單元 12: 選取樣本大小
(課本 §8.7)

樣本大小乃相當於估計的優良性. 選取樣本大小需先確認 (1) 估計誤差的界限 與 (2) 信心水準 (confidence level), 意即,

$$P(|\widehat{\theta}-\theta|<b)=1-\alpha$$

其中 b 爲估計誤差, $1-\alpha$ 爲信心水準.

接著, 根據下述二規則選取樣本大小.

(1) 經驗法則 (empirical rule).
舉例, 估計某化學肥料對農作物的平均日產量 (average daily yield) μ 時, 若假設此農作物日產量的値距 (全距, 範圍, range of daily yield) 爲 84 噸, 試求樣本大小 n 使得

$$|\widehat{\theta}-\theta|=|\overline{Y}-\mu|\leq 5 \text{ 噸}$$

的機率爲 0.95.

<解> 由經驗法則,

$$P(|\overline{Y}-\mu|\leq 2\sigma_{\overline{Y}})=0.95$$

故令 $2\sigma_{\overline{Y}} = 5$, 意即 $2\frac{\sigma}{\sqrt{n}} = 5$. 由此得

$$n = \left(\frac{2}{5}\sigma\right)^2 = \frac{4\sigma^2}{25}$$

問. 如何決定 σ?

答. (i) 根據之前的樣本, 以 S 估計 σ, 或 (ii) 採用樣本值距 (全距) 的資訊. 此處以 (ii) 決定 σ, 因爲已知給定的值距爲 84 噸. 首先, 由經驗法則,

$$P(|Y - \mu| \leq 2\sigma) = 0.95$$

此乃相當於

$$P(\mu - 2\sigma \leq Y \leq \mu + 2\sigma) = 0.95$$

意即, 有 0.95 的機率 Y 的值距爲

$$(\mu + 2\sigma) - (\mu - 2\sigma) = 4\sigma$$

故, 可設定 $4\sigma \approx 84$ 且得 $\sigma \approx 21$. 因此,

$$n = \frac{4\sigma^2}{25} \approx \frac{4(21)^2}{25} = 70.56$$

且可選取樣本大小 $n = 71$, 雖爲概略 (粗糙) 的估計, 但優於無根據的任意猜測.

(2) 大樣本法則 (large sample rule). 對於不偏估計量 $\widehat{\theta}$, 當 $n \to \infty$ 時,

$$\frac{\widehat{\theta}-\theta}{\sigma_{\widehat{\theta}}} \xrightarrow{\mathcal{D}} N(0,1)$$

因此,

$$\begin{aligned} 1-\alpha &= P(|\widehat{\theta}-\theta| \le b) = P\left(\left|\frac{\widehat{\theta}-\theta}{\sigma_{\widehat{\theta}}}\right| \le \frac{b}{\sigma_{\widehat{\theta}}}\right) \\ &\approx P\left(|Z| \le \frac{b}{\sigma_{\widehat{\theta}}}\right) \end{aligned}$$

其中 $Z \sim N(0,1)$.

接著, 由表 4, 選取 $z_{\alpha/2}$ 使得

$$P(Z > z_{\alpha/2}) = \frac{\alpha}{2}$$

如圖示, 或

$$\text{R} : z_{\alpha/2} = \texttt{qnorm}(1-\alpha/2)$$

最後, 設定 $z_{\alpha/2} = \frac{b}{\sigma_{\widehat{\theta}}}$, 即 $z_{\alpha/2}\sigma_{\widehat{\theta}} = b$, 並解 n.

例 1. 設某心理測驗會呈現出 A 或 B 二種型式的反應. 今欲估計受測者呈現 A 反應的機率 p. 試求樣本大小 n

使得在信心水準爲 0.90 下,

$$|\widehat{\theta}-\theta|=|\widehat{p}-p|\leq 0.04$$

另假設 $p\approx 0.6$.

<解> 由大樣本法則, 可設定

$$0.04=z_{\alpha/2}\sigma_{\widehat{p}}=z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

其中

$$P(Z>z_{\alpha/2})=0.05\ \left(=\frac{1-0.90}{2}\right)$$

由表 4,

$$P(Z>1.645)=0.05$$

得 $z_{\alpha/2}=1.645$, 或

$$\text{R}: z_{\alpha/2}=\texttt{qnorm}(0.95)$$

以及

$$0.04=1.645\sqrt{\frac{pq}{n}}$$

同樣的問題: 如何決定 p?

答. (i) 由假設, $p\approx 0.6$, 得

$$n=\left(\frac{1.645}{0.04}\right)^2(0.6)(0.4)=405.91$$

且選取 $n = 406$

或 (ii) 由 $p(1-p)$ 在 $p = \frac{1}{2}$ 有最大值, 得

$$\begin{aligned} n &= \left(\frac{1.645}{0.04}\right)^2 pq = \left(\frac{1.645}{0.04}\right)^2 p(1-p) \\ &\leq \left(\frac{1.645}{0.04}\right)^2 p(1-p)\Big|_{p=1/2} = 422.82 \end{aligned}$$

一個 n 的上界, 並選取 $n = 423$, 一個 n 的最大可能值.