|  |  |
| --- | --- |
| 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:

Assignment - 1

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 | Nominal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| 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 | Interval |
| 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?

**Ans:** total number of events = 8

2 heads = 2/8 1 tail = 1/8

2/8 + 1/8 = 3/8

= 0.375

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) Equal to 1 == Zero

b) Less than or equal to 4 == 6/36 = 1/6

c) Sum is divisible by 2 and 3 == 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 no pf balls = 7

2 balls picking from 7 = 7C2 = 7\*6/2\*1 = 42/2 = 21

2 balls picking from 7 except blue once = 5C2 = 5\*4/2\*1 = 20/2 = 10

Probability of event = 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:** 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: Points Score Weigh**

**Mean** 3.5965 3.2172 17.8487

**Median** 3.695 3.325 17.71

**Mode** 3.92 3.44 17.02

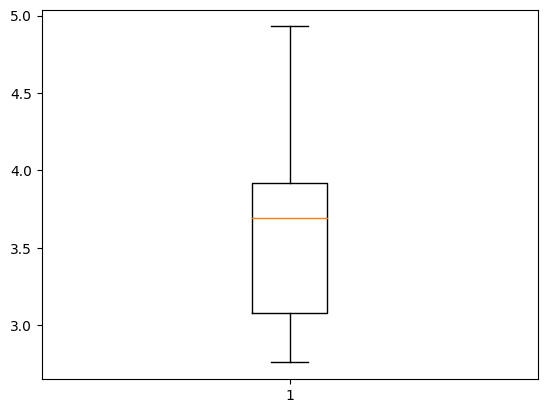
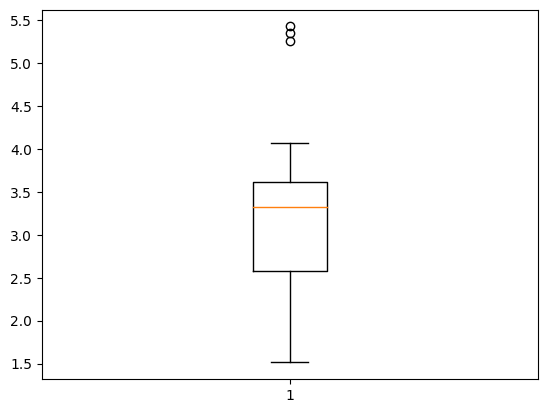
**SD** 0.5346 0.9784 1.7869

