Échantillonnage

Chapitre 5 : Choisir la taille de l'échantillon

Jaouad Madkour

30 octobre 2026

Plan du chapitre

Partir de la précision voulue **2**
La question à poser avant l'enquête . . . 2
Inverser la variance du sondage aléatoire simple . . . 2
Sur la mini-population . . . 3
Le niveau de confiance . . . 3

La précision d'une moyenne **4**
Trois marges sur la population du cours . . . 4
La taille selon la marge . . . 4
Ce que coûte la précision . . . 5
Une marge en pourcentage de la moyenne . . . 5
Pourquoi le coefficient de variation voyage bien . . . 5

La précision d'une proportion **6**
La formule pour une proportion . . . 6
La moitié, par prudence . . . 7
Une proportion rare . . . 7

Ce que l'on ignore avant l'enquête **8**
D'où vient l'écart-type S . . . 8
Une enquête pilote de trente ménages . . . 8
Se prémunir contre un pilote trop optimiste . . . 8
La non-réponse . . . 9

Plusieurs objectifs, un budget **9**
Plusieurs variables, une seule taille . . . 9
Une précision par province . . . 10
Ce que coûtent les domaines . . . 10
L'effet de plan comme multiplicateur . . . 10
Le budget et la démarche complète . . . 11

S'entraîner **12**
Exercice 1 : les loyers des étudiants . . . 12
Exercice 1 : solution . . . 12

Exercice 2 : une proportion dans une commune 12
Exercice 2 : solution 12
Exercice 3 : urbain et rural 13
Exercice 3 : solution 13

Ce qu'il faut retenir **13**
Le résumé du chapitre 13

Partir de la précision voulue

La question à poser avant l'enquête

La marge d'erreur

Pour un sondage aléatoire simple, l'intervalle de confiance à 95 % de $\overline{Y}$ s'écrit $\overline{y} \pm e$, avec la **marge d'erreur**

$$e = z\sqrt{V(\overline{y})} = z\sqrt{\left(1-\frac{n}{N}\right)\frac{S^2}{n}}, \qquad z = 1{,}96$$

Retourner la question

Le chapitre 3 partait de n pour trouver la marge. Avant une enquête, on fait l'inverse : on **fixe la marge** que le commanditaire accepte, et l'on cherche le plus petit n qui la garantit. Le chapitre 11 de Statistique inférentielle a donné la règle de base ; ce chapitre l'adapte aux sondages : population finie, proportions, non-réponse, domaines, effet de plan.

Inverser la variance du sondage aléatoire simple

Taille pour une marge donnée sur une moyenne

$$n_0 = \left(\frac{z\,S}{e}\right)^2 \qquad\qquad n = \frac{n_0}{1+n_0/N}$$

n_0 est la taille qu'exigerait une population infinie ; n tient compte de la population finie. On **arrondit à l'entier supérieur**, à la fin des calculs seulement.

La démonstration

Comme $(1-n/N)\,S^2/n = S^2/n - S^2/N$, on cherche n tel que $S^2/n - S^2/N = e^2/z^2$. En divisant par S^2 :

$$\frac{1}{n} = \frac{e^2}{z^2 S^2} + \frac{1}{N} = \frac{1}{n_0} + \frac{1}{N} \quad\Longrightarrow\quad n = \frac{n_0\,N}{N+n_0} = \frac{n_0}{1+n_0/N}$$

Sur la mini-population

Cinq ménages, une variance au plus égale à 2

A à E : 4, 6, 7, 9, 14 milliers de dh ; $S^2 = 14{,}5$ et $N = 5$. On veut $V(\overline{y}) \leq 2$: c'est la même équation avec $(e/z)^2 = 2$.

$$n_0 = \frac{S^2}{2} = 7{,}25 \qquad n = \frac{7{,}25}{1 + 7{,}25/5} = 2{,}96 \quad \Longrightarrow \quad n = 3$$

Vérification : $V(\overline{y}) = (1 - \frac{3}{5}) \times 14{,}5/3 = 1{,}93 \leq 2$, alors qu'avec $n = 2$ on avait $V(\overline{y}) = 4{,}35$.

