

## Регрессиялық талдау.

Айталық, бізде  $X$  және  $Y$  екі параметрдің  $n$  мәндер жұбы  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  бар болсын. Яғни, бір ғана объектінің екі параметрі өлшенген. Осы параметрлердің арасында қандай да бір мәнді байланыс бар ма деген сұраққа жауап беру керек.

$X$  және  $Y$  кездейсоқ шамалары өзара тәуелді немесе тәуелсіз болуы мүмкін. Тәуелділіктің екі түрі бар: **функционалды** және **статистикалық**. Математикада  $Y$  айнымалысының  $X$  айнымалысынан **функционалды тәуелділігі** деп  $X$  - тің әрбір мүмкін мәніне қандай бір белгілі ереже бойынша  $Y$  - тің бір ғана мүмкін мәні сәйкес қойылатын  $y = f(x)$  түріндегі тәуелділікті айтады.

Алайда, егер  $X$  және  $Y$  кездейсоқ шамалар болса, онда олардың арасында **статистикалық** деп аталатын тәуелділіктің басқа түрі болуы мүмкін.  $X$  және  $Y$  кездейсоқ шамаларының мәндерінің қалыптасуына әртүрлі факторлар әсер етеді. Осы факторлардың әсер етуі нәтижесінде  $X$  және  $Y$  нақты мәндері қалыптасады. Айталық,  $X$  және  $Y$  - ке бірдей факторлар, мысалы  $Z_1, Z_2, Z_3$  әсер етсін, онда  $X$  және  $Y$  бір-бірімен толық сәйкестікте болады және функционалды байланысады. Енді айталық,  $X$  - ке  $Z_1, Z_2, Z_3$ , ал  $Y$  -ке тек қана  $Z_1$  және  $Z_2$  факторлары әсер ететін болсын.  $X$  және  $Y$  екеуі де кездейсоқ шамалар, бірақ олардың екеуіне бірдей әсер ететін  $Z_1$  және  $Z_2$  ортақ факторлары бар болғандықтан, олар өзара байланысты және бұл байланыс функционалды бола алмайды, себебі: кездейсоқ шамалардың біреуіне ғана әсер ететін  $Z_3$  факторы  $X$  және  $Y$  мәндерінің арасындағы тікелей (функционалды) тәуелділікті бұзады. Бұл жағдайдағы байланыс сынаудан сынауға дейін өзгеріп отыратын кездейсоқ байланыс. Мұндай байланысты **статистикалық** деп атайды. Бұл жағдайда  $X$  - тің әрбір мәніне функционалды тәуелділіктегі сияқты  $Y$  - тің бір ғана мәні емес, қандай да бір мәндер жиыны сәйкес келуі мүмкін.

**Анықтама 1.** Егер бір кездейсоқ шаманың өзгеруіне екінші кездейсоқ шаманың таралу заңының өзгеруі сәйкес келсе, онда олардың арасындағы тәуелділік **статистикалық** деп аталады.

**Анықтама 2.** Егер бір кездейсоқ шаманың өзгеруіне екінші кездейсоқ шаманың орта мәнінің өзгеруі сәйкес келсе, онда олардың арасындағы статистикалық тәуелділік **корреляциялық** деп аталады.

Корреляциялық тәуелділікпен байланысқан кездейсоқ шамаларды **корреляцияланған** деп атайды.

**Корреляциялық тәуелділіктің** мысалдары: дененің салмағымен бойдың ұзындығының арасындағы тәуелділік: бойдың ұзындығының ( $X$ ) әрбір мәніне дене салмағының ( $Y$ ) мәндер жиыны сәйкес келеді.

- сыртқы факторлардың әсері мен аурудың арасындағы тәуелділік, мысалы, ауаның шаң – тозаңдығы, радиация деңгейі, күн көзінің өтімділігі және т. б.;

- аңдардың салмағы мен олардың қандарындағы гемоглобиннің мөлшері арасындағы тәуелділік;

- объектіге егілген препараттың мөлшерімен оның қандағы ерітіндісінің арасындағы тәуелділік;

Сол сияқты корреляция

- ионданған сәуле шығару дозасы мен мутация санының арасында;

- адам шашының пигменті мен көзінің түсі арасында;

- студенттердің жіберген дәрістері мен емтихандағы бағаларының

арасында бар.

