Ep 12 - Inferential Statistics 2

Dec 7, 2023

Type 1 and Type 2 Error

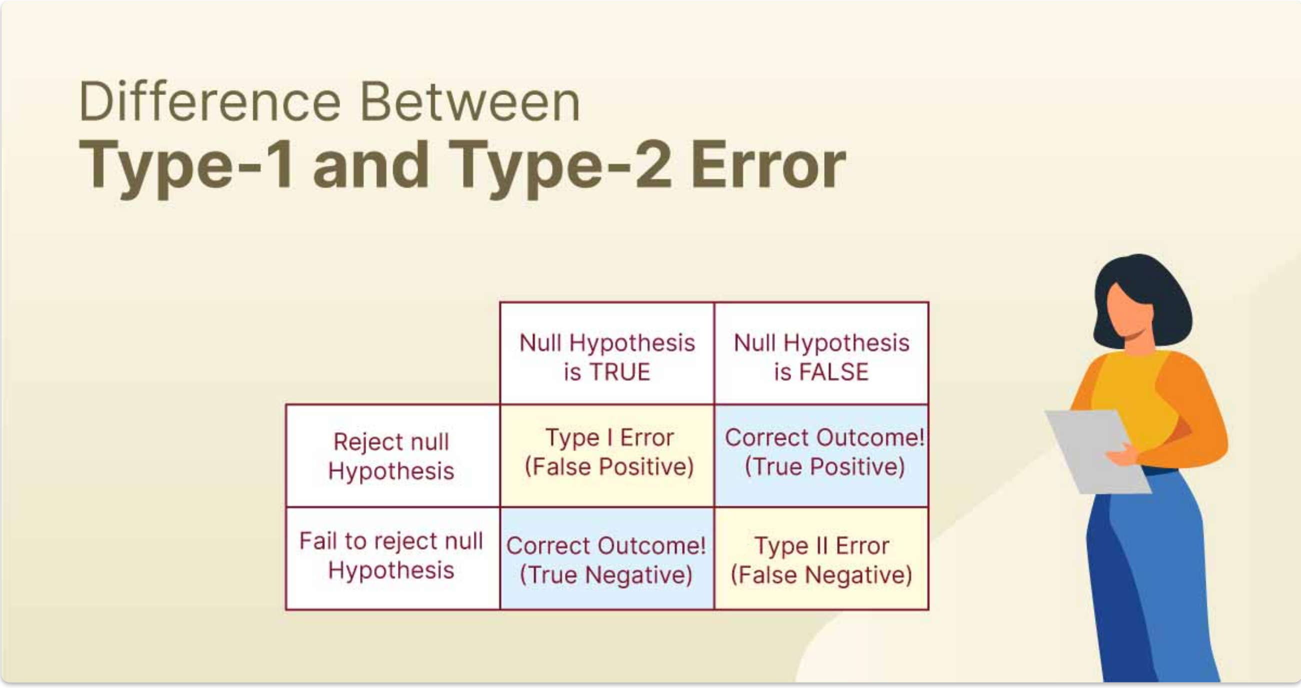
Type-I (False Positive) Error

It occurs when the sample results, lead to the rejection of the null hypothesis when it is in fact true. In other words, it's the *mistake of finding a significant effect or relationship when there is none*.
The probability of committing a Type I error is denoted by α, which is also known as the significance level. By choosing a significance level, researchers can control the risk of making a Type I error.

Type-II (False Negative) Error

It occurs when based on the sample results, the null hypothesis is not rejected when it is in fact false. This means that the researcher *fails to detect a significant effect or relationship when one actually exists*.
The probability of committing a Type II error is denoted by β.

> **Note**
>
> α and β are inversely proportional. Therefore, decreasing the chances of occurrence of an error, increases the chances of occurrence of the other.



One and Two Sided Tests

One - Sided (One - Tailed) Test

A one-sided test is used when the researcher is interested in testing the effect in a *specific direction* (either greater than or less than the value specified in the null hypothesis). The alternative hypothesis in a one-sided test contains an inequality ($>$ or $<$).

Example: A researcher wants to test whether a new medication increases the average recovery rate compared to the existing medication.

Advantages

1. More powerful
2. Directional hypothesis

Disadvantages

1. Missed effects
2. Increased effect of type - 1 error

Two - Sided (Two - Tailed) Test

A two-sided test is used when the researcher is interested in testing the effect in *both directions* (i.e., whether the value specified in the null hypothesis is different, either greater or lesser). The alternative hypothesis in a two-sided test contains a "not equal to" sign ($\neq$).

