

Assignment

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

Ans: Estimation statistics involves the use of sample data to estimate or make inferences about population parameters. There are two main types of estimates: point estimates and interval estimates.

Point Estimate:

- A point estimate is a single value that is used to estimate the corresponding population parameter.
- For example, if you want to estimate the average height of students in a school, the sample mean height would be a point estimate of the population mean height.

Interval Estimate:

- An interval estimate provides a range within which the true value of the population parameter is likely to fall.
- It recognizes the uncertainty inherent in estimating a parameter based on a sample and provides a range of plausible values.
- Commonly, interval estimates are expressed as confidence intervals. A confidence interval gives a range of values and a level of confidence that the true parameter value falls within that range.
- For instance, you might say, "I am 95% confident that the true average height of students is between X and Y inches."

In summary, estimation statistics involves using sample data to make educated guesses about population parameters. Point estimates provide a single best guess, while interval estimates provide a range of plausible values along with a level of confidence about the accuracy of that range.

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

Ans: To estimate the population mean using a sample mean and standard deviation, you can use the formula for a confidence interval. The formula for the confidence interval for the population mean (

μ) is:

Confidence Interval = $(\bar{x} - z \cdot \frac{s}{\sqrt{n}}, \bar{x} + z \cdot \frac{s}{\sqrt{n}})$

Confidence Interval=(

\bar{x}

-

$-Z$

n

s

,

\bar{x}

-

$+Z$

n

s

)

Where:

• σ^2

• \bar{x}

• $-$

• is the sample mean.

• σ^2

- s is the sample standard deviation.
- σ
- n is the sample size.
- σ
- Z is the Z-score corresponding to the desired confidence level.

Here's a Python function that takes these parameters as input and calculates the confidence interval:

python

Copy code

```
import sys
import math

def estimate_population_mean(sample_mean, sample_std_dev, sample_size, confidence_level):
    # Calculate the Z-score for the given confidence level
    z_score = 1.96  # For 95% confidence level

    # Calculate the margin of error
    margin_of_error = z_score * (sample_std_dev / math.sqrt(sample_size))

    # 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 = 25.0
sample_std_dev = 5.0
sample_size = 100
confidence_level = 0.95

print f"Estimated population mean: {sample_mean:.2f}"
print f"Confidence Interval: ({interval[0]:.2f}, {interval[1]:.2f})"
```

Make sure to adjust the values of `sample_mean`, `sample_std`, `sample_size`, and `confidence_level` according to your specific case. The result will be a confidence interval for the estimated population mean.

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

Ans: Hypothesis testing is a statistical method used to make inferences about population parameters based on a sample of data. It involves formulating two competing hypotheses about a population parameter, collecting data, and using statistical techniques to determine which hypothesis is more supported by the evidence.

The process typically involves the following steps:

Formulating Hypotheses:

- Null Hypothesis (H_0)
- H_0
- 0
-)
-): It represents a default or status quo assumption. It often asserts that there is no effect or no difference.
- Alternative Hypothesis (H_1)
- H_1
- 1
- or
- H_a
- H_a
- a
-)
-): It represents a challenge to the null hypothesis. It suggests that there is a significant effect or difference.

Collecting Data:

- Data is collected through experiments, surveys, or other methods.

Statistical Analysis:

- Statistical tests are applied to the data to determine the likelihood of observing the results under the assumption that the null hypothesis is true.

Drawing Conclusions:

- Based on the statistical analysis, a decision is made regarding whether to reject the null hypothesis in favor of the alternative hypothesis.

Interpretation:

- The results are interpreted in the context of the specific problem or research question.

Importance of Hypothesis Testing:

Scientific Rigor:

- Hypothesis testing provides a structured and systematic approach to evaluating hypotheses, contributing to the scientific rigor of research.

Decision Making:

- It helps in making informed decisions by providing a formal process for comparing competing hypotheses.

Inference to Population:

- Through hypothesis testing, conclusions drawn from a sample can be extended to the larger population.

Quality Control:

- Hypothesis testing is widely used in industries for quality control, ensuring that products meet specified standards.

Policy and Management:

- It is employed in policy analysis and management decisions to assess the effectiveness of interventions or strategies.

Scientific Research:

- In scientific research, hypothesis testing is fundamental for confirming or refuting hypotheses and theories.

Risk Management:

- It aids in assessing and managing risks by providing a statistical basis for decision-making.

Continuous Improvement:

- In various fields, hypothesis testing is integrated into processes of continuous improvement, ensuring ongoing effectiveness.