Табиғатта ең жиі кездесетін тәуелділік – ол корреляциялық тәуелділік. Себебі зерттеліп отырған көрсеткіштерді анықтайтын өте көп әртүрлі факторлар бір-біріне әсер ету арқылы әрдайым байланыста болады.

## Корреляциялық тәуелділік

$Y$  - тің  $X$  - ке тәуелділігін

$$y_x = f(x) \quad (1)$$

түріндегі теңдеу арқылы сипаттауға болады, мұндағы  $y_x$  -  $X$  шамасының  $x$  мәніне сәйкес келетін  $Y$  шамасының шартты орта мәні, ал  $f(x)$  - қандай да бір функция. (1) теңдеу  $Y$  - тің  $X$  - ке **таңдама регрессия теңдеуі** деп аталады.

$f(x)$  функциясын  $Y$  - тің  $X$  - ке **таңдама регрессиясы** деп, ал оның графигін  $Y$  - тің  $X$  - ке **регрессиясының таңдама сызығы** деп атайды.

Дәл осы сияқты  $X$  - тің  $Y$  - ке регрессиясының таңдама теңдеуі  $\bar{x}_y = \varphi(y)$  түрінде болады.

Регрессия теңдеуінің түріне және оған сәйкес сызықтық түріне байланысты қарастырылып отырған шамалардың арасындағы корреляциялық тәуелділіктің түрін анықтайды. Ол – **сызықты, квадратты, көрсеткіштік** болуы мүмкін.

$f(x)$  [немесе  $\varphi(y)$ ] регрессия функциясының түрін таңдап алу – маңызды сұрақ болып табылады. Мысалы, сызықты немесе сызықты емес (көрсеткіштік, квадраттық, логарифмдік және т.б.)

Практикада регрессия функциясының түрін координата жазықтығында бақылаудың  $(x, y)$  мәндер жұбының барлығына сәйкес келетін нүктелер жиынын тұрғызу арқылы анықтауға болады.

### Сызықтық регрессия

Регрессия функциясының, яғни  $Y$  - тің  $X$  - ке (немесе  $X$  - тің  $Y$  - ке) тәуелділігінің моделінің түрін таңдап алғаннан кейін, мәселен,  $\bar{y}_x = a + bx$  сызықтық моделін, моделдің коэффициенттерінің нақты мәндерін анықтау қажет.

$a$  және  $b$  әртүрлі мәндері бойынша  $\bar{y}_x = a + bx$  түріндегі шексіз көп тәуелділікті алуға болады, яғни координата жазықтығындағы шексіз көп түзулер тұрғызуға болады, ал бізге бақылап отырған мәндерге өте жақын, ең тиімді түрде сәйкес келетін тәуелділік қажет. Яғни, ең жақсы коэффициенттерді таңдау керек.

### Ең кіші квадраттар әдісі (ЕКӘ)

$a + bx$  сызықтық функциясын бақылау нәтижесінде алынған мәндерге қатысты іздейміз. Бақылау мәндеріне өте жақын сәйкес келетін функцияны табу үшін **ең кіші квадраттар** әдісін қолданамыз.

Мынадай белгілеулер енгіземіз:  $Y_i - Y_i = a + bx_i$  теңдеуі бойынша есептелінген мәндер,

$y_i$  - бақылау нәтижесінде өлшелініп алынған мәндер,

$\varepsilon_i = y_i - Y_i$  - өлшелініп алынған мәндер мен теңдеу бойынша есептелінген мәндердің айырмасы,

$\varepsilon_i = y_i - a - bx_i$ .

