Chapter 4

Hypothesis Testing

4.1 General Concepts

From the previous chapters on parameter estimation (both point and interval estimation), it is clear that a statistic—viewed as a random variable—serves as an estimator of a population parameter whose true theoretical value may differ from that suggested under the hypothesis.

In most decision-making situations based on a sample of observations, we must choose between two contradictory statements : the **null hypothesis** H_0 and the **alternative hypothesis** H_1. These hypotheses are formulated in terms of the theoretical parameters of the population.

To determine whether the proposed null hypothesis H_0 is supported by the data, we require a systematic method for evaluating this assumption. This procedure is known as a **hypothesis test**.

4.1.1 Statistical Hypothesis

Definition 4.1.1. *In mathematical statistics, a* statistical hypothesis *is an assumption concerning a population of interest. This information can take several forms :*

1. *A probability distribution (for example, the distribution of weights, temperatures, or the ages of players) ;*
2. *A specific numerical value, such as a proportion, a quantile, a mean, or a variance ;*
3. *An interval of possible values containing a particular statistic.*

4.1.2 Testing Hypotheses

Before performing a hypothesis test, the following statements are established :

1. **Null hypothesis** (H_0) : this represents the prior or reference value of a population parameter that we wish to test.
2. **Alternative hypothesis** (H_1) : this denotes any assumption that differs from the null hypothesis.

Remark 12. *The alternative hypothesis H_1 is the statement that we aim to support by rejecting the null hypothesis H_0.*

4.1.3 Hypothesis Test

Definition 4.1.2. *Constructing a hypothesis test involves determining the range of possible values that a random variable can take under the assumption that the null hypothesis is true, considering only random sampling variability.*

A hypothesis test can also serve to evaluate whether a particular action has successfully changed (or failed to change) the value of a population statistic.

The main steps of a hypothesis test can be summarized as follows :

1. Choose a test statistic (a function of the observations) such that we reject H_0 when its observed value provides sufficient evidence in favor of the alternative hypothesis H_1.
2. Determine the **rejection region** for H_0—the set of results so far from what is expected under H_0 that they would have an extremely small probability $\mathbb{P}$ of occurring if H_0 were true.
3. Compare the sampling distribution of the test statistic under H_0 with the value computed from the actual sample data.
4. Draw a conclusion :
   (a) If the test statistic falls within the rejection region, then H_0 is unlikely to hold, and we **reject** H_0.
   (b) If the test statistic falls within the acceptance (credible) region, then H_0 appears plausible, and we **do not reject** H_0.

Remark 13. *Rejecting H_0 means that the null hypothesis is considered false. However, failing to reject H_0 does* not *imply that it is true—only that there is insufficient evidence against it. This decision is made while accepting a certain probability of error, known as the* ***significance level*** *of the test.*

Type I Error Risk

Definition 4.1.3. *The* ***risk of a Type I error****, denoted by α, is called the significance level of the test. It is defined as*

$$\alpha = \mathbb{P}\left(\textit{reject } H_0 \mid H_0 \textit{ is true}\right).$$

Thus, the probability associated with the rejection region of the null hypothesis represents the risk of committing a Type I error.

Example 4.1.4. *If $\alpha = 0.05$, it means that we accept in advance that, in about 5% of random samples, the test statistic may fall in the rejection region of H_0, even though H_0 is actually true.*

The complementary region, known as the acceptance region of H_0, therefore has a probability equal to $1 - \alpha$.

Remark 14. *The statistic used in a hypothesis test is a random variable whose observed value determines whether we reject or fail to reject H_0. The sampling distribution of this statistic is always derived under the assumption that H_0 is true.*

4.1.4 Formulating a Hypothesis Test

The formulation of a hypothesis test involves the following steps :

1. Assume that a population parameter θ (such as a proportion, a moment, or a quantile) takes a specific theoretical value θ_0.
2. Distinguish between two types of tests for θ :

— **One-tailed test :**

$$\begin{cases} H_0 : & \theta = \theta_0, \\ H_1 : & \theta < \theta_0 \end{cases} \quad \text{or} \quad \begin{matrix} H_0 : & \theta = \theta_0, \\ H_1 : & \theta > \theta_0. \end{matrix}$$

— **Two-tailed test :**

$$\begin{cases} H_0 : & \theta = \theta_0, \\ H_1 : & \theta \neq \theta_0. \end{cases}$$

3. Represent schematically the rejection and non-rejection regions of H_0 :

H_0 rejected		H_0 accepted		H_0 rejected	
$\theta < \theta_0 \quad \theta_1$		θ_0	θ_2	$\theta > \theta_0$	
$\alpha/2$		$1-\alpha$		$\alpha/2$	

Example 4.1.5. *We are interested in the proportion p of male students in a sample of 120 candidates preparing for a Ph.D. entrance examination in mathematics.*

$$\begin{cases} H_0: & \textit{the proportion of male students is } 20\% \Rightarrow p = 0.2, \\ \textit{against} & \\ H_1: & p \neq 0.2. \end{cases}$$

Let X be the random variable representing the number of male students in the sample of 120 candidates :

$$X \rightsquigarrow \mathcal{B}(n,p), \quad \textit{where } n = 120 \textit{ and } p = 0.2.$$

Since $n = 120 \gg 30$, the binomial distribution can be approximated by a normal distribution with parameters

$$m = \mathbb{E}(X) = np = 120 \times 0.2 = 24,$$
$$\sigma = \sqrt{np(1-p)} = \sqrt{120 \times 0.2 \times 0.8} = 4.38.$$

Under H_0, these parameters are estimated as $m = 24$ and $\sigma = 4.38$. Considering a 95% confidence interval for the mean, we have

$$IC_m = \left[\bar{X} - z_{1-\alpha/2}\hat{\sigma},\ \bar{X} + z_{1-\alpha/2}\hat{\sigma}\right],$$

where $\alpha = 5\%$.

