Name – Md Anas

Course Name - Data analytics with GenAi

Topic - Hypothesis Testing

Email id - anasanwar1120@gmail.com

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

- Definition: The null hypothesis states that there is *no significant difference or relationship* between variables in a population. For example, "There is no difference in exam scores between students who study with music and those who study in silence." Notation: It is usually written as H_0.
- Contrast: The alternative hypothesis (H_1 or Ha) suggests that a difference or relationship *does exist*.

Why is the Null Hypothesis Important?

1. Provides a Starting Point:
    - Hypothesis testing begins with the assumption that H_0 is true. This makes analysis objective and avoids bias.
2. Basis for Statistical Tests:
    - Tests like the *t-test, chi-square test, ANOVA* are designed to evaluate whether data provide enough evidence to reject H_0.
3. Controls for Random Chance:
    - By assuming no effect, researchers can measure whether observed differences are statistically significant or just due to sampling error.
4. Scientific Rigor:

- It ensures that claims of effects or differences are backed by evidence rather than assumptions. Without H_0, results could be misinterpreted

Example

Imagine you want to test if a new teaching method improves student performance:

- Null Hypothesis (H_0): The new method has *no effect* on student scores.
- Alternative Hypothesis (H_1): The new method *does improve* student scores.
- If statistical analysis shows a significant difference, you reject H_0 in favor of H_1

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

Answer

- Definition: α is the cutoff point for deciding whether to reject H_0. Type I Error: It is the probability of falsely concluding that an effect exists when it doesn't.
- Common Values: Researchers often set α at 0.05 (5%), meaning they accept a 5% risk of wrongly rejecting H_0. Other common choices are 0.01 (1%) and 0.10 (10%) depending on the field and context.

Why α is Important

1. Controls Risk of False Positives:
    - By setting α, researchers limit the chance of incorrectly claiming a discovery.
2. Defines "Statistical Significance":
    - If the p-value $\leq \alpha$, results are considered statistically significant, and H_0 is rejected.
    - If the p-value $> \alpha$, results are not significant, and H_0 is not rejected.
3. Balances Sensitivity and Specificity:

- A smaller α (like 0.01) reduces false positives but increases the chance of missing real effects (Type II error).
- A larger α (like 0.10) increases sensitivity to detect effects but risks more false positives.

Example

Suppose you test whether a new medicine lowers blood pressure:

- H_0: The medicine has no effect.
- $\alpha = 0.05$: You accept a 5% risk of wrongly rejecting H_0.
- If your test produces a p-value = 0.03, since $0.03 < 0.05$, you reject H_0 and conclude the medicine likely works.

Question 3 Differentiate between Type I and Type II errors

Type I Error (False Positive)

- Definition: Rejecting the null hypothesis (H_0) when it is actually true.
- Analogy: A fire alarm goes off (positive result) when there's no fire (true state).
- Symbol: Alpha (α).
- Example: A medical test says a healthy person has a disease.

Type II Error (False Negative)

- Definition: Failing to reject the null hypothesis (H_0) when it is actually false.
- Analogy: A fire alarm doesn't go off (negative result) when there is a fire (false state).
- Symbol: Beta (β).
- Example: A medical test says a sick person is healthy, missing the disease.

Key Differences & Relationship

- False Positive vs. False Negative: Type I is a "false hit," while Type II is a "miss".
- Inverse Relationship: You can't easily minimize both; reducing the chance of one often increases the chance of the other.
- Consequences: Type I might lead to unnecessary action (e.g., useless treatment), while Type II leads to missed opportunities (e.g., untreated disease).

Quick Reference Table

	Type I Error (α)	Type II Error (β)
Action	Reject true H_0	Fail to reject false H_0
Result	False Positive	False Negative
Example	Convicting an innocent person	Letting a guilty person go free

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

Answer One-Tailed Test

- Definition: Tests for an effect in only one direction (greater than OR less than).

- Use Case: When you have a specific directional hypothesis.
- Decision Rule: Reject H_0 only if the test statistic falls into the extreme tail of the distribution in the predicted direction.

Example:
Suppose you want to test if a new teaching method improves student scores compared to the traditional method.

- H_0: The new method does not improve scores (mean ≤ traditional).
- H_1: The new method improves scores (mean > traditional).
  👉 This is a right-tailed test because you only care if scores are higher.

🔹 Two-Tailed Test

- Definition: Tests for an effect in both directions (greater than OR less than).
- Use Case: When you only want to know if there is any difference, without predicting the direction.
- Decision Rule: Reject H_0 if the test statistic falls into either tail of the distribution.

Example:
Suppose you want to test if a new drug has any effect on blood pressure (could increase or decrease).

- H_0: The drug has no effect (mean = control).
- H_1: The drug has an effect (mean ≠ control).
  👉 This is a two-tailed test because you care about both higher and lower outcome

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 α = 0.05, test the claim.

$\{\bar{x}\}$=12

n=9

S.D=3

a=0.05

h_0=10

$(\bar{x} - \mu) / (s / \sqrt{m})$

$t = (12 - 10) / (3 / \sqrt{9})$

t=2/1

t=2

Degrees of freedom (df) = n – 1 = 8

For a two-tailed test at α = 0.05

t-value ≈ ±2.306

2>2.306,

so fail to reject H_0.

Question6 When should you use a Z-test instead of a t-test

Use a Z-test when:

- Population σ is known: This is the primary condition, regardless of sample size.
- Sample size is large (n ≥ 30): Even if σ is unknown, the sample standard deviation (s) is a good enough estimate, and the Z-distribution approximates the t-distribution well enough.
- Testing proportions: Z-tests are always used for proportions (e.g., percentages).

Use a T-test when:

- Population σ is unknown: You must estimate it with the sample standard deviation (s).
- Sample size is small (n < 30): The t-distribution accounts for the increased uncertainty from estimating σ.
- Sample size is moderate: When σ is unknown, t-tests are generally preferred.

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

[assignment hypothesis.xlsx]

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

[assignment hypothesis.xlsx]