* **The set of possible outcomes when we roll a dice are {1,2,3,4,5,6}**

**When we roll two dice then possible outcomes are 6\*6 = 36**

**When we roll two dice, then possibility of getting 4 number are (1,3), (2,2), (3,1).**

**Then,**

**Number of outcomes = 3**

**Total number of possibilities = 36**

**Probability = number of outcomes / total number of possibilities**

**= 3 / 36**

**= 1 / 12**

**Thus, 1/12 is the probability of rolling two dice and getting a sum of 4.**

1. Sum is divisible by 2 and 3

* **When two dice are rolled, then probability that sum is divisible by 2 and 3 are:**

**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?

* **Given, bag contains 2 red, 3 green and 2 blue balls.**

**To find, the probability that none of the balls drawn is blue.**

**Total number of balls are: 2+3+2 = 7**

**Two balls can be drawn in 7C2 ways = 21 ways Number of ways of drawing 2 balls such that none is blue =**

**Number of ways of drawing 2 balls from 2 red and 3 green balls = 5C2 = 10 ways Probability of drawing 2 balls such that none is blue = 10/21**

**10/21 (1 red, 1 green, 2 red, 2 green).**

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

* **3.075**

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.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **POINTS** | **SCORE** | **WEIGH** |
| **MEAN** | 3.596 | 3.217 | 17.848 |
| **MEDIAN** | 3.695 | 3.325 | 17.71 |
| **MODE** | 3.92 | 3.44 | 17.02 |
| **VARIANCE** | 0.2858 | 0.9573 | 3.1931 |
| **STANDARD DEVIATION** | 0.5346 | 0.9784 | 1.7869 |
| **RANGE** | 2.764 | 1.513 | 14.5 |

* From Excel

**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?

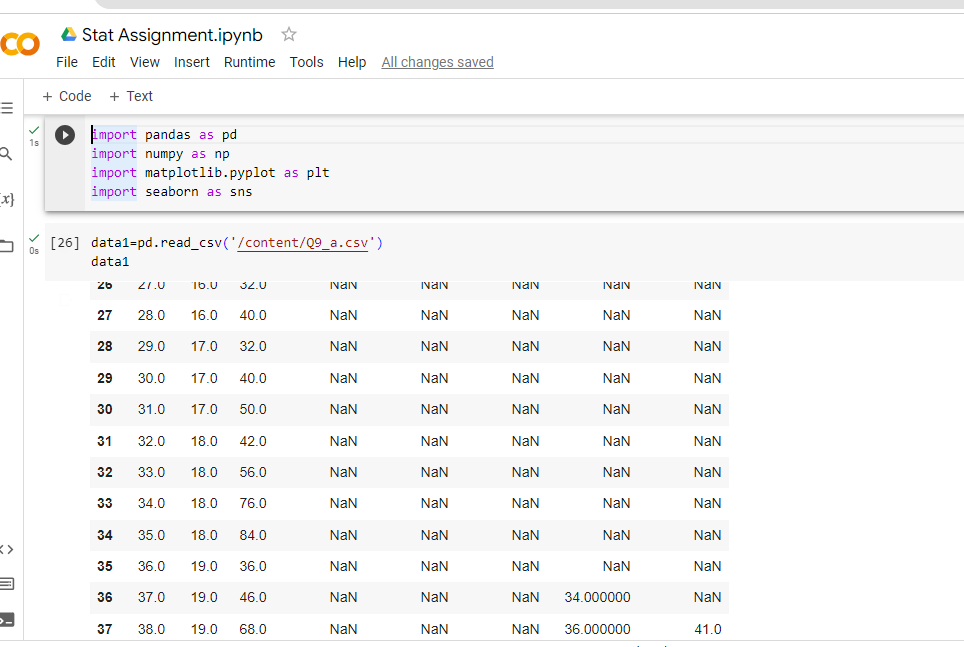
* The Expected Value of the Weight of that patient chosen at random is mean of all the patients