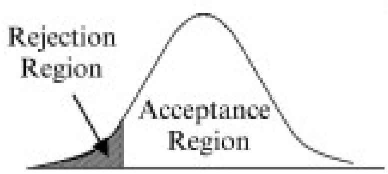
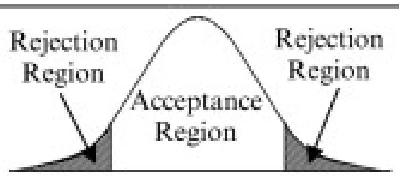
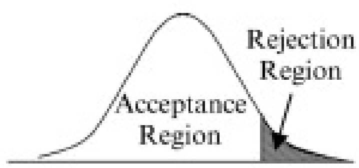
Example: A researcher wants to test whether a new medication has a different average recovery rate compared to the existing medication.

Advantages

1. Detects effects in both directions
2. More conservative

Disadvantages

1. Less powerful
2. Not appropriate for directional hypothesis

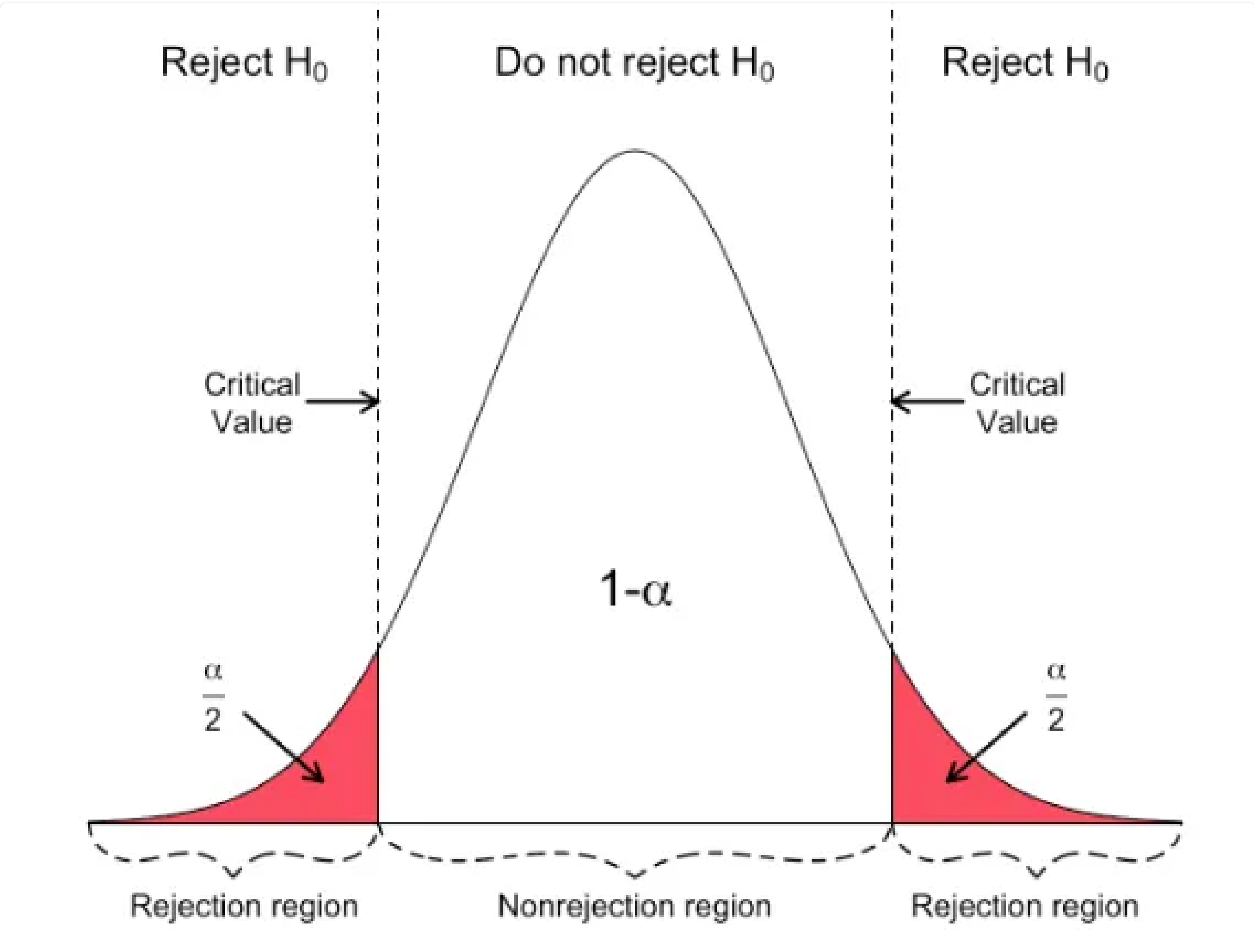
One-Tailed Test (Left Tail)	Two-Tailed Test	One-Tailed Test (Right Tail)
$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X < \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X \neq \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X > \mu_0$
		