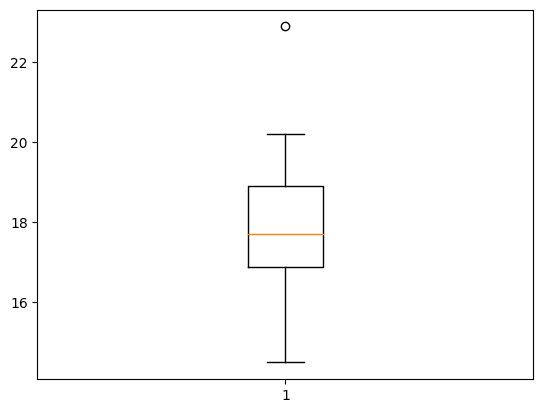
**Variance** 0.2857 0.9572 3.1930

**Min,Max** 2.76 , 4.93 1.513 , 5.424 14.5 , 22.9

**Range** 2.17 3.911 8.4

Box Plot of Points Box Plot of Score



**** Box Plot of Weigh

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:**  145.3333

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

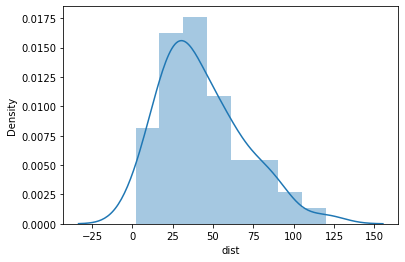
**Cars speed and distance**

**Use Q9\_a.csv**

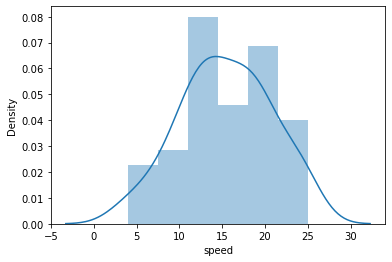
**Ans:** Speed Distance

Skewness **-0.11395477 0.78248351**

Kurtosis **-0.57714742 0.24801865**



**Distance Skewness**



**Speed Skewness**

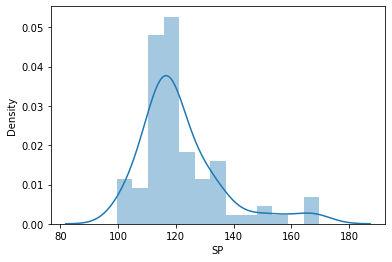
**SP and Weight(WT)**

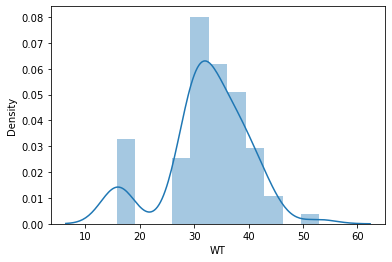
**Use Q9\_b.csv**

SP Weight

Skewness **1.581453 -0.603309**

Kurtosis **2.723521 0.819465**





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



**Ans:** The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.

**Ans:** The boxplot has outliers on the maximum side.

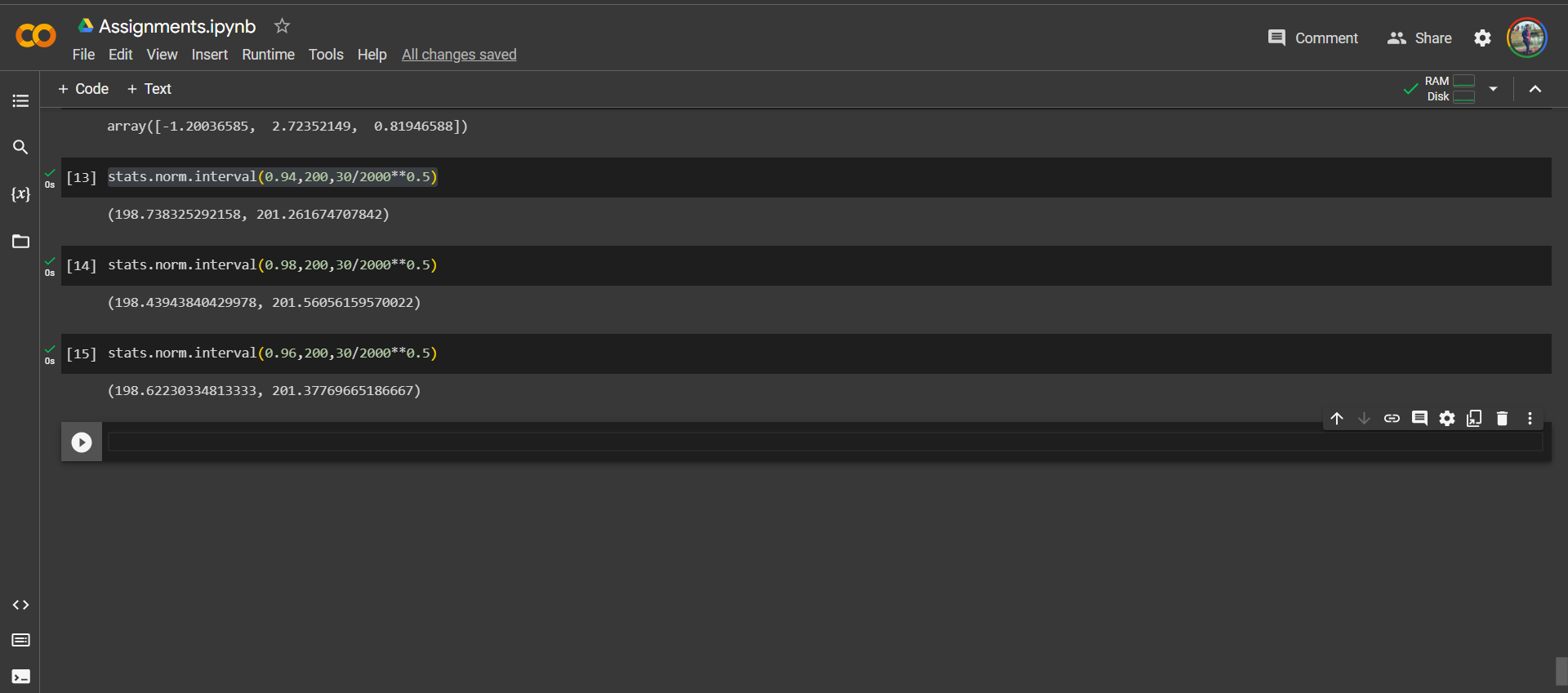
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men avnd weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate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