**Ең кіші квадраттар әдісінде** өлшелініп алынған  $y_i$  мәндері мен теңдеу бойынша есептелінген  $Y_i$  мәндерінің айырмасы ең аз (минимальды) болуы талап етіледі. Яғни,  $a$  және  $b$  коэффициенттерін бақылап отырған мәндердің түзу сызықты регрессияның мәндерінен ауытқуының квадраттарының қосындысы өте аз болатындай етіп іздейміз:

$$S = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - Y_i)^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 \rightarrow \min.$$

Бұл  $S$ - функциясы  $a$  және  $b$  аргументтеріне тәуелді екі айнымалының функциясы болып табылады.  $S$ - функциясын туындының көмегі арқылы экстремумға зерттей отырып, бұл функция минимальды мәнді, егер  $a$  және  $b$  коэффициенттері мынадай жүйенің шешімі болған жағдайда қабылдайтынын дәлелдеуге болады.

$$\begin{cases} a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \\ na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \end{cases} \quad (*)$$

Бұл жүйе **ең кіші квадраттар әдісінің қалыпты теңдеулер жүйесі** деп аталады.

Егер қалыпты теңдеулердің екі жағын да  $n$  – ге бөлсек, онда

$$\begin{cases} a \frac{\sum_{i=1}^n x_i}{n} + b \frac{\sum_{i=1}^n x_i^2}{n} = \frac{\sum_{i=1}^n x_i y_i}{n} \\ a + b \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum_{i=1}^n y_i}{n} \end{cases}$$

аламыз.

$$\frac{\sum_{i=1}^n x_i}{n} = \bar{x}, \quad \frac{\sum_{i=1}^n y_i}{n} = \bar{y}, \quad \frac{\sum_{i=1}^n x_i^2}{n} = \overline{x^2}, \quad \frac{\sum_{i=1}^n x_i y_i}{n} = \overline{xy}, \quad (**)$$

екенін ескерсек,

$$\begin{cases} a\bar{x} + b\bar{x}^2 = \bar{xy} \\ a + b\bar{x} = \bar{y} \end{cases}$$

аламыз. Бұдан  $a = \bar{y} - b\bar{x}$ ,  $a$ -ның мәнін бірінші теңдеуге қою арқылы

$$\bar{x}(\bar{y} - b\bar{x}) + b\bar{x}^2 = \bar{xy} \Rightarrow \bar{x}\bar{y} - b\bar{x}^2 + b\bar{x}^2 = \bar{xy} \Rightarrow b(\bar{x}^2 - \bar{x}^2) = \bar{xy} - \bar{x}\bar{y},$$

$$b = \frac{\bar{xy} - \bar{x}\bar{y}}{\bar{x}^2 - \bar{x}^2} \quad (2)$$

$$a = \frac{\bar{x}^2\bar{y} - \bar{x}\bar{xy}}{\bar{x}^2 - \bar{x}^2} \quad (3)$$

аламыз.  $b$ - регрессия коэффициенті,  $a$ - регрессия теңдеуінің бос мүшесі.

Алынған түзу теориялық регрессия сызығының бағасы болып табылады.

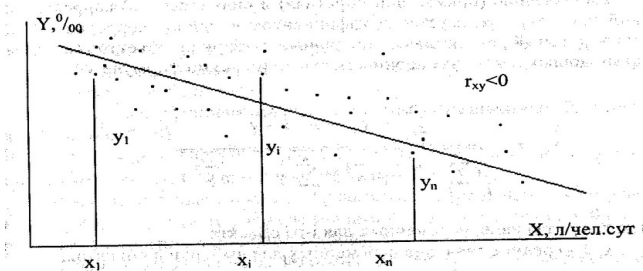
$$Y = a + bx = \bar{y} - b\bar{x} + bx = \bar{y} + b(x - \bar{x}).$$

Сонымен,  $Y = \bar{y} + b(x - \bar{x})$  сызықтық регрессияның теңдеуі.

Регрессия тура ( $b > 0$ ) және кері ( $b < 0$ ) болуы мүмкін.