Overall, hypothesis testing is a powerful tool for drawing conclusions from data, guiding decision-making, and advancing knowledge in various fields.

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

Ans: formulate the hypothesis to test whether the average weight of male college students is greater than the average weight of female college students. We can denote the average weight for males as μ_{male}

μ

male

and for females as

◆female

μ

female

.

Null Hypothesis (

◆0

H

0

):

◆0:◆male≤◆female

H

0

; μ

male

$\leq\mu$

female

The null hypothesis posits that there is no significant difference in the average weight between male and female college students. Any observed difference is due to random variability.

Alternative Hypothesis (

1

H

1

or

2

H

a

):

1: male > female

H

1

μ

male

$>\mu$

female

The alternative hypothesis suggests that the average weight of male college students is greater than the average weight of female college students.

In words, the hypotheses can be stated as follows:

- Null Hypothesis: The average weight of male college students is less than or equal to the average weight of female college students.
- Alternative Hypothesis: The average weight of male college students is greater than the average weight of female college students.

During the hypothesis testing process, if the data provides enough evidence to reject the null hypothesis in favor of the alternative hypothesis, it would imply that there is a statistically significant difference in the average weights of male and 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.

Ans: To conduct a hypothesis test on the difference between two population means using Python, you can use the `scipy.stats` module, which provides a variety of statistical tests. One common test for this purpose is the independent two-sample t-test. Here's a simple Python script using this test:

python

Copy code

```
import sys as sys
from scipy import stats

def two_sample_t_test(x, y, alpha=0.05):
    t_stat, p_value = stats.ttest_ind(x, y)
    if p_value < alpha:
        print f"Reject the null hypothesis. There is enough evidence to suggest a significant difference."
    else:
        print f"Fail to reject the null hypothesis. There is not enough evidence to suggest a significant difference."
```



```
print "Fail to reject the null hypothesis. There is not enough evidence to
suggest a significant difference."
```

```
print f"T-statistic: {t_stat}"
print f"P-value: {p_value}"
```

42

75	10	30
70	8	30

In this example, `sample_male` and `sample_female` are arrays representing the weight samples for male and female college students, respectively. The `ttest_ind` function from `scipy.stats` is used to perform the two-sample t-test. The `equal_var=False` argument assumes unequal variances between the two samples.

Adjust the sample data according to your specific scenario. The function will output whether to reject the null hypothesis based on the provided significance level (default is 0.05), along with the t-statistic and p-value.

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

Ans: In statistical hypothesis testing, the null hypothesis (

0

H

0

) and alternative hypothesis (

1

H

1

or

◆◆

H

a

) are statements about a population parameter or parameters. These hypotheses are used to make inferences based on sample data. Here are definitions and examples for both types of hypotheses:

Null Hypothesis (

◆0

H

0

):

- The null hypothesis is a statement of no effect, no difference, or no change.
- It represents the default or status quo assumption.
- Denoted as
- ◆0
- H
- 0
-
- .

Examples:

- The average height of male students is equal to 68 inches.
- The mean score on a test for two different teaching methods is the same.
- The correlation coefficient between two variables is zero.

Alternative Hypothesis (

◆1

H

1

or

❖❖

H

a

):

- The alternative hypothesis is a statement that contradicts the null hypothesis.
- It suggests the presence of an effect, a difference, or a change.
- Denoted as
- ❖1
- H
- 1
-
- or
- ❖❖
- H
- a
-
- .

Examples:

- The average height of male students is not equal to 68 inches.
- The mean score on a test for two different teaching methods is different.
- The correlation coefficient between two variables is not zero.

In the context of hypothesis testing, the goal is to collect and analyze sample data to determine whether there is enough evidence to reject the null hypothesis in favor of the alternative hypothesis. The decision is based on statistical tests that consider the observed data's likelihood under the assumption of the null hypothesis. If the data provides strong evidence against the null hypothesis, the researcher may choose to accept the alternative hypothesis.

For example, in a hypothesis test comparing the effectiveness of two drugs, the null hypothesis might state that there is no difference in the average recovery time between the two drugs, while the alternative hypothesis would assert that there is a significant difference. The analysis of the

sample data would then determine whether there is enough evidence to reject the null hypothesis and support the claim of a difference in effectiveness.

Q7: Write down the steps involved in hypothesis testing.

