Échantillonnage

Chapitre 13 : Régression, post-stratification et calage

Exercices corrigés

Jaouad Madkour
Faculté d'Économie et de Gestion de Tanger
Échantillonnage

Table des matières

Partie 1 — Énoncés 1

Partie 2 — Corrections 4

1 Exercices du chapitre

Exercice 1 — Régression et ratio sur une population de quatre unités

Une population compte quatre unités. On connaît pour chacune une variable auxiliaire x et la variable d'intérêt y :

Unité	1	2	3	4
x_k	1	2	4	5
y_k	3	4	8	9

On tire $n = 2$ unités par sondage aléatoire simple ; $\overline{X}$ est connu.

(a) Calculer $\overline{X}$, $\overline{Y}$, S^2, S_x^2, S_{xy}, puis la droite de la population (B, A), le ratio $R = Y/X$ et les résidus $e_k = y_k - A - B\,x_k$. Calculer S_e^2.

(b) Énumérer les six échantillons. Pour chacun, calculer $\overline{y}$, la pente b de la droite qui passe par les deux points tirés, l'estimation par la régression $\overline{y} + b\,(\overline{X} - \overline{x})$ et l'estimation par le ratio $\overline{y}\,\overline{X}/\overline{x}$.

(c) Calculer l'espérance, le biais, la variance et l'erreur quadratique moyenne des trois estimateurs.

(d) Comparer la variance exacte de l'estimateur par la régression à son approximation $\frac{1-f}{n}\,S_e^2$. D'où vient l'écart ? Que conclure pour les très petits échantillons ?

Exercice 2 — Post-stratifier une enquête sur les salaires

Une zone industrielle emploie $N = 5\,000$ salariés : 3 500 ouvriers, 1 000 employés et 500 cadres. Un sondage aléatoire simple de $n = 250$ salariés donne, pour le salaire mensuel (en milliers de dh) :

Catégorie	Salariés tirés n_h	Moyenne $\overline{y}_h$	Écart-type s_h
Ouvriers	160	3,2	0,8
Employés	60	5,0	1,5
Cadres	30	12,0	4,0

(a) Calculer la moyenne simple $\overline{y}$ et l'estimation post-stratifiée $\sum_h W_h\,\overline{y}_h$. Comparer les effectifs tirés aux effectifs attendus $n\,W_h$, et expliquer l'écart entre les deux estimations.

(b) Donner les poids de la post-stratification ; vérifier que leur somme vaut N. Les comparer au poids de sondage N/n.

(c) Estimer la variance de l'estimation post-stratifiée par $\sum_h W_h^2\,(1 - n_h/N_h)\,s_h^2/n_h$. En déduire un intervalle de confiance à 95 % du salaire moyen, puis une estimation de la masse salariale mensuelle.

(d) Calculer la variance s^2 de l'ensemble des 250 salaires (décomposer la somme des carrés en une partie à l'intérieur des catégories et une partie entre catégories), puis la variance estimée de $\overline{y}$ sous le sondage aléatoire simple. Comparer avec (c) et interpréter.

(e) Que faire si le tirage n'avait ramené aucun cadre ?

Exercice 3 — Caler sur un effectif connu

Un fichier administratif recense $N = 40$ entreprises d'une zone d'activités, qui emploient au total $X = 140$ salariés. On en tire $n = 4$ par sondage aléatoire simple (poids $d_k = 10$), et l'on relève leur effectif x_k et leur chiffre d'affaires mensuel y_k (centaines de milliers de dh) :

Entreprise tirée	1	2	3	4
Effectif x_k	1	2	3	6
Chiffre d'affaires y_k	5	8	9	14

(a) Estimer N et X par Horvitz-Thompson. Les estimations sont-elles cohérentes avec le fichier ?

(b) On cherche des poids calés sur la constante et l'effectif, de la forme $w_k = 10\,[1 + c\,(x_k - \overline{x})]$. Calculer $c = (\overline{X} - \overline{x})/\hat{\sigma}_x^2$, avec $\hat{\sigma}_x^2 = \frac{1}{n}\sum_S (x_k - \overline{x})^2$, puis les quatre poids. Vérifier que $\sum_S w_k = 40$ et $\sum_S w_k\, x_k = 140$.