La correction ne dépasse jamais N

Ici $n_0 = 7{,}25$ dépasse $N = 5$: une population infinie exigerait plus d'unités qu'il n'en existe ici. La correction ramène toujours $n = n_0 N/(N + n_0)$ **en dessous de** N : au pire, on recense.

Démonstration : La taille minimale, l'arrondi et les bornes de la formule

La formule. La marge $e(n) = z\,S\sqrt{\frac{1}{n} - \frac{1}{N}}$ **diminue** quand n augmente ($\frac{1}{n}$ décroît). Exiger $e(n) \leq e$ revient donc à exiger $n \geq n^*$, où n^* est la valeur qui donne l'égalité : $\frac{1}{n^*} = \frac{1}{n_0} + \frac{1}{N}$. La **plus petite taille entière** convenable est $\lceil n^* \rceil$: l'arrondi se fait vers le haut. Si l'on arrondit vers le bas, $n < n^*$ et la marge dépasse e.

Application. Pour $e = 500$ dh, $n^* = 217{,}4$. Avec 217 ménages la marge est $1{,}96 \times 3\,984{,}29\sqrt{1/217 - 1/2\,000} = 500{,}5$ dh : trop grande. Avec 218, elle est de 499,2 dh.

Les bornes. De $\frac{1}{n^*} = \frac{1}{n_0} + \frac{1}{N}$, on tire $\frac{1}{n^*} > \frac{1}{n_0}$ et $\frac{1}{n^*} > \frac{1}{N}$, donc

$$n^* < n_0 \qquad \text{et} \qquad n^* < N.$$

La correction réduit toujours la taille, et ne dépasse jamais N ; quand $n_0 = N$, $\frac{1}{n^*} = \frac{2}{N}$ donc $n^* = \frac{N}{2}$, et $n^* \to N$ quand $n_0 \to +\infty$.

Ce que cela dit. Une marge très étroite n'exige pas plus que de recenser, quelle que soit sa valeur : la taille tend vers N sans jamais le dépasser. Pour la population du cours, n_0 dépasse N dès que $e < \dfrac{1{,}96\,S}{\sqrt{N}} = 174{,}6$ dh, et $n^* = 1\,000$ exactement à ce seuil.

■

Le niveau de confiance

Trois valeurs de z

Niveau de confiance	90 %	95 %	99 %
z	1,645	1,96	2,576
n_0 pour $e = 500$ dh	171,8	243,9	421,4
n pour $N = 2\,000$, arrondi	159	218	349

Chaque ingrédient compte au carré

n_0 est proportionnel à z^2, à S^2 et à $1/e^2$. Passer de 95 % à 99 % multiplie n_0 par $(2{,}576/1{,}96)^2 = 1{,}73$. Le niveau de 95 % est la convention des enquêtes : c'est la marge e qui se discute avec le commanditaire.

La précision d'une moyenne

Trois marges sur la population du cours

La dépense moyenne des 2 000 ménages

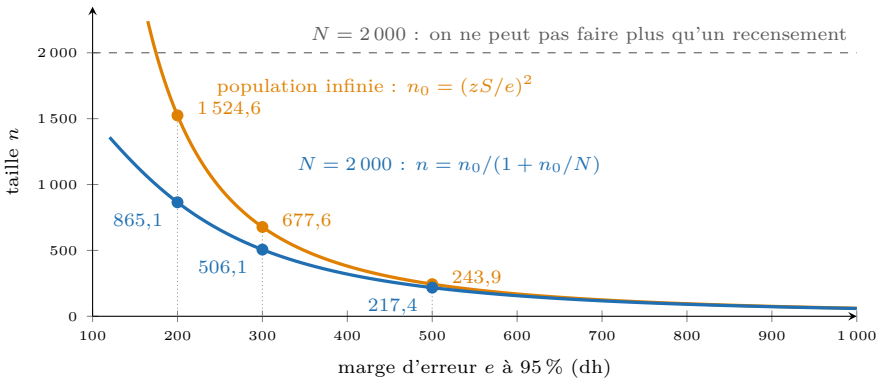
Avec $S = 3\,984{,}29$ dh et $N = 2\,000$:

