Université de Thiès

unité de formation et de recherche

en

sciences économiques et sociales

département management des organisations

Master 1 sciences des données et applications

Projet Visualisation interactive

Professeurs: Mme. Sall

présenté par:

Nogaye Fall

Ndèye Penda Diagne

Khoudia Mbodji

année académique:

2019 2020

Exercice 1:

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

P({1,2})=1/2 P({1,3})=1/4 P({2,3})=1/4

1)Ce n'est pas un sondage aléatoire simple car tous les échantillons de taille n = 2 ne sont pas tirés a des probabilités égales

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

$\pi 1 = 1/2+1/4 = 3/4$

$\pi 2 = 1/2+1/4 = 3/4$

$\pi 3 = 1/4+1/4 = 1/2$

3)Calculons $\pi 12$ et $\pi 23$ les probabilités d'inclusions d'ordre 2

$\pi 12 = 1/2$

$\pi 23 = 1/4$.

4)Quel est le π-estimateur de ¯ 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)le π-estimateur est un estimateur sans biais

6)Écrivons ce que seraient les probabilités d'échantillon P et les probabilités d'inclusions π pour un sondage aléatoire simple a probabilités égales sans remise

Si on a un sondage aléatoire simple a probabilités égales sans remise, on a alors pour chaque échantillon la même probabilité d'échantillon et d'inclusion, c'est-`a-dire :

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

Et

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

Exercice 2:

1) la taille de l'échantillon

Intervalle de confiance = [f - $\frac{1}{\sqrt{n}}$; f +$\frac{1}{\sqrt{n}}$]

Amplitude de l'intervalle de confiance: $\frac{2}{\sqrt{n}}$

$\frac{2}{\sqrt{n}} \leq 0.02$

$\frac{\sqrt{n}}{2} \geq \frac{1}{0.02}$

$\sqrt{n} \geq \frac{2}{0.02}$

$\sqrt{n} \geq 100$

n $\geq 100^2$

n ≥ 10000

La taille de l'échantillon est de 10000.

2) La proportion d'hommes

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

$P = \frac{1500}{10000} = 0.15\%$

Exercice 3:

1) Donnez une estimation du total des notes au test sur le districte

$\hat{T}_{i\text{notes}} = n_i * y_i$

$T_1=40*2=480$

$T_2=20*8=160$

$T_3=10*60=600$

$T_4=40*12=480$

$T_5=48*11=528$

$\hat{T}_{moy} = \frac{N}{n}\sum Ti = \frac{50}{n}(480+160+600+480+528)=22480$

$\hat{T}_{moy}=22480$

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

$\hat{N}_{\text{élèves}} = \frac{N}{n}\sum yi = \frac{50}{n}(40+20+60+40+48) = 2080$

$\hat{N}_{\text{élèves}}=2080$

3) Donner une estimation de la moyenne pour n=2000

$\overline{Y} = \frac{1}{2000}\hat{T} = \frac{22480}{2000} = 11.24$

$$\overline{Y} = 11.24$$

Comparons le avec

$\hat{y} = \frac{10}{50}(12+8+10+12+11)=10.6$

=10.6

4) Calculons la variance de l'estimateur total

Calculons une= estimation de cette variance

$\hat{V}(\hat{T}) = M^2(1-f)\frac{S_1^2}{m} + \frac{M}{m}\sum N_i^2\,(1-f_{2i})\frac{S_{1\,i}^2}{n_i}$

$\hat{V}(\hat{T}) = M^2(1-f)\frac{S_1^2}{m} + \frac{M}{m}\sum V_i$

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

$S_1^2 = \frac{1}{4}[(480-449.6)^2+(160-449.6)^2+(600-449.6)^2+(480-449.6)^2+(528-449.6)^2]$

$S_1^2=28620.8$

$M^2(1-f_i)\frac{S_1^2}{m} = 50^2(1-0.1)*\frac{28620.8}{5}=1279360$

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

$V_1=40^2(1-0.25)*\frac{1.5}{10}=180$

$V_2=20^2(1-0.5)*\frac{1.2}{10}=24$

$V_3=60^2(1-0.15)*\frac{1.6}{10}=480$

$V_4=40^2(1-0.25)*\frac{1.3}{10}=156$

$V_5=48^2(1-0.2)*\frac{2}{10}=364.8$

$\frac{M}{m}\sum V_i=\frac{50}{5}(180+24+480+156+364.8)=12048$

$\hat{V}(\hat{T})=1279360+12048=12891408$

En déduire la variance de la moyenne

$\text{Var}\ (\hat{y})=\frac{1}{N^2}\hat{V}(\hat{T})=\frac{12891408}{2000^2}=3.22$

$\text{Var}\ (\hat{y})=3.221$

Intervalle de confiance à 95% de la moyenne

$H\ /\ \sqrt{Var(\hat{y})}=z_{1-\alpha/2}$ => $H=z_{1-\alpha/2}*\sqrt{Var(\hat{y})}$

$z_{1-\alpha/2}=1.96$ => $H=1.96*\sqrt{3.22}=3.5$

$\hat{y}-z_{1-\alpha/2}*\sqrt{Var(\hat{y})}=11.24-3.5=7.74$

$\hat{y}+z_{1-\alpha/2}*\sqrt{Var(\hat{y})}=11.24+3.5=14.74$

Intervalle de confiance a 95% : $[7.74,14.74]$

5) Comparons avec un SAS à probabilités égales
f=50/2000=0.25
$\text{Variance}=\text{Var}_{inter}+\text{Var}_{intra}$
$\text{Var}_{inter}=\frac{1}{n}\sum_{k=1}^{H}N_h\,(\mu_k+\mu)^2$
$\text{Var}_{inter}=\frac{1}{50}(10*12^2+10*8^2+10*10^{2+}10*12^2+10*11^2)-10.6^2$