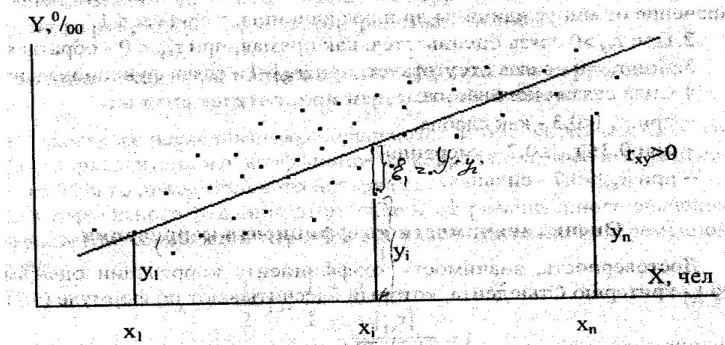
Егер бірінші параметрдің мәндері өскенде, екінші параметрдің мәндері өсетін болса, онда ол **тура регрессия**, ал бір параметрдің мәндері өскенде, екінші параметрдің мәндері кемитін болса, онда ол – **кері регрессияны** білдіреді.

**1-мысал.**  $n$  елді мекендерінде тәулігіне тұщы суды санитарлы-тұрмыстық мұқтаждыққа пайдаланудың  $X$  (л/адам) берілген деңгейіне жалпы ауру-сырқаулықтың  $Y$  (%) деңгейінің көптеген мәні сәйкес келеді. (% - ке шаққанда). Бұл жағдайда  $X$ -тің өсуіне  $Y$  - ң кемуі сәйкес келетіні байқалады. Бұл – **кері, теріс** корреляциялық байланыс. (1 сурет).



1-сурет.  $X$  факторы мен  $Y$  параметрінің арасындағы кері корреляциялық байланысты бақылау өрісі.

**2-мысал.** Өндіріс аймағындағы жұмыс орнының тығыздығы  $X$  (адам) өскен сайын жұқпалы ауру деңгейінің  $Y$  (% пен) өсуі – **тура, оң** корреляциялық байланыстың мысалы болып табылады (2 сурет).



2-сурет.  $X$  факторы мен  $Y$  параметрінің арасындағы тура корреляциялық байланысты бақылау өрісі.

### Регрессия коэффициентінің мәнділігі туралы жорамалды тексеру

Біз бақылау нәтижесінде алынған өлшемдерді сипаттайтын модельді алдық, енді оның дұрыс екендігін анықтау керек.

Регрессиялық талдаудың моделі:  $y_i = a + bx_i + \varepsilon_i$ , мұндағы  $i = 1, 2, \dots, n$  түрінде болады.  $a$ -  $Y$  бойынша ығысуды сипаттайтын параметр,  $b$ - регрессия коэффициенті, функция графигінің  $X$

бойынша ығысуын сипаттайтын параметр,  $\varepsilon_i$ - кездейсоқ айнымалының корреляциялық тәуелділікпен байланыспаған (коррелденбеген) қателіктері.

Регрессиялық талдауда  $b$  регрессия коэффициентінің және  $a$  бос мүшесінің мәнділігі туралы жорамалды тексереді.

1.  $H_0$  және  $H_1$  жорамалдарын анықтаймыз.

$H_0: b = 0$  (айнымалылар арасында сызықтық тәуелділік жоқ)

$H_1: b \neq 0$ .

2.  $\alpha$  мәнділік деңгейін береміз.

3. Критерийдің статистикасы

$$F = \frac{b^2 \sum_{i=1}^n (x_i - \bar{x})^2}{S^2}, \quad \text{мұндағы} \quad S^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 2}, \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{және} \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n}.$$

4. Критикалық нүктелер және критикалық аймақ  $K_{кр} = F_{кр}(\alpha, 1, n - 2)$ . F статистикасы 1 және (n-2) бос дәрежелерімен Фишер таралуына тең.

5. Егер  $|F_{бак}| < F_{\alpha, 1, n - 2}$ , онда  $H_0$  теріске шығарылады, яғни сызықтық тәуелділік мәнді деген қорытынды жасауға болады.

Егер  $|F_{бак}| > F_{\alpha, 1, n - 2}$ , онда бізде  $H_0$ - ді қабылдамауға негіз жоқ, яғни сызықтық тәуелділік – мәнді емес немесе берілгендерді сызықтық регрессия моделімен сипаттауға болмайды.

### Сызықтық корреляция. Корреляциялық талдау.

Корреляциялық талдау  $X$  және  $Y$  кездейсоқ шамаларының арасындағы байланыстың дәрежесін анықтаумен айналысады.