Marge e	$n_0 = (1{,}96\,S/e)^2$	n	Arrondi	$f = n/N$
±500 dh	243,9	217,4	218	10,9 %
±300 dh	677,6	506,1	507	25,4 %
±200 dh	1 524,6	865,1	866	43,3 %

Pourquoi arrondir au-dessus

Avec 217 ménages, la marge vaut $1{,}96 \times 3\,984{,}29 \times \sqrt{1/217 - 1/2\,000} = 500{,}5$ dh : un peu trop. Avec 218, elle vaut 499,2 dh. Arrondir au plus proche ferait manquer l'objectif.

La taille selon la marge



Deux courbes

Sans correction (orange), n_0 explose quand la marge se resserre et dépasse N dès que $e < 174{,}6$ dh. Avec la correction (bleu), n reste sous $N = 2\,000$; elle compte dès que l'échantillon est une part notable de la population : à $e = 174{,}6$ dh, $n_0 = N$ et $n = N/2$.

Ce que coûte la précision

Diviser la marge par 2,5 coûte quatre fois plus

De ± 500 à ± 200 dh, il faut passer de 218 à 866 ménages : 3,97 fois plus de questionnaires. Dans une population infinie, ce serait $2{,}5^2 = 6{,}25$ fois plus. Le budget d'une enquête croît comme le **carré** de la précision demandée.

Un sondage de 200 ménages

Un sondage aléatoire simple de 200 ménages, comme l'échantillon fil rouge du chapitre 3, a une marge théorique de $\pm 523{,}9$ dh. Pour descendre à ± 500 dh, il aurait fallu 18 ménages de plus ; pour ± 300 dh, 307 de plus.

Une marge en pourcentage de la moyenne

Précision relative

On demande $e = r\,\overline{Y}$, par exemple une marge de $r = 5\,\%$ de la moyenne. Alors

$$n_0 = \left(\frac{z\,S}{r\,\overline{Y}}\right)^2 = \left(\frac{z\,CV}{r}\right)^2, \qquad CV = \frac{S}{\overline{Y}}$$

Il suffit de connaître le **coefficient de variation**, ni S ni $\overline{Y}$ séparément.

Avec un coefficient de variation de 0,5036

- $r = 5\,\%$, soit $\pm 395{,}6$ dh : $n_0 = (1{,}96 \times 0{,}5036/0{,}05)^2 = 389{,}7$; $n = 326{,}2$, soit 327 ménages.
- $r = 3\,\%$, soit $\pm 237{,}3$ dh : $n_0 = 1\,082{,}6$; $n = 702{,}4$, soit 703 ménages.

Pourquoi le coefficient de variation voyage bien

Un nombre sans unité

Si toutes les dépenses augmentent de 10 %, S passe à 4 382,7 dh et la taille pour ± 500 dh passe de 217,4 à 257,2. Le CV, lui, ne bouge pas : la taille pour une marge de $\pm 5\,\%$ reste 326,2.

Démonstration : Pourquoi le coefficient de variation voyage bien

Multiplions toutes les valeurs par un facteur $c > 0$ (un changement d'unité, ou une hausse de 10 % pour $c = 1{,}1$) : $y_k \to c\,y_k$. Alors $\overline{Y} \to c\,\overline{Y}$ et $S \to c\,S$, car $\sum (c\,y_k - c\,\overline{Y})^2 = c^2 \sum (y_k - \overline{Y})^2$. Donc

$$CV = \frac{S}{\overline{Y}} \to \frac{c\,S}{c\,\overline{Y}} = CV.$$

Marge absolue. $n_0 = (zS/e)^2$ est multiplié par c^2 : pour la même marge de ± 500 dh, $n_0 = 243{,}9$ devient $243{,}9 \times 1{,}21 = 295{,}2$, et $n = 295{,}2/(1 + 295{,}2/2\,000) = 257{,}2$ (contre 217,4).

Marge relative. On demande $e = r\,\overline{Y}$, qui est aussi multipliée par c : $n_0 = \left(\frac{z\,cS}{r\,c\,\overline{Y}}\right)^2 = (z\,CV/r)^2$ ne change pas. Pour $r = 5\,\%$: $n_0 = 389{,}7$ et $n = 326{,}2$, avant comme après la

hausse.

