Introduction to Probability and Statistics

Master in Cognitive Science 2025–2026

Lecture 6

November 2025

Objectives

- Introduction to hypothesis testing.
- Definition of Type I and Type II errors, the p-value, and the power of a test.
- Examples: testing the equality of two means (`t.test`), testing the equality of variances, testing the normality of a population, Wilcoxon test, ...

Readings: Kass et al. (2014), Chapters 10.3.1–10.3.5.
Tests statistiques d'hypothèses, Doc_fr5 file in Ametice.

Testing principle

- In a court, a defendant D is either innocent (H_0) or guilty (H_1).
- The defendant is assumed to be innocent (H_0).
- The judge decides, based on the evidence, either that D is guilty (rejecting H_0) or that the evidence is insufficient to convict D (H_0 is not rejected due to lack of evidence).

Risks of a test

		True State	
		H_0 **is true**	H_1 **is true**
Decision	H_0	Correct conclusion	Type II error
	H_1	Type I error	Correct conclusion (Power)

1. Confidence level α: probability of committing a Type I error.
   - $\alpha = \mathbb{P}(\text{"Reject } H_0\text{"} | H_0)$.
   - In practice, we often take $\alpha = 0.05$.
   - Depending on how much risk we want to avoid, smaller values of α can be used (e.g., 0.01).
2. Power of a test: probability of correctly rejecting the null hypothesis H_0.
   - Probability of committing type II error $\beta = \mathbb{P}(\text{"Do not reject } H_0\text{"} | H_1)$.
   - The power of the test is $1 - \beta$.

Decision rule

Definition

A decision variable is a random variable T that *behaves differently* depending on whether H_0 or H_1 is true; that is, the distribution of T will differ according to whether H_0 is true or H_1 is true. Moreover, the distribution of T must be completely known at least under H_0.

Example:

- Assume we want to check whether the mean of a normal population is equal to 0. Suppose we have an i.i.d. sample $X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$.
- We want to test $H_0 : \mu = 0 \quad \text{vs.} \quad H_1 : \mu \neq 0$.
- Consider the statistic $T_n = \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.
- Under the null hypothesis, $T_n \sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right)$.

p-value

- Given observations $x_1, x_2, \ldots, x_n$, we compute the statistic $T_n(x_1, \ldots, x_n) = t_n$.
- The p-value is

$$p_v = \mathbb{P}\left(|T_n| \geq |t_n| \,\middle|\, H_0\right).$$

- If the p-value is smaller than the confidence level α, this is evidence against H_0.

Effect Size

Remark

Null Hypothesis Significance Testing (NHST) is widely debated, especially regarding how the p-value is interpreted. A small p-value does not imply that H_1 is more likely; it only indicates evidence against H_0. Likewise, failing to reject H_0 does not validate it; with a sufficiently large sample, even negligible differences may lead to rejection.
This motivates reporting an effect size alongside the p-value. In the two-sample case, a common measure is *Cohen's* d:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}},$$

where $\hat{\sigma}$ is the estimated standard deviation.

Confidence interval

Suppose we want to estimate the mean μ_0 of some population.

Definition

A confidence interval for a parameter μ_0 is an interval-valued statistic $(L(X_1, \ldots, X_n), U(X_1, \ldots, X_n))$ such that

$$\mathbb{P}\big(L \leq \mu_0 \leq U\big) = 1 - \alpha,$$

where $1 - \alpha$ is the confidence level. It provides a range of plausible values for μ_0 based on the data.

Examples of tests

See R follow-along 4.

- Testing equality of means: `t.test`.
- Wilcoxon test: `wilcox.test`.
- Testing equality of variances: `var.test`.
- Normality test: `shapiro.test`.
- Permutation test.
- $\dots$