UNIVERSITÉ DE THIES

UFR DES SCIENCES ECONOMIQUES ET SOCIALES

UFR DES SCIENCES ET TECHNOLOGIQUES

Master Science Des Données et Applications

Par

JOHANNA BINTA VITALE FAYE

ALMAMY YOUSSOUF LY

COUMBA SY

Sur le sujet

Techniques de sondages

Professeur : Dr. Fatou Néné Diop

Année universitaire 2019-2020

Exercice 1 $\boxed{1\_1}$

Population : $\{1, 2, 3\}$

Plan de sondage : $P(\{1,2\}) = \frac{1}{2}$, $P(\{1,3\}) = \frac{1}{4}$, $P(\{2,3\}) = \frac{1}{4}$

1- sondage aléatoire simple

Comme que tous les échantillons n'ont pas la même probabilité alors on n'a pas de sondage aléatoire simple

2- Calculons π_1, π_2 et π_3

$$\pi_1 = P(\{1,2\}) + P(\{1,3\})$$
$$= \frac{1}{2} + \frac{1}{4}$$
$$\boxed{\pi_1 = \frac{3}{4}}$$

$$\pi_2 = P(\{1,2\}) + P(\{2,3\})$$
$$= \frac{1}{2} + \frac{1}{4}$$
$$\boxed{\pi_2 = \frac{3}{4}}$$

$$\pi_3 = P(\{1,3\}) + P(\{2,3\})$$
$$= \frac{1}{4} + \frac{1}{4}$$
$$\boxed{\pi_3 = \frac{1}{2}}$$

1-2

3- Calculer π_{12} et π_{23}

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

$$\pi_{23} = \frac{1}{4} = P(\{2,3\})$$

4- π-estimateur de $\bar{Y}$

$$\hat{\bar{Y}} = \frac{1}{N} \sum_{i \in s} \frac{y_i}{\pi_i}$$

a- si l'échantillon $\{1,2\}$ est tiré

$$\hat{\bar{Y}}_1 = \frac{1}{N} \left(\frac{Y_1}{\pi_1} + \frac{Y_2}{\pi_2} \right)$$

$$= \frac{1}{3} \left(\frac{Y_1}{\frac{3}{4}} + \frac{Y_2}{\frac{3}{4}} \right)$$

$$\boxed{\hat{\bar{Y}}_1 = \frac{4}{9} (Y_1 + Y_2)}$$

b- si l'échantillon $\{1,3\}$ est tiré

$$\hat{\bar{Y}}_2 = \frac{1}{N} \left(\frac{Y_1}{\pi_1} + \frac{Y_3}{\pi_3} \right)$$

$$= \frac{1}{3} \left(\frac{Y_1}{\frac{3}{4}} + \frac{Y_3}{\frac{1}{2}} \right)$$

$$= \frac{1}{3} \left(\frac{4}{3} Y_1 + 2 Y_3 \right)$$

$$\boxed{\hat{\bar{Y}}_2 = \frac{1}{9} (4Y_1 + 6Y_3)}$$

e_ si l'échantillon $\{2, 3\}$ est tiré $\overline{H-3}$

$$\hat{Y}_3 = \frac{1}{N}\left(\frac{Y_2}{\pi_2} + \frac{Y_3}{\pi_3}\right)$$

$$= \frac{1}{3}\left(\frac{Y_2}{\frac{3}{4}} + \frac{Y_3}{\frac{1}{2}}\right)$$

$$= \frac{1}{3}\left(\frac{4}{3} Y_2 + 2 Y_3\right)$$

$$\boxed{\hat{y} = \frac{1}{9}\left(4Y_2 + 6Y_3\right)}$$

5_ Vérifions que le π_estimateur est un estimateur sans biais

$$E(\bar{Y}) = \frac{1}{2}\left(\frac{4}{9}\left(Y_1+Y_2\right)\right) + \frac{1}{4}\left(\frac{1}{9}\left(4Y_1 + 6Y_3\right)\right) + \frac{1}{4}\left(\frac{1}{9}\left(4Y_2+6Y_3\right)\right)$$