(c) Calculer l'estimation calée du chiffre d'affaires total, et la comparer à l'estimation de Horvitz-Thompson et à celle du ratio $X\,\overline{y}/\overline{x}$.

(d) Vérifier que la moyenne calée est égale à l'estimation par la régression $\overline{y} + b\,(\overline{X} - \overline{x})$.

2 Exercice cumulatif (chapitres 13, 12, 11, 10 et 9)

Exercice 4 — Cinq districts et leur taille

On tire $n_I = 5$ des $N_I = 50$ districts de la population du cours par sondage aléatoire simple, et l'on interroge tous leurs ménages. On sait que la population compte $N = 2\,000$ ménages, soit $\overline{M} = 40$ par district. Les dépenses totales Y_i sont en milliers de dh :

District tiré	41	40	42	49	4
Ménages M_i	26	29	43	30	50
Dépense totale Y_i	150,2	189,2	294,4	202,8	463,5

(a) (Chapitre 9.) Estimer la dépense moyenne par ménage par Horvitz-Thompson, $\widehat{Y}_{HT}/N$, puis par le ratio $\sum Y_i / \sum M_i$.

(b) (Chapitres 12 et 13.) Le nombre de ménages M_i joue le rôle d'une variable auxiliaire, de moyenne connue $\overline{M} = 40$. Calculer la pente b de la régression des Y_i sur les M_i dans l'échantillon, puis l'estimation par la régression du total, $N_I\,[\,\overline{y}_I + b\,(\overline{M} - \overline{m})\,]$, où $\overline{y}_I$ et $\overline{m}$ sont les moyennes des Y_i et des M_i tirés. En déduire la dépense moyenne par ménage ($\overline{Y} = 7\,911$ dh).

(c) (Chapitre 11.) Calculées sur les 50 districts, les erreurs-types de la dépense moyenne estimée (exactes pour Horvitz-Thompson et Hansen-Hurwitz, approchées pour le ratio et la régression) valent : 1 516,6 dh (Horvitz-Thompson), 773,8 dh (ratio), 753,7 dh (régression), et 754,3 dh pour un tirage de 5 districts avec probabilités proportionnelles à M_i (Hansen-Hurwitz). Calculer les effets de plan par rapport à un sondage aléatoire simple de 200 ménages (erreur-type 267,3 dh). Que montre la comparaison des trois dernières valeurs ?

(d) (Chapitre 10.) Pour réduire le coût, on n'interroge plus que $m = 10$ ménages tirés au hasard dans chaque district tiré. Calculer le poids d'un ménage de chacun des cinq districts. Le plan est-il autopondéré ? Que deviennent les poids si les districts sont tirés avec des probabilités $\pi_i = 5\,M_i/2\,000$?

(e) Pourquoi tous ces effets de plan restent-ils supérieurs à 7, alors que la régression sur la taille donnait 0,613 pour un sondage aléatoire simple de ménages ?

3 Applications en sciences économiques

Exercice 5 — La production de blé et la superficie cultivée

Une province compte $N = 800$ exploitations céréalières, dont le cadastre agricole donne la superficie moyenne : $\overline{X} = 7{,}2$ ha (chiffres fictifs). Un sondage aléatoire simple de $n = 40$ exploitations donne, pour la superficie x (ha) et la production de blé y (quintaux) :

$$\overline{x} = 6{,}5 \qquad \overline{y} = 95 \qquad s_x = 4 \qquad s_y = 60 \qquad r_{xy} = 0{,}9$$

(a) Calculer s_{xy}, la pente b, puis les estimations par la régression de la production moyenne et de la

production totale de la province.

(b) Comparer avec l'estimation de Horvitz-Thompson du total, $N\,\overline{y}$, et avec celle du ratio. Pourquoi ces deux corrections vont-elles dans le même sens ?

