Échantillonnage

Chapitre 2 : Estimateur, biais et précision

Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 21 octobre 2026

Plan du chapitre

L'estimateur de Horvitz-Thompson

La variance de l'estimateur

Estimer la variance

Biais, variance et erreur quadratique

L'intervalle de confiance

S'entraîner

Ce qu'il faut retenir

L’estimateur de Horvitz-Thompson

L'estimateur de Horvitz-Thompson du total

$$\widehat{Y}_{HT} = \sum_{k \in S} \frac{y_k}{\pi_k} = \sum_{k \in S} d_k \, y_k = \sum_{k \in U} I_k \, \frac{y_k}{\pi_k}$$

Chaque unité tirée compte pour son **poids** $d_k = 1/\pi_k$ (chapitre 1). La dernière écriture fait apparaître les indicatrices I_k : seules les unités tirées contribuent.

Un ménage tiré en représente plusieurs

Un ménage tiré avec la probabilité $0{,}1$ représente 10 ménages : sa dépense est comptée 10 fois. Un ménage tiré avec la probabilité $1/2$ n'en représente que 2. Les unités **rarement** tirées comptent **davantage** quand elles le sont : c'est ce qui compense leur rareté.

La démonstration en une ligne

Les y_k/π_k sont des nombres fixes, et $E(I_k) = \pi_k$:

$$E(\widehat{Y}_{HT}) = \sum_{k \in U} E(I_k)\,\frac{y_k}{\pi_k} = \sum_{k \in U} \pi_k\,\frac{y_k}{\pi_k} = \sum_{k \in U} y_k = Y$$

Seule l'espérance des I_k intervient : le résultat vaut pour **tout** plan, à une condition, $\pi_k > 0$ pour **chaque** unité de U.

Une unité oubliée ne se rattrape pas

Si $\pi_k = 0$, l'unité k disparaît de la somme : $E(\widehat{Y}_{HT}) = Y - \sum_{\pi_k = 0} y_k$. Une base de sondage qui omet les 134 ménages des quatre districts ruraux d'Al Hoceïma fait estimer, en moyenne, $Y - 739\,640$ dh : $4{,}7\,\%$ de moins que le vrai total, **quelle que soit** la taille de l'échantillon.

Estimer la moyenne

Quand N est connu, $\widehat{\overline{Y}}_{HT} = \widehat{Y}_{HT}/N$ est sans biais pour $\overline{Y}$. Pour un sondage aléatoire simple, $\pi_k = n/N$ et

$$\widehat{Y}_{HT} = \frac{N}{n}\sum_{k\in S} y_k = N\,\overline{y}, \qquad \widehat{\overline{Y}}_{HT} = \overline{y}$$

La moyenne d'échantillon est l'estimateur de Horvitz-Thompson de ce plan : elle est **sans biais**.

Mini-population : deux ménages sur cinq

$\pi_k = 0{,}4$ et $d_k = 2{,}5$. Échantillon $\{A, E\}$: $\widehat{Y}_{HT} = 2{,}5 \times (4 + 14) = 45$ milliers de dh, soit une moyenne de 9. Les dix échantillons donnent 25 ; $27{,}5$; $32{,}5$; 45 ; $32{,}5$; $37{,}5$; 50 ; 40 ; $52{,}5$; $57{,}5$, de moyenne exactement $Y = 40$.

Un ménage parmi A, B, C, un ménage parmi D, E

$\pi_A = \pi_B = \pi_C = 1/3$ et $\pi_D = \pi_E = 1/2$, d'où $\widehat{Y}_{HT} = 3\,y_1 + 2\,y_2$, où y_1 vient du premier groupe.

Échantillon	AD	AE	BD	BE	CD	CE
$\widehat{Y}_{HT}$	30	40	36	46	39	49

Chaque échantillon a la probabilité $1/6$: $E(\widehat{Y}_{HT}) = 240/6 = 40 = Y$.

La moyenne pondérée du chapitre 1

Divisé par $N = 5$: $\widehat{\overline{Y}}_{HT} = \frac{3}{5}\,y_1 + \frac{2}{5}\,y_2$. C'est la moyenne pondérée que le chapitre 1 avait construite « à la main ». L'estimateur de Horvitz-Thompson la retrouve **automatiquement**, à partir des seuls π_k.