The null hypothesis H_0 is accepted if $m \in IC_m$ and rejected otherwise.

The corresponding confidence interval for the proportion p is

$$IC_p = \left[p - z_{1-\alpha/2}\hat{\sigma},\ p + z_{1-\alpha/2}\hat{\sigma}\right].$$

Rejecting H_0 means that it is considered false and that the alternative H_1 is accepted, with a risk of error $\alpha = 5\%$.

Type II Error Risk

Definition 4.1.6. *The* ***Type II error risk****, denoted by β, is the probability that the computed or observed value of the test statistic does not fall within the rejection region when H_1 is actually true. In this situation, H_0 is accepted even though it is false. Thus, β represents the risk of accepting H_0 when it should be rejected :*

$$\begin{aligned}\beta &= \mathbb{P}\left(\text{accept } H_0 \mid H_0 \text{ is false}\right)\\ &= \mathbb{P}\left(\text{accept } H_0 \mid H_1 \text{ is true}\right)\\ &= \mathbb{P}\left(\text{reject } H_1 \mid H_1 \text{ is true}\right).\end{aligned}$$

Remark 15. *To determine the value of β, one must know the probability distribution of the test statistic S under the alternative hypothesis H_1.*

Example 4.1.7. *We want to test whether a coin is fair. For a biased coin, the probability of getting heads is $p = 0.6$. We apply the following decision rule :*

$$H_0 : \text{ the coin is fair} \quad \begin{cases} \text{accepted,} & \text{if } x \in [40, 60],\\ \text{rejected,} & \text{otherwise,}\end{cases}$$

where x represents the number of heads obtained in 100 tosses. The goal is to determine the risk of committing a Type II error, denoted by β.

Test Power $(1-\beta)$

Definition 4.1.8. *The* ***power of a test*** *measures its ability to reject the null hypothesis H_0 when it is false. It is given by*

$$\begin{aligned}1-\beta &= \mathbb{P}\left(\text{reject } H_0 \mid H_0 \text{ is false}\right)\\ &= \mathbb{P}\left(\text{accept } H_1 \mid H_1 \text{ is true}\right).\end{aligned}$$

Remark 16. *The power of a statistical test increases with the sample size n (for a fixed value of α), and decreases when the significance level α becomes smaller.*

Example 4.1.9. *Using the same coin-toss example as above, compute the power of the test when the probability of heads is 0.2, then when it is 0.5. What can you conclude from these results ?*

Decision Rules

The following table summarizes the two types of errors and the corresponding decisions.

	H_0 **true**	H_0 **false**
Decision	H_0 rejected	
	Incorrect decision : Type I error. α : probability of rejecting H_0 when it is true.	Correct decision : Test power. $1-\beta$: probability of rejecting H_0 when it is false.
	H_0 accepted	
	Correct decision : $1-\alpha$: probability of not rejecting H_0 when it is true.	Incorrect decision : Type II error. β : probability of not rejecting H_0 when it is false.

TABLE 4.1 – Decision outcomes in hypothesis testing

Choosing a Statistical Test

The choice of a suitable statistical test depends on :

1. The nature of the data collected.
2. The type of hypothesis to be evaluated.
3. The characteristics of the populations involved (equality or inequality of variances, for instance).

One-Sample Conformity Tests (or One-Sample Parameter Tests)

One-sample conformity tests are procedures in **inferential statistics** used to determine whether a given sample can reasonably be considered as drawn from a specific population. Put differently, they assess whether the population's **parameter** is **in conformity** with a pre-established **theoretical or standard value**.

Primary Objective

The main goal of this type of test is to compare a specific parameter calculated from the sample (the empirical value) against a known reference value (the theoretical value). They allow us to judge whether the observed difference is large enough to reject the hypothesis of conformity.

Tested Parameters

These tests apply to various quantitative parameters :
- **Mean** (μ) : Compared to a theoretical value μ_0.
- **Variance** (σ^2) : Compared to a theoretical value σ_0^2.
- **Proportion** (p) : Compared to a theoretical value p_0.

Hypothesis Formulation (Example for the Mean)

The test is structured around the following hypotheses :
- **Null Hypothesis (**H_0 **- Conformity)** : The population parameter is equal to the theoretical value.

$$H_0 : \mu = \mu_0$$

- **Alternative Hypothesis (**H_a **- Non-Conformity)** : The population parameter is different from the theoretical value.

$$H_a : \mu \neq \mu_0$$

Examples of Specific Tests

The choice of test depends on the parameter being studied and the available information :
- **For the Mean** : The **One-Sample** t**-test** (if the population standard deviation is unknown).
- **For the Variance** : The **One-Sample** χ^2 **(Chi-squared) test** for the variance.

