

Statistics Advanced - 2

Assignment

Question 1: What is hypothesis testing in statistics?

Answer: Hypothesis testing in statistics is a method used to make decisions or draw conclusions about a population based on data from a sample. It helps us decide whether the evidence we have from data is strong enough to support or reject a claim (hypothesis) about the population.

● Basic Concepts

1. Hypothesis:

A statement about a population parameter (like mean, proportion, etc.) that we want to test.

2. Types of Hypotheses:

a. Null Hypothesis (H_0):

The default assumption — there is *no effect* or *no difference*.
Example: “The average height of men is 170 cm.”

b. Alternative Hypothesis (H_1 or H_a):

What we want to prove — there *is* an effect or difference.
Example: “The average height of men is *not* 170 cm.”

● Steps in Hypothesis Testing

1. State the hypotheses:

Define H_0 and H_1 , clearly.

2. Set the significance level (α):

Common values: 0.05, 0.01

→ It represents the probability of rejecting H_0 when it is actually true (Type I error).

3. Collect data & calculate the test statistic:

Use formulas (like *z-score*, *t-score*, *chi-square*, etc.) depending on data type

and sample size.

4. Find the p-value:

The p-value tells how likely it is to get the observed data if H_0 is true.

5. Make a decision:

- If $p \leq \alpha$, reject $H_0 \rightarrow$ there's enough evidence for H_1 ,
- If $p > \alpha$, fail to reject $H_0 \rightarrow$ not enough evidence for H_1 .

Example

Suppose a company claims that their bulbs last **1000 hours on average**.

You test 50 bulbs and find a sample mean of **980 hours**, with a known population standard deviation of **60 hours**.

You can test:

- $H_0: \mu = 1000$
- $H_1: \mu \neq 1000$
at $\alpha = 0.05$ using a **z-test**.

If the p-value $< 0.05 \rightarrow$ you reject H_0 and conclude the bulbs don't last 1000 hours on average

Common Tests

- **Z-test:** large sample, population σ known
- **T-test:** small sample, population σ unknown
- **Chi-square test:** for categorical data
- **ANOVA:** compare means of more than two groups

Question 2: What is the null hypothesis, and how does it differ from the alternative hypothesis?

Answer:

- Null Hypothesis (H_0)

The null hypothesis is a statement that assumes no effect, no difference, or no change exists in the population.

It represents the status quo — the claim or condition we start with before seeing any data.

Example:

- A company claims the average battery life = 10 hours.
→ $H_0: \mu = 10$

- Alternative Hypothesis (H_1 , or H_a)

The alternative hypothesis is what we want to test or prove. It suggests that something has changed, a difference exists, or an effect is present.

Example:

- You suspect the battery life is not really 10 hours.
→ $H_1: \mu \neq 10$

- Key Differences

Feature	Null Hypothesis (H_0)	Alternative Hypothesis (H_1 , / H_a)
---------	------------------------------	---

Meaning	No effect / No difference	There is an effect / Difference exists
Purpose	Default assumption	Competing claim (what you want to prove)
Symbol	H_0	H_1 , or H_a
Example (mean test)	$\mu = 10$	$\mu \neq 10$, $\mu > 10$, or $\mu < 10$
Decision	“Fail to reject” or “Reject”	“Supported” if H_0 is rejected

● Example

Imagine a new medicine:

- H_0 : The new medicine has *no effect* (same as old one).
- H_1 : The new medicine *does work better*.

After testing:

- If data shows strong evidence → you **reject H_0** and accept H_1 .
- If not → you **fail to reject H_0** (keep assuming the old medicine works the same)

Question 3: Explain the significance level in hypothesis testing and its role in deciding the outcome of a test?

Answer:

◆ Significance Level (α)

The **significance level**, denoted by α (alpha), is the **threshold** we use in hypothesis testing to decide whether to **reject the null hypothesis (H_0)** or not.

It represents the **probability of making a Type I error**, i.e., **rejecting H_0 when it is actually true**.

The significance level tells us **how much risk we are willing to take** of being wrong when we reject the null hypothesis.

