ӘОЖ 519.862.6

ЖҰПТЫҚ СЫЗЫҚТЫҚ ЕМЕС РЕГРЕССИЯ МОДЕЛЬДЕРІН ҚҰРУ

Садуақас Талапбек Ерболұлы
talapbeksaduakas554@mail.ru
7M01501 – «Математика» білім беру бағдарламасының 2 курс магистранты
Қ.Жұбанов атындағы Ақтөбе өңірлік университеті, Ақтөбе қаласы, Қазақстан Республикасы
Ғылыми жетекшісі, PhD, қауымдастырылған профессор – **Мынбаева С.Т.**

Экономикалық және әлеуметтік құбылыстардың әртүрлі техникалық мәселелерін зерттеу кезінде модельдерді құрудың математикалық әдістерін қолданбау мүмкін емес. Модельдеу компьютерлік технологияны қолдана отырып, әртүрлі процестерді басқарудың сенімді негізін құруға мүмкіндік береді.

Жоғары мектепте экономикалық және техникалық пәндерді оқу барысында қандай да бір дәрежеде элементарлық және жоғары математиканы оқу кезінде алынған білім қолданылады. Атап айтқанда, әлеуметтік және экономикалық есептердің динамикалық модельдерін құру кезінде көбінесе берілген статистикалық мәліметтерге сәйкес айнымалылар арасындағы аналитикалық байланысты табуға тура келеді, басқаша айтқанда – белгілі бір аргумент мәндерінде тек жеке мәндері белгілі болатын кейбір функцияны жуықтап табу қажет болады.

Жұптық регрессия екі айнымалы – y және x арасындағы байланысты білдіреді, яғни келесі түрдегі модель:

$$y = \hat{f}(x),$$

мұндағы y – тәуелді айнымалы (нәтижелік белгі); x – тәуелсіз немесе түсіндірме айнымалы (белгі-фактор). «$\hat{}$» белгісі x және y айнымалылары арасында қатаң функционалдық тәуелділік жоқ екенін білдіреді, сондықтан әрбір жеке жағдайда y шамасы екі қосылғыштан тұрады:

$$y = \hat{y}_x + \varepsilon,$$

мұндағы y – нәтижелік белгінің нақты мәні; $\hat{y}$ – нәтижелік белгінің регрессия теңдеуінен табылған теориялық мәні; ε – нәтижелік белгінің нақты мәнінің теориялықтан ауытқуын сипаттайтын, регрессия теңдеуінен табылған кездейсоқ шама.

Жұптық регрессияның қарапайым моделі – **сызықтық регрессияны** қарастырайық. Сызықтық регрессия оның параметрлерін нақты экономикалық түсіндіруге байланысты эконометрикада кеңінен қолданылады [1].

Сызықтық регрессия келесі түрдегі теңдеуді табуға келтіріледі:

$$\hat{y}_x = a + b \cdot x \text{ немесе } y = a + b \cdot x + \varepsilon. \qquad (1)$$

$\hat{y}_x = a + b \cdot x$ түрдегі теңдеу x факторының берілген мәндері бойынша нәтижелік белгінің теориялық мәндерін, оған x факторының нақты мәндерін қоя отырып, табуға мүмкіндік береді.

Сызықтық регрессияны құру оның a және b параметрлерін бағалауға келтіріледі. Сызықтық регрессия параметрлерін бағалаудың классикалық тәсілі ең кіші квадраттар әдісіне негізделген.

