Université de Carthage
Ecole Supérieure de la Statistique
et de l'Analyse de l'Information à Tunis

Année Universitaire
2022-2023
Deuxième Année

Théorie des sondages
Série 1 :Sondage aléatoire simple

Exercice 1 *Soit une variable Y définie sur une population de taille $N = 4$ individus.*

i	1	2	3	4
Y	11	10	8	11

1. *Vérifier que* $\mu = E[Y] = 10$ *et* $\sigma^2 = Var[Y] = 1.5$.
2. *On tire un échantillon de taille $n = 2$ sans remise, à probabilités égales.*
   (a) *Combien d'échantillons peut-on tirer ?*
   (b) *Calculer, pour chaque échantillon possible, sa moyenne $\widehat{\overline{Y}}_{pesr}$ et sa variance corrigée s_c^2.*
   (c) *Donner la distribution de $\widehat{\overline{Y}}_{pesr}$ et s_c^2. En déduire $E[\widehat{\overline{Y}}_{pesr}]$, $Var[\widehat{\overline{Y}}_{pesr}]$ et $Var[s_c^2]$.*
   (d) *Comparer et analyser les résultats obtenus.*
3. *On tire, maintenant, un échantillon de taille $n = 2$ avec remise, à probabilités égales.*
   Effectuer le même travail que pour le cas sans remise.

à rendre sur feuille

Exercice 2 *Soit une population de 5 individus. On s'intéresse à une variable d'intérêt Y qui prend les valeurs*

$$Y_1 = Y_2 = 1 \text{ et } Y_3 = Y_4 = Y_5 = \frac{8}{3}$$

On définit le plan suivant:

$$p(\{1,2\}) = \frac{1}{2}, p(\{3,4\}) = p(\{3,5\}) = p(\{4,5\}) = \frac{1}{6}$$

1. *Calculer les probabilités d'inclusion d'ordre 1 et 2.*
2. *Donner la distribution de probabilité du π-estimateur et vérifier que cet estimateur est sans biais.*
3. *Calculer* $Var\left(\widehat{T}_\pi\right)$

4. Calculer l'estimateur de la variance de $\widehat{T}_\pi$. Cet estimateur est-il sans biais? Etait-ce prévisible?

5. On se propose d'estimer la racine carrée du total $\left(\sqrt{T}\right)$, par la racine carrée du π-estimateur $\left(\sqrt{\widehat{T}_\pi}\right)$. Donner la distribution de probabilité de cet estimateur.

6. Calculer l'espérance de $\sqrt{\widehat{T}_\pi}$. Commenter.

7. Calculer sa variance.

Exercice 3 *Exercice préliminaire aux principales méthodes de sondage (d'après Statistique inférentielle : Idées, démarches, exemples de Jean-Jacques Daudin, Stéphane Robin et Colette Vuillet).*

La figure jointe contient une population de 100 cercles.

1. **Echantillonnage empirique** : *Choisir arbitrairement 5 cercles distincts (en les marquant d'une croix par exemple). Sachant que les cercles mesurent 1cm, 2cm, 3cm, 4cm, 5cm, 6cm, 7cm, 8cm, 9cm et 10cm, calculer le diamètre moyen des cercles de votre sélection. On le notera d_1.*

2. **Echantillonnage aléatoire simple** : *A l'aide d'un générateur (ou de la table) de nombres aléatoires, tirer un echantillon de 5 cercles et calculer le diamètre moyen des cercles de votre sélection. On le notera d_2.*

3. **Recensement** :

   (a) *Calculer la distribution du diamètre dans la population des 100 cercles.*

   (b) *En déduire la vraie valeur du diamètre moyen D.*

4. **Echantillonnage systématique** :

   (a) *A l'aide d'un générateur (ou de la table) des nombres aléatoires, tirer un nombre l compris entre 1 à 20.*

   (b) *L'échantillon sera constitué par les cercles n° $l, l+20, l+40, l+60$ et $l+80$.*

   (c) *Calculer d_3, le diamètre moyen de l'échantillon.*

5. **Echantillonnage stratifié** :

   (a) *Diviser la population des 100 cercles en deux catégories. Les 51 cercles dont le diamètre est 1cm, 2cm ou 3cm forment les petits cercles et les 49 dont le diamètre est compris entre 4cm et 10cm sont appelés les grands cercles.*

(b) *Numéroter les petits cercles de 1 à 51 et, à l'aide d'un générateur (ou de la table) de nombres aléatoires tirer au hasard 3 cercles parmi les 51.*

(c) *Numéroter les grands cercles de 1 à 49 et, à l'aide d'un générateur (ou de la table) de nombres aléatoires, tirer au hasard 2 cercles parmi les 49.*