Rejection Region Approach

1. Formulate a Null and Alternate hypothesis
2. Select a **significance level (α)** (probability of rejecting the null hypothesis when it is actually true, usually set at 0.05 or 0.01)
3. Check assumptions (example distribution)
4. Decide which test is appropriate (Z-test, T-test, Chi-square test, ANOVA)
5. State the relevant **test statistic** (such as z-score, t-value)
6. Conduct the test
7. Reject or not reject the Null Hypothesis.
8. Interpret the result

Significance Level (α)

It is a predetermined threshold used in hypothesis testing to determine whether the null hypothesis should be rejected or not. It represents the *probability of rejecting the null hypothesis when it is actually true*, also known as **Type 1 error**.
The rejection region is the region of values that corresponds to the rejection of the null hypothesis at some chosen probability level.



Danger

The rejection region approach in hypothesis testing has two main flaws: it introduces subjectivity in the choice of critical values and significance levels, making the interpretation of results dependent on arbitrary decisions, and it lacks the continuous measure of evidence against the null hypothesis provided by p-values, hindering comparability across studies and experiments.

Example 1

Suppose a company is evaluating the impact of a new training program on the productivity of its employees. The company has data on the average productivity of its employees before implementing the training program. The average productivity was 50 units per day with a known population standard deviation of 5 units. After implementing the training program, the company measures the productivity of a random sample of 30 employees. The sample has an average productivity of 53 units per day. The company wants to know if the new training program has significantly increased productivity.

Solution

1. $H_0 =$ average productivity is 50 and $H_1 =$ average productivity > 50
2. $\alpha = 0.05$
3. Normality is valid and σ is also known
4. Applying Z - test
5. Z-score is

$$Z = \frac{\bar{x} - \mu}{\sigma/n} = \frac{53 - 50}{5/\sqrt{30}} = 3.28$$

6. 95% corresponds to a value of 1.65
7. As 3.28 lies beyond 1.65 i.e. in the rejection region, we have evidence to reject the null hypothesis for the alternate hypothesis

Example 2

Suppose a snack food company claims that their Lays wafer packets contain an average weight of 50 grams per packet. To verify this claim, a consumer watchdog organization decides to test a random sample of Lays wafer packets. The organization wants to determine whether the actual average weight differs significantly from the claimed 50 grams. The organization collects a random sample of 40 Lays wafer packets and measures their weights. They find that the sample has an average weight of 49 grams, with a known population standard deviation of 4 grams.

Solution

1. $H_0 =$ average weight is 50 and $H_1 =$ average weight $\neq 50$
2. $\alpha = 0.05$
3. Normality is valid and σ is also known
4. Applying Z - test
5. Z-score is

$$Z = \frac{49 - 50}{\frac{4}{\sqrt{40}}} = -1.58$$

6. $Z < -1.96 \text{ and } Z > 1.96$ is the rejection region
7. As -1.58 lies between and in the no rejection region, we have no evidence to reject null hypothesis

P - value Approach

The p-value, or **probability value**, is a measure used in statistical hypothesis testing to assess the evidence against a null hypothesis.

The p-value represents the *probability of observing a test statistic as extreme as, or more extreme than, the one computed from the sample data*, assuming that the null hypothesis is true. In other words, it *quantifies the strength of the evidence against the null hypothesis*.

Interpreting p-values

With significant value (α):

- A **small p-value** (typically below a significance level (α) like 0.05) suggests that you *have enough evidence to reject* the null hypothesis. It implies that the observed data is unlikely to occur if the null hypothesis is true.
- A **large p-value** suggests that you do *not have enough evidence to reject* the null hypothesis. It implies that the observed data is reasonably likely to occur if the null hypothesis is true.

Without significant value (α):

- **Very small p-values** $(p < 0.01)$ indicate strong evidence against the null hypothesis, suggesting that the observed effect or difference is unlikely to have occurred by chance alone.
- **Small p-values** $(0.01 \leq p < 0.05)$ indicate moderate evidence against the null hypothesis, suggesting that the observed effect or difference is less likely to have occurred by chance alone.

- **Large p-values** $(0.05 \leq p < 0.1)$ indicate weak evidence against the null hypothesis, suggesting that the observed effect or difference might have occurred by chance alone, but there is still some level of uncertainty.
- **Very large p-values** $(p \geq 0.1)$ indicate weak or no evidence against the null hypothesis, suggesting that the observed effect or difference is likely to have occurred by chance alone.

> **Note**
>
> It's important to note that a p-value does not provide the probability that either hypothesis is true; it only informs about the strength of the evidence against the null hypothesis.