Екі кездейсоқ шамалар үшін тәжірибе нәтижесінде алынған деректердің корреляциялық талдауы мынадай негізгі тәсілдерді қарастырады:

1. Корреляциялық таңдама коэффициентін есептеу.

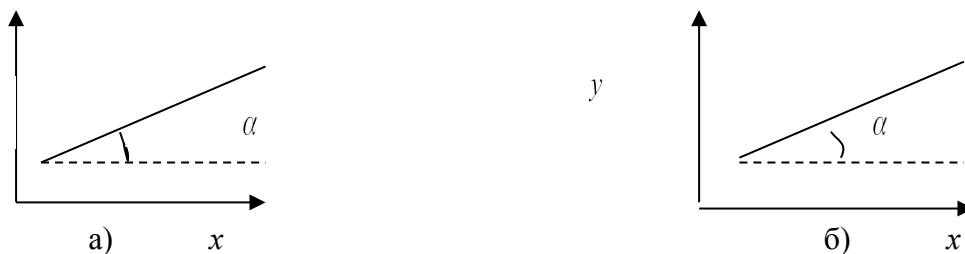
2. Корреляциялық кесте құру.

3. Байланыс мәнділігі жөніндегі статистикалық жорамалды тексеру.

**Анықтама.** Егер  $f(x)$  және  $\phi(y)$  регрессия функцияларының екеуі де сызықтық болса, онда  $X$  және  $Y$  кездейсоқ шамаларының арасындағы корреляциялық тәуелділік **сызықтық корреляция** деп аталады.

#### Корреляцияның таңдама коэффициенті.

Кездейсоқ шамалар арасындағы корреляциялық тәуелділіктің ерекшеліктерін толық сипаттау үшін бұл тәуелділіктің түрін анықтау және ол сызықтық тәуелділік болған жағдайда оның күшін регрессия коэффициентінің шамасы бойынша бағалау жеткіліксіз. 3-суретте екі сызықтық корреляциялық тәуелділіктің  $Y$ -тің  $X$ - ке графикалық суреттемесі келтірілген және регрессияның сәйкес түзу сызықтары бейнеленген.



3-сурет

Мұндағы  $Y$  және  $X$  шамаларының арасында корреляциялық тәуелділіктің күші бірдей (ол регрессия сызықтарының сәйкес түзулерінің  $OX$  осіне еңкею бұрыштарының теңдігінен шығады). Алайда 3 а) сур. кескінделген корреляциялық тәуелділік үшін бақылау нүктелерінің регрессия сызығына қарасты **шашылуы**, 3 б) сур. кескінделген корреляциялық тәуелділікке қарағанда аз. Мұндай жағдайда, қарастырылып отырған корреляциялық тәуелділіктер **тығыздығымен** ерекшеленеді деп айтады, және 3 а) сур. кескінделген корреляциялық тәуелділік 3 б) сур. келтірілген корреляциялық тәуелділікке қарағанда тығызырақ деп саналады.

Мысалы, орта мектеп оқушыларының жас мөлшері ( $Y$ ) мен олардың мектепте оқып жатқан оқу жылының ( $X$ ) арасындағы корреляциялық тәуелділік, сол сияқты жоғары оқу орнындағы студенттерінің жас мөлшерінің оқу жылына тәуелділігіне қарағанда тығыздау болатыны айқын, себебі жоғары оқу орнында бір жылда оқып жатқан студенттердің арасында жас мөлшерлерінде бір сыныпта оқитын оқушыларға қарағанда үлкен алшақтық (шашылу) болатыны байқалады.

