共分散分析

- 目的

  1 個の共変量で調整した各群の平均値に差があるかどうか検定する。

- 解析理論

  独立変数（共変量）を X，従属変数を Y とする。k 群において，各群のケース数を n_i，全ケース数を $n = \sum n_i$，また，全体の平均値を $\bar{X}$，$\bar{Y}$，各群の平均値を $\bar{X}_j$，$\bar{Y}_j$ として，$SS_{t(Y)}$，$SS_{t(X)}$ を変数 Y，X の全体の平方和（変動 sum of squares），SP を全体の積和（共変動 sum of products）としたとき，共分散分析では，全体の平方和と積和は 2 つに分割される。

$$
\begin{aligned}
SS_{t(Y)} &= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y})^2 \\
&= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_j)^2 + \sum_{j=1}^{k} n_j(\bar{Y}_j-\bar{Y})^2 \\
&= SS_{w(Y)} + SS_{b(Y)} \\
&= Y \text{ の級内平方和} + Y \text{ の級間平方和} \qquad (1)
\end{aligned}
$$

$$
\begin{aligned}
SS_{t(X)} &= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij}-\bar{X})^2 \\
&= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij}-\bar{X}_j)^2 + \sum_{j=1}^{k} n_j(\bar{X}_j-\bar{X})^2 \\
&= SS_{w(X)} + SS_{b(X)} \\
&= X \text{ の級内平方和} + X \text{ の級間平方和} \qquad (2)
\end{aligned}
$$

$$
\begin{aligned}
SP_t &= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij}-\bar{X})(Y_{ij}-\bar{Y}) \\
&= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij}-\bar{X}_j)(Y_{ij}-\bar{Y}_j) + \sum_{j=1}^{k} n_j(\bar{X}_j-\bar{X})(\bar{Y}_j-\bar{Y}) \\
&= SP_w + SP_b \\
&= \text{級内積和} + \text{級間積和} \qquad (3)
\end{aligned}
$$

  ここで，独立変数で調整された従属変数の推定誤差平方和を（4），（5），（6），（7）式のように定義する。

  - 全データに基づく回帰直線からの変動

$$
\left\{
\begin{aligned}
& SS'_{t(Y)} = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\hat{Y}^t_{ij})^2 = SS_{t(Y)} - \frac{SP_t^2}{SS_{t(X)}} \\
& \quad \hat{Y}^t_{ij} = a_t X_{ij} + b_t \\
& \quad a_t = SP_t \,/\, SS_{t(X)} \\
& \quad b_t = \bar{Y} - a_t \bar{X}
\end{aligned}
\right. \qquad (4)
$$

  - 全群に共通な傾きと各群ごとの切片を持つ回帰直線からの変動

$$
\left\{
\begin{aligned}
& SS'_{w(Y)} = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\hat{Y}^w_{ij})^2 = SS_{w(Y)} - \frac{SP_w^2}{SS_{w(X)}} \\
& \quad \hat{Y}^w_{ij} = a_w X_{ij} + b_{wj} \\
& \quad a_w = SP_w \,/\, SS_{w(X)} \\
& \quad b_{wj} = \bar{Y}_j - a_w \bar{X}_j
\end{aligned}
\right. \qquad (5)
$$

– 各群ごとのデータに基づく回帰直線からの変動

$$
\begin{cases}
SS'_{w(Y)j} = \sum_{i=1}^{n_j} (Y_{ij} - \hat{Y}_{ij})^2, \qquad (j = 1,\ 2,\ \ldots,\ k) \\
\hat{Y}_{ij} = a_j X_{ij} + b_j \\
a_j = \dfrac{\sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)(Y_{ij} - \bar{Y}_j)}{\sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2} \\
b_j = \bar{Y}_j - a_j \bar{X}_j
\end{cases}
\tag{6}
$$

– 各群の回帰の差に起因する変動

$$
SS'_{b(Y)} = SS'_{t(Y)} - SS'_{w(Y)} \tag{7}
$$

共分散分析を行う前提としては，各群において直線回帰式の傾きが同じでなければならないので，まず回帰の同質性について検定しなければならない。もしこの検定で回帰が同質とみなせるという結論が得られれば，共分散分析が行われる。

1. 回帰の同質性の検定

帰無仮説 H_0: 「各群の回帰係数が等しい」。

対立仮説 H_1: 「各群の回帰係数は異なる」。

有意水準 α で両側検定を行う（片側検定は定義できない）。

群間の推定平方和（平均回帰に基づく推定誤差）は，各群の推定誤差の和と各群の回帰の差に分解できるので，表 1 のような分散分析表を作成する。

表 1　回帰の同質性の検定のための分散分析表

要因	推定誤差平方和	自由度	推定誤差分散	F 値
各群の回帰の差	$SS'_{w(Y)} - \sum_{j=1}^{k} SS'_{w(Y)j}$	$k-1$	MS_r	MSr / MSe
各群の推定誤差の和	$\sum_{j=1}^{k} SS'_{w(Y)j}$	$n-2k$	MS_e	
平均回帰に基づく推定誤差	$SS'_{w(Y)}$	$n-k-1$	MS_w	

MS_r，MS_e，MS_w はそれぞれ推定誤差平方和を自由度で割ったもの

F 値は，第 1 自由度が $k-1$，第 2 自由度が $n-2k$ の F 分布に従う。有意確率を P_0 とすると，

– $P_0 > \alpha$ のとき，帰無仮説を採択する。「各群の回帰係数（傾き）が異なるとはいえない」。

– $P_0 \leq \alpha$ のとき，帰無仮説を棄却する。「各群の回帰係数（傾き）が異なる」。

帰無仮説が採択された場合には，（5）式の a_w が母回帰係数の推定値として得られる。

2. 要因効果の検定（共分散分析）

帰無仮説 H_0: 「独立変数で調整した従属変数の平均値に差がない」。

対立仮説 H_1: 「独立変数で調整した従属変数の平均値に差がある」。

有意水準 α で両側検定を行う（片側検定は定義できない）。

もし前項の検定で帰無仮説が採択されたならば，母回帰係数の推定値 (a_w) を用いて独立変数の影響を取り除いた要因効果の検定に進むことができる。

表 2 のような分散分析表を作る。

F 値は，第 1 自由度が $k-1$，第 2 自由度が $n-k-1$ の F 分布に従う。有意確率を P_0 とすると，

– $P_0 > \alpha$ のとき，帰無仮説を採択する。「独立変数によって調整された各群の従属変数の平均値（調整済み平均値）に差があるとはいえない」。

– $P_0 \leq \alpha$ のとき，帰無仮説を棄却する。「調整済み平均値に差がある」。

表 2　要因効果の検定のための分散分析表

要因	$SS'_{(Y)}$	自由度	$MS'_{(Y)}$	F 値
級間 (効果，群)	$SS'_{b(Y)}$	$k-1$	$MS'_{b(Y)}$	$MS'_{b(Y)} / MS'_{w(Y)}$
級内 (誤差)	$SS'_{w(Y)}$	$n-k-1$	$MS'_{w(Y)}$	
全体	$SS'_{t(Y)}$	$n-2$	$MS'_{t(Y)}$	

$MS'_{b(Y)}$，$MS'_{w(Y)}$，$MS'_{t(Y)}$ は $SS'_{b(Y)}$，$SS'_{w(Y)}$，$SS'_{t(Y)}$ を自由度で割ったもの