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

ANSWER 1:- The null hypothesis, denoted as H_0, is a formal statement in statistics that assumes no effect, no difference, or no relationship exists in the population being studied. It represents the default or baseline position that researchers test against.

- Example:
    1. If you're testing whether a new drug lowers blood pressure, H_0 would state: "The drug has no effect on blood pressure compared to the placebo."
- Role in Hypothesis Testing:
- Hypothesis testing is essentially about weighing evidence against H_0.
    1. Researchers collect data and calculate test statistics.
    2. If the evidence is strong enough (usually measured by a p-value), they reject H_0.
    3. If not, they fail to reject H_0.
- Why It's Important:
    1. Provides a baseline: It sets a clear starting point for statistical inference.
    2. Ensures objectivity: By assuming no effect initially, researchers avoid bias and only claim an effect if data strongly supports it.
    3. Controls error rates: Hypothesis testing is designed to minimize Type I errors (false positives), which means rejecting H_0 only when evidence is convincing.
    4. Framework for comparison: It allows researchers to compare observed results with what would be expected if no effect truly existed.

In short, the null hypothesis is the anchor of scientific testing—without it, we wouldn't have a structured way to decide whether our findings are meaningful or just due to chance.

Would you like me to also break down how the null hypothesis differs from the alternative hypothesis (H_1) with a simple example? That contrast often makes the concept even clearer.

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

ANSWER 2:- α is the probability of making a Type I error — rejecting the null hypothesis when it is actually true.

- Common values:
    - 0.05 (5%) → Most widely used
    - 0.01 (1%) → More stringent
    - 0.10 (10%) → More lenient

Think of α as the risk you're willing to take of falsely claiming a discovery:

- If $\alpha = 0.05$, you're saying:

"I'm okay with a 5% chance of wrongly rejecting the null hypothesis."

- You calculate a p-value from your data.
- If p-value $\leq \alpha$, you reject H_0 → result is statistically significant.
- If p-value $> \alpha$, you fail to reject H_0 → not enough evidence.
- Controls false positives: Helps ensure that findings are not just due to random chance.
- Set the bar for evidence: Higher α means you're more willing to accept risk; lower α means you demand stronger evidence.
- Used to interpret results: It's the benchmark for deciding whether your results are meaningful.

Question 3: Differentiate between Type I and Type II errors.

ANSWER 3:- These two errors represent the risks we take when making decisions based on statistical tests. Here's how they differ:

Error Type	Definition	Occurs When…	Risk Controlled By
Type I Error (False Positive)	Rejecting the null hypothesis when it is actually true	You think there's an effect or difference, but there isn't	Significance level (α)
Type II Error (False Negative)	Failing to reject the null hypothesis when it is actually false	You miss a real effect or difference	Power of the test ($1 - \beta$)

Suppose you're testing a new drug:

- H_0: The drug has no effect.
- H_1: The drug works better than placebo.
- Type I Error: You conclude the drug works when it actually doesn't.
- Type II Error: You conclude the drug doesn't work when it actually does.

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

ANSWER 4:- Hypothesis tests can be directional (one-tailed) or non-directional (two-tailed) depending on how the alternative hypothesis is framed.

ONE - Tailed Test

- Definition: Tests for an effect in one specific direction (greater than OR less than).
- Alternative Hypothesis (H_1): Specifies a direction.
- Use Case: When you only care about whether the effect goes one way.

Example:

A company claims their new battery lasts longer than 10 hours.

- H_0: Battery life ≤ 10 hours
- H_1: Battery life > 10 hours

This is a right-tailed test because we only care if it's greater.

Two-Tailed Test

- Definition: Tests for an effect in either direction (different, not equal).
- Alternative Hypothesis (H_1): States there is a difference, but not the direction.
- Use Case: When you want to detect any deviation from the null, whether higher or lower.

Example:

A pharmaceutical company tests if a new drug has a different effect on blood pressure compared to placebo.

- H_0: Drug effect = Placebo effect
- H_1: Drug effect ≠ Placebo effect

This is a two-tailed test because the drug could lower or raise blood pressure.

Question 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.

ANSWER 5:- Null Hypothesis (H_0): \mu = 10 (average resolution time is 10 minutes).

- Alternative Hypothesis (H_1): $\mu \neq 10$ (average resolution time is different from 10 minutes).

This is a two-tailed test because the claim is about equality.

Sample Statistics

- Sample mean ($\bar{x}$) = 12
- Sample size (n) = 9
- Sample standard deviation (s) = 3

Test Statistic

Formula:

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

$$t = \frac{12 - 10}{3 / \sqrt{9}} = \frac{2}{3/3} = \frac{2}{1} = 2$$

So, $t = 2.0$.

