République du Sénégal
Un people – Un but – Une foi
-=-=-=-=-=-=-=-=-=-

Ministère de l'Enseignement Supérieur de la Recherche et de l'Innovation

UNIVERSITE DE THIES

UFR Sciences Economiques et Sociales – UFR Sciences Et Technologies

Département Management des Organisations

Master en Sciences de Données et Applications

Projet Technique de Sondage

Présenté par :

Abdoulaye Djibril BA

Alioune CISSE

Bineta TALL

Yaye Sala TOURE

Professeur :

Dr Fatou Nene DIOP

Exercice 1:

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

$P(\{1,2\}) = 0,5$; $P(\{1,3\}) = 0,25$;

$P(\{2,3\}) = 0,25$

1) Non ce n'est pas un sondage aléatoire simple car les tirages ne sont pas équiprobables.

2. Calculons π_1, π_2, π_3

$$\pi_i = \sum_{S(i=1)}^{S} p(s)$$

$\pi_1 = P(\{1,2\}) + P(\{1,3\}) = 0,5 + 0,25 = 0,75 = 3/4$

$\pi_2 = P(\{1,2\}) + P(\{2,3\}) = 0,5 + 0,25 = 0,75 = 3/4$

$\pi_3 = P(\{1,3\}) + P(\{2,3\}) = 0,25 + 0,25 = 0,5 = 1/2$

3. Calculons π_{12}, π_{23}

$$\pi_{ij} = \sum_{s(i,j \in s)} p^{(s)}$$

$\pi_{12} = P(\{1,2\}) = 0,5 = 1/2$

$\pi_{13} = P(\{1,3\}) = 0,25 = 0,25 = 1/4$

$\pi_{23} = P(\{2,3\}) = 0,25 = 1/4$

4. Le π - estimateur de $\bar{Y}$

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} \frac{Y_i}{\pi_i}$$

a) Si l'échantillon $\{1,2\}$ est tiré

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} \frac{Y_i}{\pi_i} = \frac{Y_1}{3/4} + \frac{Y_2}{3/4}, \quad S = \{1,2\};$$

$$\hat{\mu} = \frac{4(Y_1 + Y_2)}{9}, \quad S = \{1,2\}$$

b) Si l'échantillon $\{1,3\}$ est tiré

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} \frac{Y_i}{\pi_i} = \frac{Y_1}{3/4} + \frac{Y_3}{1/2}, \quad S = \{1,3\}$$

$$\hat{\mu} = \frac{4Y_1 + 6Y_3}{9}, \quad S = \{1,3\}$$

c/ Si l'échantillon $\{2,3\}$ est tiré

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N}\frac{Y_1}{\pi_1} = \frac{Y_2}{3/4} + \frac{Y_3}{1/2}, \quad s=\{2,3\}$$

$$\hat{\mu} = \frac{4Y_2 + 6Y_3}{9}, \quad s=\{2,3\}$$

5. Vérifions que π- estimateur est sans biais

$$E(\bar{Y}_n) = \frac{1}{n}\sum_{i=1}^{n}\pi_i Y_i = \frac{1}{2}\times\frac{4}{9}(y_1+y_2) + \frac{1}{4} * \frac{4y_1+6y_3}{9}$$

$$+\frac{1}{4}\times\frac{4y_2+6y_3}{9} = \frac{2y_1+2y_2+y_1+1,5y_3+y_2+1,5y_3}{9}$$

$$E(\bar{Y}_n) = \frac{3y_1+3y_2+3y_3}{9} = \frac{y_1+y_2+y_3}{3} = \bar{Y}$$

6. Ecrivons ce que serait P et π pour un sondage aléatoire simple.

Pour un SAS, les tirages sont équiprobables ce qui donne: $P(\{1,2\}) = P(\{1,3\}) = P(\{2,3\}) = 1/3$

