

No.12 区間推定

区間推定

- 点推定は推定値は抽出された標本ごとに値が変化し、推定値が母数に一致することはまずない。
- 未知の母数が「ある確からしさ」で入る区間でもって推定する方法を区間推定という。
- 標本により得られるその区間を**信頼区間**といい、はじめに定める「確からしさ」を**信頼係数**という。

準備 (正規分布の再帰性)

定理 (正規分布の再帰性)

確率変数 X_1, X_2 が互いに独立で、それぞれ正規分布 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ に従う ($X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2)$) とするとき、定数 a_1, a_2 に対して $a_1X_1 + a_2X_2 \sim N(a_1\mu_1 + a_2\mu_2, a_1^2\sigma_1^2 + a_2^2\sigma_2^2)$ である。

母平均の区間推定

- 大きさ 10 の無作為標本

62.3 71.8 48.4 33.4 42.0 77.0 61.7 37.2 45.8 45.4

を抽出したとする。標本平均 $\bar{X}$ の実現値 $\bar{x}$ は 52.5

- 過去のデータから母集団分布が $N(\mu, 14.6^2)$ であると考えられるとする (μ を推定したい, 分散 14.6^2 は既知)。
- このとき大きさ 10 の無作為標本 $\bar{X}$ は $\bar{X} \sim N(\mu, 14.6^2/10)$ であるから

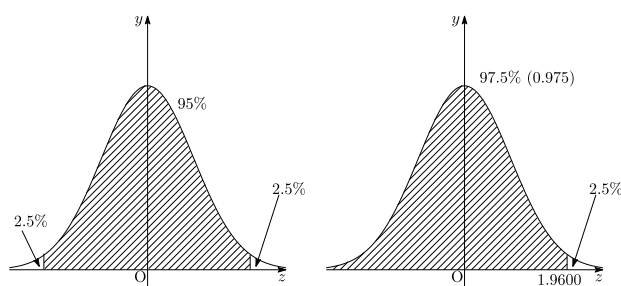
$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{14.6^2}{10}}} \sim N(0, 1)$$

- 信頼係数を 95%(0.95) とし

$$P(-z \leq Z \leq z) = 0.95$$

となる z を求める。

- このような z は正規分布表から $z = 1.9600$ が得られる：



- $$P(-1.9600 \leq Z \leq 1.9600) = 0.95 \text{ つまり } P\left(-1.9600 \leq \frac{\bar{X} - \mu}{\sqrt{\frac{14.6^2}{10}}} \leq 1.9600\right) = 0.95 \text{ より}$$

$$-1.9600 \leq \frac{\bar{X} - \mu}{\sqrt{\frac{14.6^2}{10}}} \leq 1.9600$$

を変形すると

$$\begin{aligned} -1.9600\sqrt{\frac{14.6^2}{10}} &\leq \bar{X} - \mu \leq 1.9600\sqrt{\frac{14.6^2}{10}} \\ \bar{X} - 1.9600\sqrt{\frac{14.6^2}{10}} &\leq \mu \leq \bar{X} + 1.9600\sqrt{\frac{14.6^2}{10}} \end{aligned}$$

となる確率が 0.95 である.

- おおざっぱに言って標本抽出を 100 回すれば 100 個の $\bar{X}$ の実現値 $\bar{x}$ ができ、これを上に代入することにより 100 個の不等式ができるが、そのうち 95 個は正しい不等式である.**

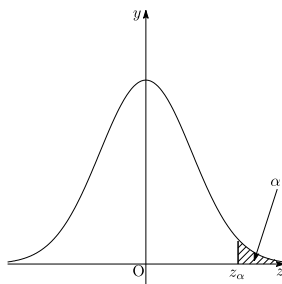
$\bar{X}$ の実現値 $\bar{x} = 52.5$ を代入するすると

$$\underset{\text{信頼限界}}{43.5} \leq \mu \leq \underset{\text{信頼限界}}{61.55} \quad \mu \text{ の } 95\% \text{ 信頼区間}$$

- 信頼係数は 95% や 99% を用いることが多い.
- $Z \sim N(0, 1)$ であるとき, $0 < \alpha < 1$ に対して

$$P(Z \geq z_\alpha) = \alpha$$

となる z_α を標準正規分布の**上側 α 点**という.



例

- $z_{0.025} = 1.9600 \approx 1.960$ (上側 2.5%点)
- $z_{0.005} = 2.5758 \approx 2.576$ (上側 0.5%点)
- 同様にして次の信頼区間の公式が得られる.

母平均の区間推定 (母分散既知)

正規母集団 $N(\mu, \sigma^2)$ (σ^2 既知) から抽出した大きさ n の無作為標本について標本平均の実現値を $\bar{x}$ とするとき, 母平均 μ の信頼係数 $1 - \alpha$ ($100(1 - \alpha)\%$) の信頼区間は

$$\bar{x} - z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}$$

母比率の区間推定

定理 (中心極限定理)

確率変数 $X_1, X_2, \dots, X_n$ が互いに独立でそれぞれ平均 μ , 分散 σ^2 をもつ**同一の確率分布**に従うとする. n が大きいとき $\bar{X} = \frac{X_1 + \dots + X_n}{n}$ は近似的に正規分布 $N\left(\mu, \frac{\sigma^2}{n}\right)$ に従う, 正確には

$$\lim_{n \rightarrow \infty} P\left(\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \leq z\right) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

つまり, 母平均 μ , 母分散 σ^2 の任意の母集団から抽出した大きさ n ば無作為標本について, その標本平均を $\bar{X}$ とするとき $\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$ は n が大きいとき近似的に $N(0, 1)$ に従う.