Degrees of Freedom

$df = n - 1 = 9 - 1 = 8$

Critical Value at α = 0.05 (two-tailed)

From the t-distribution table:

$t_{0.05, df=8} \approx 2.306$

Decision

- Calculated $t = 2.0$
- Critical value = 2.306
- Since $2.0 < 2.306$, we fail to reject H_0.

At the 5% significance level, there is not enough evidence to conclude that the average resolution time differs from 10 minutes. The company's claim stands.

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

ANSWER 6:- Z-Test vs T-Test — When to Use Each

The choice between a Z-test and a t-test depends mainly on sample size and whether the population standard deviation (σ) is known.

Use a Z-test when:

1. Population standard deviation (σ) is known.
2. Sample size is large (typically $n \geq 30$), so the sampling distribution of the mean is approximately normal by the Central Limit Theorem.
3. The population is normally distributed (or approximately normal for large samples).

Example:

A factory knows the population standard deviation of the weight of bolts is 0.5g. To test if the average weight differs from 10g, and the sample size is 50, a Z-test is appropriate.

Use a t-test when:

1. Population standard deviation (σ) is unknown.
2. Sample size is small ($n < 30$).
3. The population is normally distributed (important for small samples).

Example:

If you only have 12 samples of complaint resolution times and don't know the population standard deviation, you use a t-test.

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

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.

ANSWER 7:- Perfect — this is a paired sample t-test problem since we're comparing the same employees before and after training. Let's solve it step by step.

State Hypotheses

- Null Hypothesis (H_0): Training has no effect on productivity ($\mu_d = 0$).
- Alternative Hypothesis (H_1): Training improves productivity ($\mu_d > 0$).

This is a one-tailed test.

Compute Differences

$d = \text{After} - \text{Before}$

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

- Differences: $5, 5, 1, 3, 1, 3$

Mean and Standard Deviation of Differences

- Mean difference:

$$\bar{d} = \frac{5+5+1+3+1+3}{6} = \frac{18}{6} = 3$$

- Variance of differences:

$$s_d^2 = \frac{\sum (d_i - \bar{d})^2}{n-1}$$

Deviations:

- $(5-3)^2 = 4$
- $(5-3)^2 = 4$
- $(1-3)^2 = 4$
- $(3-3)^2 = 0$
- $(1-3)^2 = 4$
- $(3-3)^2 = 0$

Sum = 16

$$s_d^2 = \frac{16}{5} = 3.2 \quad \Rightarrow \quad s_d = \sqrt{3.2} \approx 1.789$$

Test Statistic

$$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}}$$

$$t = \frac{3}{1.789 / \sqrt{6}} = \frac{3}{1.789 / 2.449} = \frac{3}{0.730} \approx 4.11$$

Degrees of Freedom

$df = n - 1 = 6 - 1 = 5$

Critical Value at $\alpha = 0.05$ (one-tailed)

From t-table:

$t_{0.05, df=5} \approx 2.015$

Decision

- Calculated $t = 4.11$
- Critical value = 2.015
- Since $4.11 > 2.015$, reject H_0.

At the 5% significance level, there is strong evidence that the training program improved employee productivity.

Question 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 $\alpha = 0.05$, test independence

ANSWER 8:- State Hypotheses

- Null Hypothesis (H_0): Product preference is independent of gender.
- Alternative Hypothesis (H_1): Product preference depends on gender.

Observed Frequencies (O)

From the table:

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

Expected Frequencies (E)

Formula:

$$E_{ij} = \frac{(\text{Row Total})(\text{Column Total})}{\text{Grand Total}}$$

- Male, Product A: $\frac{50 \cdot 60}{100} = 30$
- Male, Product B: $\frac{50 \cdot 40}{100} = 20$
- Female, Product A: $\frac{50 \cdot 60}{100} = 30$
- Female, Product B: $\frac{50 \cdot 40}{100} = 20$

So expected table:

Gender	Product A	Product B
Male	30	20
Female	30	20

Chi-Square Test Statistic

Formula:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

- Male, Product A: $\frac{(20-30)^2}{30} = \frac{100}{30} \approx 3.33$
- Male, Product B: $\frac{(30-20)^2}{20} = \frac{100}{20} = 5$
- Female, Product A: $\frac{(40-30)^2}{30} = \frac{100}{30} \approx 3.33$
- Female, Product B: $\frac{(10-20)^2}{20} = \frac{100}{20} = 5$

$$\chi^2 = 3.33 + 5 + 3.33 + 5 = 16.66$$

Degrees of Freedom

$$df = (r-1)(c-1) = (2-1)(2-1) = 1$$

Critical Value at α = 0.05

From chi-square table:

$$\chi^2_{0.05,1} = 3.841$$

Decision

- Calculated $\chi^2 = 16.66$
- Critical value = 3.841
- Since $16.66 > 3.841$, reject H_0.

At the 5% significance level, product preference is not independent of gender. In other words, gender has a statistically significant association with product preference.