(d) *Nous obtenons ainsi un diamètre moyen pour les petits cercles et un diamètre moyen pour les grands cercles. En déduire un diamètre moyen d_4 de l'échantillon.*

6. **Echantillonnage par grappes**

   (a) *Regrouper arbitrairement les cercles par paquets de 5 cercles. On a ainsi 20 paquets que l'on numérotera de 1 à 20.*

   (b) *A l'aide d'un générateur (ou de la table) des nombres aléatoires, tirer 2 paquets parmi les 20.*

   A (c) *Nous obtenons ainsi un diamètre moyen pour chacun des deux paquets. En déduire un diamètre moyen d_5 de l'échantillon.*

7. **Echantillonnage à deux degrés :**

   (a) *Regrouper arbitrairement les cercles par paquets de 10 cercles. On a ainsi 10 paquets que l'on numérotera de 1 à 10.*

   (b) *A l'aide d'un générateur (ou de la table) des nombres aléatoires, tirer un paquet parmi les 10, puis numéroter les cercles du paquet sélectionné de 1 à 10.*

   (c) *A l'aide d'un générateur (ou de la table) des nombres aléatoires, tirer dans le paquet sélectionné un échantillon de 5 cercles parmi les 10.*

   (d) *Calculer le diamètre moyen d_6 de l'échantillon.*

   (e) *Comparer d_1, d_2, d_3, d_4, d_5, d_6 et D. Commenter.*

D = 7,7

d1 = 11,6

d2 = 6,8

d3 = 10,8

d4 = 8,26

dr = 7,8

d6 = 10cm

Série I : Sondage aléatoire simple

Exercice 1 :

1/ $E(Y) = \frac{1}{N} \cdot \sum_{i=1}^{4} Y_i = \frac{11+8+10+11}{4} = 10$

$\sigma^2 = Var(Y) = \frac{1}{N} \sum_{i=1}^{4} (Y_i - \bar{Y})^2 = \frac{1}{4} \cdot (1^2 + (-2)^2 + 1^2) = \frac{6}{4} = 1,5$

2_a/ nbr d'échantillons possible est $C_4^2 = 6$

b/ $\hat{\bar{Y}}_{pesr} = \frac{1}{m} \cdot \sum_{i=1}^{m} y_i = \bar{y}$ $\quad s_c^2 = \frac{1}{m-1} \sum_{i=1}^{m} (y_i - \bar{y})^2$

	$\{1,2\}$ (11, 10)	$\{1,3\}$ (11, 8)	$\{1,4\}$ (11, 11)	$\{2,3\}$ (10, 8)	$\{2,4\}$ (10, 11)	$\{3,4\}$ (8, 11)
$\hat{\bar{Y}}_{pesr}$	10,5	9,5	11	9	10,5	9,5
S_c^2	0,5	4,5	0	2	0,5	4,5
$P[\bar{Y}=\theta]$	1/3	1/3	1/6	1/6		
$P[S_c^2=P]$	1/3	1/3	1/6	1/6	1/3	

$(0,5)^2 + (0,5)^2 = 0,5$

c/ $\to P(E=e) = 1/6$

$E(\hat{\bar{Y}}_{psr}) = \sum_{i=1}^{m} \theta_i \, P(\hat{\bar{Y}}_{psr} = \theta_i) = 10,5 \times 1/3 + 9,5 \times 1/3 + 11 \times 1/6 + 9 \times 1/6$
$= 10$

$E(\hat{\bar{Y}}_{pesr}^2) = \sum_{i=1}^{m} \theta_i^2 \, P(\hat{\bar{Y}}_{pesr} = \theta_i) = (10,5)^2 \times 1/3 + (9,5)^2 \times 1/3 + 11^2 \times 1/6$
$+ 9^2 \times 1/6 = 36,75 + 30,08 + 20,17 + 13,5 = 100,5$

$V(\hat{\bar{Y}}_{pesr}) = E(\hat{\bar{Y}}_{pesr}^2) - (E(\hat{\bar{Y}}_{pesr}))^2 = 100,5 - 100 = 0,5$

$E(S_c^2) = \sum_{i=1}^{m} P_i \, P(S_c^2 = P_i) = 0,5 \times 1/3 + 4,5 \times 1/3 + 2 \times 1/6$
$= 1/6 + 1,5 + 2/6 = 2$

$E((S_c^2)^2) = \sum_{i=1}^{m} P_i^2 \, P(S_c^2 = P_i) = (0,5)^2 \times 1/3 + (4,5)^2 \times 1/3 + 4 \times 1/6$
$= \frac{0,25}{3} + \frac{20,25}{3} + \frac{2}{3} = 7,5$