4.2 Comparison Between an Observed Mean and a Theoretical Mean

We aim to compare the sample mean (calculated from observed data) with a theoretical population mean in two distinct situations : i) when the population variance is known ; ii) when the population variance is unknown and must be replaced by its estimator.

4.2.1 Principle of the Test

Let X be a random variable following a normal distribution, and let a random sample be drawn from this population.

Known population		Unknown population
$X \rightsquigarrow \mathcal{N}(m_0, \sigma_0)$		$X \rightsquigarrow \mathcal{N}(m, \sigma)$
$H_0 : m = m_0$		$H_1 : m \neq m_0$
	Sample	
	$n, \bar{X}, S_n^2$	

Here, $\bar{X}$ is an estimator of m. 1. If $\bar{X}$ belongs to a known reference population with mean m_0, then H_0 is true. 2. If $\bar{X}$ belongs to a population with unknown mean m, then H_1 is true.

4.2.2 Population Variance Known

Test Statistic

The sampling distribution of the sample mean $\bar{X}$ in a normal population with variance σ^2 is given by :

$$\bar{X} \rightsquigarrow \mathcal{N}\left(m, \sqrt{\frac{\sigma^2}{n}}\right).$$

We define the statistic

$$\mathcal{S} = \bar{X} - m_0 \rightsquigarrow \mathcal{N}\left(0, \sqrt{\frac{\sigma^2}{n}}\right).$$

Under the null hypothesis H_0, we have

$$\mathbb{E}(\mathcal{S}) = 0, \qquad \operatorname{Var}(\mathcal{S}) = \frac{\sigma^2}{n},$$

where σ^2 is assumed to be known.
By standardizing, we obtain the standard normal variable :

$$Z = \frac{\mathcal{S} - \mathbb{E}(\mathcal{S})}{\sqrt{\mathrm{Var}(\mathcal{S})}} = \frac{\bar{X} - m_0}{\sqrt{\frac{\sigma^2}{n}}} \rightsquigarrow \mathcal{N}(0, 1).$$

The hypotheses of interest are :

$$H_0 : m = m_0 \quad \text{versus} \quad H_1 : m \neq m_0,$$

with known σ. The computed value of the test statistic is :

$$z = \frac{|\bar{X} - m_0|}{\sqrt{\frac{\sigma^2}{n}}}.$$

This observed value, denoted ε_{obs}, is compared to the critical value $\varepsilon_{\text{crit}}$ obtained from the standard normal table for a given significance level α (decision rule 1).

Decision Rule

1. If $\varepsilon_{\text{obs}} > \varepsilon_{\text{crit}}$, reject H_0 at the significance level α. 2. If $\varepsilon_{\text{obs}} \leq \varepsilon_{\text{crit}}$, do not reject H_0.

Example 4.2.1. *Let X be the random variable representing the blood glucose level (in g/l). Assume* $X \sim \mathcal{N}(m_0 = 1, \sigma = 0.1)$. *From a sample of* $n = 9$ *patients, the observed mean is* $\bar{X} = 1.12\,\text{g/l}$. *Is this sample representative of the population ?*

$$H_0 : m = m_0 = 1\,\text{g/l} \quad \textit{versus} \quad H_1 : m \neq m_0.$$

Given $\bar{X} \rightsquigarrow \mathcal{N}\left(m, \sqrt{\frac{\sigma^2}{n}}\right)$, *we have :*

$$\mathcal{S} = \bar{X} - m_0 \rightsquigarrow \mathcal{N}\left(0, \sqrt{\frac{\sigma^2}{n}}\right).$$

Under H_0,

$$\mathbb{E}(\mathcal{S}) = 0, \qquad \mathrm{Var}(\mathcal{S}) = \frac{\sigma^2}{n}.$$

Hence, the standardized statistic is :

$$Z = \frac{\mathcal{S} - \mathbb{E}(\mathcal{S})}{\sqrt{\mathrm{Var}(\mathcal{S})}} = \frac{\bar{X} - m_0}{\sqrt{\frac{\sigma^2}{n}}} \rightsquigarrow \mathcal{N}(0,1).$$

For a significance level $\alpha = 0.05$, we compute :

$$\begin{aligned} z &= \frac{1.12 - 1}{\sqrt{\frac{0.01}{9}}} \\ &= 3.6 > \varepsilon_{crit}, \end{aligned}$$

*where $\varepsilon_{crit} = 1.96$. Therefore, H_0 is rejected — the sample is **not representative** of the population with mean $m_0 = 1$ g/l.*

Comparison Between an Observed variance and a Theoretical variance

Let $X_1, X_2, \ldots, X_n$ be a sample of n observations from a normal distribution with variance σ^2 equal to a specified value σ_0^2 (or equivalently, standard deviation σ_0). To test the hypothesis

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{against} \quad H_1 : \sigma^2 \neq \sigma_0^2,$$

we use the test statistic :

$$Z = \frac{(n-1)\tilde{S}^2}{\sigma_0^2}. \tag{4.2.1}$$

If the null hypothesis H_0 is true, then the statistic Z defined in (4.2.1) follows a chi-square distribution χ^2_{n-1} with $(n-1)$ degrees of freedom.

