|  |  |
| --- | --- |
| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Nominal |

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

**Ans:** 3/8 = 37.5 %

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

1. Equal to 1

**Ans:** Sum cannot be 1. The minimum sum would be 2 if both dices rolled 1 and 1.

1. Less than or equal to 4

**Ans:** There are total 4 instances where the sum would be less than or equal to 4. That’s why the probability is 6/36 = 1/6.

1. Sum is divisible by 2 and 3

**Ans:** When 2 dices are rolled, the minimum sum we get is 2 and the max is 12. The numbers which are divisible by 2 and 3 are 6 and 12.

There are a total 5 instances where we can get the sum 6

[(1,5), (2,4), (3,3), (4,2), (5,1)]

There is only one instance where are can get the sum of 12, which is (6,6)

So, Total of 6 possibilities.

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:** P(number of balls drawn is blue) = P(the drawn ball is not blue) / P(2 balls are drawn randomly)

= 5! / 2! \* 3! **/** 7! / 2! \* 5!

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: 0.015+0.8+1.95+0.025+0.06+0.24 = 3.09

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

**Ans:**

Mean for Points = 3.59, Score = 3.21 and Weigh = 17.84

Median for Points = 3.69, Score = 3.32 and Weigh = 17.71

Mode for Points = 3.07, Score = 3.44 and Weigh = 17.02

Variance for Points = 0.28, Score = 0.95, Weigh = 3.19

Standard Deviation for Points = 0.53, Score = 0.97, Weigh = 1.78

Range [Min-Max] for Points [3.59 – 4.93], Score [3.21 – 5.42] and Weigh [17.84 – 22.9]

**Inferences:**







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:** (1/9)(108)+ (1/9)(110)+ (1/9)(123)+ (1/9)(134)+ (1/9)(145)+ (1/9)(167)+ (1/9)(187)+ (1/9)(199)

=145.33

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

**Cars, speed and distance**

**Use Q9\_a.csv**

**Ans:** Skewness Value:

speed -0.117510

dist 0.806895

Kurtosis Value:

speed -0.508994

dist 0.405053

**SP and Weight (WT)**

**Use Q9\_b.csv**

**Ans:** Skewness Value:

SP 1.611450

WT -0.614753

Kurtosis Value:

SP 2.977329

WT 0.950291

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



**Ans :** The given histogram is right skewed.



**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 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 :**

For 94% confidence interval Range is [ 198.73 – 201.26]

For 98% confidence interval range is [198.43 – 201.56]

For 96% confidence interval range is [198.62 – 201.37]

**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, Variance =25.52 and Standard Deviation =5.05

1. What can we say about the student marks?

**Ans :** There are no outliers and the data is slightly skewed towards right.

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

**Ans:** There is no skewness. The data is symmetrically distributed.

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

**Ans:** Towards right.

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

**Ans:** Towards left.

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

**Ans:** It suggests that the peak of the curve is high.

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

**Ans:** It suggests that the curve is flatter.

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



What can we say about the distribution of the data?

**Ans:** The given boxplot is not normally distributed where the median is high.

What is nature of skewness of the data?