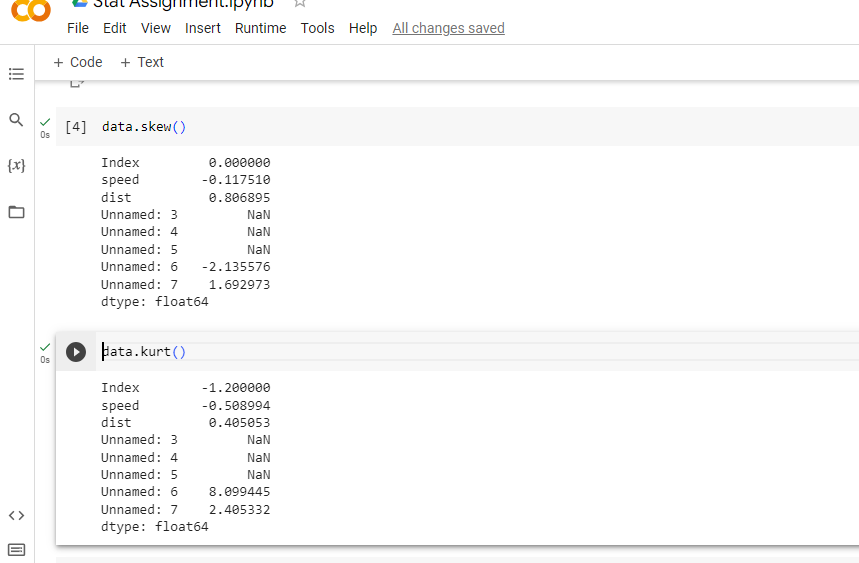
145.34 pounds

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

**Cars speed and distance**

**Use Q9\_a.csv**

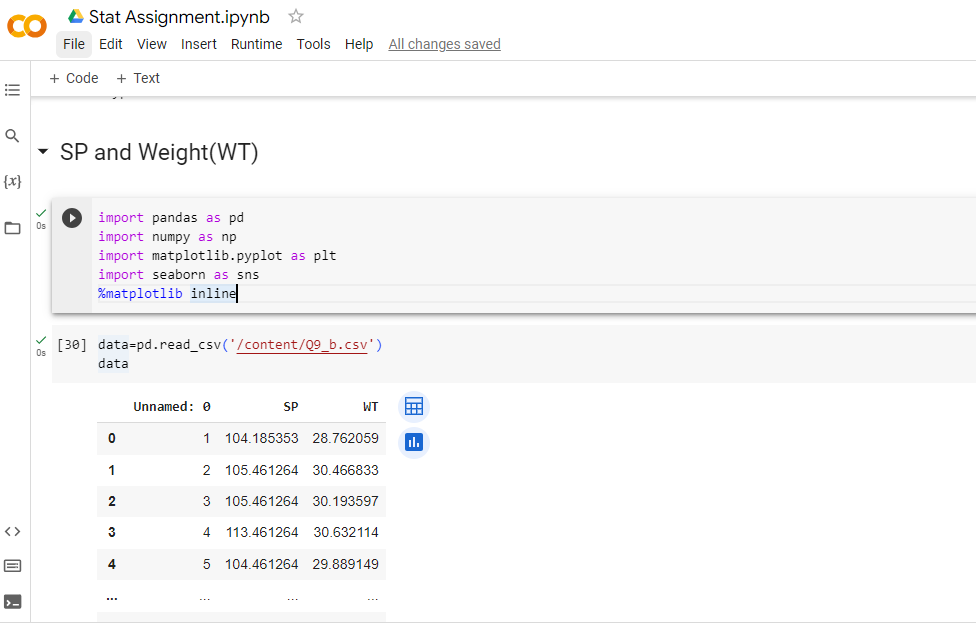


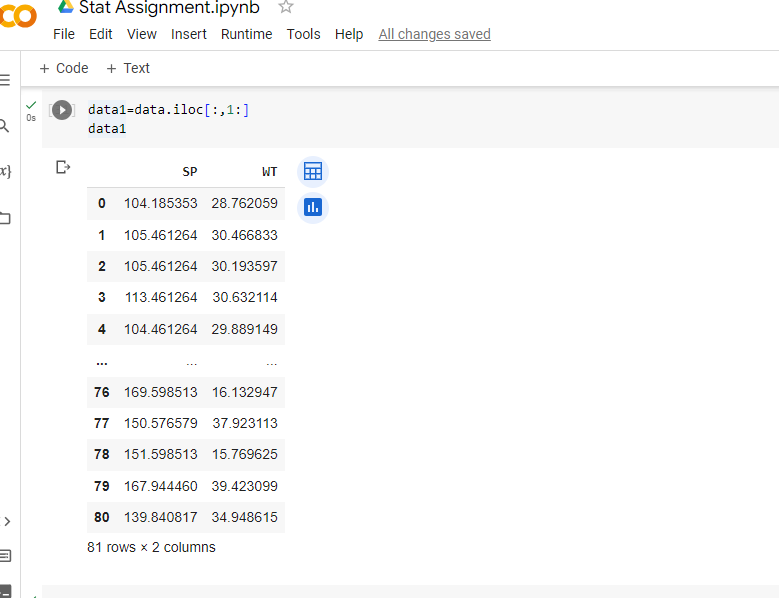


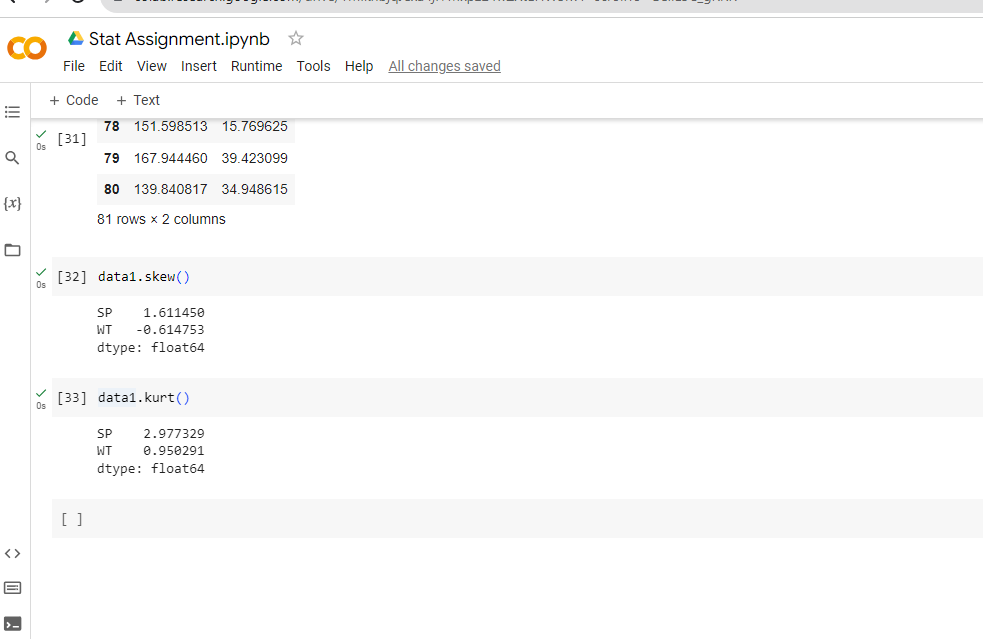
**SP and Weight(WT)**

**Use Q9\_b.csv**









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