$\pi_1 = P(\{1,2\}) + P(\{1,3\}) = 2/3$

$\pi_2 = P(\{1,2\}) + P(\{2,3\}) = 2/3$

$\pi_3 = P(\{1,3\}) + P(\{2,3\}) = 2/3$

$\pi_{12} = P(\{1,2\}) = 1/3$

$\pi_{13} = P(\{1,3\}) = 1/3$

$\pi_{23} = P(\{2,3\}) = 1/3$

Exercice 2 :

On a comme paramètre d'intérêt $p = \frac{1}{N} \sum_{k \in U} y_k$ où les y_k sont des indicatrices codant la présence ou non de la maladie. Estimons le par $\hat{p} = \frac{1}{n} \sum_{k \in s} y_k$

Sa variance est donc donnée par :

$$Var(\hat{p}) = \begin{cases} \frac{\sigma_y^2}{n} & \text{avec remise} \\ \frac{N-n}{N} \frac{S_y^2}{n} & \text{sans remise} \end{cases}$$

Mais puisque $y_k^2 = y_k$, la variance et la variance corrigée sur la population sont égales à

$$\sigma_y^2 = \frac{1}{N} \sum_{k \in U} y_k - \left(\frac{1}{N} \sum_{k \in U} y_k \right)^2 = p - p^2 = p(1-p)$$

$$S_y^2 = \frac{N}{N-1} p(1-p)$$

Ainsi on a donc $Var(\hat{p}) = \begin{cases} \frac{p(1-p)}{n} & \text{avec remise} \\ \frac{N-n}{N-1} \frac{p(1-p)}{n} & \text{sans remise} \end{cases}$

On a $1 - \alpha/2 = 1 - 0,05/2 = 0,975$ donc $z_{0,975} = 1,96$

On cherche donc n telle que $2 \times 1,96 \times \sqrt{Var(\hat{p})} \leq 0,02$

$\sqrt{Var(\hat{p})} = \frac{0,02}{2 \times 1,96} \Rightarrow Var(\hat{p}) = 2,603 \cdot 10^{-4}$

$$Var(\hat{p}) = 2,603.10^{-4} \Rightarrow \begin{cases} \frac{p(1-p)}{n} \leq 2,603.10^{-4} \text{ A.R} \\ \frac{N-n}{N-1}\frac{p(1-p)}{n} \leq 2,603.10^{-4} \text{ S.R} \end{cases}$$

Ce qui donne :

$$\begin{cases} 2,603.10^{+4} p(1-p) \leq n \text{ A.R} \\ 2,603.10^{-4} N p(1-p) / \left(N - 1 + 2,603.10^{+4} p(1-p)\right) \leq n \text{ S.R} \end{cases}$$

En prenant $p = 3/10$ et $N = 1500$ on trouve alors que

$n \geq 5466$ Avec Remise et $n \geq 1177$ Sans Remise

Exercice 3

1. Au 1er degré, on a

$M = 50$ collèges, $m = 5$ collèges, $f_1 = 0,1$

Au 2ème degré nous avons

Observation	N_i	n_i	$\bar{y}_i$	s_i^2	$\hat{T}_i$
1	40	10	12	1,5	480
2	20	10	8	1,2	160
3	60	10	10	1,6	600
4	40	10	12	1,3	480
5	48	10	11	2,0	528
Total	208	50	-	-	2248

Dans chaque collège, la note totale T_i est estimée par: $\hat{T}_i = N_i \bar{y}_i$ $\Rightarrow \hat{T}_1 = 40 \times 12 = 480$

$\hat{T}_2 = 20 \times 8 = 160$

$\hat{T}_3 = 60 \times 10 = 600$

$\hat{T}_4 = 40 \times 12 = 480$

$\hat{T}_5 = 48 \times 11 = 528$

La note totale dans le distruct est estimée par

$$\hat{T} = \frac{M}{m} \sum_{i=1}^{m} \hat{T}_i = \frac{50}{5} \times 2048 = 20480$$

2) Estimons le nombre d'élèves en 6ème du district

$$\hat{N} = \frac{M}{m} \sum_{i=1}^{m} N_i = \frac{50}{5} \times (40 + 20 + 60 + 40 + 48) = 2050$$

3. Donnons une estimation de la moyenne en supposant qu'il y ait exactement 2000 élèves.

$$\hat{Y} = \frac{1}{N} \cdot \hat{T} \; ; \quad N = 2000 \Rightarrow \hat{Y} = \frac{1}{2000} \times 20480$$

$$\hat{Y} = 11,24.$$

Comparaison avec la moyenne observée de l'échantillon.

