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とを.

統計的推定という.

点推定

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.

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

このように、未知の母数θ(平均身長)に対し.ある統計量$\Theta(X_1, \cdots, X_n)$(標本平均)

の実現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1つではない.例えば、上の例で

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

としてもよい.ただし、推定量は以下の3つの条件をみたすことが必要.

(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 : $V(\Theta(X_1, \cdots, X_n))$ がより小さい.

→ 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 人とり、

その平均身長 $\bar{X}=\frac{1}{n}(X_1+\cdots+X_n)$ を考えるとき、

μ が $\bar{X}-1$ から $\bar{X}+1$ 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 $\bar{X}$ は $N(\mu, \frac{25}{n})$ に従うとしてよい。

答. $P(\bar{X}-1 \leqq \mu \leqq \bar{X}+1) = P(\mu-1 \leqq \bar{X} \leqq \mu+1)$

とできる。ここで、$Z=\dfrac{\bar{X}-\mu}{\frac{5}{\sqrt{n}}}$ とすれば、Z は $N(0,1)$ に従い、

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

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

$P\left(0 \leqq Z \leqq \dfrac{\sqrt{n}}{5}\right) \geqq 0.495$ である。これをみたすためには、正規分布表から、

$\dfrac{\sqrt{n}}{5} \geqq 2.5758$ であればよいので、$n \geqq 165.87$ となり、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. $P\left(0 \leqq Z \leqq \dfrac{\sqrt{n}}{5}\right) \geqq 0.475$ より $\sqrt{n} \geqq 1.96\times 5 = 9.8$ となり

n は 97 以上であればよい。

問題. 確率を 90% 以上としたときはどうか。

答 $P\left(0 \leqq Z \leqq \dfrac{\sqrt{n}}{5}\right) \geqq 0.45$ より $\sqrt{n} \geqq 1.6449\times 5 = 8.2245$ となり

n は 68 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) $\overline{X}-\delta<\mu<\overline{X}+\delta$ という区間　(3) $1-\alpha$ が表す信頼度

標本の実現値 $\overline{X}$ を得たとする.

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

区間 $\overline{X}-\delta<\mu<\overline{X}+\delta$ 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、その実現値が169.8であった.

平均身長の99%信頼区間を求めよ.

答. $\overline{X}$ を平均値とし. $P(\overline{X}-\delta \leqq \mu \leqq \overline{X}+\delta)=0.99$ となる δ を求めればよい.

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

ここで $Z=\dfrac{\overline{X}-\mu}{\frac{1}{\sqrt{2}}}$ とすれば、Z は $N(0,1)$ に従い

$=P\left(\mu-\delta \leqq \frac{1}{\sqrt{2}}Z+\mu \leqq \mu+\delta\right)=P(-\sqrt{2}\delta \leqq Z \leqq \sqrt{2}\delta)$

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

$\sqrt{2}\delta=2.5758$　より　$\delta \fallingdotseq 1.8$ となる

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

問題. 上の例で、標本のサイズが100, 実現値が170.1であった.

平均身長の99%信頼区間と、95%信頼区間をそれぞれ求めよ

答. $P(\overline{X}-\delta \leqq \mu \leqq \overline{X}+\delta)=0.99$ 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 $Z=\dfrac{\overline{X}-\mu}{\frac{1}{2}}=2(\overline{X}-\mu)$ とすれば、Z は $N(0,1)$ に従い

$=P(-2\delta \leqq Z \leqq 2\delta)=2P(0 \leqq Z \leqq 2\delta)=0.99$　となる

∴ $P(0 \leqq Z \leqq 2\delta) = 0.495$ となり $2\delta = 2.5758$, $\delta = 1.2879 \fallingdotseq 1.3$ となる

∴ 99%信頼区間は $168.8 \leqq \mu \leqq 171.4$ である

同様に.

$P(0 \leqq Z \leqq 2\delta) = 0.475$ となる δ は. $2\delta = 1.96$ より $\delta = 0.98 \fallingdotseq 1.0$ より.

95%信頼区間は. $169.1 \leqq \mu \leqq 171.1$ である.

これを公式化すると次のようになる.

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

これの $1-\alpha$ 信頼区間は.

$P(\overline{X} - \delta \leqq \mu \leqq \overline{X} + \delta) = 1 - \alpha$ をみたす. これを変形する.

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

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

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

$z(\frac{\alpha}{2})$ は正規分布表から求める

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.

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

μ の $1-\alpha$ 信頼区間は.

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

となる.