統計的推定

母集団の特徴を表す数値(母数)の近似値を標本から求める.

これを統計的推定法という.

点推定

標本変量$X_1, \cdots, X_n$から
なんらかの式で計算される量.

例. 20歳男子の平均身長を調べるために.

100人の標本をとり.その平均値を近似値とする.

このように.未知の母数θ(平均身長)に対し.ある統計量$\Theta(X_1, \cdots, X_n)$(100人の標本の平均を表す確率変数)の実現値(平均値)を近似値とする方法を.点推定という.

ここで$\Theta(X_1, \cdots, X_n)$を推定量,その実現値を推定値という.

→推定量の候補は実はいろいろある.例えば、上の例で.

重みつき平均

$$\Theta(X_1, \cdots, X_{100}) = \frac{a_1X_1 + a_2X_2 + \cdots + a_{100}X_{100}}{a_1 + a_2 + \cdots + a_{100}} \quad (a_i > 0)$$

としてもよい.しかし.推定量は以下の3つの条件をみたすべきである

(1) 不偏性 : $E(\Theta(X_1, \cdots, X_n)) = \theta$

(2) 一致性 : $\lim_{n \to \infty} P(|\Theta(X_1, \cdots, X_n) - \theta| > \varepsilon) = 0$

(3) 有効性 : $V(\Theta(X_1, \cdots, X_n))$が小さい.
　　　　　　　　　　　　　　　　　　　より.

→小さい方が優れている.

区間推定.

点推定はわかりやすく、計算も単純.

⟶ 本当に、母数に近い値が出ているのか疑問.

そこで、未知の母数に対して.

"母数が $X-\delta$ から $X+\delta$ の間に入る確率が $1-\alpha$" という求め方をする.

このような考え方を区間推定法という.

例. 20歳男子の平均身長を μ, 分散を 25 とする.

標本として n 人をとりその平均身長 $\overline{X}=\frac{1}{n}(X_1+\cdots+X_n)$ を考えるとき.

μ が $\overline{X}-1$ から $\overline{X}+1$ の間に入る確率が 99% 以上になるためには.

n がいくつ以上であればよいか. ただし $\overline{X}$ は $N(\mu, \frac{25}{n})$ に従うとしてよい.

答 $P(\overline{X}-1 \leqq \mu \leqq \overline{X}+1)$

$=P(\mu-1 \leqq \overline{X} \leqq \mu+1)$ とできる. ここで. $Z=\dfrac{\overline{X}-\mu}{\frac{5}{\sqrt{n}}}$ とおくと

$=P\left(\dfrac{-1}{\frac{5}{\sqrt{n}}} \leqq Z \leqq \dfrac{1}{\frac{5}{\sqrt{n}}}\right)$

$=P\left(-\dfrac{\sqrt{n}}{5} \leqq Z \leqq \dfrac{\sqrt{n}}{5}\right)=2\cdot P\left(0 \leqq Z \leqq \dfrac{\sqrt{n}}{5}\right) \geqq 0.99.$

$\therefore P\left(0 \leqq Z \leqq \dfrac{\sqrt{n}}{5}\right) \geqq 0.495$ より

$\dfrac{\sqrt{n}}{5} \geqq 2.5758$ であればよい.

$\therefore n \geqq 165.87.$ より n は 166 以上であればよい.

問題. 上の例で確率が 95% 以上になるには、n がいくつ以上であればよいか.

答. $P(0 \leqq Z \leqq \frac{\sqrt{n}}{5}) \geqq 0.475$ より

$\sqrt{n} \geqq 1.96 \times 5$ となればよい.

∴ n は97以上であればよい.

区間推定では.3つの数値が連動している.

(1) 標本のサイズ　(2) $\bar{X}-\delta < \mu < \bar{X}+\delta$ という区間. (3) $1-\alpha$を表す信頼度.

標本の実現値 $\bar{X}$ を得たとする.