Decision Rule

For a given significance level α, we compute :

$$\begin{aligned} \varepsilon_{\text{obs}} &= \frac{(n-1)\tilde{S}^2}{\sigma_0^2} \\ &\sim \chi^2_{n-1}. \end{aligned}$$

The null hypothesis H_0 is rejected if $\varepsilon_{\text{obs}} > \chi^2_{n-1,\alpha/2}$ or $\varepsilon_{\text{obs}} < \chi^2_{n-1,1-\alpha/2}$. This rule is illustrated in the figure below :

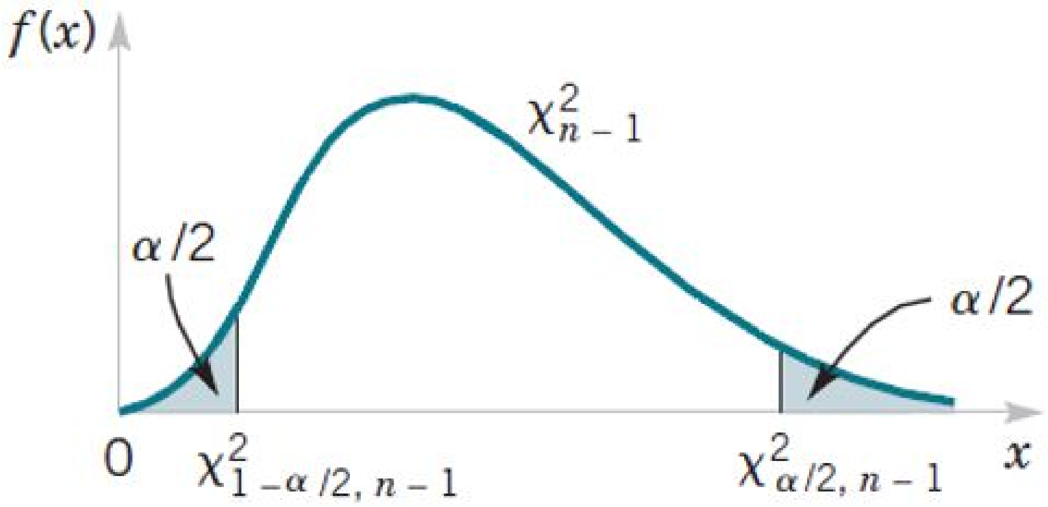


Figure 4.1 – Two-tailed Chi-Square Test

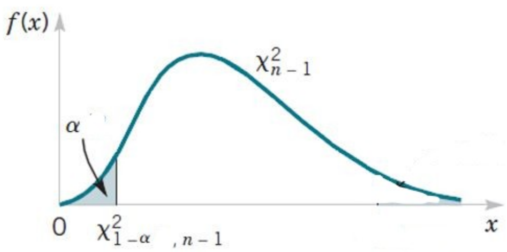
The same approach applies for one-tailed hypotheses.
For example, for the alternative hypothesis

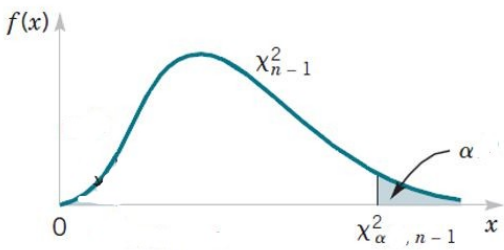
$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{against} \quad H_1 : \sigma^2 > \sigma_0^2,$$

we reject H_0 if $\varepsilon_{\text{obs}} > \chi^2_{n-1,\alpha}$.
Conversely, for

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{against} \quad H_1 : \sigma^2 < \sigma_0^2,$$

we reject H_0 if $\varepsilon_{\text{obs}} < \chi^2_{n-1,1-\alpha}$. The corresponding rejection regions are shown in the following figures :





4.3 Unknown Population Variance

When the population variance is unknown, it is estimated by :

$$\tilde{S}^2 = \frac{n}{n-1} S^2,$$

where S^2 denotes the empirical variance of the sample.

Test Statistic

By standardizing, we obtain :

$$\begin{aligned} Z &= \frac{\mathcal{S} - \mathbb{E}(\mathcal{S})}{\sqrt{\mathrm{Var}(\mathcal{S})}} \\ &= \frac{\bar{X} - m_0}{\sqrt{\frac{\tilde{S}^2}{n}}} \rightsquigarrow St_{n-1}, \end{aligned}$$

where St_{n-1} denotes the Student's t-distribution with $(n-1)$ degrees of freedom.

The tested hypotheses are :

$$H_0 : m = m_0 \quad \text{against} \quad H_1 : m \neq m_0.$$

The calculated value of the statistic is :

$$t_{\text{obs}} = \frac{|\bar{X} - m_0|}{\sqrt{\frac{\tilde{S}^2}{n}}} = \frac{|\bar{X} - m_0|}{\sqrt{\frac{S^2}{n-1}}}.$$

The observed value t_{obs} is compared to the critical value t_{crit} from the Student's t-table for a fixed significance level α and $(n-1)$ degrees of freedom.

