Statistique descriptive : Exercices sur les sondages

1. Une population est composée de N=5 individus sur lesquels on mesure une variable X, et dont voici les valeurs :
$$\{12, 18, 9, 27, 21\}.$$
   (a) Déterminer la moyenne et la variance de X dans cette population ;
   (Réponses : $\mu = 17.4$, $\sigma^2 = 41.04$)
   (b) Combien d'échantillons de taille 2 peut-on extraire de la population ? Citer-les ;
   (Réponses : 10, voir table 1)
   (c) Pour chacun de ces échantillons, calculer la moyenne et la variance ;
   (Réponses : voir table 1)
   (d) Calculer la moyenne des moyennes ainsi que la moyenne des variances ;
   (Réponses : moyenne des moyennes = 17.4 , moyenne des variances = 51.3)
   (e) En supposant que les différents échantillons sont équiprobables, donner la distribution de l'estimateur de la moyenne de la population et calculer l'espérance de cette moyenne, son biais et son erreur quadratique moyenne ;
   (Réponses : distribution voir table 1, $E[\hat{\bar{Y}}] = 17.4$, biais = 0, $V[\hat{\bar{Y}}] = 15.37$, $EQM[\hat{\bar{Y}}] = 15.37$)
   (f) En supposant maintenant que les probabilités p_i des échantillons ne contenant pas l'individu $n°2$ valent $\frac{6}{40}$ tandis que les autres valent $\frac{1}{40}$, recalculer ces différentes valeurs.
   (Réponses : distribution voir table 1, $E[\hat{\bar{Y}}] = 17.29$, biais = -0.11, $V[\hat{\bar{Y}}] = 16.65$, $EQM[\hat{\bar{Y}}] = 16.66$)

Échantillon	Moyenne	Variance	Dist.(e)	Dist.(f)
{9,12}	10.5	4.5	0.1	6/40
{9,18}	13.5	40.5	0.1	1/40
{9,21}	15	72	0.1	6/40
{9,27}	18	162	0.1	6/40
{12,18}	15	18	0.1	1/40
{12,21}	16.5	40.5	0.1	6/40
{12,27}	19.5	112.5	0.1	6/40
{18,21}	19.5	4.5	0.1	1/40
{18,27}	22.5	40.5	0.1	1/40
{21,27}	24	18	0.1	6/40

Table 1 – Exercice 1

2. Considérons une population de quatre sujets sur lesquels on mesure une variable X dont voici les valeurs :
$$\{6.2, 9.3, 9.9, 10.7\}.$$
   (a) Déterminer la moyenne et la variance dans cette population ;
   (Réponses : $\bar{X} = 9.025$, $\sigma^2 = 2.91$)

(b) Combien d'échantillons de taille 3 peut-on extraire de la population ? Citer-les ; (Réponse : Voir table 2)

(c) Pour chacun de ces échantillons, calculer la moyenne et la variance ; (Réponse : Voir table 2)

(d) Calculer la moyenne des moyennes ainsi que la moyenne des variances ; (Réponses : moyenne des moyennes = 9.025 , moyenne des variances = 2.58)

(e) En supposant que les différents échantillons sont équiprobables, donner la distribution de l'estimateur de la moyenne de la population et calculer l'espérance de cette moyenne, son biais et son erreur quadratique moyenne ;
(Réponses : distribution voir table 2, $E[\hat{\bar{Y}}] = 9.025$, biais = 0, $V[\hat{\bar{Y}}] = 0.324$, $EQM[\hat{\bar{Y}}] = 0.324$)

(f) En supposant maintenant que l'individu $n°4$ soit moins disponible que les autres pour l'enquête, recalculer ces valeurs si la probabilité de l'échantillon dans lequel il n'apparait pas est deux fois plus grande que pour les autres échantillons.
(Réponses : distribution voir table 2, $E[\hat{\bar{Y}}] = 8.914$, biais = 0, $V[\hat{\bar{Y}}] = 0.309$, $EQM[\hat{\bar{Y}}] = 0.321$)

Échantillon	Moyenne	Variance	Dist.(e)	Dist.(f)
{1,2,3}	8.47	2.57	0.25	0.4
{1,2,4}	8.73	3.59	0.25	0.2
{1,3,4}	8.93	3.9	0.25	0.2
{2,3,4}	9.97	0.26	0.25	0.2

TABLE 2 – Exercice 2

3. Les étudiants en master en statistique forment une population de quatre sujets. Soit la variable Y étant le résultat à l'un des examens qu'ils ont en commun. Les cotes obtenues (sur 20) sont {19,17,13,15}.