Un tirage de 50 ménages par province

Province	$d_k = N_h/50$	$W_h = N_h/N$	Moyenne $\overline{y}_h$
Tanger-Assilah	16,4	0,41	9 030,4
Tétouan	11,2	0,28	7 121,6
Larache	7,6	0,19	7 551,0
Al Hoceïma	4,8	0,12	6 137,4

$\widehat{Y}_{HT} = \sum_{k\in S} d_k\, y_k = \sum_h N_h\, \overline{y}_h = 15\,735\,380$ dh, soit
$\widehat{\overline{Y}}_{HT} = \sum_h W_h\, \overline{y}_h = 7\,867{,}69$ dh.

Les poids font le travail

Moyenne **brute** des 200 ménages : $7\,460{,}1$ dh, contre $\overline{Y} = 7\,911{,}44$. L'écart de l'estimation pondérée, -44 dh, vient du hasard ; celui de la brute est surtout **systématique** : on le mesurera plus loin.

La variance de l’estimateur

Variance de l'estimateur de Horvitz-Thompson

$$V(\widehat{Y}_{HT}) = \sum_{k\in U}\sum_{l\in U} \Delta_{kl}\,\frac{y_k}{\pi_k}\,\frac{y_l}{\pi_l}, \qquad \Delta_{kl} = \pi_{kl} - \pi_k\,\pi_l, \quad \Delta_{kk} = \pi_k\,(1-\pi_k)$$

La démonstration

Posons $a_k = y_k/\pi_k$, un nombre fixe. $\widehat{Y}_{HT} = \sum_k I_k\,a_k$ est une somme de variables aléatoires, dont la variance est la somme de toutes les covariances :

$$V\Big(\sum_k I_k\,a_k\Big) = \sum_k\sum_l a_k\,a_l\,\mathrm{Cov}(I_k, I_l) \quad \text{et} \quad \mathrm{Cov}(I_k, I_l) = \Delta_{kl}$$

d'après le chapitre 1 ; pour $k = l$, $\mathrm{Cov}(I_k, I_k) = V(I_k) = \pi_k(1-\pi_k)$.

Pour un plan de taille fixe

$$V(\widehat{Y}_{HT}) = \frac{1}{2} \sum_{k \in U} \sum_{l \neq k} (\pi_k \, \pi_l - \pi_{kl}) \left(\frac{y_k}{\pi_k} - \frac{y_l}{\pi_l} \right)^2$$

Elle est **nulle** si y_k / π_k est le même pour toutes les unités, c'est-à-dire si π_k est proportionnel à y_k : tirer plus souvent les grandes unités peut rendre l'estimateur très précis (chapitre 11).

D'où elle vient

Si $n = \sum_l I_l$ est fixe, $\mathrm{Cov}(I_k, n) = 0$, donc $\sum_l \Delta_{kl} = 0$ et $\Delta_{kk} = -\sum_{l \neq k} \Delta_{kl}$. En remplaçant les termes diagonaux :

$$V = -\sum_k \sum_{l \neq k} \Delta_{kl} \, a_k^2 + \sum_k \sum_{l \neq k} \Delta_{kl} \, a_k \, a_l = -\frac{1}{2} \sum_k \sum_{l \neq k} \Delta_{kl} \, (a_k - a_l)^2$$

Deux ménages sur cinq

$\pi_k = 0{,}4$, $\pi_{kl} = 0{,}1$, donc $\pi_k\,\pi_l - \pi_{kl} = 0{,}16 - 0{,}1 = 0{,}06$. Les y_k/π_k valent 10 ; 15 ; $17{,}5$; $22{,}5$; 35. Chaque paire apparaît deux fois dans la double somme, d'où

$$V(\widehat{Y}_{HT}) = 0{,}06 \sum_{k<l} \left(\frac{y_k}{\pi_k} - \frac{y_l}{\pi_l} \right)^2 = 0{,}06 \times 1\,812{,}5 = 108{,}75$$

Trois calculs, un seul résultat

- Variance des dix totaux de l'énumération : $108{,}75$.
- Formule du chapitre 3 : $N^2\,(1-f)\,S^2/n = 25 \times 4{,}35 = 108{,}75$.
- Pour la moyenne, on divise par $N^2 = 25$: $V(\overline{y}) = 4{,}35$.

