< 復元抽出による標本調査 1 >

例 2 全国の高校 3 年生の男子の身長を調べたい。全国の高校 3 年生の男子全体は N 人いるとして、その身長データを $\{y_1, y_2, \cdots, y_N\}$ とする。これを母集団とし、その平均と分散を、それぞれ

$$\mu = \frac{1}{N}\sum_{i=1}^{N} y_i \quad (\text{母平均}) \quad , \quad \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \mu)^2 \quad (\text{母分散})$$

母平均、**母分散**という。この母集団全てを調査するのは大変なので、その中から n 人だけ調べて、全体を推測したい。n 人を無作為に復元抽出し、その身長データを $X_1, X_2, \cdots, X_n$ とする。X_k は k 番目に選んだ男子の身長である。このとき X_k は

$$P(X_k = y_i) = \frac{1}{N} \qquad \begin{pmatrix} i = 1, 2, \cdots, N \\ k = 1, 2, \cdots, n \end{pmatrix}$$

である確率変数とみなせる。この標本 $X_1, X_2, \cdots, X_n$ の平均と分散を

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \quad (\text{標本平均}) \quad , \quad S^2 = \frac{1}{n}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2 \quad (\text{標本分散})$$

標本平均、**標本分散**という。このとき次の定理が成り立つ。

定理 14 (1) 各 $k = 1, 2, \cdots, n$ に対し、確率変数 X_k の平均と分散は母平均と母分散に一致する。すなわち

$E[X_k] = \mu$, $E\left[(X_k - \mu)^2\right] = \sigma^2$

(2) $E\left[\,\overline{X}\,\right] = \mu$, $E\left[\sum_{k=1}^{n}(X_k - \overline{X})^2\right] = (n-1)\sigma^2$

(注 1) この定理より標本分散の平均値は $E\left[S^2\right] = \dfrac{n-1}{n}\sigma^2$ となり母分散 σ^2 と異なる。そのため標本分散のかわりに

$$U = \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2 \qquad (\text{不偏分散})$$

を用いる。この平均は $E[U] = \sigma^2$ となる。U を**不偏分散**という。

(注 2) 統計では未知数の数を**自由度** (***degree of freedom***) という。
例えば n 個の未知数 $\{x_1,\ x_2,\ \cdots,\ x_n\}$ の自由度は n だが,
その平均 $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ を引いた値

$$\{x_1 - \overline{x},\ x_2 - \overline{x},\ \cdots,\ x_n - \overline{x}\}$$

の自由度は $n-1$ である。

問 $y_k = x_k - \overline{x}$ $\left(\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i\right)$ $(k = 1,\ 2,\ \cdots,\ n)$ とおくとき，y_n の値は $\{y_1,\ y_2,\ \cdots,\ y_{n-1}\}$ で表されることを示せ。