

Chapitre 3 : Échantillonnage

on prend un échantillon de la pop (EAS/empirique)

↑
Échantillon aléatoire simple

Si $\bar{X}$ (moyenne empirique)

$$E(\bar{X}) = \mu$$

$$E(X) = \mu$$

$$V(\bar{X}) = \frac{\sigma^2}{n}$$

$$V(X) = \sigma^2$$

$$\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

$$\sigma(X) = \sigma$$

Si $X \sim \mathcal{N}(\mu, \sigma^2)$ alors $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

Si $X \sim \mathcal{B}(n, p) \rightarrow$ approximation $X \sim \mathcal{N}(np, \sqrt{np(1-p)})$

$\hookrightarrow f_n = \frac{X}{n}$ $f_n \sim \mathcal{D}^p(p; \sqrt{\frac{p(1-p)}{n}})$ On divise par n vu que c'est $\frac{X}{n}$

Chapitre 4 : Estimations

• Estimations ponctuelle (valeur paramètre $\rightarrow \mu, \sigma$)

Propriétés d'un bon estimateur U du paramètre θ

• SANS BIAIS $E(U) = \theta$

• EFFICACITÉ $\frac{Var(U)}{Var(V)}$ Variance la + faible

• CONVERGENT $Var(U) \xrightarrow{+\infty} 0$ et $U \xrightarrow{+\infty} \theta$

• ASYMPTOTIQUEMENT SANS BIAIS $E(U) - \theta \xrightarrow{+\infty} 0$

$$E(nX) = nE(X)$$

$$V(nX) = n^2 V(X)$$

$$E(2) = 2$$

$$V(2) = 0$$

$$E(nX+2) = nE(X)+2$$

$$V(nX+2) = n^2 V(X)$$

$$E(\frac{X}{n}) = \frac{1}{n} E(X)$$

! $\frac{1}{n}$ si E / $\frac{1}{n^2}$ si V

• Estimations par IC

Les VA X_i sont iid

Si $X \sim \mathcal{B}(n, p) \rightarrow$ approximation $X \sim \mathcal{N}(np, \sqrt{np(1-p)})$
 $f_n \sim \mathcal{D}^p(p, \sqrt{\frac{pq}{n}})$

Proportion f_n

CAS GÉNÉRALE

$$f_n - u \sqrt{\frac{f_n(1-f_n)}{n}} < p < f_n + u \sqrt{\frac{f_n(1-f_n)}{n}}$$

f_n (proportion) = $\frac{260}{300}$ n = (nombre) u valeur IC

• Si 95% $\rightarrow u = 1,96$

• Si 99% $\rightarrow u = 2,58$

CAS UNILATÉRAUX

• + petit que $0 < p < f_n + u \sqrt{\frac{f_n(1-f_n)}{n}}$

• + grand que $f_n - u \sqrt{\frac{f_n(1-f_n)}{n}} < p < 1$

- $P(Z \geq \dots) \rightarrow 1 - F_Z(\dots)$
- $P(0 \leq Z \leq 2) \rightarrow 2 - 0 \rightarrow 2F_Z(\dots) - 1$
- $P(Z \geq -\dots) \rightarrow P(-Z \leq \dots) \rightarrow F_Z(\dots)$
- $P(Z \leq -\dots) \rightarrow 1 - P(Z \leq \dots) \rightarrow 1 - F_Z(\dots)$



TIPS

Δ intervalle de longueur 0,1 $\rightarrow$ 0,05 de CHAQUE CÔTÉ
3,6 $\Rightarrow$ $[3,55; 3,65]$

• tjrs ISOLER ce qu'on cherche

• $\bar{X} + u \frac{\sigma}{\sqrt{n}} = 3,65$ Or on sait que $\bar{X} = 3,6$
 $\hookrightarrow u \frac{\sigma}{\sqrt{n}} = 0,05$

$X \sim \mathcal{N}(\mu, \sigma^2) \rightarrow \mathcal{N}(0, 1)$ tq $Z = \frac{X - \mu}{\sigma}$

LOI NORMALE:

CAS σ^2 (variance connue)

CAS UNILATÉRAUX

$$\bar{X} - u \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + u \frac{\sigma}{\sqrt{n}} \text{ tq } \bar{X} \text{ (moyenne)}$$

• + petit que $p < \bar{X} + u \frac{\sigma}{\sqrt{n}}$
• + grand que $\bar{X} - u \frac{\sigma}{\sqrt{n}} < p$

$$u = 2\psi(u) - 1$$

CAS σ^2 (VARIANCE INCONNUE)

$$\bar{X} - t_{n-1} \frac{S'}{\sqrt{n-1}} \leq \mu \leq \bar{X} + t_{n-1} \frac{S'}{\sqrt{n-1}}$$

$S' = \hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2}$
 $t_{n-1} \Rightarrow$ Quartile loi Student

