ՌԵԳՐԵՍԻՈՆ ՀԱՎԱՍԱՐՄԱՆ ԳՆԱՀԱՏՈՒՄԸ.
ՓՈՔՐԱԳՈՒՅՆ ՔԱՌԱԿՈՒՍԻՆԵՐԻ ԵՂԱՆԱԿ

Գծային ռեգրեսիան լայն կիրառություն ունի էկոնոմետրիկայում՝ պայմանավորված նրա պարամետրերի տնտեսական մեկնաբանման հստակությամբ:

Գծային ռեգրեսիոն հավասարման տեսքն է. $\hat{y} = a + bx$ կամ $y = a + bx + \varepsilon$ (1.1)

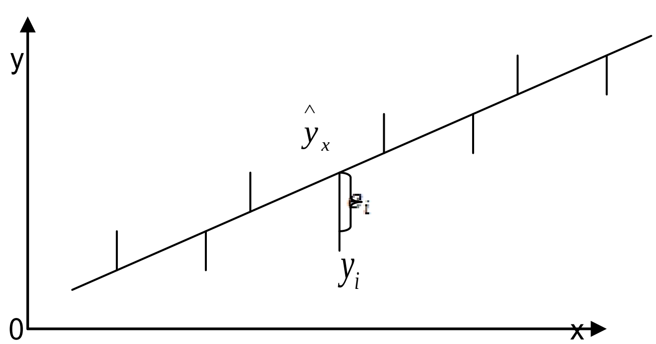
Գծային ռեգրեսիայի պարամետրերի գնահատման դասական եղանակը փոքրագույն քառակուսիների եղանակն (ՓՔԵ) է:

ՓՔԵ-ն թույլ է տալիս ստանալ a և b պարամետրերի այնպիսի արժեքներ, որոնց ժամանակ արդյունքային ցուցանիշի՝ իրական և գնահատված արժեքների շեղումների քառակուսիների գումարը ամենափոքրն է.

$$\sum_i \left(y_i - \hat{y}_i \right)^2 \to \min \qquad (1.2)$$

Այլ խոսքերով, ռեգրեսիայի բազմաթիվ գծերից ընտրվում է այն, որի դեպքում փաստացի կետերի և գնահատված գծի միջև (տե՛ս գծապատկեր) շեղումը նվազագույնն է.

$e_i = y_i - \hat{y}_i$, հետևաբար՝ $\sum e_i^2 \longrightarrow \min$



Որպեսզի գտնվի (1.2) ֆունկցիայի մինիմումը պետք է հաշվել a և b պարամետրերի նկատմամբ մասնակի ածանցյալները և հավասարեցնել 0-ի:

Նշանակենք $\sum e_i^2$-ն S-ով, այդ դեպքում.

$$S = \sum \left(y_i - \hat{y}_i \right)^2 = \sum \left(y_i - a - bx_i \right)^2 ;$$

$$\frac{\partial S}{\partial a} = -2\sum y_i + 2na + 2b\sum x_i = 0; \qquad (1.3)$$

$$\frac{\partial S}{\partial b} = -2\sum y_i x_i + 2a\sum x_i + 2b\sum x_i^{\ 2} = 0:$$

Զևափոխելով (1.5) բանաձևերը, ստանում ենք նորմալ հավասարումների հետևյալ համակարգը՝ a և b պարամետրերի գնահատման համար.
$$\begin{cases} na + b\sum x_i = \sum y_i \\ a\sum x_i + b\sum x_i^{\ 2} = \sum y_i x_i \end{cases} \qquad (1.4)$$

Լուծելով նորմալ հավասարումների (1.4) համակարգը կամ փոփոխականների հաջորդական արտաքսմամբ, կամ որոշիչների մեթոդով, գտնում ենք a և b պարամետրերի որոնվող արժեքները:

Կարելի է օգտվել հետևյալ պատրաստի բանաձևերից. $a = \bar{y} - b\bar{x}$ (1.5)

