Q1: What is Estimation Statistics? Explain point estimate and interval estimate.

Estimation Statistics is a branch of statistics that involves making informed guesses or estimates about population parameters (characteristics) based on data obtained from a sample of that population. There are two main types of estimation in statistics:

Point Estimate: A point estimate is a single value that is used to approximate the true population parameter. For example, the sample mean is a point estimate of the population mean.

Interval Estimate: An interval estimate, often referred to as a confidence interval, provides a range of values within which the true population parameter is likely to fall. It gives a level of confidence regarding the estimate's accuracy, typically represented as a range with an associated confidence level (e.g., a 95% confidence interval).

Q2. Write a Python function to estimate the population mean using a sample mean and standard deviation.

In [1]:
```
import math

def estimate_population_mean(sample_mean, sample_stddev, sample_size):
    # Calculate the standard error of the mean
    standard_error = sample_stddev / math.sqrt(sample_size)

    # Calculate the margin of error for a desired confidence level (e.g., 95%)
    confidence_level = 0.95
    z_score = 1.96  # For a 95% confidence interval (you can use a Z-table or li
    margin_of_error = z_score * standard_error

    # Calculate the confidence interval
    lower_bound = sample_mean - margin_of_error
    upper_bound = sample_mean + margin_of_error

    return (lower_bound, upper_bound)

# Example usage:
sample_mean = 100
sample_stddev = 15
sample_size = 30
confidence_interval = estimate_population_mean(sample_mean, sample_stddev, sampl
print("95% Confidence Interval:", confidence_interval)
```

```
95% Confidence Interval: (94.63231893644937, 105.36768106355063)
```

Q3: What is Hypothesis testing? Why is it used? State the importance of Hypothesis testing.

Hypothesis testing is a statistical method used to make inferences about a population based on a sample of data. It involves formulating two competing hypotheses, the null hypothesis (H0) and the alternative hypothesis (Ha), and then conducting a statistical test to determine whether there is enough evidence to reject the null hypothesis in favor of the alternative hypothesis.

Hypothesis testing is used to:

Make decisions about population parameters. Assess the significance of observed effects. Test scientific theories or claims. Compare groups or treatments.

Q4. Create a hypothesis that states whether the average weight of male college students is greater than the average weight of female college students.

Hypothesis: The average weight of male college students is greater than the average weight of female college students.

Null Hypothesis (H0): μ_male <= μ_female Alternative Hypothesis (Ha): μ_male > μ_female

Here, μ_male represents the average weight of male college students, and μ_female represents the average weight of female college students.

Q5. Write a Python script to conduct a hypothesis test on the difference between two population means, given a sample from each population.

In [ ]:
```
import scipy.stats as stats

# Sample data from two populations
sample1 = [...]
sample2 = [...]

# Perform a two-sample t-test
t_stat, p_value = stats.ttest_ind(sample1, sample2)

# Define the significance level (alpha)
alpha = 0.05

# Check if the p-value is less than alpha to determine statistical significance
if p_value < alpha:
    print("Reject the null hypothesis")
else:
    print("Fail to reject the null hypothesis")

# You can also calculate the confidence interval for the difference of means
# using stats.t.interval() if needed.
```

Q6: What is a null and alternative hypothesis? Give some examples.

In hypothesis testing, you have two main hypotheses:

Null Hypothesis (H0): This is the default or status quo hypothesis. It typically represents the absence of an effect or no difference. Researchers aim to test and potentially reject the null hypothesis.

Alternative Hypothesis (Ha): This is the hypothesis you want to support or prove. It represents a specific effect, difference, or relationship you are trying to demonstrate.

Examples:

Null Hypothesis (H0): The mean exam score of students who received tutoring is equal to the mean exam score of students who did not receive tutoring. Alternative Hypothesis (Ha): The mean exam score of students who received tutoring is not equal to the mean exam score of students who did not receive tutoring.

Null Hypothesis (H0): The new drug has no effect on blood pressure. Alternative Hypothesis (Ha): The new drug decreases blood pressure.

Q7: Write down the steps involved in hypothesis testing.

Steps involved in hypothesis testing:

Formulate Hypotheses:

Null Hypothesis (H0): Represents the default assumption. Alternative Hypothesis (Ha): Represents the claim or effect you want to test. Collect Data: Gather relevant data through experiments, surveys, or observations.