二項母集団

- ある政策について賛成か反対か, ワールドカップの中継を観たか, 名探偵コナンの映画を観たか
- 各要素が特定の性質をもつかもたないかに注目する母集団を**二項母集団**という.
- 二項母集団から大きさ 1 の無作為標本 X を抽出するとき, その性質をもつとき $X = 1$, そうでないときは $X = 0$ とする.
- $P(X = 1)$ を p で表し**母比率**という. つまり, その性質をもつ要素の割合である.
- $X_1, X_2, \dots, X_n$ を母比率 p の二項母集団からの無作為標本とすると

$$X_1 + X_2 + \dots + X_n \sim B(n, p) \quad (\text{平均 } np, \text{ 分散 } npq)$$
$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \quad \text{平均 } p, \text{ 分散 } \frac{npq}{n^2} = \frac{pq}{n}$$

- 二項母集団の場合, 標本平均は**標本比率**といい, $\hat{P}$ で表す
- $E[\hat{P}] = p, V[\hat{P}] = \frac{pq}{n}$
- 中心極限定理により n が十分大きいとき近似的に

$$Z = \frac{\hat{P} - p}{\sqrt{\frac{pq}{n}}} \sim N(0, 1)$$

例 母比率 0.5 の二項母集団から大きさ 400 の無作為標本を抽出したとき, 標本比率 $\hat{P}$ について $P(\hat{P} \geq 0.51)$ は?

中心極限定理から近似的に

$$Z = \frac{\hat{P} - 0.5}{\sqrt{\frac{0.5 \times 0.5}{400}}} = \frac{\hat{P} - 0.5}{\frac{0.5}{20}} = 40(\hat{P} - 0.5) \sim N(0, 1)$$

よって

$$P(\hat{P} \geq 0.51) = P(Z \geq 0.4) = 1 - P(Z \leq 0.4) = 1 - 0.6554 = 0.3446$$

母比率の区間推定

- 母比率 p を標本比率を用いて区間推定することを考える.

- n が大きいとき近似的に $Z = \frac{\hat{P} - p}{\sqrt{\frac{pq}{n}}} \sim N(0, 1)$

- 信頼係数を 95% とすると

$$P\left(-z_{0.025} \leq Z = \frac{\hat{P} - p}{\sqrt{\frac{pq}{n}}} \leq z_{0.025}\right) = 0.95$$

- 不等式を変形する ($q = 1 - p$ に注意)

$$\begin{aligned} -z_{0.025} &\leq \frac{\hat{P} - p}{\sqrt{\frac{pq}{n}}} \leq z_{0.025} \\ -z_{0.025} \sqrt{\frac{p(1-p)}{n}} &\leq \hat{P} - p \leq z_{0.025} \sqrt{\frac{p(1-p)}{n}} \\ -z_{0.025} \sqrt{\frac{p(1-p)}{n}} &\leq p - \hat{P} \leq z_{0.025} \sqrt{\frac{p(1-p)}{n}} \\ \hat{P} - z_{0.025} \sqrt{\frac{p(1-p)}{n}} &\leq p \leq \hat{P} + z_{0.025} \sqrt{\frac{p(1-p)}{n}} \end{aligned}$$

この不等式を p について解けば (両端にも p が!) 95%信頼区間が得られるが複雑. しかし n が十分大きく, 標本比率 (もともとは標本平均) は母比率 p の一致推定量であることに注意し, 通常は両辺の p は標本比率の実現値 $\hat{p}$ で置き換える. また, $\hat{P}$ も $\hat{p}$ で置き換えて

$$\begin{aligned} \hat{p} - z_{0.025} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &\leq p \leq \hat{p} + z_{0.025} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ \hat{p} - 1.9600 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &\leq p \leq \hat{p} + 1.9600 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \end{aligned}$$

- 信頼係数 95% $\rightarrow z_{0.025}$, 信頼係数 99% $\rightarrow z_{0.005}$, 信頼係数 $100(1 - \alpha) \%$ $\rightarrow z_{\alpha/2}$ を使う.

母比率の区間推定

母比率 p の二項母集団からの大きさ n の無作為標本を抽出し, 標本比率の実現値を $\hat{p}$ とすると, 母比率 p の信頼係数 $1 - \alpha$ ($100(1 - \alpha) \%$) 信頼区間は

$$\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

例題

A 市で, あるテレビ番組の視聴率を調べるために, 成人 500 人を無作為抽出したところ, 45 人が見ていることがわかった. A 市におけるこの番組の成人の視聴率の 95%信頼区間を求めよ.

解 標本比率は $\hat{p} = \frac{45}{500} = 0.09$ であるので, 視聴率 (母比率) p の 95%信頼区間は

$$\begin{aligned} \hat{p} - 1.9600 \sqrt{\frac{\hat{p}(1-\hat{p})}{500}} &\leq p \leq \hat{p} + 1.9600 \sqrt{\frac{\hat{p}(1-\hat{p})}{500}} \\ 0.09 - 1.9600 \sqrt{\frac{0.09 \times 0.91}{500}} &\leq p \leq 0.09 + 1.9600 \sqrt{\frac{0.09 \times 0.91}{500}} \\ 0.065 &\leq p \leq 0.115 \end{aligned}$$