Ce que cela dit. Une taille d'échantillon calculée avec un écart-type est liée à l'unité de mesure et à l'année ; calculée avec un coefficient de variation et une précision relative, elle se transpose.

■

Transposer une enquête à une autre

Un *CV* se transpose d'une année, d'une région ou d'une monnaie à l'autre bien mieux qu'un écart-type. C'est pourquoi l'on raisonne volontiers en précision relative pour les montants : dépenses, revenus, chiffres d'affaires, rendements agricoles.

La précision d'une proportion

La formule pour une proportion

Taille pour une marge donnée sur une proportion

$$n_0 = \frac{z^2\,P\,(1-P)}{e^2} \qquad\qquad n = \frac{n_0}{1+(n_0-1)/N}$$

La marge e s'exprime en **points** : $e = 0{,}05$ pour ± 5 points.

La démonstration

Pour une proportion, $S^2 = \frac{N}{N-1}\,P\,(1-P)$ (chapitre 1). L'équation $(1/n - 1/N)\,S^2 = e^2/z^2$ devient

$$\frac{N-n}{n\,(N-1)} = \frac{1}{n_0} \quad\Longrightarrow\quad n\,(N-1+n_0) = n_0\,N \quad\Longrightarrow\quad n = \frac{n_0\,N}{N-1+n_0}$$

La formule de la moyenne, $n_0/(1+n_0/N)$, ne s'en écarte que d'une fraction d'unité.

Démonstration : La taille pour une proportion, formule exacte et formule approchée

Pour une proportion, la variance du sondage aléatoire simple est $V(\widehat{P}) = \dfrac{N-n}{N-1}\cdot\dfrac{P(1-P)}{n}$ (chapitre 4). Imposer $z^2V(\widehat{P}) = e^2$ donne, avec $n_0 = z^2P(1-P)/e^2$:

$$\frac{N-n}{n\,(N-1)} = \frac{e^2}{z^2P(1-P)} = \frac{1}{n_0} \;\Longrightarrow\; n_0\,(N-n) = n\,(N-1)$$

$$\Longrightarrow\; n = \frac{n_0\,N}{N-1+n_0} = \frac{n_0}{1+(n_0-1)/N}.$$

L'écart avec la formule de la moyenne. La formule approchée $n_a = \frac{n_0 N}{N+n_0}$ est inférieure, et

$$n - n_a = \frac{n_0\,N}{N-1+n_0} - \frac{n_0\,N}{N+n_0} = \frac{n_0\,N}{(N-1+n_0)(N+n_0)} = \frac{n_a}{N-1+n_0}.$$

Comme $n_a < N$ et $n_0 \geq 1$, on a $N - 1 + n_0 \geq N$ donc $0 < n - n_a < 1$: les deux formules diffèrent de **moins d'une unité**, et après arrondi au-dessus, elles donnent presque toujours le même entier.

Le pire des cas. $P(1-P) \leq \frac{1}{4}$ avec égalité en $P = \frac{1}{2}$ (chapitre 1) : prendre $P = \frac{1}{2}$ donne un n_0 au moins aussi grand que celui de la vraie proportion, donc une marge tenue **quelle que soit** P. Pour $e = 0{,}05$: $n_0 = \frac{1{,}96^2 \times 0{,}25}{0{,}05^2} = 384{,}2$, $n = 322{,}4$ pour $N = 2\,000$, contre 322,3 par la formule approchée.

Ce que cela dit. Pour une proportion, la bonne formule contient $N - 1$, parce que la variance d'une proportion est $\frac{N}{N-1}P(1 - P)$; mais l'écart est une fraction d'unité, sans conséquence pratique.

■

La moitié, par prudence

Les ménages connectés de la population du cours

Marge	n_0, $P = 0{,}5$	n	Arrondi	Si $P = 0{,}674$
± 5 points	384,2	322,4	323	289
± 3 points	1 067,1	696,1	697	639

Avec la formule approchée $n_0/(1 + n_0/N)$: 322,3 et 695,8.

