応用数理 E 第 10 回目

永幡幸生
新潟大学工学部

5 月 17 日

区間推定：母分散が既知で正規分布に従う場合

前回の点推定は1つの値でパラメータの推定を行ったが、どのくらいの誤差があるのかは不明である。誤差を含む形で推定するために、区間を与えることにする。

区間推定：母分散が既知で正規分布に従う場合

○ 区間推定の基本的な考え方
設定は点推定と同じで、
$X_1, X_2, \ldots, X_n$ を独立同分布でその分布関数は $F_\theta(x)$ とする。（母集団分布）
誤差の範囲として α は小さい数を決める。
多くの場合 $\alpha = 0.05$ (5%) または 0.01 (1%) である。
(必要であれば、この α に合わせて）統計量
$T_1 = T_1(X_1, X_2, \ldots, X_n), \quad T_2 = T_2(X_1, X_2, \ldots, X_n), \ T_1 \leq T_2$
をとって、
$1 - \alpha = P(T_1 \leq \theta \leq T_2)$
となるように作る。
このとき実現値
$t_1 = T_1(x_1, x_2, \ldots, x_n), \quad t_2 = T_2(x_1, x_2, \ldots, x_n)$
に対して、
$\theta \in [t_1, t_2]$
と区間推定する。

区間推定：母分散が既知で正規分布に従う場合

用語の定義

前ページの設定で
$1-\alpha$：信頼係数、信頼度、信頼水準
$[t_1, t_2]$：信頼区間
t_1：下側信頼限界
t_2：上側信頼限界
と呼ぶ。

区間推定：母分散が既知で正規分布に従う場合

○ 母分散が既知で正規分布に従う母平均の区間推定
「母分散が既知」の仮定は、今まで同じような量を多く観測していて、母分散の推定ができている場合である。
よく挙げられる例として、（小学校などの）身体測定の結果などがある。
次のページにあるような考え方をする。

区間推定：母分散が既知で正規分布に従う場合

標本平均 $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$ は $N(\mu, \frac{\sigma^2}{n})$ に従うので、

$\dfrac{(\bar{X}-\mu)}{\sqrt{\frac{\sigma^2}{n}}}$ は $N(0,1)$ に従う。

与えられた α に対して $z_{\alpha/2}$ を標準正規分布の上側 $\alpha/2$ 点とすると、

$$1-\alpha = P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = P(-z_{\alpha/2} \le \frac{(\bar{X}-\mu)}{\sqrt{\frac{\sigma^2}{n}}} \le z_{\alpha/2})$$

$$= P(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \le \mu \le \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2})$$

と書け、仮定から σ は既知。また与えられた α に対して、付表を使えば $z_{\alpha/2}$ も得られる。

従って μ の信頼係数 $1-\alpha$ の信頼区間は

$\left[\bar{x} - \dfrac{\sigma}{\sqrt{n}} z_{\alpha/2}, \bar{x} + \dfrac{\sigma}{\sqrt{n}} z_{\alpha/2}\right]$ ととれる。

区間推定：母分散が既知で正規分布に従う場合

◦ 新潟の 8 月の気温の例

まず第 8 回目の資料にあるように、新潟の 8 月の気温（最高気温、最低気温共に）、ヒストグラムを見た感じでは、正規分布に従っているとは思えないので、今回、次回共に例としてはあまりふさわしくないが、適応してみる。

Book1.csv にデータをコピーしてあるので、適宜参照するように。

区間推定：母分散が既知で正規分布に従う場合

新潟の 8 月の最高気温に関して、2016 年から 2020 年までの 5 年分のデータ 31×5 日分ある。これらのデータは、年ごとに $N(\mu, \sigma^2)$ に従うが、σ^2 は 5 年分の標本分散で代用して、$\sigma^2 = v^2 \cong 9.0$ として各年の母平均を信頼係数 95% で区間推定する。

データ数 = 日数 $n = 31$、であり、各年

標本平均 $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$ の実現値 $\bar{x}$ を計算すると、

$$\bar{x}_{16} = 31.0, \quad \bar{x}_{17} = 29.8, \quad \bar{x}_{18} = 30.7, \quad \bar{x}_{19} = 31.7 \quad \bar{x}_{20} = 31.5$$

となる。信頼係数 95% としたので教科書付表 2 より

$z_{0.025} = 1.96$ を得る。$n = 31, \sigma^2 = 9.0$ より 2 ページ前の結果に代入することで、2016 年から順に

$$[29.9, 32.0], \quad [28.8, 30.9], \quad [29.7, 31.8], \quad [30.6, 32.7], \quad [30.4, 32.5]$$

と与えられる。

区間推定：母分散が既知で正規分布に従う場合

問題

最高気温と同様に最低気温に関しても 5 年分信頼係数 95% で区間推定をせよ。但し σ^2 は 5 年分の標本分散で代用して、
$\sigma^2 = v^2 \cong 4.9$ とする。（電卓等使用可、付表を使う）
前回と同様、表計算ソフトなどを使って構わないが、どのような計算をするのかや、途中経過を書き入れるように。