Choose Significance Level (α): Select a predefined level of significance that determines how much evidence is required to reject the null hypothesis (common values include 0.05 or 0.01).

Perform Statistical Test: Choose an appropriate statistical test (e.g., t-test, z-test, chi-squared test) based on the data and research question.

Calculate Test Statistic: Compute the test statistic from the sample data.

Determine P-Value: Calculate the p-value, which represents the probability of observing the data or more extreme results if the null hypothesis is true.

Make a Decision: Compare the p-value to the chosen significance level (α) and decide whether to reject the null hypothesis or fail to reject it.

Draw Conclusions: Interpret the results and make inferences about the population based on the decision made in step 7.

Q8. Define p-value and explain its significance in hypothesis testing.

: The p-value, or probability value, is a crucial concept in hypothesis testing. It represents the probability of observing the data or more extreme results if the null hypothesis is true. In other words, it quantifies the evidence against the null hypothesis. The significance of the p-value in hypothesis testing is as follows:

If the p-value is small (typically less than or equal to the chosen significance level α), it indicates strong evidence against the null hypothesis. In this case, you may reject the null hypothesis in favor of the alternative hypothesis.

If the p-value is large (greater than α), it suggests weak or insufficient evidence against the null hypothesis. In this case, you would fail to reject the null hypothesis.

The smaller the p-value, the stronger the evidence against the null hypothesis, and the more significant the results are considered.

The choice of the significance level α (e.g., 0.05 or 0.01) determines the threshold for considering a p-value as evidence against the null hypothesis.

Q10. Write a Python program to calculate the two-sample t-test for independent samples, given two random samples of equal size and a null hypothesis that the population means are equal.

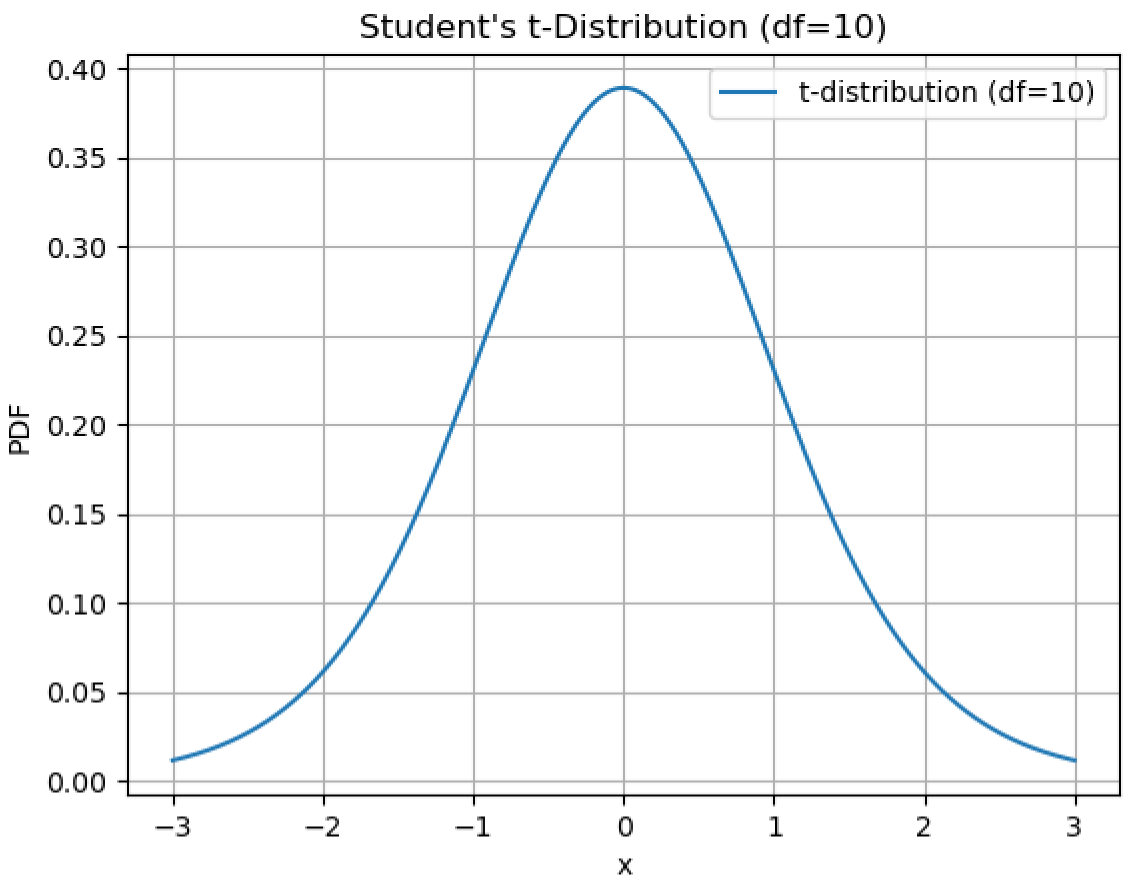
In [3]:
```
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import t

# Degrees of freedom
df = 10

# Create an array of x values
x = np.linspace(-3, 3, 1000)

# Calculate the probability density function (PDF) for the t-distribution
pdf = t.pdf(x, df)

# Plot the t-distribution
plt.plot(x, pdf, label=f't-distribution (df={df})')
plt.title(f'Student\'s t-Distribution (df={df})')
plt.xlabel('x')
plt.ylabel('PDF')
plt.legend()
plt.grid(True)
plt.show()
```