Қарапайым түрлендірулерден кейін біз a және b параметрлерін бағалау үшін келесі сызықтық теңдеулер жүйесін аламыз:

$$\begin{cases} a \cdot n + b \cdot \sum x = \sum y; \\ a \cdot \sum x + b \cdot \sum x^2 = \sum x \cdot y. \end{cases}$$

(2)

(2) теңдеулер жүйесін шеше отырып, біз a және b параметрлерінің қажетті бағаларын табамыз:

$$a = \bar{y} - b \cdot \bar{x}, \;\; b = \frac{\operatorname{cov}(x, y)}{\sigma_x^2}, \qquad (3)$$

мұндағы $\operatorname{cov}(x, y) = \overline{y \cdot x} - \bar{y} \cdot \bar{x}$ – x және y белгілерінің ковариациясы, $\sigma_x^2 = \overline{x^2} - \bar{x}^2$ – x пен

$$\bar{x} = \frac{1}{n}\sum x, \;\; \bar{y} = \frac{1}{n}\sum y, \;\; \overline{y \cdot x} = \frac{1}{n}\sum y \cdot x, \;\; \overline{x^2} = \frac{1}{n}\sum x^2$$

белгілерінің дисперсиясы.

Ковариация – осы кездейсоқ шамалардың олардың математикалық күтімдерінен ауытқуларының көбейтіндісінің математикалық күтіміне тең екі кездейсоқ шаманың бірлескен үлестірілуінің сандық сипаттамасы. **Дисперсия** – кездейсоқ шаманың математикалық күтімінен ауытқу квадратының математикалық күтімі ретінде анықталған кездейсоқ шаманың сипаттамасы. **Математикалық күтім** – кездейсоқ шама мәндерінің сәйкес ықтималдықтарға көбейтінділерінің қосындысы.

b параметрі **регрессия коэффициенті** деп аталады. Оның мәні факторды бір бірлікке өзгерту нәтижесінің орташа өзгеруін көрсетеді.

Регрессия коэффициентін нақты экономикалық тұрғыдан түсіндіру мүмкіндігін эконометрикалық зерттеулерде кеңінен таралған сызықтық регрессия теңдеуі көрсетті.

Ресми түрде а – х = 0 кезіндегі у мәні. Егер х факторының белгісі нөлдік мәнге ие бола алмайтын болса, онда а бос мүшесінің жоғарыда аталған түсіндірмесінің мәні жоқ, яғни а параметрі экономикалық мазмұнға ие болмауы мүмкін.

Регрессия теңдеуі әрқашан байланыс тығыздығы көрсеткішімен толықтырылады. Сызықтық регрессияны қолданған кезде r_{xy} сызықтық корреляция коэффициенті осындай көрсеткіш ретінде қолданылады, оны келесі формулалар бойынша есептеуге болады:

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = \frac{\operatorname{cov}(x, y)}{\sigma_x \cdot \sigma_y}. \qquad (4)$$

Сызықтық корреляция коэффициенті: $-1 \le r_{xy} \le 1$ аралығында орналасады. r_{xy} абсолютті мәні бірге неғұрлым жақын болса, факторлар арасындағы сызықтық байланыс соғұрлым күшті болады ($r_{xy} = \pm 1$ болғанда қатаң функционалдық тәуелділік аламыз). Бірақ сызықтық корреляция коэффициентінің абсолютті шамасының нөлге жақындығы белгілер арасында ешқандай байланыс жоқ дегенді білдірмейтінін есте ұстаған жөн. Модельдің басқа (сызықтық емес) спецификациясында белгілер арасындағы байланыс өте тығыз болуы мүмкін.

Сызықтық функцияны таңдау сапасын бағалау үшін **детерминация коэффициенті** деп аталатын r_{xy}^2 сызықтық корреляция коэффициентінің квадраты есептеледі. Детерминация коэффициенті нәтижелі белгінің жалпы дисперсиясындағы регрессиямен түсіндірілетін у нәтижелі белгісінің дисперсиясының үлесін сипаттайды:

$$r_{xy}^2 = 1 - \frac{\sigma_{калд}^2}{\sigma_y^2}, \qquad (5)$$

мұндағы $\sigma_{калд}^2 = \frac{1}{n}\sum (y - \hat{y}_x)^2, \ \sigma_y^2 = \frac{1}{n}\sum (y - \bar{y})^2 = \overline{y^2} - \bar{y}^2.$

Сәйкесінше, $1 - r_{xy}^2$ шамасы модельде ескерілмеген факторлардың әсерінен туындаған у дисперсиясының үлесін сипаттайды.