Ans: Hypothesis testing involves a series of steps to make inferences about population parameters based on sample data. The following are the general steps involved in hypothesis testing:

Formulate the Hypotheses:

- Null Hypothesis (
- H_0
- H
- 0
-) : Represents the default assumption of no effect, no difference, or no change.
- Alternative Hypothesis (
- H_1
- H
- 1
- or
- H_a
- H
- a
-) : Represents the claim or the presence of an effect, a difference, or a change.

Choose the Significance Level (

?

 $\alpha):$

- The significance level (
- α
- α) is the probability of rejecting the null hypothesis when it is true.
- Common choices include 0.05, 0.01, or 0.10.

Select the Appropriate Statistical Test:

- The choice of the statistical test depends on the type of data and the research question.
- Common tests include t-tests, chi-square tests, ANOVA, regression analysis, etc.

Collect and Analyze the Data:

- Collect a sample of data relevant to the research question.
- Perform the chosen statistical test on the sample data.

Calculate the Test Statistic:

- The test statistic is a numerical value computed from the sample data that is used to assess the evidence against the null hypothesis.

Determine the Critical Region or Critical Value:

- Based on the chosen significance level, determine the critical region (the range of values that would lead to rejecting the null hypothesis) or the critical value (the specific value that separates the critical region from the non-critical region).

Make a Decision:

- Compare the calculated test statistic to the critical value.
- If the test statistic falls in the critical region or is less than the critical value, reject the null hypothesis.
- If the test statistic is in the non-critical region or greater than the critical value, fail to reject the null hypothesis.

Draw a Conclusion:

- Based on the decision in the previous step, draw a conclusion about the null hypothesis.
- State the results in the context of the research question.

Consider Practical Significance:

- Even if the null hypothesis is rejected, consider whether the observed effect is practically significant and has real-world importance.

Report the Results:

- Clearly communicate the findings, including the test statistic, p-value, and the decision regarding the null hypothesis.

It's important to note that hypothesis testing is a formalized process, and the results should be interpreted with care, considering the context of the research and the limitations of the study. Additionally, the formulation of hypotheses, choice of significance level, and interpretation of results require careful consideration and domain knowledge.

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

Ans: The p-value, or probability value, is a measure that helps in assessing the evidence against the null hypothesis in hypothesis testing. It quantifies the probability of observing a test statistic as extreme as, or more extreme than, the one calculated from the sample data, assuming that the null hypothesis is true.

Here's how to interpret the p-value:

- A small p-value (typically less than the chosen significance level, often 0.05) indicates that the observed data is unlikely under the assumption of the null hypothesis. This leads to the rejection of the null hypothesis in favor of the alternative hypothesis.
- A large p-value suggests that the observed data is likely under the assumption of the null hypothesis. This leads to a failure to reject the null hypothesis.

In summary:

- Small p-value: Strong evidence against the null hypothesis; favor the alternative hypothesis.
- Large p-value: Weak evidence against the null hypothesis; do not reject the null hypothesis.

The significance of the p-value can be understood by comparing it to the chosen significance level (



α). If the p-value is less than or equal to



α , you reject the null hypothesis. If the p-value is greater than



α , you fail to reject the null hypothesis.

Here are some key points about the p-value:

Significance Level (



α):

- The significance level is the threshold used to make the decision about the null hypothesis.
- Commonly used values are 0.05, 0.01, or 0.10.

Interpretation:

- A p-value does not provide the probability that either hypothesis is true; it only indicates the strength of evidence against the null hypothesis.

- The smaller the p-value, the stronger the evidence against the null hypothesis.

Not a Measure of Effect Size:

- The p-value does not measure the size or importance of the observed effect; it only indicates the likelihood of observing such an effect if the null hypothesis is true.

Dependence on Sample Size:

- Larger sample sizes may result in smaller p-values, even for small effect sizes, due to increased statistical power.

Decision Rule:

- The decision to reject or fail to reject the null hypothesis is based on comparing the p-value to the significance level.

It's crucial to interpret p-values carefully and consider them in conjunction with other relevant information. Additionally, the p-value is not a guarantee of making the correct decision; it is a tool for statistical inference in the context of hypothesis testing.

Q9. Generate a Student's t-distribution plot using Python's matplotlib library, with the degrees of freedom parameter set to 10.

