DEPARTMENT OF SOFTWARE ENGINEERING

STATISTICS 1

CHAPTER 10: HYPOTHESIS TESTING

MD. SABBIR AHMED MAYEN
BSc., MS in Statistics, Shahjalal University Of Science & Technology, Sylhet
Lecturer (Statistics), Department Of Software Engineering,
Daffodil International University, Dhaka-1216

HYPOTHESIS AND HYPOTHESIS TESTING

- **Hypothesis:** A hypothesis is an assumption that is made based on some evidence.

- **Hypothesis Testing:** Hypothesis testing involves making an assumption (the hypothesis) about a population parameter and testing whether the sample data provides sufficient evidence to support or reject that assumption. It is based on probability theory and is used to make inferences about population parameters from sample statistics.

TYPES OF HYPOTHESIS

- **Null Hypothesis:** This is the default assumption that there is no effect or no difference in the population. It is tested directly and often represents a statement of no relationship or no change.

✓ Example: H_0: *"Offering a discount does not increase the number of customers visiting the store"*. This suggests that the discount has no impact on the number of customers.

TYPES OF HYPOTHESIS

- **Alternative Hypothesis:** The hypothesis that contradicts the null hypothesis, suggesting that there is an effect or difference.
- Example: H_1: *"Offering a discount increases the number of customers visiting the store"*. This suggests that offering a discount will lead to more customers.

TYPES OF HYPOTHESIS

- **One-Tailed Test:** Tests whether a parameter is greater than or less than a certain value.

✓Example: H_0: *"The average monthly sales are $10,000 or more"*.

H_1: *"The average monthly sales are less than $10,000"*.

This is a one-tailed test because we are only interested in whether sales have dropped below $10,000.

- **Two-Tailed Test:** Tests whether a parameter is different from a certain value, without specifying the direction.

✓Example: H_0: "The average customer satisfaction score is 8 out of 10".

H_1: "The average customer satisfaction score is not 8 out of 10".

This is a two-tailed test because we are testing if the score is either greater or less than 8.

HYPOTHESIS FORMULATION

1. **Identify the Research Problem:** Start with a clear research question.

Example: "Does increasing the price of a product affect sales?"

2. **Conduct Preliminary Research:** Review existing literature to understand what is already known.
3. **Define Variables:**

- *Independent Variable (IV):* What you change (e.g., "Price of the product").
- *Dependent Variable (DV):* What you measure (e.g., "Sales volume").

4. **Formulate the Hypothesis:**

- *Null Hypothesis (H_0):* "There is no effect" (e.g., "Price increase has no effect on sales").
- *Alternative Hypothesis (H_1):* "There is an effect" (e.g., "Price increase decreases sales").

5. **Ensure Testability:** Make sure the hypothesis can be tested through data.
6. **Falsifiability:** The hypothesis must be proven wrong through evidence.

STEPS IN HYPOTHESIS TESTING

- **Steps in Hypothesis Testing**

1. **State the hypotheses** (null and alternative).
2. **Choose the significance level** (α), typically 0.05 or 0.01.
3. **Collect the sample data** and calculate the test statistic (z-test, t-test, chi-square test, F-test).
4. **Make a decision** based on the p-value:
    1. If the p-value is less than α, reject the null hypothesis.
    2. If the p-value is greater than α, fail to reject the null hypothesis.
5. **Draw a conclusion** based on the results.

ERRORS IN HYPOTHESIS TESTING

In hypothesis testing, there are two types of errors that can occur when making decisions about the null hypothesis (H_0). These are called **Type I Error** and **Type II Error**.

- Let's use a criminal court case as an example:

✓**Null Hypothesis (H_0):** The person is innocent (not guilty).

✓**Alternative Hypothesis (H_1):** The person is guilty.

ERRORS IN HYPOTHESIS TESTING

1. **Type I Error (False Positive):** Rejecting the null hypothesis when it is actually true.

- ***In the court case:***
  - *The jury finds the person **guilty** when he is actually **innocent**.*

  This is a **Type I Error**—a false positive, because they incorrectly rejected the null hypothesis (innocence) and convicted someone who didn't commit the crime.

2. **Type II Error (False Negative):** **Accepting** the null hypothesis when it is actually false.

- ***In the court case:***
  - *The jury finds the person **innocent** when he is actually **guilty**.*

  This is a **Type II Error**—a false negative, because they accept the null hypothesis (innocence) and let a guilty person go free.

CRITICAL REGION & LEVEL OF SIGNIFICANCE

- **Critical Region:** A critical region is that region in which if the value of the test statistic falls, we reject the null hypothesis.
- **Level of Significance:** The probability level for which the null hypothesis is rejected even it is true is called level of significance. It is denoted by $\boldsymbol{\alpha}$.

PARAMETRIC TESTS

- **T-Test**: Compares the means of two groups to determine if they are significantly different, assuming normally distributed and continuous data. Types include one-sample, independent two-sample, and paired T-tests, suitable for small or large sample sizes.
- **Z-Test**: Used for comparing sample and population means when the sample size is large (typically n>30) and the population variance is known. It assumes normal distribution and is commonly applied in quality control and survey analysis.
- **F-Test**: Compares the variances of two groups to check if they differ significantly, often as a preliminary test for ANOVA. It assumes normal distribution and helps determine equal or unequal variance handling in further analysis.

NON-PARAMETRIC TESTS

- **Mann-Whitney U Test (U-Test)**: A nonparametric test comparing differences between two independent groups without assuming normality. It ranks data across both groups and is ideal for ordinal or non-normal continuous data.
- **Wilcoxon Rank-Sum Test (Rank-Sum Test)**: Similar to the U-Test, this test compares two independent groups based on median ranks, making it suitable for ordinal or non-normal data without relying on mean values.
- **Kruskal-Wallis Test (K-W Test)**: A nonparametric alternative to one-way ANOVA, used for testing differences among more than two groups. It ranks data across groups and works well with ordinal data or when normality assumptions are violated.

