Линеен регресионен модел. Метод на най-малките квадрати. Теорема на Гаус-Марков.

П.Матеев

София, 2012

Формулира се задачата и модела на линейна регресия в общия случай или в случая с проста линейна регресия. Показва се, че при нормално разпределение на грешките в модела, максимално правдоподобните оценки съвпадат с решението на нормалната система уравнения от „метод на най-малките квадрати“. Показва се неизместеност и се извежда дисперсията на оценките.

Доказва се теоремата на Гаус – Марков, че оценката на параметрите, получени като решение на нормалната система уравнения са с минимална дисперсия (относно всички неизместени оценки).

Съдържание

1 Линеен регресионен модел.

1.1 Регресионен експеримент.

Данните от проведено изследване или експеримент са наблюдения на

- отклик y и
- предиктори $\mathbf{x} = (x(1), x(2), \ldots, x(m))'$.

Предикторите описват условията, при които се провежда отделно наблюдение. Задачата е да предсказваме резултата от наблюдение на количествена променлива, наричана отклик, при известни условия на наблюдение. С други думи да установим зависимост на отклика от конкретните стойности на предикторите.

1.2 Регресионен модел.

Регресионен модел наричаме случайна величина Y, чието разпределение зависи от предикторите, съответства на променливата отклик и удовлетворява условията:

- $\mathbb{E}\,(Y \mid \mathbf{x}) = f(\mathbf{x})$ - математичското очакване на Y е функция на предикторите, $f(\mathbf{x})$ наричаме *регресионна функция*;
- $\mathbb{V}ar\,Y = \sigma^2$ - дисперсията на Y е константа (не зависи от $\mathbf{x}$).

Разликата $e = Y - f(\mathbf{x})$ е случайна величина и се интерпретира като *случайна грешка* при наблюдаване на отклика, чиято "истинска" стойност е $f(\mathbf{x})$. Разпределението на грешката е с очакване $\mathbb{E}\,e = 0$ и дисперсия $\mathbb{V}ar\,e = \sigma^2$.

Регресионният модел се представя чрез *регресионно уравнение*

$$Y = f(\mathbf{x}) + e, \quad (\mathbb{E}\,e = 0, \quad \mathbb{V}ar\,e = \sigma^2).$$

1.3 Линеен регресионен модел.

Регресионен модел е *линеен*, когато функцията $f(\mathbf{x})$ е зададена с точност до неизвестен параметър $\mathbf{b}' = (b_1, \ldots, b_p)$ и зависи от този параметър линейно

$$f(\mathbf{x}) = b_1 f_1(\mathbf{x}) + \ldots + b_p f_p(\mathbf{x}) = \sum_{j=1}^{p} b_j f_j(\mathbf{x}).$$

Без загуба на общност, можем да считаме, че условията се описват от $f_j(\mathbf{x})$ или нови променливи, получени чрез трансформация на оригиналните предиктори. За краткост пишем,

$$f(\mathbf{x}) = b_1 x(1) + \ldots + b_p x(p) = \sum_{j=1}^{p} x(j) b_j = \mathbf{x}'\mathbf{b}.$$

Линеен регресионен модел се представя чрез линейно регресионно уравнение

$$Y = b_1 x(1) + \ldots + b_p x(p) + e, \quad (\mathbb{E}\,e = 0, \quad \mathbb{V}ar\,e = \sigma^2),$$

линейно спрямо неизвестните параметри b_j!

2 Метод на най-малките квадрати.

2.1 Данни от регресионен експеримент.

Нека регресионен експеримент е проведен n пъти при различни условия $\mathbf{x_i} = (x_i(1), \ldots, x_i(p))', \ i = 1, \ldots, n$, и наблюдаваните стойности на отклика са били $y_i, \ i = 1, \ldots, n$.

Наблюденията удовлетворяват n регресионни уравнения

$$y_i = b_1 x_i(1) + \ldots + b_p x_i(p) + e_i, \quad i = 1, \ldots, n.$$

Данните записваме в матричен вид, като наблюденията на всяка променлива е n-мерен вектор

- $\mathbf{y} = (y_1, \ldots, y_n)'$ са наблюденията на отклика, а
- $\mathbf{x}(j) = (x_1(j), \ldots, x_n(j))'$ са наблюденията на j -тия от предикторите $(j = 1, \ldots, p)$.