Сызықтық регрессия теңдеуі табылғаннан кейін жалпы теңдеудің де, оның жеке параметрлерінің де маңыздылығы бағаланады.

Регрессия теңдеуінің маңыздылығын **тексеру** дегеніміз – айнымалылар арасындағы байланысты білдіретін математикалық модельдің эксперименттік мәліметтерге сәйкес келетіндігін және тәуелді айнымалыны сипаттау үшін теңдеуге енгізілген түсіндірме айнымалылардың (бір немесе бірнеше) жеткілікті екендігін анықтау.

Әр бақылау бойынша салыстырмалы ауытқулардан модельдің сапасы туралы жалпы пікірге ие болу үшін жуықтаудың орташа қатесін анықтаймыз:

$$\overline{A} = \frac{1}{n}\sum \left| \frac{y - \hat{y}_x}{y} \right| \cdot 100\% . \qquad (6)$$

Жуықтаудың орташа қателігі 8-10% - дан аспауы керек.

Егер экономикалық құбылыстар арасында сызықтық емес қатынастар болса, онда олар тиісті сызықтық емес функциялардың көмегімен өрнектеледі [1].

Сызықтық емес регрессияның екі классы бөліп қарастырылады:

1. Түсіндірме айнымалыларды талдауға қатысты сызықтық емес, бірақ бағаланатын параметрлер бойынша сызықтық регрессиялар, мысалы:

- әртүрлі дәрежедегі көпмүшеліктер – $\hat{y}_x = a + b \cdot x + c \cdot x^2$, $\hat{y}_x = a + b \cdot x + c \cdot x^2 + d \cdot x^3$;
- теңқабырғалы гипербола – $\hat{y}_x = a + \frac{b}{x}$;
- жартылай логарифмдік функция – $\hat{y}_x = a + b \cdot \ln x$.

2. Бағаланатын параметрлер бойынша сызықтық емес регрессиялар, мысалы:

- дәрежелік – $\hat{y}_x = a \cdot x^b$;
- көрсеткіштік – $\hat{y}_x = a \cdot b^x$;
- экспоненциалдық – $\hat{y}_x = e^{a + b \cdot x}$.

Қосылған айнымалылар бойынша сызықтық емес регрессиялар айнымалыларды жай ауыстыру арқылы сызықтық түрге келеді, ал параметрлерді одан әрі бағалау ең кіші квадраттар әдісімен жүзеге асырылады. Кейбір функцияларды қарастырайық.

Екінші дәрежелі парабола $\hat{y}_x = a + b \cdot x + c \cdot x^2$ келесі ауыстыру: $x = x_1$, $x^2 = x_2$ арқылы сызықтық түрге келеді. Нәтижесінде, біз екі факторлы $\hat{y}_x = a + b \cdot x_1 + c \cdot x_2$ теңдеуге келеміз, оның параметрлерін ең кіші квадраттар әдісі көмегімен бағалау, келесі қалыпты теңдеулер жүйесіне әкеледі:

$$\begin{cases} a \cdot n + b \cdot \sum x_1 + c \cdot \sum x_2 = \sum y; \\ a \cdot \sum x_1 + b \cdot \sum x_1^2 + c \cdot \sum x_1 \cdot x_2 = \sum x_1 \cdot y; \\ a \cdot \sum x_2 + b \cdot \sum x_1 \cdot x_2 + c \cdot \sum x_2^2 = \sum x_2 \cdot y. \end{cases}$$

Айнымалыларды кері ауыстырғаннан кейін мынаны аламыз:

$$\begin{cases} a \cdot n + b \cdot \sum x + c \cdot \sum x^2 = \sum y; \\ a \cdot \sum x + b \cdot \sum x^2 + c \cdot \sum x^3 = \sum x \cdot y; \\ a \cdot \sum x^2 + b \cdot \sum x^3 + c \cdot \sum x^4 = \sum x^2 \cdot y. \end{cases} \qquad (7)$$

Екінші дәрежелі Парабола әдетте фактор мәндерінің белгілі бір аралығы үшін қарастырылып отырған белгілердің байланыс сипаты өзгерген жағдайларда: тура байланыс керіге немесе кері байланыс тураға өзгергенде қолданылады.