$\bar{X}-\delta < \mu < \bar{X}+\delta$ である確率が $1-\alpha$ であるとき.

区間 $\bar{X}-\delta < \mu < \bar{X}+\delta$ を母平均 μ の $1-\alpha$ 信頼区間 という.

例. さきほどの例で. 標本のサイズが50であるとき. その平均値が169.8であった.

平均身長の 0.99信頼区間 (99%信頼区間)を求めよ.

答. $\bar{X}$ を平均値とすると.

$P(\bar{X}-\delta \leqq \mu \leqq \bar{X}+\delta) = 0.99$ となる δ を求めればよい.

$P(\bar{X}-\delta \leqq \mu \leqq \bar{X}+\delta) = P(\mu-\delta \leqq \bar{X} \leqq \mu+\delta)$

ここで. $\bar{X}$ は $N(\mu, \frac{25}{50})$ に従うので. $Z = \frac{\bar{X}-\mu}{\frac{1}{\sqrt{2}}}$ とおくと.

$= P(-\sqrt{2}\delta \leqq Z \leqq \sqrt{2}\delta)$

$= 2P(0 \leqq Z \leqq \sqrt{2}\delta) = 0.99$ となる

∴ $P(0 \leqq Z \leqq \sqrt{2}\delta) = 0.495$ より

$\sqrt{2}\delta = 2.5758$ となり　$\delta = 1.8$ となる.

∴ 平均身長の99%信頼区間は. $168.0 \leqq \mu \leqq 171.6$ である

<u>問題</u>. 標本のサイズが100で その平均値が170.1であった.

平均身長の95%信頼区間を求めよ.

答. $P(\overline{X}-\delta \leqq \mu \leqq \overline{X}+\delta)=0.95$ となる δ を求めると.

$P(\overline{X}-\delta \leqq \mu \leqq \overline{X}+\delta)=P(\mu-\delta \leqq \overline{X} \leqq \mu+\delta)$

ここで $\overline{X}$ は $N(\mu, \frac{25}{100})$ に従うので $Z=\frac{\overline{X}-\mu}{\frac{1}{2}}$ とおくと

$=P(-2\delta \leqq Z \leqq 2\delta)$

$=2P(0 \leqq Z \leqq 2\delta)=0.95$

$\therefore P(0 \leqq Z \leqq 2\delta)=0.475$ となり

$2\delta=1.96$ より $\delta=0.98 \fallingdotseq 1.0$

$\therefore$ 平均身長の95%信頼区間は $169.1 \leqq \mu \leqq 171.1$ である

これを公式化してみる.

母平均が μ, 母分散が σ^2 とすると 標本平均 $\overline{X}$ は $N(\mu, \frac{\sigma^2}{n})$ に従う.

これの $1-\alpha$ 信頼区間を求めると.

$P(\overline{X}-\delta \leqq \mu \leqq \overline{X}+\delta)=1-\alpha$.

$P(\mu-\delta \leqq \overline{X} \leqq \mu-\delta)=1-\alpha$ ここで $Z=\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ とおくと.

$P(-\frac{\sqrt{n}\delta}{\sigma} \leqq Z \leqq \frac{\sqrt{n}\delta}{\sigma})=1-\alpha$ となる

ここで $z(\frac{\alpha}{2})$ を $P(-z(\frac{\alpha}{2}) \leqq Z \leqq z(\frac{\alpha}{2}))=1-\alpha$ をみたす値とすると

$\frac{\sqrt{n}\delta}{\sigma}=z(\frac{\alpha}{2})$ より. $\delta=\frac{\sigma}{\sqrt{n}}z(\frac{\alpha}{2})$ となり.

μ の $1-\alpha$ 信頼区間は.

$\overline{X}-\frac{\sigma}{\sqrt{n}}z(\frac{\alpha}{2}) < \mu < \overline{X}+\frac{\sigma}{\sqrt{n}}z(\frac{\alpha}{2})$ となる