🡪 The most of the data points are concerted in the range 50-100 with frequency 200 and least range of weight is 400 some were around 0-10.

So the expected value the above distribution is 75.

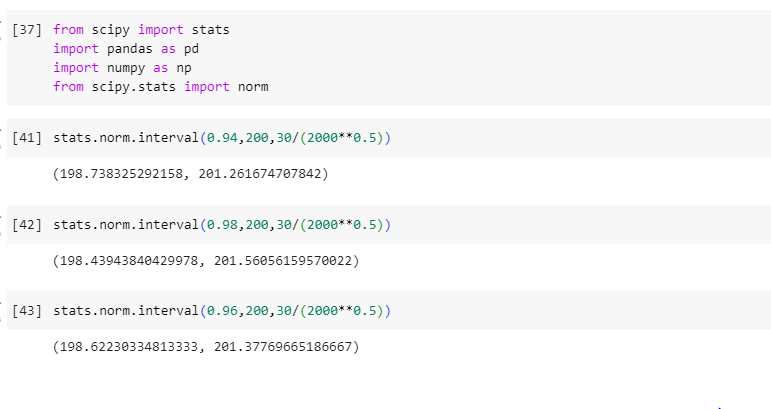
Skewness- we can notice a long tail towards right so it is heavily right skewed.



* Median is less than mean right skewed and we have outlier on the upper side of box plot.