(c) Estimer la variance de $\overline{y}$, puis celle de l'estimateur par la régression, avec $s_e^2 = \frac{n-1}{n-2}\, s_y^2\,(1 - r_{xy}^2)$. Donner l'intervalle de confiance à 95 % de la production moyenne, puis l'erreur-type de la production totale.

(d) Combien d'exploitations un sondage aléatoire simple sans information auxiliaire devrait-il interroger pour atteindre la même précision ? Qu'apporte ici le cadastre ?

Exercice 6 — Caler une enquête auprès des commerces

Une province compte $N = 2\,000$ commerces (chiffres fictifs). Le recensement des établissements donne deux marges : 800 commerces ruraux et 1 200 urbains ; 900 dans l'alimentation, 500 dans l'habillement, 600 dans les services. Le tableau croisé n'est pas connu. Une enquête par sondage aléatoire simple de $n = 200$ commerces (poids $d_k = 10$) donne les effectifs tirés et le chiffre d'affaires mensuel moyen par case (en milliers de dh) :

Commerces tirés (CA moyen)	Alimentation	Habillement	Services
Rural	50 (40)	30 (55)	10 (70)
Urbain	40 (60)	30 (90)	40 (80)

(a) Calculer les marges estimées avec les poids $d_k = 10$, et les comparer aux marges connues.

(b) Faire un tour d'ajustement proportionnel itératif : d'abord le milieu, puis le secteur. Donner les facteurs, les poids de chaque case et les marges du milieu après ce tour.

(c) Après convergence, les poids valent 9,288, 7,672 et 10,545 pour les commerces ruraux, 10,890, 8,995 et 12,364 pour les urbains. Vérifier qu'ils reproduisent les deux marges.

(d) Estimer le chiffre d'affaires mensuel moyen avec les poids d_k, puis avec les poids calés. Expliquer l'écart.

(e) Pourquoi ne peut-on pas post-stratifier directement sur les six cases ? À quelle condition sur l'échantillon le calage sur marges est-il possible ?

1 Exercices du chapitre

Correction — Exercice 1

(a) $\overline{X} = 12/4 = 3$ et $\overline{Y} = 24/4 = 6$. Écarts : $-2, -1, 1, 2$ pour x ; $-3, -2, 2, 3$ pour y. Sommes : $\sum(x_k - \overline{X})^2 = 10$, $\sum(y_k - \overline{Y})^2 = 26$, $\sum(x_k - \overline{X})(y_k - \overline{Y}) = 6 + 2 + 2 + 6 = 16$. D'où $S_x^2 = 10/3 = 3{,}333$, $S^2 = 26/3 = 8{,}667$, $S_{xy} = 16/3 = 5{,}333$.
$B = S_{xy}/S_x^2 = 16/10 = 1{,}6$; $A = 6 - 1{,}6 \times 3 = 1{,}2$; $R = 24/12 = 2$. Résidus : $3 - 1{,}2 - 1{,}6 = 0{,}2$; $4 - 1{,}2 - 3{,}2 = -0{,}4$; $8 - 1{,}2 - 6{,}4 = 0{,}4$; $9 - 1{,}2 - 8 = -0{,}2$. Somme des carrés 0,4, d'où $S_e^2 = 0{,}4/3 = 0{,}1333$. La droite colle aux points : $\rho^2 = 1 - S_e^2/S^2 = 0{,}985$.
(b) Avec deux points, la pente des moindres carrés est celle de la droite qui les joint, $b = (y_2 - y_1)/(x_2 - x_1)$:

Échantillon	$\{1,2\}$	$\{1,3\}$	$\{1,4\}$	$\{2,3\}$	$\{2,4\}$	$\{3,4\}$
$\overline{x}$	1,5	2,5	3	3	3,5	4,5
$\overline{y}$	3,5	5,5	6	6	6,5	8,5
b	1	1,667	1,5	2	1,667	1
Régression	5	6,333	6	6	5,667	7
Ratio	7	6,6	6	6	5,571	5,667

Par exemple, pour $\{1,2\}$: $3{,}5 + 1 \times (3 - 1{,}5) = 5$ et $3{,}5 \times 3/1{,}5 = 7$.
(c) Chaque échantillon a la probabilité $1/6$.