$Var(S_c^2) = 7,5 - 2^2 = 3,5$

d/ Oma $\sigma_c^2 = \frac{1,5 \times 4}{3} = 2$ $\quad \sigma^2 = 1,5$

Oma $E(S_c^2) = \sigma^2$ (prévisible car c'est un estimateur sans biais)

$E(\hat{\bar{Y}}_{pesr}) = \bar{Y}$ estimateur sans biais

$Var(\hat{\bar{Y}}_{pesr}) = 0,5$ $\left((1-f)\frac{\sigma_c^2}{m} = (1-\frac{1}{2}) \cdot \frac{2}{2} = 0,5 \right)$

$Var(S_c^2) = 3,5$

Exercice 2 :

Université de Carthage
Ecole Supérieure de la Statistique
et de l'Analyse de l'Information à Tunis

Année Universitaire
2022-2023
Deuxième Année

Théorie des sondages

Série 2 :Sondage stratifié

Exercice 1 *L'objectif est d'estimer le poids total de 100 éléphants afin de les embarquer pour traverser le Gange. on dispose des résultats d'une pesée effectuée l'année précédente. Les valeurs sont consignées dans le tableau qui suit (valeurs en tonnes):*

	Effectifs N_h	Moyennes $\overline{Y}_h$	Dispersions σ_h^2
Mâles	60	6	4
Femelles	40	4	2,25

1. *Calculer, pour l'année précédente, la dispersion dans la population de la variable Y (poids de l'éléphant).* $\sigma^2 = \underbrace{\sum_{h=1}^{n} \frac{N_h}{N} \sigma_h^2}_{intra} + \underbrace{\sum_{h=1}^{n} \frac{N_h}{N} (\overline{Y}_h - \overline{Y})^2}_{inter}$ (2) $\quad \overline{Y}$? (1)
2. *On procède à un tirage aléatoire simple sans remise de 10 éléphants. En partant de l'hypothèse que la dispersion n'évolue pas sensiblement d'une année à l'autre, quelle serait la variance de l'estimateur du poids total des éléphants?*
3. *Qu'en est-il dans le cas d'un sondage stratifié à allocations proportionnelles (n vaut toujours 10)?* $Var(\hat{\overline{Y}}_{pop}) = \frac{1}{m}(1-f) \sum \frac{N_h}{N} \sigma_{hc}^2 = \frac{1}{m}(1-f) \sum \frac{N_h}{N} \cdot \frac{N_h}{N_h - 1} \sigma_h^2$ $\Rightarrow Var(\hat{T}_{pop}) = N^2 Var(\hat{\overline{Y}}_{pop})$
4. *On opte maintenant pour un sondage stratifié optimal de 10 éléphants. Donner les effectifs de l'échantillon dans chacune des strates et calculer la variance de l'estimateur du total.*

$Var(\hat{\overline{Y}}_\pi)$ (1) $= \frac{1}{m}(1-f)\sigma_c^2 = \frac{1}{m}(1-f) \frac{N\sigma^2}{N-1}$

$\Rightarrow Var(\hat{T}_\pi)$ (2) $= N^2 Var(\hat{\overline{Y}}_\pi)$

$m_h = \frac{N_h \sigma_{hc}}{\frac{1}{m} \sum_{h=1}^{m} N_h \sigma_{hc}}$

$\Rightarrow Var(\hat{T}_{opt}) = \frac{1}{m} \left(\sum_{h=1}^{m} N_h \sqrt{\sigma_{hc}^2} \right)^2 - \sum_{h=1}^{m} N_h \sigma_{hc}^2$

Exercice 2 *L'ordre des médecins a procédé à la classification des médecins d'une ville en 3 groupes distincts:*

- *les débutants ou classe 1 (500 médecins)*
- *les confirmés ou classe 2 (1000 médecins)*
- *les très expérimentés ou classe 3 (2500 médecins)*

  Dans chaque classe, on tire, par sondage aléatoire simple sans remise, un échantillon de 200 médecins. On calcule alors le nombre moyen de patients par jour et par médecin dans chacun des échantillons tirés. Les résultats obtenus sont comme suit:

- *Classe 1 :* $\hat{\overline{Y}}_1 = 10 \qquad s_{1c}^2 = 4$

- *Classe 2 :* $\widehat{\overline{Y}}_2 = 15 \qquad s^2_{2c} = 7$

- *Classe 3 :* $\widehat{\overline{Y}}_3 = 20 \qquad s^2_{3c} = 10$

1. *Estimer le nombre moyen de patients soignés par jour et par médecin.*

2. *Donner un intervalle de confiance de niveau 95% pour ce paramètre.*

Exercice 3 *On souhaite estimer la moyenne d'une variable d'intérêt relative à l'ensemble des entreprises d'un département. Ces entreprises sont classées selon leur chiffre d'affaires et répertoirées en trois classes. Les données issues d'un recensement antérieur sont les suivantes :*

