

Université de Thiès
UFR de Sciences Economiques et Sociales &
UFR de Sciences et Technologies
ANNEE 2019-2020

Projet Techniques de Sondage
A rendre avant Lundi 11 Mai à 12h
Master 1 de Sciences de Données et Applications

Exercice 1.

Probabilité d'inclusion. Soit la population $\{1, 2, 3\}$ et le plan de sondage suivant :

$$\mathbf{P}(\{1, 2\}) = \frac{1}{2} \quad \mathbf{P}(\{1, 3\}) = \frac{1}{4} \quad \mathbf{P}(\{2, 3\}) = \frac{1}{4}.$$

1. Est-ce un sondage aléatoire simple ?
2. Calculer π_1 , π_2 , et π_3 , les probabilités d'inclusion d'ordre 1.
3. Calculer π_{12} , et π_{23} les probabilités d'inclusion d'ordre 2.
4. Quel est le π -estimateur de \bar{Y} ? Nous noterons Y_1 , Y_2 et Y_3 les valeurs respectives de la variable Y .
 - a) si l'échantillon $\{1, 2\}$ est tiré ?
 - b) si l'échantillon $\{1, 3\}$ est tiré ?
 - c) si l'échantillon $\{2, 3\}$ est tiré ?
5. Vérifier que le π -estimateur est un estimateur sans biais.
6. Écrire ce que seraient les probabilités d'échantillon \mathbf{P} et les probabilités d'inclusion π pour un sondage aléatoire simple à probabilités égales sans remise.

Exercice 2.

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 3.

Pour faire un test de niveau sur les élèves des collèges de 6ième d'un district, on va mettre en place une enquête par sondage. Au premier degré on tire 5 collèges parmi les 50 selon un sondage aléatoire simple à Probabilités égales sans remise. Puis au deuxième degré on effectue les tests sur un échantillon de 10 élèves par collège. à l'issue du sondage, on a calculé pour chaque collège la note moyenne des 10 élèves testés ainsi que leur variance corrigée.

Collège	1	2	3	4	5
Effectif des classes de 6ième du collège	40	20	60	40	48
Note moyenne	12	8	10	12	11
Variance corrigée des notes	1,5	1,2	1,6	1,3	2

1. Donner une estimation du total des notes au test sur le district.
2. Estimer le nombres d'élèves en 6ième du district.
3. En supposant qu'il y ait exactement 2000 élèves en 6ième dans le district, donner une estimation de la note moyenne. Comparer avec la moyenne observée sur l'échantillon.
4. Calculer la variance de l'estimateur du total. En déduire la variance de la moyenne. Donner un intervalle de confiance de la moyenne.
5. Comparer avec un sondage aléatoire simple à Probabilités égales sans remise sur les mêmes données.

Une étude portant sur 25 voitures identiques indique une consommation moyenne de 8,5 litres aux 100 km avec un écart type de 0.8 litre.

a) Donner une estimation ensembliste de la consommation moyenne de ce type de véhicule au seuil de 5%.

b) Quel devrait être le nombre minimal d'observation auxquelles on devrait procéder pour connaître la consommation moyenne à plus ou moins 2 décilitres près (au plus) aux seuils de 5% et de 1% ?

Exercice 4.

Un vérificateur est chargé de contrôler la qualité d'un certain nombre d'enregistrements comptables. L'hypothèse est faite sur le vu de vérifications antérieures, d'un niveau d'acceptation $p=0,05$.

1. On prévoit de contrôler $n=200$ enregistrements. Quel nombre maximum d'erreurs peut-on accepter dans cet échantillon sans remettre en cause le niveau d'acceptation 0,05.
Même question avec $n=400$, $n=600$, $n=1000$.
2. Avec le même niveau d'acceptation, on a fait un sondage sur $n=400$ enregistrements. On a constaté 7 erreurs. Le contrôle est poursuivi. Combien doit-on examiner d'enregistrements supplémentaires en tolérant, au cours de ce contrôle 4 nouvelles erreurs, pour que l'hypothèse d'un niveau d'acceptation de 5% puisse être raisonnablement retenue.

Exercice 5.

La variable d'intérêt est le chiffre d'affaire moyen réalisé par un ensemble de 1060 entreprises. Celles-ci étant de tailles très différentes, on a constitué 5 strates en fonction du nombre de salariés dans chaque entreprise.

Nombre de salariés	0 à 9	10 à 19	20 à 49	50 à 499	500 et +
Nombre d'entreprises	500	300	150	100	10

1. A l'intérieur de chaque strate on réalise un sondage aléatoire simple à Probabilités égales sans remise avec les tailles d'échantillon suivantes :

$$n_1 = 130, n_2 = 80, n_3 = 60, n_4 = 25, n_5 = 5.$$

Les résultats sont les suivants :

h	1	2	3	4	5
y_h	5	12	30	150	600
s_2^h	1,5	4	8	100	2500

Donner une estimation du chiffre d'affaire moyen avec un intervalle de confiance.

2. En conservant toujours la même taille globale d'échantillon, quels effectifs d'échantillon faut-il considérer dans chaque strate

(a) Pour une allocation proportionnelle ?

(b) Pour une allocation optimale ? (Faire attention que l'on a toujours évidemment $n_h \leq N_h$.)

3. En supposant inchangées les variances dans les strates, calculer les variances de l'estimateur pour le plan avec allocation proportionnelle et celui avec allocation optimale.