
Exercices supplémentaires sur les sondages : 2^{ème} partie

1. Même si le droit de vote est obligatoire en Belgique, il existe certaines exceptions et dérogations. En réalité, le pourcentage de la population belge ($N = 10666866$) qui s'exprime vraiment aux élections n'est donc pas connu directement. Par conséquent, une enquête a été réalisée aux dernières élections afin de se faire une meilleure idée de cette proportion et de la moyenne d'âge des électeurs belges. Pour des raisons de coûts, trois groupes ont été définis au préalable : les Flamands (60% de la population), les Wallons (30%) et les Bruxellois (10%). 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 des électeurs était respectivement de 46.9 ans, 45.7 ans et 41.2 ans dans les trois groupes avec des écarts-types respectifs de 18.78 ans, 18.47 ans et 17.77 ans.
- Calculer le taux de sondage dans les 3 groupes ; (rép : $f_F = 0.0001762$; $f_W = 0.0001784$; $f_B = 0.0000928$)
 - Dans chacune des strates, donner une estimation non-biaisée de la proportion de personnes ayant voté lors des dernières élections ; (rép : $p_F = 0.84$; $p_W = 0.77$; $p_B = 0.63$)
 - Donner une estimation non-biaisée de la proportion de Belges ayant voté lors des dernières élections ; (rép : $p_{st} = 0.798$)
 - Donner un intervalle de confiance à 95% pour la proportion de Belges ayant voté lors des dernières élections ; (rép : $IC(\pi_{st}) = [0.779 ; 0.817]$)
 - Donner un intervalle de confiance à 95% pour la moyenne d'âge des Wallons ; (rép : $IC(\bar{Y}_W) = [44.15 ; 47.25]$)
 - Donner une estimation non-biaisée de la moyenne d'âge en Belgique ; (rép : $\bar{y}_{st} = 45.97$)
 - Donner un intervalle de confiance à 95% pour la moyenne d'âge en Belgique. (rép : $IC(\bar{Y}_{st}) = [45.08 ; 46.86]$)
2. Depuis l'an dernier, la faculté de Sciences de l'Université de Liège a lancé le "tronc commun" regroupant tous les étudiants de premier bachelier de la faculté (hormis les étudiants en mathématiques) pour les 6 premiers mois d'enseignement. Ce n'est qu'après la session de janvier que les étudiants doivent choisir "leur" science (biologie, chimie, géologie, géographie ou physique). Ce choix d'opter pour le "tronc commun" a été posé suite à une étude réalisée au début de l'année académique 2007-2008 auprès de 20% des étudiants du moment. Cette année-là, on comptait 342 inscrits en premier bachelier à la faculté des Sciences (sans les étudiants de mathématiques) qui se répartissaient dans les différentes sections comme suit : 124 en biologie, 58 en chimie, 33 en géologie, 70 en géographie et 57 en physique. Une des questions du sondage concernait la participation à une interrogation sur les prérequis organisée chaque année durant la première semaine de cours. Les résultats variant d'une section à l'autre, un sondage stratifié est préférable ; 20% des étudiants de chaque section ont donc été sélectionnés.
- Voici les taux de participation au test **dans l'ordre alphabétique des sections** : 0.88, 0.84, 0.73, 0.93 et 0.68.
- Voici également les moyennes (et écart-types) des résultats obtenus : 12.95 (4.41), 13.63 (4.22), 13.08 (3.80), 13.38 (3.67), 17.54 (4.40).

- a. Quel est le nombre d'étudiants sondés dans chaque strate ? (rép : $f_1 = 25$; $f_2 = 12$; $f_3 = 6$; $f_4 = 14$; $f_5 = 11$)
- b. Estimer la proportion d'étudiants prenant part à l'interrogation en début d'année académique ; (rép : $p_{st} = 0.84$)
- c. Donner un intervalle de confiance à 95% pour cette proportion ; (rép : $IC(\pi_{st}) = [0.76 ; 0.92]$)
- d. Donner un intervalle de confiance à 95% pour la moyenne du résultat obtenu lors du test. (rép : $IC(\bar{Y}_{st}) = [13.02 ; 14.84]$)
3. 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 Wallons, de Flamands et de Bruxellois y a-t-il en Belgique ? (rép : $N_W = 3.333.000$; $N_F = 5.927.000$; $N_B = 955.000$)
- b. Estimer, sans biais, la proportion de Belges favorables à la séparation de la Belgique ; (rép : $p_{st} = 0.3$)
- c. Calculer la variance de cet estimateur, ainsi que son erreur quadratique moyenne ; (rép : $V(p_{st}) = 0.00001913 = EQM$)
- d. Donner un intervalle de confiance à 95% pour la proportion de Belges favorables à la séparation. (rép : $IC(\pi_{st}) = [0.291 ; 0.309]$)
4. 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 60 ans et plus (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.