$$= \frac{2}{9}\left(Y_1 + Y_2\right) + \frac{1}{9} Y_1 + \frac{1}{6} Y_3 + \frac{1}{9} Y_2 + \frac{1}{6} Y_3$$

$$= \frac{1}{3} Y_1 + \frac{1}{3} Y_2 + \frac{1}{3} Y_3$$

$$= \frac{1}{3}\left(Y_1 + Y_2 + Y_3\right)$$

$$= \frac{1}{3} \sum_{i=1}^{3} Y_i$$

$$E(\bar{Y}) = \overline{Y}$$

<u>Conclusion</u> : le π_estimateur est alors sans biais

6- Probabilités de P et π $\boxed{1-4}$

Si on a les probabilités d'échantillon P et les probabilités d'inclusion π pour un sondage aléatoire simple à probabilités égales sans remise, on aurait :

- Toutes les probabilités d'échantillons auraient la même valeur :

$$P(\{1,2\}) = P(\{1,3\}) = P(\{2,3\}) = \frac{1}{3}$$

- $\pi_{12} = \pi_{13} = \pi_{23} = \frac{1}{3}$

- $\pi_1 = \pi_2 = \pi_3 = \frac{2}{3}$

Exercice 2 $\boxed{2.1}$

1- Déterminer la taille de l'échantillon

- Plan avec remise

$$IC(0{,}95) = \left[\hat{\bar{Y}} - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{S_y^2}{n}} \, , \, \hat{\bar{Y}} + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{S_y^2}{n}} \right]$$

Soit $\hat{P}_{AR}$ l'estimateur de la proportion pour le plan avec remise

$$IC(0{,}95) = \left[\hat{P}_{AR} - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{P}_{AR}(1-\hat{P}_{AR})}{n-1}} \, , \right.$$

$$\left. \hat{P}_{AR} + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{P}_{AR}(1-\hat{P}_{AR})}{n-1}} \right].$$

$$VAR_p(\hat{P}_{AR}) = \frac{\hat{P}_{AR}(1-\hat{P}_{AR})}{n-1}$$

d'où $2 Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{P}_{AR}(1-\hat{P}_{AR})}{n-1}} \leq 0{,}02$

par suite $Z^2_{1-\frac{\alpha}{2}} \frac{\hat{P}_{AR}(1-\hat{P}_{AR})}{n-1} \leq 0{,}0001$

$$n-1 \geq Z^2_{1-\frac{\alpha}{2}} \frac{\hat{P}_{AR}(1-\hat{P}_{AR})}{0{,}0001}$$

$$n \geq 1 + Z^2_{1-\frac{\alpha}{2}} \frac{\hat{P}_{AR}(1-\hat{P}_{AR})}{0{,}0001}$$

2.2

n $95\% = 100(1-\alpha)\%$

$\Rightarrow \frac{95}{100} = 1-\alpha$

$\Rightarrow \alpha = \frac{95}{100} - 1$

$\Rightarrow \boxed{\alpha = 0{,}05}$

On a $P(|Z| \geq Z_\alpha) = \alpha = 0{,}05$

$Z \sim N(0,1)$ avec $Z_\alpha = 1{,}96$ $(*)$

d'où $n \geq 1 + 1{,}96^2 \times \frac{0{,}3 \times 0{,}7}{0{,}0001}$

$n \geq 8068{,}36$

donc $\boxed{n \simeq 8068}$

- Plan sans remise

$$IC(0{,}95) = \left[\hat{\bar{Y}} - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{N-n}{N} \times \frac{S_y^2}{n}} , \hat{\bar{Y}} + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{N-n}{N} \times \frac{S_y^2}{n}} \right]$$

$$IC(0{,}95) = \left[\hat{P}_{SR} - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{N-n}{N} \frac{\hat{P}_{SR}(1-\hat{P}_{SR})}{n-1}} , \hat{P}_{SR} + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{N-n}{N} \frac{\hat{P}_{SR}(1-\hat{P}_{SR})}{n-1}} \right]$$

$$2 Z_{1-\frac{\alpha}{2}} \sqrt{\frac{N-n}{N} \times \frac{\hat{P}_{SR}(1-\hat{P}_{SR})}{n-1}} \leq 0{,}02$$