**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?



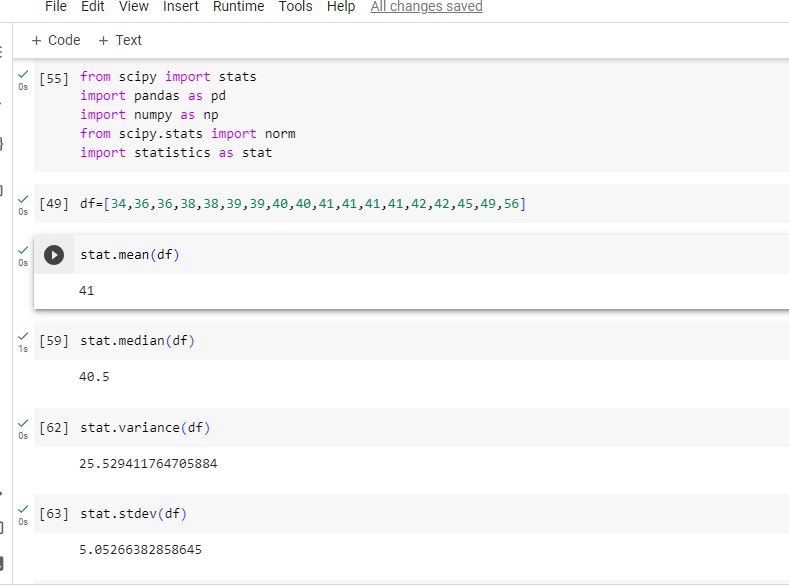


**Q12)** Below are the scores obtained by a student in tests

**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.

* Mean=41, Median=40.5, Variance=25.52, Standard deviation=5.052



1. What can we say about the student marks?

* Most of the students in the class are having an average percentage of 65 and there are only a few students securing value above 90%.

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

* Data is normalized and there is skewness.

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

* Negative skewness.

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

* Positive skewness.

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

* Thinner peak and wider tails.

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

* Thinner tails and wider peak.

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



What can we say about the distribution of the data?

* Not normally distributed.

What is nature of skewness of the data?

* Negative skewness.

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

* 10-18

Q19) Comment on the below Boxplot visualizations?



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

* First there are no outliers.

Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

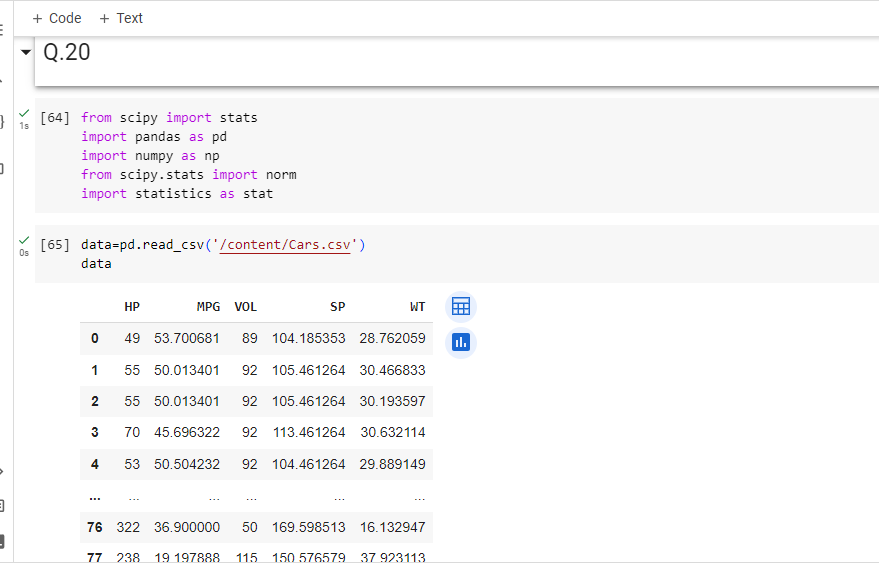
Q 20) Calculate probability from the given dataset for the below cases.