Le pire des cas

$P\,(1-P)$ vaut au plus 0,25, pour $P = 0{,}5$: en le prenant, la marge est tenue **quelle que soit** la vraie proportion. Savoir que $P \approx 0{,}674$ donne $P\,(1 - P) = 0{,}2197$: n_0 ne baisse que de 12 %, et il faut encore 289 ménages au lieu de 323.

Une proportion rare

Trois points sur une proportion de cinq pour cent

Pour une proportion voisine de 5 %, une marge de ± 3 points ($n_0 = 202{,}8$) laisse un intervalle de 2 % à 8 % : du simple au quadruple. La marge absolue n'a plus de sens.

Raisonner en précision relative

Exiger ± 20 % de P, soit ± 1 point :

$$n_0 = \frac{1{,}96^2 \times 0{,}05 \times 0{,}95}{0{,}01^2} = 1\,824{,}8$$

et $n = 954{,}4$ dans la population du cours. Les phénomènes rares (pauvreté extrême, entreprises exportatrices, maladies) exigent de **grands** échantillons.

Ce que l'on ignore avant l'enquête

D'où vient l'écart-type S

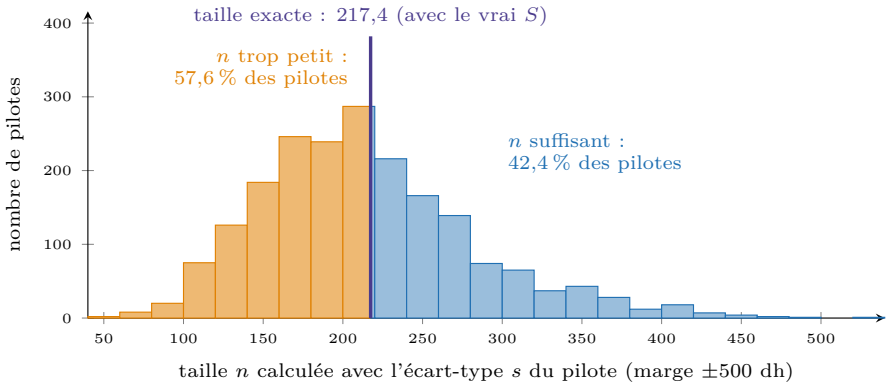
Trois sources

1. Une **enquête antérieure** sur la même population, par exemple la vague précédente d'une enquête répétée.
2. Une **enquête pilote** : quelques dizaines d'unités interrogées pour tester le questionnaire, et mesurer la dispersion.
3. Un **coefficient de variation** connu d'enquêtes semblables, qui suffit pour une précision relative.

Un cercle, et comment en sortir

Il faut connaître S pour dimensionner l'enquête qui va mesurer la population. Un **ordre de grandeur** suffit, mais il compte au carré : une erreur de 10 % sur S change n_0 de 21 %, car $1{,}1^2 = 1{,}21$.

Une enquête pilote de trente ménages



Deux mille pilotes simulés

Chaque pilote tire 30 ménages, calcule leur écart-type s, puis la taille n pour ±500 dh. Les tailles conseillées vont de 54 à 525 ; dans 57,6 % des cas, elles sont **inférieures** à la taille exacte 217,4.

Se prémunir contre un pilote trop optimiste

Pourquoi le pilote se trompe plutôt par défaut

Les dépenses ont une longue queue à droite : un petit échantillon manque souvent les gros budgets, et sous-estime S. L'écart-type médian des pilotes vaut 3 875 dh, contre $S = 3\,984$ dh.

Majorer l'écart-type par prudence

Écart-type utilisé	s	$1{,}1\,s$	$1{,}2\,s$	$1{,}3\,s$
Taille trop petite (%)	57,6	35,2	18,8	9,1
Taille médiane conseillée	207	245	285	326

Première ligne : part des 2 000 pilotes qui conseillent moins de 217,4 ménages. La sécurité se paie : majorer S de 20 % multiplie n_0 par 1,44.

La non-réponse

Tirer plus pour obtenir assez

Si l'on attend un taux de réponse τ, on tire n/τ unités pour obtenir environ n réponses.

