< 母平均の区間推定 2 >

例 ある大都市の満 17 歳の男子の平均身長 μ(cm) を求めるのに、50 名の標本調査を無作為復元抽出により行った。そのデータを $X_1,\ X_2,\ \cdots,\ X_{50}$ としたとき、平均と標準偏差は

$$\overline{X} = \frac{1}{50}\sum_{i=1}^{50} X_i = 169.5(\mathrm{cm}) \quad : \quad \text{標本平均}$$

$$s = \sqrt{\frac{1}{50}\sum_{i=1}^{50}(X_i - 169.5)^2} = 5.6 \quad : \quad \text{標本標準偏差}$$

であった。過去のデータからみて、この都市における 17 歳男子の身長の分布は正規分布 $N(\mu, \sigma^2)$ で近似できる。また標本の大きさ (= 50) が大きいので、σ^2 を $s^2 = 5.6^2$ で代用する。正規分布の性質より、平均 $\overline{X} = \dfrac{1}{50}\displaystyle\sum_{i=1}^{50} X_i$ は正規分布 $N(\mu, \frac{5.6^2}{50})$ に従うので、前ページより、$n = 50,\ \alpha = 1.96$ のとき

$$P\left(\left|\overline{X} - \mu\right| \leqq \frac{\sigma}{\sqrt{50}} \times 1.96\right) = \int_{-1.96}^{1.96} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0.95(= 95\%)$$

であるから、$\overline{X} = 169.5,\ \sigma = 5.6$ を代入すると、確率 $p = 0.95$ で

$$|169.5 - \mu| \leqq \frac{5.6}{\sqrt{50}} \times 1.96 \Leftrightarrow 169.5 - \frac{5.6}{\sqrt{50}} \times 1.96 \leqq \mu \leqq 169.5 + \frac{5.6}{\sqrt{50}} \times 1.96$$

$$\Leftrightarrow \boxed{167.95 \leqq \mu \leqq 171.05} \quad \cdots \quad (*)$$

と推定できる。(*) が確率 95% の信頼区間である。

問 付録の正規分布表を見て、この例の場合に、確率 99% の信頼区間を求めよ。

(注) 95% の確率で区間推定を行うとき、95% を**信頼度**という。

一般に正規母集団 $N(\mu, \sigma^2)$ からの標本数 n が大きいとき

信頼度 $f(\alpha) = \displaystyle\int_{-\alpha}^{\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ の信頼区間は

$$\boxed{\overline{X} - \alpha \times \frac{s}{\sqrt{n}} \leqq \mu \leqq \overline{X} + \alpha \times \frac{s}{\sqrt{n}}} \quad \cdots \quad (*)'$$

である。ここで $s = \sqrt{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$ は標本標準偏差である。