Обобщенные обратные матрицы в однофакторном дисперсионном анализе

Белоусов Ю.С., студент, СПбГУ, bus99@ya.ru

Алексеева Н.П., к.ф.-м.н., доцент, СПбГУ, ninaalexeyeva@mail.ru

Аннотация

Рассматривается задача оценки параметров в модели однофакторного дисперсионного анализа. Известные методы решения позволяют построить несмещенную оценку с минимальной возможной дисперсией. Однако, если ослабить требование минимальности дисперсии, то можно получать оценки, удовлетворяющие дополнительным свойствам (например, меньшие корреляции). В данной работе предлагается метод оценивания параметров, основанный на обобщенно обратных матрицах, практический способ для их вычисления в этом случае. Кроме того, демонстрируется использование обобщенных обратных для достижения дополнительных свойств оценок.

Введение

Рассмотрим классическую задача однофакторного дисперсионного анализа [1]

$$\begin{cases} y_{ij} = \mu + \beta_i + \varepsilon_{ij} & i = 1, \ldots, r;\ j = 1, \ldots, n_i \\ \{\varepsilon_{ij}\} & \text{независимы и распределены } N(0, \sigma^2). \end{cases}$$

В матричном виде она выглядит как

$$Y = X\beta + \varepsilon, \tag{1}$$

где Y — вектор наблюдений, X — матрица плана, B — вектор параметров, а ε — матрица ошибок. Причем,

$$Y = \begin{pmatrix} y_{11} \\ \dots \\ y_{1n_1} \\ y_{21} \\ \dots \\ y_{2n_2} \\ y_{r1} \\ \dots \\ y_{rn_r} \end{pmatrix},\ X = \left(\begin{array}{ccccc} 1 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & 0 & \dots & 0 \\ \hline 1 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 1 & \dots & 0 \\ \hline \dots & \dots & \dots & \dots & \dots \\ \hline 1 & 0 & 0 & \dots & 1 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & \dots & 1 \end{array} \right),\ \beta = \begin{pmatrix} \mu \\ \beta_1 \\ \dots \\ \beta_r \end{pmatrix}.$$

Задача состоит в оценке вектора параметров. Естественно искать ее методом наименьших квадратов, т.е. как решение задачи

$$||Y - X\beta|| \to \min_{\beta}. \tag{2}$$

Откуда получается, что оценкой является решение системы $X^{\mathrm{T}}X\beta = X^{\mathrm{T}}Y$. Из необратимости матрицы $X^{\mathrm{T}}X$ следует, что решение не единственное. Однако, если рассматривать минор матрицы X без последнего столбца (иначе говоря, для последней группы не вводить дополнительную градацию β_r) и решать аналогичное уравнение, то оно имеет единственное решение. Кроме того, эта оценка будет несмещенной и с минимальной дисперсией[2], в следующем смысле: обозначив это решение за $\hat{\beta}$, а соответствующую ему ковариационную матрицу за $\Sigma_{\hat{\beta}}$, для любой другой несмещенной оценки $\tilde{\beta}$ матрица $\Sigma_{\tilde{\beta}} - \Sigma_{\hat{\beta}}$ положительно определена.

Цель данной работы состоит в исследовании остальных решений (2), с учетом введения последней градации.

Теоретические сведения и результаты

Наложение дополнительных линейных ограничений

Предположение, описанное выше, является частным случаем следующей идеи: от решения (2) также требуется выполнение условия $H\beta = Y_2$, где $H \in \mathbb{R}^{p \times t}$, а $Y_2 \in \mathbb{R}^p$. Известна следующая

Теорема. *[1] Пусть имеется совместная система $Xb = z$, где $X \in \mathbb{R}_r^{p\times n}, b \in \mathbb{R}^p, z \in \mathbb{R}^n$. Кроме того, дана матрица $H \in \mathbb{R}^{p\times t}$, $t \le p - r$. Тогда имеется единственное решение b_0 системы уравнений*

$$\begin{cases} Xb = z \\ Hb = 0 \end{cases}$$

в том и только в том случае, когда выполняются два условия:

1. *Ранг составной матрицы $\begin{pmatrix} X \\ H \end{pmatrix}$ равен p.*

2. *Никакая линейная комбинация строк H (кроме 0) не представляется в виде линейной комбинации строк X.*

Конкретно, для случая, описанного выше имеем линейное ограничение следующего вида

$$\begin{pmatrix} 0 & \dots & 0 & 1 \\ 0 & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 0 \end{pmatrix} \beta = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Однако, как будет показано далее, этот подход так же может быть выражен в терминах обобщенных обратных.