Матрица на експеримента наричаме матрицата, чиито стълбове са наблюденията на p-те предиктора, а редовете са транспонирани векторите на условията на наблюдение $\mathbf{x_i}' = (x_i(1), \ldots, x_i(p)) \ \ (i = 1, \ldots, n)$:

$$\mathbf{X}_{(n \times p)} = \Big(\mathbf{x}(1) \ldots \mathbf{x}(p)\Big) = \begin{pmatrix} \mathbf{x}'_1 \\ \cdots \\ \mathbf{x}'_n \end{pmatrix}.$$

Регресионните уравнения за n-те наблюдения в матричен запис са:

$$\mathbf{y}_{(n \times 1)} = \mathbf{X}_{(n \times p)}.\mathbf{b}_{(p \times 1)} + \mathbf{e}_{(n \times 1)} \quad \text{или накратко} \quad \mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}.$$

Сега e_i е неизвестната "грешка" при $i-$тото наблюдение и $\mathbf{e}' = (e_1, e_2, \ldots, e_n)$.

В системата от n уравнения неизвестните са $p+n$ на брой – параметрите $\mathbf{b}$ и грешките $\mathbf{e}$.

Задачата е да оценим $\mathbf{b}$ като наложим подходящи и разумни ограничения за грешките $\mathbf{e}$.

2.2 Метод на най-малките квадрати.

Карл Фридрих Гаус mостулира *метод на най-малките квадрати* като определя условие за грешките

> Неизвестните параметри се определят така,
> че сумата от квадратите на грешките да е минимална!

Разликите $y_i - \sum_{j=1}^{p} x_i(j) b_j = e_i, \ i = 1, \ldots, n$, наричаме още *остатъци (residuals)* и по традиция сумата от квадратите им (Sum of Squares) означаваме с

$$SS(\mathbf{b}) = \sum_{i=1}^{n} e_i^2 = \mathbf{e}'\mathbf{e}.$$

Теорема 1 *Ако* $\hat{\mathbf{b}}$ *е решение на системата*

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y},$$

наричана **системата от нормални уравнения**, *то за* $\hat{\mathbf{b}}$ *се достига минимума на сумата от квадрати* $SS(\mathbf{b})$*:*

$$SS(\mathbf{b}) \geq SS(\hat{\mathbf{b}}).$$

Забележка 1. Нормалната система уравнения е друга форма на запис на условието за некорелираност на $\mathbf{e}$ и $\mathbf{x}(j)$, $j = 1, \ldots, p$, т.е.

$$\mathbf{X}'\mathbf{e} = \mathbf{X}'(\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{0}.$$

Забележка 2. Нормалната система уравнения е еквивалентна на анулиране на първите производни на $SS(\mathbf{b})$ по параметъра $\mathbf{b}$:

$$\begin{aligned}
\frac{\partial SS}{\partial \mathbf{b}} &= \frac{\partial \mathbf{e}'\mathbf{e}}{\partial \mathbf{b}} = \frac{\partial (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})}{\partial \mathbf{b}} = \\
&= \frac{\partial (\mathbf{y}'\mathbf{y} - \mathbf{2}\mathbf{b}\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b})}{\partial \mathbf{b}} = \\
&= -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b}.
\end{aligned}$$

Директно доказателство следва от равенствата

$$\begin{aligned}
SS(\mathbf{b}) &= (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) = \\
&= (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}} + \mathbf{X}\hat{\mathbf{b}} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}} + \mathbf{X}\hat{\mathbf{b}} - \mathbf{X}\mathbf{b}) = \\
&= (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) + (\mathbf{X}\hat{\mathbf{b}} - \mathbf{X}\mathbf{b})'(\mathbf{X}\hat{\mathbf{b}} - \mathbf{X}\mathbf{b}) + \\
&\quad +2(\mathbf{X}\hat{\mathbf{b}} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \\
&= SS(\hat{\mathbf{b}}) + (\mathbf{X}\hat{\mathbf{b}} - \mathbf{X}\mathbf{b})'(\mathbf{X}\hat{\mathbf{b}} - \mathbf{X}\mathbf{b}),
\end{aligned}$$

