République du Sénégal
Un Peuple – Un But – Une Foi

-*-*-*-*-*-*-*-

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

UNIVERSITE DE THIES
UFR SES/SET

Master Science des données et Applications

ESTIMATION

PROJET TECHNIQUE DE SONDAGE

Par :

Ousmane Dia

Abdoulaye Bara DIAW

Ndeye Fatou Diaw

<u>Exercice1</u>

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

P ({1, 2}) = ½ **P** ({1, 3}) = ¼ **P** ({2, 3}) = ¼

1) **Ce sondage n'est pas un sondage aléatoire simple puisque les échantillons ne sont pas équiprobables.**

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

   On a

   $\pi 1 = \sum$ **P (s) = P ({1, 2}) + P ({1, 3})** = ¾ ;

   $\pi 2 = \sum$ **P (s) = P ({1, 2}) + P ({2, 3})** = ¾ ;

   $\pi 3 = \sum$ **P (s) = P ({1, 3}) + P ({2, 3})** = ½

3) **Calcul de $\pi 12$ et $\pi 23$, les probabilités d'inclusion d'ordre 2 :**

$\pi 12 = \pi 1\, \pi 2$ = **[P ({1, 2}) + P ({1, 3})] + [P ({1, 2}) + P ({2, 3})]** = $\frac{3}{2}$

$\pi 23 = \pi 2\, \pi 3$ = **[P ({1, 2}) + P ({2, 3})] + [P ({1, 3}) + P ({2, 3})]** = $\frac{5}{4}$

4) **Le** π- estimateur de Y

   a. **Si l'échantillon {1, 2} est tiré :**

$\frac{y1}{3/4} + \frac{y2}{3/4}$; S = {1, 2} = $\frac{4(y1+y2)}{9}$

   b. **Si l'échantillon {1, 3} est tiré :**

$\frac{y1}{3/4} + \frac{y3}{1/2}$; S = {1, 3} = $\frac{4y1+6y3}{9}$

c. **Si l'échantillon {2, 3} est tiré :**

$$\frac{y2}{3/4} + \frac{y3}{1/2} \; ; \qquad S = \{2, 3\} = \frac{4y2 + 6y3}{9}$$

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

$$E(Y) = \tfrac{1}{2} * \frac{4(y1 + y2)}{9} + \tfrac{1}{4} * \frac{4y1 + 6y3}{9} + \tfrac{1}{4} * \frac{4y2 + 6y3}{9}$$

$$E(Y) = \frac{2y1 + 2y2 + y1 + 1.5y3 + y2 + 1.5y3}{9}$$

$$E(Y) = \frac{3y1 + 3y2 + 3y3}{9}$$

$$E(Y) = \frac{y1 + y2 + y3}{3} = Y$$

Ce qui montre que le π- estimateur est un estimateur sans biais.

Exercice2

1) La taille d'échantillon à sélectionner

La Variance V(p) de l'estimateur est $\frac{\sigma 2\ y}{n}$, avec remise (AR)

$\frac{N-n}{N}\ \frac{S2y}{n}$, sans remise (SR)

Puisque *y2k* = *yk*, la variance et la variance corrigée sur la population sont égales à

$\sigma^2 y = \frac{1}{N} \sum_{U}^{k\,\in} yk = (\frac{1}{N} \sum_{U}^{k\,\in} yk)^2$ = p - p^2 = p (1- p), $S^2 y = \frac{\mathrm{N}}{N-1}$ p (1 -p)

Ainsi, on a Var (p) = $\frac{\mathrm{p(1 - p)}}{n}$, avec remise

= $\frac{\mathrm{N - n}}{N-1}\ \frac{\mathrm{p(1 - p)}}{n}$, sans remise

Supposons que la taille de l'échantillon est suffisamment grande pour que l'approximation soit selon la loi normale soit acceptable, on a donc un intervalle de confiance à 95% de la forme.

$$p \mp 1.96 \times \sqrt{Var}(p).$$

Ainsi on cherche donc la taille de l'échantillon n telle que

$$2 \times 1.96 \times \sqrt{Var(p)} \leq 0.02 \ \Rightarrow Var(p) \leq 196^{-2}$$

$$\Rightarrow \ \frac{\mathrm{p(1-p)}}{n} \leq 196^{-2} \ \text{AR}$$

$$\frac{\mathrm{N-n}}{N-1} \ \frac{\mathrm{p(1-p)}}{n} \leq 196^{-2} \ \text{SR}$$

$$\Rightarrow n \geq 196^2 \, p(1-p) \quad \text{avec remise}$$

$$n \geq 196^2 N \, p(1-p) \ / \ \{N-1+196^2 \ p(1p)\}$$

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

$$n > 8067 \ , \ AR \quad \text{et}$$

$$n > 1264, \ \ SR$$

Avec remise la taille de l'échantillon est supérieure à la taille de la population mère.