Ans: can use the `scipy.stats` module to generate a Student's t-distribution and the `matplotlib` library to plot it. Here's an example code to create a Student's t-distribution plot with degrees of freedom set to 10:

python

Copy code

```
import sys as sys
import matplotlib.pyplot as plt
from scipy.stats import t

df = 10

fig, ax = plt.subplots(1, 1)
ax.plot(t.pdf(x, df), label='Student's t-distribution')
ax.plot(t.cdf(x, df), label='Cumulative Distribution Function (CDF)')
ax.plot(t.ppf(x, df), label='Probability Density Function (PDF)')
```

True

In this example, `df` is the degrees of freedom parameter for the t-distribution, and `pdf` is used to calculate the probability density function. The resulting plot shows the probability density function for the Student's t-distribution with degrees of freedom set to 10.

You can adjust the range of x values (`np.linspace(-4, 4, 1000)`) or other parameters based on your preferences and requirements. The plot will display the shape of the t-distribution for the specified degrees of freedom.

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.

Ans: can use the `scipy.stats` module in Python to perform a two-sample t-test for independent samples. Here's an example Python program:

python

Copy code

```
import sys as sys
from scipy import
```

```
def two_sample_t_test
```

```
print f"Two-sample t-test results:"
print f"T-statistic: {t_statistic}"
print f"P-value: {p_value}"
```

42

50	10	30
55	10	30

In this example:

- `sample1` and `sample2` are two random samples of equal size.
- The `ttest_ind` function from `scipy.stats` is used to perform the two-sample t-test.
- The function prints the t-statistic and p-value.

Adjust the values of `loc`, `scale`, `size`, and any other parameters according to your specific scenario.

The interpretation of the results is based on the p-value. If the p-value is less than the chosen significance level (e.g., 0.05), you may reject the null hypothesis that the population means are equal. If the p-value is greater than the significance level, you fail to reject the null hypothesis.

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

Ans: Student's t-distribution, often simply referred to as the t-distribution, is a probability distribution that is used in statistical hypothesis testing when the sample size is small and the population standard deviation is unknown. It is named after William Sealy Gosset, who published under the pseudonym "Student."

Key Characteristics of the t-Distribution:

Shape:

- The t-distribution is symmetric and bell-shaped, similar to the normal distribution.
- However, it has heavier tails, which means it has more probability in the tails than the normal distribution.

Parameter:

- The shape of the t-distribution is determined by the degrees of freedom (df).
- As the degrees of freedom increase, the t-distribution approaches the standard normal distribution.

Degrees of Freedom:

- The degrees of freedom in the t-distribution depend on the sample size. For a sample of size n , the degrees of freedom are $n-1$.

- As the degrees of freedom increase, the t-distribution becomes closer to the standard normal distribution.

When to Use the t-Distribution:

The t-distribution is commonly used in the following situations:

Small Sample Sizes:

- When dealing with small sample sizes (typically
- $n < 30$
- $n < 30$) where the population standard deviation is unknown.

Population Standard Deviation Unknown:

- When the population standard deviation is unknown, and the sample standard deviation is used as an estimate.

Estimation and Confidence Intervals:

- For constructing confidence intervals for population parameters (e.g., mean) when the sample size is small.

Two-Sample t-Test:

- In the two-sample t-test when comparing the means of two independent samples.

Paired Samples:

- In the paired samples t-test when comparing the means of two related samples (e.g., before and after measurements).

Regression Analysis:

- In situations where the residuals of a regression model are assumed to be normally distributed.

Use Case Example:

Suppose you want to estimate the average height difference between two groups of people. If you have small sample sizes and don't know the population standard deviation, you would use a t-test. The t-distribution would provide a more accurate assessment of uncertainty in your estimate compared to a normal distribution in such cases.

In summary, the t-distribution is particularly valuable in situations where the sample size is small, and the population standard deviation is unknown, providing a more accurate representation of uncertainty in statistical inference.

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

Ans: The t-statistic is a measure used in hypothesis testing to assess the evidence against a null hypothesis. It indicates how many standard errors a sample mean is from the hypothesized population mean. The t-statistic is commonly used when dealing with small sample sizes and situations where the population standard deviation is unknown.

The formula for the t-statistic depends on the context, but a common formula for the t-statistic in the context of testing the difference between a sample mean (

\bar{x}

x

-

) and a population mean (

μ

μ) is as follows:

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

$t =$

$s /$

n

x

-

μ

Where:

- \bar{x}

- μ
- σ

- \bar{x} is the sample mean.

- μ

- μ is the hypothesized population mean under the null hypothesis.

- s

- s is the sample standard deviation.

- n

- n is the sample size.

The t-statistic measures how far the sample mean is from the population mean in terms of standard errors. If the t-statistic is large (far from zero), it suggests that the sample mean is unlikely to have occurred by chance, providing evidence against the null hypothesis.