• Si 95% $\rightarrow u = 1,96$

• Si 99% $\rightarrow u = 2,58$

CAS ON VEUT L'IC de σ^2 en connaissant μ

$$\frac{nS^2}{K_1} \leq \sigma^2 \leq \frac{nS^2}{K_2}$$

$$\sigma^2 \geq n \frac{S^2}{K_1} \text{ (UNILATÉRAL)}$$

CAS μ CONNUE

CAS μ INCONNUE

$$n \frac{\hat{\sigma}^2}{R_1} \leq \sigma^2 \leq n \frac{\hat{\sigma}^2}{R_2}$$

$$\frac{nS^2}{R_1} \leq \sigma^2 \leq \frac{nS^2}{R_2}$$

$$n \frac{\hat{\sigma}^2}{R_1} \leq \sigma^2 \leq n \frac{\hat{\sigma}^2}{R_2}$$

$R_1/R_2 =$ Quartile loi $\chi^2_{k=2}$

$$\hat{\sigma}^2 = S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2$$

$$\frac{(n-1)S^2}{R_1} < \sigma^2 < \frac{(n-1)S^2}{R_2}$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$$

THÉORIE DES TESTS

Différents tests

Test unilatéral à gauche

Test bilatéral

Test unilatéral à droite

$$\begin{cases} H_0: \rho = \dots \\ H_A: \rho < \dots \end{cases}$$

$$\begin{cases} H_0: \rho = \dots \\ H_A: \rho \neq \dots \end{cases}$$

$$\begin{cases} H_0: \rho = \dots \\ H_A: \rho > \dots \end{cases}$$

$$C\alpha = F^{-1}(1-\alpha)$$

$$C\alpha = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

$$C\alpha = F^{-1}(1-\alpha)$$

Règles de décision

si $T_0 \leq -C\alpha$ on rejette H_0 au seuil α
 si $T_0 \geq C\alpha$ on conserve H_0 au risque de second espèce β

si $Z_0 \geq C\alpha$ on rejette H_0 au seuil α
 si $Z_0 \leq -C\alpha$ on conserve H_0 au risque de second espèce β

Si test de proportion

"si échantillon de X pers dont Y personnes"

Estimateur de $p = \hat{p} = \frac{1}{n} \sum X_i$ où $X_i \sim \text{Ber}(p)$

avec $X_i = \begin{cases} 1 & \text{si l'ind } i \text{ est un } \dots \\ 0 & \text{sinon} \end{cases}$

la stat de test est $z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$

où $C\alpha = \Phi^{-1}(1-\alpha)$ est le fractile de la loi $N(0,1)$

voir dans table Appendix B.1.

si seuil 5% = 1,64

si seuil 1% = 2,32

Si on cherche $H_0: m$ ou $u = \dots$ avec $X \sim N(m \text{ ou } u; \sigma)$

$\bar{X} \sim N\left(m; \frac{\sigma}{\sqrt{n}}\right)$ avec $t \sim \text{St}(n-1)$

la stat de test est

$$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

calculs:

$$\bar{x} = \frac{1}{n} \sum x_i = A$$

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 = B$$

$$s = \sqrt{B} = C$$

$$t = \frac{A - \mu_0}{\frac{C}{\sqrt{n}}}$$

$$C\alpha = F^{-1}(1-\alpha)$$

voir dans table Appendix B2

Si $\bar{X}$ et σ connues

la stat de test est $T = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

avec $t \sim \text{St}(n-1)$

voir dans table Appendix B2

$C\alpha = F^{-1}(1-\alpha)$ pour $\alpha = 5\%$ ou 1%

Si on cherche $H_0: \sigma = \dots$ avec $X \sim N(m; \sigma)$

la stat de test est $K = \frac{(n-1)S^2}{\sigma^2}$

$K \sim \chi^2(n-1)$

IC au seuil de la variance
 ↳ Δ bilatéral $(1 - \frac{\alpha}{2})$ et $(\frac{\alpha}{2})$

cf. calculs:

$C\alpha_1 = F^{-1}(1-\alpha)$ puis vérifier avec $C\alpha_2 = F^{-1}\left(\frac{\alpha}{2}\right)$

voir dans table Appendix B3

Hypothèses

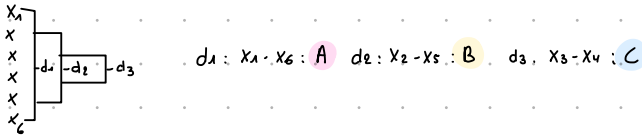
Hypothèse de normalité de X

- ① On veut tester au seuil 1% → on connaît S^2
 $\hookrightarrow \begin{cases} H_0 : X \sim N(\mu_x; \sigma_x^2) \\ H_1 : X \text{ ne suit pas une loi normale} \end{cases}$