La moyenne observée sur un échantillon de taille 50 donne $\bar{y} = \frac{1}{50}(10 \times 12 + 10 \times 8 + 10 \times 10 + 10 \times 12 + 10 \times 11)$

$$\bar{y} = 10,6$$

4. Calculons la variance de l'estimateur total

$$\widehat{Var}(\hat{T}) = M^2(1-f_1)\frac{S_1^2}{m} + \frac{M}{m}\sum_i N_i^2(1-f_{2,i})\frac{S_{2,i}^2}{n_i}$$

où $S_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(\hat{T}_i - \frac{\hat{T}}{M}\right)^2$

$$S_{2,i}^2 = \frac{1}{n_i - 1}\sum_j (y_{i,j} - \bar{y}_i)^2$$

$S_{2,i}^2$ est donnée dans le tableau.

$$S_1^2 = \frac{1}{4}\left[(480-449,6)^2 + \dots + (428-449,6)^2\right] = 28260,8$$

On calcule le 1er terme de l'estimation de la var

$$M^2(1-f_1)\frac{S_1^2}{m} = 50^2 \times 0,9 \times \frac{28260,8}{5} = 12879360$$

En posant $V_i = N_i^2(1-f_{2,i})\frac{S_{2,i}^2}{n_i}$, on peut calculer le second terme de l'estimation de la variance de l'estimateur du total qui vaudra alors la somme des quantités suivantes pondérée par la qualité M/m.

$$V_1 = 40^2\left(1 - \frac{10}{40}\right) \times \frac{1,5}{10} = 180$$

$$V_2 = 20^2\left(1 - \frac{10}{20}\right) \times \frac{1,2}{10} = 24$$

$V_3 = 60^2 \left(1 - \frac{10}{60}\right) \times \frac{1,6}{10} = 480$

$V_4 = 40^2 \left(1 - \frac{10}{40}\right) \times \frac{13}{10} = 156$

$V_5 = 48^2 \left(1 - \frac{10}{48}\right) \times \frac{8}{10} = 364,8$

Ainsi en multipliant par M/m, on obtient que la quantité cherchée est égale à

$$\frac{M}{m} \sum_{i=1}^{5} V_i = \frac{50}{5} \times 1204,8 = 12048$$

Finalement, l'estimation de la variance de l'estimateur du total est égale à $\widehat{Var}(\hat{T})$

$\widehat{Var}(\hat{T}) = 12879360 + 12048 = 12891408$

On en déduit $\widehat{Var}(\bar{Y}) = \frac{1}{N^2} \widehat{Var}(\hat{T}) = \frac{1}{2000^2} 12891408$

$\widehat{Var}(\hat{T}) = 3,22$.

On obtient ainsi une précision égale à

$\pm 1,96\sqrt{3,22} = 3,5$ et qui va nous permettre de calculer un intervalle de confiance de la moyenne à 95% qui vaut $11,2 \pm 3,5$ et qui va nous permettre de calculer un I.C de la moy à 95% qui vaut $11,2 \pm 3,5$

5. Comparaison avec un SAS à proba égales sur les mêmes données.

$\hat{\bar{Y}} = \bar{y} = 10,6 \quad n = 50$ et $N = 2000$

Le taux de sondage est égal à $f = \frac{50}{2000} = 0,25$.