Estimateur	Espérance	Biais	Variance	Erreur quadratique moyenne
Moyenne simple $\overline{y}$	6	0	2,167	2,167
Régression	6	0	0,370	0,370
Ratio	6,140	0,140	0,256	0,276

Pour la régression, les écarts à 6 sont -1, $\frac{1}{3}$, 0, 0, $-\frac{1}{3}$, 1 : $V = (1 + \frac{1}{9} + \frac{1}{9} + 1)/6 = 0{,}370$. Pour $\overline{y}$, on retrouve $(1 - f)\, S^2/n = 0{,}5 \times 8{,}667/2 = 2{,}167$. Les deux estimateurs qui utilisent x divisent l'erreur quadratique moyenne de $\overline{y}$ par près de 6 (régression) et près de 8 (ratio).
(d) L'approximation donne $\frac{1-f}{n} S_e^2 = \frac{0{,}5}{2} \times 0{,}1333 = 0{,}0333$, onze fois moins que la variance exacte. Elle suppose $b \approx B$; or, avec deux points, la pente varie de 1 à 2 autour de $B = 1{,}6$, et son erreur, multipliée par $\overline{X} - \overline{x}$, s'ajoute à celle de $\overline{y}$. Sur de très petits échantillons, la formule est trop optimiste, et le ratio, qui n'estime qu'un coefficient, peut faire mieux (ici, 0,276 contre 0,370). L'estimateur par la régression est un outil de **grands** échantillons.

Correction — Exercice 2

(a) $\overline{y} = (160 \times 3{,}2 + 60 \times 5 + 30 \times 12)/250 = 1\,172/250 = 4{,}688$ milliers de dh. Avec $W_h = 0{,}7\,;\, 0{,}2\,;\, 0{,}1$:

$$\widehat{\overline{Y}}_{\text{post}} = 0{,}7 \times 3{,}2 + 0{,}2 \times 5 + 0{,}1 \times 12 = 2{,}24 + 1 + 1{,}2 = 4{,}44 \text{ milliers de dh.}$$

Effectifs attendus : $n\,W_h = 175$, 50 et 25. Le tirage a ramené trop d'employés (60) et de cadres (30), les mieux payés, et trop peu d'ouvriers (160) : $\overline{y}$ dépasse l'estimation post-stratifiée de 248 dh. La post-stratification rend à chaque catégorie sa part dans la population.
(b) Poids N_h/n_h : ouvriers $3\,500/160 = 21{,}875$; employés $1\,000/60 = 16{,}667$; cadres $500/30 = 16{,}667$. Somme : $160 \times 21{,}875 + 60 \times 16{,}667 + 30 \times 16{,}667 = 3\,500 + 1\,000 + 500 = 5\,000 = N$. Le poids de sondage valait $N/n = 20$ pour tous : un ouvrier tiré compte davantage, un employé ou un cadre tiré compte moins.
(c) Les trois termes :

$$0{,}7^2\left(1 - \tfrac{160}{3\,500}\right)\tfrac{0{,}8^2}{160} + 0{,}2^2\left(1 - \tfrac{60}{1\,000}\right)\tfrac{1{,}5^2}{60} + 0{,}1^2\left(1 - \tfrac{30}{500}\right)\tfrac{4^2}{30} = 0{,}00187 + 0{,}00141 + 0{,}00501 = 0{,}00829$$

