

DEVOIR MAISON

Ce devoir, à rendre pour le 27 février 2017, compte pour 1/3 de la note totale du cours. Il peut être réalisé en groupe jusqu'à quatre personnes ; dans ce cas, veuillez ne rendre qu'une copie par groupe avec tous vos noms. Il comporte un problème et un exercice plus court. La notation tiendra compte de la présentation et de l'orthographe.

1 Problème

La Poste dispose de trois gammes de tarification pour l'envoi des courriers : la lettre Classique, qui a un coût modéré et qui met environ 2 jours à arriver, la lettre Verte, moins chère mais plus lente, et le courrier Prioritaire, qui coûte plus cher mais est plus rapide. Les courriers ont des formats (longueur, largeur, épaisseur) qui varient, ce qui complique le stockage et la distribution. Un responsable d'un centre de tri souhaite mieux connaître le courrier qui transite par sa plate-forme.

1.1 Première approche

Il veut tout d'abord estimer, parmi les 1 000 courriers qui sont arrivés ce jour, combien sont de chacun des trois types. Il sélectionne complètement au hasard 3 lettres dans le tas et obtient les résultats suivants :

Type	Nombre
Classique	2
Verte	1
Prioritaire	0

TABLE 1 – Répartition des trois lettres

1.1.1 Comment s'appelle le sondage utilisé ? Rappeler la probabilité d'inclusion simple de chacune des lettres.

1.1.2 Combien vaut l'estimateur d'Horvitz-Thompson du nombre total de lettres classiques, vertes puis prioritaires ? Qu'en déduire sur la présence de lettres prioritaires dans le centre de tri ?

1.1.3 Calculer l'estimateur de la variance de l'estimateur d'Horvitz-Thompson du nombre total de lettres vertes.

1.1.4 Donner un intervalle de confiance à 95 % du nombre total de lettres vertes dans le centre de tri.

1.2 Étude approfondie

Le manager trouve que les résultats obtenus précédemment ne sont pas assez précis. Il décide donc d'organiser une opération à grande échelle dans le centre de tri pour mieux connaître le courrier qui transite. Le tableau suivant résume le nombre de courriers, selon le type et l'épaisseur du courrier, parmi les 1 000 courriers présents au centre de tri.

1.2.1 Quel est le vrai nombre total de lettres vertes ? Commenter le résultat obtenu à l'aide de la réponse à la question 1.1.4.

Epaisseur \ Type	Vert	Classique	Prioritaire
1 cm	105	235	75
2 cm	80	200	0
3 cm	40	95	0
4 cm	30	85	0
5 cm	25	15	0
8 cm	15	0	0

TABLE 2 – Tableau croisé de données

1.2.2 Combien y a-t-il d'échantillons possibles dans le cadre d'un sondage aléatoire simple de 3 lettres? Que peut-on dire de leurs probabilités d'inclusion?

1.2.3 Remplir le tableau suivant, où la configuration $(x \text{ V}, y \text{ C}, z \text{ P})$ signifie que l'échantillon de 3 lettres comporte x lettres vertes, y lettres classiques et z lettres prioritaires.

Configuration	Probabilité que l'échantillon présente cette configuration	Valeur de l'estimateur d'Horvitz-Thompson du nombre total de lettres vertes
(1 V, 1 C, 1 P)		
(2 V, 1 C, 0 P)		
(1 V, 2 C, 0 P)		
(2 V, 0 C, 1 P)		
(1 V, 0 C, 2 P)		
(0 V, 2 C, 1 P)		
(0 V, 1 C, 2 P)		
(3 V, 0 C, 0 P)		
(0 V, 3 C, 0 P)		
(0 V, 0 C, 3 P)		

1.2.4 En utilisant le résultat de la question précédente, montrer que l'estimateur d'Horvitz-Thompson du nombre de lettres vertes est sans biais.

1.2.5 Quelle est la vraie variance de l'estimateur d'Horvitz-Thompson du total de nombre de lettres vertes dans ce cadre?

1.3 Étude de l'épaisseur des courriers

1.3.1 Le responsable du centre souhaite savoir quelle hauteur atteindra la pile des 1000 courriers pour la distribution. Donnez la formule de l'estimateur d'Horvitz-Thompson de l'épaisseur totale des courriers, pour un échantillon s tiré par sondage aléatoire simple de n courriers. Combien vaut sa *vraie* variance?

1.3.2 Le responsable du centre va organiser une distribution différente pour chaque type de courrier. Donnez la formule de l'estimateur d'Horvitz-Thompson de l'épaisseur totale des courriers de type prioritaires, pour un échantillon s tiré par sondage aléatoire simple de n courriers. Combien vaut sa *vraie* variance? Commentez le résultat obtenu.

2 Exercice

Si je veux estimer la proportion de personnes nées un 29 février en France, combien de personnes dois-je interroger via un sondage aléatoire simple pour obtenir un coefficient de variation (CV) de 5%? *On fera une hypothèse raisonnable sur le vrai pourcentage.*