**Ans:** Skewed towards left.

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

**Ans:** 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.

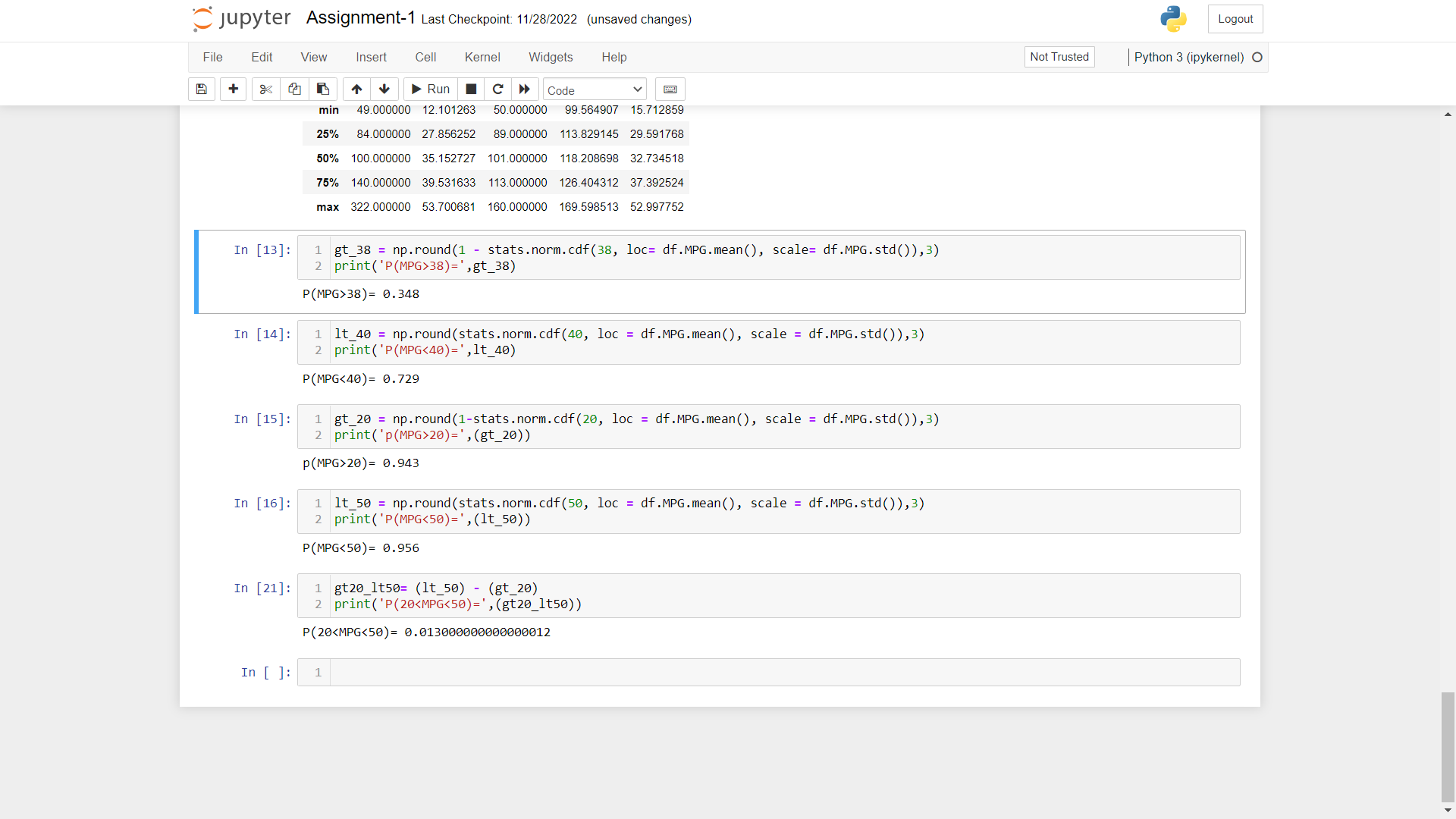
**Ans:** There are no outliers in the given distribution. The median of both of the boxplots is the same, ranging between 250-275. No skewness present.

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:** 0.013

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans:** The MPG of Cars 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:** Adipose Tissue (AT) and Waist does not follow Normal Distribution.