Decision Rule

1. If $t_{\text{obs}} > t_{\text{crit}}$, reject H_0 at the significance level α.

2. If $t_{\text{obs}} \leq t_{\text{crit}}$, do not reject H_0.

Example 4.3.1. *To assess a production batch of tablets, a random sample of 10 tablets is selected from a lot of 30,000 units, and their masses (in grams) are measured as follows :*

0.80	*0.81*	*0.84*	*0.83*	*0.85*	*0.86*	*0.85*	*0.83*	*0.84*	*0.80*

Is the observed mean weight consistent with the production mean of 0.83 g at the 98% confidence level ?

Solution 1

We test the hypothesis

$$H_0 : \mu = 0.83 \qquad \text{vs} \qquad H_1 : \mu \neq 0.83$$

at the significance level $\alpha = 0.02$ (i.e. 98% confidence).

Step 1 — Sample mean.

$$\bar{X} = 0.831.$$

Step 2 — Sample standard deviation (unbiased).

$$\tilde{S}_n = 0.02132.$$

Step 3 — Compute the test statistic.

$$t_{\text{obs}} = \frac{|\bar{X} - \mu_0|}{\tilde{S}_n / \sqrt{n}} = \frac{|0.831 - 0.83|}{0.02132/\sqrt{10}} \approx \frac{0.001}{0.006737} \approx 0.148.$$

Step 4 — Critical value.
Degrees of freedom : $df = n - 1 = 9$.
Two-sided test at $\alpha = 0.02$:

$$t_{\text{crit}} = t_{9,1-\alpha/2} = t_{9,0.99} \approx 2.821.$$

Step 5 — Decision rule.

$$\text{Reject } H_0 \quad \text{if} \quad t_{\text{obs}} > t_{\text{crit}}.$$

Since

$$t_{\text{obs}} = 0.148 \quad \text{and} \quad t_{\text{crit}} = 2.821,$$

we have

$$t_{\text{obs}} \leq t_{\text{crit}}.$$

Conclusion. We do *not* reject H_0. The sample mean is consistent with the production mean value 0.83 at the 98% confidence level.

Remark 17. *When $n < 30$, the random variable X must follow a normal distribution $\mathcal{N}(0, \sigma)$. For $n \geq 30$, the Student's t-distribution approximates the standard normal distribution.*

4.4 Comparison Between an Observed and a Theoretical Proportion

4.4.1 Principle of the Test

Consider the random variable X following a Bernoulli distribution :

X	0	1
$\mathbb{P}_X$	$1-p$	p

This variable is observed within a population. A random sample of size n drawn from this population follows a binomial distribution $\mathcal{B}(n,p)$.

Known Population		Unknown Population
$X \rightsquigarrow \mathcal{B}(n,p)$		$X \rightsquigarrow \mathcal{B}(n,p_0)$
$H_0 : p = p_0$		$H_1 : p \neq p_0$
	Random sample $\underset{n,f_i}{\overset{f_i=\frac{n_i}{n}}{}}$	

If the observed frequency $f_i = \frac{n_i}{n}$, which estimates the proportion p, corresponds to a reference population with known probability p_0, then the null hypothesis H_0 holds. Otherwise, if f_i significantly deviates from p_0 (that is, $f_i \neq p_0$), we reject H_0 in favor of the alternative H_1.

4.4.2 Sampling Distribution

Assume that the population variances of both the reference and the observed populations are equal. Then the sample frequency f_i approximately follows a normal distribution :

$$f_i \rightsquigarrow \mathcal{N}\left(p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right).$$

The test statistic is defined as :

$$\mathcal{S} = f_i - p_0,$$

which follows the normal distribution :

$$\mathcal{S} \rightsquigarrow \mathcal{N}\left(0, \sqrt{\frac{p_0(1-p_0)}{n}}\right),$$

with

$$\mathbb{E}(\mathcal{S}) = 0 \quad \text{and} \quad \mathrm{Var}(\mathcal{S}) = \sqrt{\frac{p_0(1-p_0)}{n}}.$$

By standardizing, we obtain the following statistic :

$$\begin{aligned} Z &= \frac{\mathcal{S} - \mathbb{E}(\mathcal{S})}{\sqrt{\frac{p_0(1-p_0)}{n}}} \\ &= \frac{f_i - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \\ &\rightsquigarrow \mathcal{N}(0,1). \end{aligned}$$

Decision Rule

The computed statistic is given by :

$$\varepsilon_{\text{obs}} = \frac{f_i - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}.$$

This observed value is compared with the critical threshold $\varepsilon_{\text{seuil}}$ obtained from the standard normal table for a predetermined significance level α.

Decision :

1. If $\varepsilon_{\text{obs}} > \varepsilon_{\text{seuil}}$, the sample lies outside the reference population. Therefore, the sample is not representative, and the null hypothesis H_0 is rejected.
2. If $\varepsilon_{\text{obs}} \leq \varepsilon_{\text{seuil}}$, the sample can be considered representative of the reference population, and the null hypothesis H_0 is accepted.

4.5 Comparison of Two Empirical Means

4.5.1 Principle of the Test

We consider two independent samples extracted from two populations :

$$X_1, \ldots, X_{n_1} \rightsquigarrow \mathcal{N}(m_1, \sigma_1^2),$$
$$Y_1, \ldots, Y_{n_2} \rightsquigarrow \mathcal{N}(m_2, \sigma_2^2).$$

Let $\bar{X}$ and $\bar{Y}$ denote the sample means, and let S_1^2 and S_2^2 denote the empirical variances.