$\boxed{2\text{-}3}$

$$Z^2_{1-\frac{\alpha}{2}} \frac{N-n}{N} \times \frac{\hat{p}_{SR}(1-\hat{p}_{SR})}{n-1} \leq 0,0001$$

$$(n-1)\times 0,0001 - Z^2_{1-\frac{\alpha}{2}} \times \frac{N-n}{N} \hat{p}_{SR}\left(1-\hat{p}_{SR}\right) \geq 0,001$$

$$n\left[0,0001 + Z^2_{1-\frac{\alpha}{2}} \times \frac{1}{N} \hat{p}_{SR}\left(1-\hat{p}_{SR}\right)\right] \geq 0,0001$$

$$+ Z^2_{1-\frac{\alpha}{2}} \times \hat{p}_{SR}\left(1-\hat{p}_{SR}\right)$$

$$n \geq \frac{0,0001 + Z^2_{1-\frac{\alpha}{2}} \times \hat{p}_{SR}\left(1-\hat{p}_{SR}\right)}{\left[0,0001 + Z^2_{1-\frac{\alpha}{2}} \times \frac{1}{N} \hat{p}_{SR}\left(1-\hat{p}_{SR}\right)\right]}$$

or d'après (*) : $Z_{1-\frac{\alpha}{2}} = 1,96$

d'où $n \geq \dfrac{0,0001 + 1,96^2 \times 0,30 \times 0,70}{\left[0,0001 + 1,96^2 \times \frac{1}{1500} \times 0,30 \times 0,70\right.}$

$n \geq 1264,95$

$$\boxed{n \simeq 1265}$$

2-/ Que faire si nous ne connaissons pas la proportion ?

$\boxed{2-4}$

Si l'on ne disposait pas d'information sur la proportion d'hommes habituellement touchés par la maladie, pour le cas du plan sans remise alors on utiliserait $p = 0,5$ ce qui correspondrait au cas le plus défavorable car la dépression est plus grande.

3 1

Exercice 3

1.| Donnons une estimation du total des notes

On a $\hat{T}_N = N \bar{y}_\omega$; N : nombre total de collèges

$$= \frac{N}{n} \sum_{i=1}^{n} y_i \, \mathbb{1}_{\{u_i \in \omega\}}$$

$$= \frac{50}{5} \left(12 \times 40 + 8 \times 20 + 10 \times 60 + 12 \times 40 + 11 \times 48 \right)$$

$$\boxed{\hat{T}_N = 22480}$$

2.| Estimons le nombre d'élèves en 6ièmes

$$nbre\_eleves = \frac{N}{n} \sum_{i=1}^{n} Ef_i \, \mathbb{1}_{\{u_i \in \omega\}}$$

Ef_i : effectif de la classe i dans l'échantillon prélevé

d'où $nbre\_élèves = \frac{50}{5} \left(40 + 20 + 60 + 40 + 48 \right)$

$$\boxed{nbre\_élèves = 2080}$$

3°/ Estimation $\boxed{3\_2}$

$note\_moy_{w} = \frac{1}{2000} \times \hat{T}_{N}$, $note\_moy_{w}$: note moyenne estimée

$= \frac{22480}{2000}$

$\boxed{note\_moy_{w} = 11,24}$ (*)

$note\_moy\_ech = \frac{1}{n} \sum_{i=1}^{n} N_{w_i} \mathbb{1}_{\{w_i \in w\}}$

note_moy_ech : note moyenne de l'échantillon prélevée

$note\_moy\_ech = \frac{1}{5} (12 + 8 + 10 + 12 + 11)$

$\boxed{note\_moy\_ech = 10,6}$

Conclusion : L'estimation de la note moyenne (note_moy) est supérieure à la moyenne observée par l'échantillon (note_moy_ech). Donc on peut dire que la note