Обобщенно обратные матрицы

Приведем необходимые теоретические сведения. Пусть $A \in \mathbb{R}^{r\times n}$ и $A^- \in \mathbb{R}^{n\times r}$. Рассмотрим 4 соотношения:

$$\begin{array}{rccc} O1: & AA^-A & = & A, \\ O2: & A^-AA^- & = & A^-, \\ O3: & (AA^-)^* & = & AA^-, \\ O4: & (A^-A)^* & = & A^-A, \end{array}$$

где A^* — сопряженная матрица.

Определение. $A^{(i,j,\dots,k)}$ называться $\{(i),(j),\dots,(k)\}$-обратной к A, если она удовлетворяет соотношениям $(i),(j),\dots,(k)$ из $O1 - O4$. $\{1\}$-обратная к A так же будет называться обобщенной обратной.

Так же приведем здесь без доказательства теорему, являющуюся основой построения оценок в модели (2) через обобщенный обратные матрицы.

***Теорема* 1.** *В задаче* $||Ax-b|| \to \min$, *где* $A \in \mathbb{R}^{k\times n}$ *и* $b \in \mathbb{R}^k$ *и норма евклидова, минимум достигается при* $x = A^{(1,3)}b$. *Обратно, если матрица* $X \in \mathbb{R}^{n\times k}$ *обладает свойством, что для любого* b *значение* $||Ax-b||$ *минимально при* $x = Xb$, *то* $X \in A\{1,3\}$.

Укажем теперь на связь построения оценок через введение линейных ограничений, и через обобщенные обратные. Пусть есть задача

$$\begin{aligned} &||X\beta - Y|| \to \min \\ &H\beta = Y_2. \end{aligned}$$

Тогда, положив $\overline{H} = X(\mathbb{1} - H^{(1)}H)$, $\overline{Y} = Y - XH^{(1)}Y_2$ и, взяв любое z, оценка получается как

$$\hat{\beta} = H^{(1)}Y_2 + \left(\mathbb{1} - H^{(1)}H\right)\left(\overline{H}^{(1,3)}\overline{Y} + (\mathbb{1} - \overline{H}^{(1,3)}\overline{H})z\right).$$

В случае, когда линейные ограничения обеспечивают единственность решения, можно положить $z = 0$ и все матрицы брать псевдообратными. Стандартный случай здесь примет следующий вид: $\hat{\beta} = (\mathbb{1} - H^+H)((X(\mathbb{1} - H^+H)^+Y)$.

Построение обобщенно обратных

Построение обобщенных обратных матриц может осуществляться, например, с помощью матричной параметризации, предложенной в [3]. Она строится следующим образом. Пусть $\mathbb{N}_n$ — начальный отрезок натурального ряда длины n. Обозначим $\nu(l|n) = \{(\nu_1, \nu_2, \ldots, \nu_l) \mid \nu_1 < \nu_2 < \cdots < \nu_l;\ \nu_i \in \mathbb{N}_n\}$. Матрица, составленная из строк ($\mathbb{1}_\nu(l|n)$) или из столбцов ($\mathbb{1}^\nu(l|n)$) матрицы $\mathbb{1}_n$, соответствующих множеству $\nu(l|n)$, называется ν-частичной. Кроме того $\mathbb{1}_\nu^\lambda = \mathbb{1}_\nu \mathbb{1}^\lambda$. Тогда для $A \in \mathbb{R}^{k\times n}$ ранга r, и таких $\lambda = \lambda(r|k)$, $\nu = \nu(r|n)$, что $\det A_\nu^\lambda \neq 0$, обобщенная обратная к A может быть записана в виде

$$A^- = (A_\lambda)^- A_\lambda^\nu (A^\nu)^-,$$

где

$$\begin{aligned} (A_\lambda)^- &= \mathbb{1}^\nu (A_\lambda^\nu)^{-1} + (\mathbb{1}_n - \mathbb{1}^\nu (A_\lambda^\nu)^{-1} A_\lambda) Q, \\ (A^\nu)^- &= (A_\lambda^\nu)^{-1} \mathbb{1}_\lambda + P(\mathbb{1}_k - A^\nu (A_\lambda^\nu)^{-1} \mathbb{1}_\lambda), \end{aligned}$$

а параметрами обращения являются миноры $P^{(\lambda)}$ и $Q_{(\nu)}$ матриц P и Q из $\mathbb{R}^{r\times k}$ и $\mathbb{R}^{n\times r}$ соответственно. При этом принято следующее обозначение $(\lambda) := \mathbb{N}_k \setminus \{\lambda\}$.

