Statistiques et Applications - TD1
Echantillonage

1 Exercice 1 : Echantillonage aléatoire simple

Des cercles, des triangles et des carrés sont représentés sur la fiche “population d'objets”. On souhaite estimer la moyenne des aires de ces formes.

1. Construisez un échantillon de 12 objets par une méthode quelconque (lancer de crayons, choix raisonné, formes ayant une intersection avec une droite quelconque,...). Noter les numéros des objets choisis.
2. Les aires de chacun des cercles, triangles et carrés sont notés sur la table `PopAire.csv`. A l'aide du logiciel de votre choix, **R**, excel, libre office, calculer l'aire moyenne de votre échantillon. Calculer ensuite la moyenne et l'écart-type corrigé de la population.
3. Calculer la moyenne et l'écart-type des estimations obtenues dans le groupe de TD. Construire l'histogramme des estimations obtenues dans le groupe de TD. Que peut-on dire sur la méthode d'échantillonnage “au hasard” mise en place ?
4. Reprendre les questions 1 à 3, en réalisant cette fois un échantillonnage aléatoire simple.

2 Exercice 2 : Echantillonage aléatoire stratifié

On sait maintenant que notre population est composée de $H = 3$ strates, une pour les cercles, une pour les triangles et une pour les carrés.

1. Construisez un échantillon de 12 objets en procédant à un échantillonnage aléatoire stratifié proportionnel (le taux de sondage dans la strate h est égal au poids de la strate h). En déduire une estimation de l'aire moyenne de la population. Calculer la moyenne et l'écart-type des estimations obtenues dans le groupe de TD. Comparer ces résultats à ceux obtenus avec l'échantillonnage aléatoire simple.
2. Soit $x_{\alpha h}$ l'aire de l' α^{e} objet de la strate h. Soit $N = \sum_{h=1}^{H} N_h$, $\mu_h = \frac{1}{N_h}\sum_{\alpha=1}^{N_h} x_{\alpha h}$. On définit la variance corrigée de la strate h, $\sigma_h^{*2} = \frac{1}{N_h-1}\sum_{\alpha=1}^{N_h}(x_{\alpha h} - \mu_h)^2$.
   Démontrer la relation suivante :
   $$(N-1)\sigma^{*2} = \sum_{h=1}^{H}(N_h - 1)\sigma_h^{*2} + \sum_{h=1}^{H} N_h(\mu_h - \mu)^2.$$
3. Dans cette question, on recherche quel taux de sondage appliquer dans chaque strate pour minimiser la variance de $\bar{X}$. On note f_h le rapport n_h/n.
   — Donner la variance $V(\bar{X})$ en fonction de N, n, N_h, σ_h^* et f_h.
   — Que vaut $\sum_{h=1}^{H} f_h$?
   — On considère $a_1, a_2, \cdots, a_H$ une suite de H nombres réels positifs. Appliquer l'inégalité de Cauchy-Scwartz aux suites $\sqrt{f_1}, \sqrt{f_2}, \cdots, \sqrt{f_H}$ et $\sqrt{a_1/f_1}, \sqrt{a_2/f_2}, \cdots, \sqrt{a_H/f_H}$. Dans quel cas a-ton égalité ?

— Appliquer le résultat de la question précédente au cas où $a_h = N_h^2 \sigma_h^{*2}$. En déduire les valeurs de f_h qui minimisent la variance $V(\bar{X})$.

4. Reprendre la question 1, en appliquant les taux de sondages optimaux trouvés à la question 3.