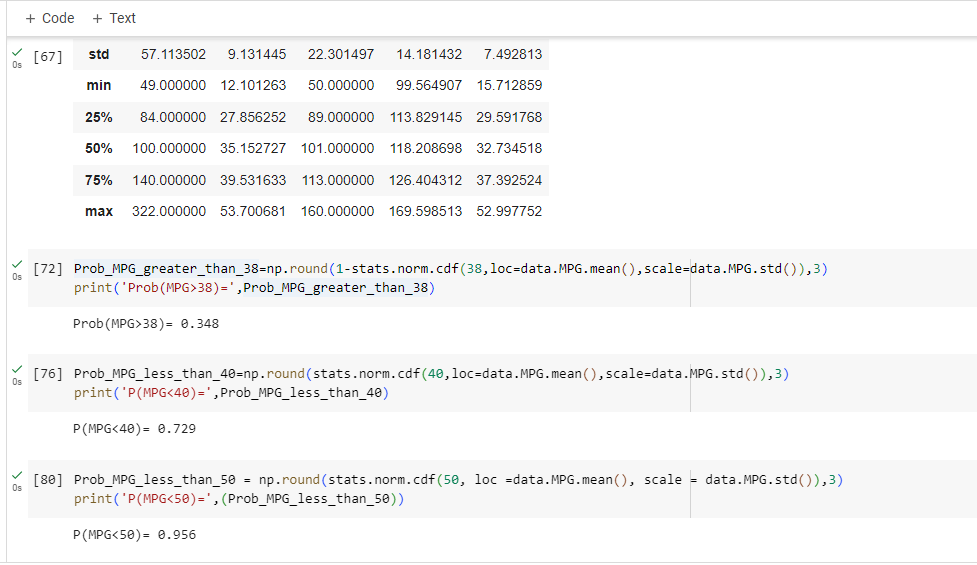
Data \_set: Cars.csv

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)