(a) Déterminer la moyenne et la variance dans cette population ; (Réponses : $\bar{Y} = 16$ et $\sigma^2 = 5$)

(b) Combien d'échantillons de taille 3 peut-on extraire de la population ? Citer-les ; (Réponses : voir table 3)

(c) Pour chacun de ces échantillons, calculer la moyenne et la variance ; (Réponses : voir table 3)

(d) En supposant que les différents échantillons sont équiprobables, donner la distribution de l'estimateur de la moyenne de la population et calculer l'espérance de cette moyenne, son biais et son erreur quadratique moyenne ;
(Réponses : distribution voir table 3, $E[\hat{\bar{Y}}] = 16$, biais = 0, $V[\hat{\bar{Y}}] = 0.5546$, $EQM[\hat{\bar{Y}}] = 0.5546$)

(e) L'individu $n°3$ étant Erasmus, il fréquente moins souvent le bâtiment où se donnent les cours. On a ainsi moins de chance de le croiser dans les couloirs et donc moins de chance de le recenser que les trois autres. Recalculer les valeurs de l'exercice précédent si la probabilité de l'échantillon dans lequel il n'apparaît pas est deux fois plus grande que pour les autres échantillons.

(Réponses : distribution voir table 3, $E[\hat{\hat{Y}}] = 16.2$, biais $= 0.2$, $V[\hat{\hat{Y}}] = 0.60356$, $EQM[\hat{\hat{Y}}] = 0.64356$)

Échantillon	Moyenne	Variance	Dist.(e)	Dist.(f)
{1,2,3}	16.33	6.33	0.25	0.2
{1,2,4}	17	2.67	0.25	0.4
{1,3,4}	15.67	6.12	0.25	0.2
{2,3,4}	15	2.676	0.25	0.2

Table 3 – Exercice 3

4. La vitesse des voitures à l'entrée du tunnel de Cointe a été mesurée durant 1 heure. Le nombre de véhicules concernés est de 250. Voici un échantillon, obtenu par sondage aléatoire simple, des 250 vitesses mesurées :

$$\{78, 81, 95, 87, 108, 91, 89, 80, 98, 97, 72, 88, 75, 92, 89\}.$$

   (a) Quel est le taux de sondage ?
   (Réponse : $f = 0.06$)
   (b) Estimer (sans biais) la moyenne des vitesses des 250 automobilistes ;
   (Réponse : $\hat{\bar{Y}} = 88$)
   (c) Estimer la variance de cet estimateur ;
   (Réponse : $\hat{V}(\hat{\bar{Y}}) = 5.8$)
   (d) Donner un intervalle de confiance à 95% pour la vitesse moyenne des 250 automobilistes ;
   (Réponse : $IC(\bar{Y}) = [83.18, 92.82]$)
5. Le gérant d'une petite station-service étudie le prix que paie chaque automobiliste à l'unique pompe de sa station. Il estime à 100 le nombre de véhicules faisant le plein chez lui et il a relevé les montants payés par 4 d'entre eux : {55, 42, 73, 20}.
   (a) Calculer le taux de sondage ;
   (Réponse : $f = 0.04$)
   (b) Estimer la moyenne du prix payé par les automobilistes à cette station-service ;
   (Réponse : $\hat{\bar{Y}} = 47.5$ euros)
   (c) Estimer la variance de l'estimation donnée juste avant ;
   (Réponse : $\hat{V}(\hat{\bar{Y}}) = 119.44$)
   (d) Donner une intervalle de confiance à 95% pour la moyenne du prix payé par les automobilistes à cette station-service.
   (Réponse : $IC(\bar{Y}) = [25.64, 69.36]$)
6. Les tailles d'enfants de la crèche du Sart-Tilman ont été mésurées (N=40). Voici un échantillon (constitué par sondage aléatoire simple) des tailles (en cm) observées

