Université Thiès

Master MSDA

Projet Techniques de Sondages

Realisé par : **Khadim CISSE- Seyni KAIRE – Abdoulahi MBENGUE**

<u>Exercice 1 :</u>

Probabilité d'inclusion

Soit la population {1 ,2 ,3}
Et le plan de sondage : P ({1,2}) = ½, P ({1,3})= ¼, P({2,3})= ¼

1) **Est-ce un sondage aléatoire simple ?**

Non, Ce n'est pas un sondage aléatoire simple car tous les échantillons ne sont pas tirés aux même probabilités.

2) **Calculons $\pi 1$, $\pi 2$ et $\pi 3$ les probabilités d'inclusion d'ordre 1**

$\pi 1$ = P ({1,2}) + P ({1,3})= ½ + ¼ = ¾

$\pi 2$ = P ({1,2}) + P ({2,3})= ½ + ¼ = ¾

$\pi 3$ = P ({1,3}) + P ({2,3})= ¼ + ¼ = ¼

3) **Calculons $\pi 12$, $\pi 23$ les probabilités d'inclusion d'ordre 2**

$\pi 12$ = P ({1,2}) = ½

$\pi 23$ = P ({2,3}) = ¼

4) **Déterminons l'estimateur π de $\bar{Y}$**

a) **si l'échantillon {1,2} est tiré**

$Y1 = 1/N \ (Y1/\pi 1 + Y2/\pi 2) = 4/9(Y1 + Y2)$

b) **si l'échantillon {1,3} est tiré**

$Y2 = 1/N \ (Y1/\pi 1 + Y3/\pi 3) = 1/9 \ (4Y1 + 6Y2)$

c)si l'échantillon {2,3} est tiré

$Y3 = 1/N \ (Y2/\pi 2) + (Y3/\pi 3) = 1/9 \ (4Y2 + 6Y3)$

5) **Vérifions que le π-estimateur est un estimateur sans biais**

E($\bar{Y}$)= ½ (4/9 (y1+y2)) + ¼ (1/9 (4y1 + 6y3)) + ¼ (1/9 (4y2 +6y3))
= 1/3 Y1+ 1/3 Y2 + 1/3 Y3
=1/3 (Y1 + Y2 + Y3)
= $\bar{Y}$

Alors le π-estimateur est un estimateur sans biais.

6) **Écrivons ce que seraient les probabilités d'échantillon P et les probabilités d'inclusion π pour un sondage aléatoire simple à probabilités égales sans remise.**

Avec un sondage aléatoire simple a probabilités égales sans remise, on a pour chaque échantillon la même probabilité d'échantillon :

$\pi_{12} = \pi_{13} = \pi_{23} = 1/3$

$\pi_1 = \pi_2 = \pi_3 = 2/3$

EXERCICE2

P^ ± 1.96 √ Var (p^)
Par suite : on cherche la taille de l'échantillon n telle que :
2 x 1,96 √ var (p) ≤ 0,02 è
Or var (p) (sans remise) = (N – n) p (1 – p)
/ (N – 1) n Var (p) (avec remise) = p (1 –
p) / n

P = 3 / 10 ; N = 1500
Par application :
Tirage avec remise :
2 x 1,96√ p (1 – p) / n ≤ 0,02
2 x 196√ p (1 – p) / n ≤ 2

√ p (1 – p) / n = 2 / 2 x 196 ; √ p (1 – p) / n = 1/ 196
√ p (1 – p) / n = 196-1 donc
(√ p (1 – p) / n) 2 = (196-1)2 donc p (1 –
p) / n ≤ 196-2 Or a-n = 1 / an donc on
obtient :
n ≥ 1962 p (1 – p) ;n ≥ 8 067 **(tirage avec
remise)**
Tirage sans remise :
Var (p) = (N – n) p (1- p) / (N – 1) n
(N – n) p (1- p) / (N – 1) n ≤ 196-2

$n \geq 1962\ N\ p\ (1 - p)\ /\ (N - 1 +$
$1962\ p(1 - p)$ $p = 3 / 30$; $N = 1500$
$n \geq 1264$ (tirage sans remise)

2- Que faire si nous ne connaissons pas la proportion :
Si l'on ne connait pas a priori la proportion de personnes affectées, il faudrait alors remplacer Var [pˆ] par son estimation.
Dans ce cas on obtient :
$Var (p) = p (1-p) / (n- p)$: tirage avec remise (AR)
$Var (p) = (N - n) /N * p (1 - p) / n - 1$: tirage sans remise (SR) Notons que le p est un p avec chapeau comme défini ici Var [pˆ].
Une autre approche consisterait à prendre le cas le plus mauvais (ou pessimiste), c'est à dire la valeur théorique de p telle que Var [pˆ] soit le plus grand possible. Clairement le cas le plus pessimiste correspond au choix $p = 0.5$.
Cas sur la consommation des 25 automobilistes au 100km Les hypothèses se traduisent par :

n=25; x_bar=8,5; S(l'écarttype)=0,8

Nous utiliserons la table de student à n-1 degrés deliberté :n−1=24 L'écart type de la population est inconnu:
Si $\alpha = 0{,}05$ alors sur la table de student $t\alpha = 2{,}064$ pour $n = 24$
L'intervalle de confiance ayant 95 chances sur 100 de contenir la valeur de la moyenne est : $X\_bar - t\alpha * S / \sqrt{n-1} \leq m \leq X\_bar + t\alpha * S / \sqrt{n-1}$
$8{,}5 - 2{,}064 * 0{,}8 / \sqrt{24} \leq m \leq 8{,}5 + 2{,}064 * 0{,}8/\sqrt{24}$
$8{,}16 \leq m \leq 8{,}83$
La probabilité que la consommation moyenne soit comprise entre [8,16 ; 8,83] est égale a 95
%.
2 – pour une marge d'erreurs de 2 décilitres : 2 dl = 0,2 L
Or $0{,}2 = t\alpha * S / \sqrt{n - 1}$
$0{,}2 (\sqrt{n-1}) = t\alpha * S$ donc $0{,}2\ 2\ (n-1) = (t\alpha\ S)$
2 par suite : $n = (t\alpha\ S)\ 2 / 0{,}04 + 1$
Avec $t\alpha = 1{,}96$ pour tout $n > 30$ et $\alpha = 0{,}05$