**Ans:**

Confidence Interval Z Range

94% 201.261674 198.738325

98% 201.560561 198.436438

 96% 201.377696 198.622303

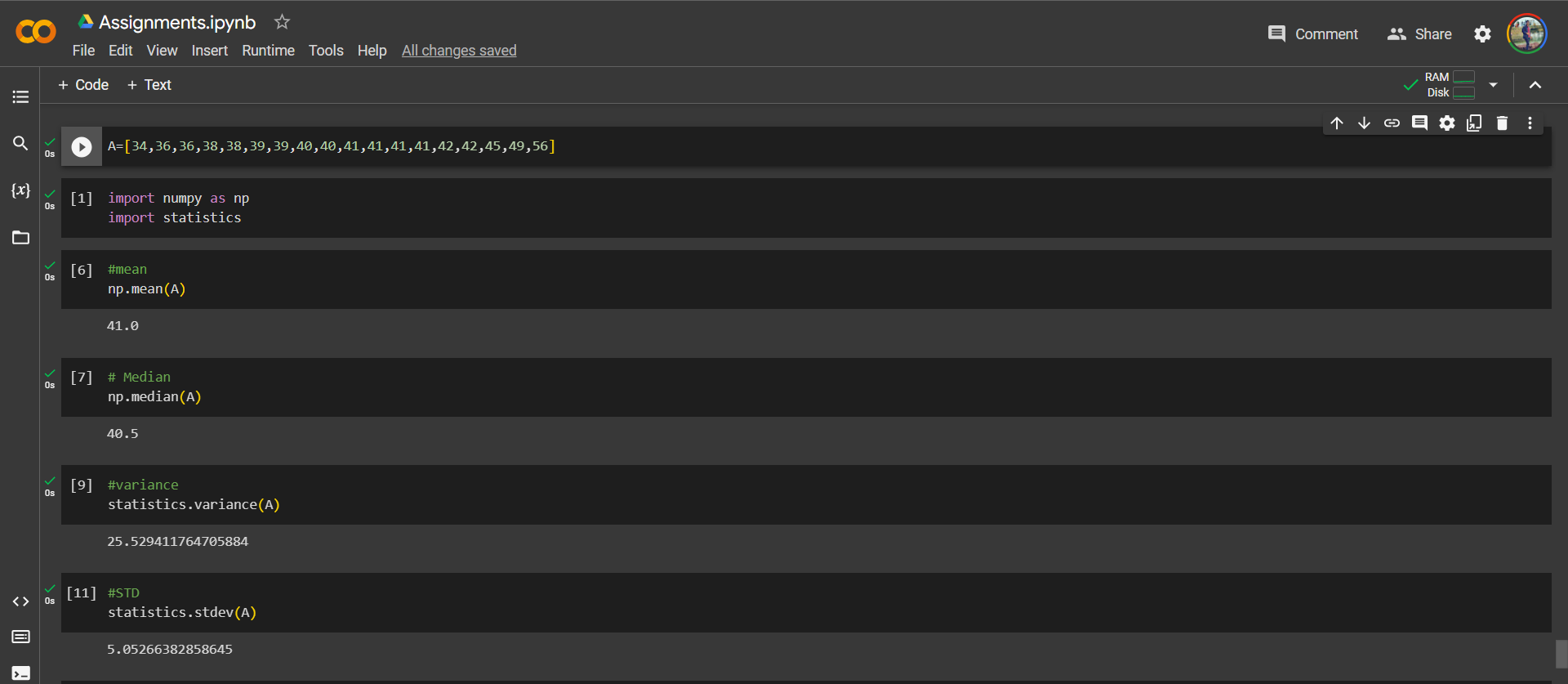
**Q12)** Below are the scores obtained by a student in tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