$$\{76.1, 77.0, 78.1, 78.2, 78.8, 79.7, 79.9, 81.1, 81.2, 81.8, 82.8, 83.5\}.$$

   (a) Quel est le taux de sondage ?
   (Réponse : $f = 0,3$)

(b) Estimer (sans biais) la moyenne des tailles ;
(Réponse : $\hat{\bar{Y}} = 79.85$)

(c) Estimer la variance de l'estimateur précédent ;
(Réponse : $\hat{V}(\hat{\bar{Y}}) = 0.31$)

(d) Donner un intervalle de confiance à 95% pour la taille moyenne des enfants de cette crèche.
(Réponse : $IC(\bar{Y}) = [78.74; 80.96]$)

7. Lors d'une vaste enquête sur les prix d'achat, la ville de Liège souhaite connaître le prix d'un litre de soda au cola. La liste complète de tous les distributeurs liégeois compte 5400 magasins (des grandes surfaces au "night-shop"). Pour réaliser ce sondage, un échantillon de 135 magasins est déterminé. Sur celui-ci, le prix moyen observé pour un litre de soda au cola est de 1.5 euros avec une variance de 0.25 euros^2.

(a) Quel est le taux de sondage ?
(Réponse : $f = 0.025$)

(b) Estimer la variance et l'erreur quadratique moyenne de l'estimateur proposé pour la moyenne de population ;
(Réponses : $\hat{V}[\bar{y}] = 0.0018 = \mathrm{EQM}[\bar{y}]$)

(c) Donner un intervalle de confiance à 95% pour le prix d'un litre de soda au cola en région liégeoise ;
(Réponse : $IC(\bar{Y}) = [1.42; 1.58]$)

(d) Sachant que le prix maximal conseillé est de 1.56 euros, est-il plausible que le prix moyen sur les 5400 distributeurs excède cette limite ? Et si la limite est de 1.6 euros ?
(Réponse : L'intervalle de confiance calculé contient la vraie valeur du prix moyen avec 95% de chance. Il est donc plausible que ce prix moyen dépasse 1.56. Par contre, il est moins probable que 1.6 soit dépasser.)

8. Un sondage aléatoire simple réalisé parmi les étudiants d'une université a révélé que 879 des 1690 personnes interrogées fument. En supposant que cette université compte 10 000 étudiants,

(a) Calculer le taux de sondage ;
(Réponse : 0.169)

(b) Estimer la proportion de fumeurs au sein des étudiants de l'université ;
(Réponse : $p = 0.52$)

(c) Estimer la variance et l'erreur quadratique moyenne pour cet estimateur ;
(Réponses : $\hat{V(p)} = 0.000123 = \mathrm{EQM}(p)$)

(d) Donner un intervalle de confiance à 95% pour la proportion d'étudiants fumeurs dans cette université.
(Réponse : $IC(\pi) = [0.498, 0.542]$)

9. Comme chaque année, le Recteur de l'université de Liège s'intéresse au taux de réussite en premier bachelier dans son université. Les étudiants concernés étant estimés à 5000, il est plus rapide pour lui d'effectuer un sondage afin de donner une estimation de ce taux de réussite. Il choisit alors 200 étudiants. Parmi ceux-ci, 60 seulement ont réussi.