тъй като

$$(\mathbf{X}\hat{\mathbf{b}} - \mathbf{X}\mathbf{b})'(\mathbf{X}\hat{\mathbf{b}} - \mathbf{X}\mathbf{b}) \geq 0,$$

а

$$\begin{aligned}
(\mathbf{X}\hat{\mathbf{b}} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) &= (\hat{\mathbf{b}} - \mathbf{b})'\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \\
&= (\hat{\mathbf{b}} - \mathbf{b})'(\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\hat{\mathbf{b}}) = 0.
\end{aligned}$$

Нормалната система уравнения има единствено решение, ако ранга на $\mathbf{X}$ е равен на p, т.е. стълбовете $\mathbf{x}(\mathbf{j})$ са линейно независими. Решението на системата е търсената оценка:

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Веднага можем да получим и оценки за математическите очаквания на отклика за всяко едно наблюдение $\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}}$. Разликите между съответните наблюдавани стойности и оценките на средните са оценки за грешките на наблюденита (на остатъците или residuals) при така избрания модел с така оценени параметри. Ще ги означим с r_i, $i = 1, \ldots, n$ или

$$\mathbf{r} \stackrel{def}{=} \mathbf{y} - \hat{\mathbf{y}}.$$

Освен това

$$SS(\hat{\mathbf{b}}) = \mathbf{r}'\mathbf{r}.$$

2.3 Нормално разпределение на грешките.

Нека за грешките предположим, че са независими и нормално разпределени

$$e_i \sim N(0, \sigma^2),\ i = 1, \ldots, n,\ \text{или } \mathbf{e} \sim \mathbf{N}(\mathbf{0}, \sigma^\mathbf{2}.\mathbf{I}).$$

Тогава нормално е разпределението и на наблюденията $\mathbf{y} \sim \mathbf{N}(\mathbf{Xb}, \sigma^2\mathbf{I})$ и оценките, получени по МНК, съвпадат с оценките получени по метода на максималното правдоподобие и притежават всички техни оптимални свойства.

Системата има единствено решение, ако ранга на $\mathbf{X}$ е равен на p, т.е. стълбовете $\mathbf{x(j)}$ са линейно независими. Решението на системата е търсената оценка:

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Веднага можем да получим и оценки за математическите очаквания на отклика за всяко едно наблюдение $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{b}}$. Разликите между съответните наблюдавани стойности и оценките на средните им наричаме *остатъци* (*Residuals*). Те са оценки за грешките на наблюденита при така избрания модел с така оценени параметри. Ще ги означим с r_i, $i = 1, \ldots, n$ или

$$\begin{aligned}\mathbf{r} &\stackrel{def}{=} \mathbf{y} - \hat{\mathbf{y}} = \\ &= \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}.\end{aligned}$$

3 Теорема на Гаус-Марков

3.1 Твърдение

Теорема 2 (Гаус-Марков) *Нека неизвестните грешки се подчиняват на условията:*

- $\mathbb{E}\, e_i = 0, \quad i = 1, \ldots, n;$
- $\mathbb{V}ar\, e_i = \sigma^2 > 0, \quad i = 1, \ldots, n;$
- $\mathrm{cov}(e_i, e_j) = 0, \quad i \neq j.$

и нека $\hat{\mathbf{b}}$ е решение на системата нормални уравнения

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}.$$

Тогава $\hat{\mathbf{b}}$ е BLUE (Best Linear Unbeased Estimator):

1. *линейна по случайния вектор $\mathbf{y}$: $\hat{\mathbf{b}} = \mathbf{C}\mathbf{y}$ и*
2. *неизместена $\mathbb{E}\,\hat{\mathbf{b}} = \mathbf{b}$,*
3. *с минимална дисперсия – за всяка друга линейна по y неизместена оценка $\tilde{\mathbf{b}}$ е изпълнено неравенство:*

$$\mathbb{V}ar\,\hat{\mathbf{b}} \leq \mathbb{V}ar\,\tilde{\mathbf{b}}.$$

Неравенството е в смисъл, че разликата на двете ковариационни матрици $\mathbf{D} = \mathbb{V}ar\,\tilde{\mathbf{b}} - \mathbb{V}ar\,\hat{\mathbf{b}}$ е неотрицателно определена матрица:

$$\text{за } \forall \mathbf{x} \quad \mathbf{x}'\mathbf{D}\mathbf{x} \geq 0.$$

3.2 Доказателство

Последователно доказваме трите свойства на $\hat{\mathbf{b}}$:

