1. What is a null hypothesis (H_0) and why is it important in hypothesis testing?

➔ The **null hypothesis (H_0)** is a statement of "no effect," "no difference," or "no change". It represents the current belief or the status quo that you are trying to challenge.

Why it is important:

- **Starting Point:** It acts as the default assumption or the "innocent until proven guilty" stance in a statistical test.
- **Testing Framework:** Statistical tests are designed to determine whether there is enough evidence in your data to **reject** the null hypothesis.
- **Clarity:** It provides a specific, quantifiable value (like a mean or proportion) to test against, allowing you to calculate the probability (p-value) of seeing your results if the null were actually true.

2. What does the significance level (α) represent in hypothesis testing?

➔ The **significance level**, denoted by the Greek letter alpha (α), is the threshold you set to decide whether the evidence is strong enough to reject the null hypothesis.

   a. **The Probability of Error**: It represents the maximum risk you are willing to take of committing a **Type I error** (rejecting the null hypothesis when it is actually true).
   b. **Common Values**: In data analysis, α is typically set at **0.05 (5%)**, **0.01 (1%)**, or **0.10 (10%)** before the test begins.
   c. **The "P-value" Gatekeeper**: If your calculated p-value is less than or equal to α ($p \le \alpha$), you reject the null hypothesis. If it is greater, you fail to reject it.

3. Differentiate between Type I and Type II errors.

Feature	Type I Error (α)	Type II Error (β)
Definition	Rejecting H0 when it is actually true.	Failing to reject H0 when it is actually false.
Common Name	A "False Positive".	A "False Negative".
Analogy	Convicting an innocent person.	Letting a guilty person go free.
Control	Controlled by setting the significance level (α).	Reduced by increasing the sample size or the "power" of the test.

4. Explain the difference between a one-tailed and two-tailed test. Give an example of each.

➔ The "tails" refer to the regions of the probability distribution where the null hypothesis is rejected.

   a. **One-Tailed Test**: You are looking for a change in a **specific direction** (either an increase or a decrease).

      Example: Testing if a new fertilizer makes plants grow **taller** than the current average.

   b. **Two-Tailed Test**: You are looking for **any difference**, regardless of direction.

      Example: Testing if a machine is filling cereal boxes with an amount **different** (either more or less) than the labeled 500g.

5. A company claims that the average time to resolve a customer complaint is 10 minutes. A random sample of 9 complaints gives an average time of 12 minutes and a standard deviation of 3 minutes. At $\alpha = 0.05$, test the claim.

➔ At $\alpha = 0.05$, test the claim.
To solve this, we use a **t-test** because the sample size is small ($n < 30$) and the population standard deviation is unknown.

1. State the Hypotheses:

- H0: u = 10 (The claim is true)
- Ha: u /10 (The average time is different—Two-tailed test)

2. Identify the Givens:

- Sample Mean (bar{x}) = 12
- Population Mean (u) = 10
- Sample Standard Deviation (s) = 3
- Sample Size (n) = 9
- Significance Level (α) = 0.05

3. Calculate the Test Statistic (t):

t = (bar{x} - u)/(s / sqrt(n)) = (12 - 10)/(3 / sqrt(9)) = (2)/(3/3) = 2.0
4. Determine the Critical Value:

With degrees of freedom df = n - 1 = 8 and $\alpha = 0.05$ (two-tailed), the critical t-value from the table is approximately **2.306**.

5. Conclusion:

Since our calculated t (2.0) is **less than** the critical value (2.306), we **fail to reject the null hypothesis**. There is not enough evidence at the 5% level to say the average resolution time is different from 10 minutes.

6. When should you use a Z-test instead of a t-test?

➔ Choosing between these two tests depends on what you know about the population and how much data you have.

- **Use a Z-test when:**
    a. You know the **population standard deviation** (σ).
    b. Your **sample size is large** ($n >= 30$), even if the population standard deviation is unknown (the sample standard deviation becomes a good proxy).
- **Use a t-test when:**
    a. The **population standard deviation is unknown**.
    b. Your **sample size is small** ($n < 30$).

7. The productivity of 6 employees was measured before and after a training program.

Employee	Before	After
1	50	55
2	60	65
3	58	59
4	55	58
5	62	63
6	56	59

At $\alpha = 0.05$, test if the training improved productivity.

➔ This is a **Paired t-test** because we are comparing the same subjects (employees) at two different times.

1. State the Hypotheses:

- H0: ud <= 0 (The training did not improve productivity).
- Ha: ud > 0 (The training improved productivity—One-tailed test).

2. Calculate the Differences (d = After - Before):

- Employee 1: 55 - 50 = 5
- Employee 2: 65 - 60 = 5
- Employee 3: 59 - 58 = 1
- Employee 4: 58 - 55 = 3
- Employee 5: 63 - 62 = 1
- Employee 6: 59 - 56 = 3

3. Find the Mean (bar(d)) and Standard Deviation (sd) of Differences:

- bar(d) = (5+5+1+3+1+3) / 6 = 18 / 6 = 3
- sd ~1.897$

4. Calculate the Test Statistic (t):

t = (bar(d)/(sd / sqrt(n)) = (3)/(1.897 / sqrt(6)) ~(3)/(0.774) ~3.87

5. Conclusion:

With df = 5 and α= 0.05 (one-tailed), the critical value is **2.015**. Since 3.87 > 2.015, we **reject the null hypothesis**. The training significantly improved productivity.

8. A company wants to test if product preference is independent of gender.

Gender	Product A	Product B	Total
Male	30	20	50
Female	10	40	50
Total	40	60	100

At α = 0.05, test independence

➔ To determine if two categorical variables (Gender and Product Preference) are independent, we use a **Chi-Square Test for Independence**.

1. State the Hypotheses

- **H0**: Product preference is independent of gender.
- **Ha**: Product preference is dependent on gender.

2. The Data Table

Gender	Product A	Product B	Total
Male	30	20	50
Female	10	40	50
Total	40	60	100

3. Calculate Expected Values (E)

The formula for each cell is: E = [(Row Total) * (Column Total)]/(Grand Total)

- **Male / Product A**: (50 * 40) / 100 = 20
- **Male / Product B**: (50 * 60) / 100 = 30
- **Female / Product A**: (50 * 40) / 100 = 20
- **Female / Product B**: (50 * 60) / 100 = 30

4. Calculate the Chi-Square Statistic (chi^2)

The formula is: chi^2 = ∑(O-E)^2/E

- {(30 - 20)^2}/{20} = {100}/{20} = 5
- {(20 - 30)^2}/{30} = {100}/{30} ~= 3.33
- {(10 - 20)^2}/{20} = {100}/{20} = 5
- {(40 - 30)^2}/{30} = {100}/{30} ~= 3.33
- **Total chi^2** ~= 16.66

5. Conclusion

- **Degrees of Freedom (df):** (R-1)(C-1) = (2-1)(2-1) = 1.
- **Critical Value:** For α = 0.05 and df = 1, the critical value is **3.841**.
- **Result:** Since 16.66 is much greater than 3.841, we **reject the null hypothesis**.