Q11: What is Student’s t distribution? When to use the t-Distribution.

In [ ]:
```python
import scipy.stats as stats

# Sample data from two populations
sample1 = [...]
sample2 = [...]

# Perform a two-sample t-test for independent samples
t_stat, p_value = stats.ttest_ind(sample1, sample2)

# Define the significance level (alpha)
alpha = 0.05

# Check if the p-value is less than alpha to determine statistical significance
if p_value < alpha:
    print("Reject the null hypothesis")
else:
    print("Fail to reject the null hypothesis")
```

Q12: What is t-statistic? State the formula for t-statistic.

Student's t-distribution, often referred to as the t-distribution, is a probability distribution that is similar in shape to the standard normal distribution (z-distribution) but has heavier tails. It is used when the sample size is small, and the population standard deviation is unknown. The t-distribution is characterized by its degrees of freedom (df), which determine its shape.

When to use the t-Distribution:

The t-distribution is used when dealing with small sample sizes (typically $n < 30$) and when the population standard deviation is unknown. It is commonly employed in hypothesis testing and confidence interval estimation when working with sample data.

Q13. A coffee shop owner wants to estimate the average daily revenue for their shop. They take a random sample of 50 days and find the sample mean revenue to be $500 with a standard deviation of $50. Estimate the population mean revenue with a 95% confidence interval.

The t-statistic is a measure used in hypothesis testing to determine if the sample data provides enough evidence to reject the null hypothesis. The formula for the t-statistic when comparing a sample mean (x̄) to a population mean (μ) is:

� = � ¯ − � � / √ � t= s/√n x ¯ −μ

Where:

� ¯ x ¯ is the sample mean. � μ is the population mean under the null hypothesis. � s is the sample standard deviation. � n is the sample size. The t-statistic quantifies how many standard errors the sample mean is away from the population mean. A larger t-statistic indicates stronger evidence against the null hypothesis.

Q14. A researcher hypothesizes that a new drug will decrease blood pressure by 10 mmHg. They conduct a clinical trial with 100 patients and find that the sample mean decrease in blood pressure is 8 mmHg with a standard deviation of 3 mmHg. Test the hypothesis with a significance level of 0.05.

To estimate the population mean revenue with a 95% confidence interval, you can use the formula for a confidence interval:

\text{Confidence Interval} = \text{Sample Mean} ± \left(\frac{{\text{Critical Value} × \text{Standard Error}}}{{\sqrt{\text{Sample Size}}}}\right)

Given the data:

Sample Mean (� ¯ x ¯): $500 Standard Deviation (� s): $50 Sample Size (� n): 50 Confidence Level: 95% First, you need to find the critical value for a 95% confidence level. You can look up this value in a t-table or use a calculator, or you can use the Z-score since the sample size is reasonably large (for a Z-distribution, the critical value for 95% confidence is approximately 1.96).

Now, calculate the standard error: Standard Error = � � = 50 50 Standard Error= n

s= 50

50

Substitute these values into the formula to calculate the confidence interval:

\text{Confidence Interval} = $500 ± (1.96 × \frac{50}{\sqrt{50}})

Calculate the upper and lower bounds of the confidence interval to estimate the population mean revenue.