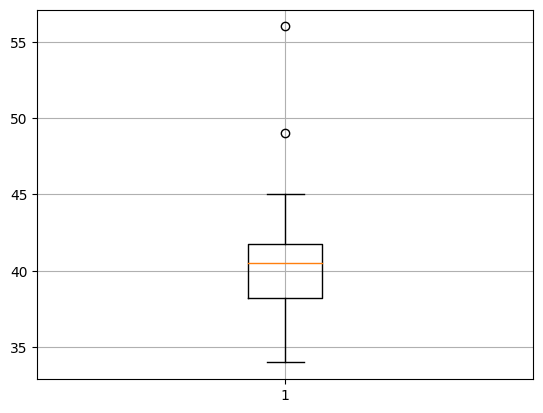
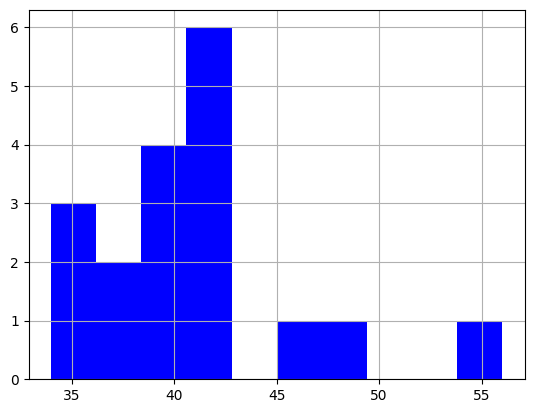
1. Find mean, median, variance, standard deviation.
2. What can we say about the student marks?

**Ans 1:** Mean = 41.0 Variance = 25.529411

Median = 40.5 SD = 5.0526638



**Ans 2:** Maximum number of students have obtained marks 41 to 42.

 Outliers are 49 and 59.

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

**Ans:** If the mean and median of a data is equal that means the data is symmetrically distributed so there will be no skewness.

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

**Ans:** When mean > median the distribution of data is positively skewed, so the positive skewness implies mass of the distribution concentrated on left side.

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

**Ans:** When mean < median the distribution of data is negatively skewed, so the negative skewness implies mass of the distribution concentrated on right side.

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

**Ans:** An extremely positive kurtosis indicates a distribution where more numbers are located in the tails of the distribution instead of around the mean.

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

**Ans:** A distribution with a negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

**Ans:**  Data is not normally distributed.

What is nature of skewness of the data?

**Ans:** This data has negative skewness.

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

**Ans:** IQR is 10 to 18.

Q19) Comment on the below Boxplot visualizations?



