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

* Линеен регресионен модел.

* Регресионен експеримент. - Данните са проведено изследване или експеримент са наблюдения на
  - отклик y и
  - предиктор $x = (x(1), x(2), \ldots, x(m))'$

Предикторите описват условията, при които се провежда отделно наблюдение. Задачата е да предсказване резултата от наблюдение на количествена променлива, нарича отклик при известни условия на наблюдение. Търси се да установим зависимости на отклика от контролираните стойности на предикторите.

* Регресионен модел. - Регресионен модел нарича се Y, чието разпределение зависи от предикторите, свързана посредством променливата отклик и удовлетворява условията.
  - $E(Y|x) = f(x)$ - математическо очакване на Y е ф-я на предикторите, $f(x)$ - нарича регресионна ф-я
  - $Var\ Y = \sigma^2$ - дисперсията на Y е константа (не зависи от x)

Разликата $e = Y - f(x)$ е сл.в. и се интерпретира като случайна грешка при наблюдаване на отклика, чиято „истинска" стойност е $f(x)$. Разпределението на грешката е с очакване $Ee = 0$ и $Var\ e = \sigma^2$

Регресионният модел се представя чрез регресионно у-е:

$$Y = f(x) + e \quad (Ee = 0,\ Var\ e = \sigma^2)$$

* Линеен регресионен модел - Регресионният модел е линеен, когато функцията $f(x)$ е зададена с точността до неизвестен параметър $b' = (b_1 \dots b_p)$ и зависи от този параметър линейно.

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

За краткост пишем

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

Линейния регресионен модел се представя чрез линейно регресионно уравнение.

$$y = b_1 x(1) + \dots + b_p x(p) + e \quad (Ee = 0 \quad Vare = \sigma^2),$$

линейно спрямо неизвестния параметър b_j!

* Метод на най-малките квадрати

Нека регресионния експеримент е проведен n пъти при различни условия

$x_i = (x_i(1), \dots, x_i(p))'$ $i = 1 \dots n$ и наблюдаваните стойности на отклика са били y_i, $i = 1 \dots n$.

Наблюденията удовлетворяват регресионни уравнения

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

Данните записване в матричен вид, като наблюдението на p променливи е n-мерен вектор

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

Матрицата на експерименталните данни е матрица, чиито стълбове са наблюденията на p-те предиктора, а редовете са транспонираните вектори на условията на наблюдение.

$$x_i' = (x_i(1) \dots x_i(p)) \quad i = 1 \dots n$$

$X(n \times p) = (x(1) \dots x(p)) = \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}$ Регресионните уравнения

за n-те наблюдения в матричен запис са:

$Y(n\times 1) = X(n \times p) . b(p \times 1) + e(n \times 1)$ или накратко $Y = Xb + e$

Сега e_i – неизвестната „грешка" при i-тото наблюдение

и $e' = (e_1, e_2 \dots e_n)$

В системата от n уравнения неизвестни са p+n на бр – параметрите b и грешката e.

Задачата е да оценим ~~грешката~~ b, като наложим подходящи и разумни ограничения за грешките e.

Гаус построил метод на най-малките квадрати като определя условие за грешките.

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

Разликата $y_i - \sum_{j=1}^{p} x_i(j) b_j = e_i$, $i = 1 \dots n$ наричаме още остатъци и сумата от квадратите им означ. с

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

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

$$X'Xb = X'Y$$ – наричана системата от нормални

(малко y)

уравнения, то за $\hat{b}$ се достига ~~минимална~~ минимума на сумата от квадрати SS(b):

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

Забележка 1. Нормалната система уравнения е друга форма на запис на условието за некорелираност на e и x_j, $j=1..p$, т.е.

$$x'e = x'(y-xb) = x'y - x'xb = 0$$

Забележка 2. Нормалната система уравнения е еквивалентна на анулиране на първите производни на $SS(b)$ по параметъра b.

$$\frac{\partial SS}{\partial b} = \frac{\partial e'e}{\partial b} = \frac{\partial (y-xb)'(y-xb)}{\partial b} = \frac{\partial (y'y - 2bx'y + b'x'xb}{\partial b}$$

$$= -2x'y + 2x'xb$$

Директно доказателство следва от равенствата

$$SS(b) = (y-xb)'(y-xb) = (y - x\hat{b} + x\hat{b} - xb)'(y - x\hat{b} + x\hat{b} -$$

$$= (y-x\hat{b})'(y-x\hat{b}) + (x\hat{b}-xb)'(x\hat{b}-xb) + 2(x\hat{b}-xb)'(y-x\hat{b})$$

$$= SS(\hat{b}) + (x\hat{b}-xb)'(x\hat{b}-xb) \text{ тъй като}$$

$$(x\hat{b}-xb)'(x\hat{b}-xb) \geq 0, \text{ а}$$

$$(x\hat{b}-xb)'(y-x\hat{b}) = (\hat{b}-b)'x'(y-x\hat{b}) = (\hat{b}-b)'(x'y - x'x\hat{b}) = 0$$

