Méthodes statistiques
Réponses Chapitre 10

Problème 10.1

a)	$\overline{y}_{st}$	122,588	
b)	T_{st}	391668,6	
c)	$\hat{\sigma}_{\overline{y}_{st}}$	17,63076	
d)	$\hat{\sigma}_{T_{st}}$	56330,29	
e)	Intervalle de confiance pour μ	87,32647	157,8495
f)	Intervalle de confiance pour τ	279008,1	504329,2

g) Allocation optimale : $n_1 = 3$; $n_2 = 1$; $n_3 = 38$. Avec cette allocation, l'écart-type de l'estimateur de la moyenne aurait été de 10,17 (au lieu de 17,63) si on suppose que les s_h obtenus dans cet échantillon sont de bonnes estimations des S_h.

Problème 10.2

a) Répartition optimale : 21, 23, 33, 40, 63.

b) Allocation proportionnelle : 5, 9, 32, 47, 87. L'écart-type de $\overline{y}_{st}$ est 1,2200 avec l'allocation proportionnelle et 1,0150 avec l'allocation optimale.

Problème 10.3

Allocation optimale : 10, 26 et 64.

Problème 10.4

a) (i) 875,2563 ; ii) 207,2841; iii) 3172,004.

b) Répartition proportionnelle : $n_1 = 2$, $n_2 = 6$, $n_3 = 12$. Elle donne un écart-type de 2117,746.

c) Répartition optimale : $n_1 = 10$, $n_2 = 8$, $n_3 = 2$. Elle donne un écart-type de 207,28.

Problème 10.5

a) $\hat{p}_{st} = 0{,}633341$.

b) $\hat{\sigma}_{\hat{p}_{st}} = 0{,}02612$. Intervalle de confiance : $0{,}5811092 \le p \le 0{,}68558$.

c) 9329.

d) $8\ 560 \le N_c \le 10\ 099$.

e) Écart-type : 0,028.

f) 491. L'allocation doit être (à peu près, les erreurs d'arrondi pouvant changer les choses) la suivante : 36 ; 102 ; 55 ; 297.

Problème 10.6

$\overline{u} = 25{,}84239$; $\hat{\sigma}_{\overline{u}} = 1{,}280819$.

roblème 10.7

a) Intervalle de confiance pour la moyenne par 100m^2 : (8,718787 ; 10,077934).

b) Dépenses moyennes par propriété : 107,54 $. Écart-type de l'estimateur : 3,888075. Intervalle de confiance est (99,77 ; 115,32).

Problème 10.8

a) Solde total : 10 135 367; Écart-type : 378 958,7. Intervalle de confiance pour le montant total : (9 377 450 ; 10 893 285).

b) Moyenne par compte : 101,3669; écart-type de l'estimateur : 3,79008. Intervalle de confiance pour la moyenne par compte : (93,78669 ; 108,94701).

Problème 10.9

a) Écart-type de $\overline{y}$ (la moyenne des 12 unités primaires sélectionnées) : 49,68273

b) Moyenne par unité secondaire : 0,954293. Écart-type de l'estimateur = 70,8256.

c) On considère ici les propriétés d'un échantillon tiré avec remise d'une population dont les valeurs sont les $u_i = y_i/x_i$. L'estimateur de la moyenne par unité secondaire est $\overline{u}$, l'estimateur de la moyenne par unité primaire est $M_0\overline{u}/N$. S_u = 0,3202935; $\sigma_{\overline{u}}$ = 0,09246076; l'écart-type de l'estimateur sera donc $M_0\sigma_{\overline{u}}/N$ = 92,89162.

d) Les unités à tirer sont les unités 2, 11, 21, 25, 28 (2 fois), 30, 41, 42, 46 (2 fois), 49. Ou, dans l'ordre de présentation des numéros tirés, 41, 49, 28, 11, 46, 2, 25, 21, 42, 30, 46, 28.