Deux groupes, des probabilités doubles nulles

- Même groupe : $\pi_{AB} = 0$, donc $\Delta_{AB} = -\frac{1}{9}$; $\pi_{DE} = 0$, donc $\Delta_{DE} = -\frac{1}{4}$.
- Groupes différents : $\pi_{AD} = \frac{1}{6} = \frac{1}{3} \times \frac{1}{2}$, donc $\Delta_{AD} = 0$.

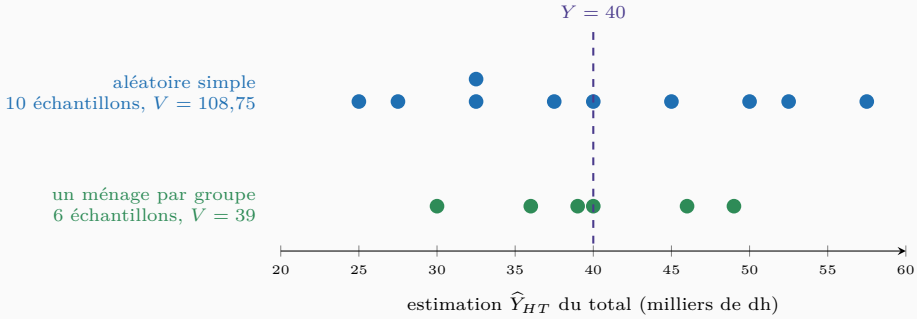
Les y_k/π_k valent 12 ; 18 ; 21 dans le premier groupe, 18 ; 28 dans le second :

$$V(\widehat{Y}_{HT}) = \tfrac{1}{9}\left[6^2 + 9^2 + 3^2\right] + \tfrac{1}{4} \times 10^2 = 14 + 25 = 39$$

L'énumération confirme

Les six totaux 30 ; 40 ; 36 ; 46 ; 39 ; 49 s'écartent de 40 de -10 ; 0 ; -4 ; 6 ; -1 ; 9. Somme des carrés : 234, et $234/6 = 39$.

Deux plans, deux précisions



Même taille, variance presque trois fois plus faible

Avec deux ménages dans les deux cas, la variance tombe de $108{,}75$ à 39. Le plan par groupes **interdit** les échantillons extrêmes, $\{A, B\}$ ou $\{D, E\}$: il associe toujours un petit budget à un grand. C'est le principe de la **stratification** (chapitre 6).

Estimer la variance

Deux estimateurs sans biais

$$\widehat{V}(\widehat{Y}_{HT}) = \sum_{k\in S}\sum_{l\in S}\frac{\Delta_{kl}}{\pi_{kl}}\frac{y_k}{\pi_k}\frac{y_l}{\pi_l} \qquad (\pi_{kk} = \pi_k)$$

et, pour un plan de taille fixe, la forme de Sen-Yates-Grundy :

$$\widehat{V}_{SYG} = \frac{1}{2}\sum_{k\in S}\sum_{l\in S,\, l\neq k}\frac{\pi_k\pi_l - \pi_{kl}}{\pi_{kl}}\left(\frac{y_k}{\pi_k} - \frac{y_l}{\pi_l}\right)^2$$

Le même argument que pour le total

Chaque terme de V concerne une **paire** (k, l) ; la paire est observée avec la probabilité $E(I_k I_l) = \pi_{kl}$. On divise donc chaque terme par π_{kl}, comme on divisait y_k par π_k.

Condition : $\pi_{kl} > 0$ pour **toutes** les paires.

Les dix estimations de la variance

Pour l'aléatoire simple, $(\pi_k\pi_l - \pi_{kl})/\pi_{kl} = 0{,}6$:
$\widehat{V} = 0{,}6 \times (2{,}5\,y_k - 2{,}5\,y_l)^2 = 3{,}75\,(y_k - y_l)^2$.

Échantillon	AB	AC	AD	AE	BC
$\widehat{V}$	15	33,75	93,75	375	3,75
Échantillon	**BD**	**BE**	**CD**	**CE**	**DE**
$\widehat{V}$	33,75	240	15	183,75	93,75

Moyenne des dix : $1\,087{,}5/10 = 108{,}75 = V(\widehat{Y}_{HT})$.

