Hypothesis Testing

T-Test: One Sample and Two Sample

The **t-test** is a statistical test used to determine if there is a significant difference between the means of two groups or if a sample mean differs significantly from a known population mean. The **t-test** is divided into two types:

1. **One-Sample t-test**
2. **Two-Sample t-test**

1. One-Sample t-test

A **one-sample t-test** is used to compare the mean of a single sample to a known population mean. The goal is to determine whether the sample mean is significantly different from the population mean.

Hypotheses:

- **Null hypothesis ($\mathbf{H}_0$)**: The sample mean is equal to the population mean (μ_0).
- **Alternative hypothesis ($\mathbf{H}_1$)**: The sample mean is not equal to the population mean.

Formula:

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

Where:

- $\bar{X}$ = sample mean
- μ_0 = population mean
- s = sample standard deviation
- n = sample size

Example Problem:

A teacher claims that the average score of students in a math test is 75. A random sample of 10 students gave the following scores: 80, 85, 70, 75, 90, 85, 65, 78, 72, 82. Can we conclude that the mean score differs from 75?

- **Hypotheses:**

$$H_0 : \mu = 75$$
$$H_1 : \mu \neq 75$$

- **Calculations:**
  - Sample mean $\bar{X} = 78.2$
  - Sample standard deviation $s = 7.33$
  - $n = 10$

– Population mean $\mu_0 = 75$

Using the formula:

$$t = \frac{78.2 - 75}{\frac{7.33}{\sqrt{10}}} \approx 1.37$$

- **Conclusion:** Compare the calculated t-value to the critical t-value from the t-distribution table for $n - 1 = 9$ degrees of freedom at a significance level (e.g., $\alpha = 0.05$). If $|t|$ is greater than the critical value, reject the null hypothesis.

2. Two-Sample t-test

A **two-sample t-test** (independent t-test) is used to compare the means of two independent groups to determine if there is a statistically significant difference between them.

Types of Two-Sample t-test:

- **Equal Variances Assumed (Pooled t-test):** Assumes that both groups have the same variance.
- **Unequal Variances (Welch's t-test):** Does not assume equal variance between the two groups.

Hypotheses:

- **Null hypothesis (H_0):** The means of the two groups are equal ($\mu_1 = \mu_2$).
- **Alternative hypothesis (H_1):** The means of the two groups are not equal.

Formula (for equal variances):

$$t = \frac{\bar{X_1} - \bar{X_2}}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Where:

- $\bar{X_1}, \bar{X_2}$ = sample means of the two groups
- s_p^2 = pooled variance
- n_1, n_2 = sample sizes of the two groups

The **pooled variance** is calculated as:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Where s_1^2 and s_2^2 are the variances of the two groups.

Example Problem:

A researcher wants to know if there is a significant difference between the exam scores of students from two different schools. The exam scores for 10 students from School A are: 85, 88, 84, 90, 91, 89, 87, 85, 83, 86, and for 10 students from School B are: 78, 75, 80, 82, 77, 81, 79, 76, 74, 78.

- **Hypotheses:**

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

- **Calculations:**
  – Mean of School A $\bar{X_1} = 86.8$

- Mean of School B $\bar{X}_2 = 77.8$
- Standard deviation of School A $s_1 = 2.49$
- Standard deviation of School B $s_2 = 2.41$
- $n_1 = n_2 = 10$.

Pooled variance:

$$s_p^2 = \frac{(10-1)(2.49)^2 + (10-1)(2.41)^2}{10+10-2} = 6.03$$

Now calculate the t-value:

$$t = \frac{86.8 - 77.8}{\sqrt{6.03\left(\frac{1}{10} + \frac{1}{10}\right)}} \approx 9.06$$

- **Conclusion:** Compare the calculated t-value to the critical t-value for $n_1 + n_2 - 2 = 18$ degrees of freedom. If the calculated t-value is greater than the critical t-value, reject the null hypothesis.

Summary

- **One-Sample t-test:** Tests if a sample mean is different from a known population mean.
- **Two-Sample t-test:** Tests if the means of two independent groups are significantly different.

Both tests help in making inferences about the population based on sample data.

F-Test

The **F-test** is a statistical test used to compare the variances of two populations to determine if they are significantly different. It is commonly used in analysis of variance (ANOVA), regression analysis, and hypothesis testing to evaluate the ratio of two variances. The F-test assumes that the populations from which the samples are drawn are normally distributed.

When to Use the F-Test?

- To compare two population variances.
- To test the hypothesis that the variances of two independent samples are equal.
- As part of ANOVA (Analysis of Variance) to test the differences between group means when there are more than two groups.

Hypotheses in an F-Test

- **Null Hypothesis (H_0):** The variances of two populations are equal, i.e., $\sigma_1^2 = \sigma_2^2$.
- **Alternative Hypothesis (H_1):** The variances of two populations are not equal, i.e., $\sigma_1^2 \neq \sigma_2^2$.

Formula for the F-Test

The F-statistic is the ratio of two sample variances:

$$F = \frac{s_1^2}{s_2^2}$$

Where:

- s_1^2 is the variance of the first sample.
- s_2^2 is the variance of the second sample.

The F-distribution is used to compare these sample variances, with degrees of freedom $\mathrm{df}_1 = n_1 - 1$ and $\mathrm{df}_2 = n_2 - 1$, where n_1 and n_2 are the sizes of the two samples.

Key Assumptions

- The populations are normally distributed.
- The samples are independent of each other.

Example Problem

Suppose a quality control inspector wants to know if the variability in the production of two machines is significantly different. He takes a random sample of 10 products from Machine A and 12 products from Machine B. The sample variances for Machine A and Machine B are calculated as:

$$s_1^2 = 4.5, \quad s_2^2 = 2.8$$

Using the F-test, we can determine if there is a significant difference in the variances.

1. **Hypotheses:**

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

2. **Calculations:** Using the formula for the F-statistic:

$$F = \frac{s_1^2}{s_2^2} = \frac{4.5}{2.8} \approx 1.61$$

3. **Degrees of Freedom:**

$$\mathrm{df}_1 = n_1 - 1 = 10 - 1 = 9$$
$$\mathrm{df}_2 = n_2 - 1 = 12 - 1 = 11$$

4. **Conclusion:** To determine if the calculated F-value is significant, compare it to the critical value from the F-distribution table at a certain significance level (e.g., $\alpha = 0.05$) with $\mathrm{df}_1 = 9$ and $\mathrm{df}_2 = 11$. If the calculated F-value is greater than the critical value, reject the null hypothesis and conclude that the variances are significantly different.

Summary

- The F-test is used to compare the variances of two independent populations.
- It is widely used in ANOVA and regression analysis to assess the homogeneity of variances.
- The F-test compares the ratio of two sample variances and makes conclusions based on the F-distribution.