1. $\hat{\mathbf{b}}$ е линейна по вектора $\mathbf{y}$

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{y} = \mathbf{C}\mathbf{y} \text{ като } \mathbf{C} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}';$$

2. $\hat{\mathbf{b}}$ е неизместена оценка на $\mathbf{b}$

$$\mathbb{E}\,\hat{\mathbf{b}} = \mathbb{E}\,(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'.\mathbb{E}\,\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}.\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{b};$$

3. с минимална дисперсия:

Дисперсията на $\hat{\mathbf{b}}$ е

$$\begin{aligned}
\mathbb{V}ar\,\hat{\mathbf{b}} &= \mathbb{V}ar\,(\mathbf{C}.\mathbf{y}) = \mathbf{C}.(\mathbb{V}ar\,\mathbf{y}).\mathbf{C}' = \mathbf{C}.\sigma^2.\mathbf{I_n}.\mathbf{C}' = \\
&= \sigma^2((\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}').((\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}')' = \sigma^2.(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}} = \\
&= \sigma^2.(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}.
\end{aligned}$$

Преди да изследваме диспресията на $\tilde{\mathbf{b}}$ ще означим $\mathbf{U} = \mathbf{B} - \mathbf{C}$ и ще покажем, че $\mathbf{U}\mathbf{X} = 0$. Наистина $\mathbb{E}\,(\mathbf{U}\mathbf{y}) = \mathbb{E}\,((\mathbf{B}-\mathbf{C})\mathbf{y}) = \mathbb{E}\,(\mathbf{B}\mathbf{y}) - \mathbb{E}\,(\mathbf{C}\mathbf{y}) = \mathbf{b} - \mathbf{b} = 0$.

Освен това $\mathbb{E}\,(\mathbf{U}\mathbf{y}) = \mathbf{U}\mathbb{E}\,\mathbf{y} = \mathbf{U}.\mathbf{X}\mathbf{b}$. Следователно $\mathbf{U}\mathbf{X}\mathbf{b} = 0$ за произволни стойности на параметрите $\mathbf{b}$, което е възможно само при $\mathbf{U}\mathbf{X} = 0$. В следващото равенство използваме, че $\mathbf{U}\mathbf{X}.(\mathbf{X}'\mathbf{X})^{-\mathbf{1}} = \mathbf{U}\mathbf{C}' = 0$ и също $\mathbf{C}\mathbf{U}' = 0$.

$$\begin{aligned}
\mathbb{V}ar\,\tilde{\mathbf{b}} &= \mathbb{V}ar\,(\mathbf{B}.\mathbf{y}) = \mathbf{B}.(\mathbb{V}ar\,y).\mathbf{B}' = \mathbf{B}.(\sigma^2\mathbf{I_n}).\mathbf{B}' = \\
&= \sigma^2.(\mathbf{C}+\mathbf{U})(\mathbf{C}+\mathbf{U})' = \sigma^2.(\mathbf{C}\mathbf{C}' + \mathbf{U}\mathbf{C}' + \mathbf{C}\mathbf{U}' + \mathbf{U}\mathbf{U}') = \\
&= \mathbb{V}ar\,\hat{\mathbf{b}} + \sigma^2.\mathbf{U}\mathbf{U}'.
\end{aligned}$$

С което теоремата е доказана, защото $\mathbf{U}\mathbf{U}' = 0$ е неотрицателно определена матрица.

3.3 Коментар

Забележка 1. Теоремата е в сила дори когато матрицата $\mathbf{X}'\mathbf{X}$ е изродена и не съществува единствена обратна матрица $(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}$. Тогава в стъпките на доказателството вместо обратна матрица заместваме обобщена обратна матрица.

Забележка 2. При предположение за нормално разпределение на отклика и съответно на грешките, оценките, решение на нормалната система $\hat{\mathbf{b}}$ съвпадат с оценки получени по метода на максималното правдоподобие, достигат равенство в *Неравенството на Рао-Крамер* и следователно са *ефективни*, дисперсията им е долна граница не само за линейните, а за всички неизместени оценки.