② Calculer w :

$$w = \frac{(\sum a_j d_j)^2}{(n-1)S^2}$$

→ ranger toutes les valeurs par ordre croissant
 $d_1; d_2; d_3 \dots \rightarrow$ si impair une valeur = 0



→ Table Shapiro avec d en colonne & n en ligne

n	d	S	W
1	0,6646	0,6431	
2	0,2413	0,2806	
3	0,0000	0,0875	
4			
5			
6			
7			
8			

$$(A \times 0,6431 + B \times 0,2806 + C \times 0,0875)^2 = W$$

$(6-1) \times S^2$
 $\hookrightarrow$ variance empirique

③ Calculer w^*

→ Table Shapiro avec n en colonne et $\alpha\%$ en ligne

	Risque 5%	Risque 1%
n	$W_{5,00}$	$W_{1,00}$
5	0,762	0,686
6	0,788	0,713
7	0,803	0,730

- $\hookrightarrow$ si $w \leq w^*$ on rejette H_0 au seuil α
 si $w \geq w^*$ on conserve H_0 au risque de second espèce β

Hypothèse variance Y = variance X

- ① On veut tester au seuil 1% → on connaît S^2_x
 $\hookrightarrow \begin{cases} H_0 : \sigma_y^2 = \sigma_x^2 \\ H_1 : \sigma_y^2 \neq \sigma_x^2 \end{cases}$

→ $\sigma_y^2 = (\text{donné ds la question})^2 = A = S^2_y$
 → $\sigma_x^2 = B \quad \Delta \neq$ même que S^2_x

② Stat de test : $F = \frac{S^2_y}{S^2_x} = \frac{A}{C} \rightarrow$ calculé avant

③ Table Fisher : $F(n_y - 1; n_x - 1)$
 $\hookrightarrow 0,99$ ou $0,95 \quad \hookrightarrow$ colonne $\hookrightarrow$ ligne

n_2	10
1	6055.85
2	99.40
3	27.23
4	14.55
5	10.05
6	7.87
7	6.62
8	5.81
9	5.26
10	4.85

$$\text{Ex : } F(11-1; 7-1) = F(10; 6)$$

- $\hookrightarrow$ si $F \geq C_\alpha$ on rejette H_0 au seuil α
 si $F \leq C_\alpha$ on conserve H_0 au risque de second espèce β

Hypothèse moyenne Y = moyenne X

- ① On veut tester au seuil 5% → on connaît $\bar{X}$ & $\bar{Y}$
 $\hookrightarrow \begin{cases} H_0 : \bar{Y} = \bar{X} \\ H_1 : \bar{Y} \neq \bar{X} \end{cases}$

② Stat de test : $T = \frac{\bar{X} - \bar{Y}}{\sqrt{S^2 \times \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}}$

$$S^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$$

③ Table Student : 0,950 $\text{St}(n_x + n_y - 2)$

n	0,750	0,900	0,950
1	1,000	3,078	6,314
2	0,816	1,886	2,920
3	0,765	1,638	2,353
4	0,741	1,533	2,132
5	0,727	1,476	2,015
6	0,718	1,440	1,943
7	0,711	1,415	1,895
8	0,706	1,397	1,860
9	0,703	1,383	1,833
10	0,700	1,372	1,812
11	0,697	1,363	1,796
12	0,695	1,356	1,782
13	0,694	1,350	1,771
14	0,692	1,345	1,761
15	0,691	1,341	1,753
16	0,690	1,337	1,746
17	0,689	1,333	1,740

$$\text{Ex : } \text{St}(11 + 7 - 2) = \text{St}(16)$$

- $\hookrightarrow$ si $T \geq C_\alpha$ on rejette H_0
 si $T \leq C_\alpha$ on conserve H_0
 $\hookrightarrow$ m moyenne

Formulaire des principaux tests paramétriques

I. Tests d'hypothèses simples (comparaison à une valeur de référence), au seuil α (Exemples donnés pour test bilatéral, à adapter si besoin pour test unilatéral)