(a) Quel est le taux de sondage ?
(Réponse : $f = 0.04$)
(b) Estimer la proportion d'étudiants de l'université de Liège ayant réussi leur année ;
(Réponse : $p = 0.3$)
(c) Estimer la variance et l'erreur quadratique moyenne pour cet estimateur ;
(Réponse : $\hat{V}[p] = 0.001008 = \text{EQM(p)}$)
(d) Donner un intervalle de confiance à 95% pour la proportion d'étudiants de l'université de Liège ayant réussi leur année.
(Réponse : $IC(\pi) = [0.24, 0.36]$)

Comme le taux de réussite varie entre les différentes facultés, il choisit ensuite 200 étudiants répartis dans les grandes facultés : sciences du vivant, sciences humaines, sciences exactes et sciences de la matière. Voici comment se répartissent les étudiants de premier bachelier dans ces facultés : il y en a 1502 en sciences du vivant, 1202 en sciences humaines, 925 en sciences exactes et 1371 en sciences de la matière. Dans chacune de ces facultés, il choisit respectivement 65, 40, 37 et 58 étudiants pour réaliser le sondage. Les estimations obtenues pour le taux de réussite dans chaque faculté sont : 0.4 en sciences du vivant, 0.2 en sciences humaines, 0.3 en sciences exactes et 0.28 en sciences de la matière.

(a) Quel est le taux de sondage dans chaque strate ?
(Réponses : dans le même ordre que précédemment, $f_1 = 0.043$, $f_2 = 0.033$, $f_3 = 0.04$ et $f_4 = 0.042$)
(b) Si le Recteur avait opté pour un sondage à allocation proportionnelle, combien d'étudiants aurait-il dû interroger dans chaque strate ?
(Réponses : dans le même ordre que précédemment, $n_1 = 60$, $n_2 = 48$, $n_3 = 37$ et $n_4 = 55$)
(c) Estimer, sans biais, la proportion d'étudiants de premier bachelier ayant réussi leur année ;
(Réponse : $p_{\text{st}} = 0.301$)
(d) Calculer la variance de cet estimateur ;
(Réponse : $\hat{V}[p_{\text{st}}] = 0.000979$)
(e) Donner un intervalle de confiance à 95% pour la proportion d'étudiants de premier bachelier ayant réussi leur année.
(Réponse : $IC(\pi) = [0.238, 0.364]$)

10. Le volume (en cm^3) de la langue dans une population de patients du CHU a été mesuré par imagerie par raisonnance magnétique chez 130 femmes et 150 hommes (que l'on peut assimiler à des échantillons aléatoires simples des sous-populations respectives). La moyenne et la variance de ces mesures sont reprises dans le tableau 4.

Population		Echantillon		
Groupe	Effectif	Effectif	Moyenne	Variance
Femmes	800	130	70.15	57.46
Hommes	900	150	90.17	106.3

TABLE 4 – Exercice 10

(a) Déterminer un intervalle de confiance à 95% pour le volume moyen de la langue dans chaque sous-population.
(Réponses : IC(femmes)=[68.93; 71.37], IC(hommes)=[88.63; 91.71])

(b) Déterminer un intervalle de confiance à 95% pour le volume moyen de la langue dans cette population de 1700 patients.
(Réponse : $IC(\bar{Y}) = [79.76; 81.74]$)

11. On a interrogé un certain nombre de belges à propos d'une éventuelle séparation de la Belgique. Pour effectuer ce sondage, les strates "Wallonie", "Flandre" et "Bruxelles" ont été considérées. Ce sondage a touché 3333 wallons, 5927 flamands et 955 bruxellois. Les résultats obtenus sont :

Région	En faveur de la séparation
Wallons	15%
Flamands	40%
Bruxellois	20%

(a) Sachant qu'une allocation proportionnelle a été considérée et que le taux de sondage est de 0.001, combien de bruxellois y a-t-il en Belgique ?
(Réponses : $N_B = 955000$, $N_W = 3333000$ et $N_F = 5927000$)

(b) Estimer, sans biais, la proportion de belges favorables à la séparation de la Belgique ;
(Réponse $= p_{st} = 0.3$)

(c) Calculer la variance de cet estimateur, ainsi que son erreur quadratique moyenne ;
(Réponses : $\hat{V}(p_{st}) = 0.00001913 = \text{EQM}(p_{st})$)

(d) Donner un intervalle de confiance à 95% pour la proportion de belges favorables à la séparation.
(Réponse : $IC(\pi) = [0.291, 0.309]$

12. Lors des dernières élections présidentielles en Belgovie (1200000 électeurs), deux candidats s'affrontaient lors du dernier tour : A et B. Un sondage (aléatoire simple) a été réalisé auprès de 3000 électeurs : 63% de ces personnes ont exprimé une préférence pour le candidat A.

