Méthodes statistiques
Réponses - Chapitre 7

Problème 7.1

a) $T = 43\,550$; $\hat{\sigma}_{\bar{y}} = 9$. L'écart-type du total T est estimé par $\hat{\sigma}_T = 7650$.

b) $33{,}23 \le \mu \le 69{,}242 \Leftrightarrow 28250 \le \tau \le 58850$.

Problème 7.2

a) Estimation du total du domaine: 848,3125. Écart-type estimé de l'estimateur : 276,0254. Intervalle de confiance pour le total du domaine : $296 \le \tau_d \le 1400$.

b) Estimation du total du domaine: 775,6. Écart-type estimé de l'estimateur : $\hat{\sigma}_{T_d} = 155{,}8781$.

Problème 7.3

a) Estimation du total du domaine : $T' = 31137{,}04$. $s' = 54{,}39585$. Écart-type estimé de l'estimateur : 8589,263. Intervalle de confiance pour le total du domaine : $13959 \le \tau_d \le 48316$.

b) Estimation du total du domaine : $T_d = 28723$. Écart-type estimé de l'estimateur : $\hat{\sigma}_{T_d} = 6540$.

Problème 7.4

a) Intervalle de confiance pour le total : $48879 \le \tau \le 140366$.

b) L'intervalle de confiance pour la différence entre le montant réel est la valeur aux livres : [-87341 ; 4146]. On ne peut pas conclure que la valeur aux livres est erronée.

Problème 7.5

a) $\hat{p} = 0{,}52$. $\hat{\sigma}_{\hat{p}} = 0{,}06181$. Intervalle de confiance pour p : $0{,}396 \le p \le 0{,}644$.

b) Estimation de l'effectif : $\hat{N}_c = 104$. Écart-type estimé de l'estimateur : 12,36189. Intervalle de confiance pour l'effectif : $79 \le N_c \le 129$.

c) Estimation de la moyenne : $\bar{y}_d = 89653$. Écart-type estimé de l'estimateur : $\hat{\sigma}_{\bar{y}_d} = 1364{,}145$.

d) (i) $n_d = 26$; $N_d = 120$, $\bar{y}_d = 89\,653$, $T_d = 10\,758\,365$, $\hat{\sigma}_{\bar{y}_d} = 1394{,}13$; $\hat{\sigma}_{T_d} = 167\,296$. Intervalle de confiance pour la moyenne du domaine: $86\,925 \le \mu_d \le 92\,441$.
Intervalle de confiance pour le total du domaine : $10\,423\,778 \le \tau_d \le 11\,092\,956$.

(ii) $\bar{y}' = 46620$; $s' = 45\,608$; $\hat{\sigma}_{\bar{y}'} = 5585{,}774$. $T' = N\bar{y}' = 9\,323\,916$. $\hat{\sigma}_{T'} = 1\,117\,155$.
Intervalle de confiance pour le total du domaine : $7\,089\,606 \le \tau_d \le 11\,558\,226$.

Problème 7.6

N_d connue L'estimation dans le domaine 2 est meilleure (cv = 3,786 contre cv = 7,383 dans le domaine 1). C'est la plus grande taille de l'échantillon qui explique la différence.

N_d inconnue Cette qualité s'accentue lorsqu'on ne connaît pas N_d.
Il est certainement utile de connaître la taille du domaine, comme le montrent les écarts-types et les coefficients de variation. Ce qui s'explique bien, d'ailleurs.

Problème 7.7

a) Estimation du total du domaine : $T' = 340$; Estimation du total du domaine : $\hat{\sigma}_{T'} = 103{,}2715$. Intervalle de confiance pour le total : $133 \le \tau \le 547$.

b) Estimation du total du domaine : $T_d = 232$; Écart-type estimé de l'estimateur : $\hat{\sigma}_{T_d} = 48{,}949$. Intervalle de confiance pour le total : $133 \le \tau \le 330$. L'intervalle de confiance ici est plus court, ce qui signifie qu'il est utile de connaître le nombre de librairies qui vendent des livres espagnols.

Problème 7.8

$\hat{\sigma}_{\hat{R}} = 0{,}3498522$. Le coefficient de variation de $\hat{R}$ est estimé par $\hat{\sigma}_{\hat{R}} / \hat{R} = 0{,}083$,

Intervalle de confiance pour le taux d'augmentation de salaire : 250,7 % $\le$ augmentation de salaire $\le$ 390,7 %

Problème 7.9

a) $\bar{y}_d = 3{,}885714$; $\hat{\sigma}_{\bar{y}_d} = 0{,}1596711$.

b) 4,3345.
c) Estimation du total : 8761; Estimation de l'écart-type de l'estimateur: 672.

