Hypothesis Testing

Hypothesis testing is a fundamental method in statistics used to evaluate the plausibility of a claim (hypothesis) by analyzing sample data.

There are 5 main steps in hypothesis testing:

1. **Formulating Hypotheses:**
    - **Null Hypothesis (H_o):** This represents the "no effect" scenario. It assumes there's no significant difference between the groups being compared, or that a certain variable doesn't have an impact.
    - **Alternative Hypothesis (H_1):** This is the opposite of the null hypothesis. It suggests there might be a difference between the groups or an effect of the variable you're interested in.
2. **Data Collection:**
   You gather data through experiments, surveys, or other means relevant to your research question. This data serves as the evidence to evaluate the hypotheses.
3. **Choosing a Statistical Test:**
   The type of test you choose depends on your data (categorical, numerical), the number of groups you're comparing, and the nature of your hypothesis. Common tests include:
   * T-tests (for comparing means of two groups)
   * F-tests (for comparing variances of two groups)
   * Chi-square tests (for assessing associations between categorical variables)
   * ANOVA (for comparing means of three or more groups)
4. **Hypothesis Testing Procedure:**
    - You calculate a test statistic based on your chosen test and sample data.
    - You set a significance level (alpha), usually 0.05, which represents the maximum acceptable probability of rejecting the null hypothesis when it's actually true (a type I error).

- Based on the test statistic and the chosen significance level, you determine a p-value. The p-value represents the probability of observing your data (or more extreme data) assuming the null hypothesis is true.

5. **Decision Making:**
    - **Reject H_o:** If the p-value is less than your significance level (e.g., p-value < 0.05), you reject the null hypothesis. This suggests there's evidence against the "no effect" scenario and supports the alternative hypothesis.
    - **Fail to Reject H_o:** If the p-value is greater than your significance level, you fail to reject the null hypothesis. This doesn't necessarily mean the null hypothesis is true, but rather that you don't have enough evidence to reject it with your current data.

Statistical significance

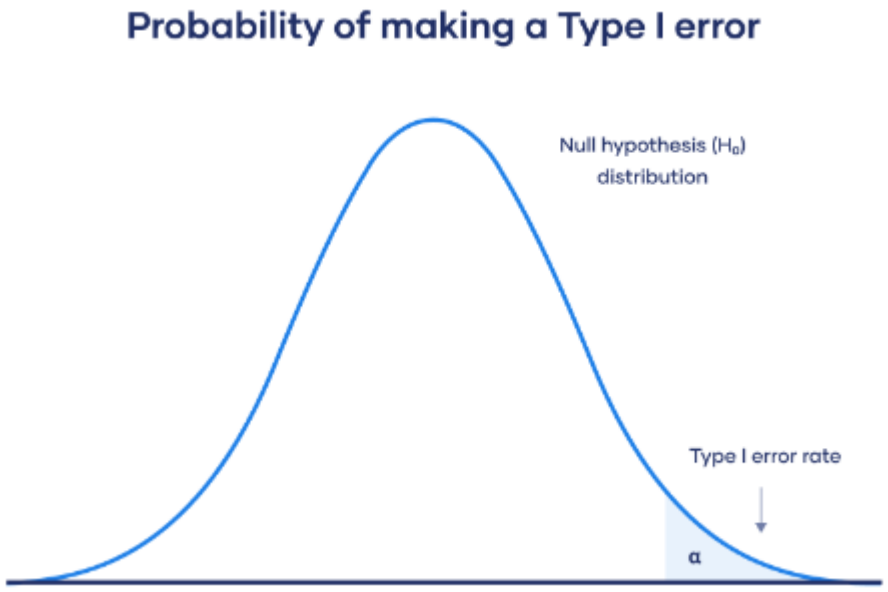
- **The Decision:**
    - **Statistically Significant:** If the p-value is less than your chosen significance level (e.g., p-value < 0.05), we say the result is statistically significant.
    - **Not Statistically Significant:** If the p-value is greater than the significance level (e.g., p-value > 0.05), we fail to reject the null hypothesis. This doesn't necessarily mean there's no effect, but rather that we don't have strong enough evidence to reject the possibility of random chance with the current data.

Type I and Type II errors

Type I error

A Type I error means rejecting the null hypothesis when it’s actually true. It means concluding that results are **statistically significant** when, in reality, they came about purely by chance or because of unrelated factors.

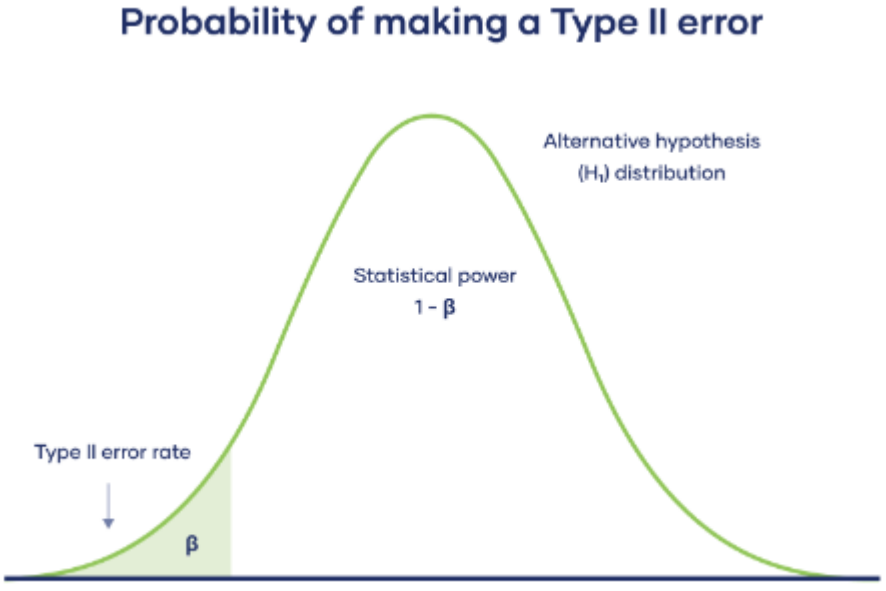
The risk of committing this error is the significance level (alpha or α) you choose. That’s a value that you set at the beginning of your study to assess the statistical probability of obtaining your results (*p* value).



Type II error

A Type II error means not rejecting the null hypothesis when it's actually false. This is not quite the same as "accepting" the null hypothesis, because hypothesis testing can only tell you whether to reject the null hypothesis.

Instead, a Type II error means failing to conclude there was an effect when there actually was. In reality, your study may not have had enough statistical power to detect an effect of a certain size.



This means there's an important tradeoff between Type I and Type II errors:

- Setting a lower significance level decreases a Type I error risk, but increases a Type II error risk.
- Increasing the power of a test decreases a Type II error risk, but increases a Type I error risk.