Méthodes statistiques
Réponses Chapitre 6

Problème 6.1
a) $\bar{y}$ =51,235; $s_{y\,=}$ 40,7313

b) Intervalle de confiance :[33,235 ; 69,235]

Problème 6.2
L'intervalle de confiance ne contredit pas l'hypothèse que μ =38,63, puisqu'il inclut cette valeur.

Problème 6.3
a) $\bar{y}$ = 18526,42; $s_{y\,=}$ 9323,831

b) Intervalle de confiance :[16242,56 ; 20810,28]

Problème 6.4
a) $\bar{y}$ = 1,74; $s_{y\,=}$ 1,70006

b) Intervalle de confiance : [1,32 ; 2,16]

Résultat inattendu et décevant : l'intervalle ne contient pas la valeur connue de la moyenne.

Problème 6.5
Voici les données brutes, hommes et femmes :

	Intervalle de confiance	Largeur de l'intervalle
Tous	71,93 ; 74,60	2,68
Femmes	70,65 ; 75,60	4,95
Hommes	71,91 ; 74,83	2,91

Comparaison femmes/hommes L'estimation de la moyenne des femmes est moins précise que pour les hommes, 1) parce que, parmi les femmes, la dispersion des valeurs est apparemment supérieure à celle des hommes et 2) l'échantillon de femmes est plus petit.
Comparaison population/hommes. L'estimation de la moyenne globale est néanmoins plus précise que l'estimation de la moyenne des hommes alors que les données de la population entière sont plus dispersées que celles des hommes seuls. Cela est dû à la plus grande taille de l'échantillon.

Problème 6.6
Intervalle de confiance : [5,32 ; 9,12]

Problème 6.7
On calcule les différences y = B1-A1 et on obtient l'intervalle de confiance suivant pour la moyenne de y :
[-2,42 ; -0,929]. Contrairement à ce à quoi on s'attendrait, on conclut que l'apprentissage fait diminuer le score.

Problème 6.8
Intervalle de confiance : [3820,06 ; 9925,54]. Puisque l'intervalle recouvre la moyenne de la population (4099 $), vous ne pouvez pas confirmer vos soupçons.

Problème 6.9
a) Intervalle de confiance : [0,2178 ; 0,8322]

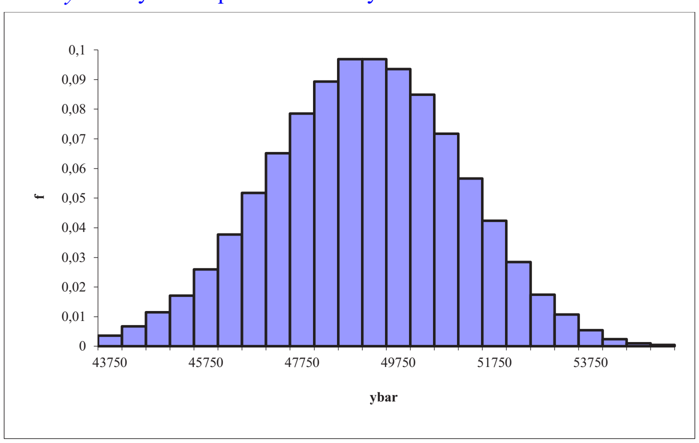
b) Oui.

Problème 6.10
Les inégalités approximatives « ≈ » dans les réponses ci-dessous reflètent le fait que la normalité de la moyenne échantillonnale est nécessairement approximative.