Problème 10.10

a) $S_y = 15,837$; écart-type de $N\bar{y}$: $28\sqrt{1-10/28}(15,837)/\sqrt{10} = 112,43$.

b) On a $S_x = 1,13797$ et $S_{xy} = 14,92328$ et $R = 14,77$. (i) Écart-type de l'estimateur par la différence = 105,8386 ; (ii) Écart-type de l'estimateur par le quotient = 68,28.

c) (i) Les variances de strates sont 77,9 ; 41,2 ; 103,5 ; et 154,9 pour les strates avec x = 1, 2, 3 et 4, respectivement. Les tailles des échantillons sont 5, 2, 2, et 1, pour les strates avec x = 1, 2, 3 et 4, respectivement. L'écart-type est 70,93.
(ii) Les tailles des échantillons sont 4, 1, 2, et 3, pour les strates avec x = 1, 2, 3 et 4, respectivement. L'écart-type est 67,47.

Problème 10.11

a) $S_y = 203,4798995$; $\sigma_{\bar{y}} = \sqrt{1-\frac{50}{102}}\frac{203,48}{\sqrt{50}} = 20,54$.

b) Quelques données sur les strates :
Allocation proportionnelle : 6 ; 16 ; 9 ; 19. Écart-type de l'estimateur de la moyenne est 20,19.
Allocation optimale : 5 ; 13 ; 10 ; 22. Écart-type de l'estimateur de la moyenne est 19,70.

Il y a très peu de différences entre les trois approches (pourquoi?)

Problème 10.12

a) Prix moyen d'une maison : 289,4853; et Écart-type de l'estimateur : 213,6392.

b) Prix moyen d'une maison : 281,2143. Écart-type de l'estimateur : 95,34. Notez bien que le but de cet exercice est purement pédagogique; l'utilisation de l'estimateur par le quotient serait en pratique fortement déconseillée étant donné la petitesse de l'échantillon.

Problème 10.13

Intervalle de confiance : [130,7587 ; 329,2622]

Problème 10.14

Moyenne de la population : 85 038,82; écart-type de l'estimateur : 4152,937. Intervalle de confiance est [76 733 ; 93 345].

Problème 10.15

a) Allocation optimale : $n_1 = 4$, $n_2 = 3$; et $n_3 = 5$.

b) $\hat{\sigma}_{\bar{y}_{st}} = 21,464$

c) La répartition optimale basée sur les y est la même que celle basée sur les x : 4; 3; et 5.

d) Écart-type de $\bar{y}_{st}$ sous l'allocation proportionnelle : 21,96, un peu supérieur à l'écart-type sous l'allocation optimale.

e) L'écart-type de l'estimateur dans un échantillon aléatoire simple de taille 12 est 30,9, ce qui est bien plus grand qu'avec un échantillon stratifié. La différence importante entre l'échantillon aléatoire simple et l'échantillon stratifié est due au fait que les strates ont des moyennes très différentes. C'est la situation dans laquelle la stratification est le plus efficace.

f) L'écart-type de l'estimateur par le quotient est 17,28, calculé à l'aide des données suivantes:

μ_y	μ_x	R	s_y	s_x	s_{xy}
796,13	696,23	1,1435	138,16	112,95	13029,49

On découvre finalement, que le mieux est d'utiliser la variable x comme variable auxiliaire. La corrélation entre les deux variables étant de l'ordre de 0,83, il est clair que l'estimation par le quotient est efficace. La stratification par rapport à x est une façon d'exploiter cette corrélation. L'estimation par le quotient en est une autre.

Problème 10.16

Les 20 valeurs possibles de l'estimateur deviennent alors

```
439,4152  411,9298  403,5088  461,0526  469,7076  447,0175  414,8538  414,0351
414,7368  420,3509  412,2807  353,8012  388,1871  400,4678  398,1287  407,7193
429,1228  416,8421  371,3450  353,8012
```

Leur moyenne, 411,4152, est bien égale à la moyenne de la population.