Avec le taux de réponse de la population du cours

$\tau = 0{,}8085$. Pour ± 5 points : $322{,}4/0{,}8085 = 398{,}8$, soit 399 ménages à tirer. Pour ± 3 points : $696{,}1/0{,}8085 = 860{,}9$, soit 861.

La taille, pas le biais

Diviser par τ assure le **nombre** de réponses, pas leur représentativité. Les répondants de la population du cours dépensent en moyenne 7 692,00 dh, contre $\overline{Y} = 7\,911{,}44$: un biais de $-219{,}43$ dh que la taille ne réduit pas (chapitre 14).

Plusieurs objectifs, un budget

Plusieurs variables, une seule taille

Deux objectifs pour une même enquête

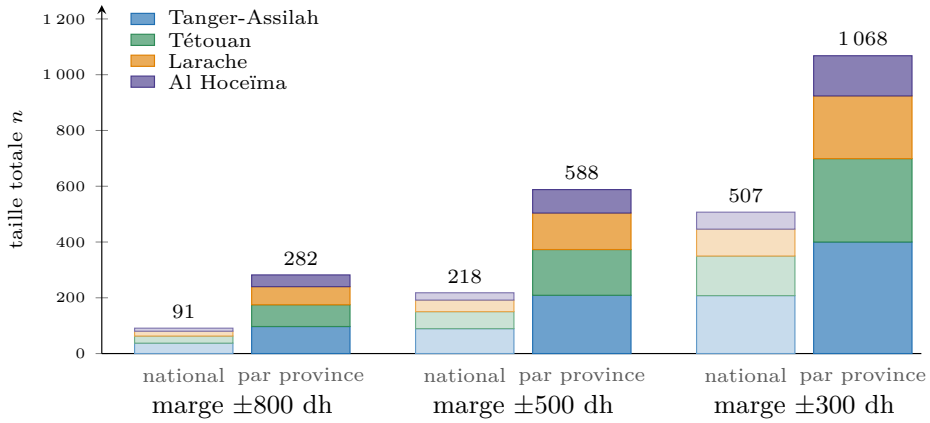
- Dépense moyenne à ± 500 dh : 218 ménages.
- Part des ménages connectés à ± 5 points : 323 ménages.

On retient 323, la plus grande des deux. La dépense est alors connue à $\pm 397{,}9$ dh, mieux que demandé.

L'objectif le plus exigeant commande

Une enquête porte sur des dizaines de variables. On la dimensionne sur **quelques indicateurs clés**, et l'on retient la plus grande des tailles obtenues ; les autres variables profitent de cette précision supplémentaire.

Une précision par province



Une marge dans chaque province

Barres pâles : la taille nationale, répartie au prorata des provinces. Barres foncées : la taille qui donne la **même marge dans chaque province**. Pour ±500 dh, il faut 588 ménages au lieu de 218.

Ce que coûtent les domaines

Une marge de 500 dh dans chaque province

Province	N_h	S_h	n_0	n_h	f_h
Tanger-Assilah	820	4 271	280,3	209	25,5 %
Tétouan	560	3 873	230,5	164	29,3 %
Larache	380	3 593	198,4	131	34,5 %
Al Hoceïma	240	2 900	129,2	84	35,0 %
Ensemble	2 000			588	29,4 %

Chaque domaine coûte presque une enquête

La taille dépend peu de N_h : Tanger-Assilah seule demande 209 ménages, presque autant que toute la région (218). Avec 218 ménages répartis au prorata, Al Hoceïma n'en aurait que 26, et une marge de ±1 053 dh. L'échantillon n'est plus proportionnel : il faudra des poids (chapitre 1) et une allocation réfléchie (chapitre 7).

