Exercices supplémentaires sur les sondages : $1^{ère}$ partie

Corrigés des exercices 2 et 4 du TP5 (non vus en classe)

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 ; (rép : $\bar{Y}$ = 17.4 ; σ^2 = 41.04)

b. Combien d'échantillons de taille 2 peut-on extraire de la population ? Citer-les ; (K = 10)

c. Pour chacun de ces échantillons, calculer la moyenne et la variance ;

Échantillon : s_k	Moyenne : $\bar{y}(s_k)$	Variance : $\hat{\sigma}^2(s_k)$
s_1 = {12 ; 18}	$\bar{y}_{(s1)}$ = 15	$\hat{\sigma}^2_{(s1)}$ = 9
s_2 = {12 ; 9}	$\bar{y}_{(s2)}$ = 10.5	$\hat{\sigma}^2_{(s2)}$ = 2.25
s_3 = {12 ; 27}	$\bar{y}_{(s3)}$ = 19.5	$\hat{\sigma}^2_{(s3)}$ = 56.25
s_4 = {12 ; 21}	$\bar{y}_{(s4)}$ = 16.5	$\hat{\sigma}^2_{(s4)}$ = 20.25
s_5 = {18 ; 9}	$\bar{y}_{(s5)}$ = 13.5	$\hat{\sigma}^2_{(s5)}$ = 20.25
s_6 = {18 ; 27}	$\bar{y}_{(s6)}$ = 22.5	$\hat{\sigma}^2_{(s6)}$ = 20.25
s_7 = {18 ; 21}	$\bar{y}_{(s7)}$ = 19.5	$\hat{\sigma}^2_{(s7)}$ = 2.25
s_8 = {9 ; 27}	$\bar{y}_{(s8)}$ = 18	$\hat{\sigma}^2_{(s8)}$ = 81
s_9 = {9 ; 21}	$\bar{y}_{(s9)}$ = 15	$\hat{\sigma}^2_{(s9)}$ = 36
s_{10} = {27 ; 21}	$\bar{y}_{(s10)}$ = 24	$\hat{\sigma}^2_{(s10)}$ = 9

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 ; (rép : $p(s_k)$ = 1/10 ;

$\bar{y}(s_k)$	15	10.5	19.5	16.5	13.5	22.5	19.5	18	15	24
$p(s_k)$	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

E($\bar{y}$) = 17.4 ; Biais = 0 ; EQM =15.39)

e. En supposant maintenant que les probabilités p_i des échantillons ne contenant pas l'individu n°2 valent 6/40 tandis que les autres valent 1/40, recalculer ces différentes valeurs.

$\bar{y}(s_k)$	15	10.5	19.5	16.5	13.5	22.5	19.5	18	15	24
$p(s_k)$	0.025	0.15	0.15	0.15	0.025	0.025	0.025	0.15	0.15	0.15

E($\bar{y}$) = 17.29 ; Biais = -0.11 ; EQM = 16.66)

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" sciences (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). Une des questions du sondage concernait la participation à une interrogation sur les prérequis organisée chaque année durant la première semaine de cours. Il s'est avéré que 57 des étudiants sondés avaient participé à cette interrogation. La moyenne du score obtenu était de 14.17 sur 20 avec une variance de 21.09.

a. Quel est le taux de sondage ? Quelle est la taille de l'échantillon ? (Remarque : dans un échantillon, il y a toujours un nombre entier d'individus) il faudra probablement arrondir la valeur trouvée). (rép : $f = 0.2;\ \mathrm{n} = 68$)
b. Estimer la proportion d'étudiants prenant part à l'interrogation en début d'année académique ; (rép : p = 0.84)
c. Donner un intervalle de confiance à 95% pour cette proportion ; (rép : IC (π) = [0.76 ; 0.92]
d. Donner un intervalle de confiance à 95% pour la moyenne (sur 20) du résultat obtenu lors du test. (rép : IC $(\bar{Y})$ = [13.17 ; 15.17])

Exercices supplémentaires avec les corrigés

1. Les 4 pandas du zoo de Vienne forment une population de quatre sujets. Soit Y la variable donnant leur poids et {98.7, 102.6, 108.5, 120.3} les valeurs observées.
   a. Déterminer la moyenne et la variance dans cette population ; (rép : $\bar{Y}$ = 107.53 ; σ^2 = 66.57187)
   b. Combien d'échantillons de taille 3 peut-on extraire de la population ? Citer-les; (rép : K = 4)
   c. Pour chacun de ces échantillons, calculer la moyenne et la variance ;

Échantillon : s_k	Moyenne : $\bar{y}(s_k)$	Variance : $\hat{\sigma}^2(s_k)$
s_1 = {102.6 ; 108.5 ; 120.3}	$\bar{y}_{(s1)}$ = 110.47	$\hat{\sigma}^2_{(s1)}$ = 53.41
s_2 = {98.7 ; 108.5 ; 120.3}	$\bar{y}_{(s2)}$ = 109.17	$\hat{\sigma}^2_{(s2)}$ = 77.25
s_3 = {98.7 ; 102.6 ; 120.3}	$\bar{y}_{(s3)}$ = 107.2	$\hat{\sigma}^2_{(s3)}$ = 88.34
s_4 = {98.7 ; 102.6 ; 108.5}	$\bar{y}_{(s4)}$ = 103.27	$\hat{\sigma}^2_{(s4)}$ = 15.54

   d. En supposant que les différents échantillons soient 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 ; (rép : $p(s_k)$ = 1/4 ;

$\bar{y}(s_k)$	110.47	109.17	107.2	103.27
$p(s_k)$	0.25	0.25	0.25	0.25

   $E(\bar{y})$ = 107.53 ; Biais = 0 ; EQM = 7.397)

   e. Les pandas femelles se laissent plus difficilement attraper que les mâles. La population étant constituée de trois mâles et d'une femelle, l'échantillon masculin a une probabilité qui vaut 2/5 contre 1/5 seulement pour les échantillons contenant la femelle. Donner alors 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.