Erreur-type 0,0911 millier de dh, soit 91 dh. Intervalle à 95 % : $4{,}44 \pm 1{,}96 \times 0{,}0911 = 4{,}44 \pm 0{,}18$, soit $[4{,}26\,;4{,}62]$ milliers de dh. Masse salariale mensuelle : $5\,000 \times 4{,}44 = 22\,200$ milliers de dh, soit 22,2 millions de dh, avec l'intervalle [21,3 ; 23,1] millions.
(d) Somme des carrés à l'intérieur des catégories : $\sum (n_h - 1)\, s_h^2 = 159 \times 0{,}64 + 59 \times 2{,}25 + 29 \times 16 = 101{,}76 + 132{,}75 + 464 = 698{,}51$. Entre catégories : $\sum n_h (\overline{y}_h - \overline{y})^2 = 354{,}26 + 5{,}84 + 1\,603{,}96 = 1\,964{,}06$. D'où $s^2 = (698{,}51 + 1\,964{,}06)/249 = 10{,}693$ et

$$\widehat{V}(\overline{y}) = (1 - 0{,}05) \times 10{,}693/250 = 0{,}0406, \quad \text{erreur-type } 0{,}202.$$

La variance post-stratifiée est cinq fois plus petite (rapport 0,204) : la catégorie explique 74 % de la dispersion des salaires ($1\,964/2\,663$). Post-stratifier élimine la part de l'erreur due au hasard de la composition de l'échantillon.
(e) On ne pourrait pas calculer $\overline{y}_h$ des cadres. On regroupe alors la catégorie vide avec la plus proche (employés et cadres, $W = 0{,}3$), ou l'on renonce à cette variable. En pratique, on exige au moins une vingtaine d'unités tirées par post-strate.

Correction — Exercice 3

(a) $\widehat{N} = 4 \times 10 = 40 = N$: cohérent. $\widehat{X} = 10 \times (1 + 2 + 3 + 6) = 120$ salariés, contre 140 : il en manque 20. Les entreprises tirées sont plus petites que la moyenne ($\overline{x} = 3$ contre $\overline{X} = 140/40 = 3{,}5$).
(b) Écarts $x_k - \overline{x}$: $-2, -1, 0, 3$; $\hat{\sigma}_x^2 = (4 + 1 + 0 + 9)/4 = 3{,}5$ et $c = (3{,}5 - 3)/3{,}5 = 1/7 = 0{,}1429$. Les poids $w_k = 10\,[1 + (x_k - 3)/7]$ valent

$$\frac{50}{7} = 7{,}143 \qquad \frac{60}{7} = 8{,}571 \qquad \frac{70}{7} = 10 \qquad \frac{100}{7} = 14{,}286$$

$\sum_S w_k = 280/7 = 40$ et $\sum_S w_k\, x_k = (50 + 120 + 210 + 600)/7 = 980/7 = 140$: les deux totaux du fichier sont retrouvés. La plus grande entreprise, qui porte l'écart, voit son poids monter.
(c) $\widehat{Y}_{\text{cal}} = (50 \times 5 + 60 \times 8 + 70 \times 9 + 100 \times 14)/7 = 2\,760/7 = 394{,}3$ centaines de milliers de dh, soit 39,4 millions de dh par mois. Horvitz-Thompson : $10 \times 36 = 360$; ratio : $140 \times 9/3 = 420$. Le calage corrige l'estimation de Horvitz-Thompson vers le haut, comme le ratio, mais moins fort.
(d) Moyenne calée : $394{,}3/40 = 9{,}857$. Écarts $y_k - \overline{y}$: $-4, -1, 0, 5$; $\sum (x_k - \overline{x})(y_k - \overline{y}) = 8 + 1 + 0 + 15 = 24$, d'où $b = 24/14 = 1{,}714$ et

$$\overline{y} + b\,(\overline{X} - \overline{x}) = 9 + 1{,}714 \times 0{,}5 = 9{,}857.$$

C'est la démonstration du cours : le calage linéaire sur la constante et une variable redonne l'estimateur par la régression.

