T. D. n^{o} 6
Estimation d'un ratio

Exercice 1. *R*atio. **Extrait du livre « Exercices corrigés de méthodes de sondage » de P. Ardilly et de Y. Tillé.**

Dans une population de 10 000 entreprises, nous voulons estimer le chiffre d'affaires moyen. Pour cela, nous échantillonnons $n = 100$ entreprises par SAS. Nous disposons par ailleurs de l'information auxiliaire « nombre de salariés », notée x, par entreprise. Les données issues du sondage sont :

— $\mu_X = 50$ salariés (vraie moyenne sur les x_k),
— $\widehat{\mu}_Y = 5,2 \times 10^6$ euros (moyenne du chiffres d'affaires dans l'échantillon),
— $\widehat{\mu}_X = 45$ salariés (moyenne des effectifs dans l'échantillon),
— $s^2_{y,c} = 25 \times 10^{10}$ (dispersion corrigée des y_k calculée dans l'échantillon)
— $s^2_{x,c} = 15$ (dispersion corrigée des x_k calculée dans l'échantillon)
— $\widehat{\rho(X,Y)} = 0,80$ (coefficient de corrélation linéaire entre X et Y calculé dans l'échantillon).

1. Que vaut l'estimateur par ratio ? Cet estimateur est-il biaisé ?
2. Rappeler la formule de variance « vraie » de cet estimateur.
3. Calculer une estimation de la variance vraie. L'estimateur de variance utilisé est-il biaisé ?
4. Donner un intervalle de confiance pour $\widehat{\mu}_Y$ à 95%.

Exercice 2. Estimation d'un ratio. D'après l'examen de Juin 2005, M1-IMSV.

Nous souhaitons estimer :

$$R = \frac{\overline{Y}}{\overline{X}}$$

par

$$r = \frac{\overline{y}}{\overline{x}}\cdot$$

Nous supposons que :

$$\mathbb{E}\left[\overline{y}\right] = \overline{Y} \quad \text{et} \quad \mathbb{E}\left[\overline{x}\right] = \overline{X}.$$

1. Écrire r sous la forme :

$$r = R\left(1 + \frac{\overline{y} - \overline{Y}}{\overline{Y}}\right)\left(\frac{1}{1 + \dfrac{\overline{x} - \overline{X}}{\overline{X}}}\right).$$

.

2. En utilisant le développement limité de $\dfrac{1}{1+u}$ à l'ordre 2, montrer que :
$$\mathbb{E}\left[r\right] \simeq R\left(1 - \frac{\mathrm{Cov}(\overline{x},\overline{y})}{\overline{X}\cdot\overline{Y}} + \frac{\mathrm{Var}\left[\overline{x}\right]}{\overline{X}^2}\right).$$
3. En déduire l'expression du biais et montrer, qu'en négligeant les termes d'ordre 2, l'erreur quadratique moyenne s'écrit :
$$\begin{aligned}\mathrm{EQM}(r) &= \mathbb{E}\left[(r-R)^2\right]\\ &\simeq \frac{1}{\overline{X}^2}\left(\mathrm{Var}\left[\overline{y}\right] - 2R\cdot\mathrm{Cov}(\overline{x},\overline{y}) + R^2\mathrm{Var}\left[\overline{x}\right]\right).\end{aligned}$$
4. Que valent les quantités $B(r)$ et $\mathrm{EQM}(r)$ dans le cas particulier d'un sondage aléatoire simple à probabilités égales sans remise ?
5. Application : nous cherchons à estimer par un sondage aléatoire simple à probabilités égales sans remise la quantité $R = \dfrac{1}{\pi}$ par $r = \dfrac{1}{\widehat{\pi}}$ où π et $\widehat{\pi}$ sont des proportions.
   (a) Montrer que dans ce cas :
   $$B(r) \simeq \frac{1-\pi}{\pi^2}\cdot\frac{(1-f)}{n}$$
   et
   $$\mathrm{EQM}(r) \simeq \frac{B(r)}{\pi}.$$
   (b) Quel est le biais si $n = 100$, $f = 1$ % et $\widehat{\pi} = 0,10$? Commenter.