Inference Worksheet

Step 1: State the Hypotheses

We begin by specifying the hypotheses:

- **Null Hypothesis (H_0):** The population mean is equal to 3 hours ($\mu = 3$).
  Example: Children watch TV for 3 hours per day.
- **Alternative Hypothesis (H_1):** The population mean differs from 3. Depending on research focus, H_1 can be:
  - **Two-tailed:** $\mu \neq 3$
  - **One-tailed greater:** $\mu > 3$
  - **One-tailed less:** $\mu < 3$

This is analogous to a courtroom: the null hypothesis is assumed “innocent” (true) until evidence suggests otherwise (Emmert-Streib & Dehmer, 2019).

Step 2: Set the Criteria for the Decision

We choose the **significance level (α)**, usually 5% (0.05).

- This represents the maximum probability of committing a **Type I error** (rejecting a true H_0) (Kim, 2019; Huecker & Shreffler, 2022).
- In the figure provided, α is split into **two tails** ($\alpha/2$ each side), creating rejection regions.
- In high-stakes research (e.g., medical) a stricter α (e.g., 0.01) may be chosen to reduce false positives (Emmert-Streib & Dehmer, 2019).

Definition:

- **Level of significance (α):** The threshold for deciding whether sample evidence is strong enough to reject H_0.

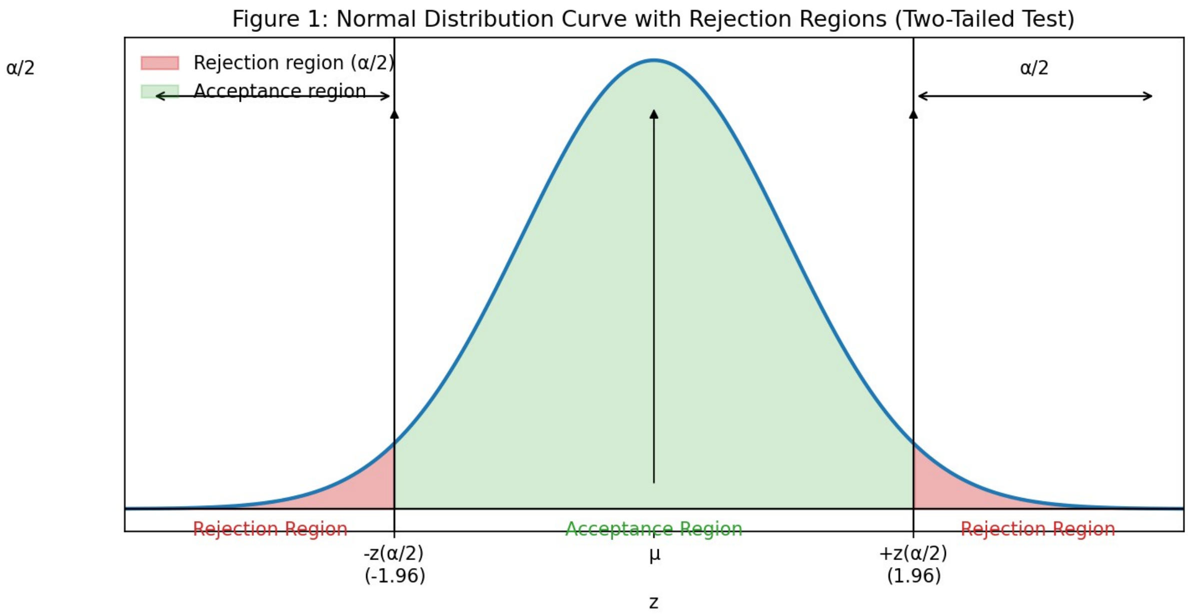

Figure 1: *The shaded areas on both tails represent the critical regions where we would reject H_0 if the test statistic falls there.*

Step 3: Compute the Test Statistic

We compare the observed sample mean to the hypothesised population mean.

Suppose we collected a sample and found the average TV viewing = **4 hours**.

