D.8 / Simpson's paradox

We keep things simple and assume Uppsala has only two parts – Gottsunda and Luthagen. Here is a list of sold flats:

	Gottsunda	Luthagen	Uppsala
2011	50 kr 60 kr	200 kr 250 kr	50 kr ; 60 kr ; 200 kr ; 250 kr
	⌀ 55 kr	⌀ 225 kr	⌀ 140 kr
2012	50 kr	200 kr 220 kr	50 kr ; 200 kr ; 220 kr
	⌀ 50 kr	⌀ 210 kr	⌀ 156,67 kr

Problem on estimators

We have two independent samples of a $Exp(3\alpha)$ - resp. $Exp(4\alpha)$ distribution. The first of size 60 and mean $\overline{x} = 4.3$, the second of size 80 and mean $\overline{y} = 5.76$. In order to estimate α we have estimators $\hat{\alpha}_1 = \overline{y} - \overline{x}$ and $\hat{\alpha}_2 = \frac{\overline{x} + \overline{y}}{7}$.

Show unbiasedness of estimators. Which of the two is more effective?

Repetition: $Z \sim Exp(m) \leadsto E Z = m$; $Var Z = m^2$.

We have $X_1, \ldots, X_{60} \sim Exp(3\alpha)$ independent.

Put $\overline{X} = \frac{1}{60} \sum_{i=1}^{60} X_i$. Then

$$E\overline{X} = \frac{1}{60} \sum_{i=1}^{60} EX_i = \frac{1}{60} \sum_{i=1}^{60} 3\alpha = 3\alpha$$

and

$$Var\,\overline{X} = Var\left(\frac{1}{60} \sum_{i=1}^{60} X_i\right) \underset{\uparrow}{=} \frac{1}{60^2} \sum_{i=1}^{60} Var(X_i) = \frac{(3\alpha)^2}{60}$$

X_i indep.

$$= \frac{3}{20}\alpha^2$$

In the same way: $Y_1, \ldots, Y_{80} \sim Exp(4\alpha)$ indep.

$$\overline{Y} = \frac{1}{80} \sum_{i=1}^{80} Y_i \leadsto E\overline{Y} = \ldots = 4\alpha$$

$$Var\,\overline{Y} = \ldots = \frac{(4\alpha)^2}{80} = \frac{1}{5}\alpha^2$$

Now consider $\hat{\alpha}_1 = \overline{Y} - \overline{X}$.

we have

$$\mathbb{E}\hat{\alpha}_1 = \mathbb{E}\overline{Y} - \mathbb{E}\overline{X} = 4\alpha - 3\alpha = \alpha$$

$\hookrightarrow$ $\hat{\alpha}_1$ is unbiased //

Moreover,

$$\mathrm{Var}\,\hat{\alpha}_1 = \mathrm{Var}(\overline{Y} - \overline{X})$$

$$\underset{\overline{X},\overline{Y}\text{ indep.}}{=} \mathrm{Var}(\overline{Y}) + \mathrm{Var}(\overline{X})$$

$$= \frac{3}{20}\alpha^2 + \frac{1}{5}\alpha^2 = \frac{7}{20}\alpha^2 \quad //$$

Consider $\hat{\alpha}_2 = \dfrac{\overline{X} + \overline{Y}}{7}$. we have

$$\mathbb{E}\hat{\alpha}_2 = \mathbb{E}\left(\frac{\overline{X}+\overline{Y}}{7}\right) = \frac{1}{7}(\mathbb{E}\overline{X} + \mathbb{E}\overline{Y})$$

$$= \frac{1}{7}(3\alpha + 4\alpha) = \alpha$$ $\hookrightarrow$ $\hat{\alpha}_2$ unbiased //

Moreover,

$$\mathrm{Var}\,\hat{\alpha}_2 = \mathrm{Var}\left(\frac{\overline{X}+\overline{Y}}{7}\right) = \frac{1}{49}(\mathrm{Var}\,\overline{X} + \mathrm{Var}\,\overline{Y})$$

$$= \frac{1}{49}\cdot\frac{7}{20}\alpha^2 = \frac{1}{140}\alpha^2 \quad //$$

It follows $\mathrm{Var}\,\hat{\alpha}_2 < \mathrm{Var}\,\hat{\alpha}_1$ $\hookrightarrow$ $\hat{\alpha}_2$ is more effective than $\hat{\alpha}_1$.

Note: We never used the numbers 4,3 and 5,76 !