# Intervalles de confiance - tests statistiques

**G. Marot-Briend**
guillemette.marot@univ-lille.fr

2021-2022

# Estimation ponctuelle

Vocabulaire :

- écart-type de l'échantillon

$$s_{\text{ech}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- déviation standard (anglicisme)

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- erreur standard de la moyenne

$$\text{esm} = \frac{\sigma}{\sqrt{n}}$$

# Plan

# Estimation par intervalle

L'estimation par intervalle associe à un échantillon aléatoire un intervalle $[\widehat{\theta_1}, \widehat{\theta_2}]$ qui recouvre $\theta$ avec une certaine probabilité.
Cet intervalle est appelé intervalle de confiance.

On appelle risque d'erreur la probabilité $\alpha$ que l'intervalle de confiance ne contienne pas la vraie valeur du paramètre.
On appelle niveau de confiance la probabilité $1 - \alpha$ que l'intervalle de confiance contienne la vraie valeur du paramètre.

$$P(\widehat{\theta_1} < \theta < \widehat{\theta_2}) = 1 - \alpha$$

# Estimation par intervalle

Exercice :

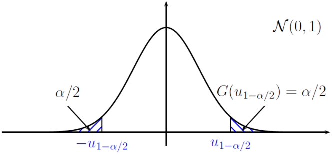
On cherche l'intervalle de confiance à 95% du poids des bébés.
$X \sim \mathcal{N}(3,3; 0,6)$

# Estimation par intervalle

Exercice :

On cherche l'intervalle de confiance à 95% du poids des bébés.
$X \sim \mathcal{N}(3,3;0,6)$

On veut $P(a < X < b) = 0.95$.
En centrant réduisant, on cherche a* et b* tels que
$P(a^* < X^* < b^*) = 0.95$
$\Rightarrow a^* = -u_{1-0.05/2}; b^* = u_{1-0.05/2}$




$$-u_{1-\alpha/2} < \frac{X-\mu}{\sigma} < u_{1-\alpha/2}$$

$$\mu - u_{1-\alpha/2}\sigma < X < \mu + u_{1-\alpha/2}\sigma$$

$$3,3 - 1,96 * 0,6 < X < 3,3 + 1,96 * 0,6$$

$$IC_{95\%} = [2,1;4,5]$$

# Intervalles de confiance

## Intervalles de confiance d'une moyenne

- si $\sigma$ est connu (rare)

$$IC(\mu) = \left[\bar{x} - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} ; \bar{x} + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

- si $\sigma$ est inconnu

$$IC(\mu) = \left[\bar{x} - t_{1-\alpha/2;n-1}\frac{s}{\sqrt{n}} ; \bar{x} + t_{1-\alpha/2;n-1}\frac{s}{\sqrt{n}}\right]$$

$n \leq 30$ petits échantillons : on doit supposer que X suit une loi normale
$n > 30$ grands échantillons : l'intervalle de confiance est valable quelle que soit la loi de X.

# Intervalles de confiance

## Intervalle de confiance d'une proportion

si $n\widehat{\pi} \geq 10$ et $n(1-\widehat{\pi}) \geq 10$, alors

$$\mathrm{IC}(\pi) = \left[\widehat{\pi} - u_{1-\alpha/2}\sqrt{\frac{\widehat{\pi}(1-\widehat{\pi})}{n}} ; \widehat{\pi} + u_{1-\alpha/2}\sqrt{\frac{\widehat{\pi}(1-\widehat{\pi})}{n}}\right]$$

# Plan

# Tests

Question fréquente :

**Est-ce que deux variables sont liées ?**

# Tests

Question fréquente :
**Est-ce que deux variables sont liées ?**

La réponse dépend du type des variables.

- deux variables qualitatives : test du $\chi_2$
- deux variables quantitatives : test du coefficient de corrélation
- une quantitative, une qualitative : t-test ou ANOVA

Tout comme les intervalles de confiance, les tests sont utilisés pour généraliser des résultats dans une population à partir d'observations d'un échantillon.