We then calculate the **test statistic**:

$$t = \frac{\acute{x} - \mu}{SE}$$

Where:

- $\acute{x}$= sample mean
- μ = population mean under H_0
- SE = standard error of the mean

This test statistic tells us how far our sample mean is from the hypothesised mean in standard-error units (Emmert-Streib & Dehmer, 2019).

Concrete Example Using Diet Data

Using the data from the diet report (Exercise 9.3 - Unit 9 - Diet Weight Loss Analysis). Suppose a national health body states that a useful diet should result in an average weight loss of at least 5 kg. We want to test if **Diet A** meets this standard.

- **Step 1: State the Hypotheses**
    - H_0: The mean weight loss for Diet A is 5 kg or less ($\mu \leq 5$). (It is not effective enough).
    - H_1: The mean weight loss for Diet A is greater than 5 kg ($\mu > 5$). (It is effective). *This is a one-tailed test.*
- **Step 2: Set the Criteria for the Decision**
    - We set $\alpha = 0.05$. For a one-tailed test, the entire rejection region is in the right tail of the distribution.
- **Step 3: Compute the Test Statistic**
    - From the diet report, let's assume the sample mean for Diet A ($\bar{x}$) was calculated as **6.2 kg**, with a standard deviation (s) of **2.9 kg** and a sample size (n) of **50**.
    - **Standard Error (SE) = $s / \sqrt{n} = 2.9 / \sqrt{50} \approx 0.41$**
    - **Test Statistic (t) = $(\bar{x} - \mu) / SE = (6.2 - 5) / 0.41 \approx 2.93$**
- **Step 4: Make a Decision**
    - We compare our calculated t-statistic (2.93) to a critical t-value from a table (or software) with 49 degrees of freedom at α=0.05 (one-tailed). The critical value is approximately 1.68.
    - Since $2.93 > 1.68$, our test statistic falls in the rejection region.
    - **Conclusion:** We reject the null hypothesis. There is statistically significant evidence at the 0.05 level to conclude that the average weight loss from Diet A is greater than 5 kg.

Step 4: Make a Decision

Using the computed test statistic and sampling distribution:

- If the **p-value < α**, reject H_0 (evidence against the null).
- If the **p-value ≥ α**, fail to reject H_0 (evidence is insufficient).

Example:

- If $p = 0.03$ at $\alpha = 0.05 \rightarrow$ reject H_0.

- If $p = 0.10$ at $\alpha = 0.05$ → fail to reject H_0.

Thus, there are two decisions possible:

1. Reject the null hypothesis.
2. Accept (fail to reject) the null hypothesis.

Hypothesis Testing and Sampling Distributions

- The **sample mean** is an unbiased estimator of the population mean. (Yadav, Singh & Gupta, 2019).
- The **sampling distribution of the mean is** normally distributed (by the Central Limit Theorem) regardless of the population distribution (Kim, 2019).
- To locate the probability of observing a sample mean, we need:
    1. The hypothesised population mean.
    2. The standard error of the mean.

Because we observe only samples (not the entire population), there is always a chance of error in conclusions.