(a) Calculer le taux de sondage ; que mesure-t-il ?
(Réponse : $f = 0.0025$)

(b) Estimer (sans biais) la proportion de personnes favorables au candidat A ;
(Réponse : $p_A = 0.63$)

(c) Estimer (sans biais) la proportion de personnes favorables au candidat B ;
(Réponse : $p_B = 0.37$)

(d) Estimer la variance de ces deux estimateurs. Que constatez-vous ? Est-ce normal ?
(Réponses : $\hat{V}(p_A) = \hat{V}(p_B) = 0.0000775$, *vu la formule de la variance, cette égalité est normale*)

(e) Donner un intervalle de confiance à 95% pour chacune de ces proportions.
(Réponses : $IC(\pi_A) = [0,61; 0,65]$ et $IC(\pi_B) = [0,35; 0,39]$)

Ce pays est constitué de trois régions comptant respectivement 350 000, 600 000 et 250 000 électeurs. Un autre sondage (cette fois) à allocation proportionnelle a été réalisé sur un total de 3000 électeurs. Les opinions favorables au candidat A dans chacune des régions représentent, respectivement, 40%, 67% et 53% des personnes interrogées.

(a) Calculer le nombre de personnes interrogées dans chaque région ;
(Réponses : $n_1 = 875$, $n_2 = 1500$, $n_3 = 625$)

(b) Quel est le taux de sondage ?
(Réponse : $f = 0.0025$)

(c) Quelle est, dans la population globale, l'estimation (sans biais) de la proportion d'électeurs favorables au candidat A ?
(Réponse : $p_{st} = 0.562$)

(d) Estimer la variance de cet estimateur ;
(Réponse : $\hat{V}(p_{st}) = 0,0000773$)

(e) Donner un intervalle de confiance à 95% pour la proportion de personnes favorables à A.
(Réponse : $IC(\pi) = [0.54; 0.58]$)

13. Un sondage concernant l'activité sportive des membres d'un club de sport a été réalisé. Ces membres ont été classés en 4 catégories d'âges : les moins de 25 ans (300), les 25-40 ans (220), les 40-60 ans (130) et les plus de 60 ans (170). Il a été demandé à 25 personnes choisies (par sondage aléatoire simple) dans chaque classe d'âges si elles pratiquaient plus de 2 heures de sport par semaine. Les résultats sont repris dans le tableau ci-dessous.

	- de 25 ans	25 - 40 ans	40 - 60 ans	+ de 60 ans
plus de 2 heures	64%	52%	49%	68%

Table 5 – Exercice 13

(a) Donner un intervalle de confiance à 95% pour la proportion de membres actifs (pratiquant plus de 2 heures par semaine) dans ce club ?
(Réponses : $f_1 = 0.083$, $f_2 = 0.114$, $f_3 = 0.192$, $f_4 = 0.147$, $p_{st} = 0.592$, $\hat{V}(p_{st}) = 0.00229$, $IC(\pi) = [0.497, 0.688]$).

(b) Si les pourcentages obtenus dans le tableau 5 l'avaient été dans le cadre d'un sondage stratifié à allocation proportionnelle (avec toujours 100 personnes sondées au total), que vaudrait cet intervalle ?
Arrondissez le nombre de personnes interrogées par strate correctement pour conserver l'allocation proportionnelle quitte à interroger approximativement 100 personnes.
(Réponses : $n_1 = 37$, $n_2 = 27$, $n_3 = 16$, $n_4 = 21$, $n = 101$, $p_{st} = 0.592$, $\hat{V}(p_{st}) = 0.00205$, $IC(\pi) = [0.502, 0.683]$)

14. Dans une forêt, trois types d'arbres sont rencontrés : des hêtres (260), des chênes (210) et des sapins (180). La hauteur de certains d'entre eux a été mésurée et répertoriée dans la table ci-dessous.

