2003 年 05 月 09 日

共分散分析のしくみ

文責：尾崎寿昭、稲葉太一（神戸大学）

0. 問題内容

２変数の直線関係を見る際に、これらの関係に影響を及ぼす第３の変数がある場合、これを共変量として解析する方法として、共分散分析がある。

このレポートでは、これらが、最小二乗法によって求められることと、その際に導出される平方和で F 検定できることを示すのが目的である。

今回は、最小二乗解を求めるための前提となる「データの構造式」をダミー変数を用いて表現し、これを解くことが目的である。

1. 具体的状況

数値例として、「毒性・薬効データの統計解析 p.106」の表 3-3 にある「肝重量と体重の母獣単位の平均値」を紹介する。

対照群と処理群の違いを検定したいとき、「肝重量と体重」の関係を考慮に入れて解析するときに、体重を共変量として薬物の影響を見たいとする。

2. 最小二乗法

群番号を表すために、次のダミー変数を考える。

$$z_1 = 0, \quad z_2 = 1$$

このダミー変数を用いて、この問題を定式化すると以下のようになる。データの構造式は

$$y_{ij} = z_i\alpha_0 + \alpha_1 + \beta(x_{ij} - \overline{x}_{i\cdot}) + \varepsilon_{ij}$$

である。ここで、α_1 は第２群（処理群）の母平均であり、α_0 は、２群の母平均の違いである。β は、２つの群に共通の傾き、即ち、肝重量と体重の関係を表す。このとき、予測値 $\widehat{y}_{ij}$ は、

$$\widehat{y}_{ij} = \widehat{\alpha}_0 z_i + \widehat{\alpha}_1 + \widehat{\beta}(x_{ij} - \overline{x}_{i\cdot})$$

となる。したがって、残差 $e_{ij} = y_{ij} - \widehat{y}_{ij}$ は

$$e_{ij} = y_{ij} - \widehat{y}_{ij} = y_{ij} - \{\widehat{\alpha}_0 z_i + \widehat{\alpha}_1 + \widehat{\beta}(x_{ij} - \overline{x}_{i\cdot})\}$$

となる。最小二乗法は、残差平方和を最小にするのであるから、次の量が最小化される対象である。

$$Q = \sum_{i=1}^{2}\sum_{j=1}^{n_i} e_{ij}^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_i}[y_{ij} - \{\widehat{\alpha}_0 z_i + \widehat{\alpha}_1 + \widehat{\beta}(x_{ij} - \overline{x}_{i\cdot})\}]^2$$

3. 正規方程式とその解

残差平方和を、偏微分することで、正規方程式を作る。

$$\frac{\partial}{\partial \widehat{\alpha}_0} Q = -2\sum_{i=1}^{2}\sum_{j=1}^{n_i} z_i [y_{ij} - \{\widehat{\alpha}_0 z_i + \widehat{\alpha}_1 + \widehat{\beta}(x_{ij} - \overline{x}_i.)\}] = 0 \quad \cdots\cdots \tag{3.1}$$

$$\frac{\partial}{\partial \widehat{\alpha}_1} Q = -2\sum_{i=1}^{2}\sum_{j=1}^{n_i} [y_{ij} - \{\widehat{\alpha}_0 z_i + \widehat{\alpha}_1 + \widehat{\beta}(x_{ij} - \overline{x}_i.)\}] = 0 \quad \cdots\cdots \tag{3.2}$$

$$\frac{\partial}{\partial \widehat{\beta}} Q = -2\sum_{i=1}^{2}\sum_{j=1}^{n_i} (x_{ij} - \overline{x}_i.)[y_{ij} - \{\widehat{\alpha}_0 z_i + \widehat{\alpha}_1 + \widehat{\beta}(x_{ij} - \overline{x}_i.)\}] = 0 \quad \cdots\cdots \tag{3.3}$$

まず、(3.1) より、$z_1 = 0,\ z_2 = 1$ を考慮すると、

$$\begin{aligned} & z_1 \sum_{j=1}^{n_1} [y_{1j} - \{\widehat{\alpha}_0 z_1 + \widehat{\alpha}_1 + \widehat{\beta}(x_{1j} - \overline{x}_1.)\}] \\ & \qquad + z_2 \sum_{j=1}^{n_2} [y_{2j} - \{\widehat{\alpha}_0 z_2 + \widehat{\alpha}_1 + \widehat{\beta}(x_{2j} - \overline{x}_2.)\}] = 0 \quad \cdots\cdots \end{aligned} \tag{3.4}$$

よって、以下を得る。

$$\sum_{j=1}^{n_2} [y_{2j} - \{\widehat{\alpha}_0 + \widehat{\alpha}_1 + \widehat{\beta}(x_{2j} - \overline{x}_2.)\}] = 0 \quad \cdots\cdots \tag{3.4$'$}$$

(3.4)' と (3.2) より

$$\sum_{j=1}^{n_1} [y_{1j} - \{\widehat{\alpha}_1 + \widehat{\beta}(x_{1j} - \overline{x}_1.)\}] = 0 \quad \cdots\cdots \tag{3.5}$$

また、(3.4)' と $\displaystyle\sum_{j=1}^{n_2}(x_{2j} - \overline{x}_2.) = 0$ より

$$\sum_{j=1}^{n_2} y_{2j} = n_2(\widehat{\alpha}_0 + \widehat{\alpha}_1)$$

よって、

$$\overline{y}_2 = \widehat{\alpha}_0 + \widehat{\alpha}_1 \quad \cdots\cdots \tag{3.4$''$}$$

同様に、(3.5) より

$$\overline{y}_1 = \widehat{\alpha}_1 \quad \cdots\cdots \tag{3.5$'$}$$

次に、$\displaystyle\sum_{j=1}^{n_j}(x_{ij} - \overline{x}_i.) = 0$ であるから (3.3) より

$$\begin{aligned} \sum_{i=1}^{2}\sum_{j=1}^{n_j}(x_{ij} - \overline{x}_i.) y_{ij} &= \sum_{i=1}^{2}\sum_{j=1}^{n_j}(x_{ij} - \overline{x}_i.)\{\widehat{\alpha}_0 z_i + \widehat{\alpha}_1 + \widehat{\beta}(x_{ij} - \overline{x}_i.)\} \\ &= \widehat{\beta}\sum_{i=1}^{2}\sum_{j=1}^{n_j}(x_{ij} - \overline{x}_i.)^2 = \widehat{\beta}(S_{xx_1} + S_{xx_2}) \quad \cdots\cdots \end{aligned} \tag{3.6}$$

これらを解くと、以下が導かれる。

$$\widehat{\alpha}_0 = \overline{y}_1 - \overline{y}_2, \quad \widehat{\alpha}_1 = \overline{y}_1$$

$$\widehat{\beta} = \frac{S_{xy_1} + S_{xy_2}}{S_{xx_1} + S_{xx_2}}$$ （これは (3.6) より分かる）