Sans biais, mais très dispersé

Les estimations vont de $3{,}75$ à 375 : avec deux unités, la variance est mal connue. Pour l'aléatoire simple, la formule se réduit à $N^2\,(1-f)\,s^2/n$; le chapitre 3 l'exploitera avec $n = 200$.

Le plan par groupes ne voit jamais deux ménages d'un même groupe

$\pi_{AB} = \pi_{AC} = \pi_{BC} = \pi_{DE} = 0$: aucun échantillon ne renseigne sur la dispersion **à l'intérieur** d'un groupe, celle qui fait la variance. Les deux estimateurs échouent : $\widehat{V}$ vaut 479 en moyenne, $\widehat{V}_{SYG}$ vaut toujours 0, contre $V = 39$.

Un problème qui reviendra

Le tirage systématique (chapitre 8) et les plans à une seule unité par strate ont des π_{kl} nuls : leur estimateur reste sans biais, mais sa variance ne peut qu'être **approchée**. Un bon plan permet de mesurer sa propre précision.

Biais, variance et erreur quadratique

Biais, variance, erreur quadratique moyenne

Pour un estimateur $\hat{\theta}$ d'un paramètre θ :

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta \qquad V(\hat{\theta}) = E\big[(\hat{\theta} - E\hat{\theta})^2\big]$$

$$EQM(\hat{\theta}) = E\big[(\hat{\theta} - \theta)^2\big] = V(\hat{\theta}) + B(\hat{\theta})^2$$

La démonstration

On écrit $\hat{\theta} - \theta = (\hat{\theta} - E\hat{\theta}) + B$, où B est un nombre fixe. En élevant au carré puis en prenant l'espérance, le double produit $2B\,E(\hat{\theta} - E\hat{\theta})$ est nul, car $E(\hat{\theta} - E\hat{\theta}) = 0$. Il reste $V + B^2$.



Lire les quatre cibles

Le centre est le paramètre ; chaque point est l'estimation d'un échantillon. Le **biais** est le décalage du nuage, la **variance** sa dispersion, l'**EQM** la distance moyenne au centre, au carré. Un estimateur biaisé mais précis peut tomber plus près, en moyenne, qu'un estimateur sans biais très dispersé.

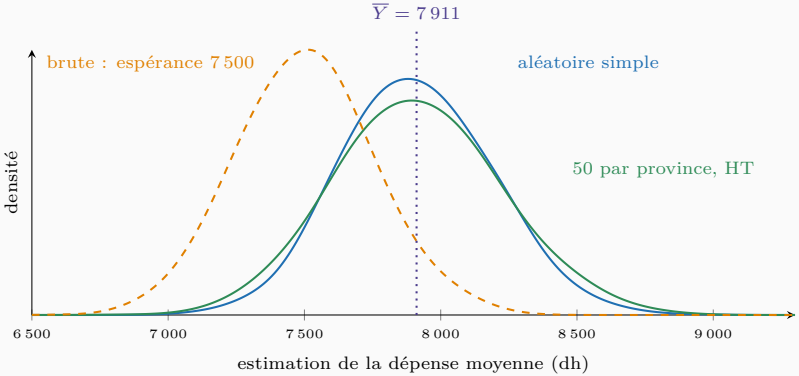
Trois estimateurs de la moyenne, n = 2

Estimateur	Espérance	Biais	Variance	EQM
Aléatoire simple, $\overline{y}$	8	0	$4{,}35$	$4{,}35$
Groupes, moyenne brute	$8{,}58$	$0{,}58$	$1{,}95$	$2{,}29$
Groupes, HT pondérée	8	0	$1{,}56$	$1{,}56$

$\overline{Y} = 8$. Brute : $E = 51{,}5/6$; $EQM = 1{,}95 + 0{,}58^2 = 2{,}29$. Pondérée : $V = 39/25 = 1{,}56$.

Deux leçons

Sous le plan par groupes, la pondération supprime le biais **et** réduit la variance. Et la moyenne brute, pourtant biaisée, fait mieux que l'aléatoire simple sans biais : $2{,}29$ contre $4{,}35$. Le biais n'est pas tout ; c'est l'EQM qui mesure l'erreur.



