Hypothesis Testing

(Hypothesis Testing, Test Statistics, p-Values, t-Tests, Two-Sample Tests, Matched Pairs, and Practical Significance)

---

1. The Idea Behind Hypothesis Testing

Hypothesis testing is a core statistical method used to determine whether an observed effect is real or could reasonably be due to random chance. It is used to:

- Assess a single sample
- Compare two samples
- Detect whether an intervention or treatment has a true effect

Every hypothesis test is built on two competing statements:

- **Null hypothesis (H_0):** Assumes no special effect; represents the status quo.
- **Alternative hypothesis (H_1 or H_a):** Claims that a real effect or difference exists.

Hypothesis testing follows **indirect logic**:

1. Assume H_0 is true.
2. Determine whether the observed data are so unlikely under H_0 that H_0 should be rejected.

The central question is:

> Are the observed data consistent with H_0, or are they so unusual that H_0 should be rejected?

---

2. Test Statistics

A **test statistic** measures how far the observed data deviates from what is expected under the null hypothesis.

A commonly used form is the **Z-statistic**:

Z = (Observed – Expected) / (SE)

Components:

- **Observed:** Value calculated from the sample
- **Expected:** Value predicted under H_0
- **Standard Error (SE):** Natural variability of the statistic under H_0

The Z-value quantifies how unusual the observed result is if H_0 were true. The choice of test statistic must match the structure of the hypothesis (counts, proportions, means, etc.).

---

3. p-Values as Measures of Evidence

The **p-value** converts the test statistic into a probability-based measure of evidence.

Definition:
The p-value is the probability of obtaining a result as extreme as or more extreme than the observed result, **assuming that H_0 is true**.

Key properties:

- Under H_0, the Z-statistic follows the **standard normal distribution**.
- The p-value is the **tail area** beyond the observed $|Z|$.
- **Larger $|Z|$ → smaller p-value → stronger evidence against H_0.**

Common decision rule:

- If **$p \le 0.05$**, reject H_0 (statistically significant)
- If **$p > 0.05$**, do not reject H_0

Important warning:
The p-value is **not** the probability that H_0 is true. H_0 is either true or false; the p-value only measures how unlikely the data are **if H_0 were true**.

---

4. One-Sided vs Two-Sided Tests

- **Two-sided test:** Tests for any difference (H_1: parameter ≠ value).
- **One-sided test:** Tests only for improvement in one direction (H_1: parameter > value or < value).

In one-sided tests, the p-value is taken from **one tail** of the distribution.
In two-sided tests, the p-value is **double** the one-sided tail area.

A critical rule:

The test type (one-sided or two-sided) **must be chosen before seeing the data**.
It is invalid to change the test type after checking significance.

---

5. The t-Test for Small Sample Sizes

When:

- The sample size is **small** (typically $n < 30$), and
- The population standard deviation **σ is unknown**,

the **z-test is no longer valid**. Instead, the **Student's t-test** is used.

The test statistic becomes:

$t = \bar{x} - \mu 0 / s \sqrt{n}$

where:

- (s) = sample standard deviation
- Degrees of freedom = (n - 1)

The **t-distribution**:

- Has a lower center and **fatter tails** than the normal distribution
- Reflects additional uncertainty from estimating σ
- Approaches the normal distribution as sample size increases

For small samples:

- Use **t-tests instead of z-tests**
- Use **t-based confidence intervals instead of z-intervals**

---

6. Statistical Significance vs Practical Importance

Statistical significance does NOT imply practical importance.

With very large sample sizes:

- The standard error becomes extremely small (square-root law),
- Tiny differences can produce very large test statistics,
- Results may be **highly significant yet practically meaningless**.

To assess **practical importance**, one must examine the **confidence interval**:

- A narrow CI centered near the reference value implies a small real-world effect.

Connection between CIs and hypothesis tests:

- A value is rejected by a two-sided 5% test **if and only if** it lies **outside the 95% confidence interval**.

Types of errors:

- **Type I error (False Positive):** Rejecting a true H_0
- **Type II error (False Negative):** Failing to reject a false H_0

At $\alpha = 0.05$, the probability of a Type I error is at most 5%.

---

7. Two-Sample z-Tests

The **two-sample z-test** is used to compare:

- Two population proportions, or
- Two population means, when **samples are independent** and σ is known or sample sizes are large.

Hypotheses are written as:

$H0: p1 - p2 = 0,$
$H1: p1 - p2 \neq 0$

Test statistic:

Z = Observed difference – 0 / SE of the difference

Standard error of the difference:

$SE = \sqrt{}(SE_1{}^2 + SE_2{}^{2)}$

A 95% confidence interval for the difference:

(Difference) $\pm z \times SE$

If 0 lies inside the interval, the difference is not statistically significant.

Pooling proportions under H is mathematically valid but usually yields very

similar results to unpooled SEs.

Key assumption:

All two-sample z-tests require independent samples.