(a) Calculer le taux de sondage dans chaque strate ;
(Réponses : voir table 7)

(b) Estimer (sans biais) la moyenne et la variance de la hauteur des arbres dans chaque strate ;
(Réponses : voir table 7)

(c) Estimer (sans biais) la hauteur moyenne des arbres dans la forêt ;
(Réponse : $\hat{\bar{Y}}_{st} = 24.934$

Hêtres	Chênes	Sapins
23,4	22,5	18,9
24,4	22,9	21,1
24,6	23,7	,21,2
24,9	24	22,1
25	24,4	22,5
26,2	24,5	23,6
26,3	25,3	24,5
26,8	26	24,6
26,8	26,2	26,2
26,9	26,4	26,7
27	26,7	
27,6	26,9	
27,7	27,4	
	25,8	

Table 6 – Exercice 14

(d) Estimer la variance de l'estimateur précédent ;
(Réponse : $\hat{V}(\hat{\bar{Y}}_{st}) = 0.0818$)

(e) Donner un intervalle de confiance à 95% pour la hauteur moyenne des arbres de cette forêt.
(Réponse : $IC(\bar{Y}) = [24.362, 25.506]$)

Hêtres	Chênes	Sapins
$n_1 = 13$	$n_2 = 14$	$n_3 = 10$
$N_1 = 260$	$N_2 = 210$	$N_3 = 180$
$f_1 = 0.05$	$f_2 = 0.0667$	$f_3 = 0.0556$
$\bar{y}_1 = 25.97$	$\bar{y}_2 = 25.19$	$\bar{y}_3 = 23.14$
$s_1^2 = 1.803$	$s_2^2 = 2.527$	$s_3^2 = 5.958$

Table 7 – Exercice 14

15. On a mesuré la glycémie à jeun dans quatre groupes de personnes adultes choisies au hasard (par sondage aléatoire simple) dans 4 sous-populations caractérisées par leur niveau d'activité physique ('très élevé', 'élevé', 'modéré' ou 'faible'). Les résultats sont repris dans le tableau 8. Si une allocation proportionnelle avec un taux de sondage de 0.2 est considéré, déterminer un intervalle de confiance à 95% du taux de glycémie moyen dans l'ensemble de la population.
(Réponse : $IC(\bar{Y}) = [90.11; 93.09]$; résultats intermédiaires : $\bar{y}_1 = 82.38, s_1^2 = 28.554$; $\bar{y}_2 = 91.92, s_2^2 = 7.356$; $\bar{y}_3 = 93.33, s_3^2 = 42.242$; $\bar{y}_4 = 97.75, s_4^2 = 36.214$).

16. Le comité des riverains d'un village de la région liégeoise luttant contre l'expansion du nombre de véhicules de type 4×4 dans le village a étudié la consommation de ces véhicules en les comparant à d'autres types de voitures (les urbaines et les familiales). Pour cela, un sondage a été réalisé au départ des 20 "4×4", 60 urbaines et 40 familiales que possèdent les habitants du village. Les mesures de la consommation sur 100 km

Très élevé	Élevé	Modéré	Faible
92	90	91	103
88	89	99	96
84	93	88	96
78	92	102	93
78	86	87	108
80	90	86	102
82	95	95	90
77	95	82	94
	94	96	
	93	97	
	92	102	
	94	95	

TABLE 8 – Exercice 15

faites sur l'échantillon de 12 véhicules sont données dans la table suivante :

4×4	Urbaines	Familiales
18.2	5.8	8.6
24.5	3.6	9.4
16.3	4.5	7.8
22	7.1	8.2

(a) Calculer le taux de sondage dans chaque strate ;
(Réponses : dans le même ordre que précédemment, $f_1 = 0.02$, $f_2 = 0.066$ et $f_3 = 0.1$)