We want to test whether the two population means are equal :

$$H_0 : m_1 = m_2 \qquad \text{against} \qquad H_1 : m_1 \neq m_2.$$

Different test statistics are used depending on whether the population variances are known or unknown.

4.5.2 Case 1 : Variances Known

When σ_1^2 and σ_2^2 are known, the statistic

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

follows the standard normal distribution under H_0.

Decision Rule

Let $\varepsilon_{\text{obs}} = |Z|$. Reject H_0 if

$$\varepsilon_{\text{obs}} > \varepsilon_{\text{seuil}} \quad \text{with } \varepsilon_{\text{seuil}} = z_{\alpha/2}.$$

Otherwise, do not reject H_0.

Example 4.5.1. *We have two samples of tube diameters produced by two manufacturing processes, A and B. The diameters (in millimeters) are as follows :*

Process A	51.90	50.90	52.80	52.90	53.40
Process B	51.10	51.30	51.50	52.10	

We assume that tube diameters follow a normal distribution and that the theoretical standard deviations are known :

$$\sigma_A = 1\ mm, \qquad \sigma_B = 0.45\ mm.$$

We want to test, at the 5% *significance level, whether there is a significant difference between the mean diameters produced by processes A and B.*

Let X_A and Y_B be the random variables representing the diameters of tubes from processes A and B, respectively. We test the hypotheses :

$$H_0 : \mu_A - \mu_B = 0 \qquad against \qquad H_1 : \mu_A - \mu_B \neq 0.$$

The test statistic for two independent samples with known variances is

$$Z = \frac{\bar{X}_{n_1} - \bar{Y}_{n_2}}{\sqrt{\dfrac{\sigma_A^2}{n_1} + \dfrac{\sigma_B^2}{n_2}}} \quad \rightsquigarrow \quad \mathcal{N}(0,1) \quad under\ H_0.$$

From the samples we obtain :

$$\bar{X}_{n_1} = 52.38, \qquad \bar{Y}_{n_2} = 51.50.$$

Thus

$$Z_{obs} = \frac{52.38 - 51.50}{\sqrt{\dfrac{1^2}{5} + \dfrac{0.45^2}{4}}} = 1.76.$$

For a two-sided test *at significance level $\alpha = 5\%$, the critical value is :*

$$z_{0.025} = 1.96.$$

Decision : *Since*

$$|Z_{obs}| = 1.76 < 1.96,$$

*we **fail to reject** the null hypothesis H_0.*

Conclusion. *At the* 5% *significance level, the observed data do not provide enough evidence to conclude that the mean diameters produced by processes A and B are different.*

4.5.3 Case 2 : Variances Unknown but Equal

If the two populations share the same variance σ^2, we use the pooled estimator

$$\tilde{S}_p^2 = \frac{(n_1 - 1)\tilde{S}_1^2 + (n_2 - 1)\tilde{S}_2^2}{n_1 + n_2 - 2}.$$

The statistic

$$T = \frac{\bar{X} - \bar{Y}}{\tilde{S}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

follows a Student distribution with $n_1 + n_2 - 2$ degrees of freedom.

Decision Rule

Reject H_0 if

$$|T| > t_{n_1+n_2-2,\,1-\alpha/2}.$$

4.5.4 Case 3 : Variances Unknown and Unequal (Welch's Test)

When variances differ, the statistic becomes

$$T_w = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\tilde{S}_1^2}{n_1} + \frac{\tilde{S}_2^2}{n_2}}}$$

and follows approximately a Student distribution with Welch–Satterthwaite degrees of freedom

$$\nu \approx \frac{\left(\frac{\tilde{S}_1^2}{n_1} + \frac{\tilde{S}_2^2}{n_2}\right)^2}{\frac{\tilde{S}_1^4}{n_1^2(n_1-1)} + \frac{\tilde{S}_2^4}{n_2^2(n_2-1)}}.$$

Remark 18. *Welch's test is recommended when $\tilde{S}_1^2$ and $\tilde{S}_2^2$ differ significantly, as it does not assume equality of variances.*

4.6 Comparison of Two Empirical Variances

4.6.1 Principle of the Test

Let two independent samples come from two normal populations :

$$X_1, \ldots, X_{n_1} \rightsquigarrow \mathcal{N}(m_1, \sigma_1^2), \qquad Y_1, \ldots, Y_{n_2} \rightsquigarrow \mathcal{N}(m_2, \sigma_2^2).$$

We want to check whether the two variances are equal :

$$H_0 : \sigma_1^2 = \sigma_2^2 \qquad \text{vs} \qquad H_1 : \sigma_1^2 \neq \sigma_2^2.$$

4.6.2 Test Statistic

The statistic used is the ratio of sample variances :

$$F = \frac{\tilde{S}_1^2}{\tilde{S}_2^2},$$

where by convention $\tilde{S}_1^2$ is the larger variance (to ensure $F \geq 1$). Under H_0,

$$F \sim F_{(n_1-1,\ n_2-1)}.$$

4.6.3 Decision Rule (Bilateral Test)

Reject H_0 if

$$F > F_{1-\alpha/2}(n_1 - 1, n_2 - 1) \quad \text{or} \quad F < F_{\alpha/2}(n_1 - 1, n_2 - 1).$$

Otherwise, do not reject H_0.

