T. D. n$^{\text{o}}$ 1 bis
Complément sur le SAS à probabilités égales

Exercice 1. D'après le livre « Exercices corrigés de méthodes de sondage » de P. Ardilly et de Y. Tillé, Ellipses.

En sondage, il arrive que la taille N de la population U soit ignorée du sondeur. Une méthode pour y remédier est la suivante : nous identifions, parmi la population totale U de taille N (inconnue), M individus de cette population U. Nous laissons ensuite ces M individus se « mélanger » à la population totale U. Nous tirons ensuite n individus dans cette population U par sondage aléatoire simple dans la population totale U après mélange. Nous repérons alors, dans cet échantillon, m individus appartenant à la population préalablement « marquée ».

1. Quelle est la loi de la variable aléatoire m ? En déduire son espérance et sa variance ?
2. Quelle est la probabilité que m soit égal à zéro ? Nous supposons n petit devant M et devant $N - M$.
3. Considérant l'espérance de m, donner un estimateur $\widehat{N}$ de la taille N dans le cas où m est non nul. Nous vérifions qu'en pratique ceci se produit si n et M sont « suffisamment grands ».
4. Calculer $\mathcal{M} = \mathbb{E}\left[m|m > 0\right]$ et $\mathcal{V} = \mathrm{Var}\left[m|m > 0\right]$. En utilisant un développement limité de m autour de $\mathcal{M}$, approcher $\mathbb{E}\left[\widehat{N}|m > 0\right]$ en considérant n « grand » (et, en conséquence, N « particulièrement grand »).
5. Conclure sur le biais éventuel de l'estimateur $\widehat{N}$ de la taille N.

Exercice 2. D'après le livre « Exercices corrigés de méthodes de sondage » de P. Ardilly et de Y. Tillé, Ellipses.

Un échantillon de 100 étudiants est constitué au moyen d'un plan aléatoire simple sans remise dans une population de 1 000 étudiants. Nous nous intéressons alors aux résultats obtenus par ces étudiants à un examen. Il y a deux cas possibles : réussite ou échec. Le bilan est présenté dans le tableau ci-dessous :

	Femmes	Hommes	Total
Réussite	$n_{11} = 25$	$n_{12} = 35$	$n_{1.} = 60$
Échec	$n_{21} = 20$	$n_{22} = 20$	$n_{2.} = 40$
Total	$n_{.1} = 45$	$n_{.2} = 55$	$n_{..} = 100$

1. Estimer le taux de réussite des femmes et celui des hommes.
2. Calculer le biais approché des taux de réussite estimés.

3. Estimer l'erreur quadratique moyenne de ces taux de réussite.
4. Proposer les intervalles de confiance à 95% pour les taux de réussite des femmes R_F et des hommes R_H. Que pouvons-nous dire de leurs positions respectives ?
5. Quels intervalles de confiance faut-il considérer pour que les vraies valeurs R_F et R_H soient encadrées par des intervalles de confiance disjoints ? Commenter.

Exercice 3. D'après le livre « Exercices corrigés de méthodes de sondage » de P. Ardilly et de Y. Tillé, Ellipses.

Nous nous intéressons à l'estimation de la proportion d'hommes P atteints par une maladie professionnelle dans une entreprise de 1 500 travailleurs. Nous savons par ailleurs que trois travailleurs sur dix sont ordinairement touchés par cette maladie dans des entreprises du même type. Nous nous proposons de sélectionner un échantillon au moyen d'un sondage aléatoire simple.

1. Quelle taille d'échantillon faut-il sélectionner pour que la longueur totale d'un intervalle de confiance, avec un niveau de confiance de 95% soit inférieure à $0,02$ pour les plans simples avec et sans remise ?
2. Que faire si nous ne connaissons pas la proportion d'hommes habituellement touchés par la maladie, pour le cas du plan sans remise ?

Pour éviter les confusions de notation, nous mettrons l'indice AR aux estimateurs avec remise, et l'indice SR aux estimateurs sans remise.

Exercice 4. D'après le livre « Exercices corrigés de méthodes de sondage » de P. Ardilly et de Y. Tillé, Ellipses.

Dans une population de 4 000 personnes, nous nous intéressons à deux proportions :

— P_1 =proportion des individus possédant un lave-vaisselle,

— P_2 =proportion des individus possédant un ordinateur portable.

D'après des renseignements « sûrs », on sait qu'à priori :

$$45\% \leqslant P_1 \leqslant 65\% \quad \text{et} \quad 5\% \leqslant P_2 \leqslant 10\%.$$

Quelle doit être la taille de l'échantillon n dans le cadre d'un sondage aléatoire simple si nous voulons connaître à la fois P_1 à 2% près et P_2 à 1% près, avec un niveau de confiance de 95% ?