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

**Ans:** 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)

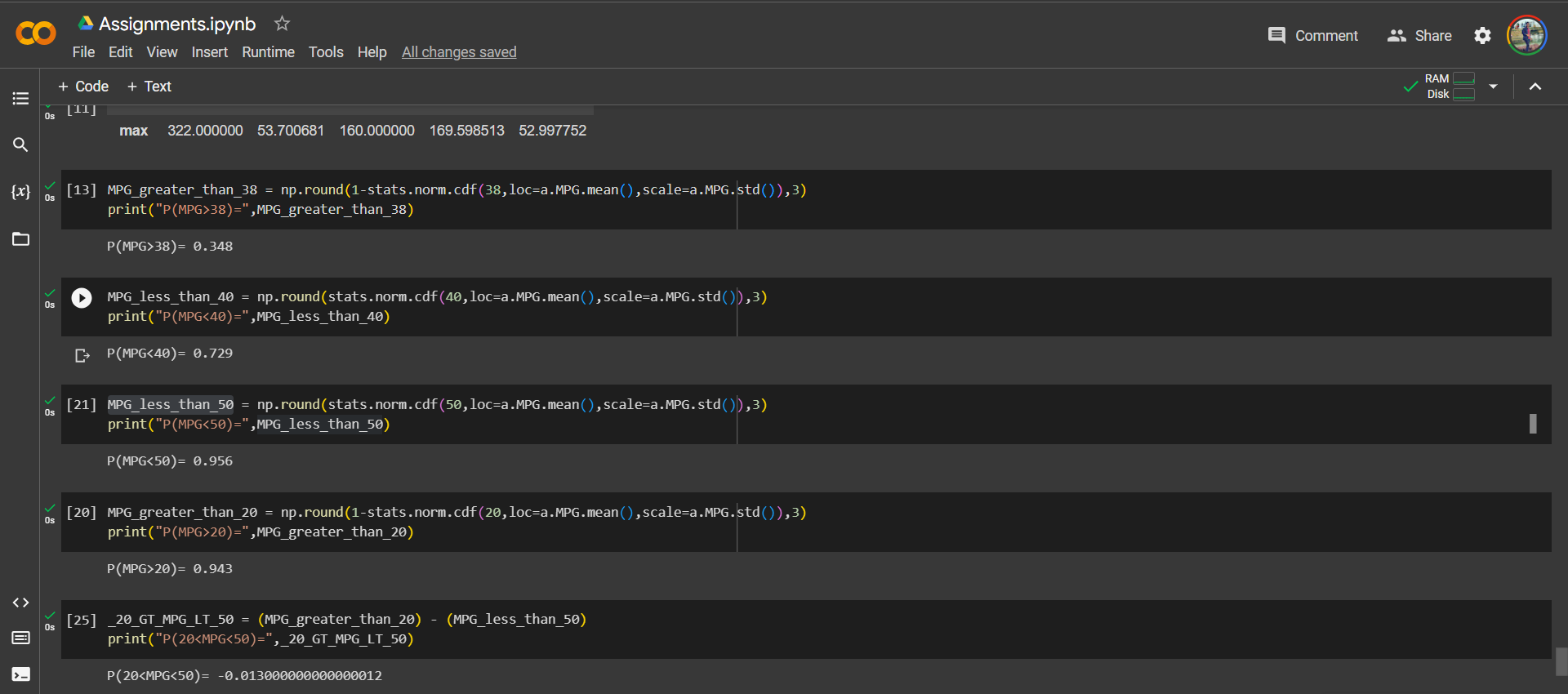
**Ans:** 0.348

* 1. P(MPG<40)

**Ans:** 0.729

* 1. P (20<MPG<50)

