

Devoir 2
 STT2000 : Échantillonnage
 À remettre le 7 décembre 2009

1. (30 points) Une agence publicitaire s'interroge sur l'effet d'une nouvelle campagne publicitaire régionale sur le total des ventes d'un produit particulier. Un EASSR de taille $n = 20$ magasins est tiré dans la population composée $N = 452$ magasins de la région qui vendent le produit en question. Les ventes trimestrielles sont obtenues pour le trimestre actuel mais également pour le trimestre précédent la campagne publicitaire. On cherche à estimer t_y , le total des ventes pour le trimestre actuel. On supposera que le total des ventes au trimestre précédent, t_x , est connu et est égal à 216256.

Magasin	Ventes (en \$) au trimestre précédent	Ventes (en \$) au trimestre actuel	Magasin	Ventes (en \$) au trimestre précédent	Ventes (en \$) au trimestre actuel
1	208	239	11	599	626
2	400	428	12	510	538
3	440	472	13	828	888
4	259	276	14	473	510
5	351	363	15	924	998
6	880	942	16	110	171
7	273	294	17	829	889
8	487	514	18	257	265
9	183	195	19	388	419
10	863	897	20	244	257

De plus on obtient les résultats suivant:

Analyse de régression: Ventes actuelles vs. ventes pré-campagne

The regression equation is
 Present sales = 10.1 + 1.05 Precampaign sales

Predictor	Coef	SE Coef	T	P
Constant	10.105	7.083	1.43	0.171
Precampaign sales	1.04975	0.01314	79.87	0.000

S = 14.9321 R-Sq = 99.7% R-Sq(adj) = 99.7%

- a) (4 points) i) L'estimateur par le ratio est-il approprié dans ce cas? Pourquoi?
 (3 points) ii) Estimer le total des ventes pour le trimestre actuel en utilisant l'estimateur par le ratio
 (3 points) iii) Construire un IC de niveau 95% pour le total des ventes pour le trimestre actuel
- b) (4 points) i) L'estimateur par la régression est-il approprié dans ce cas? Pourquoi?
 (3 points) ii) Estimer le total des ventes pour le trimestre actuel en utilisant l'estimateur par la régression.
 (3 points) iii) Construire un IC de niveau 95% pour le total des ventes pour le trimestre actuel.
- c) (4 points) i) L'estimateur par la différence est-il approprié dans ce cas? Pourquoi?
 (3 points) ii) Estimer le total des ventes pour le trimestre actuel en utilisant l'estimateur par la différence.
 (3 points) iii) Construire un IC de niveau 95% pour le total des ventes pour le trimestre actuel.

Remarque : pour a)-iii), b)-iii) et c)-iii), vous devez utiliser *proc surveymeans*. Vous devez fournir le output de sas ainsi que votre code en annexe.

2. (20 points) Supposons que les valeurs de deux variables y et x sont observées pour les n unités tirées d'une population de taille N selon l'EASSR. On cherche à estimer le ratio $B = \frac{\bar{y}_U}{\bar{x}_U}$. Une source administrative nous permet de connaître la vraie valeur de la variable x dans la population, \bar{x}_U . Quelle stratégie d'estimation utiliseriez-vous? (a) utiliser **toujours** l'estimateur $\frac{\bar{y}_s}{\bar{x}_U}$. (b) Utiliser **parfois**

l'estimateur $\frac{\bar{y}_s}{\bar{x}_U}$. (c) utiliser **toujours** l'estimateur $\frac{\bar{y}_s}{\bar{x}_s}$. (d) utiliser **parfois** l'estimateur $\frac{\bar{y}_s}{\bar{x}_s}$. *Attention !*

Faites une réponse complète. Considérez le biais, la variance, l'EQM. Démontrez vos résultats et discutez ce que vous obtenez.

3. (15 points) Supposons que l'on cherche à estimer le ratio de deux ratios

$$B = B_1 / B_2$$

$$\text{où } B_1 = \frac{\bar{y}_{1U}}{\bar{x}_{1U}} \text{ et } B_2 = \frac{\bar{y}_{2U}}{\bar{x}_{2U}} .$$

- a) (3 points) Exprimer B comme une fonction de totaux.
 b) (8 points) Afin d'estimer B , On tire un EASSR, s , de taille n . Un estimateur de B est défini selon

$$\hat{B} = \hat{B}_1 / \hat{B}_2$$

Utiliser une approximation par séries de Taylor du premier ordre pour montrer que

$$\hat{B} - B \approx \frac{N}{n} \sum_{i \in s} e_i, \text{ et donner l'expression pour } e_i .$$

- c) (4 points) Donner une expression de la variance approximative de \hat{B} .