Q15. An electronics company produces a certain type of product with a mean weight of 5 pounds and a standard deviation of 0.5 pounds. A random sample of 25 products is taken, and the sample mean weight is found to be 4.8 pounds. Test the hypothesis that the true mean weight of the products is less than 5 pounds with a significance level of 0.01.

: To test the hypothesis that the new drug decreases blood pressure by 10 mmHg with a significance level of 0.05, you can perform a one-sample t-test.

Given data:

Sample Mean (� ¯ x ¯): 8 mmHg Standard Deviation (� s): 3 mmHg Sample Size (� n): 100 Null Hypothesis (H0): � = 10 μ=10 (No change in blood pressure) Significance Level (α): 0.05 Perform the t-test as follows:

Calculate the t-statistic: � = � ¯ − � � / √ � = 8 − 10 3 / √ 100 = − 10 t= s/√n x ¯ −μ= 3/√100 8−10=−10

Find the critical t-value for a one-tailed test at α = 0.05 with 99 degrees of freedom (n - 1): You can use a t-table or a t-distribution calculator to find the critical t-value. For a one-tailed test at α = 0.05 and 99 degrees of freedom, it's approximately -1.6607.

Compare the calculated t-statistic to the critical t-value: Since -10 is more extreme than -1.6607, you would reject the null hypothesis.

Draw the conclusion: You have enough evidence to conclude that the new drug decreases blood pressure by more than 10 mmHg.

Q16. Two groups of students are given different study materials to prepare for a test. The first group (n1 = 30) has a mean score of 80 with a standard deviation of 10, and the second group (n2 = 40) has a mean score of 75 with a standard deviation of 8. Test the hypothesis that the population means for the two groups are equal with a significance level of 0.01.

To test the hypothesis that the population means for the two groups are equal with a significance level of 0.01, you can perform a two-sample t-test.

Given data for Group 1:

Sample Mean (� ¯ 1 x ¯

1): 80 Standard Deviation (� 1 s 1): 10 Sample Size (� 1 n 1): 30 Given data for Group 2:

Sample Mean (� ¯ 2 x ¯

2): 75 Standard Deviation (� 2 s 2): 8 Sample Size (� 2 n 2): 40 Null Hypothesis (H0): � 1 = � 2 μ 1=μ 2(The population means for the two groups are equal) Alternative

Hypothesis (Ha): � 1 ≠ � 2 μ 1

 =μ 2(The population means for the two groups are not equal)

Significance Level (α): 0.01

Perform the two-sample t-test as follows:

Calculate the pooled standard error: SE = \sqrt{\frac{{(s_1^2 / n_1) + (s_2^2 / n_2)}}}

Calculate the t-statistic: � = (� ˉ 1 − � ˉ 2) � � t= SE (x ˉ

1− x ˉ

2)

Find the critical t-value for a two-tailed test at α = 0.01 with degrees of freedom df = \frac{{(n_1 + n_2 - 2)}}: Use a t-table or a t-distribution calculator to find the critical t-value.

Compare the calculated t-statistic to the critical t-value: If the calculated t-statistic is more extreme (either significantly larger or significantly smaller) than the critical t-value, you would reject the null hypothesis.

Draw the conclusion: Based on the comparison, you can determine whether there is enough evidence to conclude that the population means for the two groups are not equal at the 0.01 significance level.

Q17. A marketing company wants to estimate the average number of ads watched by viewers during a TV program. They take a random sample of 50 viewers and find that the sample mean is 4 with a standard deviation of 1.5. Estimate the population mean with a 99% confidence interval.

To estimate the average number of ads watched by viewers during a TV program with a 99% confidence interval, you can use the formula for a confidence interval as follows:

\text{Confidence Interval} = \text{Sample Mean} ± \left(\frac{{\text{Critical Value} × \text{Standard Error}}}{{\sqrt{\text{Sample Size}}}\right)

Given the data:

Sample Mean (� ˉ x ˉ): 4 Standard Deviation (� s): 1.5 Sample Size (� n): 50 Confidence Level: 99% First, you need to find the critical value for a 99% confidence level. You can look up this value in a t-table or use a calculator.

Calculate the standard error: Standard Error = � � = 1.5 50 Standard Error= n

s= 50

1.5

Substitute these values into the formula to calculate the confidence interval:

Confidence Interval = 4 ± (Critical Value × 1.5 50) Confidence Interval=4±(Critical Value× 50

1.5)

Calculate the upper and lower bounds of the confidence interval to estimate the population mean.