ONE-SAMPLE MEAN TEST

- **Hypotheses**:

Null Hypothesis (H_0): $\mu = \mu_0$ (The population mean is equal to the hypothesized mean.)
Alternative Hypothesis (H_1): $\mu \neq \mu_0$ (The population mean is different from the hypothesized mean.)

- **Test Statistic**:

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

Here, $\bar{x}$ is the sample mean, μ_0 is the population mean, s is the standard deviation and n is the sample size

- **Degrees of Freedom:** n-1
- **Critical Region**: For a two-tailed test at significance level α, the critical region is: $|T| > t_{\frac{\alpha}{2}, df}$
- **Interpretation:** After calculating the test statistic, we compare it to the tabulated value for the chosen significance level. If the calculated value exceeds the tabulated value, we reject the null hypothesis, suggesting a significant difference between the sample mean and the hypothesized population mean. If it does not exceed the tabulated value, we fail to reject null hypothesis, indicating no significant difference.

TWO-SAMPLE MEAN TEST (INDEPENDENT T TEST)

- **Hypotheses**:

Null Hypothesis (H_0): $\mu_1 = \mu_2$ (The means of the two groups are equal.)
Alternative Hypothesis (H_1): $\mu_1 \neq \mu_2$ (The means of the two groups are different.)

- **Test Statistic**:

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

Here, $\bar{x}_1$ and $\bar{x}_2$ are the mean of the 1st and 2nd group. s_1^2 and s_2^2 are the variance of the 1st and 2nd group. n_1 and n_2 are the sample size of 1st and 2nd group.

- **Degrees of Freedom:** $n_1 + n_2 - 2$
- **Critical Region**: For a two-tailed test at significance level α, the critical region is: $|T| > t_{\frac{\alpha}{2}, \text{df}}$
- **Interpretation:** After calculating the test statistic, we compare it to the tabulated value for the chosen significance level. If the calculated value exceeds the tabulated value, we reject the null hypothesis, suggesting a significant difference between means of the two groups. If it does not exceed the tabulated value, we fail to reject null hypothesis, indicating no significant difference.

PAIRED MEAN TEST (PAIRED T TEST)

- **Hypotheses**:

Null Hypothesis (H_0): $\mu_d = 0$ (The mean difference between paired observations is zero.)
Alternative Hypothesis (H_1): $\mu_d \neq 0$ (The mean difference is not zero.)

- **Test Statistic**:

$$T = \frac{D}{\frac{S_D}{\sqrt{n}}}$$

D is the mean of the differences between each pair of observations and S_D is the standard deviation of the differences

- **Degrees of Freedom:** n-1
- **Critical Region**: For a two-tailed test at significance level α, the critical region is: |T|> $t_{\frac{\alpha}{2}}$, df
- **Interpretation:** After calculating the test statistic, we compare it to the tabulated value for the chosen significance level. If the calculated value exceeds the tabulated value, we reject the null hypothesis, suggesting a significant difference between the paired observations. If it does not exceed the tabulated value, we fail to reject null hypothesis, indicating no significant difference.

SEVERAL MEAN TEST (ANOVA)

- **Hypotheses**:

Null Hypothesis (H_0): $\mu_1 = \mu_2 = \ldots\ldots = \mu_k$ (All group means are equal.)
Alternative Hypothesis (H_1): At least one mean is different from the others.

- **Test Statistic**:

$$F = \frac{Between - group\ variance}{Within - group\ variance}$$

- **Degrees of Freedom:** Between-group (k−1), within group (N-k)
- **Critical Region**: For a significance level α, the critical region is: F > F_{α}, df (between), df (within)
- **Interpretation:** After calculating the test statistic, we compare it to the tabulated value for the chosen significance level. If the calculated value exceeds the tabulated value, we reject the null hypothesis, suggesting a significant difference among the group means. If it does not exceed the tabulated value, we fail to reject null hypothesis, indicating no significant difference.

PROPORTION TEST

- **Hypotheses**:

Null Hypothesis (H_0): p= p_0 (The population proportion is equal to the hypothesized proportion)
Alternative Hypothesis (H_1): p ≠ p_0 (The population prop. is different from the hypothesized prop.)

- **Test Statistic**:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0\,(1 - p_0)}{n}}}$$

Here, $\hat{p}$ is the sample proportion, p_0 is the population proportion and n is the sample size

- **Critical Region**: For a two-tailed test at significance level α, the critical region is: |Z|> $Z_{\frac{\alpha}{2}}$
- **Interpretation:** After calculating the test statistic, we compare it to the tabulated value for the chosen significance level. If the calculated value exceeds the tabulated value, we reject the null hypothesis, suggesting a significant difference between the sample prop and the hypothesized population prop. If it does not exceed the tabulated value, we fail to reject null hypothesis, indicating no significant difference.

VARIANCE TEST

- **Hypotheses**:

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

- **Test Statistic**:

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

- **Degrees of Freedom:** df_1 **= numerator** $(n_1 - 1)$, df_2 **= denominator** $(n_2 - 1)$
- **Critical Region**: For a significance level α, the critical region is: F > F_{α}, df_1, df_2
- **Interpretation:** After calculating the test statistic, we compare it to the tabulated value for the chosen significance level. If the calculated value exceeds the tabulated value, we reject the null hypothesis, suggesting a significant difference between the variances. If it does not exceed the tabulated value, we fail to reject null hypothesis, indicating no significant difference.

- **P value < 0.05 --- null hypothesis is rejected (Have difference)**
- P value >= 0.05 --- null hypothesis is accepted (No difference)