



Université PANTHÉON - ASSAS (PARIS II)

U.E.C.1

Droit - Economie - Sciences Sociales

4615

Session :	Janvier 2017
Année d'étude :	Première année de Master économie-gestion mention ingénierie économique et statistique
Discipline :	Sondages (Unité d'Enseignements Complémentaires 1)
Titulaire(s) du cours :	M. Philippe PERIE
Document(s) autorisé(s) :	Formulaire

Les exercices sont indépendants et peuvent être résolus dans n'importe quel ordre. Barème indicatif : question de cours : 7 points, exercice 1 : 6 points, exercice 2 : 6 points

Question de cours : Estimation par le ratio. LAPLACE et le recensement

Depuis août 1539 et François Ier, il est obligatoire en France de tenir des registres de naissances par paroisse (commune). C'est donc une information connue de manière sûre.

Pour le recensement de la population, la technique d'extrapolation en vigueur à l'époque de LAPLACE (Pierre-Simon Laplace, 'Essai Philosophique sur les Probabilités' 1814) était la suivante : un sondage était effectué dans un certain nombre de paroisses d'une province, l'intendant notait y = le nombre d'habitants comptés tête par tête, x = nombre moyen des naissances par année sur les 10 dernières années dans ces paroisses. Ces résultats permettaient de calculer le multiplicateur y/x . Ensuite, pour connaître Y = nombre d'habitants dans une région, il suffisait de connaître X = nombre moyen des naissances sur les 10 dernières années dans cette région considérée et d'effectuer $Y = X * y/x$. Le ratio est donc le nombre d'habitants vivants par naissance sur l'année.

En fait l'idée initiale de Laplace remonte à 1802, et lors de sa première mise en œuvre, il avait sélectionné aléatoirement un échantillon de 30 communes réparties dans tout le pays. Pour ces communes, on a compté un total de 2037615 habitants, le 23 Septembre 1802. Jugeant les données des registres paroissiaux peu fiables avant 1800, il se limita à 3 ans et dans les trois années précédant le 23 Septembre 1802, il compta 215599 naissances dans les 30 communes. Laplace détermina ainsi le nombre annuel de naissances enregistrées dans les 30 communes soit $2037615 \times 3 / 215599 = 28.35$, Laplace estime donc que chaque année il y avait une naissance enregistrée pour tous les 28.35 personnes. Il a ensuite supposé que le rapport de la population aux naissances annuelles dans son échantillon serait probablement similaire dans toute la France, et conclut que l'on pouvait estimer la population totale de France en multipliant le nombre total de naissances annuelles dans toute la France par 28,35.

Bien entendu, un intendant d'une autre province pouvait contester la valeur de ce multiplicateur dans la mesure où celui-ci dépendait des paroisses sondées, mais aussi des caractéristiques sociologiques de la province. Ainsi, pour la généralité d'Auvergne, après sondage dans certaines paroisses, il avait été conclu que le multiplicateur devait être égal à 25.

C'est ainsi qu'un secrétaire de l'intendant de Lyon, Louis MESSANCE établit la règle suivante : *"25 pour les paroisses de la campagne et 28 pour celles des villes"*. En effet ; il remarquait que : *"les grandes villes renferment dans leurs murs des ordres de citoyens qui ne se trouvent pas dans les petites villes et paroisses de la campagne, tels que les corps ecclésiastiques des deux sexes, les séminaires, les hôpitaux, les collèges et un grand nombre de domestiques de tout genre attachés au service des personnes riches ; et comme ces différentes classes de citoyens vivent dans le célibat, et ne contribuent pas au renouvellement de la population, il est nécessaire d'adopter un facteur plus fort."*

Questions :

1. La méthode proposée par LAPLACE s'appelle une estimation par la méthode du ratio : quels sont ses avantages dans ce cas précis d'une opération lourde comme un recensement ?
2. Que pensez-vous du passage de 3 ans (en 1802) à 10 ans (en 1814) sur la méthode de Laplace. En quoi rend-elle la procédure plus fiable ?
3. Que pensez-vous de l'échantillon initial de 30 communes pour tout le pays de Laplace. Quel est le principe derrière la proposition de Louis MESSANCE ?

Exercice 1 : Mise en place d'une enquête mixte, allocation optimale dans les strates : (Lohr, 2010, exercice no. 8 p. 104). Dans certains cas il est difficile de concilier dans un seul mode de recueil tous les avantages, et on est donc amené à mettre en place des stratégies qui mélangent les modes de recueil.

Un chercheur dispose d'un budget de 20000 € pour effectuer une enquête. Il sait que, dans la population, seuls 90% des ménages ont un numéro de téléphone fixe. Par contre tous les ménages sont accessibles en face à face. Les interviews téléphoniques coûtent 10€ par ménage ; les interviews sur place coûtent, quant à elles, 30 € chacune si toutes les interviews sont effectuées sur place et coûtent 40€ si elles ne sont utilisées que pour les ménages ne possédant pas un numéro de téléphone (car dans ce cas, les coûts de déplacement sont plus élevés car amortis sur une base plus faible). Supposons que la dispersion dans les strates des ménages ayant un téléphone est égale à celle dans la strate des ménages n'ayant pas de téléphone. De plus, supposons que le coût fixe de base $C_0 = 5000$ €

Combien de ménages peut-on au mieux interviewer dans chaque strate si :

- Le chercheur veut une couverture complète de la population avec un mode de collecte, donc tous les ménages bénéficient d'une interview en face à face.
- Le chercheur met en place une stratégie mixte de collecte pour obtenir cette couverture complète : les ménages ayant un téléphone sont interviewés par téléphone et les ménages n'ayant pas de téléphone sont interviewés sur place.

Exercice 2 : Strates selon les disciplines : (Lohr, 1999, exercice no. 8 p. 121) Le « American Council of Learned Societies » (ACLS) a tiré un échantillon de sociétés dans sept disciplines faisant parties de l'ACLS selon un plan stratifié aléatoire simple. Le but était d'étudier les habitudes de publication et l'utilisation des bibliothèques et du matériel informatique parmi les membres (étudiants, chercheurs et professeurs) des différents départements appartenant à une société de l'ACLS. On a demandé si le répondant est d'accord ou en désaccord avec la question suivante : « Lorsque je feuillette les pages du nouveau numéro de la revue la plus importante dans mon champ académique, je trouve rarement un article qui m'intéresse ». Les résultats par champs sont exhibés ci-dessous.

- Estimez la proportion d'individus dans la population qui sont en accord avec la question posée et estimez la variance de votre estimateur et donnez en un intervalle de confiance à 95%
- Calculez la variance de l'estimateur si l'on n'avait pas fait la stratification et un intervalle de confiance à 95%
- Commentez

On note :

N_h = taille de la strate h

n_h = taille de l'échantillon dans la strate h (sondage aléatoire simple dans chaque strate)

\hat{p}_h = estimateur de la proportion d'individus possédant un véhicule dans la strate h

Champs	Taille du département N_h	Taille de l'échantillon n_h	En accord (%) \hat{p}_h
Littérature	9100	915	37
Etudes classiques	1950	633	23
Philosophie	5500	658	23
Histoire	10850	855	29
Langues	2100	667	19
Sciences politiques	5500	833	43
Sociologie	9000	824	41