Test	Hyp.	Statistique de test	Loi	Décision
Espérance d'une loi normale, variance connue	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$	$Z \sim N(0, 1)$	On lit la valeur critique C_α dans la table appropriée. Si $ Z \geq C_\alpha$, on rejette H_0 Si $ Z < C_\alpha$, on accepte H_0 .
Espérance d'une loi normale, variance inconnue	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$T = \frac{\bar{X} - \mu_0}{S / \sqrt{n}}, \text{ avec}$ $S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$	$T \sim St(n-1)$	On lit la valeur critique C_α dans la table appropriée. Si $ T \geq C_\alpha$, on rejette H_0 Si $ T < C_\alpha$, on accepte H_0 .
Variance d'une loi normale, espérance connue (rare !!)	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$	$K = \frac{n\hat{\sigma}^2}{\sigma_0^2}, \text{ avec}$ $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$	$K \sim \chi^2(n)$ NB : loi non symétrique !	On lit la valeur critique C_α dans la table appropriée. Si $K \geq C_\alpha$, on rejette H_0 Si $K < C_\alpha$, on accepte H_0 .
Variance d'une loi normale, espérance inconnue	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$	$K = \frac{n\hat{\sigma}^2}{\sigma_0^2} = \frac{(n-1)S^2}{\sigma_0^2}$ avec $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2$ $S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$	$K \sim \chi^2(n-1)$ NB : loi non symétrique !	On lit la valeur critique C_α dans la table appropriée. Si $K \geq C_\alpha$, on rejette H_0 Si $K < C_\alpha$, on accepte H_0 .
Proportion (pour n réalisations indépendantes d'une V.A. $B(1, p)$)	$H_0: p = p_0$ $H_1: p \neq p_0$	$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	$Z \sim N(0, 1)$	On lit la valeur critique C_α dans la table appropriée. Si $ Z \geq C_\alpha$, on rejette H_0 Si $ Z < C_\alpha$, on accepte H_0 .

II. Test de comparaisons de valeurs entre deux populations indépendantes (Exemple donné pour test bilatéral, à adapter si besoin pour test unilatéral)

Comparaison	Hyp.	Statistique de test	Loi	Décision
Espérances de lois normales, variances connues	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$Z \sim N(0, 1)$	Si $ Z \geq C_\alpha$, rejeter H_0 Si $ Z < C_\alpha$, accepter H_0 .
Espérances de lois normales, variances inconnues égales ($\sigma_1^2 = \sigma_2^2 = \sigma^2$)	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	$T = \frac{\bar{X}_1 - \bar{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \text{ avec}$ $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$ $= \frac{n_1 \hat{\sigma}_1^2 + n_2 \hat{\sigma}_2^2}{n_1 + n_2 - 2}$	$T \sim St(n_1 + n_2 - 2)$	Si $ T \geq C_\alpha$, rejet de H_0 Si $ T < C_\alpha$, accepter H_0 .
Espérances de lois normales, variances inconnues	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1 - 1} + \frac{\hat{\sigma}_2^2}{n_2 - 1}}}$	$T \sim St(\nu)$ où ν est l'entier le + proche de $\frac{\left(\frac{\hat{\sigma}_1^2}{n_1 - 1} + \frac{\hat{\sigma}_2^2}{n_2 - 1}\right)^2}{\frac{\hat{\sigma}_1^4}{(n_1 - 1)n_1^2} + \frac{\hat{\sigma}_2^4}{(n_2 - 1)n_2^2}}$	Si $ T \geq C_\alpha$, rejet de H_0 Si $ T < C_\alpha$, accepter H_0 .
Espérances de lois quelconques, variances inconnues	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1 - 1} + \frac{\hat{\sigma}_2^2}{n_2 - 1}}}$	$Z \sim N(0, 1)$ en approximation	Si $ Z \geq C_\alpha$, rejet de H_0 Si $ Z < C_\alpha$, accepter H_0 .
Variances de lois normales, espérance inconnue	$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$	$F = \frac{S_1^2}{S_2^2} \text{ avec } S_1^2 > S_2^2$ <u>NB</u> : si $S_2^2 > S_1^2$, on prend $F = S_2^2/S_1^2$ pour avoir $F > 1$. Prendre $(n_2 - 1, n_1 - 1)$ comme d.d.l. dans la loi de Fisher.	$F \sim \text{Fisher } \mathfrak{F}(n_1 - 1, n_2 - 1)$ <u>NB</u> : densité non nulle si $F \geq 0$, mais nulle si $F < 0$.	Trouver C_α dans la table de Fisher pour α . Si $F \geq C_\alpha$ rejeter H_0 Si $F < C_\alpha$ accepter H_0 . <u>NB</u> : attention, ce test est bilatéral mais se lit ici comme unilatéral !
Proportions, pour caractère B(1, p) observé dans deux populations indépendantes.	$H_0: p_1 = p_2$ $H_1: p_1 \neq p_2$	$Z = \frac{n_{p1} - \frac{n_p \times n_1}{n}}{\sqrt{\frac{n_1 \times n_2 \times n_p \times n_{(1-p)}}{n^2(n-1)}}}$ n_1 = taille de l'échantillon 1 n_2 = taille de l'échantillon 2 $n = n_1 + n_2$ n_{p1} = nombre de succès dans l'échantillon 1 n_p = nombre total de succès dans les deux échantillons $n_{(1-p)}$ = nombre total d'échecs dans les deux échantillons	$Z \sim N(0, 1)$	Si $ Z \geq C_\alpha$ rejet de H_0 Si $ Z < C_\alpha$ accepter H_0 .