Figure 2: Sampling Distribution of the Mean

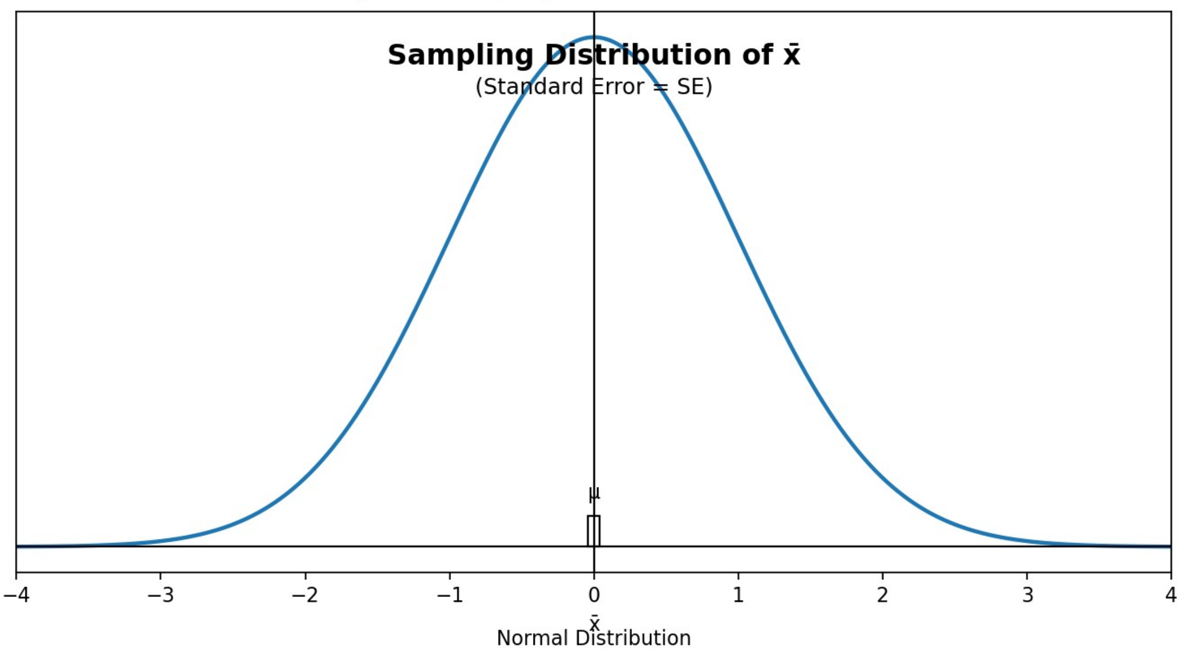

Figure 2: *This figure shows the distribution of sample means around the population mean μ.*

Errors in Hypothesis Testing

Decision	H_0 is True	H_0 is False

Accept H_0	Correct ($1-\alpha$)	Type II Error (β)
Reject H_0	Type I Error (α)	Correct (Power = $1-\beta$)

- **Type I Error (α):** Rejecting H_0 when it is actually true (false positive) (Curbing Type I & II Errors, 2011).
- **Type II Error (β):** Failing to reject H_0 when it is actually false (false negative).
- **Power ($1-\beta$):** The probability of correctly rejecting a false H_0 (Emmert-Streib & Dehmer, 2019).

Analogy:

- **Type I error:** Convicting an innocent person.
- **Type II error:** Letting a guilty person go free.