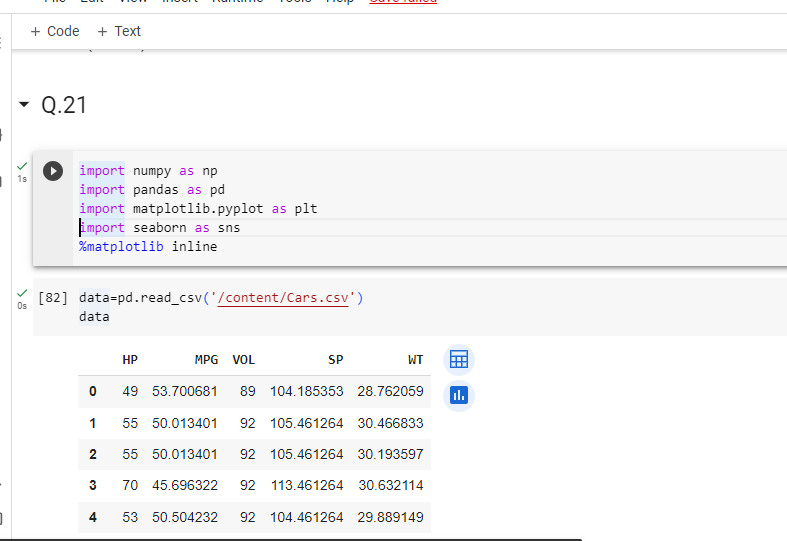


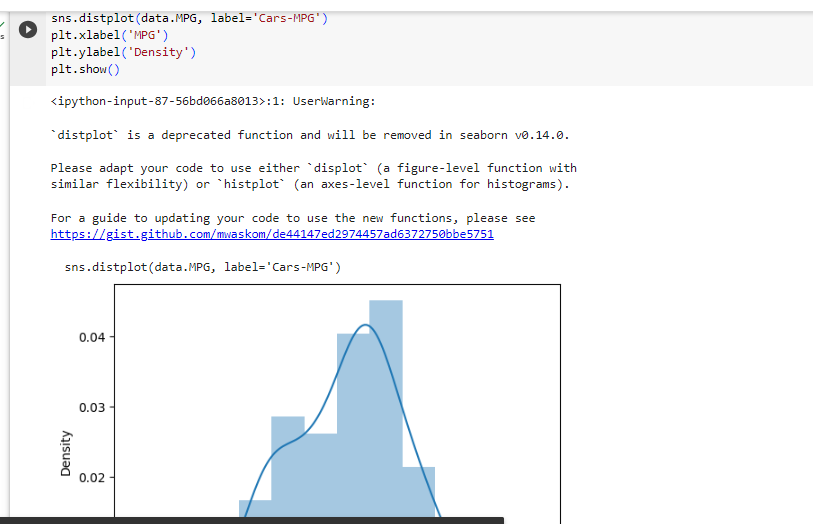


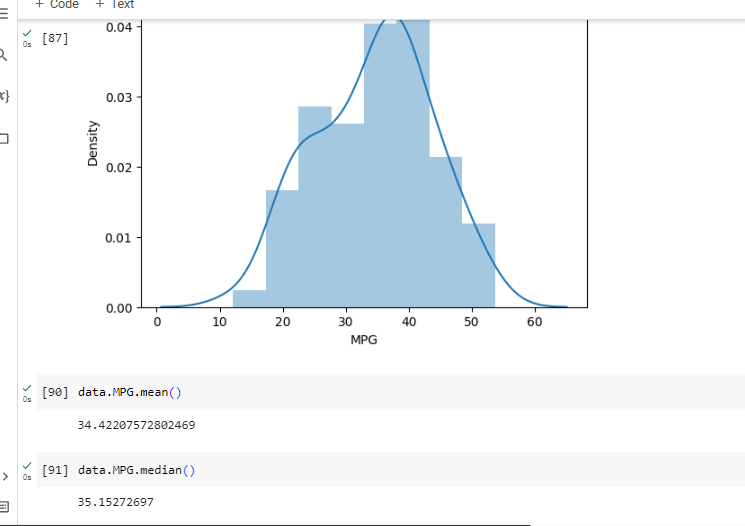
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



