**Data Science Assignments**

**Assignment No. 02 – Basic Stats Level 02**

**Name : Pravin Dattatray Shinde**

**Batch :** 5th Aug 2024, Syeda Samreen

**Email : shindepravin1611@gmail.com**

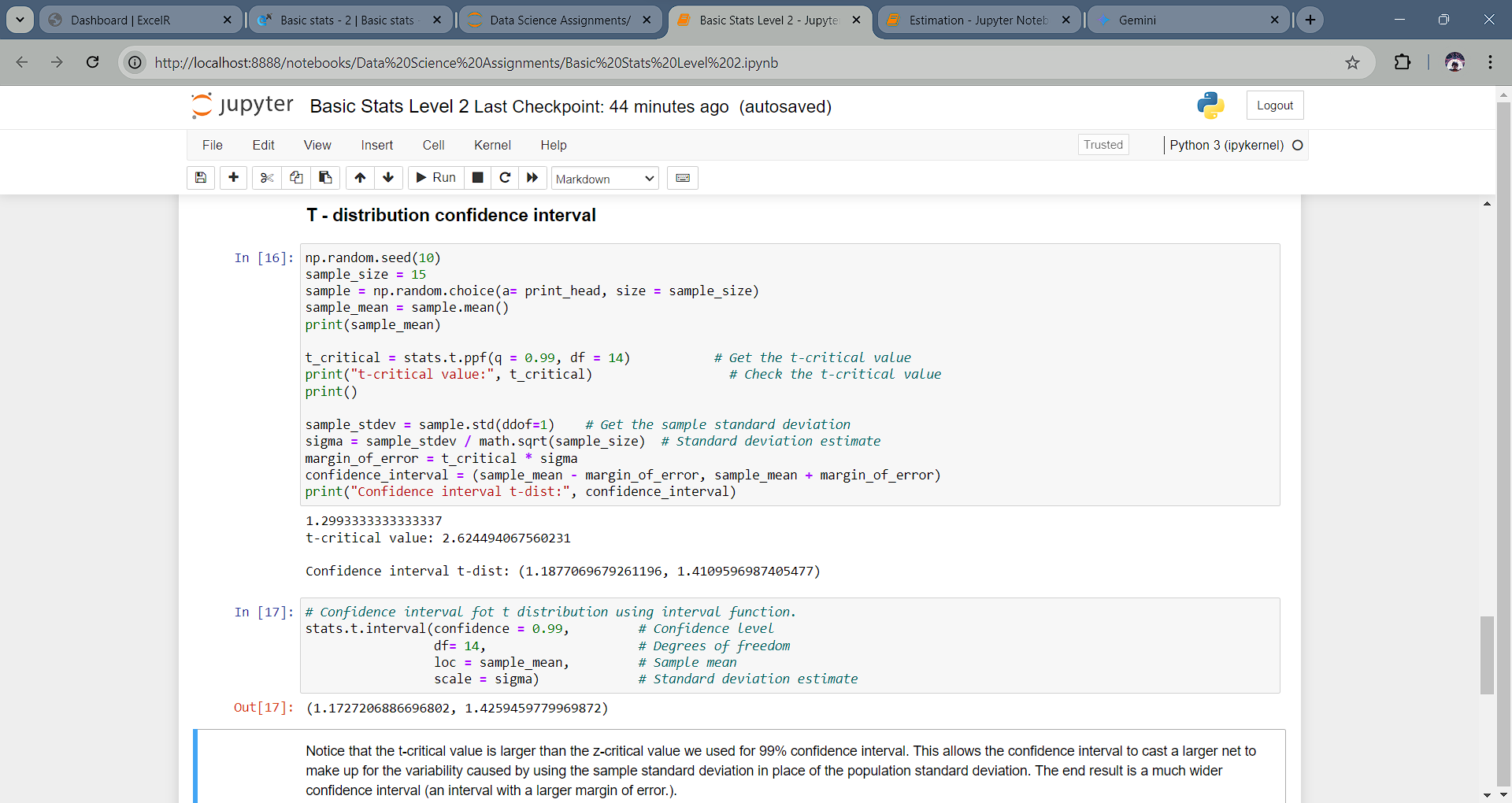
**Task 01 :** To build a 99% Confidence Interval Using Sample Standard Deviation.

Skewness and Its Impact on the Confidence Interval

The **slight left skewness** in the data, indicated by a skewness value of -0.4407, means that the data contains more extreme values on the lower end. While this skewness is not severe, it does suggest that the sample might not be perfectly symmetrical. In small samples, even slight skewness can influence the accuracy of statistical estimates, making it essential to use a distribution like the t-distribution, which adjusts for such variations.

Why the t-Critical Value is Larger than the z-Critical Value

For a given confidence level (e.g., 99%), the **t-critical value** is always larger than the corresponding **z-critical value**. This reflects the additional uncertainty when estimating the population standard deviation using the sample. In practical terms, the larger t-critical value results in a wider confidence interval, which provides a higher degree of coverage for the true population mean. This wider interval compensates for the greater variability inherent in smaller sample sizes.



* Rationale for using the t-distribution :

The t-distribution is used for constructing confidence intervals when:

1. The population standard deviation is unknown.
2. The sample size is small (typically less than 30).

In this case, we have a small sample of 15 print-heads and the population standard deviation is unknown. Therefore, the t-distribution is the appropriate choice for constructing the confidence interval.

* Steps Involved :

Calculate the Sample Mean (xˉ**xˉ)**: This provides the central point estimate for the population mean.