Chiffres d'affaires	*Nombre d'entreprises*
de 0 à 1	*1000*
de 1 à 10	*100*
de 10 à 100	*10*

Nous avons fixé une taille d'échantillon de 111 entreprises. En supposant que la distribution du chiffre d'affaires est une loi uniforme au sein de chaque classe, donner la variance de l'estimateur de la moyenne du chiffre d'affaires pour un plan stratifié avec allocations proportionnelles puis optimales. Commenter. (on utilisera un sondage aléatoire simple dans chaque strate).

Mâle $\bar{Y}_h = 6$, $\sigma_h^2 = 4$ — 60 — 100 — 40 — Femelle $\bar{Y}_h = 4$, $\sigma_h^2 = 2,25$

1/ $\bar{Y} \longrightarrow \sigma^2 = \sum_{h=1}^{n} \frac{N_h}{N}\cdot\sigma_h^2 + \sum_{h=1}^{n}\frac{N_h}{N}(\bar{Y}_h - \bar{Y})^2$

On a $\bar{Y} = \frac{1}{100}\times(60\times6 + 40\times4) = 5,2$

alors $\sum_{h=1}^{2}\frac{N_h}{N}\sigma_h^2 = \frac{60}{100}\times4 + \frac{40}{100}\times2,25 = 3,3$

$\sum_{h=1}^{2}\frac{N_h}{N}(\bar{Y}_h - \bar{Y})^2 = \frac{60}{100}(6-5,2)^2 + \frac{40}{100}(4-5,2)^2 = 0,96$

d'où $\sigma^2 = 3,3 + 0,96 = 4,26$

2/ $Var(\hat{T}_\pi) = N^2\cdot Var(\hat{\bar{Y}}_\pi) = N^2\cdot\frac{(1-f)}{m}\cdot\sigma_c^2 = \frac{1}{m}(1-f)\cdot N^2\cdot\frac{N}{N-1}\sigma^2$

alors $Var(\hat{T}_\pi) = \frac{1}{m}\cdot N^2(1-f)\cdot\frac{N}{N-1}\sigma^2 = \frac{1}{10}\cdot100^2(1-0,1)\cdot\frac{100}{99}\times4,26$

$= 3872,73$

3/ $Var(\hat{T}_{prop}) = N^2\cdot var(\hat{\bar{Y}}_{prop}) = N^2\cdot\frac{1}{m}(1-f)\cdot\sum\frac{N_h}{N}\cdot\sigma_{hc}^2$

$= N^2\cdot\frac{1}{m}(1-f)\cdot\sum\frac{N_h}{N}\frac{N_h}{N_h-1}\sigma_h^2$

alors $Var(\hat{T}_{prop}) = \frac{N^2}{m}\cdot(1-f)\cdot\sum\frac{N_h^2}{N(N_h-1)}\cdot\sigma_h^2$

$= \frac{100^2}{10}(1-0,1)\cdot\left(\frac{40^2}{100(39)}\cdot2,25 + \frac{60^2}{100(59)}\cdot4\right) = 3027,38$

4/ $m_h = \frac{N_h\cdot\sigma_{hc}}{\frac{1}{m}\sum N_h\cdot\sigma_{hc}}$ où $\sigma_{hc} = \sqrt{\sigma_{hc}^2} = \sqrt{\frac{N_h}{N_h-1}\cdot\sigma_h^2} = \sqrt{\frac{N_h}{N_h-1}}\cdot\sigma_h$

$Var(\hat{T}_{opt}) = \frac{1}{m}\cdot\left(\sum N_h\cdot\sigma_{hc}\right)^2 - \sum N_h\cdot\sigma_{hc}^2$ valeur réel $= \sum_{h=1}^{n} N_h^2\cdot\frac{1}{m_h}(1-f_h)$ valeur approchée

$m_{hF} = \frac{N_{hF}\,\sigma_{hcF}}{\frac{1}{m}\cdot\sum N_h\cdot\sigma_{hc}} = \frac{40\cdot\sqrt{\frac{40}{39}\times2,25}}{\frac{\left(60\cdot\sqrt{\frac{60}{59}\cdot4} + 40\sqrt{\frac{40}{39}\cdot2,25}\right)}{10}} = 3,34 \simeq 3$

$m_{hH} = \frac{60\sqrt{\frac{60}{59}\times4}}{*} = 6,66 \simeq 7$

Exercice 2 :