# Principe d'un test statistique

- Définir l'hypothèse nulle $H_0$ et l'hypothèse alternative $H_1$
- Choisir une statistique de test appropriée $T$ et sa distribution *sous l'hypothèse nulle*
- Choisir un niveau de signification $\alpha$, un seuil de probabilité en dessous duquel l'hypothèse nulle est rejetée.
- Calculer la réalisation de la statistique de test à partir des données observées et le degré de signification associé.
- Conclure au rejet ou non rejet de l'hypothèse nulle.

Remarque : $H_0$ est toujours l'hypothèse préférée (comme la présomption d'innocence). Pas de preuve suffisante $\Rightarrow$ pas de rejet. Quand on ne peut pas rejeter $H0$, cela ne veut pas dire qu'$H0$ est vraie.

# Exemples de tests statistiques

On pose souvent comme hypothèse nulle le contraire de ce que l'on cherche à prouver.

Exemples :

- les variables taille et poids ne sont pas corrélées
- l'âge moyen des personnes malades et celui des personnes saines sont égaux
- la proportion de fumeurs est la même dans le groupe sain et le groupe malade

# Vocabulaire :

Risque de 1ère espèce (niveau de significativité) :

$$\alpha = P(\text{rejet H0/H0 vraie})$$

Risque de 2ème espèce :

$$\beta = P(\text{non rejet H0/H1 vraie})$$

Puissance (capacité à obtenir un résultat statistiquement significatif dans le cas d'une réelle différence) :

$$1 - \beta = P(\text{rejeter H0/H1 vraie})$$

# Notion de risques $(\alpha, \beta)$ - Rappel

| | Réalité | |
|---|---|---|
| Décision | $H_O$ | $H_1$ |
| non rejet $H_0$ | conclusion correcte | risque de 2ème espèce $(\beta)$ |
| rejet $H_0$ | risque de 1ère espèce $(\alpha)$ | conclusion correcte |

| | Réalité | |
|---|---|---|
| Décision | $H_0$ | $H_1$ |
| non rejet $H_0$ | Niveau de confiance $1-\alpha$ | $\beta$ |
| rejet $H_0$ | $\alpha$ | Puissance $1-\beta$ |

# Degré de signification (p-value)

Ne pas confondre niveau de signification (risque de 1ère espèce $\alpha$), niveau de confiance $(1-\alpha)$ et degré de signification (p-value).

**Degré de signification ou p-value p(t) :**

pour une réalisation t d'une statistique de test T, probabilité (calculée sous l'hypothèse nulle) d'obtenir une statistique de test au moins aussi extrême que celle réellement observée.

Autrement dit, la p-value correspond à la plus petite valeur de risque $\alpha$ accepté pour cette réalisation.
p-value $< \alpha \Rightarrow$ on rejette $H_0$
p-value $\geq \alpha \Rightarrow$ on ne rejette pas $H_0$

# Degré de signification (p-value)

Plus la réalisation de la statistique de test est grande en valeur absolue, plus la p-value est petite.

Cas bilatéral : $p(t) = P_{H0}(|T| \geq |t| = 2.(1 - F(|t|))$
Cas unilatéral à droite : $p(t) = P_{H0}(T \geq |t|) = 1 - F(t)$
Cas unilatéral à gauche : $p(t) = P_{H0}(T \leq t) = F(t)$

Lien avec la région critique R : $P_{H0}(T \in R) = P(p(t) \leq \alpha)$

# Degré de signification (p-value)

Plus la réalisation de la statistique de test est grande en valeur absolue, plus la p-value est petite.