3-3

4- Calcule de la variance de

$$Var(\hat{T}_N) = N^2 (1-f) \frac{S^2_w}{n}$$

$$f = \frac{n}{N}$$

$$Var(\hat{T}_N) = N^2 \left(1 - \frac{n}{N}\right) \frac{S^2_w}{n}$$

$$S^2_w = \frac{1}{n-1} \sum_{i=1}^{N} (y_i - \bar{y}_0)^2 \mathbb{1}_{\{w_i \in w\}}$$

$$= \frac{1}{4} \left[(12 - 10,6)^2 + (8 - 10,6)^2 + (10 - 10,6)^2 + (12 - 10,6)^2 + (11 - 12,6)^2 \right]$$

$$\boxed{S^2_w = 3,4} \quad (*)$$

$$Var(\hat{T}_N) = 50^2 \left(1 - \frac{1}{10}\right) \times \frac{3,4}{5} = 1530$$

- Déduire la variance de la moyenne

$$Var(\hat{T}_N) = N^2 \left(1 - \frac{n}{N}\right) \frac{S_w^2}{n}$$

$$\left(1 - \frac{n}{N}\right) \frac{S^2_w}{n} = \frac{Var(\hat{T}_N)}{N^2}$$

$$Var(\hat{A}_{moy\,w}) = \frac{Var(\hat{T}_N)}{N^2}$$

3-4

$$Var(\hat{N}_{moy_{10}}) = \frac{1530}{50^2}$$

$$\boxed{Var(\hat{N}_{moy_{10}}) = 0{,}612} \quad (**)$$

- Intervalle de confiance

$100(1-\alpha)\%$, $\alpha \in]0,1[$

$$IC(95\%) = [note\_moy_{10} - Z_\alpha \sqrt{Var(note\_moy_{10})} \; ; \; note\_moy_{10} + Z_\alpha \sqrt{Var(note\_moy_{10})}]$$

Prenons $\alpha = 0{,}05$

$95\% = 100(1-\alpha)\% \Rightarrow \alpha = 0{,}05$

et $Z_\alpha = 1{,}96$

$$IC(95\%) = [11{,}24 - 1{,}96\sqrt{0{,}612} \; ; \; 11{,}24 + 1{,}96\sqrt{0{,}612}]$$

$$IC(95\%) = [9{,}70 \; ; \; 12{,}77]$$

$$\boxed{IC(95\%) = [9{,}7 \; ; \; 12{,}8]}$$

5_ Comparaison avec un sondage PESR $\boxed{3\_5}$

$$\hat{T}_N = N\,\bar{y}_w$$

$$= \frac{N}{n}\sum_{i\in s} y_i$$

$$= \frac{50}{5}\left(12\times 40 + 8\times 20 + 10\times 60 + 12\times 4 + 11\times 48\right)$$

$\boxed{\hat{T}_N = 22480}$ (d'après 1.1)

D'après 1.1 : le nbre d'élèves en 6ième ;

nbre_élèves = 2080

note_moy$_w$ = 11,24 (d'après 3.1(*))

D'après (*) : $s^2_w = 3,4$

$$\text{Var}\left(\hat{T}_N\right) = 50^2\left(\frac{50-1}{50}\right)\frac{3,4}{5} = 1666$$

et Var(note_moy$_w$) = 0,6124 (d'après 4(**))