Classe 1	Classe 2	Classe 3
$N_1 = 500$	$N_2 = 1000$	$N_3 = 2500$
$\hat{\bar{Y}}_1 = 10$	$\hat{\bar{Y}}_2 = 15$	$\hat{\bar{Y}}_3 = 20$
$S_{1c}^2 = 4$	$S_{2c}^2 = 7$	$S_{3c}^2 = 10$

$\underbrace{m_1 = m_2 = m_3 = 200}_{\text{Sondage à allocation fixe}}$

1/ $\hat{\bar{Y}}_{strat} = \sum_{h=1}^{3} \frac{N_h}{N} \hat{\bar{Y}}_{h pers}$ $\quad N = 4000$

$\hat{\bar{Y}}_{strat} = \frac{1}{4000} \left(500 \times 10 + 1000 \times 15 + 2500 \times 20 \right) = 17,5$

2/ $IC_{0,95}(\bar{Y}) = \left[\hat{\bar{Y}}_{strat} \pm 1,96 . \sqrt{\hat{Var}(\hat{\bar{Y}}_{strat})} \right]$

$Var(\hat{\bar{Y}}_{strat}) = \sum_{h=1}^{3} \frac{N_h^2}{N^2} (1 - f_h) \sigma_{hc}^2 = \sum_{h=1}^{3} \frac{N_h^2}{N^2} (1 - f_h) \frac{S_{hc}^2}{m_h} \quad f_h = \frac{m_h}{N}$

$= \frac{500^2}{4000^2} \left(1 - \frac{200}{500}\right) . \frac{4}{200} + \frac{1000^2}{4000^2} \left(1 - \frac{200}{1000}\right) . \frac{7}{200} + \frac{2500^2}{4000^2} \left(1 - \frac{200}{2500}\right) \frac{10}{200}$

$= 3,983$

alors $IC(\bar{Y}) = \left[17,5 \pm 1,96 . \sqrt{3,983} \right] = \left[13,588 ; 21,412 \right]$

Exercice 3 :

...ations proportionnelles :

$$\sigma^2_{CA1} = \frac{1}{12} \quad \sigma^2_{CA2} = \frac{9^2}{12} \quad \sigma^2_{CA3} = \frac{90^2}{12}$$

$$Var(\hat{\bar{Y}}_{prop}) = \sum_{h=1}^{3} \frac{N_h^2}{N^2}(1-f)\frac{\sigma_{hc}^2}{m_h} = \frac{1-f}{mN}\sum_{h=1}^{3} N_h \frac{\sigma_h^2}{(N_h-1)}$$

$$= \frac{1-\frac{111}{1110}}{111.\left(\frac{111}{1110}\right)}\left(\frac{1000^2}{1000-1}\times\frac{1}{12} + \frac{100^2}{(100-1)}\times\frac{9^2}{12} + \frac{10^2}{10-1}\times\frac{90^2}{12}\right)$$

$$= 669,48$$

Cas Optimal :

$$m_h = \frac{N_h.\sigma_{hc}}{\frac{1}{m}\sum_{K=1}^{m} N_h.\sigma_{hc}} \Rightarrow N_1.\sigma_{1c} = 1000\sqrt{\frac{1000}{999}.\frac{1}{12}} = 288,81$$

$$N_2.\sigma_{2c} = 100\sqrt{\frac{100}{99}.\frac{9^2}{12}} = 26,11$$

$$N_3.\sigma_{3c} = 10.\sqrt{\frac{10}{9}.\frac{90^2}{12}} = 273,86$$

$$m_1 = 111 \times \frac{288,81}{261,1 + 288,81 + 273,86} = 38,91$$

$$m_2 = 111 \times \frac{261,1}{261,1 + 288,81 + 273,86} = 35,18$$

$$m_3 = 111 \times \frac{273,86}{261,1 + 288,81 + 273,86} = 36,9$$

parce que la valeur depasse la taille $\Rightarrow$ max $=10$

on pose $m'_3 = 10 \Rightarrow m' = 111 - 10 = 101$

alors $m'_1 = 101 \times \frac{288,81}{261,1 + 288,81} = 53,11$

$$m'_2 = 101 \times \frac{261,1}{261,1 + 288,81} = 48,01$$

on prend $m'_1 = 53 \quad m'_2 = 48 \quad m'_3 = 10$

$$Var(\hat{\bar{Y}}_{opt}) = \sum_{h=1}^{3} \frac{N_h^2}{N^2}\left(1 - \frac{m_h}{N_h}\right)\frac{\sigma_h^2}{m_h}$$

$$= \frac{1}{1110^2}\left(1000^2\left(1-\frac{53}{1000}\right)\left(\frac{1}{12}\right).\frac{1}{53} + \frac{100^2}{48}\left(1-\frac{48}{100}\right)\frac{9^2}{12}\right)$$

