SHRI VILEPARLE KELAVANI MANDAL'S
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC ACCREDITED with "A" GRADE (CGPA : 3.18)

DEPARTMENT OF INFORMATION TECHNOLOGY

COURSE CODE: DJS22ITL5013 **DATE: 28/08/2024**

COURSE NAME: Statistical Analysis Lab **CLASS: T.Y. Btech**

NAME: Anish Sharma **SAP: 60003220045**

EXPERIMENT NO.04

CO 2: Perform Test of Hypothesis for independence and appropriateness of distribution using various statistical

techniques.

AIM / OBJECTIVE: To implement single population tests for mean, proportion and variance

DESCRIPTION OF EXPERIMENT:

This experiment aims to implement single population tests for mean, proportion, and variance. The single population tests are used to make inferences about the population parameters based on a sample drawn from the population. The three types of tests covered in this experiment are:

1. Single Population Mean Test: This test is used to determine if the sample mean is significantly different from a hypothesized population mean.
2. Single Population Proportion Test: This test is used to determine if the sample proportion is significantly different from a hypothesized population proportion.
3. Single Population Variance Test: This test is used to determine if the sample variance is significantly different from a hypothesized population variance.

The key differences between the single population tests and the two-sample tests (or independent samples tests) are:

1. **Single Population Tests**: These tests are used to make inferences about a single population parameter based on a sample from that population.
2. **Two-Sample Tests**: These tests are used to make inferences about the difference between two population parameters based on two independent samples.

For the single population tests, the hypotheses and test statistics are formulated differently compared to the two-sample tests. The single population tests focus on comparing a sample statistic (mean, proportion, or variance) to a hypothesized population parameter, while the two- sample tests focus on comparing the differences between two sample statistics.

INPUT DATA / DATASET:
Steps:

1. Perform Hypothesis Testing

a. Testing a Population Mean with Known Standard Deviation (z-test)

Exam Scores: [85, 90, 88, 92, 87, 91, 93, 89, 84, 86]

Step 1: Hypotheses

- Null Hypothesis (H0H_0H0): The mean exam score is 90 (μ=90\mu = 90μ=90).
- Alternative Hypothesis (H1H_1H1): The mean exam score is not 90 (μ≠90\mu \neq 90μ□=90).

Step 2: Statistical Test

- Use the z-test for a population mean with a known standard deviation.

Step 3: Set Alpha

- Alpha (α\alphaα) is commonly set to 0.05 (5%).

Step 4: Decision Rule

- Use the z-table to find the critical z-value for α=0.05\alpha = 0.05α=0.05. For a two-tailed test, the critical z-values are ±1.96\pm 1.96±1.96.

Step 5: Gather Sample Data

- Sample Mean (x¯\bar{x}x¯) = Average of the exam scores.
- Population Standard Deviation (σ\sigmaσ) = Assume known.

Step 6: Analyze the Data

- Calculate the z-statistic: z=x¯−μσnz = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}z=nσx¯−μ where x¯\bar{x}x¯ is the sample mean, μ\muμ is the population mean, σ\sigmaσ is the population standard deviation, and nnn is the sample size.

Step 7: Statistical Conclusion

- Compare the calculated z-value with the critical z-values. If the z-value falls within the range [−1.96,1.96][-1.96, 1.96][−1.96,1.96], fail to reject H0H_0H0. Otherwise, reject H0H_0H0.

Step 8: Business Decision

- Based on the statistical conclusion, decide if the mean score is significantly different from 90.

```
import numpy as np
from scipy import stats

data = [85, 90, 88, 92, 87, 91, 93, 89, 84, 86]
sample_mean = np.mean(data)
sample_std = np.std(data, ddof=0)  # Assume known population std deviation
population_mean = 90
n = len(data)

z = (sample_mean - population_mean) / (sample_std / np.sqrt(n))
p_value = 2 * (1 - stats.norm.cdf(abs(z)))

print("Z-value:", z)
print("P-value:", p_value)
```

b. Testing a Population Mean with Unknown Standard Deviation (t-test)

Step 1: Hypotheses

- Null Hypothesis (H0H_0H0): The mean exam score is 90 (μ=90\mu = 90μ=90).
- Alternative Hypothesis (H1H_1H1): The mean exam score is not 90 (μ≠90\mu \neq 90μ□=90).

Step 2: Statistical Test

- Use the t-test for a population mean with an unknown standard deviation.

Step 3: Set Alpha

- Alpha (α\alphaα) is commonly set to 0.05 (5%).

Step 4: Decision Rule

- Use the t-table to find the critical t-value for α=0.05\alpha = 0.05α=0.05 and degrees of freedom (df=n−1df = n - 1df=n−1).

Step 5: Gather Sample Data

- Sample Mean (x¯\bar{x}x¯) and Sample Standard Deviation (sss).

Step 6: Analyze the Data

- Calculate the t-statistic: t=x¯−μsnt = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}t=nsx¯−μ

Step 7: Statistical Conclusion

- Compare the calculated t-value with the critical t-values. If the t-value is beyond the critical value, reject H0H_0H0.