a) $\mu_{\bar{y}_1} = \mu_{\bar{y}_2}$	b) $\sigma_{\bar{y}_1} > \sigma_{\bar{y}_2}$
c) $\mu_{\bar{y}_1} = \mu$	d) $\mu_{\bar{y}_2} = \mu$
e) s_1 ? S	f) P($\bar{y}_1 > \mu+2$) > P($\bar{y}_2 > \mu+2$)
g) P($\bar{y}_1 > \mu + 2\sigma_{\bar{y}_1}$) > P($\bar{y}_1 > \mu + 2S$)	h) P($\bar{y}_1 > \mu + 2\sigma_{\bar{y}_1}$) > P($\bar{y}_2 > \mu + 2\sigma_{\bar{y}_1}$)
i) $\sigma_{\bar{y}_1} < S$	j) P($\bar{y}_1 > \mu + 2\sigma_{\bar{y}_1}$) ≈ P($\bar{y}_2 > \mu + 2\sigma_{\bar{y}_2}$)
k) s_1 ? s_2	l) P($\lvert \bar{y}_1 - \mu \rvert > 2\sigma_{\bar{y}_1}$) ≈ P($\lvert \bar{y}_2 - \mu \rvert > 2\sigma_{\bar{y}_2}$)
m) P($\bar{y}_1 > \mu + 2S$) > P($\bar{y}_2 > \mu + 2S$)	n) P($\lvert \bar{y}_1 - \mu \rvert > 2\sigma_{\bar{y}_1}$) ≈ 0,05
o) P($\bar{y}_1 > \mu + 2\sigma_{\bar{y}_1}$) ≈ 0,05	p) P($\bar{y}_1 > \mu - 2\sigma_{\bar{y}_1}$) ≈ P($\bar{y}_2 > \mu - 2\sigma_{\bar{y}_2}$)
q) P($\lvert \bar{y}_1 - \mu \rvert < 2$) < P($\lvert \bar{y}_2 - \mu \rvert < 2$)	r) P($\lvert \bar{y}_1 - \mu \rvert > 2\sigma_{\bar{y}_2}$) > 0,05
s) $\hat{\sigma}_{\bar{y}_1}$? $\sigma_{\bar{y}_1}$	t) $\hat{\sigma}_{\bar{y}_1} < s_1$
u) $\hat{\sigma}_{\bar{y}_1}$? $\hat{\sigma}_{\bar{y}_2}$	v) $\hat{\sigma}_{\bar{y}_1}$? S

Problème 6.11

a) Voici la distribution des $\bar{y}$. On y remarque une bonne symétrie :



b) i) 19,38 % ; ii) 37,675 % ; iii) 86,55 %.

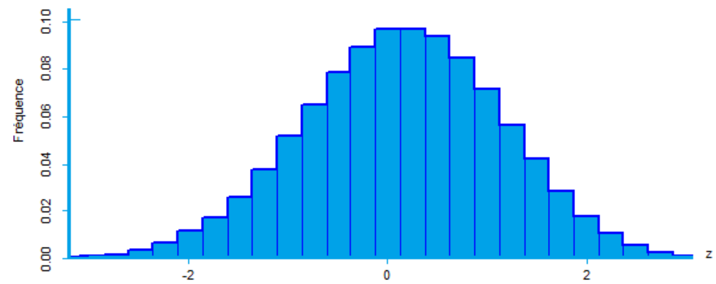
c) $\sigma_{\bar{y}} = 2010$. Théoriquement, l'écart-type est censé être égal à 2012.

d) À peu près 95,45 %, ce qui n'est pas trop éloigné de la probabilité prévue par la théorie.

e) Voici les probabilités, calculées avec les deux valeurs

	Calculs avec la loi normale		Calcul (approximatif) avec les données simulées
	$\sigma_{\bar{y}} = 2012$	$\sigma_{\bar{y}} = 2010$	
P(48500 ≤ $\bar{y}$ ≤ 49500)	0,19623	0,19642	0,19382
P(48000 ≤ $\bar{y}$ ≤ 50000)	0,38077	0,38112	0,37675
P(| $\bar{y}$ -μ|≤ 3000) ≈ P(46034,9 ≤ $\bar{y}$ ≤ 52034,9)	0,86405	0,86444	0,8655

f) Voici l'histogramme des valeurs de Z :



Voici quelques comparaisons de probabilités :