Cas bilatéral : $p(t) = P_{H0}(|T| \geq |t| = 2.(1 - F(|t|))$
Cas unilatéral à droite : $p(t) = P_{H0}(T \geq |t|) = 1 - F(t)$
Cas unilatéral à gauche : $p(t) = P_{H0}(T \leq t) = F(t)$

Lien avec la région critique R : $P_{H0}(T \in R) = P(p(t) \leq \alpha)$

Explications détaillées et rappel des formules sur les tests usuels dans le cours disponible en ligne de M. Genin (Univ Lille, METRICS) :
https ://pro.univ-lille.fr/michael-genin/enseignements/

# Tests paramétriques

## Test de comparaison de moyennes pour échantillons indépendants

1. Si $\sigma_1^2 = \sigma_2^2 = \sigma^2$, alors sous $H_O$ (test de Student) :

$$T = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \mathcal{T}_{n_1+n_2-2 \ ddl}$$

Avec $S^2$ l'estimateur de la variance commune $\sigma^2$.

$$S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

2. Si $\sigma_1^2 \neq \sigma_2^2$, alors sous $H_0$ :

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim \mathcal{T}_\nu \text{ ddl}$$

# Tests paramétriques

## Test de Student pour échantillons appariés

$$T = \frac{\bar{D}}{S_D/\sqrt{n}} \sim \mathcal{T}_{n-1\ ddl}$$

Avec

$$\bar{D} = \bar{X}_1 - \bar{X}_2$$

et

$$S_D = \sqrt{\frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n} D_i^2 - \left(\bar{D}\right)^2\right]}$$

# Tests paramétriques

## Comparaison de deux proportions / échantillons indépendants

Conditions de validité :

$$Z = \frac{\widehat{\pi_1} - \widehat{\pi_2}}{\sqrt{\frac{\widehat{\pi_1}(1-\widehat{\pi_1})}{n_1} + \frac{\widehat{\pi_2}(1-\widehat{\pi_2})}{n_2}}} \sim \mathcal{N}(0,1)$$

# Tests paramétriques

## Tests du $\chi^2$ : variables qualitatives

### Test du $\chi^2$ d'ajustement (1 variable qualitative / 1 échantillon)

Objectif : comparaison d'une distribution observée à une distribution théorique

Exemple : Soit un échantillon de 100 français. La distribution observée (sur l'échantillon) de l'âge regroupé en classes est-elle identique à celle de la population française ?

# Tests paramétriques

## Tests du $\chi^2$ : variables qualitatives

### Test du $\chi^2$ d'ajustement (1 variable qualitative / 1 échantillon)

Objectif : comparaison d'une distribution observée à une distribution théorique

Exemple : Soit un échantillon de 100 français. La distribution observée (sur l'échantillon) de l'âge regroupé en classes est-elle identique à celle de la population française ?

### Test du $\chi^2$ d'homogénéité (1 variable qualitative / plusieurs échantillons)

Objectif : comparaison de $k \geq 2$ distributions observées sur $k$ échantillons

Exemple : Soient trois échantillons de 100 français, 100 belges et 100 anglais. La distribution observée de l'âge regroupé en classes est-elle différente entre les populations ?

# Tests paramétriques

## Test du $\chi^2$ d'indépendance (2 variables qualitatives / 1 échantillon)

Objectif : Etudier la liaison entre deux variables qualitatives

Exemple : Soit un échantillon de 100 français. Existe-t-il un lien entre le sexe (Homme / Femme) et la couleur des yeux (Marrons, Bleus, Vert, ...) ?

$$K = \sum \frac{(O_{ij} - T_{ij})^2}{T_{ij}}$$

# Tests non paramétriques

Les tests paramétriques nécessitent des conditions de validité :

- Hypothèses sur la distribution des observations (ex : $X \sim \mathcal{N}(\mu, \sigma)$)
- Distributions caractérisées par des paramètres (moyenne, variance, ...)
- Ces paramètres sont estimés

# Tests non paramétriques

Les tests paramétriques nécessitent des conditions de validité :

- Hypothèses sur la distribution des observations (ex : $X \sim \mathcal{N}(\mu, \sigma)$)
- Distributions caractérisées par des paramètres (moyenne, variance, . . . )
- Ces paramètres sont estimés

**Question :**

Que faire lorsque les conditions de validité ne sont pas respectées ?