Quatre mille tirages de 200 ménages pour chaque estimateur

Les deux estimateurs de Horvitz-Thompson sont centrés sur $\overline{Y}$. La moyenne brute du plan à 50 ménages par province est la plus resserrée, mais centrée sur $7\,500$ dh.

Le biais ne diminue pas avec n

Valeurs exactes, en dh

Plan et estimateur	n	Biais	Écart-type	$\sqrt{EQM}$
Aléatoire simple	200	0	$267{,}3$	$267{,}3$
50 par province, HT	200	0	$298{,}4$	$298{,}4$
50 par province, brute	200	$-411{,}1$	$246{,}6$	$479{,}4$
200 par province, brute	800	$-411{,}1$	$98{,}4$	$422{,}7$

Plus de ménages ne corrige pas un biais

Quadrupler l'échantillon divise l'écart-type de la moyenne brute par $2{,}5$, mais laisse le biais intact : il représente $74\,\%$ de l'EQM avec 200 ménages, $95\,\%$ avec 800. Au-delà d'une certaine taille, **seul le biais compte** : c'est la leçon du Literary Digest.

L’intervalle de confiance

Intervalle de confiance à 95 pour cent

Quand n est grand, $\widehat{Y}_{HT}$ suit approximativement une loi normale (théorème central limite, chapitre 20 de Probabilités ; Statistique inférentielle, chapitres 8 à 10) :

$$\left[\, \widehat{Y}_{HT} - 1{,}96\,\sqrt{\widehat{V}}\ ;\ \widehat{Y}_{HT} + 1{,}96\,\sqrt{\widehat{V}} \,\right]$$

Pour la moyenne, on divise les bornes par N.

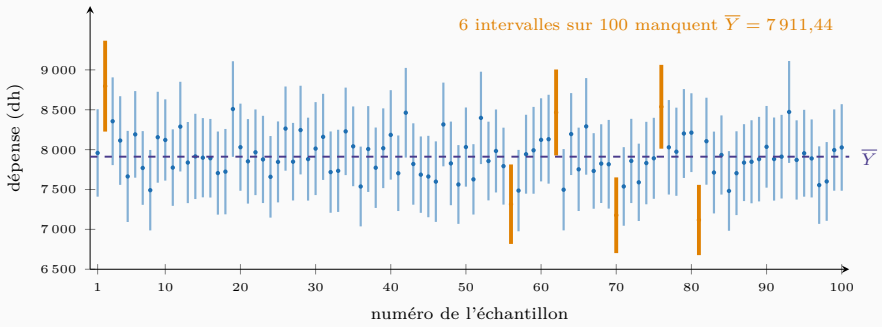
L'échantillon fil rouge : aléatoire simple de 200 ménages

$\overline{y} = 7\,843{,}05$ dh, $s = 3\,829{,}16$ dh, erreur-type estimée $\sqrt{(1-f)\,s^2/n} = 256{,}87$ dh :

$$\text{IC}_{95\,\%}(\overline{Y}) = 7\,843{,}05 \pm 1{,}96 \times 256{,}87 = [\,7\,339{,}6\ ;\ 8\,346{,}5\,]$$

Total : $\widehat{Y}_{HT} = 2\,000 \times 7\,843{,}05 = 15\,686\,100$ dh, intervalle $[\,14\,679\,178\ ;\ 16\,693\,022\,]$.

Cent échantillons, cent intervalles



L'intervalle bouge, le paramètre reste

Chaque segment est l'intervalle à $95\,\%$ d'un échantillon de 200 ménages. Six d'entre eux, en orange, manquent $\overline{Y}$: on en attendait environ cinq.

Ce que garantit le niveau de confiance

Avant le tirage, l'intervalle est **aléatoire** : la probabilité qu'il contienne $\overline{Y}$ vaut environ $0{,}95$. Le niveau de confiance qualifie la **méthode** : sur un grand nombre d'échantillons, $95\,\%$ des intervalles contiennent $\overline{Y}$.