Université de Carthage
Ecole Supérieure de la Statistique
et de l'Analyse de l'Information à Tunis

Année Universitaire
2022-2023
Deuxième Année

Théorie des sondages

Série 3 :Sondage à Probabilités inégales

Exercice 1 *Soient une population* $\mathcal{P} = \{1,2,3\}$ *et le plan suivant :*

$$P\{1,2\} = \frac{1}{2},\ P\{1,3\} = \frac{1}{4},\ P\{2,3\} = \frac{1}{4}$$

Donner les probabilités d'inclusion d'ordre 1 *et la matrice de variance-covariance des indicatrices d'appartenance à l'échantillon.*

Exercice 2 *Soit la matrice de covariance* Γ *des indicatrices de la présence des unités dans l'échantillon pour un plan sans remise* Π *donnée par*

$$\Gamma = \frac{6}{25}\begin{bmatrix} 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 \end{bmatrix}$$

1. *Peut-on dire que le plan est de taille fixe? Justifier.*
2. *Calculer les probabilités d'inclusion d'ordre 1 sachant que* $\pi_1 = \pi_2 = \pi_3 > \pi_4 = \pi_5$
3. *En déduire la matrice des probabilités d'inclusion d'ordre 2.*
4. *Calculer, pour tous les échantillons possibles de taille* $n \geq 2$, *les probabilité qu'ils soient sélectionnés.*

Sondage à probabilités inégales

Exercice 1 :

$\pi_1 = P\{1,3\} + P\{1,2\} = 3/4$

$\pi_2 = P\{1,2\} + P\{2,3\} = 3/4$

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

$\Delta_{11} = \pi_1 - \pi_1^2 = (3/4) - (3/4)^2 = \frac{3}{16}$

$\Delta_{22} = \pi_2 - \pi_2^2 = 3/16$

$\Delta_{33} = \pi_3 - \pi_3^2 = (1/2) - (1/2)^2 = 1/4$

Calculons les probabilités d'inclusion d'ordre 2

$\pi_{12} = 1/2 \quad \pi_{13} = 1/4 \quad \pi_{23} = 1/4$

$\Delta_{12} = \pi_{12} - \pi_1\pi_2 = -1/16 \quad \Delta_{13} = \pi_{13} - \pi_1 . \pi_3 = -1/8 \quad \Delta_{23} = -1/8$

$\Rightarrow$ La matrice de variance-covariance

$$= 1/16 \begin{pmatrix} 3 & -1 & -2 \\ -1 & 3 & -2 \\ -2 & -2 & 4 \end{pmatrix}$$

Exercice 2 :

1) $\sum_{i=1}^{m} \pi_i = m$, si le plan est de taille fixe, on a $\# E = m = \sum_{i \in E} 1\!\!1_{\{i \in E\}}$

$Cov(1\!\!1_{\{i\in E\}}, 1\!\!1_{\{j \in E\}}) = \sum_{j \in P} Cov(1\!\!1_{\{i \in E\}} 1\!\!1_{\{j \in E\}})$; soit $i \in P$

$= Cov(1\!\!1_{\{i\in E\}}, \sum_{j \in P} 1\!\!1_{\{j \in E\}}) = Cov(1\!\!1_{\{i \in E\}}, m) = 0$

donc $Cov(1\!\!1_{\{i\in E\}}, 1\!\!1_{\{j \in E\}}) = 0 \neq \sum L_i$

donc le plan n'est pas de taille fixe

2) $\Delta_i = \pi_i(1-\pi_i) = 6/25 \Rightarrow \pi_i - \pi_i^2 - \frac{6}{25} = 0$ donc $\pi_i \in \{2/5, 3/5\}$

or $\pi_1 = \pi_2 = \pi_3 > \pi_4 = \pi_5$ alors $\pi_1 = \pi_2 = \pi_3 = 3/5$ et $\pi_4 = \pi_5 = 2/5$

3) $\Delta_{ij} = \pi_{ij} - \pi_i\pi_j \quad \forall i \in P, j \in P, i \neq j \qquad \pi_{ij} = \pi_{ji} = P\{\underbrace{i,j,\dots}_{Ech}\}$

$\pi_{ij} = \Delta_{ij} + \pi_i . \pi_j \qquad \pi_{12} = \frac{6}{25} + \frac{9}{25} = \frac{15}{25}$