Step 8: Business Decision

- Make decisions based on whether H0H_0H0 is rejected or not.

```
t_stat, p_value = stats.ttest_1samp(data, population_mean)

print("T-statistic:", t_stat)

print("P-value:", p_value)
```

c. Testing a Population Proportion (z-test)

Step 1: Hypotheses

- Null Hypothesis (H0H_0H0): The proportion is 0.5 (p=0.5p = 0.5p=0.5).
- Alternative Hypothesis (H1H_1H1): The proportion is not 0.5 (p≠0.5p \neq 0.5p□=0.5).

Step 2: Statistical Test

- Use the z-test for a population proportion.

Step 3: Set Alpha

- Alpha (α\alphaα) is commonly set to 0.05 (5%).

Step 4: Decision Rule

- Use the z-table for critical values.

Step 5: Gather Sample Data

- Proportion in the sample.

Step 6: Analyze the Data

- Calculate the z-statistic for proportion: z=p^−p0p0(1−p0)nz = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 (1 - p_0)}{n}}}z=np0(1−p0)p^−p0

Step 7: Statistical Conclusion

- Compare z-value with critical values.

Step 8: Business Decision

- Based on the result, decide whether the proportion differs significantly from 0.5.

```
p_hat = 0.6  # Example proportion
p0 = 0.5
n = 100  # Sample size

z = (p_hat - p0) / np.sqrt(p0 * (1 - p0) / n)
p_value = 2 * (1 - stats.norm.cdf(abs(z)))

print("Z-value:", z)
print("P-value:", p_value)
```

d. Testing a Population Variance (Chi-Square Test)

Step 1: Hypotheses

- Null Hypothesis (H_0): The variance is σ^2_0 (e.g., 25).
- Alternative Hypothesis (H_1): The variance is not σ^2_0.

Step 2: Statistical Test

- Use the chi-square test for variance.

Step 3: Set Alpha

- Alpha (α) is commonly set to 0.05 (5%).

Step 4: Decision Rule

- Use the chi-square distribution table to find critical values.

Step 5: Gather Sample Data

- Sample Variance (s^2).

Step 6: Analyze the Data

- Calculate the chi-square statistic: $\chi^2 = \frac{(n - 1) s^2}{\sigma^2_0}$

Step 7: Statistical Conclusion

- Compare the chi-square statistic with the critical chi-square values.

Step 8: Business Decision

- Make decisions based on whether the variance is significantly different.

```
df = len(data) - 1

s2 = np.var(data, ddof=1)  # Sample variance

sigma2_0 = 25  # Hypothesized variance



chi2_stat = (df * s2) / sigma2_0

p_value = 2 * (1 - stats.chi2.cdf(chi2_stat, df))



print("Chi-Square Statistic:", chi2_stat)

print("P-value:", p_value)
```

a. z-test for Population Mean with Known Standard Deviation

Observations/Discussion of Result: The z-test showed whether the sample mean differed from 90. If the z-value was within the critical range, the sample mean was consistent with the hypothesized mean.

Conclusion: A significant result indicates the average exam score deviates from 90, suggesting potential issues in student performance or instructional methods may need to be addressed.

---

b. t-test for Population Mean with Unknown Standard Deviation

Observations/Discussion of Result: The t-test revealed if the sample mean differed significantly from 90. A significant t-statistic suggested differences in performance compared to the hypothesized mean.

Conclusion: A significant t-test result means the exam score mean differs from 90, indicating possible need for teaching adjustments or further analysis of student performance.

---

c. z-test for Population Proportion

Observations/Discussion of Result: The z-test assessed if the observed proportion differed from an expected proportion (e.g., 0.5). A significant z-value indicated a meaningful difference in proportions.

Conclusion: A significant result suggests the observed proportion differs from the expected, potentially requiring changes in educational strategies or policies to address performance gaps.

---

d. Chi-Square Test for Population Variance

Observations/Discussion of Result: The chi-square test determined if the sample variance was consistent with the hypothesized variance. A significant result indicated a deviation in variance.

Conclusion: A significant chi-square result implies the variance in exam scores differs from expectations, suggesting a review of grading practices or further investigation into score variability may be necessary.

REFERENCES:

Bhattacharyya, G. K., and R. A. Johnson, (1997). Statistical Concepts and Methods, John Wiley andSons, New York.

Website References:

1. https://www.geo.fu-berlin.de/en/v/soga-r/Basics-of-statistics/Hypothesis-Tests/Hypothesis-Tests-for-One-Population-Mean/Sigma-Is-Known/index.html
2. https://medium.com/%E8%89%BE%E8%9C%9C%E8%8E%89%E8%AE%80%E8%AE%80%E5%AF%AB%E5%AF%AB/datacamp-hypothesis-testing-in-python-21427a987352
3. https://www.statsmodels.org/dev/generated/statsmodels.stats.proportion.proportions_ztest.html
4. https://openstax.org/books/introductory-statistics-2e/pages/9-6-hypothesis-testing-of-a-single-mean-and-single-proportion