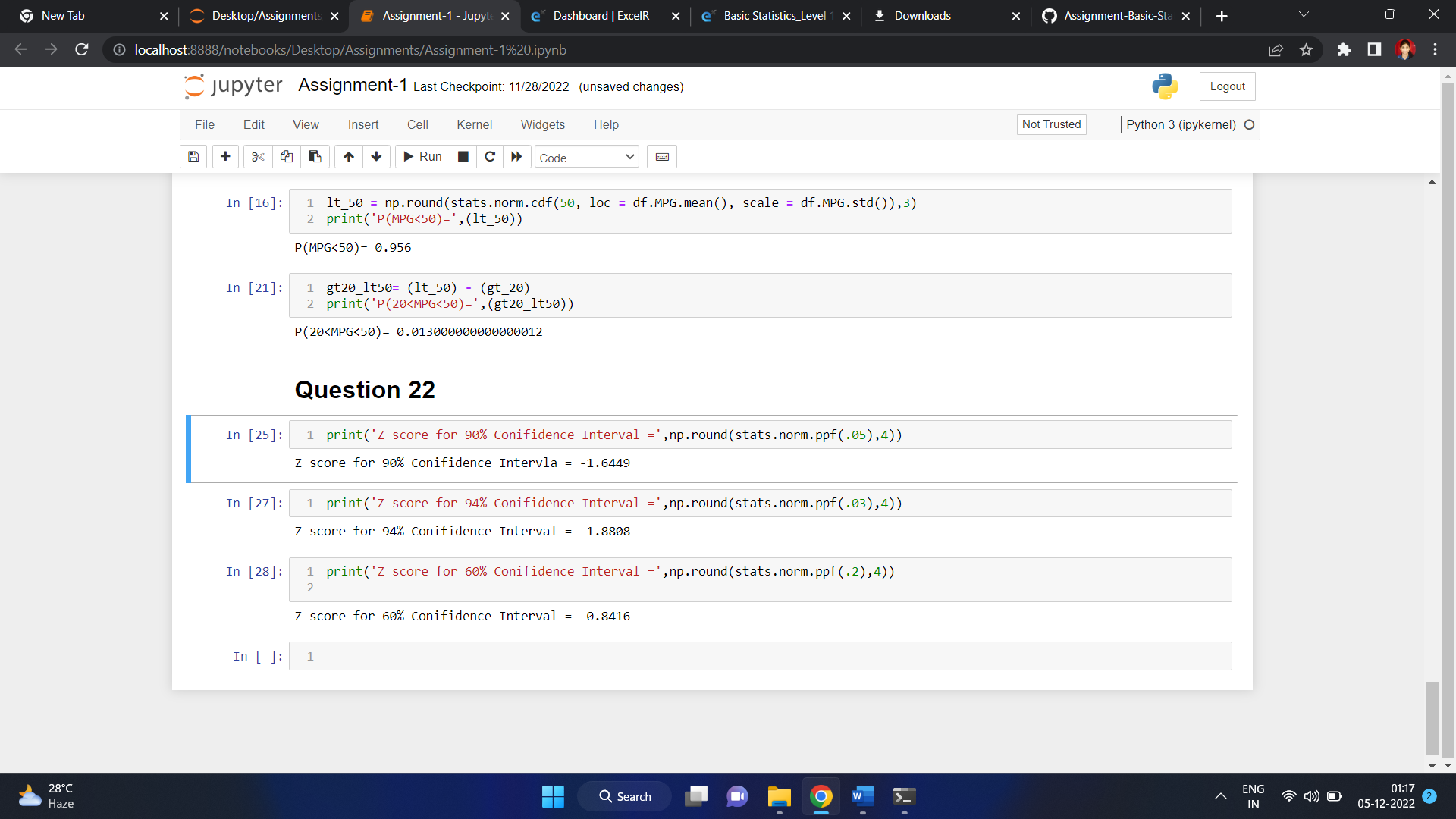
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval.

