

Feuille Cours_14 : Test d'hypothèse de conformité à la moyenne.

Dans toute cette feuille on écrira (*) pour rappeler les hypothèses du théorème limite :

Soit $(X_1, \dots, X_n)$ une liste de variables aléatoires indépendantes de même loi possédant une espérance μ et une variance σ^2 ; on note : $M_n = \frac{1}{n}(X_1 + \dots + X_n)$ et $S_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - M_n)^2$

Rappels :

- Dans les conditions de (*) on a : $\lim_{n \rightarrow +\infty} \mathbb{P} \left(\left[-u_{1-\frac{\alpha}{2}} \leq \frac{M_n - \mu}{\frac{S_n}{\sqrt{n}}} \leq u_{1-\frac{\alpha}{2}} \right] \right) = 1 - \alpha$
 - Si $X \hookrightarrow \mathcal{N}(0, 1)$ alors pour $x \in \mathbb{R}$, $P(|X| \leq x) = 2\Phi(x) - 1$, $P(|X| > x) = 2 - 2\Phi(x)$
 - Si $X \hookrightarrow \mathcal{N}(0, 1)$ alors $P(|X| \geq 1.96) \approx 0,05$
 - Si $X \hookrightarrow \mathcal{N}(\mu, \sigma^2)$ alors $P\left(\left|\frac{X - \mu}{\sigma}\right| \geq 1.96\right) \approx 0,05$
 - Dans les conditions de (*) et pour n assez grand, $P\left(\left|\frac{M_n - \mu}{\frac{\sigma}{\sqrt{n}}}\right| \geq 1.96\right) \approx 0,05$
 - Dans les conditions de (*) et pour n assez grand, $P\left(\left|\frac{M_n - \mu}{\frac{S_n}{\sqrt{n}}}\right| \geq 1.96\right) \approx 0,05$
-

Situation :

On a une suite de résultats : $(x_1, \dots, x_n)$ (réalisation d'un échantillon $(X_1, \dots, X_n)$ d'une loi mère d'espérance μ) et on veut savoir si $\mu = \mu_0$.

On note : $H_0 : \mu = \mu_0$ (hypothèse) et $H_1 : \mu \neq \mu_0$ (l'hypothèse alternative)

Le **test de conformité à la moyenne** consiste en :

- 1) Calculer une estimation des deux paramètres :

$$m_n = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{et} \quad s_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - m_n)^2}$$

Remarque : m_n et s_n sont respectivement la réalisation des variables aléatoires M_n et de S_n .

- 2) Calculer la statistique : $T_n = \frac{m_n - \mu_0}{\frac{s_n}{\sqrt{n}}}$

- 3) Tester si $\left| \frac{m_n - \mu_0}{\frac{s_n}{\sqrt{n}}} \right| > 1,96$

- 4) Décider :

- Si $|T_n| > 1,96$ (zone critique) alors on rejette l'hypothèse H_0 au seuil de 5%.
- Sinon on ne rejette pas H_0 .

Remarque : Le seuil de 5% correspond au risque de rejeter H_0 alors qu'elle est vraie.

- 5) *Pour aller plus loin :* calculer la p -value.

Sous l'hypothèse H_0 , la p -value est la probabilité d'avoir une statistique plus ou aussi loin de zéro

- a. Expliquer pourquoi la p -value est égale à $2(1 - \Phi(|T_n|))$.
- b. Faire une illustration graphique.

1 Exercices

- 1) On mesure la taille (en cm) de 50 plantes. On obtient : $m_n = 102$ $s_n = 10$
On souhaite tester si la taille moyenne est égale à 100 cm.
 - a. Formuler les hypothèses H_0 et H_1 et calculer la statistique T_n .
 - b. Conclure au seuil de 5%.
 - c. Proposer une interprétation biologique du résultat.
 - d. Calculer la p -value associée et l'interpréter.
- 2) On mesure une production enzymatique (en unité arbitraire) sur 40 essais. On obtient : $m_n = 12,5$; $s_n = 2$
On teste l'hypothèse $\mu = 11$.
 - a. Faire un test de conformité à la moyenne au seuil de 5%.
 - b. Interpréter le résultat dans le contexte expérimental.
 - c. Calculer la p -value. Que représente cette p -value dans le contexte ?
- 3) On mesure la masse (en mg) de 100 graines. On obtient : $m_n = 48$ $s_n = 12$
On souhaite tester si la masse moyenne vaut 50 mg.
 - a. Calculer la statistique T_n .
 - b. Conclure au seuil de 5%.
 - c. Expliquer pourquoi un écart entre m_n et μ_0 ne suffit pas à rejeter H_0 .
- 4) On mesure les temps d'attente (en minutes) d'un bus pour 80 usagers. On obtient : $m_n = 6,5$ $s_n = 3$
On teste si le temps moyen est de 5 minutes.
 - a. Calculer T_n .
 - b. Conclure au seuil de 5%.
 - c. Interpréter le résultat dans le cadre du modèle théorique étudié.
- 5) On observe un phénomène sur 200 individus. On modélise les résultats par une variable réelle.
On obtient : $m_n = 0,42$ et $s_n = 0,1$ et on souhaite tester si la moyenne vaut 0,5.
 - a. Calculer la statistique T_n .
 - b. Conclure au seuil de 5%.
 - c. Comment interpréter un rejet très net de H_0 ?
- 6) On veut tester dans les deux cas suivants l'hypothèse : $\mu = 100$.
Expérience A : $n = 25$, $m_n = 102$, $s_n = 10$ **Expérience B** : $n = 100$, $m_n = 102$, $s_n = 10$
 - a. Calculer T_n dans chaque cas.
 - b. Comparer les décisions.
 - c. Expliquer l'influence de la taille de l'échantillon sur le test.
- 7)
 - a. On lance 1000 fois un dé et on obtient 150 fois la face 6. Le dé semble-t-il équilibré ?
 - b. On lance 100000 fois un dé et on obtient 15000 fois la face 6. Le dé semble-t-il équilibré ?
- 8) Un agriculteur a mis en culture 100 plants de soja pour lesquels l'industriel affirme qu'au bout de 6 semaines ils mesurent en moyenne 40 cm.
 Au bout de 6 semaines il mesure les hauteurs suivantes :

hauteurs	36	37	38	39	40	41
effectifs	6	11	26	32	14	11

 - a. Calculer la moyenne empirique m_n et la variance empirique s_n^2 de cette série statistique.
 - b. Calculer : $\frac{m_n\sqrt{n} - 40\sqrt{n}}{s_n}$.
 - c. Peut-il, à la vue de ces valeurs, mettre en doute la valeur annoncée par l'industriel ?
- 9) *Questions de synthèse.*
 - a. Expliquer pourquoi : “Ne pas rejeter H_0 ne signifie pas que H_0 est vraie.”
 - b. Comparer le test de conformité à la moyenne avec le test du χ^2 :
nature des données ; statistique utilisée ; type de conclusion.
 - c. Préciser les conditions d'application du test de conformité à la moyenne.
 - d. Compléter la phrase : “Une petite p -value signifie que ...”
 - e. Expliquer pourquoi la p -value permet une prise de décision plus fine qu'un seuil fixé à 1,96.