Université PANTHÉON - ASSAS (PARIS II) **U.E.F.1**

Droit - Economie - Sciences Sociales **4390**

Melun

Session : Janvier 2019

Année d'étude : Troisième année de licence économie-gestion mention administration économique et sociale

Discipline : ***Sondages techniques quantitatives***
(Unité d'Enseignements Fondamentaux 1)

Titulaire(s) du cours :
M. Antoine AUBERGER

Document(s) autorisé(s) : Aucun document n'est autorisé. L'usage d'une calculatrice est autorisé.

Questions de cours :

1) Donner trois noms d'instituts de sondages.
Pour un sondage d'intention de vote (exemple : élections Européennes de 2019 en France), quelle est la méthode utilisée pour sélectionner l'échantillon ? Donner des précisions sur la méthode, sur l'échantillon et sur la stratification.

2) Soit une population avec N éléments et un échantillon avec n éléments (sondage sans remise).
Soit $S^2 = \frac{1}{N-1}\sum_1^N (x_i - \bar{X})^2$. Que représente S^2 ?
Montrer que : $S^2 = \frac{N}{N-1}(\frac{1}{N}\sum_1^N x_i^2 - \bar{X}^2)$
Si S^2 est inconnue, quelle estimation de S^2 utilise-t-on pour remplacer S^2 ?

3) Présenter l'analyse en composantes principales : objectif, principes, méthode et proposer un exemple de données. Préciser notamment la nature des variables.

Exercice 1 :
On veut estimer la superficie moyenne cultivée dans les fermes d'un canton. Sur les 1480 fermes que compte le canton, on en tire 100 par sondage aléatoire simple. On mesure (en hectares) la surface cultivée x_i par la ferme numéro i de l'échantillon et on trouve :

$$\sum_{i=1}^{100} x_i = 2148 \quad \text{et} \quad \sum_{i=1}^{100} x_i^2 = 121451$$

1) Traduire en quelques mots l'information contenue dans la formule : $\sum_{i=1}^{100} x_i = 2148$
2) Donner la valeur de l'estimateur de la moyenne $\hat{\mu}$ (= $\bar{x}$).

3) Donner un intervalle de confiance à 95% pour la superficie moyenne cultivée dans les fermes de ce canton. Ecrire une phrase de conclusion.

Exercice 2 :

On considère le caractère Y = "âge" en années dans la population de 5 individus :
U = {Marc, Jean, Augustin, Yves, Martial} = $\{u_1, \dots , u_5\}$. Pour tout i ϵ {1, … , 5}, soit y_i la valeur de Y pour l'individu u_i. Les résultats, en années, sont :

y_1	y_2	y_3	y_4	y_5
19	25	34	41,5	23,5

1) Calculer la moyenne de la population $\bar{y}_U$ et l'écart-type corrigé de la population S_U.
2) Dans un premier temps, on prélève un échantillon de 3 individus suivant un plan de sondage aléatoire simple de type PESR.
a) Quel est le taux de sondage ? Combien d'échantillons peut-on former ? Expliciter les.
b) Pour chaque échantillon ω, calculer la moyenne de l'échantillon $\bar{y}_\omega$.
c) Soit $\bar{y}_W$ la variable aléatoire réelle égale à la moyenne de l'échantillon, l'aléatoire étant dans l'échantillon considéré.
Déterminer sa loi, puis calculer son espérance et sa variance.
d) Avec quelle formule peut-on calculer la variance de l'estimation de $\bar{y}_U$ ($\hat{\bar{y}}_U$) ? Faire le calcul.
3) Dans un deuxième temps, on prélève un échantillon de 3 individus suivant un plan de sondage aléatoire de type stratifié avec :
- les 3 strates : U_1 = {Marc, Jean, Augustin}, U_2 = {Yves}, U_3 = {Martial}
- un individu par strate.
a) Combien d'échantillons peut-on former ? Expliciter les.
b) Pour chaque échantillon ω, calculer la moyenne de l'échantillon $\bar{y}_\omega$.
c) Soit $\bar{y}_W$ la variable aléatoire réelle égale à la moyenne de l'échantillon stratifié, l'aléatoire étant dans l'échantillon considéré.
Déterminer sa loi, puis calculer son espérance et sa variance.
d) Calculer la variance de $\hat{\bar{y}}_W$ à l'aide de la formule : $V(\hat{\mu}) = \frac{1}{N^2}\sum_{h=1}^{H} N_h \frac{N_h - n_h}{n_h} S_h^2$

Exercice 3 :

Dans une population de très grande taille N = 20000, on souhaite estimer l'âge moyen μ des individus. Pour cela, on stratifie la population en trois catégories d'âge, et on tire un échantillon par sondage aléatoire simple dans chaque catégorie. De plus, grâce à une enquête précédente, on dispose d'estimations pour les variances corrigées de chaque strate.
L'ensemble des informations dont on dispose sont résumées dans le tableau suivant :

Strate	Population		Echantillon	
	N_h	s_h^2	n_h	$\bar{x}_h$
Moins de 30 ans	6000	17	60	19
De 30 à 50 ans	8500	15	90	43
Plus de 50 ans	5500	24	50	65

1) Préciser la signification de : N_h , S_h^2 , n (et donner sa valeur), n_h , $\bar{x}_h$.
1) Quelle est la valeur de l'estimateur stratifié de l'âge moyen μ ? ($\hat{\mu}$)
2) Calculer la variance de cet estimateur.
3) Quelles tailles d'échantillons n_h doit-on choisir pour chaque strate si on souhaite réaliser une allocation proportionnelle afin de constituer un échantillon de n = 200 individus ?
Calculer alors la variance de l'estimateur stratifié que l'on obtient avec ce plan de sondage.
4) On souhaite maintenant réaliser une allocation optimale (toujours avec n = 200).

Calculer alors la valeur des n_h ainsi que la variance de l'estimateur stratifié que l'on obtient avec ce plan de sondage.
5) Parmi les trois plans de sondage proposés, lequel vous semble le plus approprié ? (expliquer)

Exercice 4 :

Exemple d'un tableau de distances

	a	b	c	d
a	0			
b	1	0		
c	3	2		
d	7	6	4	0

On utilise l'agrégation par l'ultramétrique supérieure minima :

$$D(I_1, I_2) = \max\{d(i, i'), i \epsilon I_1, i' \epsilon I_2\}$$

On agrège les deux parties ayant les deux points les plus proches et qui vont former une classe de diamètre maxima, les classes sont longues et fines, il y a un effet de chaînage (critère du plus proche voisin).

Construire une hiérarchie : algorithme ascendant ou agrégatif.