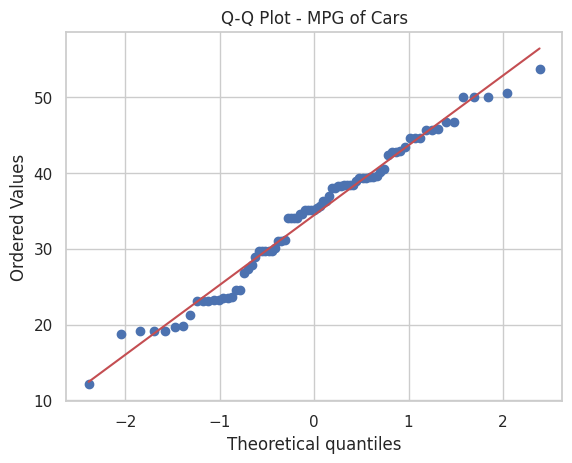
**Ans:** -0.013000000000000012



Q 21) Check whether the data follows normal distribution

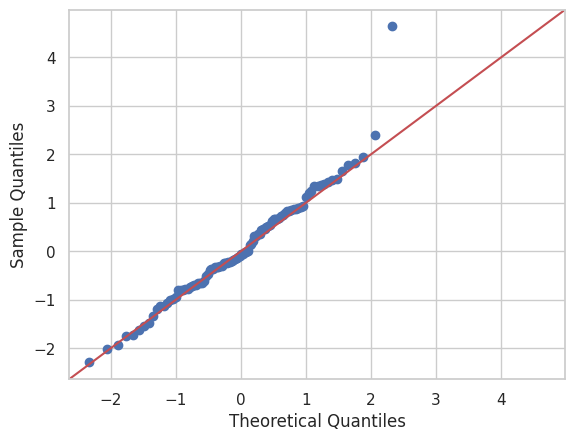
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans:**

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

Dataset: wc-at.csv

**Ans: **

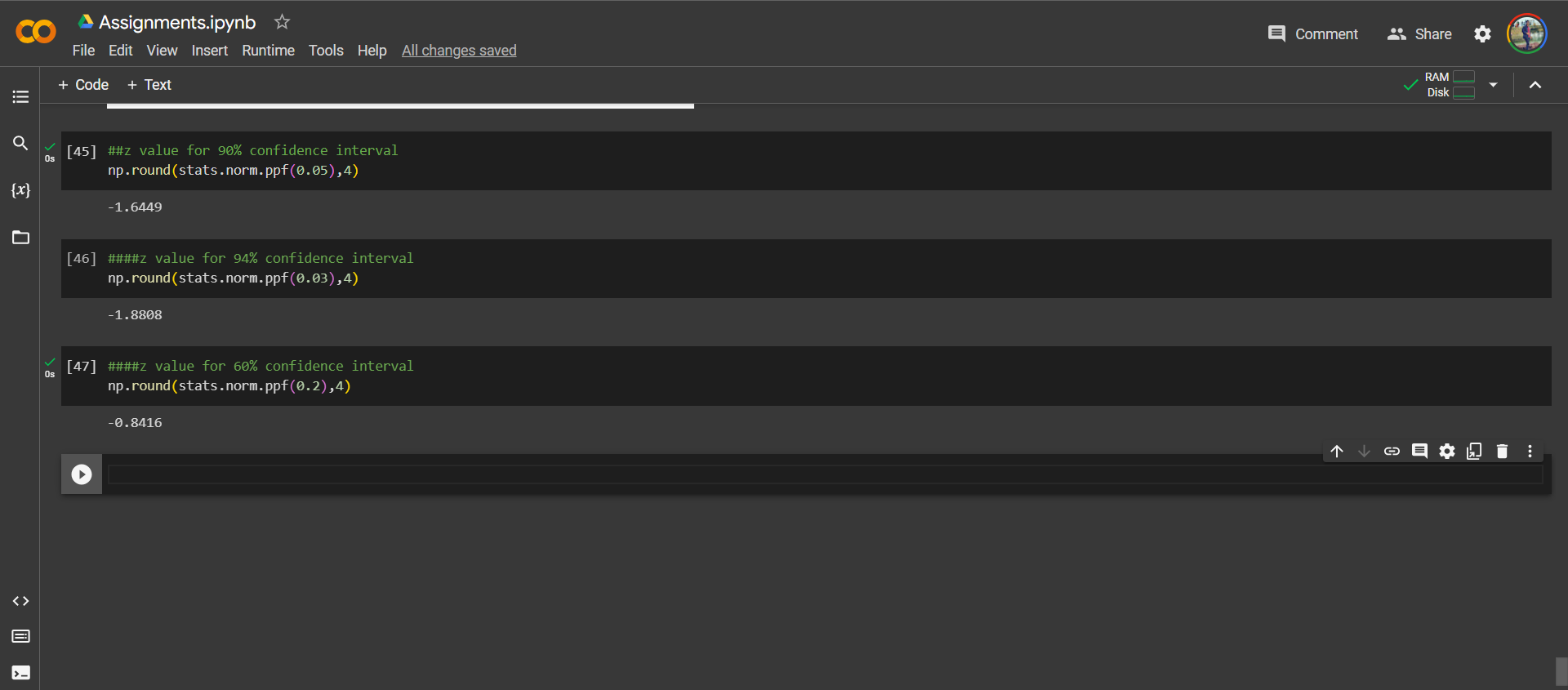
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**Ans:** Confidence Interval Z Score