	< 25 ans	[25 ; 40 ans [[40 ; 60ans [≥ 60 ans
>2 h de sport/semaine	64%	52%	49%	68%

- a. Donner un intervalle de confiance à 95% pour la proportion de membres actifs (pratiquant plus de 2 heures par semaine) dans ce club ? (rép : $IC(\pi_{st}) = [0.497 ; 0.688]$)
- b. Si les pourcentages obtenus dans le tableau l'avaient été dans le cadre d'un sondage stratifié à allocation proportionnelle (avec toujours 100 personnes sondées au total), que vaudrait cet intervalle ? (rép : $IC(\pi_{st}) = [0.502 ; 0.683]$)

5. On souhaite connaître les revenus moyens des ménages belges afin de baser les discussions autour du pouvoir d'achat sur des chiffres concrets. Étant donné que les revenus sont susceptibles de varier fortement en fonction du nombre de personnes actives au sein du ménage, une enquête est organisée dans les trois groupes suivants :
- Aucune personne active au sein du ménage (23%);
 - 1 personne active au sein du ménage (33%);
 - 2 personnes actives ou plus au sein du ménage (44%).

En moyenne, plus il y a de personnes actives au sein du ménage, plus il est difficile pour l'enquêteur de les atteindre (le nombre de tentatives de contacts avant d'avoir un contact fructueux est plus important). C'est pourquoi on estime le coût moyen au sein des trois groupes de, respectivement, 15€, 25€ et 30€ par personne interrogée. Sachant que l'on dispose d'un budget de 30000€, combien de personnes va-t-on pouvoir interroger dans chacun des groupes ?

(rép : $n_1 = 277$; $n_2 = 397$; $n_3 = 530$)

6. Le comité des riverains d'un village de la région liégeoise luttant contre l'expansion du nombre de véhicules de type 4x4 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 "4x4", 60 urbaines et 40 familiales que possèdent les habitants du village. Les mesures de la consommation sur 100 km faites sur l'échantillon de 12 véhicules sont données dans la table suivante :

4x4 (strate 1)	Urbaines (strate 2)	Familiales (strate 3)
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 ; (rép : $f_1 = 0.2$; $f_2 = 0.066$; $f_3 = 0.1$)
- b. Calculer les effectifs des échantillons de chaque strate si l'on avait opté pour un sondage à allocation proportionnelle ; (rép : $n_1 = 2$; $n_2 = 6$; $n_3 = 4$)
- c. Estimer (sans biais) la moyenne et la variance de la consommation sur 100 km dans chaque strate ; (rép : $\bar{y}_1 = 20.25$; $\bar{y}_2 = 5.25$; $\bar{y}_3 = 8.5$; $\hat{\sigma}_1^2 = 10.2325$ et $s_1^2 = 13.643$; $\hat{\sigma}_2^2 = 1.7524$ et $s_2^2 = 2.337$; $\hat{\sigma}_3^2 = 0.35$ et $s_3^2 = 0.467$)
- d. Estimer (sans biais) la consommation moyenne sur 100 km des véhicules du village ; (rép : $\bar{y}_{st} = 8.833$)
- e. Estimer la variance de l'estimateur précédent ; (rép : $V(\bar{y}_{st}) = 0.2239$)
- f. Donner un intervalle de confiance à 95% pour la consommation moyenne sur 100km des véhicules du village ; (rép : IC (\bar{Y}_{st}) = [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) ? (rép : le taux de sondage a doublé → l'IC plus précis).

7. 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 suivant :

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	

Si une allocation proportionnelle avec un taux de sondage de 0.2 est considérée, déterminer un intervalle de confiance à 95% du taux de glycémie moyen dans l'ensemble de la population. (rép : $IC(\bar{Y}_{st}) = [90.11 ; 93.09]$)

8. 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 mesurée et répertoriée ci-dessous :

Hêtres (strate 1)	Chênes (strate 2)	Sapins (strate 3)
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	

- Calculer le taux de sondage dans chaque strate ; (rép : $f_1 = 0.05 ; f_2 = 0.0667 ; f_3 = 0.0556$)
- Estimer (sans biais) la moyenne et la variance de la hauteur des arbres dans chaque strate ; (rép : $\bar{y}_1 = 25.97 ; \bar{y}_2 = 25.19 ; \bar{y}_3 = 23.14 ; s^2_1 = 1.803 ; s^2_2 = 2.527 ; s^2_3 = 5.958$)
- Estimer (sans biais) la hauteur moyenne des arbres dans la forêt ; (rép : $\bar{y}_{st} = 24.934$)
- Estimer la variance de l'estimateur précédent ; (rép : $V(\bar{y}_{st}) = 0.0818$)
- Donner un intervalle de confiance à 95% pour la hauteur moyenne des arbres de cette forêt. (rép : $IC(\bar{Y}_{st}) = [24.362 ; 25.506]$).