**Hypothesis Testing:**

* **Hypothesis testing** is a key tool in business analytics used to **evaluate assumptions or claims** about business metrics based on data samples.
* It’s a **structured process** that helps businesses make informed decisions using data from a portion (sample) of their customers, operations, or market.
* Businesses use it to decide whether a **claim—like a new marketing strategy increasing sales—is supported by data**.
* It helps determine whether the observed results are **statistically significant** or just due to random chance, guiding whether to **continue, stop, or change a strategy**.

**The process of hypothesis testing can be broken down into several steps, which include:**

1. **Formulating Hypotheses:**

* **Null hypothesis (H0):**The null hypothesis is the starting assumption in statistics. It says there is no relationship between groups.
* **Alternative hypothesis (H1):** The alternative hypothesis is the opposite of the null hypothesis it suggests there is a difference between groups.
* **Example**: Suppose a company claims that its employees work an average of 40 hours a week.
  + H0​: The average workweek is 40 hours.
  + H1: The average workweek is not 40 hours.

1. **Choosing the Significance Level (α\alpha):**
   * It’s the probability of rejecting the null hypothesis when it's actually true. Commonly set at 0.05 (5%), meaning you're 95% confident in your results.
2. **Selecting the Appropriate Test:**
   * Depending on your **business question** and the **type of data**, you choose a test:
   * For comparing two groups (like sales before and after a campaign), use a **t-test**.
   * For comparing categories (like customer preferences), use a **chi-square test**.
3. **Compare p-value with α:**
   * The p-value tells you how likely your results are if the original assumption (null hypothesis) is true.
   * If p-value < α, your results are unlikely under the current assumption—so you reject the assumption.
   * If p-value ≥ α, there’s not enough evidence to say the assumption is wrong—so you keep it.
4. **Making a Decision:**

 If **p < α** → **Reject the null hypothesis** (your new idea might be better).

 If **p ≥ α** → **Do not reject** the null hypothesis (no strong evidence for change).

1. **Conclusion:**

Based on the test, you decide whether the data supports **making a change** (like launching a new product) or **keeping things the same** (like sticking with your current strategy).

**Types of Hypothesis Tests (for Business Decisions)**

1. **Z-test**

* **When to use:** Large sample size (n > 30) and known population variance.
* **Use case:** Checking if average customer spending matches a set target.

1. **T-test** *(more flexible for smaller samples)*

* **When to use:** Small sample size (n ≤ 30) and unknown variance.
* **Types:**
  + **One-sample t-test:** Is the average sales this month different from last year’s average?
  + **Two-sample t-test:** Do two stores have different average sales?
  + **Paired t-test:** Did average sales improve after a new strategy?

1. **Chi-square Test** *(for categories, not numbers)*

* **When to use:** Testing relationships between **categorical data**.
* **Use cases:**
  + **Goodness-of-fit:** Do customer preferences match what we expected?
  + **Independence:** Is purchase behavior linked to age group?

1. **ANOVA (Analysis of Variance)**

* **When to use:** Comparing **3 or more group averages**.
* **Use case:** Which of several marketing campaigns led to higher customer engagement?

1. **F-test**

* **When to use:** Comparing **variances** (spread or consistency) of two groups.
* **Use case:** Is sales variability different between two regions?

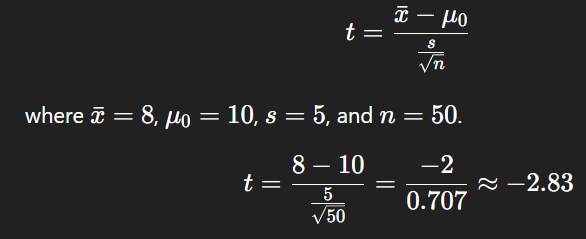
**Types of Errors in Hypothesis Testing:**

1. **Type I Error (False Positive):**
   * Occurs when the null hypothesis is rejected when it is actually true.
   * **What it means:** You think there’s an effect or difference, but there really isn’t.
   * **Example:** Believing a new ad increases sales when it actually doesn’t.
   * **Controlled by:** **α** (significance level, usually 0.05).
2. **Type II Error (False Negative):**
   * Occurs when the null hypothesis is not rejected when it is actually false.
   * **What it means:** You miss a real effect or difference.
   * **Example:** Ignoring a truly effective marketing campaign because the data wasn’t strong enough.
   * **Denoted by:** **β**

**Example to Illustrate Hypothesis Testing:**

**Problem:**  
A pharmaceutical company claims that their new drug lowers blood pressure by at least 10 mmHg. A clinical trial on a sample of 50 patients shows a mean reduction of 8 mmHg with a standard deviation of 5 mmHg. Test the claim at the 5% significance level.

1. **Step 1: Formulate Hypotheses**
   * Null Hypothesis (H0): μ = 10 (The drug lowers blood pressure by 10 mmHg on average.)
   * Alternative Hypothesis (H1): μ < 10(The drug does not lower blood pressure by 10 mmHg on average.)
2. **Step 2: Choose Significance Level**
   * α=0.05
3. **Step 3: Select the Test**
   * Since the sample size is large and the population variance is unknown, use a t-test for a one-sample test.
4. **Step 4: Calculate the Test Statistic**
   * The formula for the t-statistic is:



1. **Step 5: Compare the p-value**
   * Using a t-table or calculator, find the p-value for t=−2.83 with df=49.
   * The p-value is approximately 0.005.
2. **Step 6: Conclusion**
   * Since the p-value (0.005) is less than α=0.05, we reject the null hypothesis.
   * There is sufficient evidence to conclude that the drug does not lower blood pressure by 10 mmHg.

**Limitations of hypothesis testing**:

1. **Limited Focus**: It only tests specific assumptions and may not cover the full complexity of the problem.
2. **Depends on Data Quality**: If the data is poor or inaccurate, the results can be misleading.
3. **May Miss Other Patterns**: It might overlook important trends or relationships not included in the hypothesis.
4. **Lacks Broader Context**: It may ignore the bigger picture, leading to incomplete or oversimplified conclusions.
5. **Needs Other Methods**: To fully understand the data, hypothesis testing should be combined with tools like data visualization or machine learning.