
ASSIGNMENT TITLE:
HYPOTHESIS TESTING

Submitted by:

Name: Hajera Shams

Course: Data Analytics With AI

Batch: September 2025 Live (Pro Batch)

Date of submission

November 26,2025

ASSIGNMENT QUESTIONS

HYPOTHESIS TESTING

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

Ans. The null hypothesis (H_0) is a statement that there is no effect, no difference, or no relationship between the parameters of interest. It represents the "status quo" or the widely accepted belief. For example, $H_0: \mu = 10$ (the population mean is 10).

Importance of H_0

H_0 is crucial because hypothesis testing is a proof by contradiction. The entire process revolves around testing the assumption that H_0 is true. We collect sample data and use statistical methods to determine if there is enough evidence to reject H_0 in favor of an alternative hypothesis (H_a or H_1). If the sample evidence is very unlikely to occur under the assumption that H_0 is true, we reject H_0 .

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

Ans. The significance level (α) is the maximum probability of committing a Type I error that a researcher is willing to accept. It is the probability of rejecting the null hypothesis (H_0) when it is actually true.

Common values for α are 0.05 (5%), 0.01 (1%), or 0.10 (10%).

1. If the p-value of the test is less than or equal to α ($p \leq \alpha$),

α), we reject H_0 .

2. α defines the critical region (or rejection region) in the sampling distribution.

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

Ans.

Feature	Type I Error (α)	Type II Error (β)
Definition	Rejecting the null hypothesis (H_0) when it is actually true.	Failing to reject the null hypothesis (H_0) when it is actually false.

Mistake	False Positive	False Negative
Consequence	Concluding an effect/difference exists when it doesn't.	Failing to detect an effect/difference that does exist.
Probability	The significance level (α)	Denoted by β

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

Ans. The difference lies in the alternative hypothesis (H_a) and the

rejection region.

One-Tailed Test (Directional)

1. Alternative Hypothesis (H_a): Specifies a direction (either greater than $>$ or less than $<$ the hypothesized value).

2. Rejection Region: Located entirely in one tail (either the right or the left) of the sampling distribution.

3. Example: Testing if the average product lifespan is greater than 5 years ($H_a: \mu > 5$).

Two-Tailed Test (Non-Directional)

1. Alternative Hypothesis (H_a): States that the parameter is simply not equal to the hypothesized value (\neq).

2. Rejection Region: Split between both tails (the far right and the far left) of the sampling distribution.

Example: Testing if the average salary is different from \$50,000 ($H_a: \mu \neq \$50,000$).

Question 5: A company claims that the average time to resolve a The 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.

Ans. This is a t-test for a single mean since the population standard deviation (σ) is unknown and the sample size is small ($n < 30$).

Given: $\mu_0 = 10$, $n = 9$, $\bar{x} = 12$, $s = 3$, $\alpha =$

0.05\$.

Degrees of Freedom (df): $n - 1 = 9 - 1 = 8$.

1. State Hypotheses

$H_0: \mu = 10$ (The average time is 10 minutes)

$H_a: \mu \neq 10$ (The average time is different from 10 minutes) \rightarrow Two-tailed test

2. Calculate the Test Statistic (t)

The formula for the t-test statistic is:

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

$$t = \frac{12 - 10}{3 / \sqrt{9}} = \frac{2}{3 / 3} = \frac{2}{1} = 2.0$$

\$\$

3. Determine the Critical Value (or P-value)

1. For a two-tailed t-test with $df = 8$ and $\alpha = 0.05$, the critical t-values are found to be ± 2.306 .

4. Make a Decision

1. Critical Value Method: Since $|t_{\text{calc}}| = 2.0$ is less than $|t_{\text{crit}}| = 2.306$, the test statistic falls in the non-rejection region.

2. Decision: Do not reject H_0 .

5. Conclusion

At the 0.05 significance level, there is not enough evidence to reject the company's claim that the average time to resolve a customer

The complaint is 10 minutes long.

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

Ans. A Z-test for a mean should be used instead of a t-test when at least one of the following two conditions is met:

1. The population standard deviation (σ) is known.

Note: The t-test is used when σ is unknown and the sample standard deviation (s) is used as an Estimate.

Question 7: The productivity of 6 employees was measured before and after a training program. At $\alpha = 0.05$, test if the training improved productivity.