$\bar{y}(s_k)$	110.47	109.17	107.2	103.27
$p(s_k)$	0.4	0.2	0.2	0.2

   $E(\bar{y})$ = 108.116 ; Biais = 0.591 ; EQM = 7.65)

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 ; (rép : $\bar{Y}$ = 9.025 ; σ^2 = 2.91)
   b. Combien d'échantillons de taille 3 peut-on extraire de la population ? Citer-les (rép : K=4)
   c. Pour chacun de ces échantillons, calculer la moyenne et la variance ;

Échantillon : s_k	Moyenne : $\bar{y}(s_k)$	Variance : $\hat{\sigma}^2(s_k)$
s_1 = {6.2 ; 9.3 ; 9.9}	$\bar{y}_{(s1)}$ = 8.47	$\hat{\sigma}^2_{(s1)}$ = 2.57
s_2 = {6.2 ; 9.3 ; 10.7}	$\bar{y}_{(s2)}$ = 8.73	$\hat{\sigma}^2_{(s2)}$ = 3.59
s_3 = {6.2 ; 9.9 ; 10.7}	$\bar{y}_{(s3)}$ = 8.93	$\hat{\sigma}^2_{(s3)}$ = 3.9
s_4 = {9.3 ; 9.9 ; 10.7}	$\bar{y}_{(s4)}$ = 9.97	$\hat{\sigma}^2_{(s4)}$ = 0.26

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 ; (rép : $p(s_k) = 1/4$;

$\bar{y}(s_k)$	8.47	8.73	8.93	9.97
$p(s_k)$	0.25	0.25	0.25	0.25

$E(\bar{y}) = 9.025$; Biais = 0 ; EQM = 0.324)

e. 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. (rép : $p(s_k) = 1/4$;

$\bar{y}(s_k)$	8.47	8.73	8.93	9.97
$p(s_k)$	0.4	0.2	0.2	0.2

$E(\bar{y}) = 8.914$; Biais = 0 ; EQM = 0.321)

3. 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}
   e. Quel est le taux de sondage ? (rép : $f = 0.06$)
   a. Estimer (sans biais) la moyenne des vitesses des 250 automobilistes ; (rép : $\bar{y} = 88$)
   b. Estimer la variance de cet estimateur ; (rép : $V(\bar{y}) = 5.8$)
   c. Donner un intervalle de confiance à 95% pour la vitesse moyenne des 250 automobilistes. (rép : IC $(\bar{Y})$ = [83.18 ; 92.82])

4. 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 ; (rép : $f = 0.04$)
   b. Estimer la moyenne du prix payé par les automobilistes à cette station-service ; (rép : $\bar{y}$ = 47.5)
   c. Estimer la variance de l'estimation donnée juste avant ; (rép : $V(\bar{y}) = 119.44$)
   d. Donner un intervalle de confiance à 95% pour la moyenne du prix payé par les automobilistes à cette station-service. (rép : IC $(\bar{Y})$ = [25.64 ; 69.36])

5. Les tailles d'enfants de la crèche du Sart-Tilman ont été mesurées (N=40). Voici un échantillon (constitué par sondage aléatoire simple) des tailles (en cm) observées : {76.1; 77; 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 ? (rép : $f = 0.3$)
   b. Estimer (sans biais) la moyenne des tailles ; (rép : $\bar{y} = 79.85$)
   c. Estimer la variance de l'estimateur précédent ; (rép : $V(\bar{y}) = 0.31$)
   d. Donner un intervalle de confiance à 95% pour la taille moyenne des enfants de cette crèche. (rép : IC $(\bar{Y})$ = [78.74 ; 80.96])

6. 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 ? (rép : $f = 0.025$)
   b. Estimer la variance et l'erreur quadratique moyenne de l'estimateur proposé pour la moyenne de population ; (rép : $V(\bar{y}) = 0.0018$; EQM = $V(\bar{y})$ car estimateur non biaisé)
   c. Donner un intervalle de confiance à 95% pour le prix d'un litre de soda au cola en région liégeoise; (rép : 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 ? (rép : il est possible que le prix moyen excède 1.56 car cette valeur fait partie de l'IC – qui a 95% de chances de contenir la vraie valeur de $\bar{Y}$. Par contre, il est moins probable que le prix moyen excède 1.6 puisque cette valeur n'est pas comprise dans l'intervalle).

7. 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 10000 étudiants,
   a. Calculer le taux de sondage ; (rép : $f = 0.169$)
   b. Estimer la proportion de fumeurs au sein des étudiants de l'université ; (rép : p = 0.52)
   c. Estimer la variance et l'erreur quadratique moyenne pour cet estimateur ; (rép : V(p) = 0.000123 = EQM)
   d. Donner un intervalle de confiance à 95% pour la proportion d'étudiants fumeurs dans cette université. (rép : IC (π) = [0.498 ; 0.542])