Intervalle	Probabilités d'intervalles			Probabilités cumulatives		
	Fréquence observée	Selon la loi normale	Pourcentage d'écart	Cumulative observée	Selon la loi normale	Pourcentage d'écart
-1,9 ; -1,8	0,0056	0,0072	22,4 %	0,0263	0,0359	26,8 %
-1,7; -1,6	0,0094	0,0102	8,1 %	0,0428	0,0548	21,9 %
-1,5; -1,4	0,0100	0,0139	28,3 %	0,0628	0,0808	22,2 %
-1,4; -1,3	0,0128	0,0160	20,2 %	0,0756	0,0968	21,9 %
-1,2; -1,1	0,0201	0,0206	2,4 %	0,1131	0,1357	16,6 %
-1,1; -1,0	0,0228	0,0230	0,8 %	0,1359	0,1587	14,3 %
-0,5; -0,4	0,0413	0,0360	14,6 %	0,3315	0,3446	3,8 %
-0,3; -0,2	0,0428	0,0387	10,7 %	0,4127	0,4207	1,9 %
-0,2; -0,1	0,0417	0,0394	5,8 %	0,4544	0,4602	1,3 %
-0,1; 0,0	0,0416	0,0398	4,4 %	0,4960	0,5000	0,8 %
0,1; 0,2	0,0425	0,0394	7,8 %	0,5820	0,5793	0,5 %
0,3; 0,4	0,0368	0,0375	1,9 %	0,6584	0,6554	0,5 %
0,5; 0,6	0,0334	0,0343	2,6 %	0,7282	0,7257	0,3 %
0,7; 0,8	0,0281	0,0301	6,7 %	0,7853	0,7881	0,4 %
0,9; 1,0	0,0221	0,0254	13,0 %	0,8352	0,8413	0,7 %
1,1; 1,2	0,0196	0,0206	4,8 %	0,8765	0,8849	1,0 %
1,3; 1,4	0,0132	0,0160	17,7 %	0,9065	0,9192	1,4 %
1,5; 1,6	0,0114	0,0120	5,1 %	0,9286	0,9452	1,8 %
1,9; 2,0	0,006	0,0060	0,6%	0,9634	0,9772	1,4%
2,0; 6,0	0,0366	0,0228	0,6088	1,0000	1,0000	9,9E-10

Problème 6.12

a) $\mu = 29{,}25$ et $S = 19{,}09188$.

b) On trouve en effet que la moyenne de tous les $\overline{y}$ est égale à 29,25 — la moyenne de la population; et que l'écart-type (non corrigé) de tous les $\overline{y}$ est bien 8,7142125.

c) On calcule la moyenne μ_s de tous les s. On trouve $\mu_s = 17{,}905649$, ce qui n'est pas égal à $S = 19{,}09188$. Donc s est un estimateur biaisé de S.

d) (i) 75 %; (ii) 62,5 %; (iii) 46,4286 %

e) 71,4286 %.

f) 89,2857 %. Théoriquement, cette probabilité est censée être de 95 %

g) 96,4286 %

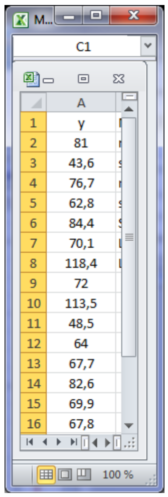
h) 96,4286 %

Commandes Excel
Intervalle de confiance

Considérons un échantillon de taille n = 15 tiré d'une population de taille N = 5000.

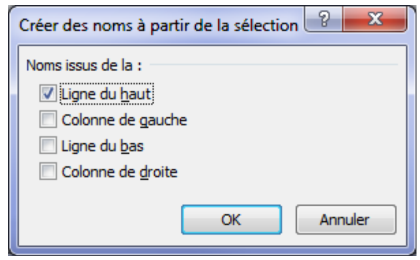
1 Inscrire les données

Désignez par `y` les observations de l'échantillon; placez-en les valeurs dans la plage `A2-A16` et sélectionnez la plage `A1–A16`.