Problème 7.10

a) $\hat{R} = 0{,}425$; $\hat{\sigma}_{\hat{R}} = 0{,}0720$. Intervalle de confiance : $0{,}281 \le R \le 0{,}569$.

b) $\hat{p} = 0{,}425$. $\hat{\sigma}_{\hat{p}} = 0{,}0392$. Intervalle de confiance pour p : $0{,}347 \le p \le 0{,}503$.

Le deuxième intervalle n'est pas correct, car il suppose que les 160 observations sont indépendantes

Problème 7.11

a) $\hat{p} = 0{,}42$; $\hat{\sigma}_{\hat{p}} = 0{,}06154474$; $0{,}297 \le p \le 0{,}543$.

b) $\hat{N}_C = 88{,}2$; $\hat{\sigma}_{\hat{N}_c} = 12{,}924$; $62{,}35 \le N_c \le 114{,}05$.

c) $T = 365$; $\hat{\sigma}_T = 44{,}07$; $277 \le \tau \le 454$.

d) $\hat{R} = 1{,}412475$; $\hat{\sigma}_{\hat{R}} = 0{,}1744621$; $1{,}063551 \le R \le 1{,}761399$.

e) $T_d = 474{,}60$; $331 \le \tau_d \le 618$.

Problème 7.12

a) $E(\hat{R}) = 1{,}900607$, $R = 1{,}876068$. Donc un biais de 0,0245, soit 1,308 %.

b) Variance réelle de $\hat{R}$: 0,044525. La variance calculée à l'aide de la formule approximative : 0,030347, une erreur d'environ 32 %

c) 0,054.

Problème 7.13

a) Variance de $\hat{R}$: 0,044525, alors que la formule $F = 0{,}0302$. Donc la formule ne donne pas la variance de $\hat{R}$. C'est une approximation. Alors que $\hat{\sigma}_{\hat{R}}$ estime son espérance, qui est 0,03532575, relativement éloignée de la vraie variance $\sigma^2_{\hat{R}} = 0{,}044525$.

b) Tout porte à croire que s_{xy} est un estimateur sans biais de S_{xy}, sauf que ce n'est pas vraiment démontré (pourquoi?).

Problème 7.14

a) $\hat{p} = 250/500 = 0{,}5$; $\hat{\sigma}_{\hat{p}} = 0{,}0221$. Intervalle de confiance : $0{,}456 \le p \le 0{,}544$.

b) $\hat{p} = 350/500 = 0{,}7$; $\hat{\sigma}_{\hat{p}} = 0{,}0202$. Intervalle de confiance : $0{,}660 \le p \le 0{,}740$.

c) $\hat{p} = 250/350 = 0{,}714$. $\hat{\sigma}_{\hat{p}} = 0{,}0238$. Intervalle de confiance : $0{,}667 \le p \le 0{,}762$.

d) $\hat{N}_c = 8662$. Intervalle de confiance : $7898 \le p \le 9426$.

e) $\hat{p} = 0{,}714$; $\hat{\sigma}_{\hat{p}} = 0{,}02385$. Intervalle de confiance pour N_c : $8666 \le N_c \le 9906$.

Problème 7.15

b) On estime donc que le poids moyen des pommes de terre dans le lot est de 9,76 kg. C'est aussi l'espérance du poids moyen des 4000 sacs avant l'élimination des sacs de moins de 9 kg.

b) Intervalle de confiance pour p : [0,0389 ; 0,0521]. Intervalle de confiance pour μ : $9{,}731 \le \mu \le 9{,}794$.

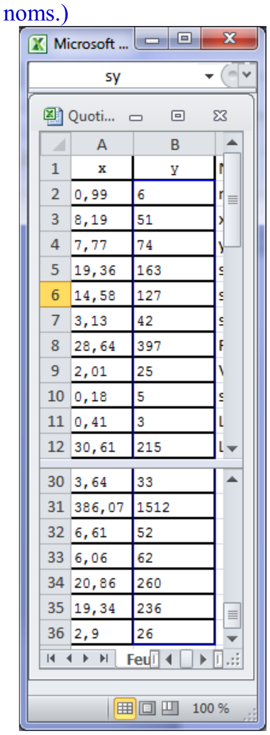
Problème 7.16

a) Faux b) Vrai c) Non d) Non e) Bien sûr que non.

Commandes Excel
Intervalle de confiance pour un quotient

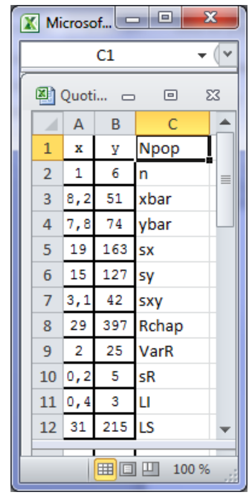
Dans l'exemple suivant, la variable x est le nombre d'habitants dans un échantillon de taille 35 tiré d'une population de 176 paroisses québécoises. La variable y est le nombre de naissances. Le problème est de déterminer un intervalle de confiance pour le quotient $R = \mu_y/\mu_x$.