In a two-sample t-test, where you are comparing the means of two independent samples, the formula for the t-statistic is more complex and involves the difference between the sample means (

$$\bar{x}_1 - \bar{x}_2$$

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$=$$

$$1$$

$$=$$

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$=$$

$$2$$

) and the standard error of the difference.

In summary, the t-statistic is a standardized measure that helps assess the significance of the difference between sample and population means or between two sample means. The larger the t-statistic, the more likely it is that the observed difference is statistically significant.

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.

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

μ):

$$\text{Confidence Interval} = (\bar{x} - z^* \frac{s}{\sqrt{n}}, \bar{x} + z^* \frac{s}{\sqrt{n}})$$

$$\text{Confidence Interval} = (\bar{x} - z^* \frac{s}{\sqrt{n}}, \bar{x} + z^* \frac{s}{\sqrt{n}})$$

\bar{x}

-

$-z^*$

n

s

,

\bar{x}

-

$+Z$

n

s

)

Where:

- \bar{x}
 - s
 - n
 - Z
- \bar{x} is the sample mean.
 - s is the sample standard deviation.
 - n is the sample size.
 - Z is the Z-score corresponding to the desired confidence level.

For a 95% confidence interval, the Z-score is approximately 1.96 (you can find this value from a standard normal distribution table or use a calculator).

Given the information provided:

- $\bar{x} = \$500$ (sample mean)
- $s = \$50$ (sample standard deviation)
- $n = 50$
- $n = 50$ (sample size)
- $Z \approx 1.96$
- $Z \approx 1.96$ (for a 95% confidence interval)

Let's calculate the confidence interval:

Confidence Interval=(500-1.965050,500+1.965050)

Confidence Interval=(500-1.96

50

50

,500+1.96

50

50

)

python

Copy code

import

500

50

50

0.95

1.96

```
print f"95% Confidence Interval for Population Mean Revenue: ({lower_bound:.2f},  
{upper_bound:.2f})"
```

The 95% confidence interval for the population mean revenue is approximately \$493.78, \$506.22. This means that we are 95% confident that the true average daily revenue for the coffee shop falls within this interval.

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.

Ans: To test the hypothesis that the new drug decreases blood pressure by 10 mmHg, you can perform a one-sample t-test. The null hypothesis (

0

H_0

0

) is that the mean decrease in blood pressure (

?

μ) is equal to 10 mmHg, and the alternative hypothesis (

1

H_1

1

or

??

H_2

α

) is that the mean decrease is not equal to 10 mmHg.

The formula for the t-statistic in a one-sample t-test is given by:

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

$t =$

$s /$

n

\bar{x}

$-$

μ_0

where:

- \bar{x}

- \bar{x}

- $-$

- \bar{x} is the sample mean decrease in blood pressure.

- μ_0

- μ_0 is the hypothesized population mean decrease (in this case, 10 mmHg).

- s

- s is the sample standard deviation.

- n

- n is the sample size.

Let's perform the calculations and test the hypothesis:

python

Copy code

```
import sys
from statistics import mean, stdev

n = 8
sample_mean = 10
population_mean = 3
population_std_dev = 100
significance_level = 0.05

df = n - 1
critical_value = 1

t_statistic = (sample_mean - population_mean) / (stdev / (n ** 0.5))

print f"t-statistic: {t_statistic}"
print f"Critical value: {critical_value}"
print f"P-value: {p_value}"

if abs(t_statistic) > critical_value:
    print "Reject the null hypothesis. There is enough evidence to suggest a significant difference."
else:
    print "Fail to reject the null hypothesis. There is not enough evidence to suggest a significant difference."
```

In this example, the t-statistic is calculated, and the critical value is determined for a two-tailed test with a significance level of 0.05. The results are then compared, and a conclusion is drawn regarding whether to reject the null hypothesis.

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.

ans: To test the hypothesis that the true mean weight of the products is less than 5 pounds, you can perform a one-sample t-test. The null hypothesis (

H_0

H

0

) is that the mean weight (

μ

μ) is equal to 5 pounds, and the alternative hypothesis (

H_1

H

1

or

H_a

H

a

) is that the mean weight is less than 5 pounds.