n = (1, 96 * 0,8) 2 / 0, 04 +1 = 62, 5
Pour α = 0,01, dans la table de student tα = 2,576 pour tout n> 30 n = (2,576 * 0,8) 2 / 0, 04
+1 = 106

EXERCICE 3 :

1- **Donnons une estimation totale des notes dans le district** :

M = 50 ; m= 5 par suite f = 5/50= 1/10= 0,1
Dans chaque collège, la note est estimée par :

$T_i = N_i * y_i$ bar : on obtient dans les 05 collèges :
T_1 = 40 * 12 = 480
T_2= 20 * 8 = 160
T_3 = 60 * 10 =600
T_4 = 40 * 12 =480
T_5 = 48 * 11 =528
La note totale du district est estimée par :

$$T = M/m *(Ti) :$$

T = 50/5 (480 + 160 + 600 + 480 + 528) = 22 480

La note totale estimée est à : 22 480
2 – le nombre d'élèves estimé est de :

$N = M/m *(Ni)$:
N = 50 /5 (40 + 20 + 60 + 40 + 48) = 2 080

Le nombre d'élèves estimé est de : 2 080
3-pour N = 2000, donnons une estimation de la moyenne et comparons :

$$Y\_bar = 1 /N * T$$

Y_bar = 1/2000 * 22 480 = 11,24
Par conséquent : la moyenne observée sur N = 50 est de :
: y_bar = 1/50 (10* 12 + 10*8 + 10* 10 + 10* 12 + 10 * 11) = 10,6
Comparons : y_bar n'est un bon estimateur de Y_bar : Y_bar ≠ y_bar 4 – S21 = 1/ 4

Calculons la variance de l'estimateur total :

(480−449, 6)2+ (160-449,6)2+ (600-449,6)2+ (480-449,6)2+
(528−449, 6)2
= 28 620,8 4
M2 (1− fi) S21/ m = 502 * (1-0,1) * 28620,8 4/ 5 = 12 879 360
Maintenant en posant :
Vi =N2i (1−f2, i) S22/ ni Par application :
V1 = 402 * (1 – 10/40) * 1,5 / 10 = 180
V2 = 202 * (1 – 10 / 20) * 1,2 / 10 = 24 donc dans la même logique,
V3 = 480, V4 = 156, V5 = 364,8.
Ainsi en multipliant par M/m, on obtient que la quantité cherchée est égale
: M/m (v1 + v2 + v3 + v4 +v5)
= 50 /5 (180+24+480+156+364,5) = 10 * 1204,8 = 12 048

L'estimation de la variance de l'estimateur du total est égale : Var (T)= 12 879 360 + 12 048 = 12 891 408.
On peut en déduire la variance de la moyenne :
Var (y_bar) = 1/N2 * Var (T)= 1/ 2000 * 12 891 408 = 3,22
5 - Comparaison avec un sondage aléatoire simple à probabilité est égale sur les mêmes données :
Y_bar=y_bar =10,6 ; n=50 et N=2000.
Donc le taux de sondage est égal à :
f= 50/2000 =0,25
L'estimation de la variance de l'estimateur de la moyenne est égale à : Var Y = (1−f) * S2/n, ou S2 est la variance corrigée de l'échantillon.
Dans notre échantillon de taille 50, on a :
Variance totale = variance inter + variance intra
Calculons maintenant chaque terme qui compose la variance totale : Variance inter = 1/50 (10*122 +10*82 + 10*102+10*122+10*112) -10,62= 2,24
Variance intra = 1/ 50 * 0,9*10 (1,5 + 1,2 + 1,6 + 1,3 + 2,0) = 1,368
Donc Variance totale = 2,24 + 1,368 = 3,608
La variance corrigée est de : S2 = 50 / (50 -1) * 3,608 = 3,68
Et Var (Y_bar) = (1 – 0,25) * 3,68/ 50 = 0, 07.
La précision d'un sondage aléatoire simple à probabilité égale sans remise est supérieure à celle d'un sondage à plusieurs degrés :
Pour un intervalle de confiance de 95%, on a :
Plus ou moins : 1,96√var (Y_bar) donc on a les précisions suivantes 0,52 et 3,25.

Exercice 4:

1) **le nombre maximum d'erreurs qu'on peut acceptation dans cet échantillon sans remettre en cause le niveau d'acceptation**

***POUR n=200 on a :**

Nombre d'erreurs= 0,05*200=10 erreurs

Même question avec n=400, n=600 et n=1000

***POUR n=400 on a :**

Nombre d'erreurs =0,05*400= 20 erreurs

***POUR n=600 on a :**

Nombre d'erreurs =0,05*600= 30 erreurs

***POUR n=1000 on a :**

Nombre d'erreurs = 0,05*1000= 50 erreurs

2) le nombre d'enregistrement supplémentaire qu'on doit effectuer pour que l'hypothèse soit acceptée

0 ,05*n= (7+4) => n=11/0,05=180

Donc on doit faire 180 enregistrements supplémentaires pour que l'hypothèse d'un niveau d'acceptation de 5% puisse être raisonnement raisonnablement retenue