● Common Values of α

α (Significance Level)	Confidence Level	Meaning
0.05	95%	5% risk of being wrong
0.01	99%	1% risk of being wrong
0.10	90%	10% risk of being wrong

👉 Most research studies use $\alpha = 0.05$ (5% level of significance).

◆ Role in Hypothesis Testing

1. You set α before testing (example: 0.05).
2. You calculate the **p-value** from your sample data.
3. Then you **compare p-value with α** :

Decision Rule	Interpretation
If $p \leq \alpha$	Reject $H_0 \rightarrow$ evidence is strong enough to support H_1 ,
If $p > \alpha$	Fail to reject $H_0 \rightarrow$ not enough evidence against H_0

◆ Example

A company claims its bulbs last 1000 hours.
You test a sample and get **p-value = 0.03**, using $\alpha = 0.05$.

→ Since $0.03 < 0.05$,
You **reject H_0** , meaning there is **significant evidence** that the bulbs don't last 1000 hours.

◆ Summary Table

Concept	Description
Symbol	α
Typical values	0.05, 0.01, 0.10
Meaning	Probability of rejecting H_0 when it's actually true
Decision rule	Compare p-value to α
Role	Sets the cutoff for determining statistical significance

Question 4: What are Type I and Type II errors? Give examples of each?

Answer:

• Type I and Type II Errors in Hypothesis Testing

Whenever we make a decision based on sample data, there's a chance of being wrong.
These wrong decisions are called Type I and Type II errors.

• The Two Possible Errors

Decision	Reality (H_0 true)	Reality (H_0 false)
Reject H_0	Type I Error	Correct decision
Fail to reject H_0	Correct decision	Type II Error

◆ Type I Error (α error)

- Meaning: Rejecting the null hypothesis when it is actually true.
- Symbol: α (same as the significance level)
- In simple words:
You think there's an effect or difference — but actually, there isn't.

Example:

A company tests a new medicine.

- H_0 : The medicine has no effect.
- Reality: The medicine truly has no effect.
- Error: You reject H_0 and claim it works — but it doesn't.

 → You conclude the medicine works (wrongly).

◆ Type II Error (β error)

- **Meaning:** Failing to reject the null hypothesis when it is actually false.
- **Symbol:** β
- **In simple words:**
You miss a real effect — you fail to detect something that actually exists.

Example:

Same medicine test:

- H_0 : The medicine has no effect.
- **Reality:** The medicine *does* work.
- **Error:** You fail to reject H_0 — and say it doesn't work.

 → You miss out on a truly effective medicine.

◆ Summary Table

Error Type	Definition	Symbol	Example
Type I Error	Rejecting H_0 when it's true	α	Approving a useless medicine
Type II Error	Failing to reject H_0 when it's false	β	Rejecting a useful medicine

Question 5: What is the difference between a Z-test and a T-test? Explain when to use each?

Answer:

● Z-Test vs T-Test — Overview

Both Z-tests and T-tests are used in hypothesis testing to compare sample data with population parameters —
but they differ mainly in sample size and knowledge of population standard deviation (σ).

1. Z-Test

● When to Use

- The sample size is large ($n \geq 30$)
- The population standard deviation (σ) is known
- Data follows an approximately normal distribution

Examples

- Testing whether the mean weight of a large shipment = 50 kg
- Checking if the average exam score differs from 70 when $\sigma = 10$ is known

2. T-Test

• When to Use

- The sample size is small ($n < 30$)
- The population standard deviation (σ) is unknown (we use sample s instead)
- The population is approximately normal

Examples

- Comparing the average marks of 20 students with the expected mean
- Checking if a small group's average height differs from a known value

• Key Differences Table

Feature	Z-Test	T-Test
Population standard deviation (σ)	Known	Unknown (use sample s)
Sample size	$n \geq 30$ (large)	$n < 30$ (small)
Distribution used	Standard Normal (Z)	Student's T-distribution
Shape of curve	Fixed (bell-shaped)	Changes with sample size (wider for small n)
Example	Quality control in large batches	Testing new teaching method on small class