$\pi_{13} = \frac{6}{25} + \frac{9}{25} = \frac{15}{25} \qquad \pi_{14} = \frac{-6}{25} + \frac{6}{25} = 0 \qquad \pi_{15} = \frac{-6}{25} + \frac{6}{25} = 0$

$\pi_{23} = \frac{6}{25} + \frac{3}{5} \times \frac{3}{5} = \frac{15}{25}$ $\pi_{24} = \pi_{25} = 0$ $\pi_{34} = \pi_{35} = 0$

$\pi_{45} = \frac{6}{25} + \frac{4}{25} = \frac{10}{25}$

$$\begin{pmatrix} 0 & 15/25 & 15/25 & 0 & 0 \\ \frac{15}{25} & 0 & \frac{15}{25} & 0 & 0 \\ 15/25 & \frac{15}{25} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{10}{25} \\ 0 & 0 & 0 & \frac{10}{25} & 0 \end{pmatrix}$$

Matrice des proba d'inclusion d'ordre 2

4) $\{1,2\}$; $\{1,3\}$; $\{2,3\}$; $\{4,5\}$; $\{1,2,3\}$;

On a $\pi_1 = \pi_2 = \pi_3 = \pi_{12} = \pi_{13} = \pi_{23}$

$\Rightarrow$ La proba de sélectionner $\{1\}$ = la proba de sélectionner $\{1,2\}$

La proba de sélectionner $\{2\}$ = la proba de sélectionner $\{1,2\}$

La proba de sélectionner $\{3\}$ = la proba de sélectionner $\{1,3\}$ et $\{2,3\}$

La proba de sélectionner $\{1,2,3\}$ = la proba de sélectionner $\{1\}$ ou $\{2\}$ ou $\{3\} = 3/5$

Il s'agit donc d'un plan par grappe où on peut choisir la grappe $\{1,2,3\}$ ou $\{4,5\}$

3980 Ag

$| m_1 | m_2 | \dots | \dots | \dots | \dots | \dots | m_K |$

$m_1 = \dots = m_K = 10$ $\qquad \sum m_K = 39800$

Université de Carthage
Ecole Supérieure de la Statistique
et de l'Analyse de l'Information à Tunis

Année Universitaire
2022-2023
Deuxième Année

Théorie des sondages

Série 4 :Sondage à 2 degrés

40 Ag
&
400 Clt

Exercice 1 *Une banque a 39800 clients dans ses fichiers répartis en 3980 agences gérant chacune exactement 10 clients. On désire estimer la proportion de clients ayant octroyé un crédit auprès de la banque. On sélectionne par sondage aléatoire simple 40 agences et on dénombre dans chaque agence h, A_h clients bénéficiaires d'un prêt. On alors*

$1 \leq h \leq 40$

échantillon $\underline{\sum_{h \in E_1}} A_h = 185 \qquad \sum_{h \in E} A_h^2 = 1263$

$0 \leq A_h \leq 10$

1. *Quelle est la nature de ce sondage?*
   Par Grappe
2. *Donner l'expression du paramètre à estimer et de son estimateur.*
3. *Estimer sans biais la variance de cet estimateur et en déduire un intervalle de confiance de niveau asymptotique 95%*

Exercice 2 *Le ministère de l'éducation souhaite évaluer le niveau des élèves lors de leur entrée en collège. A ce titre, il met à l'essai une enquête par sondage sur un seul gouvernorat. Dans un premier temps, on tire 5 collèges parmi les 50 collèges du gouvernorat désigné, selon un sondage aléatoire simple à probabilités égales sans remise. Ensuite, dans chacun des 5 collèges sélectionnés, on effectue un test de niveau sur un échantillon de 10 élèves (On suppose qu'il y a 1 seule classe de 7ème par collège).*

A l'issue du sondage, on a calculé pour chaque collège la note moyenne des 10 élèves évalués ainsi que la variance corrigée dans l'échantillon:

Collège	1	2	3	4	5
Effectif des classes de 7ème	40	20	60	40	48
Note moyenne	12	8	10	12	11
Variance corrigée des notes	1,5	1,2	1,6	1,3	2,0

Rappels :

- *Dans le cas d'un SAS, on a* $Var\left(\widehat{\overline{Y}}\right) = (1-f)\,\frac{\sigma_c^2}{n}$.
- $Var\,(U_p) = \sum_{h=1}^{M} \frac{T_h^2}{\pi_{1h}}(1-\pi_{1h}) + \sum_{h=1}^{M}\sum_{\substack{k=1\\k\neq h}}^{M} \frac{T_h}{\pi_{1h}}\frac{T_k}{\pi_{1k}}(\pi_{1hk} - \pi_{1h}\pi_{1k})$