2 Nommer la variable

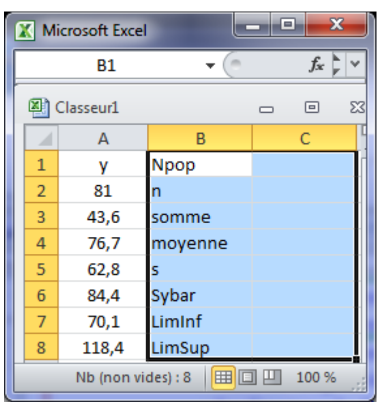
Dans l'onglet **Formules**, groupe **Noms définis**, Sélectionnez **Depuis la sélection**. Cliquez sur **Ligne du haut**.



3 Inscrire les noms des éléments à calculer

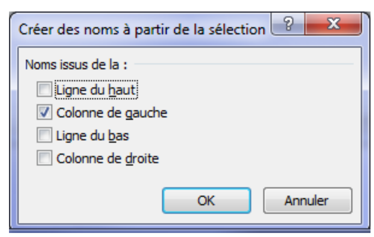
Inscrivez les noms des quantités que vous avez l'intention de calculer. Ici ils sont inscrits dans la plage `B1–B8`.

[Sybar désigne $\hat{\sigma}_{\bar{y}}$, LinInf et LimSup désignent les limites d'un intervalle de confiance pour la moyenne.]

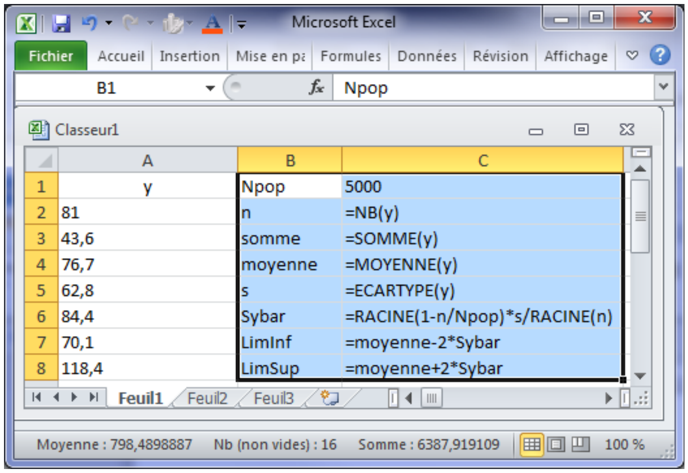


4 Définir les noms

Sélectionnez la plage `B1–C8`. Dans l'onglet **Formules**, groupe **Noms définis**, Sélectionnez **Depuis la sélection**. Cliquez sur **Colonne de gauche** :

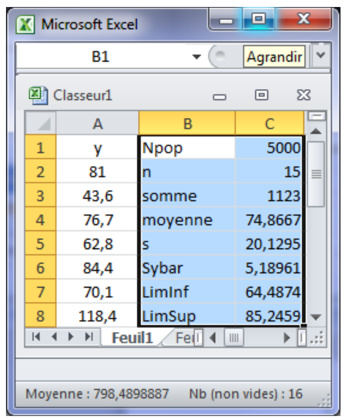


5 Écrire les formules



	A	B	C
1	y	Npop	5000
2	81	n	=NB(y)
3	43,6	somme	=SOMME(y)
4	76,7	moyenne	=MOYENNE(y)
5	62,8	s	=ECARTYPE(y)
6	84,4	Sybar	=RACINE(1-n/Npop)*s/RACINE(n)
7	70,1	LimInf	=moyenne-2*Sybar
8	118,4	LimSup	=moyenne+2*Sybar

6 Résultats



	A	B	C
1	y	Npop	5000
2	81	n	15
3	43,6	somme	1123
4	76,7	moyenne	74,8667
5	62,8	s	20,1295
6	84,4	Sybar	5,18961
7	70,1	LimInf	64,4874
8	118,4	LimSup	85,2459