The formula for the t-statistic in a one-sample t-test is given by:

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

$t =$

$s /$

n

\bar{x}

$-$

μ

where:

- \bar{x} is the sample mean weight.
- μ is the hypothesized population mean weight (in this case, 5 pounds).
- s is the sample standard deviation.
- n is the sample size.
- \bar{x}
- μ

Let's perform the calculations and test the hypothesis:

python

Copy code

```
import
from          import
```

4.8
5
0.5
25
0.01

1

```
print f"t-statistic: {t_statistic}"
print f"Critical value: {critical_value}"
print f"P-value: {p_value}"

if
    print "Reject the null hypothesis. There is enough evidence to suggest the true
mean weight is less than 5 pounds."
else
    print "Fail to reject the null hypothesis. There is not enough evidence to
suggest the true mean weight is less than 5 pounds."
```

In this example, the t-statistic is calculated, and the critical value is determined for a one-tailed test with a significance level of 0.01. The results are then compared, and a conclusion is drawn regarding whether to reject the null hypothesis.

Q16. Two groups of students are given different study materials to prepare for a test. The first group ($n_1 = 30$) has a mean score of 80 with a standard deviation of 10, and the second group ($n_2 = 40$) 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.

Ans: To test the hypothesis that the population means for the two groups are equal, you can perform an independent two-sample t-test. The null hypothesis (

H_0

H

0

) is that the means are equal, and the alternative hypothesis (

H_1

H

1

or

H_a

H

a

) is that the means are not equal.

The formula for the t-statistic in an independent two-sample t-test is given by:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$t =$

n

1

s

1

2

+

n

2

s

2

2

x

-

1

-

x

-

2

where:

- \bar{x}_1

- \bar{x}
- $-$

- 1
-
- and
- \bar{x}_2

- \bar{x}
- $-$

- 2
-
- are the sample means for Group 1 and Group 2, respectively.
- \bar{x}_1
- s_1
- 1
-
- and
- \bar{x}_2
- s_2
- 2
-
- are the sample standard deviations for Group 1 and Group 2, respectively.
- \bar{x}_1
- n_1
- 1
-
- and
- \bar{x}_2
- n_2
- 2
-
- are the sample sizes for Group 1 and Group 2, respectively.

Let's perform the calculations and test the hypothesis:

python

Copy code

import


```

from statistics import NormalDist

n = 30
x_bar = 75
s = 10
alpha = 0.01

df = n - 1
t_statistic = (x_bar - mu_0) / (s / (df ** 0.5))

critical_value = NormalDist(0, 1).ppf(1 - alpha / 2)

p_value = 2 * (1 - NormalDist(t_statistic, 1).cdf(abs(t_statistic)))

print f"t-statistic: {t_statistic}"
print f"Critical value: {critical_value}"
print f"P-value: {p_value}"

if abs(t_statistic) > critical_value:
    print "Reject the null hypothesis. There is enough evidence to suggest the
population means are not equal."
else:
    print "Fail to reject the null hypothesis. There is not enough evidence to
suggest the population means are not equal."

```

In this example, the t-statistic is calculated, and the critical value is determined for a two-tailed test with a significance level of 0.01. The results are then compared, and a conclusion is drawn regarding whether to reject the null hypothesis.

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.

Ans: To estimate the population mean with a 99% confidence interval, you can use the formula for the confidence interval for the population mean (

μ):

Confidence Interval = $(\bar{x} - Z_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + Z_{\alpha/2} \frac{s}{\sqrt{n}})$

Confidence Interval =

\bar{x}

-

$-Z$

n

s

,

\bar{x}

-

$+Z$

n

s

)

where:

- \bar{x}
- s is the sample standard deviation.
- n is the sample size.
- Z is the Z-score corresponding to the desired confidence level.

For a 99% confidence interval, the Z-score is approximately 2.576 (you can find this value from a standard normal distribution table or use a calculator).

Given the information provided:

- $\bar{x}=4$
- $s=1.5$ (sample standard deviation)
- $n=50$ (sample size)

- $\hat{\sigma} \approx 2.576$
- $Z \approx 2.576$ (for a 99% confidence interval)

Let's calculate the confidence interval:

Confidence Interval = $(4 - 2.576 \cdot 1.550, 4 + 2.576 \cdot 1.550)$

Confidence Interval = $(4 - 2.576$

50

1.5

, $4 + 2.576$

50

1.5

)

python

Copy code

import

```

4
1.5
50
0.99

2.576

```

```
print f"99% Confidence Interval for Population Mean: ({lower_bound:.2f},  
{upper_bound:.2f})"
```

The 99% confidence interval for the population mean number of ads watched by viewers is approximately (3.59, 4.41). This means that we are 99% confident that the true average number of ads watched falls within this interval.