L'effet de plan comme multiplicateur

Corriger la taille selon le plan

Si un plan a une variance *deff* fois celle d'un sondage aléatoire simple de même taille (l'**effet de plan**, chapitre 7), on multiplie n_0 par *deff*, puis on corrige pour la population finie :

$$n_0' = \mathit{deff} \times n_0 \qquad\qquad n = \frac{n_0'}{1 + n_0'/N}$$

Deux plans à venir, pour une marge de 500 dh

- **Stratifié** par province, *deff* $=$ 0,954 (chapitre 6) : $n'_0 = 243{,}9 \times 0{,}954 = 232{,}7$; $n = 208{,}5$, soit 209 ménages au lieu de 218.
- **Par grappes** de 40 ménages, $deff \approx 1 + 39 \times 0{,}162 = 7{,}32$ (chapitre 9) : $n'_0 = 1\,785{,}1$; $n = 943{,}2$, soit 24 districts entiers.

Démonstration : L'effet de plan multiplie la taille

Supposons que la variance du plan soit, pour tout échantillon de taille n de ce plan, égale à *deff* fois celle du sondage aléatoire simple de même taille : $V_{\text{plan}}(n) = deff \times \left(\frac{1}{n} - \frac{1}{N}\right) S^2$ (hypothèse de travail : *deff* varie peu avec n). La marge $e = z\sqrt{V_{\text{plan}}(n)}$ exige

$$deff\left(\frac{1}{n} - \frac{1}{N}\right) S^2 = \frac{e^2}{z^2} \implies \frac{1}{n} - \frac{1}{N} = \frac{e^2}{z^2 S^2\, deff} = \frac{1}{deff \times n_0} = \frac{1}{n'_0}.$$

Donc $n'_0 = deff \times n_0$ joue le rôle de n_0, et $\frac{1}{n} = \frac{1}{n'_0} + \frac{1}{N}$ donne $n = \dfrac{n'_0}{1 + n'_0/N}$.

Pour les deux plans du texte, marge $\pm$500 *dh* ($n_0 = 243{,}9$). Stratifié, *deff* $= 0{,}954$: $n'_0 = 232{,}7$ et $n = 232{,}7/(1 + 232{,}7/2\,000) = 208{,}5$, soit 209 ménages. Par grappes de 40, *deff* $= 1 + 39 \times 0{,}162 = 7{,}318$: $n'_0 = 1\,785{,}1$ et $n = 1\,785{,}1/(1 + 1\,785{,}1/2\,000) = 943{,}2$, soit $943{,}2/40 \approx 24$ districts.

Ce que cela dit. Un effet de plan inférieur à 1 économise des questionnaires, supérieur à 1 en coûte : la correction de population finie intervient **après** le multiplicateur, ce qui explique que n augmente moins vite que *deff* (7,3 fois plus pour n_0, 4,3 fois pour n, par rapport à 217,4).

■

Le budget et la démarche complète

Cinq étapes

1. Fixer l'indicateur clé, la marge e et le niveau de confiance.
2. Trouver S, CV ou P ; calculer n_0.
3. Multiplier par l'effet de plan ; corriger pour la population finie.
4. Diviser par le taux de réponse attendu ; arrondir au-dessus.
5. Chiffrer le coût $C = c_0 + c\,n$; si le budget ne suffit pas, revoir la marge.

Un exemple chiffré, avec des coûts supposés

Dépense à $\pm$300 dh, plan stratifié (*deff* $= 0{,}954$), $\tau = 0{,}8085$: $677{,}6 \times 0{,}954 = 646{,}4$; $n = 488{,}5$; $488{,}5/0{,}8085 = 604{,}2$, soit 605 questionnaires. À 20 000 dh de frais fixes et 150 dh par questionnaire : 110 750 dh. Avec 80 000 dh, on n'envoie que 400 questionnaires, pour une marge d'environ $\pm$388 dh.

S'entraîner

Exercice 1 : les loyers des étudiants

Une enquête dans une ville universitaire

Une association veut estimer le loyer mensuel moyen des 4 000 étudiants locataires d'une ville universitaire. Une enquête de l'an dernier donne un écart-type d'environ 600 dh.

1. Quelle taille faut-il pour une marge de ± 50 dh à 95 % ?
2. Et pour ± 25 dh ? Pourquoi la taille ne quadruple-t-elle pas ?
3. Quelle taille faut-il pour ± 50 dh à 99 % ?