Нормалната система уравнения има единствено решение, ако ранга на Xe равен на p, т.е. стълбовете $x(j)$ са ЛНЗ (линейнонезависими). Решението на системата е търсената оценка: $\hat{b} = (X'X)^{-1}X'y$.

Нормално разпределение на грешките.

Нека за грешките предположим, че са независими и нормално разпределени

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

Тогава нормално е разпределението и на наблюденията $y \sim N(Xb, \sigma^2 I)$ и оценките получени по МНК съвпадат

с оценките получени по метода на максималното правдоподобие и приемливаи всички тези оптимални свойства.

Теорема на Гаус-Марков.

Нека неизвестните грешки се подчиняват на условията:

- $E\varepsilon_i = 0$, $i = 1 \dots n$;
- $Var\ \varepsilon_i = \sigma^2 > 0$, $i = 1 \dots n$
- $cov(\varepsilon_i, \varepsilon_j) = 0$ $i \neq j$ и нека $\hat{\beta}$ е решение на системата нормални уравнения

$$X'X\beta = X'y$$

Тогава $\hat{\beta}$ е BLUE:

1. Линейна по случайния вектор y: $\hat{\beta} = Cy$ и
2. Неизместена $E\hat{\beta} = \beta$
3. с минимална дисперсия - за ∀ друга линейна по y неизместена оценка $\tilde{\beta}$ е изпълнено неравенството:

$$Var\ \hat{\beta} \leq Var\ \beta$$

Неравенството е в смисъл, че разликата на двете ковариационни матрици

$\Delta = Var\ \tilde{\beta} - Var\ \hat{\beta}$ е неотрицателно определена матрица за ∀ x $x'\Delta x \geq 0$

Доказателство: Последователно доказваме трите свойства на $\hat{\beta}$:

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

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

2. $\hat{\beta}$ е неизместена оценка на β

$$E\hat{\beta} = E(X'X)^{-1}X'y = (X'X)^{-1}X'Ey = (X'X)^{-1}X'X\beta = \beta$$

3. С минимална дисперсия

Дисперсията на $\hat{b}$ е:

$$Var\,\hat{b} = Var(Cy) = C.(Var\,y).C' = C\,\sigma^2 I_n\,C' =$$

$$= \sigma^2((X'X)^{-1}X')((X'X)^{-1}X')' = \sigma^2(X'X)^{-1}X'X(X'X)^{-1} =$$

$$= \sigma^2(X'X)^{-1}$$

Преди да изследваме дисперсията на $\tilde{b}$ ще означим $U = B - C$ и ще покажем, че $UX = 0$.

$$E(Uy) = E((B-C)y) = E(By) - E(Cy) = b - b = 0$$

Освен това $E(Uy) = UEy = U.Xb \Rightarrow UXb = 0$ за произволна стойност на параметрите b, което е възможно само при $UX = 0$. В следващото равенство използваме, че $UX(X'X)^{-1} = UC' = 0$ и също $CU' = 0$

$$Var\,\tilde{b} = Var(By) = B.(Var\,y).B' = B.(\sigma^2 I_n)B' =$$

$$\sigma^2(C+U)(C+U)' = \sigma^2(CC' + UC' + CU' + UU') =$$

$$= Var\,\hat{b} + \sigma^2\,U.U'.$$ Тъй като теоремата е доказана за $UU' = 0$ е неотрицателно определена матрица.

*Коментари -

Заб 1. Теорема 1 е вярна дори когато матрицата $X'X$ е изродена и не съществува единствена обратна матрица $(X'X)^{-1}$

Заб 2. При предположение за нормално разпределение на отклоненията и съответно на грешките, оценките решени по метода на най-малките квадрати съвпадат с оценките получени по метода на максимално правдоподобие, достигат равенство в неравенството на Рао-Крамер и $\Rightarrow$ са ефективни, дисперсията е долна граница не само за линейните, а за всички несмещени оценки.