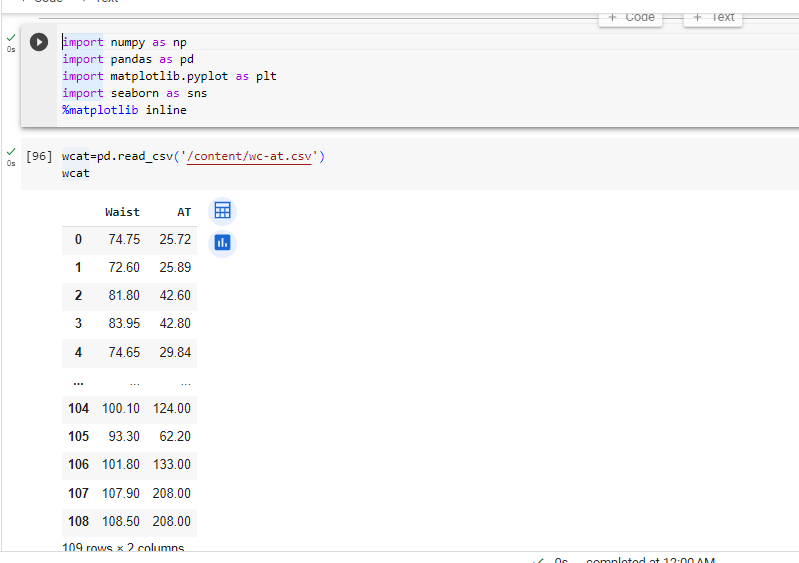


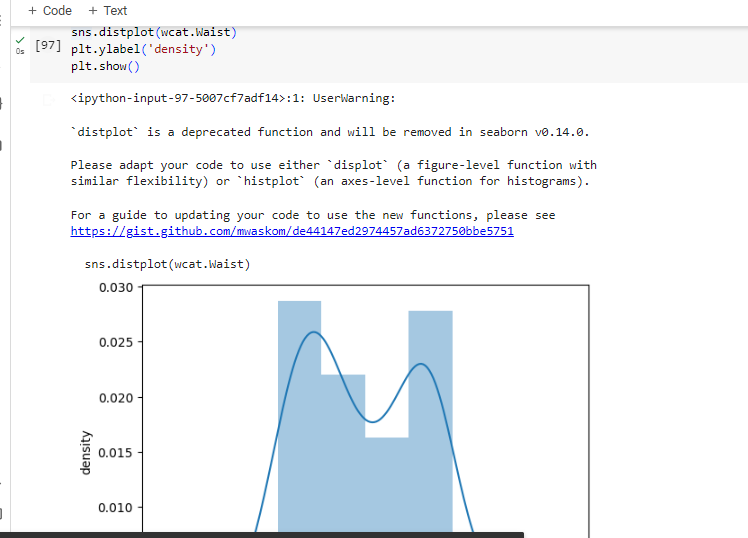


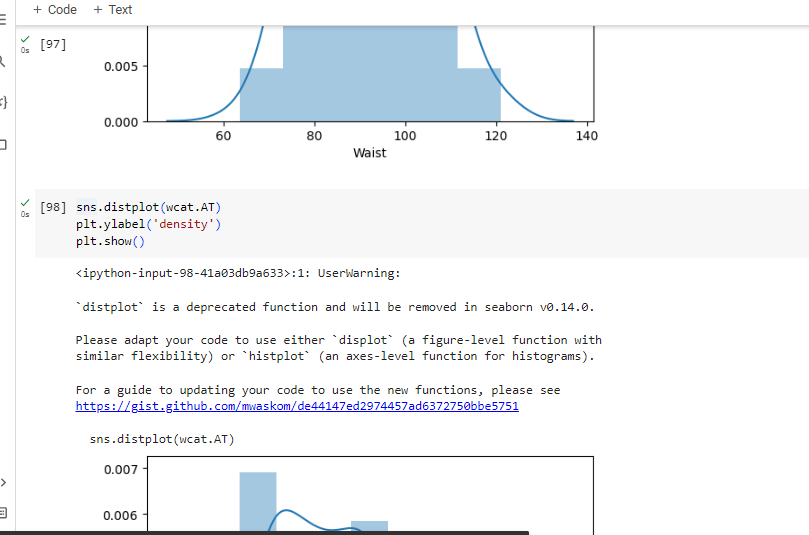
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

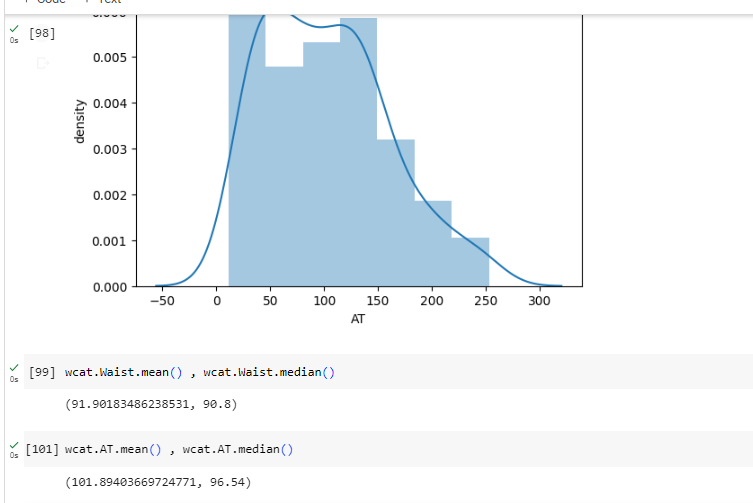
Dataset: wc-at.csv











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

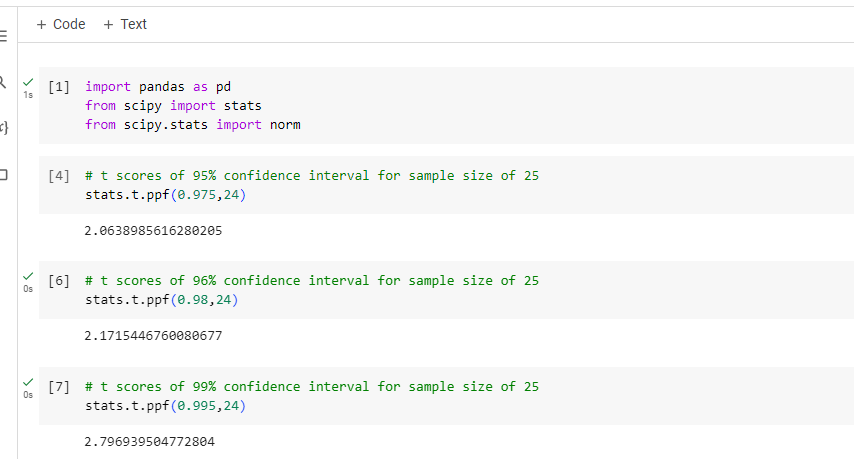




Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25.

* n=25

n-1=24



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



µ=270, n-18, x bar=260, s=90

df=n-1=18-1=17

t score= (x bar- µ)/(s/sqrt(n))

= (260-270)/ (90/sqrt (18))

= -10/21.23

=-0.47

Required probability=0.32