4.6.4 Unilateral Alternatives

- If $H_1 : \sigma_1^2 > \sigma_2^2$, reject H_0 if

$$F > F_{1-\alpha}(n_1 - 1, n_2 - 1).$$

- If $H_1 : \sigma_1^2 < \sigma_2^2$, reject H_0 if

$$F < F_\alpha(n_1 - 1, n_2 - 1).$$

Remark 19. *The F-test is highly sensitive to normality. If the data are not approximately normal, alternative robust tests are preferred.*

4.7 independence test

Notations :

- T_i : theoretical frequencies
- O_i : observed frequencies
- n : total sample size
- k : number of rows
- l : number of columns

Objective : to study the relationship between two variables within the same sample. **Steps of the Test**

1. **Formulation of Hypotheses** :

$$H_0 : \text{The variables are independent.}$$
$$H_1 : \text{There is a relationship between the variables.}$$

2. **Calculation of theoretical frequencies** :

$$n_{i\cdot} = \sum_{j=1}^{l} O_{ij} \quad \text{for } i = 1, \ldots, k$$
$$n_{\cdot j} = \sum_{i=1}^{k} O_{ij} \quad \text{for } j = 1, \ldots, l$$
$$N = \sum_{i=1}^{k} \sum_{j=1}^{l} O_{ij}$$
$$T_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{N}$$

Validity Condition : All $T_{ij} \geq 5$.

3. **Test Statistic** :

$$\chi_c^2 = \sum_{j=1}^{l} \sum_{i=1}^{k} \frac{(O_{ij} - T_{ij})^2}{T_{ij}}$$

4. **Critical value** : The critical value is found in Table (chi-square distribution) using :
   - Row : Significance level α.
   - Column : Degrees of freedom $ddl = (c-1)(l-1)$.

$$\chi^2_T = \chi^2_{(ddl,1-\alpha)}$$

5. **Decision**
   - If $\chi^2_c \leq \chi^2_T$: Accept H_0.
   - If $\chi^2_c > \chi^2_T$: Reject H_0.

Example 4.7.1. *An episode of collective food poisoning occurred among the students of a primary school. A doctor was assigned to investigate and prepared the following table :*
Is the food poisoning linked to the consumption of chocolate ice cream ?

	Sick	**Healthy**
Chocolate Ice Cream	69	83
No Ice Cream	31	17

Solution
Choice of test : *Chi-squared test of independence*

1. ***Formulation of hypotheses***
   - H_0 *: There is no relationship between the food poisoning and the consumption of chocolate ice cream.*
   - H_1 *: There is a relationship between the food poisoning and the consumption of chocolate ice cream.*

2. ***Calculation of theoretical frequencies :*** *contingency table*

	Outcome		***Total*** $n_{i,.}$
	Sick	***Healthy***	
Chocolate Ice Cream	$O_{1,1} = 69$ $T_{1,1} = 76$	$O_{1,2} = 83$ $T_{1,2} = 76$	*152*
No Ice Cream	$O_{2,1} = 31$ $T_{2,1} = 24$	$O_{2,2} = 17$ $T_{2,2} = 24$	*48*
Total $n_{.,j}$	*100*	*100*	$n = 200$

All $T_{ij} \geq 5$, so the test is valid.

3. ***Test Statistic***

$$\chi_c^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(O_{ij}-T_{ij})^2}{T_{ij}} = 5.4$$

4. ***Critical value***
   - *Degrees of freedom :*

$$ddl = (c-1)(l-1) = (2-1)(2-1) = 1$$

   - *Significance level :*

$$1-\alpha = 1-0.05 = 0.95$$

   - *Critical value :*

$$\chi_T^2 = \chi^2_{(1,0.95)} = 3.841$$

5. ***Decision*** *Since $\chi_c^2 > \chi_T^2$:*
   - ***Statistical decision*** *: Reject H_0.*
   - ***Practical conclusion*** *: There is a link between food poisoning and the consumption of chocolate ice cream.*

4.8 Likelihood Ratio Test

This test is used when the previous methods do not apply.

(a) Test of $H_0 : \theta = \theta_0$ **against** $H_1 : \theta \neq \theta_0$

Here, θ may be a vector of dimension p. We define

$$\lambda = \frac{L(\underline{x}, \theta_0)}{\sup_{\theta\in\Theta} L(\underline{x}, \theta)}.$$

Remark 20. *We have $0 \leq \lambda \leq 1$. The larger λ is, the more plausible H_0 becomes. This corresponds to replacing θ under H_1 by its maximum likelihood estimator.*

The critical region of the test is

$$W = \left\{\underline{x} \in \mathbb{R}^n : \lambda < k_\alpha\right\}.$$

i.e. H_0 is rejected whenever $\lambda < k_\alpha$.