- $Var\left(U_s\right) = \sum_{h=1}^{M} \frac{1}{\pi_{1h}} Var\left(\widehat{T}_h\right)$

1. *Quelle est la nature de ce sondage aléatoire? Expliquer.*

2. *Donner une estimation du total des notes sur tout le gouvernorat.*

3. *Estimer le nombre total d'élèves inscrits en 7ème dans le gouvernorat.*

4. *En supposant qu'il y a exactement 2000 élèves en 7ème dans le gouvernorat, donner une estimation de la note moyenne. Comparer avec la moyenne observée sur l'échantillon.*

5. *Exprimer la variance de l'estimateur du total des notes en fonction des totaux et de la dispersion des notes par collège.*

6. *L'estimation de la variance de* $\widehat{T}$ *donne* $\widehat{Var}\left(\widehat{T}\right) = 12\ 891\ 408$

   En déduire un intervalle de confiance de niveau 95% *pour le niveau moyen des élèves de 7ème.*

Serie 4 : Sondage à 2 degrés

Exercice 1 :

$\sum_{h\in E} A_h = 185$ $\quad \sum_{h\in E} A_h^2 = 1263$ $\quad N = 39.800$ clients

$n = 3980$ Agences

$m = 40$ Agences

$N_h = 10$ Clients $\forall h$ (taille d'une grappe)

1/ $\pi_{1h} = \frac{m}{n} = \frac{40}{3980} = \frac{2}{199}$ La probabilité de choisir une Agence

Puis on dénombre chaque agence $\Rightarrow$ Sondage par grappe

où la grappe désigne une agence

2/ $P = \frac{1}{N} \cdot \sum_{h=1}^{3980} \sum_{i\in S_h} \mathbb{1}_{\{c = oui\}}$

A_h : nombre des clients qui ont des prêts

$A_h = \sum_{i\in h} \mathbb{1}_{\{c=oui\}}$ $\quad P = \frac{1}{N} \cdot \sum_{h=1}^{3980} A_h \Rightarrow \hat{P}_{grap} = \frac{1}{N} \cdot \sum_{h\in E_I} \frac{A_h}{\pi_{Ih}}$

$\hat{P}_{grap\ posr} = 0{,}46$

3/ $Var(\hat{P}_{grap\ posr}) = Var\left(\frac{1}{N} \cdot \sum_{h\in E_I} \frac{A_h}{\pi_I}\right)$

$$= \frac{1}{N^2} \cdot Var\left(\sum_{h\in E_I} \frac{A_h}{\pi_{Ih}}\right)$$

$$= \frac{1}{N^2} \cdot \frac{n}{m} \cdot \frac{n-m}{n-1} \sum_{h=1}^{n} \left(A_h - \frac{\sum_{h=1}^{n} A_h}{n}\right)^2$$

$$= \frac{1}{(39800)^2} \frac{3980}{40} \cdot \frac{3980-40}{3980-1} \sum_{h=1}^{3980} \left(A_h - \frac{\sum_{h=1}^{3980} A_h}{3980}\right)$$

$$= 24.10^{-5} \left[\sum_{h=1}^{3980} \left(A_h - \frac{\sum_{h=1}^{3980} A_h}{n}\right)^2\right] \cdot \frac{1}{n-1}$$

$$\Rightarrow \hat{Var}(\hat{P}_{grap\ posr}) = 24.10^{-5} \cdot \frac{1}{m-1} \cdot \left(\sum_{h=1}^{40} \left(A_h - \frac{\sum_{h=1}^{40} A_h}{m}\right)^2\right)$$

$$= \frac{24.10^{-5}}{39} \sum_{h=1}^{40} \left(A_h^2 + \left(\frac{\sum A_h}{m}\right)^2 - 2A_h \cdot \frac{\sum A_h}{m}\right)$$

$$= \frac{24.10^{-5}}{39} \cdot \left(\sum A_h^2 + \sum \left(\frac{\sum A_h}{m}\right)^2 - 2 \cdot \sum \left(A_h \cdot \frac{\sum A_h}{n}\right)\right)$$

$$= \frac{24.10^{-5}}{39} \left(1263 + 40 \times \cdot \left(\frac{185}{3980}\right)^2 - 2 \cdot \frac{(185)^2}{3980}\right) = 7{,}87 \times 10^{-3}$$

$0{,}086425$

$$IC_{gr\%}(P) = \left[\hat{P}_{grp\ pur} \pm 1{,}96\ \sigma_{\hat{P}_{grp\ pur}}\right] = \left[0{,}46 \pm 0{,}17\right]$$
$$= \left[0{,}29\ ;\ 0{,}63\right]$$