**90% -1.6449**

**60% -0.8416**

**94% -0.8808**



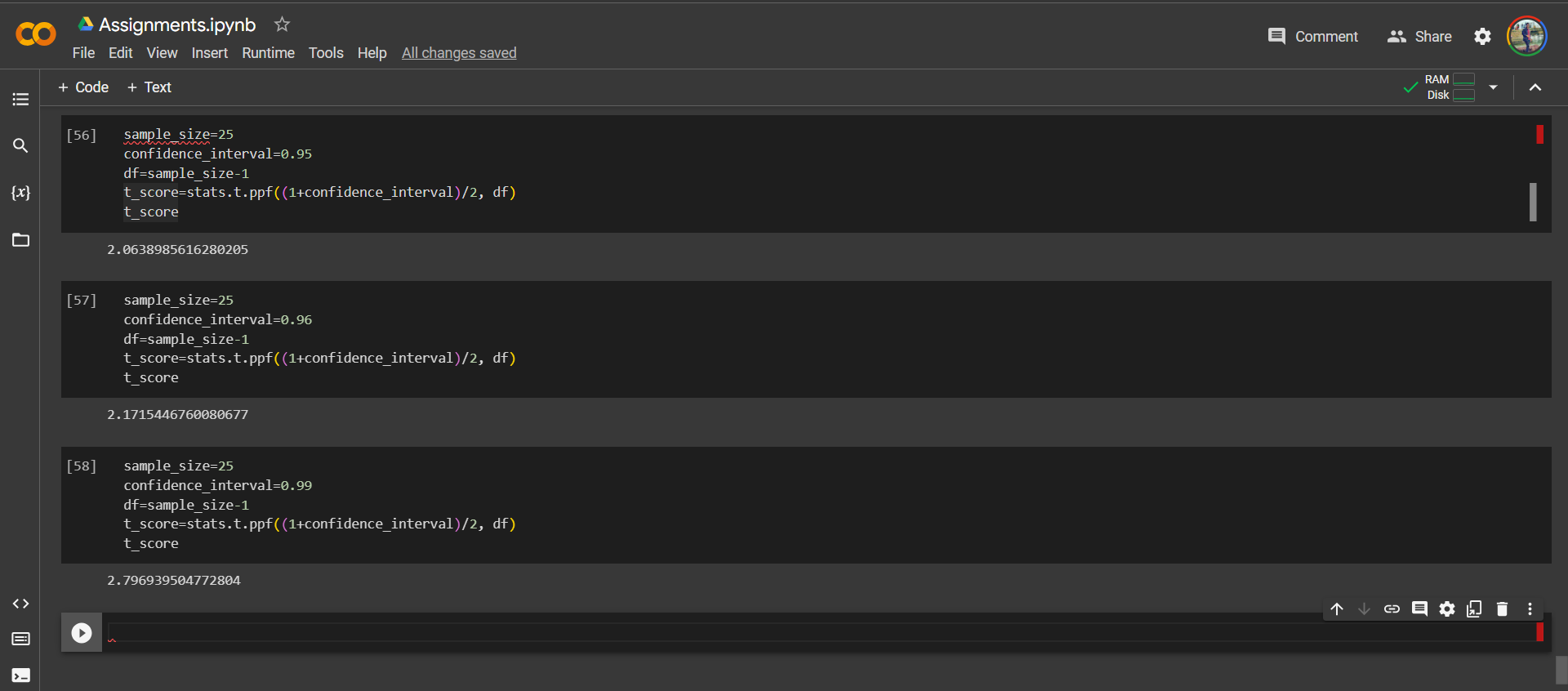
Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