Теңқабырғалы гипербола $\hat{y}_x = a + \frac{b}{x}$ шикізаттың, материалдардың, отынның шығарылатын өнім көлемімен, тауар айналымының уақытпен, жұмыссыздық деңгейінің жалақы өсімінің пайызымен (мысалы, А. В. Филлипс қисығы), азық-түлікке жатпайтын тауарларға арналған шығыстардың кірістермен немесе шығындардың жалпы сомасымен байланысын (мысалы, Э. Энгель қисықтары) сипаттау үшін және басқа жағдайларда пайдаланылуы мүмкін. Гипербола қарапайым $z = \frac{1}{x}$ ауыстыру арқылы сызықтық теңдеуге келтіріледі. Ең кіші квадраттар әдісін қолданған кезде сызықтық теңдеулер жүйесі келесідей болады:

$$\begin{cases} a \cdot n + b \cdot \sum \frac{1}{x} = \sum y; \\ a \cdot \sum \frac{1}{x} + b \cdot \sum \frac{1}{x^2} = \sum \frac{1}{x} \cdot y. \end{cases} \qquad (8)$$

Осыған ұқсас $\hat{y}_x = a + b \cdot \ln x,\ \hat{y}_x = a + b \cdot \sqrt{x}$ және басқа да тәуелділіктер сызықтық түрге келтіріледі.

Бағаланатын параметрлерге байланысты сызықтық емес регрессияларда біршама өзгешелік бар, олар екі түрге бөлінеді: ішкі сызықты сызықтық емес модельдер (сызықтық түрлендірулер арқылы сызықтық түрге әкеледі, мысалы, логарифмдеу) және ішкі сызықты емес сызықтық емес модельдер (сызықтық түрге келтірілмейді).

Ішкі сызықты модельдерге мысалы дәрежелік – $\hat{y}_x = a \cdot x^b$, көрсеткіштік – $\hat{y}_x = a \cdot b^x$, экспоненциалдық – $\hat{y}_x = e^{a+b\cdot x}$, логистикалық – $\hat{y}_x = \frac{a}{1 + b \cdot e^{-c\cdot x}}$, кері – $\hat{y}_x = \frac{1}{a + b \cdot x}$ функциялар жатады.

Ішкі сызықты емес модельдерге, мысалы, келесі модельдерді жатқызуға болады: $\hat{y}_x = a + b \cdot x^c,\ \hat{y}_x = a \cdot \left(1 + \frac{1}{1 - x^b}\right)$.

Сызықтық емес модельдер арасында көбінесе $\hat{y}_x = a \cdot x^b \cdot \varepsilon$ дәрежелік функциясы қолданылады, ол сызықтық түрге логарифмдеу арқылы келтіріледі:

$$\ln y = \ln(a \cdot x^b \cdot \varepsilon);$$

$$\ln y = \ln a + b \cdot \ln x + \ln \varepsilon;$$

$$Y = A + b \cdot X + E,$$

мұндағы $Y = \ln y,\ X = \ln x,\ A = \ln a,\ E = \ln \varepsilon$. Яғни, біз ең кіші квадраттар әдісін түрлендірілген мәліметтер үшін қолданамыз:

$$\begin{cases} A \cdot n + b \cdot \sum X = \sum Y; \\ A \cdot \sum X + b \cdot \sum X^2 = \sum X \cdot Y, \end{cases}$$

одан әрі, потенциалдау арқылы ізделінді теңдеуді табамыз.