(1.5) բանաձևը ստացվել է (1.4) համակարգի առաջին հավասարումից, եթե նրա բոլոր անդամները բաժանենք n-ի վրա. $b = \dfrac{\operatorname{cov}(x, y)}{\sigma_x^{\ 2}}$,

որտեղ՝

cov(x,y) ցուցանիշների կովարիացիան է,

σ_x^2 -ն x ցուցանիշի դիսպերսիան է:

Նկատի ունենալով, որ $\operatorname{cov}(x, y) = \overline{yx} - \bar{y} \cdot \bar{x}$, իսկ $\sigma_x^2 = \overline{x^2} - \bar{x}^2$, կստանանք b պարամետրի արժեքի հաշվարկման համար հետևյալ բանաձևը. $b = \dfrac{\overline{yx} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2}$ (1.6)

b պարամետրը կոչվում է **ռեգրեսիայի գործակից**: Նրա մեծությունը **ցույց է տալիս արդյունքի միջին փոփոխությունը կախված գործոնի միավոր փոփոխությունից:**

Ռեգրեսիայի գործակցի հստակ տնտեսական մեկնաբանումը գծային ռեգրեսիայի հավասարումը դարձրել է էկոնոմետրիկական հետազոտություններում բավականաչափ տարածված: a-ն y-ի արժեքն է, երբ x=0: Եթե x գործոնը 0 չի կարող դառնալ, ապա a ազատ անդամի նման մեկնաբանությունը իմաստ չունի: a պարամետրը կարող է չունենալ տնտեսական բովանդակություն: a պարամետրի տնտեսապես մեկնաբանության փորձերը կարող են երբեմն անհեթեթության հասցնել, հատկապես, երբ a<0:

Մեկնաբանել կարելի է a պարամետրի նշանը:

Կոռելյացիայի գծային գործակից: Ռեգրեսիայի հավասարումը միշտ լրացվում է կապի սերտության ցուցանիշով: Գծային ռեգրեսիայի օգտագործման ժամանակ որպես այդպիսի ցուցանիշ հանդես է գալիս կոռելյացիայի գծային գործակիցը r_{xy}: Գոյություն ունեն կոռելյացիայի գծային գործակցի բանաձևի տարբեր ձևափոխություններ. $r_{xy} = b\frac{\sigma_x}{\sigma_y} = \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y} = \frac{\overline{yx} - \bar{y}\cdot\bar{x}}{\sigma_x \sigma_y}$: (1.7)

Կոռելյացիայի գծային գործակիցը գտնվում է հետևյալ միջակայքում՝ $-1 \le r_{xy} \le 1$: Եթե ռեգրեսիայի գործակիցը b>0, ապա $0 \le r_{xy} \le 1$ և հակառակը, b<0 դեպքում՝ $-1 \le r_{xy} \le 0$:

Պետք է հաշվի առնել, որ կոռելյացիայի գծային գործակցի մեծությունը գնահատում է դիտարկվող ցուցանիշների միջև գծային կապի սերտությունը: Այդ պատճառով կոռելյացիայի գծային գործակցի բացարձակ մեծության մոտիկությունը 0-ին դեռ չի նշանակում ցուցանիշների միջև կապի բացակայություն: Այլ մոդելի (ոչ գծային) ընտրության դեպքում ցուցանիշների միջև կապը կարող է նույնիսկ բավականաչափ սերտ լինել:

Ապրոկսիմացիայի միջին սխալ: Ինչպես հասկացացանք, արդյունքային ցուցանիշի փաստացի և գնահատված արժեքների տարբերությունը որքան լինի փոքր, մոդելը կարելի է համարել այդքանով ավելի որակյալ: Որպեսզի հետազոտությունների ժամանակ ունենանք ընդհանուր պատկերացում մոդելի հարաբերականորեն էական շեղումների մասին, կարող ենք հաշվել ապրոկսիմացիայի միջին սխալը.

$$A = \frac{1}{n}\sum\left|\frac{(y_i - \hat{y}_i)}{y_i}\right|\cdot 100 \qquad (1.8)$$