Une phrase à ne pas écrire

« Il y a $95\,\%$ de chances que $\overline{Y}$ soit entre $7\,339{,}6$ et $8\,346{,}5$ » est **faux** : une fois l'échantillon tiré, l'intervalle et $\overline{Y}$ sont deux choses fixes. Ou bien il le contient, ou bien non. Ici il le contient ($7\,911{,}44$) ; l'enquêteur, lui, ne le saura jamais.

Cinq mille sondages aléatoires simples

Taille n	50	200	800
Couverture	$0{,}9420$	$0{,}9424$	$0{,}9512$

La **couverture** est la part des intervalles qui contiennent $\overline{Y}$. Avec $n = 200$: 288 échecs, dont 188 intervalles **trop bas** et 100 trop hauts. L'écart à $0{,}95$ vaut $2{,}5$ erreurs-types de simulation $(0{,}0031)$: il est réel.

Pourquoi un peu moins de 95

La dépense est **asymétrique** : quelques ménages dépensent beaucoup. Un échantillon qui en contient peu a une moyenne basse **et** un s faible : son intervalle est trop court précisément quand il est décalé. L'approximation normale s'améliore quand n grandit.

Une précision relative

$$CV(\hat{\theta}) = \frac{\sqrt{V(\hat{\theta})}}{\theta}, \qquad \text{estimé par } \frac{\sqrt{\widehat{V}}}{\hat{\theta}}$$

Sans unité, il est le même pour le total et pour la moyenne ($\widehat{Y}_{HT} = N\,\widehat{\overline{Y}}_{HT}$). La marge à $95\,\%$ vaut environ $2\,CV$ en valeur relative.

Dépense moyenne, aléatoire simple

Taille n	50	200	1 000
Erreur-type de $\overline{y}$	556,38	267,27	89,09
CV	7,0 %	3,4 %	1,1 %

Fil rouge : CV estimé $= 256{,}87/7\,843{,}05 = 3{,}3\,\%$, marge relative $\pm 6{,}4\,\%$.

S’entraîner

Trois unités, trois échantillons

Une population compte trois unités, de valeurs $y_1 = 10$, $y_2 = 21$ et $y_3 = 40$ ($Y = 71$). Un plan de taille 2 donne aux échantillons $\{1, 2\}$, $\{1, 3\}$ et $\{2, 3\}$ les probabilités $0{,}2$, $0{,}3$ et $0{,}5$.

1. Calculer les π_k et vérifier que leur somme vaut n.
2. Calculer $\widehat{Y}_{HT}$ pour chaque échantillon, puis son espérance.
3. Calculer $V(\widehat{Y}_{HT})$.
4. On propose plutôt $N\,\overline{y} = 3\,\overline{y}$. Calculer son biais, sa variance et son EQM. Conclure.

Exercice 1 : solution

Horvitz-Thompson

1. $\pi_1 = 0{,}2 + 0{,}3 = 0{,}5$; $\pi_2 = 0{,}2 + 0{,}5 = 0{,}7$; $\pi_3 = 0{,}3 + 0{,}5 = 0{,}8$; somme $2 = n$.
2. y_k/π_k : 20 ; 30 ; 50. D'où $\widehat{Y}_{HT} = 50$; 70 ; 80 et $E = 0{,}2 \times 50 + 0{,}3 \times 70 + 0{,}5 \times 80 = 71 = Y$.
3. $E(\widehat{Y}_{HT}^2) = 0{,}2 \times 2\,500 + 0{,}3 \times 4\,900 + 0{,}5 \times 6\,400 = 5\,170$; $V = 5\,170 - 71^2 = 129$.

La moyenne brute extrapolée

4. $3\,\overline{y}$ vaut $46{,}5$; 75 ; $91{,}5$. $E = 77{,}55$, biais $+6{,}55$: l'unité 3, la plus grande, est tirée quatre fois sur cinq. $V = 6\,306{,}075 - 77{,}55^2 = 292{,}07$ et $EQM = 292{,}07 + 6{,}55^2 = 334{,}98$, contre 129 : Horvitz-Thompson l'emporte sur les deux tableaux.

Les quatre boutiques du chapitre 1

Les chiffres d'affaires sont 2, 5, 8 et 9 milliers de dh ($S^2 = 10$) ; on en tire deux par sondage aléatoire simple.