Дәрежелік функцияны кеңінен қолдану ондағы b параметрінің нақты экономикалық түсіндірілуіне байланысты – бұл иілгіштік коэффициенті болып табылады. (Иілгіштік коэффициенті, егер фактор 1%-ға өзгерсе, орташа есеппен нәтиже қанша пайызға өзгеретінін көрсетеді.) Иілгіштік коэффициентін есептеу формуласы келесідей:

$$Э = f'(x) \cdot \frac{x}{y}. \qquad (9)$$

Қалған функциялар үшін иілгіштік коэффициенті тұрақты шама емес, бірақ x факторының сәйкес мәніне байланысты болғандықтан, әдетте орташа иілгіштік коэффициенті есептеледі:

$$\overline{Э} = f'(\overline{x}) \cdot \frac{\overline{x}}{\overline{y}}. \qquad (10)$$

Регрессия теңдеулерінің ең көп қолданылатын түрлері үшін орташа иілгіштік коэффициенттерін есептеу формулаларын келтірейік (кесте):

Кесте 1 – Орташа икемділік коэффициенттерін есептеу формулалары

Функция түрі, у	Бірінші туынды, у’	Орташа икемділік коэффициенті, Э
$\hat{y}_x = a + b \cdot x + \varepsilon$	b	$\dfrac{b \cdot \bar{x}}{a + b \cdot \bar{x}}$
$\hat{y}_x = a + b \cdot x + c \cdot x^2 + \varepsilon$	$b + 2c \cdot x$	$\dfrac{(b + 2c \cdot \bar{x}) \cdot \bar{x}}{a + b \cdot \bar{x} + c \cdot \bar{x}^2}$
$\hat{y}_x = a + \dfrac{b}{x} + \varepsilon$	$-\dfrac{b}{x^2}$	$-\dfrac{b}{a \cdot \bar{x} + b}$
$\hat{y}_x = a \cdot x^b \cdot \varepsilon$	$a \cdot b \cdot x^{b-1}$	b
$\hat{y}_x = a \cdot b^x \cdot \varepsilon$	$a \cdot \ln b \cdot b^x$	$\bar{x} \cdot \ln b$
$\hat{y}_x = a + b \cdot \ln x + \varepsilon$	$\dfrac{b}{x}$	$\dfrac{b}{a + b \cdot \ln \bar{x}}$
$\hat{y}_x = \dfrac{a}{1 + b \cdot e^{-c \cdot x + \varepsilon}}$	$\dfrac{a \cdot b \cdot c \cdot e^{-c \cdot x}}{(1 + b \cdot e^{-c \cdot x})^2}$	$\dfrac{b \cdot c \cdot \bar{x}}{b + e^{c \cdot \bar{x}}}$
$\hat{y}_x = \dfrac{1}{a + b \cdot x + \varepsilon}$	$-\dfrac{b}{(a + b \cdot x)^2}$	$-\dfrac{b \cdot \bar{x}}{a + b \cdot \bar{x}}$

Икемділік коэффициентін есептеудің мағынасы жоқ жағдайлар болуы мүмкін. Бұл қарастырылып отырған белгілер үшін пайыздық өзгерісті анықтау мағынасыз болған кезде пайда болады.

Сызықтық емес регрессия теңдеуі, сызықтық тәуелділік жағдайындағыдай, байланыстың тығыздығы көрсеткішімен толықтырылады. Бұл жағдайда корреляция индексі мынадай:

$$\rho_{xy} = \sqrt{1 - \frac{\sigma^2_{калд}}{\sigma^2_y}}, \qquad (11)$$

мұндағы $\sigma^2_y = \frac{1}{n}\sum (y - \bar{y})^2$ – у нәтижелік белгісінің жалпы дисперсиясы,

$\sigma^2_{калд} = \frac{1}{n}\sum (y - \hat{y}_x)^2$ – қалдық дисперсия.