Exercice 3

1) Une estimation du total des notes

- Au 1er degré, on a

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

- Au 2eme` degré, on a

Observation	N_i	n_i	$\overline{y}_i$	$s2_i$	$\widehat{T}_i$
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, on estime la note totale $\boldsymbol{T_i}$ par : $Ti = N_i y_i$

Ce qui donne avec les valeurs numériques :

Tb1 = 40×12 = 480, Tb2 = …

On estime la note totale dans le district par

$$\hat{T} = \frac{M}{m}\sum_{i=1} \hat{T}_i$$

$$\hat{T} = \frac{50}{5} * 2248$$

$\hat{T}$ = **22480**

2) Le nombre estimé d'élèves en 6ième

$$\hat{N} = \frac{M}{m}\sum i = 1\, N_i$$

$$= \frac{50}{5} \times (40 + 20 + \cdots + 48)$$

$\hat{N}$ = 2080

3) Estimation de la note moyenne :

On sait que N = 2000 alors

$$\boldsymbol{\overline{Y} = \frac{1}{N} \times \hat{T}}$$

$$\boldsymbol{= \frac{1}{2000} \times 22480}$$

$$\boldsymbol{\overline{Y} = 11.24}$$

Par conséquent, la moyenne observée sur l'échantillon de taille 50 est égale à

$$\bar{y} = \frac{1}{50} \times (10.12 + \cdots + 10.11) = 10.6.$$

En général, $\bar{y}$ n'est pas un bon estimateur de $\overline{\boldsymbol{Y}}$ **.** Il n'y a que dans le cas particulier ou les taux de sondage $f_i = \frac{n_i}{N_i}$ sont constants et si toutes les unités primaires ont la meme taille que $\bar{Y} = \bar{y}$.

4) Calcul de la variance de l'estimateur du total et de la moyenne

Vu qu'on ne peut pas calculer la variance de l'estimateur du total. Donc, on calculera une estimation de cette variance :

$\widehat{Var}(\hat{T}) = M^2(1-f_1)\frac{s^2 1}{m} + \frac{M}{m}\sum N^2 i\left(1-f_{2,i}\right)\frac{s^2 2,i}{n_i}$;

d'ou

$s^2_1 = \frac{1}{m-1}\sum i=1\left(\hat{T}_i - \frac{\hat{T}}{M}\right)^2$

La variance observée entre les unités primaires et

$s^2_{2,i} = \frac{1}{n_i - 1}\sum\left(y_{i,j} - \bar{y}_i\right)^2$

La variance observée entre les unités secondaires.

$s_1^2 = \frac{1}{4}[(480-449{,}6)^2 + \cdots + (528-449{,}6)^2 +] = 28620{,}8$

Qui peut se calculer également grâce à la formule développée suivante

$\frac{1}{m-1}\sum i = 1\,\hat{T}_i^2 - \frac{m}{m-1}\hat{T}^2.$

Maintenant, on peut calculer le premier terme de l'estimation de la variance de l'estimateur de total qui vaut alors

$$M^2(1-f_1)\frac{s_1^2}{m} = 50^2.0{,}9.\frac{28620{,}8}{5} = 12879360$$

On pose

$V_i = N_i^2(1-f_{2,i}\frac{s_{1,i}^2}{n_i}$

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

$V_2 = 24,$

$V_3 = 480,$

$V_4 = 156,$

$V_5 = 364{,}8$

Ce qui implique

$\frac{M}{m}\sum i = 1\,V_i = \frac{50}{5}.1204{,}8 = 12048$.

La variance de l'estimateur total :

$$\widehat{Var}(\hat{T}) = 12\,879\,360 + 12048 = 12\,891\,408$$

La variance de la moyenne :

$$\widehat{Var}(\bar{Y}) = \frac{1}{N^2}\widehat{Var}(\bar{T})$$

$$= \frac{1}{(2000)^2}.\,12\,891\,408$$

$$= 3{,}22$$

La précision est ainsi égale à $\pm 1{,}96\sqrt{3{,}22} = 3{,}5$

Intervalle de confiance de la moyenne $11{,}2 \pm 3{,}5$

La précision n'est donc pas bonne.

5) Comparons

$\bar{Y} = \bar{y} = 10{,}6$; n = 50 et N = 2 000

$$f = \frac{50}{2\,000} = 0{,}25$$

$$Var(\bar{Y}) = (1-f)\frac{s^2}{n}$$

Dans l'échantillon de taille 50, on a
Variance totale = Variance inter + Variance intra

- Variance inter = $\frac{1}{50}.(10.12^2 + \cdots + 10.11^{2)} - 10{,}6^2$
- Variance intra = $\frac{1}{50}.(10.\frac{9}{10}.1{,}5 + \cdots + 10.\frac{9}{10}.2{,}0)$
  = 1.368.

Donc Variance totale égale à $2{,}24 + 1{,}368 = 3{,}608$

La Variance corrigée est égale à $s^2 = \frac{50}{49}.3{,}608 = 3{,}68$

On peut en déduire que $Var\,(\bar{Y}) = (1-0{,}25).\frac{3{,}68}{50} = 0{,}07$

La précision est $\pm 1{,}96\sqrt{0{,}07} \approx \pm\, 0{,}52$

$$\bar{Y} = 10{,}6 \pm 0{,}52$$

On peut en conclure que la précision d'un sondage aléatoire simple à probabilités égales sans remise est supérieure à celle d'un sondage à plusieurs degrés.

Exercice 5 :

1-Donnons une estimation du chiffre d'affaire moyen avec un intervalle de confiance.

La taille des échantillons étant suffisamment grande, on peut supposer sans trop de risque µˆy suit approximativement une loi normale. L'intervalle de confiance en découle et s'écrit

$$\left[\hat{\mu}_y \pm 1.96\sqrt{\frac{N-n}{N}\frac{\hat{S}_y^2}{n}}\right] :$$

2-L'effectif pour chaque strate