1. Calculer π_k, π_{kl}, $\pi_k \pi_l - \pi_{kl}$ et les y_k / π_k.
2. Calculer $V(\widehat{Y}_{HT})$ par la forme de Sen-Yates-Grundy ; comparer à $N^2 (1 - f)\, S^2 / n$.
3. Calculer $\widehat{V}_{SYG}$ pour chacun des six échantillons. Vérifier qu'il est sans biais.

La variance

1. $\pi_k = 0{,}5$, $\pi_{kl} = 1/6$, $\pi_k\pi_l - \pi_{kl} = \frac{1}{4} - \frac{1}{6} = \frac{1}{12}$; $y_k/\pi_k : 4 ; 10 ; 16 ; 18$.
2. Carrés des écarts sur les six paires : $36 ; 144 ; 196 ; 36 ; 64 ; 4$, de somme 480.
   $V = \frac{1}{12} \times 480 = 40$, et $16 \times 0{,}5 \times 10/2 = 40$.

Son estimation

3. Une seule paire par échantillon, et $\frac{1/12}{1/6} = 0{,}5$: $\widehat{V}_{SYG} = 0{,}5\,(a_k - a_l)^2$. Pour $\{2,5\}$, $\{2,8\}$, $\{2,9\}$, $\{5,8\}$, $\{5,9\}$, $\{8,9\}$: $18 ; 72 ; 98 ; 18 ; 32 ; 2$. Moyenne : $240/6 = 40 = V$.

Intervalle, coefficient de variation et EQM

Un sondage aléatoire simple de 400 ménages de la population du cours donne $\overline{y} = 7\,952$ dh et $s = 4\,074$ dh.

1. Calculer l'erreur-type estimée $\sqrt{(1-f)\,s^2/n}$, l'intervalle de confiance à $95\,\%$ et le CV.
2. Vrai ou faux : « il y a $95\,\%$ de chances que $\overline{Y}$ soit dans cet intervalle » ?
3. Un plan moins coûteux donne un estimateur de biais 200 dh et d'écart-type 120 dh. Comparer son EQM à celle du sondage ($V = 31\,749{,}1$).

Précision et lecture

1. $f = 0{,}2$; $\sqrt{0{,}8 \times 4\,074^2/400} = 182{,}19$ dh ; IC : $7\,952 \pm 357{,}1 = [\,7\,594{,}9\ ;\ 8\,309{,}1\,]$; $CV = 182{,}19/7\,952 = 2{,}3\,\%$.
2. Faux : l'intervalle tiré contient $\overline{Y}$ ou non (ici oui : $7\,911{,}44$). Les $95\,\%$ décrivent la méthode, sur l'ensemble des échantillons possibles.

Le biais, un plancher

3. $EQM = 120^2 + 200^2 = 54\,400$ (racine $233{,}2$ dh), contre $31\,749{,}1$ (racine $178{,}2$ dh). Même avec un écart-type réduit à 60 dh, $EQM = 43\,600$: la racine de l'EQM ne descend jamais sous le biais, 200 dh.

Ce qu’il faut retenir

Le résumé du chapitre

Sept points

1. L'estimateur de **Horvitz-Thompson** $\widehat{Y}_{HT} = \sum_{k \in S} y_k / \pi_k$ compte chaque unité tirée pour son poids d_k.
2. Il est **sans biais** pour tout plan, pourvu que $\pi_k > 0$ pour toute unité.
3. $V(\widehat{Y}_{HT}) = \sum_k \sum_l \Delta_{kl} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l}$; à taille fixe, elle est nulle si y_k / π_k est constant.
4. La variance s'estime sur l'échantillon si $\pi_{kl} > 0$ pour toutes les paires.
5. $EQM = V + B^2$: la variance baisse avec n, **le biais ne baisse pas**.
6. L'intervalle $\hat{\theta} \pm 1{,}96\sqrt{\widehat{V}}$ couvre le paramètre dans environ $95\,\%$ des échantillons ($0{,}9424$ simulé pour $n = 200$).
7. Le $CV = \sqrt{V}/\theta$ mesure la précision relative.

La suite

Le chapitre 3 applique ces résultats au sondage aléatoire simple : il démontre $V(\overline{y}) = (1 - f)\, S^2/n$ et montre comment l'estimer.