**Ans:** Confidence Interval T Score

**95% 2.063898**

**96% 2.171544**

**99% 2.796939**

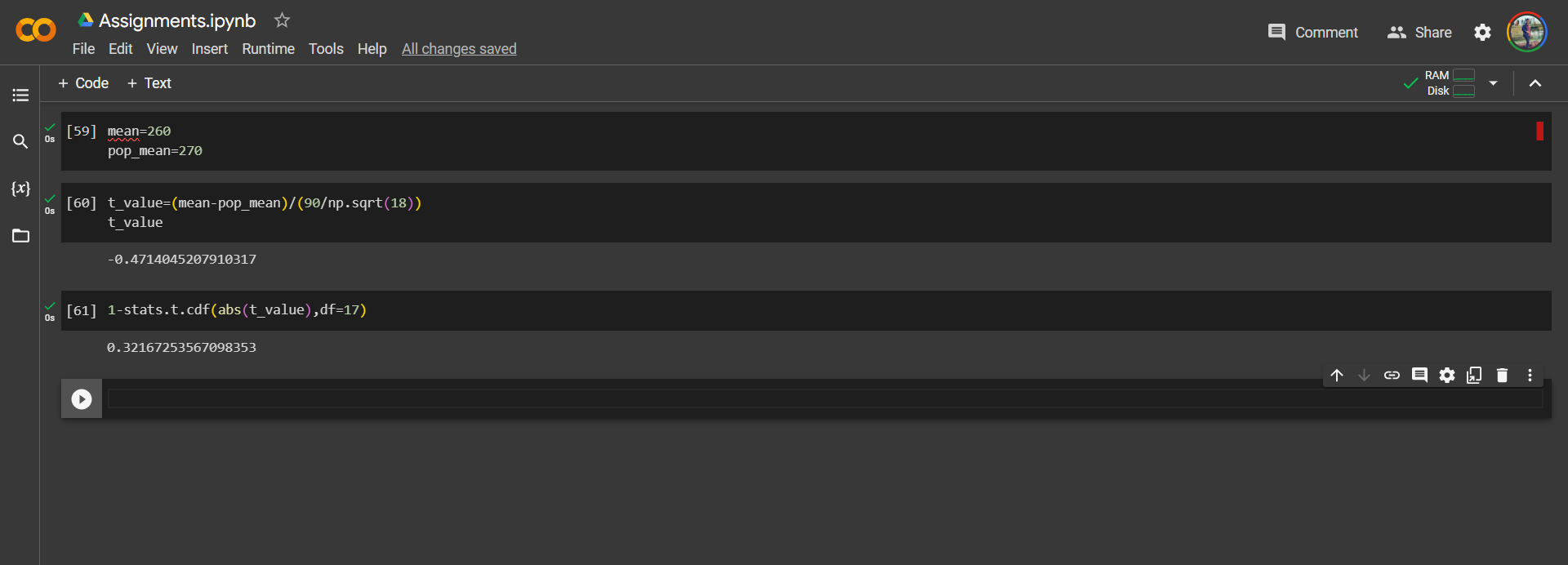
****

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 24: **