2 Exercice cumulatif (chapitres 13, 12, 11, 10 et 9)

Correction — Exercice 4

(a) $\sum Y_i = 1\,300{,}1$ milliers de dh et $\sum M_i = 178$ ménages. Horvitz-Thompson : $\widehat{Y} = \frac{50}{5} \times 1\,300{,}1 = 13\,001$ milliers de dh, soit $13\,001\,000/2\,000 = 6\,500{,}5$ dh par ménage. Ratio : $1\,300\,100/178 = 7\,303{,}9$ dh. Les cinq districts tirés sont petits (35,6 ménages en moyenne au lieu de 40) : l'estimateur de Horvitz-Thompson, qui ignore leur taille, sous-estime fortement ($-1\,410{,}9$ dh) ; le ratio corrige ce défaut ($-607{,}5$ dh).
(b) $\overline{m} = 35{,}6$ et $\overline{y}_I = 260{,}02$. Écarts des M_i : $-9{,}6$; $-6{,}6$; $7{,}4$; $-5{,}6$; $14{,}4$, de somme des carrés 429,2 ; somme des produits avec les écarts des Y_i : 5 026,64. D'où $b = 5\,026{,}64/429{,}2 = 11{,}712$ milliers de dh par ménage, et

$$\widehat{Y}_{\text{reg}} = 50 \times [260{,}02 + 11{,}712 \times (40 - 35{,}6)] = 50 \times 311{,}55 = 15\,577{,}6 \text{ milliers de dh},$$

soit 7 788,8 dh par ménage, à 123 dh de $\overline{Y}$. Ici, la régression fait mieux que le ratio : le district le plus grand (le 4, 50 ménages) est aussi le plus dépensier, et la pente b dépasse la moyenne par ménage.

(**c**) *deff* = (erreur-type/267,3)^2 : 32,2 (Horvitz-Thompson), 8,38 (ratio), 7,95 (régression), 7,96 (Hansen-Hurwitz). Les trois dernières méthodes font presque jeu égal : toutes trois exploitent la même information, la taille M_i, l'une au moment de l'estimation (ratio, régression), l'autre au moment du tirage (probabilités proportionnelles à la taille). Mettre M_i dans l'estimateur ou dans le plan revient ici au même.
(**d**) Poids d'un ménage du district i : $\frac{N_I}{n_I} \cdot \frac{M_i}{m} = 10 \times \frac{M_i}{10} = M_i$, soit 26, 29, 43, 30 et 50 : le plan n'est pas autopondéré. Avec $\pi_i = 5\,M_i/2\,000$ (de 0,065 à 0,125 ici, au plus 0,178 pour le plus grand district) : poids $\frac{2\,000}{5\,M_i} \cdot \frac{M_i}{10} = 40$ pour tous. Le plan devient autopondéré : c'est le plan à deux degrés du chapitre 11.
(**e**) Ratio et régression ne corrigent que l'effet des **tailles** inégales. Il reste la ressemblance des ménages d'un même district : $deff \approx 1 + (\overline{M} - 1)\,\rho = 1 + 39 \times 0{,}162 = 7{,}3$ (chapitre 9). La taille du ménage (0,613) agissait sur la dispersion entre ménages ; ici, l'erreur vient des différences entre districts, que seuls davantage de districts, moins de ménages par district (chapitre 10) ou une stratification des districts peuvent réduire.

3 Applications en sciences économiques

Correction — Exercice 5

(**a**) $s_{xy} = r_{xy}\,s_x\,s_y = 0{,}9 \times 4 \times 60 = 216$; $b = 216/16 = 13{,}5$ quintaux par hectare. Production moyenne : $95 + 13{,}5 \times (7{,}2 - 6{,}5) = 95 + 9{,}45 = 104{,}45$ quintaux ; production totale : $800 \times 104{,}45 = 83\,560$ quintaux.
(**b**) Horvitz-Thompson : $800 \times 95 = 76\,000$ quintaux. Ratio : $\widehat{R} = 95/6{,}5 = 14{,}615$ quintaux par hectare, $14{,}615 \times 7{,}2 = 105{,}23$, total 84 185 quintaux. Les deux corrections montent l'estimation : les exploitations tirées sont trop petites (6,5 ha contre 7,2), et la production croît avec la superficie. Le ratio corrige un peu plus (+10,23 contre +9,45 quintaux), car sa pente 14,6 dépasse $b = 13{,}5$.
(**c**) $\widehat{V}(\overline{y}) = (1 - 0{,}05) \times 3\,600/40 = 85{,}5$, erreur-type 9,25 quintaux. Régression : $s_e^2 = \frac{39}{38} \times 3\,600 \times (1 - 0{,}81) = 702{,}0$ et

$$\widehat{V} = \frac{0{,}95}{40} \times 702{,}0 = 16{,}67, \qquad \text{erreur-type } 4{,}08 \text{ quintaux.}$$