d'où $IC(0,95) = \left[11,24 - 1,96\sqrt{0,612}\ ;\ 11,24 + \sqrt{0,612}\right]$

$\boxed{IC(0,95) = [9,71\ ;\ 12,04]}$

<u>Conclusion</u> : On obtient sensiblement le même intervalle de confiance

$\boxed{3\_6}$

a) Estimation ensembliste

seuil = 5% $\Rightarrow$ on a 95% de niveau de confiance

$100\,(1-\alpha)\%$

$\Rightarrow \alpha = 0{,}05$

$P(|Z| \geq Z_\alpha) = \alpha$

$Z \sim N(0,1)$ alors $Z_\alpha = 1{,}96$

$IC = [8{,}5 - 1{,}96 \times S \;;\; 8{,}5 + 1{,}96]$

$= [6{,}932 \;;\; 10{,}068]$

$$\boxed{IC \simeq [6{,}9 \;;\; 10{,}1]}$$

b) Nombre minimal d'observations

On a $100\,(1-\alpha)\%$, $\alpha \in]0,1[$

$d_\omega = Z_\alpha\, S\,(p_\omega)$

$= Z_\alpha \sqrt{(1-f)\,\dfrac{p_\omega(1-p_\omega)}{n-1}}$

* Taille de n pour que $d_\omega \leq 0{,}02$

$$n \geq \frac{N Z_\alpha^2\, p_\omega (1-p_\omega)}{N d_\omega^2 + Z_\alpha^2\, p_\omega (1-p_\omega)}$$

$$n \geq \frac{25 \times (1{,}96)^2 \times 8{,}5(1-8{,}5)}{25 \times 0{,}2^2 + (1{,}96)^2 \times 8{,}5(1-8{,}5)}$$

$\boxed{3\text{-}7}$

$$\boxed{n \geq 25}$$

* Pour seuil = 1 %

$Z_\alpha = 2{,}576 \quad ; \alpha = 0{,}01$

$$n \geq \frac{25\,(2{,}576)^2 \times 8{,}5 \times (1-8{,}5)}{25\,(0{,}2)^2 + (2{,}576)^2 \times 8{,}5(1-8{,}5)}$$

$$\boxed{n \geq 25}$$

<u>Conclusion</u> :

Le nombre minimal d'observation auxquelles on devrait procéder pour connaitre la consommation moyenne à plus ~~ou~~ moins 2 dL près (au plus) aux seuils de 5 % et de 1 % devrait être de 25 observations.

$\boxed{4-1}$

Exercice 4

1_ Nombre maximum d'erreurs

erreur = $n \times p$

* Pour $n = 200$: erreur = $200 \times 0{,}05 = 10$
* Pour $n = 400$: erreur = $400 \times 0{,}05 = 20$
* Pour $n = 600$: erreur = $600 \times 0{,}05 = 30$
* Pour $n = 1000$: erreur = $1000 \times 0{,}05 = 50$

2_ Le nombre d'enregistrement en tolérant 4 nouvelles erreurs

$n \times p$ = erreurs

$n \times 0{,}05 = 4$

$$n = \frac{4}{0{,}05}$$

$$\boxed{n = 80}$$

Exercice 5

1°) Estimation du chiffre d'affaire moyen avec un intervalle de confiance

$$Var(\hat{\mu}) = \frac{1}{N^2} \sum_{h=1}^{4} \frac{N_h (N_h - n_h)}{n_h} \times S_h^2$$

$$= \frac{1}{1060^2} \left[\frac{500(500-130)}{130} \times 1,5^2 + \frac{300(300-80)}{80} \times 4^2 \right.$$

$$+ \frac{150(150-60)}{60} \times 8 + \frac{100(100-25)}{25} \times 100$$

$$+ \frac{10(10-5)}{5} \times 2500$$

$$\boxed{Var(\hat{\mu}) = 0,055}$$

Pour $Z_{0,90} = 1,64$, on a : $\hat{\mu} = \frac{1}{N^2} \sum_{h=1}^{4} N_h \times y_h$

$\hat{\mu}$: estimation du chiffre d'affaire moyen

$$N = \sum_{h=1}^{4} n_h \quad 300$$

$$N = 130 + 80 + 60 + 25 + 5$$

$$\boxed{N = 300}$$

$$\hat{\mu} = \frac{1}{300} (130 \times 5 + 12 \times 80 + 30 \times 60 + 150 \times 25 + 600 \times 5)$$

$$\boxed{\hat{\mu} = 29,81}$$

$$IC_{(0,90)} = \left[\hat{u} - Z_{0,90}\sqrt{V(\hat{u})}\ ,\ \hat{u} + Z_{0,90}\sqrt{V(\hat{u})}\right]$$