**Ans:**

90% Confidence interval = -1.6449

94% Confidence interval= -1.8808

60% Confidence interval= -0.8416



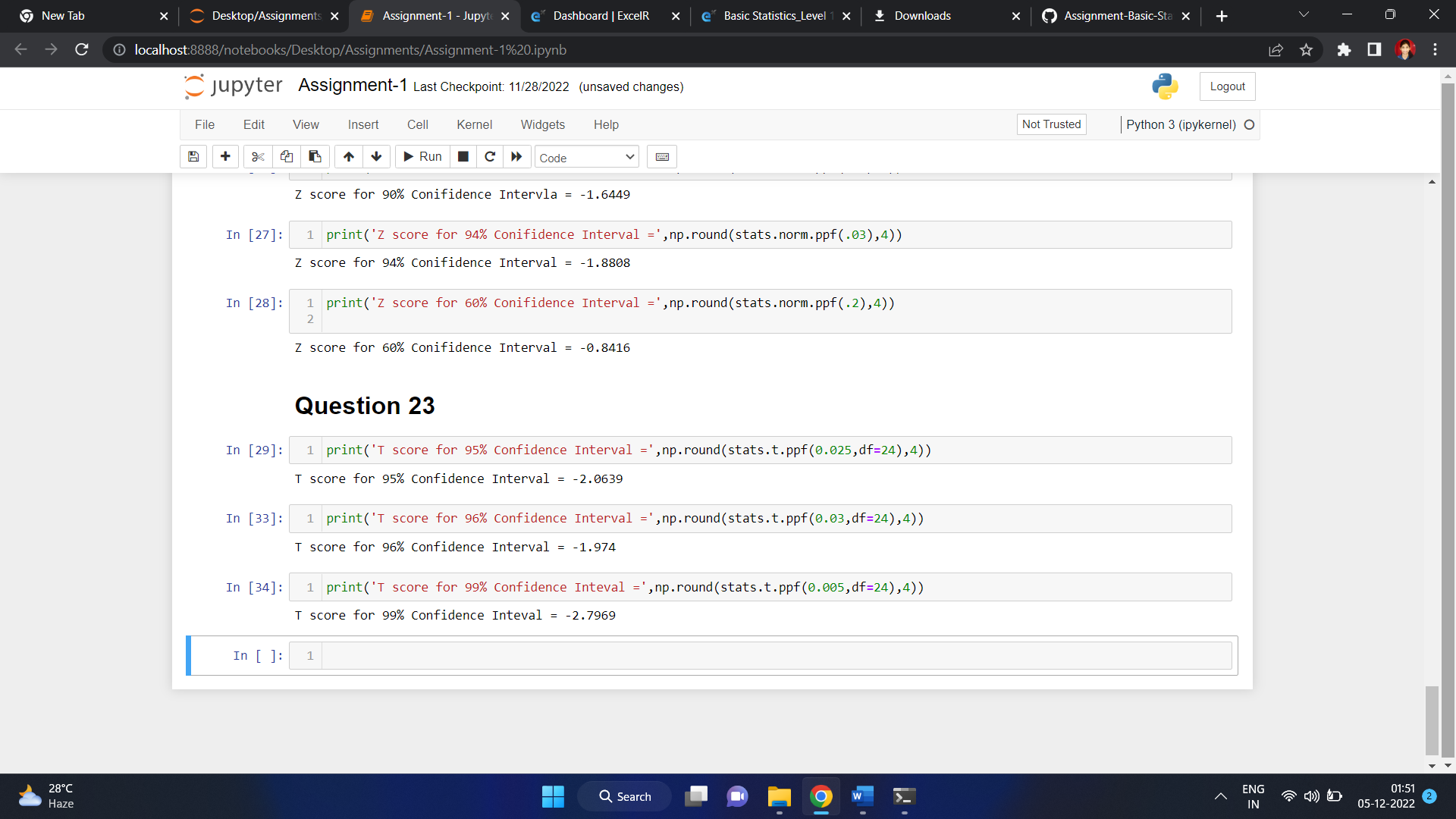
Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25.

**Ans:**

95% Confidence interval= -2.064

96% Confidence interval= -1.974

99% Confidence interval= -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:** 0.32 = 32%