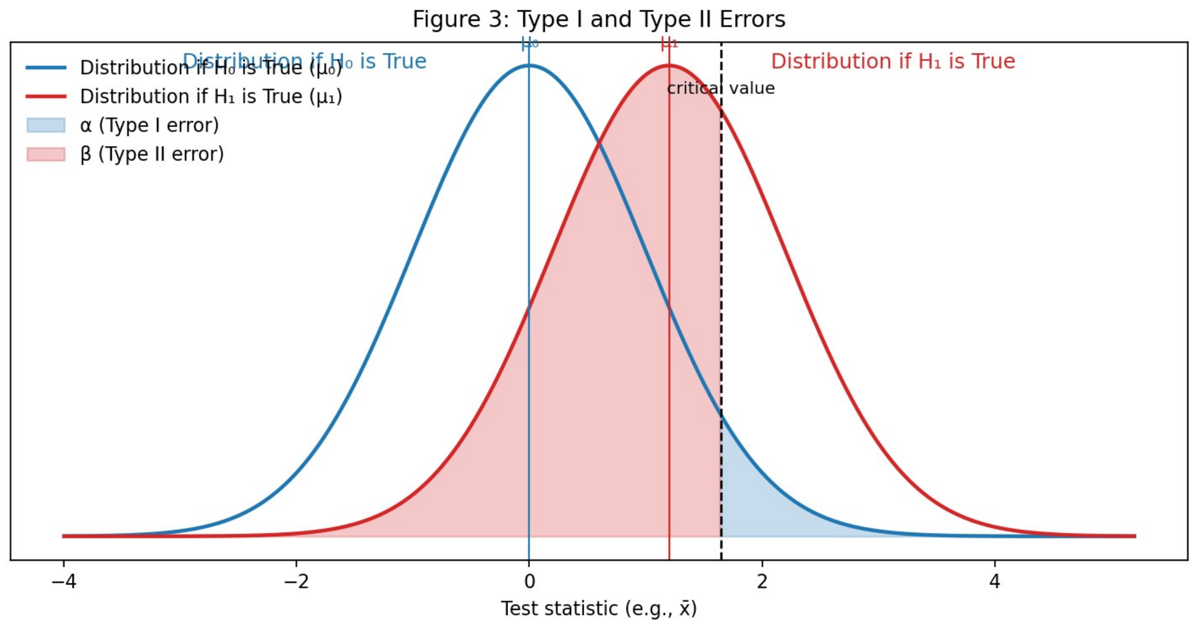

Figure 3: *This diagram illustrates the overlap between H_0 and H_1 distributions, showing where Type I and II errors occur.*

The Decision to Reject H_0

When we reject H_0, we may be correct or incorrect. The probability of a Type I error is set by α — that is why stricter α levels (e.g., 0.01) are chosen when false positives have serious consequences (Huecker & Shreffler, 2022).

Reporting Significance Levels

Rejecting H_0 at $\alpha = 0.05$ does not mean there is a 95% chance that H_0 is false. Rather, it means:

"If H_0 were true, the probability of observing the result (or one more extreme) is less than 5%." (Emmert-Streib & Dehmer, 2019)"

Importance of Setting Hypotheses in Advance

According to Emmert-Streib and Dehmer (2019), Understanding Statistical Hypothesis Testing;

- Hypotheses (H_0 and H_1) must be defined **before** inspecting the data.
- Switching to a one-tailed test after seeing the data inflates the risk of false positives (equivalent to a two-tailed test with α doubled).
- Doing so is misleading and can invalidate results.

Hypothesis Tests and "Proof"

Statistical hypothesis testing cannot prove or disprove a hypothesis with certainty. It assesses whether sample evidence is consistent with H_0 or not. A p-value = 0 is essentially impossible in practice, so absolute proof is unattainable (Understanding Hypothesis Testing, 2019).

Related vs. Unrelated Samples

- **Related samples tests (paired samples)** are more powerful since individual differences cancel out.
- **Unrelated samples tests (independent samples)** may require larger sample sizes.
- However, repeated measurements in related samples may risk participant dropout, which must be considered in study design.

Conclusion

Hypothesis testing is a structured process involving:

1. Stating H_0 and H_1.

2. Choosing a significance level (α).
3. Computing a test statistic.
4. Making a decision based on p-values.

Errors are possible, but careful choice of significance level and study design reduces risks. Hypothesis testing does not prove or disprove but instead provides evidence for or against the null hypothesis.

References

Curbing Type I and Type II Errors. (2011) *BMC Medicine.* Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2850991/ (Accessed: 12 October 2025).

Emmert-Streib, F. & Dehmer, M. (2019) ‘Understanding Statistical Hypothesis Testing: The Logic of Statistical Inference’, *AIMS Proceedings*, 1(3), pp. 54–72.

Huecker, M. & Shreffler, J. (2022) ‘Type I and Type II Errors and Statistical Power’, *StatPearls.* Available at: https://www.ncbi.nlm.nih.gov/books/NBK557530/ (Accessed: 13 October 2025).

Kim, T.K. (2019) ‘About the Basic Assumptions of t-Test: Normality and Sample Size’, *Journal of Educational Evaluation for Health Professions*, 16, p. 18.

Yadav, S.K., Singh, S. & Gupta, R. (2019) ‘Sampling Distribution and Hypothesis Testing’, in *Biomedical Statistics*. Singapore: Springer, pp. 95-98.