Example 1

Suppose a company is evaluating the impact of a new training program on the productivity of its employees. The company has data on the average productivity of its employees before implementing the training program. The average productivity was 50 units per day with a known population standard deviation of 4 units. After implementing the training program, the company measures the productivity of a random sample of 30 employees. The sample has an average productivity of 53 units per day. The company wants to know if the new training program has significantly increased productivity.

Solution

1. $H_0 =$ average productivity is 50 and $H_1 =$ average productivity > 50
2. $\alpha = 0.05$
3. Normality is valid and σ is also known
4. Applying Z - test
5. Z-score is

$$Z = \frac{53 - 50}{\frac{4}{\sqrt{30}}} = 4.10$$

6. Plot z-score on normal distribution, and find the area using z-table, in our case it is 0.9998.
7. To get p-value, we just subtract it from 1, i.e. p-value = 0.0002.
8. As p-value $<< 0.05$, we have strong evidence to reject the null hypothesis

Example 2

Suppose a snack food company claims that their Lays wafer packets contain an average weight of 50 grams per packet. To verify this claim, a consumer watchdog organization decides to test a random sample of Lays wafer packets. The organization wants to determine whether the actual average weight differs significantly from the claimed 50 grams. The organization collects a random sample of 40 Lays wafer packets and measures their weights. They find that the sample has an average weight of 49 grams, with a known population standard deviation of 5 grams.

Solution

1. $H_0 =$ average weight is 50 and $H_1 =$ average weight $\neq 50$
2. $\alpha = 0.05$
3. Normality is valid and σ is also known
4. Applying Z - test
5. Z-score is

$$Z = \frac{49 - 50}{\frac{5}{\sqrt{40}}} = -1.26$$

6. Plot z-score on normal distribution, and find the area using z-table, in our case it is 0.103.
7. As it is two sided test, we have to add both areas. Therefore p-value = 0.206.
8. As p-value > 0.05, we do not have evidence to reject null hypothesis.

How To Interpret p-values
How to Calculate p-values
p-values in research
Tailed p-values

t - test

A t-test is a statistical test used in hypothesis testing to compare the means of two samples or to compare a sample mean to a known population mean. The t-test is based on the t-distribution, which is used when the *population standard deviation is unknown* and the *sample size is small*.

Following are the types of t-tests:

1. **One-sample t-test**: The one-sample t-test is used to *compare the mean of a single sample to a known population mean*.
2. **Independent two-sample t-test**: The independent two-sample t-test is used to *compare the means of two independent samples*.

3. **Paired t-test (dependent two-sample t-test)**: The paired t-test is used to *compare the means of two samples that are dependent or paired, such as pre-test and post-test scores for the same group* of subjects or measurements taken on the same subjects under two different conditions.

Single Sample t - test

A one-sample t-test checks whether a sample mean differs from the population mean.

Assumptions

1. Normality or sample size > 30
2. Independence
3. Random Sampling
4. Unknown Population Standard Deviation

Example

Suppose a manufacturer claims that the average weight of their new chocolate bars is
50
grams, we highly doubt that and want to check this so we drew out a sample of 25
chocolate
bars and measured their weight, the sample mean came out to be 49.7 grams and the
sample
std deviation was 1.2 grams. Consider the significance level to be 0.05

Solution

1. $H_0 : \mu = 50$ and $H_1 : \mu \neq 50$
2. $$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = -1.25$$
3. Using `scipy.stats.t` we have calculated the p-value to be $0.11 * 2 = 0.22$.
4. As p-value lies in no reject region, we cannot reject null hypothesis.

Independent Two Sample t - test

An independent two-sample t-test, also known as an unpaired t-test, is a statistical method used to compare the means of two independent groups to determine if there is a significant difference between them.

Assumptions

1. Independence of observations

2. Normality
3. Homoscedasticity (Equal variances)
4. Random sampling

Example

Suppose a website owner claims that there is no difference in the average time spent on their website between desktop and mobile users. To test this claim, we collect data from 30 desktop users and 30 mobile users regarding the time spent on the website in minutes.

Solution

1. $H_0 : \mu_x = \mu_y$ and $H_0 : \mu_x \neq \mu_y$
2. $$t = \frac{\bar{x_a} - \bar{x_b}}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}} = \frac{18.5 - 14.3}{\sqrt{\frac{3.5^2}{30} + \frac{2.7^2}{30}}} = 5.25$$
3. $\text{dof} = n_a + n_b - 2 = 58$
4. Using `scipy.stats.t` we have calculated the p-value to be ≈ 0.
5. As p-value $<< 0.05$, we have evidence to reject null hypothesis

Paired Two Sample t - test

> **See also**
>
> Shapiro Wilks test
> Mann Whitney U test
> F test
> Levene test
> Welch's t test