**Calculate the Sample Standard Deviation (s)**: The sample standard deviation is used as a substitute for the unknown population standard deviation.

Determine the Degrees of Freedom (df): For a sample of size n, the degrees of freedom are n−1. This is used to find the t-critical value.

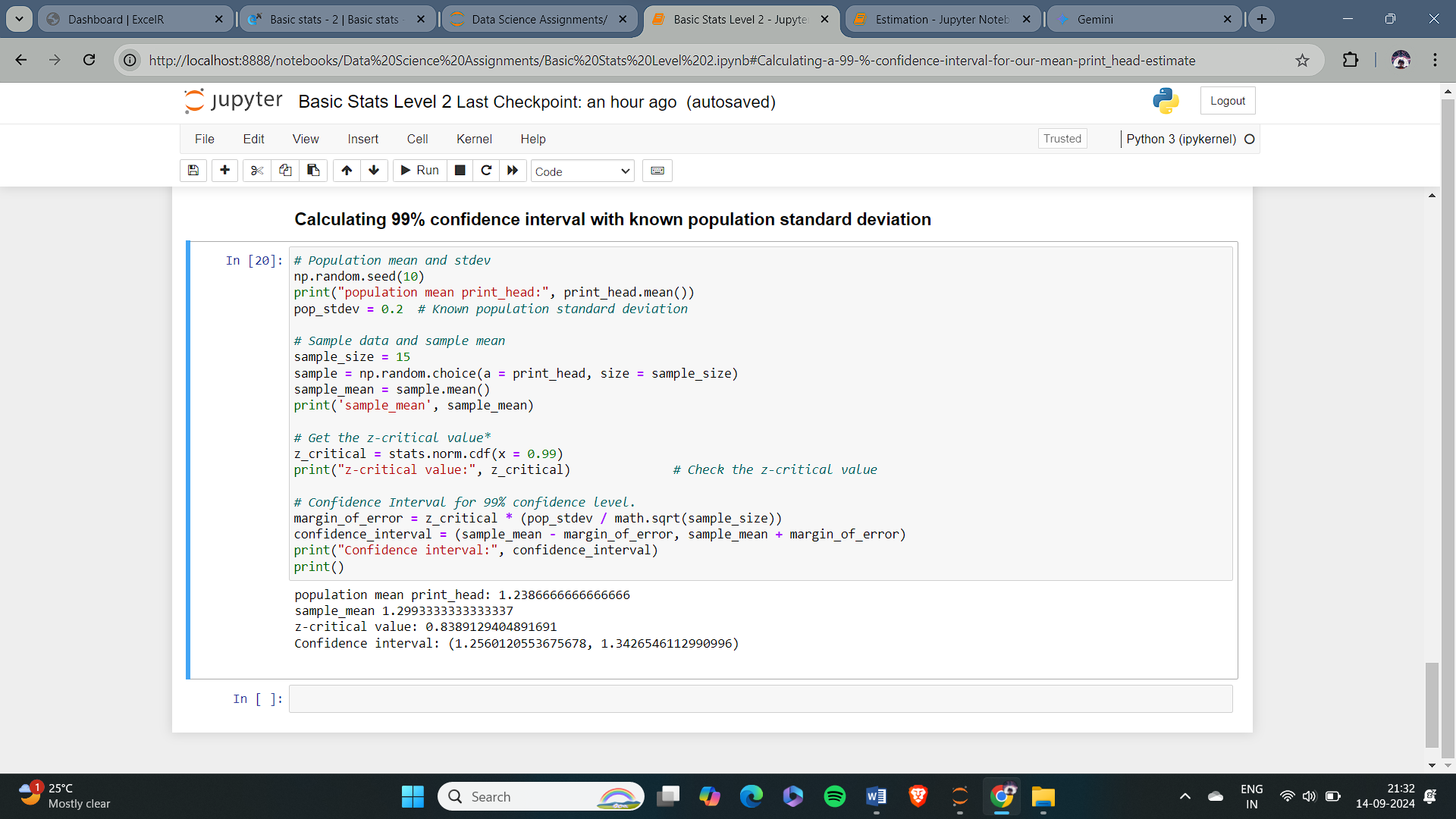
**Find the t-Critical Value**: Using the degrees of freedom and the desired confidence level (in this case, 99%), the t-critical value is obtained from a t-distribution table.

**Calculate the Margin of Error (ME)**: The margin of error is the product of the t-critical value and the standard error (the standard deviation of the sample divided by the square root of the sample size).

**Construct the Confidence Interval**: The confidence interval is given by xˉ±MExˉ±ME, providing the range within which the true population mean is likely to lie.

**Task 02 :** To build a 99% Confidence Interval Using Known Population Standard Deviation.

* We have a sample of 15 print-heads.
* We know the population standard deviation (σ) is 0.2 million characters.
* We want to construct a 99% confidence interval for the mean number of characters printed before failure.
* The z-critical value was found to be : 0.8389129404891691
* Confidence interval for the normal distribution: (1.2560120553675678, 1.3426546112990996)
* When the population standard deviation is known, regardless of sample size, we use the normal distribution to construct confidence intervals.



* Steps Involved :

1. Calculate the sample mean (x̄):
2. Determine the z-value for the desired confidence level:
3. Calculate the margin of error (ME):
4. Construct the confidence interval:

* Key Difference :

When the population standard deviation is known, we use the z-distribution and the margin of error formula is based on the z-value.

When the population standard deviation is unknown, we use the t-distribution and the margin of error formula is based on the t-value.