L'estimation de la variance de l'estimateur de la moyenne est égal à $var(\hat{\bar{Y}}) = (1-f)\frac{s^2}{n}$,

$S^2 = \frac{N}{N-1} s^2$

Cherchons d'abord s^2

$s^2 = v1 + v2$

$v1 = \frac{1}{10}(10\times 12^2 + 10\times 8^2 + 10\times 10^2 + 10\times 12^2 + 10\times 11^2) - 10,6^2$

$v1 = 2,24$ ①

$v2 = \frac{1}{50}\left(10\times\frac{9}{10}\times 1,5 + 10\times\frac{9}{10}\times 1,2 + 10\times\frac{9}{10}\times 1,6 + 10\times\frac{9}{10}\times 1,3 + 10\times\frac{9}{10}\times 2\right)$

$v2 = 1,368$ ②

① + ② $\Rightarrow s^2 = 2,24 + 1,368 = 3,608 \Rightarrow S^2 = \frac{50}{49}\times 3,608 = 3,68$

$var(\hat{\bar{Y}}) = (1-0,25)\times\frac{3,68}{10} = 0,07$

Exercice 4 :

1. Nombre maximum d'erreurs

n = 200 enregistrements p = 0,05

On a un plan de sondage aléatoire simple avec remise (PEAR)

P (Wm = ui) = 1/N i ∈ {1, . . . , N} et P : Probabilité uniforme

Or p= 0,05 donc N=1/P => N=1/0,05 N=20

EQM[P EAR] = N – 1/N * s^2U/n avec s^2U= 0,25

Pour n= 200

EQM= 20-1/20 * (0.25*0.25)/200 = 1,2

Pour n= 400

EQM= 20-1/20 * (0.25*0.25)/400 =6,5

Pour n= 600

EQM= 20-1/20 * (0.25*0.25)/600 = 6,9

Pour n= 1000

EQM= 20-1/20 * (0.25*0.25)/1000 = 8,1

2. Nombre d'enregistrements supplémentaires

On a 95% = 100(1 − α)% avec α = 0.05. On souhaite déterminer n tel que $n \geq z^2 p\omega (1 - p\omega)/(p\omega d1)^2$, avec d1 = 0.05, zα = 1.96 ω est l'échantillon considéré précédemment et pω = 0.05

On a : $n \geq 1.96^2 \times 0.05 (1 - 0.05)/(0.5 \times 0.05)^2$= 291.9616

Le nombre d'enregistrements supplémentaires est 291.

Exercice 5

1) Donner une estimation du CA Moyen avec un intervalle de confiance :

IC 1−α = [μ ± z1−α √s2]

Avec un risque α=10

Nous avons ICO,9 = [29,43 ;30,19]

y1 5 5 12 12 30 30

y2 150 600 150 600 150 600

µst 63 243 67,2 247,2 78 258

µst= 3/5 y1 + 2/5 y2

L'estimation du chiffre d'affaires moyen µ=159,4

2) Effectifs d'échantillon

a) Allocation proportionnelle :

Soit nh/Nh = n/N

 nh = n* Nh /N

n1 = (300*500)/1060 = 142

n2 = (300* 300)/1060 = 85

n3 = (300* 150)/1060 = 42

n4 = (300* 100)/1060 = 28

n5 = (300* 10)/1060 = 3

b) Allocation optimale ou répartition de Neyman avec nh<= Nh

Posons Nh/Nh Sh = n / $\sum Hk$=1 NhSh

$\sum Hk$=1 NhSh = (500*√1,5) + (300*√4 + (150 *√8) + (100*√100) + (10*√2500)

 $\sum Hk$=1 NhSh = 4460

Nh = n * Nh Sh / $\sum Hk$=1 NhSh

n1 = (300*500*√1,5) / 4460 = 41

n2 = (300*300*√4) / 4460 = 40

n3 = (300*150*√8) / 4460 = 28

n4 = (300*100*√100) / 4460 = 67

n5 = (10*500*√2500) / 4460 = 56

3) Calcul des variances de l'estimateur

V (µst) =(1-f)1/n∑Hk=1 Nh/N Sh

Pour l'allocation optimale :

V(µst) = 0,01

Pour l'allocation proportionnelle :

V(µst) = 0,08