(b) Calculer les effectifs des échantillons de chaque strate si l'on avait opté pour un sondage à allocation proportionnelle ;
(Réponses : dans le même ordre que précédemment, $n_1 = 2$, $n_2 = 6$ et $n_3 = 4$)

(c) Estimer (sans biais) la moyenne et la variance de la consommation sur 100 km dans chaque strate ;
(Réponses : $\bar{y}_1 = 20.25$, $\bar{y}_2 = 5.25$, $\bar{y}_3 = 8.5$, $s_1^2 = 13.643$, $s_2^2 = 2.337$ et$s_3^2 = 0.467$)

(d) Estimer (sans biais) la consommation moyenne sur 100 km des véhicules du village ;
(Réponse : $\hat{\bar{Y}}_{\text{st}} = 8.833$)

(e) Estimer la variance de l'estimateur précédent ;
(Réponse : $\hat{V}[\hat{\bar{Y}}_{\text{st}}] = 0.2239$)

(f) Donner un intervalle de confiance à 95% pour la consommation moyenne sur 100 km des véhicules du village ;
(Réponse : $\text{IC}(\bar{Y}) = [7.89; 9.78]$)

(g) Quel serait l'impact d'un doublement du nombre de véhicules dans l'échantillon sur la largeur de cet intervalle de confiance (si l'on suppose que la variance de la consommation reste inchangée) ?
(Réponse : les taux de sondage doublent, et cela avec l'augmentation des effectifs, entraine une variance plus petite et donc un intervalle plus petit, plus précis.)

17. En Belgique, le droit de vote est une obligation ... pour autant que l'on soit majeur et que l'on ne soit pas privé de ce droit de vote. Hors des 10 666 866 belges que compte notre pays, il n'est donc pas facile de connaître le pourcentage de la population s'exprimant vraiment aux élections et encore moins d'avoir une idée sur leur moyenne d'âge. Sachant d'emblée que la population belge se compose d'environ 60% de flamands, 30% de wallons et 10% de bruxellois, il paraissait moins coûteux d'échantillonner dans chacun de ces sous-groupes. C'est pourquoi 1128 flamands, 571 wallons et 99 bruxellois ont participé à cette enquête. Dans ces groupes respectifs, 944, 442 et 62 personnes ont voté lors des dernières élections. L'âge moyen vaut respectivement 46.9 ans, 45.7 ans et 41.2 ans avec des écarts-types respectifs de 18.78 ans, 18.47 ans et 17.77 ans.

    (a) Calculer le taux de sondage dans les 3 groupes ;
    (Réponses : $f_F = 0.0001762$, $f_W = 0.0001784$ et $f_B = 0.0000928$)

    (b) Dans chacune des strates, donner une estimation non-biaisée de la proportion de personnes ayant voté lors des dernières élections ;
    (Réponses : $p_F = 0.84$, $p_W = 0.77$ et $p_B = 0.63$)

    (c) Donner une estimation non-biaisée de la proportion de belges ayant voté lors des dernières élections ;
    (Réponse : $p_{\text{st}} = 0.798$)

    (d) Donner un intervalle de confiance à 95% pour la proportion de belges ayant voté lors des dernières élections ;
    (Réponses : $\hat{V}[p_{\text{st}}] = 0.0000943$ et $\text{IC}(\pi) = [0.779; 0.817]$)

    (e) Donner un intervalle de confiance à 95% pour la moyenne d'âge des wallons ;
    (Réponse : $\text{IC}(\bar{Y}_W) = [44.15; 47.25]$)

    (f) Donner une estimation non-biaisée de la moyenne d'âge en Belgique ;
    (Réponse : $\hat{\bar{Y}}_{\text{st}} = 45.97$)

    (g) Donner un intervalle de confiance à 95% pour la moyenne d'âge en Belgique.
    (Réponses : $\hat{V}[\hat{\bar{Y}}_{\text{st}}] = 0.1968$ et $\text{IC}(\bar{Y}) = [45.08; 46.86]$)