Exercice 1 : solution

Marge absolue

1. $n_0 = (1{,}96 \times 600/50)^2 = 553{,}2$; $n = 553{,}2/(1+553{,}2/4\,000) = 485{,}98$, soit 486 étudiants.
2. $n_0 = 2\,212{,}8$, quatre fois plus ; mais $n = 1\,424{,}7$, soit 1 425 : 2,9 fois plus seulement. L'échantillon devient une part notable (36 %) de la population, et la correction joue davantage.

Niveau de confiance

3. $n_0 = (2{,}576 \times 600/50)^2 = 955{,}6$; $n = 771{,}3$, soit 772 étudiants : 59 % de plus qu'à 95 %.

Exercice 2 : une proportion dans une commune

Les chauffe-eau solaires

Une commune de 3 000 ménages veut connaître la part des ménages équipés d'un chauffe-eau solaire, à ± 4 points près, à 95 %.

1. Quelle taille retenir si l'on ne sait rien de cette part ?
2. Une enquête voisine indique une part proche de 20 %. Quelle taille suffit alors ?
3. On attend 75 % de réponses. Combien de ménages faut-il tirer dans le cas de la question 1 ?

Exercice 2 : solution

Le pire des cas, puis l'information

1. $P = 0{,}5$ par prudence : $n_0 = 1{,}96^2 \times 0{,}25/0{,}04^2 = 600{,}25$; $n = \dfrac{600{,}25}{1 + 599{,}25/3\,000} = 500{,}3$, soit 501 ménages.
2. $P(1-P) = 0{,}2 \times 0{,}8 = 0{,}16$: $n_0 = 384{,}16$; $n = 340{,}6$, soit 341 ménages, un tiers de moins.

La non-réponse

3. $500{,}3/0{,}75 = 667{,}1$: il faut tirer 668 ménages pour obtenir environ 501 réponses. Si les non-répondants diffèrent des répondants, ce surcroît ne corrige pas le biais.

Exercice 3 : urbain et rural

Une marge dans chaque milieu

Dans la population du cours, les 1 438 ménages urbains ont un écart-type de dépense de 4 134 dh, les 562 ménages ruraux de 3 215 dh. On veut une marge de ± 600 dh à 95 %.

1. Quelle taille faut-il pour la seule moyenne d'ensemble ($S = 3\,984{,}29$ dh) ?
2. Quelle taille faut-il dans chaque milieu pour y obtenir cette marge ? Quel est le total ?
3. Si l'on répartit la taille de la question 1 au prorata des milieux, quelle est la marge pour les ruraux ?

Exercice 3 : solution

Ensemble et milieux

1. $n_0 = (1{,}96 \times 3\,984{,}29/600)^2 = 169{,}4$; $n = 156{,}2$, soit 157 ménages.
2. Urbain : $n_0 = 182{,}4$, $n = 161{,}9$, soit 162. Rural : $n_0 = 110{,}3$, $n = 92{,}2$, soit 93. Total : 255 ménages, 1,62 fois plus.

Le prorata ne suffit pas

3. Les ruraux recevraient $157 \times 562/2\,000 = 44$ ménages, d'où une marge de $1{,}96 \times 3\,215 \times \sqrt{1/44 - 1/562} = 912$ dh : une fois et demie l'objectif. Pour publier par milieu, il faut **suréchantillonner** le rural (chapitre 7).

Ce qu'il faut retenir

Le résumé du chapitre

Sept points

1. On part de la **marge** e voulue et l'on inverse la variance : $n_0 = (zS/e)^2$.
2. La population finie réduit la taille : $n = n_0/(1 + n_0/N) < N$; on arrondit au-dessus.
3. La taille croît comme le **carré** de la précision : diviser e par deux multiplie n_0 par quatre.
4. En précision relative, seul compte le coefficient de variation : $n_0 = (z\,CV/r)^2$.
5. Pour une proportion, $P = 0{,}5$ par prudence : ± 5 points demandent $n_0 = 384{,}2$.
6. On majore le S du pilote par prudence ; on divise par le taux de réponse.
7. Chaque domaine coûte presque une enquête ; l'effet de plan multiplie n_0.

La suite

Le chapitre 6 tire séparément dans chaque province : c'est le **sondage stratifié**, plus précis à taille égale.