4. (15 points) Deux dentistes font une enquête sur l'état des dents des 200 enfants d'un village. Le premier dentiste sélectionne selon un plan de sondage aléatoire simple sans remise 20 enfants. Les résultats sont exhibés dans le tableau qui suit :

Nombre de dents cariées	0	1	2	3	4	5	6	7	8
Nombre d'enfants	8	4	2	2	1	2	0	0	1

Le second dentiste examine les 200 enfants, mais dans le seul but de déterminer ceux qui n'ont aucune carie. Il constate que 50 enfants sont dans ce cas.

- a) (5 points) Estimer le nombre moyen de dents cariées par enfant dans le village en utilisant seulement les résultats du premier dentiste. Estimer la variance de votre estimateur.
 b) (10 points) Estimer le nombre moyen de dents cariées par enfant dans le village en utilisant les résultats des deux dentistes. Comment s'appelle l'estimateur que vous avez utilisé? Estimer la variance de votre estimateur.

5. (27 points) Une population est répartie en deux sous-populations de tailles M_1 et M_2 , de moyennes \bar{y}_{u_1} et \bar{y}_{u_2} et de variances $S_{u_1}^2$ et $S_{u_2}^2$, respectivement. Au hasard, on tire l'une des deux populations, et dans cette population, on tire un échantillon aléatoire simple de taille n . Soit \bar{y} la moyenne et s_y l'écart-type de l'échantillon.
- (6 points) Soit X une variable aléatoire qui prend la valeur i si la sous-population choisie est i ($i = 1, 2$). Déterminer (i) $E(\bar{y})$, (ii) $V(\bar{y} | X=1)$; (iii) $E[V(\bar{y} | X)]$
 - (6 points) \bar{y} est-il un estimateur sans biais pour la moyenne de la population? Déterminez $V(\bar{y})$ et l'erreur quadratique moyenne de \bar{y} .
 - (15 points) Considérez l'estimateur $2M\bar{y}/(M_1+M_2)$, où M est la taille de la population choisie.
 - (3 points) Montrez que cet estimateur est sans biais pour la moyenne de la population.
 - (4 points) Déterminez la variance de cet estimateur.
 - (8 points) Supposez cette fois-ci que les sous-populations sont choisies avec probabilités $M_i/(M_1+M_2)$, $i = 1, 2$. Montrez que \bar{y} est sans biais, et déterminez sa variance.

6. (15 points) Dans le cadre d'un projet de rénovation de lits d'hôpitaux, on tire un EASSR de 4 hôpitaux d'une population de 700 hôpitaux; et dans chacun de ces hôpitaux, on tire un EASSR de chambres. Dans chaque chambre on compte le nombre de lits vétustes. On sait que dans la population il y a en tout 86 200 chambres. Les données sont présentées dans le tableau de la page suivante.
- (5 points) Estimer par l'estimateur sans biais le nombre total de lits vétustes dans la population.
 - (5 points) Estimer l'écart-type de l'estimateur en a).
 - (5 points) Estimer par le ratio le nombre total de lits vétustes dans la population sachant qu'il y a en tout 86 200 chambres dans les hôpitaux de la population et estimez l'écart-type de votre estimateur..

Hôpital		Nombre de chambres dans l'hôpital M_i	Nombre de chambres tirées dans l'hôpital m_i	Moyenne échantillonnale \bar{y}_i	Variance échantillonnale s_i^2
1	3 2 2 1 1 0 0 0 0	300	9	1	1,25
2	2 2 1 1 1 1 0 0 0 0	350	10	0,8	0,622222
3	4 4 4 3 3 3 3 2	220	8	3,25	0,5
4	3 2 1 1 1 0 0 0 0 0	330	10	0,8	1,066667

7. (15 points) On tire un EASSR de taille n d'une population de N compagnies afin d'estimer le salaire moyen des employés. Dans chacune des n compagnies de l'échantillon, on tire un EASSR d'employés (m_i employés tirés parmi les M_i de la compagnie i). On propose l'estimateur $\hat{y}_U = \frac{1}{n} \sum_{i \in s} \bar{y}_i$, où s est l'ensemble des compagnies de l'échantillon et \bar{y}_i est la moyenne de l'échantillon tiré dans la compagnie i .
- (6 points) L'estimateur \hat{y}_U est-il sans biais? Si non, donnez l'expression du biais et donnez une condition sous laquelle il est sans biais. [Prenez soin de bien justifier chaque étape de votre développement]
 - (6 points) Déterminer $V(\hat{y}_U)$ [Prenez soin de bien justifier chaque étape de votre développement]
 - (3 points) Déterminez l'erreur quadratique moyenne de \hat{y}_U .
8. a) (4 points) Démontrer la Proposition 5.7.
 b) (6 points) Démontrer la Proposition 5.8.

9. (13 points) Exercice 2 page 169 (Lohr)

10. (15 points) Démontrer la Proposition 6.3. Exhiber tous les détails!

11. (20 points) Exercice 6 page 171 (Lohr)

David Haziza