Дополнительные свойства оценок

Теперь формализуем задачу. В условиях модели (2), оценка вектора параметров β строится как решение задачи

$$\begin{cases} ||Y - X\beta|| \to \min \\ F(\beta) \to \min, \end{cases} \tag{3}$$

где F — некоторый функционал. Далее, используя теорему 1, первое условие переписывается как $\beta = X^{(1,3)}Y$, а вся задача принимает вид

$$F(X^{(1,3)}Y) \to \min_{X^{(1,3)} \in X\{1,3\}}. \tag{3.1}$$

Примеры

Минимальная норма решения

Если взять $F(\beta) = ||\beta||_2$, то решение может быть получено на основе известной теоремы

Теорема 2. *[4] Пусть $A \in \mathbb{R}^{k\times n}$ и $b \in \mathbb{R}^k$, тогда МНК решение уравнения $Ax = b$ $x = A^+b$ — одно из решений с минимальной нормой. Обратно, если для любого b $x = Xb$ — решение $Ax = b$ с минимальной нормой, то $X = A^+$.*

Откуда следует, что в качестве решения этой задачи нужно взять $\beta = X^+Y$.

Диагональность ковариационной матрицы

Полагая $\sigma^2 = 1$, рассмотрим следующий случай: $r = 4$, $n_1 = n_2 = n_3 = n_4 = 3$. Обозначив $cov(\beta) = \{\sigma_{ij}\}_{0\le i,j\le 4}$, определим $F(\beta) = \sum_{i\neq j}|\sigma_{ij}|$. Иначе говоря, минимизируем сумму модулей внедиагональных элементов ковариационной матрицы оценок. Сначала найдем выражение для ковариационной матрицы: $cov(\beta) = cov(X^{(1,3)}Y) = cov\left(X^{(1,3)}(X+\varepsilon)\right) = X^{(1,3)}\left(X^{(1,3)}\right)^{\mathrm{T}}$. Взяв $\nu = \{1,2,3,4\}$, $\lambda = \{1,4,7,10\}$, принимаем параметры

$$P^{(\lambda)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \end{pmatrix}, \quad Q_{(\nu)} = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix}.$$

Финально получаем

$$cov(\beta) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} \end{pmatrix}.$$

Здесь параметр $P^{(\lambda)}$ обеспечивает выполнение свойства $O3$, а параметр $Q_{(\nu)}$ — диагональность матрицы.

Если рассмотреть данную задачу для произвольных r и набора $\{n_i\}_{1\le i\le r}$, можно взять параметры обращения $\nu = \{1, 2, \ldots, r\}$, $\lambda = \{1, 1 + n_1, 1 + n_1 + n_2, \ldots, 1 + \sum_{i=1}^{r-1} n_i\}$ и

$$P^{(\lambda)} = \left(\begin{array}{ccc|c|ccc|ccc} 0 & \ldots & 0 & \ldots & 0 & \ldots & 0 & \frac{1}{n_r} & \ldots & \frac{1}{n_r} \\ \frac{1}{n_1} & \ldots & \frac{1}{n_1} & \ldots & 0 & \ldots & 0 & -\frac{1}{n_r} & \ldots & -\frac{1}{n_r} \\ 0 & \ldots & 0 & \ldots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots & \ldots & 0 & \ldots & 0 & -\frac{1}{n_r} & \ldots & -\frac{1}{n_r} \\ 0 & \ldots & 0 & \ldots & \frac{1}{n_{r-1}} & \ldots & \frac{1}{n_{r-1}} & -\frac{1}{n_r} & \ldots & -\frac{1}{n_r} \end{array}\right),$$

$$Q_{(\nu)} = \begin{pmatrix} 0 & \ldots & 0 & 1 \end{pmatrix}.$$

В i-ом блоке $P^{(\lambda)}$ ровно $n_i - 1$ столбцов.

При таких параметрах $cov(\beta) = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 \\ 0 & \frac{1}{n_1} & 0 & \ldots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \ldots & 0 & \frac{1}{n_r} \end{pmatrix}.$

Заключение

В работе был представлен метод построения оценок через обобщенные обратные матрицы с учетом дополнительных свойств, показан метод построения обобщенных обратных, а также приведены примеры для демонстрации этого метода.

Литература

[1] Г. Шеффе. Дисперсионный анализ. Москва «Наука», 1980.

[2] Е.З. Демиденко Линейная и нелинейная регрессия. Москва «Финансы и статистика», 1981.

[3] А. Г. Барт. Анализ медико-биологических систем. Издательство Санкт-Петербургского университета, 2003.

[4] R. Penrose. On best approximate solutions of linear matrix equations. Mathematical Proceedings of the Cambridge Philosophical Society, 1956.