Ans. This is a Paired Samples t-test (Dependent Samples t-test) because the same employees are measured *before* and *after* the training. We analyze the difference (d) between the paired observations.

Employee	Before (XB)	After(XA)	Difference ($d=XA-XB$)	d^2
1	50	55	5	25
2	60	65	5	25
3	58	59	1	1
4	55	58	3	9

--	--	--	--	--

Employee	Before (XB)	After(XA)	Difference (d=XA-XB)	d ²
5	62	63	1	25
6	56	59	3	25
Sum (\$\sum\$)			18	70

Given: $n = 6$, $\alpha = 0.05$.

Degrees of Freedom (df): $n - 1 = 6 - 1 = 5$.

Calculations

Mean Difference (\bar{d}): $\bar{d} = \frac{\sum d}{n} = \frac{18}{6} = 3$

Standard Deviation of Differences (s_d):

$$s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}} = \sqrt{\frac{70 - \frac{(18)^2}{6}}{5}} = \sqrt{\frac{70 - \frac{324}{6}}{5}}$$

$$s_d = \sqrt{\frac{70 - 54}{5}} = \sqrt{\frac{16}{5}} = \sqrt{3.2} \approx 1.7889$$

1. State Hypotheses

We are testing if the training improved productivity, meaning "After" should be greater than "Before", or $d = X_A - X_B$ should be

greater than 0.

$H_0: \mu_d \leq 0$ (Training did not improve productivity)

$H_a: \mu_d > 0$ (Training improved productivity)

\rightarrow Right-tailed test

2. Calculate the Test Statistic (t)

The formula for the paired t-test statistic is:

$$t = \frac{\bar{d} - \mu_{d0}}{s_d / \sqrt{n}}$$

(Assuming $\mu_{d0} = 0$)

$$t = \frac{3 - 0}{1.7889 / \sqrt{6}} = \frac{3}{1.7889 / 2.4495} = \frac{3}{0.7303} \approx 4.108$$

3. Determine the Critical Value

For a right-tailed t-test with $df = 5$ and $\alpha = 0.05$, the critical t-value is $t_{\text{crit}} = 2.015$.

4. Make a Decision

Critical Value Method: Since $t_{\text{calc}} = 4.108$ is greater than $t_{\text{crit}} = 2.015$, the test statistic falls in the rejection region.

Decision: Reject H_0 .

5. Conclusion

At the 0.05 significance level, there is sufficient evidence to conclude 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.

Ans. This is a Chi-Square Test for Independence (χ^2).

1. State Hypotheses

H_0 : Product preference is independent of gender.

H_a : Product preference is dependent on gender.

2. Calculate Expected Frequencies (E)

The expected frequency for each cell is $E = (\text{Row Total} \times \text{Column Total}) / \text{Grand Total}$.

$$E_{\{\text{Male, A}\}} = (50 \times 40) / 100 = 20$$

$$E_{\{\text{Male, B}\}} = (50 \times 60) / 100 = 30$$

$$E_{\text{Female, A}} = (50 \times 40) / 100 = 20$$

$$E_{\text{Female, B}} = (50 \times 60) / 100 = 30$$

3. Calculate the Test Statistic (χ^2)

The formula is:

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

Where O is the Observed frequency and E is the Expected Frequency.

Observed(O)	Expected (E)	O-E	(O-E) ²	(O-E) ² /E
30	20	10	100	$100 / 20 = 5.0$
20	30	-10	100	$100 / 30 \approx 3.333$
10	20	-10	100	$100 / 20 = 5.0$
40	30	10	100	$100 / 30 \approx 3.333$
Sum (\sum)				16.666

$$\chi^2_{\text{calc}} = 5.0 + 3.333 + 5.0 + 3.333 = 16.666$$

4. Determine the Critical Value

Degrees of Freedom (df): $df = (r - 1)(c - 1) = (2 - 1)(2 - 1) = 1$.

For χ^2 test with $df = 1$ and $\alpha = 0.05$, the critical value is $\chi^2_{\text{crit}} = 3.841$.

5. Make a Decision

Critical Value Method: Since $\chi^2_{\text{calc}} = 16.666$ is greater than $\chi^2_{\text{crit}} = 3.841$, the test statistic falls in the rejection region.

Decision: Reject H_0 .

6. Conclusion

At the 0.05 significance level, there is sufficient evidence to conclude that product preference is dependent on gender.
