**Devoir 2**
STT2000 : Échantillonnage
À remettre le 9 décembre 2010

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) Considérons une population sur laquelle sont définies deux variables, $x$ et $y$, prenant les valeurs $(x_1, y_1), \ldots, (x_N, y_N)$. Soit $B = \dfrac{\overline{y}_U}{\overline{x}_U}$ un paramètre à estimer. On sait que $\hat{B} = \dfrac{\overline{y}_s}{\overline{x}_s}$ est biaisé dans un tirage aléatoire simple sans remise. Il existe un mode de tirage, la méthode de Lahiri, pour lequel $\hat{B}$ est sans biais pour $B$. Elle consiste à tirer l'échantillon $s$ avec probabilités proportionnelles aux $\overline{x}_s$ : $P(s) = k\overline{x}_s$. Dans ce problème, je vous conseille d'utiliser les définitions originales de l'espérance et de la variance d'un estimateur (voir chapitre 2).

a) (4 points) Montrez que $k = 1/[\binom{N}{n}\ \overline{x}_U]$

b) (5 points) Montrez que $\hat{B}$ est sans biais : $E_L(\hat{B}) = B$, où $E_L(.)$ dénote l'espérance sous un tirage de Lahiri.

c) (6 points) Montrez que $V_L(\hat{B}) = E_{EASSR}\left(\overline{y}^2 / \overline{x}_s\overline{x}_U\right) - B^2$, où $V_L(.)$ dénote la variance sous un tirage de Lahiri et $E_{EASSR}$ dénote l'espérance sous un tirage aléatoire simple sans remise [alors que $V_{EASSR}(\hat{B}) = E_{EASSR}\left(\overline{y}^2 / \overline{x}_s^2\right) - B^2$].

d) (5 points) Montrez qu'une façon de réaliser un tirage de Lahiri est la suivante : on tire une première unité avec probabilités proportionnelles aux $x_i$ ; puis un tirage aléatoire simple de $n$ -1 unités parmi les $N$ - 1 qui restent.

3. (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.

4. (4+ 5+ 5+ 7+ 4 points) Exercice 11 page 157 dans Lohr.

5. (5+ 3+5+7 points) Exercice 35 page 162 dans Lohr (sauf (e)).

6. (25 points) Une population est répartie en deux sous-populations de tailles $M_1$ et $M_2$, de moyennes $\overline{y}_{U_1}$ et $\overline{y}_{U_2}$ et de variances $S_1^2$ et $S_2^2$, respectivement. Au hasard, on tire l'une des deux populations, et dans cette population, on tire un échantillon aléatoire simple de taille $n$. Soit $\overline{y}$ la moyenne et $s_y$ l'écart-type de l'échantillon.

a) (5 points) Soit $X$ une variable aléatoire qui prend la valeur $i$ si la sous-population choisie est $i$ ($i$ = 1, 2). Déterminer (*i*) $E(\overline{y})$, (*ii*) $V(\overline{y} \mid X=1)$; (*iii*) $E[V(\overline{y} \mid X)]$

b) (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 $\bar{y}$.

c) (15 points) Considérez l'estimateur $2M\bar{y}/(M_1+M_2)$, où $M$ est la taille de la population choisie.

(i) (3 points) Montrez que cet estimateur est sans biais pour la moyenne de la population.

(ii) (4 points) Déterminez la variance de cet estimateur.

d) (7 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.

7. (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.

a) (5 points) Estimer par l'estimateur sans biais le nombre total de lits vétustes dans la population.

b) (5 points) Estimer l'écart-type de l'estimateur en a).

c) (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 échantil-lonnale $\bar{y}_i$* | *Variance échantil-lonnale $s_i^2$* |
|---|---|---|---|---|---|
| 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 |

8. (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 $\hat{\bar{y}}_U = \frac{1}{n}\sum_{i\in s}\bar{y}_i$, où $s$ est l'ensemble des compagnies de l'échantillon et $\bar{y}_i$ est la moyenne de l'échantillon tiré dans la compagnie $i$.

a) (6 points) L'estimateur $\hat{\bar{y}}_U$ 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]

b) (6 points) Déterminer V( $\hat{\bar{y}}_U$ ) [Prenez soin de bien justifier chaque étape de votre développement]

c) (3 points) Déterminez l'erreur quadratique moyenne de $\hat{\bar{y}}_U$.

9. a) (*4 points*) Démontrer la Proposition 5.7.
   b) (*6 points*) Démontrer la Proposition 5.8.

10. (4+ 8+ 8 points) Exercice 8 page 209 (Lohr)

11. (*15 points*) Démontrer la Proposition 6.3. Exhiber tous les détails!

12. (6+ 4+6+6+3 points) Exercice 36 page 217 (Lohr)

13. a) (7 points) Exercice 4 page 268 (Lohr)
    b) (13 points) Exercice 5 page 268 (Lohr)

**David Haziza**