Theorem 6. *The distribution of $-2\ln\lambda$ is asymptotically χ^2_p under H_0.*

Remark 21. *The support $X(\Omega)$ must be independent of θ.*

Proof. For $p = 1$, we want to show that

$$-2\ln\lambda \xrightarrow[n\to\infty]{\mathcal{L}} \chi^2_1.$$

We write

$$\ln\lambda = \ln L(\underline{x}, \theta_0) - \ln L(\underline{x}, \hat{\theta}).$$

The Taylor expansion of $\ln L(\underline{x}, \theta_0)$ around $\hat{\theta}$ gives

$$\ln\lambda = \ln L(\underline{x}, \hat{\theta}) + (\theta_0 - \hat{\theta})\frac{\partial \ln L(\underline{x}, \hat{\theta})}{\partial\theta} + \frac{(\theta_0 - \hat{\theta})^2}{2}\frac{\partial^2 \ln L(\underline{x}, \theta^*)}{\partial\theta^2} - \ln L(\underline{x}, \hat{\theta}),$$
$$\theta^* \in [\theta_0, \hat{\theta}].$$

Thus,

$$\ln\lambda = \frac{1}{2}(\theta_0 - \hat{\theta})^2\frac{\partial^2 \ln L(\underline{x}, \theta^*)}{\partial\theta^2}, \qquad \theta^* \in [\theta_0, \hat{\theta}].$$

Under $H_0 : \theta = \theta_0$, the MLE satisfies

$$\hat{\theta} \xrightarrow[n\to\infty]{p.s.} \theta_0,$$

and therefore

$$\theta^* \xrightarrow[n\to\infty]{p.s.} \theta_0.$$

Moreover,

$$\frac{\partial^2 \ln L(\underline{x}, \theta^*)}{\partial\theta^2} = n\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \ln f(x_i, \theta)}{\partial\theta^2}\right).$$

By the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \ln f(X_i, \theta)}{\partial\theta^2} \xrightarrow[n\to\infty]{\mathcal{L}} \mathbb{E}\left[\frac{\partial^2 \ln f(X, \theta)}{\partial\theta^2}\right] = -I_1(\theta).$$

Thus,

$$\frac{\partial^2 \ln L(\underline{x}, \theta^*)}{\partial\theta^2} \xrightarrow[n\to\infty]{\mathcal{L}} -nI_1(\theta) = -I_n(\theta).$$

We obtain

$$-2\ln\lambda \longrightarrow (\theta_0 - \hat{\theta})^2 I_n(\theta).$$

Since

$$\frac{\theta_0 - \hat{\theta}}{\sqrt{1/I_n(\theta)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

it follows that

$$(\theta_0 - \hat{\theta})^2 I_n(\theta) \xrightarrow{\mathcal{L}} \chi_1^2.$$

Therefore,

$$-2\ln\lambda \xrightarrow{\mathcal{L}} \chi_1^2.$$

□

(b) Test between two composite hypotheses

We define

$$\lambda = \frac{\sup_{\theta\in H_0} L(\underline{x},\theta)}{\sup_{\theta\in H_1} L(\underline{x},\theta)},$$

and the same asymptotic properties as in part (a) apply.

Example 4.8.1. *Let $(X_1,\ldots,X_n)$ be an i.i.d. sample from $X \sim \mathcal{N}(\mu,1)$. Test at level α the hypotheses*

$$H_0:\ \mu = \mu_0 = 3 \qquad \textit{versus} \qquad H_1:\ \mu \neq \mu_0.$$

The likelihood ratio is

$$\lambda = \frac{L(\underline{x},\mu_0)}{L(\underline{x},\hat{\mu})}, \qquad W = \{\underline{x} \in \mathbb{R}^n:\ \lambda < k\}.$$

Compute λ explicitly :

$$\lambda = \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\sum_{i=1}^n (x_i-\mu_0)^2\right)}{\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\sum_{i=1}^n (x_i-\bar{x})^2\right)} = \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^n (x_i-\mu_0)^2 - \sum_{i=1}^n (x_i-\bar{x})^2\right]\right\}.$$

Hence

$$-2\ln\lambda = \sum_{i=1}^n (x_i-\mu_0)^2 - \sum_{i=1}^n (x_i-\bar{x})^2.$$

A short algebraic simplification yields

$$-2\ln\lambda = n(\bar{x}-\mu_0)^2.$$

The rejection rule $\lambda < k$ is equivalent to

$$-2\ln\lambda > -2\ln k \quad\Longleftrightarrow\quad n(\bar{x}-\mu_0)^2 > -2\ln k.$$

Under H_0 we have asymptotically $-2\ln\lambda \sim \chi^2_1$, so choose k so that

$$\mathbb{P}\Big(n(\bar{X}-\mu_0)^2 > -2\ln k\Big) = \alpha \quad \Longleftrightarrow \quad \mathbb{P}\Big(\chi^2_1 > -2\ln k\Big) = \alpha.$$

Therefore $-2\ln k$ is the $(1-\alpha)$-quantile of the χ^2_1 distribution, denoted $\chi^2_{1,1-\alpha}$, and

$$k = \exp\Big(-\tfrac{1}{2}\,\chi^2_{1,1-\alpha}\Big).$$

The critical region is thus

$$W = \Big\{\underline{x} \in \mathbb{R}^n : \ \lambda < \exp\Big(-\tfrac{1}{2}\,\chi^2_{1,1-\alpha}\Big)\Big\},$$

or equivalently

$$W = \Big\{\underline{x} \in \mathbb{R}^n : \ n(\bar{x}-\mu_0)^2 > \chi^2_{1,1-\alpha}\Big\}.$$