$\text{Var}_{inter}=2.24$

$\text{Var}_{intra}=\frac{1}{n}\sum N_h\,\sigma_h^2$

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

$\text{Var}_{intra}=\frac{1}{n}\sum N_h\,\frac{N-1}{N}S^2$

$\text{Var}_{intra}=\frac{1}{50}*10*\frac{9}{10}(1.5+1.2+1.6+1.3+2)$

$\text{Var}_{intra}=1.368$

$\text{Variance}=2.24+1.368=3.608$

$S^2=\frac{50}{49}*3.608=3.68$

$V\hat{a}r(\hat{y})$=(1-f)$\frac{S^2}{n}$=(1-0.25)$\frac{3.68}{50}$

$V\hat{a}r(\hat{y})$=0.07

H=1.96*$\sqrt{0.07}$=0.52

$\hat{y}$=[10.16, 11.03]

La précision est plus nette avec la SAS

a-Donnons une estimation ensembliste de la consommation moyenne de ce type de vehicule au seuil de 5%

1.96* $\frac{0.8}{\sqrt{25}}$=0.314

$\hat{y}$=[8.186, 8.8136]

b- quel devrait etre le le nombre minimal d'observation au quel on devrait proceder pour conaitre la consommation moyenne a plus ou moin 2dl pres au seuil de 5% et de 1%

- 5% :
  y+ $\frac{\sigma}{\sqrt{n}}$
  $\frac{\sigma}{\sqrt{n}}$*1.96=0.2
  $\frac{\sigma}{\sqrt{n}}$=$\frac{0.2}{1.96}$
  $\sqrt{n}$=8
  $$n = 64$$
- 1% :
  y+ $\frac{\sigma}{\sqrt{n}}$
  $\frac{\sigma}{\sqrt{n}}$*1.96=0.2
  $\frac{\sigma}{\sqrt{n}}$=$\frac{0.2}{1.96}$
  $\sqrt{n}$=8
  $$n = 64$$

EXERCICE 4:

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

NE=0,05*200= 10 erreurs

MEME QUESTIONS AVEC n=400, n=600 et n=1000

*POUR n=400

NE=0,05*400= 20 erreurs

*POUR n=600

NE=0,05*600= 30 erreurs

*POUR n=1000

NE= 0,05*1000= 50 erreurs

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

0 ,05* x = (7+4) => x=(7+4)/0,05=180 enregistrements

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

Exercice 5 :

Un intervalle de confiance de niveau 0.90 est donné par

V(µ̂) = 0.055

µ̂ = 29.81

Intervalle de Confiance : 0.90 = [29.43; 30.19].

2)effectif des échantillons dans chaque strate

a-pour une allocation proportionnelle :

$\frac{n_h}{n}= \frac{N_h}{N} => n_h = n*\frac{N_h}{N}$

$n_1=300*\frac{500}{1060}=141.50=142$

$n_2=300*\frac{300}{1060}=84.90=85$

n_3=300*$\frac{150}{1060}$=42.45=42

n_4=300*$\frac{10}{1060}$=2.83=3

b-pour une allocation optimale avec $n_h \leq N_h$

n_h= $\frac{N_h * S_h}{\sum_{k=1}^{H} N_h S_h}$

Calculons $\sum_{k=1}^{H} N_h S_h$:

$\sum_{k=1}^{H} N_h S_h$ = ($\sqrt{1.5}$*500) + ($\sqrt{4}$*300) + (150 *$\sqrt{8}$) + ($\sqrt{100}$*100) + ($\sqrt{10}$*2500)=3136.64

n_1=300*$\frac{500*\sqrt{1.5}}{3136.64}$=58.57=59

n_2=300*$\frac{300*\sqrt{4}}{3136.64}$=57.38=57

n_3=300*$\frac{150*\sqrt{8}}{3136.64}$=40.57=40

n_4=300*$\frac{100*\sqrt{100}}{3136.64}$=95.64=96

n_5=300*$\frac{10*\sqrt{2500}}{3136.64}$=47.64=48

on constate que $n_5 > N_5$ donc on choisit les 10 personnes dans la strate 5

n devient 300-10=290

n_1=290*$\frac{500*\sqrt{1.5}}{2636.64}$=67.35=67

n_2=290*$\frac{300*\sqrt{4}}{2636.64}$=65.99=66

n_3=290*$\frac{150*\sqrt{8}}{2636.64}$=46.66=47

n_4=290*$\frac{100*\sqrt{10}}{2636.64}$=109.9=110

on que n_4 > N_4 donc les 100 personnes dans l'echantillon 4

n devient : 290-100=190

n_1=190*$\frac{500*\sqrt{1.5}}{1636.64}$=71.09=71

n_2=190*$\frac{300*\sqrt{4}}{1636.64}$=70.65=70

n_3=190*$\frac{150*\sqrt{8}}{1636.64}$=49.26=49

donc n_1=71, n_2=70, n_3=49, n_4=100, n_1=10

3)

$$V(\hat{u}) = \frac{1}{n}\left(1 - \frac{n}{N}\right)\frac{1}{N}\sum_{k=1}^{H} N_h {S^2}_h$$

-Pour l’allocation proportionnelle on obtient

V(µ̂) = 0.0819

-Pour l’allocation optimale on obtient

V(µ̂) = 0.00974