Statistique descriptive : TP n°5 (sondages)

1. 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 ;
   b. Combien d'échantillons de taille 3 peut-on extraire de la population ? Citer-les;
   c. Pour chacun de ces échantillons, calculer la moyenne et la variance ;
   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 ;
   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.

2. 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 ;
   b. Combien d'échantillons de taille 2 peut-on extraire de la population ? Citer-les ;
   c. Pour chacun de ces échantillons, calculer la moyenne et la variance ;
   d. Calculer la moyenne des moyennes ainsi que la moyenne des variances ;
   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 ;
   f. 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.

3. Le magazine Touring s'intéresse au taux de réussite à la partie pratique du permis de conduire (premier essai) des jeunes entre 18 et 25 ans. Parmi les 1 336 099 jeunes ayant déjà tenté au moins une fois cet examen, un échantillon de taille 300 000 est prélevé. Il s'avère que 59 786 d'entre eux ont eu leur permis du premier coup.
    a. Quel est le taux de sondage ?
    b. Estimer la proportion de jeunes entre 18 et 25 ans ayant réussi l'examen pratique du permis de conduire à la première tentative ;
    c. Estimer la variance et l'erreur quadratique moyenne de cet estimateur ;
    d. Donner un intervalle de confiance à 95% pour la proportion de jeunes entre 18 et 25 ans ayant obtenu leur permis de conduire dès la première tentative.

4. 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 pré-requis 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).
   b. Estimer la proportion d'étudiants prenant part à l'interrogation en début d'année académique ;
   c. Donner un intervalle de confiance à 95% pour cette proportion ;
   d. Donner un intervalle de confiance à 95% pour la moyenne (sur 20) du résultat obtenu lors du test.

5. La commune de Liège souhaite renouveler le stock de mobiles de jeu des cours de récréation de ses écoles. Toutes écoles fondamentales confondues, la commune de Liège compte 9578 écoliers. Afin de se conformer aux normes européennes, une étude concernant la taille des enfants est réalisée sur un échantillon de 500 écoliers. Sur cet échantillon, la taille moyenne est de 127.9 cm avec un écart-type de 18.4 cm.
    a. Quel est le taux de sondage ?
    b. Estimer la variance et l'erreur quadratique moyenne de l'estimateur proposé pour la moyenne de population ;
    c. Donner un intervalle de confiance à 95% pour la taille des enfants de l'enseignement primaire de la commune de Liège ;
    d. Sachant que la taille minimale pour le mobile envisagé est de 125 cm, est-il plausible qu'il ne convienne pas à un écolier de taille moyenne ? Et si la limite est de 120 cm ?
    e. Sur les 500 écoliers interrogés, 260 étaient des garçons, donner un intervalle de confiance à 95% pour la proportion de filles dans la population ;

6. La Région Wallonne souhaite organiser un sondage auprès de sa population (N = 3.500.000) pour estimer la proportion de citoyens ayant entamé des travaux de rénovation ou d'aménagement de leur habitation dans le but de réaliser des économies d'énergie au cours des trois dernières années (elle ne dispose d'aucune information *a priori* sur cette proportion - paramètre d'intérêt). Afin de trouver le meilleur compromis entre la précision de l'estimation attendue et le budget disponible, elle demande à un expert de déterminer quelle devrait être la taille d'échantillon (n) à prendre en compte afin d'estimer π avec une marge d'erreur maximale respectivement égale à 1%, 3% et 5%.

Quid si l'enquête n'était plus réalisée au niveau de la Région Wallonne mais de la Province de Liège (N = 1.050.000) ? Et de la Ville de Liège (N = 200.000) ?