1 Inscrire les titres et les valeurs de x et de y (`A1-B36`) et leur attribuer des noms, x et y ici (voir le solutionnaire du chapitre 6 pour voir comment attribuer des noms.)



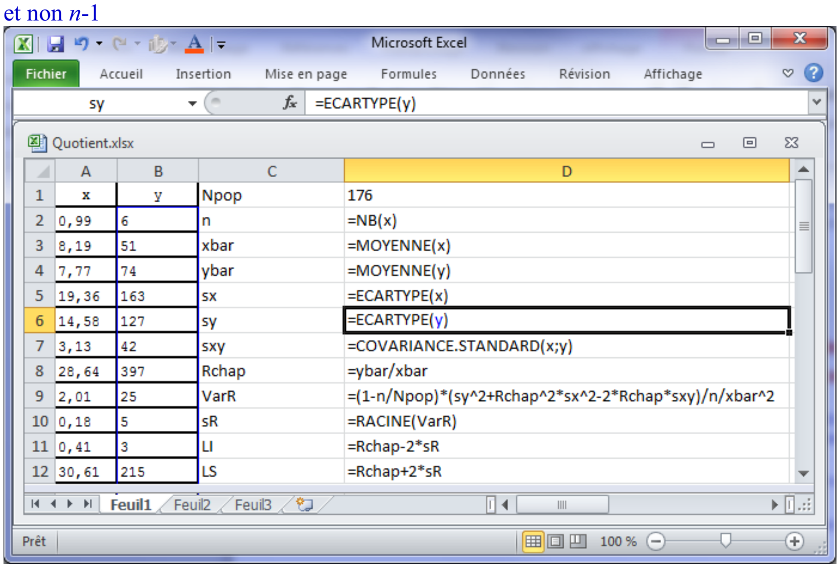
	A	B
1	x	y
2	0,99	6
3	8,19	51
4	7,77	74
5	19,36	163
6	14,58	127
7	3,13	42
8	28,64	397
9	2,01	25
10	0,18	5
11	0,41	3
12	30,61	215
30	3,64	33
31	386,07	1512
32	6,61	52
33	6,06	62
34	20,86	260
35	19,34	236
36	2,9	26

2 Inscrire (dans `C1:C12` ici) les noms des quantités que vous voudrez calculer



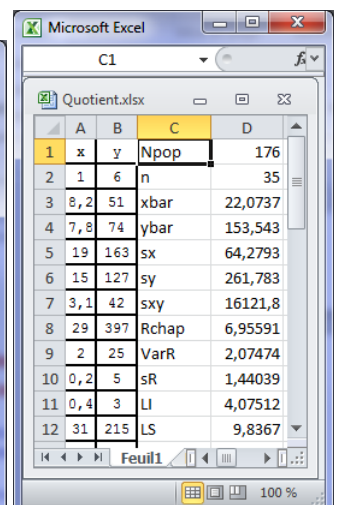
	A	B	C
1	x	y	Npop
2	1	6	n
3	8,2	51	xbar
4	7,8	74	ybar
5	19	163	sx
6	15	127	sy
7	3,1	42	sxy
8	29	397	Rchap
9	2	25	VarR
10	0,2	5	sR
11	0,4	3	LI
12	31	215	LS

3 Effectuer le calculs (`D1-D12`). Notez bien la commande pour la covariance : `COVARIANCE.STANDARD()` et non `COVARIANCE()` (la deuxième divise par n et non n-1



	A	B	C	D
1	x	y	Npop	176
2	0,99	6	n	=NB(x)
3	8,19	51	xbar	=MOYENNE(x)
4	7,77	74	ybar	=MOYENNE(y)
5	19,36	163	sx	=ECARTYPE(x)
6	14,58	127	sy	=ECARTYPE(y)
7	3,13	42	sxy	=COVARIANCE.STANDARD(x;y)
8	28,64	397	Rchap	=ybar/xbar
9	2,01	25	VarR	=(1-n/Npop)*(sy^2+Rchap^2*sx^2-2*Rchap*sxy)/n/xbar^2
10	0,18	5	sR	=RACINE(VarR)
11	0,41	3	LI	=Rchap-2*sR
12	30,61	215	LS	=Rchap+2*sR

4 Voici les résultats :



	A	B	C	D
1	x	y	Npop	176
2	1	6	n	35
3	8,2	51	xbar	22,0737
4	7,8	74	ybar	153,543
5	19	163	sx	64,2793
6	15	127	sy	261,783
7	3,1	42	sxy	16121,8
8	29	397	Rchap	6,95591
9	2	25	VarR	2,07474
10	0,2	5	sR	1,44039
11	0,4	3	LI	4,07512
12	31	215	LS	9,8367