5-2

$$= \left[29,81 - 1,64\sqrt{0,055}\ ,\ 29,81 + 1,64\sqrt{0,055}\right]$$

$$\boxed{IC_{(0,90)} = [29,43\ ;\ 30,19]}$$

2. a/ Effectifs d'échantillon pour une allocation proportionnelle

ou : $n_h = \dfrac{n \times N_H}{N}$, avec $N = 1060$; $n = 300$

d'où : $n1 = 300 \times \dfrac{500}{1060} = 142$

$n2 = 300 \times \dfrac{300}{1060} = 85$

$n3 = 300 \times \dfrac{150}{1060} = 42$

$n4 = 300 \times \dfrac{100}{1060} = 28$

$n5 = 300 \times \dfrac{10}{1060} = 3$

b/ Pour une allocation optimale

ou : $n_h = n \times \dfrac{N_h \times S_h}{\sum N_h \times S_h}$

5-2

d'où : $n_1 = \dfrac{300 \times 500\sqrt{1,5}}{(500\sqrt{1,5} + 300\sqrt{4} + 150\sqrt{8} + 100\sqrt{100} + 10\sqrt{2500}} = 59$

$n_2 = 57$

$n_3 = 40$

$n_4 = 96$

$n_5 = 48$

On doit interroger 48 personnes dans la strate 5 ($N_5 = 10$) alors qu'elle n'en contient que 10. c'est impossible, on choisit alors d'interroger les 10 personnes de la strate 5 ($n = 10$) et on recalcule les tailles d'échantillons des autres strates

On a $n = 300 - 10 = 290$

$n_1 = 290 \times \dfrac{500\sqrt{1,5}}{500\sqrt{1,5} + 300\sqrt{4} + 150\sqrt{8} + 100\sqrt{100}}$

$n_1 = 67,35$

$n_2 = 65,99$

$n_3 = 46,66$

$n_4 = 109,98$

|5-4|

On doit interroger $n4 = 110$ individus dans la strate 4 qui en contient 100.

On les interroge donc toutes ($n4 = 100$) et on recalcule $n1$, $n2$ et $n3$

$n = 290 - 100 = 190$

$$n1 = \frac{190 \times 500\sqrt{1,5}}{500\sqrt{1,5} + 300\sqrt{4} + 150\sqrt{8} + 100\sqrt{100}}$$

$\boxed{n1 = 71}$

$n2 = 70$

$n3 = 49$

$n4 = 100$

$n5 = 10$

3°| Allocation proportionnelle

$$Var(\hat{u}) = \frac{1}{N^2} \sum N_h \frac{(N_h - n_h)}{n_h} \times S^2_h$$

$$Var(\hat{u}) = \frac{1}{1060^2} \Big[500 \times 1,5 \frac{(500 - 142)}{142} + 300 \times 4 \times \frac{(300-85)}{85}$$

$$+ 150 \times 8 \times \frac{(150 - 42)}{42} + 100 \times 100 \frac{(100 - 28)}{28}$$

5-5

$$+ 10 \times 2500 \frac{(10-3)}{3} \Big]$$

$$\boxed{Var(\hat{u}) = 0{,}0819}$$

Allocation optimale

$$Var(\hat{u}) = \frac{1}{N^2} \sum N_h \frac{(N_h - n_h)}{n_h} \times S_h^2$$

$$= \frac{1}{1060^2} \Big[500 \times 1{,}5 \frac{(500-71)}{71} + 300 \times 4 \times \frac{(300-70)}{70}$$

$$+ 150 \times 8 \times \frac{(150-49)}{49}$$

$$+ 100 \times 100 \times \frac{(100-100)}{100}$$

$$+ 10 \times 2500 \times \frac{(10-10)}{10} \Big]$$

$$\boxed{Var(\hat{u}) = 0{,}00974}$$