# Tests non paramétriques

Les tests paramétriques nécessitent des conditions de validité :

- Hypothèses sur la distribution des observations (ex : $X \sim \mathcal{N}(\mu, \sigma)$)
- Distributions caractérisées par des paramètres (moyenne, variance, . . . )
- Ces paramètres sont estimés

**Question :**

Que faire lorsque les conditions de validité ne sont pas respectées ?

$\Rightarrow$ **Tests non-paramétriques**

- Pas d'hypothèse sur la distribution
- La plupart du temps : tests basés sur la notion de rangs

# Tests non paramétriques

## Avantages

- Pas d'hypothèse sur la distribution $\Rightarrow$ champ d'application *a priori* plus large, tests notamment adaptés aux petits échantillons ($n < 30$)
- Tests adaptés aux variables ordinales (ex : degré de satisfaction), pas seulement aux variables quantitatives
- Tests peu sensibles aux valeurs extrêmes

## Inconvénients

- Lorsque les conditions d'applications sont vérifiées :

  Tests non paramétriques moins puissants que les tests paramétriques
- Difficultés d'interprétation : on ne compare plus des paramètres (moyenne, proportion, variance, . . . )

# Tests non paramétriques

**Comparaison de 2 échantillons**

Test de Mann-Whitney-Wilcoxon (échantillons indépendants) ou test des rangs signés de Wilcoxon (échantillons appariés) :

Soient

$F_1(X)$ la fonction de répartition de $X$ dans la population 1

$F_2(X)$ la fonction de répartition de $X$ dans la population 2

**Hypothèses**

$$\begin{cases} H_0 \ : \ F_1(X) = F_2(X+\theta) \ ; \ \theta = 0 & \text{Distributions identiques} \\ H_1 \ : \ F_1(X) = F_2(X+\theta) \ ; \ \theta \neq 0 & \text{Distributions différentes} \end{cases}$$

$\theta$ paramètre de translation : décalage entre les fonctions de répartition

# Tests non paramétriques

## Comparaison de plus de 2 échantillons indépendants

### Test de Kruskal-Wallis

Objectif : comparer la distribution d'une variable quantitative $X$ entre $K$ groupes indépendants (Equivalent non paramétrique de l'ANOVA à un facteur).

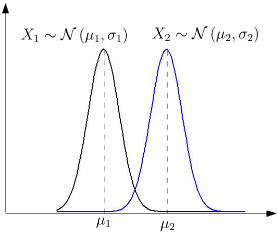
**Hypothèses**

$$\begin{cases} H_0 \ : \ F_1(X) = F_2(X+\theta) = \ldots = F_K(X+\theta) \ ; \ \theta = 0 \\ \ \text{Distributions identiques} \\ H_1 \ : \ \exists i \neq j / F_i(X) = F_j(X+\theta); \ \theta \neq 0 \\ \ \text{Distributions différentes} \end{cases}$$

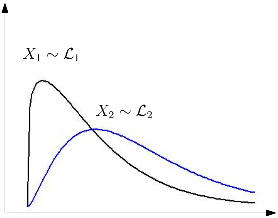
$\theta$ paramètre de translation : décalage entre les fonctions de répartition

# Tests non paramétriques

Tests paramétrique vs. test non paramétrique




$$\begin{aligned} \mathrm{H}_0 &: \mu_1 = \mu_2 \\ \mathrm{H}_1 &: \mu_1 \neq \mu_2 \end{aligned}$$




$$\begin{aligned} \mathrm{H}_0 &: \mathcal{L}_1 = \mathcal{L}_2 \\ \mathrm{H}_1 &: \mathcal{L}_1 \neq \mathcal{L}_2 \end{aligned}$$

Remarques : Les tests de Wilcoxon, Mann-Whitney-Wilcoxon sont plus facilement interprétables quand les distributions sont symétriques, de même forme et nécessitent des amplitudes ou des rangs disponibles.