Intervalle à 95 % : $104{,}45 \pm 1{,}96 \times 4{,}08 = 104{,}45 \pm 8{,}00$, soit [96,45 ; 112,45] quintaux par exploitation. Erreur-type du total : $800 \times 4{,}083 = 3\,267$ quintaux.
(**d**) Effet de plan : $16{,}67/85{,}5 = 0{,}195$, proche de $1 - r_{xy}^2 = 0{,}19$. Taille effective (chapitre 7) : $40/0{,}195 = 205$ exploitations. En tenant compte du taux de sondage (chapitre 5) : $n_0 = 3\,600/16{,}67 = 215{,}9$ et $n = 215{,}9/(1 + 215{,}9/800) = 170$ exploitations. Le cadastre, gratuit pour l'enquête, vaut plus de quatre fois l'échantillon : une forte corrélation (0,9) divise la variance par cinq.

Correction — Exercice 6

(**a**) Milieu : rural $10 \times 90 = 900$ (connu 800), urbain $10 \times 110 = 1\,100$ (connu 1 200). Secteur : alimentation 900 (900), habillement 600 (500), services 500 (600). L'échantillon a trop de commerces ruraux et d'habillement, trop peu d'urbains et de services.
(**b**) Milieu : facteurs $800/900 = 0{,}8889$ (rural) et $1\,200/1\,100 = 1{,}0909$ (urbain) ; poids 8,889 et 10,909. Secteurs estimés : alimentation $50 \times 8{,}889 + 40 \times 10{,}909 = 880{,}8$; habillement 593,9 ; services 525,3. Facteurs : $900/880{,}8 = 1{,}0218$; $500/593{,}9 = 0{,}8418$; $600/525{,}3 = 1{,}1423$. Poids après ce tour :

Poids	Alimentation	Habillement	Services
Rural	9,083	7,483	10,154
Urbain	11,147	9,184	12,462

Les secteurs sont justes ; le milieu devient 780,2 ruraux et 1 219,8 urbains, à 20 commerces près. Un deuxième tour ramène l'écart à 1,8 commerce.
(**c**) On multiplie chaque poids par l'effectif tiré de sa case (par exemple $50 \times 9{,}288 = 464{,}40$) :

Effectif × poids	Alimentation	Habillement	Services	Total
Rural	464,40	230,16	105,45	800,01
Urbain	435,60	269,85	494,56	1 200,01
Total	900,00	500,01	600,01	2 000,02

Les deux marges sont reproduites, au centième près dû aux arrondis des poids. Dans chaque secteur, un commerce rural pèse 0,853 fois un commerce urbain.

(d) Avec $d_k = 10$: $\frac{1}{2\,000}\sum_S 10\,y_k = 12\,650/200 = 63{,}25$ milliers de dh. Avec les poids calés : $\frac{1}{2\,000}\sum_S w_k\,y_k = 128\,604/2\,000 = 64{,}30$ milliers de dh, soit +1,05 (+1,7 %). Le calage réduit le poids des commerces ruraux (48,3 en moyenne dans l'échantillon, contre 75,5 pour les urbains) et augmente celui des services (78,0), deux corrections qui montent l'estimation.

(e) Post-stratifier sur les six cases exigerait les effectifs N_{ph} de chaque case, que le recensement ne donne pas : on ne connaît que les marges. Le calage sur marges n'utilise que ces marges ; il faut au moins un commerce tiré dans chaque modalité de chaque marge (une modalité absente ne peut pas être repondérée), et des marges de même total (2 000). En pratique, on évite les cases vides ou presque, qui produisent des poids extrêmes.