Бұл көрсеткіштің шамасы келесі аралықта болады: $0 \le \rho_{xy} \le 1$. Корреляция индексінің мәні бірге неғұрлым жақын болса, қарастырылып отырған белгілердің байланысы соғұрлым тығыз болады, регрессия теңдеуі соғұрлым сенімді болады.

Корреляция индексінің квадраты **детерминация индексі** деп аталады және нәтижелік белгінің жалпы дисперсиясындағы регрессиямен түсіндірілетін нәтижелік у белгісінің дисперсиясының үлесін сипаттайды:

$$\rho^2_{xy} = 1 - \frac{\sigma^2_{калд}}{\sigma^2_y} = \frac{\sigma^2_{түсінд}}{\sigma^2_y}, \qquad (12)$$

яғни сызықтық регрессиядағыдай мәнге ие болады; $\sigma^2_{түсінд} = \frac{1}{n}\sum(\hat{y}_x - \bar{y})^2$.

ρ^2_{xy} детерминация индексін сызықтық функцияны қолдану мүмкіндігін негіздеу үшін r^2_{xy} детерминация коэффициентімен салыстыруға болады. Регрессия сызығының қисықтығы неғұрлым үлкен болса, r^2_{xy} шамасы ρ^2_{xy}-тан аз болады. Бұл көрсеткіштердің жақындығы регрессия теңдеуінің формасын қиындатудың қажеті жоқ екенін және сызықтық функцияны қолдануға болатындығын көрсетеді.

Детерминация индексі Фишердің F-критерийі бойынша жалпы регрессия теңдеуінің маңыздылығын тексеру үшін қолданылады:

$$F = \frac{\rho^2_{xy}}{1-\rho^2_{xy}} \cdot \frac{n-m-1}{m}, \qquad (13)$$

мұндағы ρ^2_{xy} – детерминация индексі, n – бақылау саны, m – x айнымалысы кезіндегі параметрлер саны. F-критерийінің (13) нақты мәні α маңыздылық деңгейіндегі және $k_2 = n - m - 1$ (квадраттардың қалдық қосындысы үшін) мен $k_1 = m$ (квадраттардың факторлық қосындысы үшін) еркіндік дәрежелерінің санына сәйкес кесте мәнімен салыстырылады.

Сызықтық емес регрессия теңдеуінің сапасын жуықтаудың орташа қателігінен де бағалауға болады, ол сызықтық жағдайда да (6) формула бойынша есептеледі.

Регрессиялық талдаудың негізгі міндеттерінің бірі – экзогендік белгілердің берілген мәндерінде эндогендік айнымалының жеке мәндері белгілі болғандағы айнымалылар арасындағы аналитикалық тәуелділік түрін табу. Эконометрика бойынша көптеген әдістемелік құралдарда жұптық регрессияның ең дұрыс теңдеуін табу үшін олардың ішінен қолда бар деректерден ауытқуы минималды болатын функцияны таңдау ұсынылады. Бұл тәсіл әрдайым оңтайлы бола бермейді, өйткені таңдалған функция процесс моделіне экономикалық немесе дифференциалды-геометриялық талаптарды қанағаттандырмауы мүмкін.

Қолданылған әдебиеттер тізімі:

1. Касьянов В.А. «Эконометрика», учебное электронное текстовое издание, Екатеринбург, 2008.

2. Гмурман В. Е. Теория вероятностей и математическая статистика: учебное пособие для

вузов / В. Е. Гмурман. – М.: Высш. шк., 2003. – 479 с.
3. Новиков Д. А. Статистические методы в педагогических исследованиях (типовые случаи)
/ Д. А. Новиков. – М.: МЗ-Пресс, 2004. – 67 с.
4. Шилова З. В. Теория вероятностей и математическая статистика: учебное пособие / З. В. Шилова, О. И. Шилов. – Киров: Изд-во ВГГУ, 2015. – 158 с.