Problème 10.17

a) $\sigma_{\bar{y}} = \sqrt{1-\frac{25}{180}}\frac{932,821}{\sqrt{25}} = 186,5642$

b) $\sigma^2_{\bar{y}_{st}} = \left(\frac{176}{180}\right)^2\left(1-\frac{21}{176}\right)\frac{(151,1947)^2}{21} + \left(\frac{4}{180}\right)^2\left(1-\frac{4}{4}\right)\frac{(4975,9620)^2}{4} = 916,5434$ et donc $\sigma_{\bar{y}_{st}} = 30,27447$.

c) $S_y^2 + R^2 S_x^2 - 2RS_{xy} = 51075{,}738$. $\hat{\sigma}^2_{\hat{\mu}_{yq}} = (1 - 25/180)\dfrac{S_y^2 + R^2 S_x^2 - 2RS_{xy}}{25}$ =1759,27; $\hat{\sigma}_{\hat{\mu}_{yq}} = 41{,}94$. L'estimateur par le quotient est nettement meilleur que la moyenne, mais la stratification demeure la meilleure stratégie.

d) $\mathrm{Var}(\hat{\mu}_{yq}) = (176/180)^2 \mathrm{Var}(\hat{\mu}_{yq_1}) + (4/180)^2 \mathrm{Var}(\hat{\mu}_{yq_2})$. $\mathrm{Var}(\hat{\mu}_{yq_1}) = (1 - n_1/N_1)\dfrac{(S_{y1}^2 + R_1^2 S_{x1}^2 - 2R_h S_{xy1})}{n_1}$ = 482,7865,

alors que $\mathrm{Var}(\hat{\mu}_{yq_2}) = 0$. Donc $\mathrm{Var}(\hat{\mu}_{yq})$ = 482,7865 et l'écart-type est 21,97. Il est clair qu'on obtient les meilleurs résultats en combinant les deux techniques.

e) i) Les moyennes des strates sont estimées par 100,5238 et 4576,75, respectivement. La moyenne pondérée est 199,9955. C'est l'estimation de la moyenne

ii) Les moyennes des strates sont estimées à 113,7753 et 4576,75. La moyenne pondérée est 233,2480.

iii) La variance de la moyenne échantillonnale de la première strate est estimée à 701,0966. La variance de l'estimateur est nulle dans la deuxième strate est nulle. L'écart-type de l'estimateur stratifié est estimé à 25,8898.

iv) La variance de l'estimateur par le quotient dans la première strate est estimée à 804,1577. Elle est nulle dans la deuxième strate. L'écart-type de l'estimateur stratifié est estimée à 27,7275.

Ces résultats disent, contrairement à ce que nous avons trouvé en d), que les résultats sont meilleurs lorsqu'on estime les moyennes dans chaque strate par la moyenne plutôt que par le quotient. Mais cette conclusion est basée sur une estimation de l'écart-type et peut facilement être erronée.

Problème 10.18

a) Disposition 1: E($\bar{y}$) = 411,5556 alors que la moyenne de la population est μ = 411,5556, un biais négligeable de 0,034 %.

Disposition 2: E($\bar{y}$) = 414,6465, un biais relatif de 0,785 %.

Le biais est peu important dans les deux cas.

b) Disposition 1: L'espérance de s^2 est 6212,492, alors que la variance de la population est S^2 = 5620,644. Donc s^2 a tendance à surestimer la variance de la population. Le biais relatif est de 10,53 %.

Disposition 2: L'espérance de s^2 est 38,02, alors que la variance de la population est S^2 = 5620,644. Donc s^2 a tendance à sous-estimer la variance de la population. Le biais relatif est de 99 %.

c) Disposition 1: L'espérance de $\hat{\sigma}_{\bar{y}}$ est 26,12158, alors que l'écart-type de $\bar{y}$ est 6,1545. Donc $\hat{\sigma}_{\bar{y}}$ a tendance à surestimer l'écart-type de l'estimateur. Le biais relatif est de 76 %.

Disposition 2: L'espérance de $\hat{\sigma}_{\bar{y}}$ est 1,6517, alors que l'écart-type de $\bar{y}$ est 74.24. Donc $\hat{\sigma}_{\bar{y}}$ a tendance à sous-estimer grossièrement l'écart-type de l'estimateur. Le biais relatif est de 97,78 %.