Таңдама бақылау нәтижелері бойынша  $X$  және  $Y$  шамаларының арасындағы сызықтық корреляциялық тәуелділіктердің тығыздығын бағалау үшін **сызықтық корреляцияның таңдама коэффициенті** ұғымы енгізіледі, және ол

$$r_T = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\sigma_X \sigma_Y} \quad (4)$$

формуласымен анықталады, мұндағы  $\sigma_X$  және  $\sigma_Y$  -  $X$  және  $Y$  шамаларының сәйкес таңдама орта квадраттық ауытқулары және олар

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{\overline{X^2} - (\bar{X})^2} \quad \sigma_Y = \sqrt{\sigma_Y^2} = \sqrt{\overline{Y^2} - (\bar{Y})^2} \quad (5)$$

$$\overline{Y^2} = \frac{1}{n} \sum_{j=1}^k n_{y_j} y_j^2$$

формулаларымен есептелінеді. Айта кету керек, сызықтық корреляцияның таңдама коэффициенті  $r_T$  -ң негізгі мағынасы, ол сәйкес сызықтық корреляцияның бас коэффициентінің  $r$  эмпирикалық бағасын береді:

$$r = r_0 \quad (6)$$

$$\frac{\sigma_X}{\sigma_Y} = \frac{\sqrt{\overline{X^2} - \bar{X}^2}}{\sqrt{\overline{Y^2} - \bar{Y}^2}}, \quad r_T = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\sqrt{(\overline{X^2} - \bar{X}^2)(\overline{Y^2} - \bar{Y}^2)}}$$

формулаларын ескерсек,  $Y$  - тің  $X$  - ке сызықтық регрессиясының таңдама теңдеуі мына түрде болады:

$$Y - \bar{Y} = r_T \frac{\sigma_Y}{\sigma_X} (X - \bar{X}) \quad (7)$$

мұндағы  $r_T \frac{\sigma_Y}{\sigma_X} = b$  (регрессия коэффициенті)

Сол сияқты  $X$  - тің  $Y$  - ке сызықтық регрессияның теңдеуін жазуға болады:

$$X - \bar{X} = r_T \frac{\sigma_X}{\sigma_Y} (Y - \bar{Y}) \quad (8)$$

### Сызықтық корреляцияның таңдама коэффициентінің негізгі қасиеттері

1. Сызықтық корреляциялық тәуелділікпен байланыспаған екі шаманың корреляция коэффициенті нөлге тең.
2. Сызықтық корреляциялық тәуелділікпен байланысқан екі шаманың корреляциялық коэффициенті 1-ге тең, егер тәуелділік тура (өспелі) болса және -1-ге тең, егер тәуелділік кері (кемімелі) болса;
3. Сызықтық корреляциялық тәуелділікпен байланысқан екі шаманың корреляциялық коэффициентінің абсолюттік шамасы  $|r| < 1$  теңсіздігін қанағаттандырады. Егер корреляциялық тәуелділік өспелі болса, онда корреляция коэффициенті оң, егер корреляциялық тәуелділік кемімелі болса, онда корреляция коэффициенті теріс.
4.  $|r|$  1-ге неғұрлым жақын болса,  $Y, X$  шамаларының арасындағы түзу сызықты корреляция соғұрлым тығыздық.

Корреляциялық байланыс сипатына қарай **тура** және **кері**, күшіне қарай **күшті**, **орташа**, **әлсіз** болуы мүмкін.

**Параметрлер арасындағы байланыстың күші мен сипаты.**

Байланыс күші	Байланыс сипаты	
	Тура (+)	Кері (-)
Толық	1	-1
Күшті	0,7 ден 0,999 ге дейін	-0,7 ден -0,999 ге дейін
Орта	0,699 ден 0,3 ге дейін	-0,699 ден -0,3 ге дейін
Әлсіз	0,299 ден 0 ге дейін	-0,299 ден 0 ге дейін
Байланыс жоқ	0	0

### Корреляцияның таңдама коэффициентінің мәнділігі туралы жорамалды тексеру.

Статистикалық өзара байланыс жөнінде сөз болғанда, оның **бар болуы** немесе **жоқ болуы**, бағытының болуы (оң немесе теріс) және күшінің сипатымен (күшті, әлсіз) анықталады. Егер зерттеу нәтижесінде нөлдік жорамал теріске шығарылмаса, онда «өзара байланыс жоқ». Нөлдік жорамал теріске шығарылған жағдайда кездейсоқ шамалардың арасында байланыс бар делінеді.

1.  $H_0$  және  $H_1$  жорамалдарын тағайындайық:

$H_0 : r = 0$  (яғни, корреляция жоқ),

$H_1 : r \neq 0$ .

2.  $\alpha$  мәнділік деңгейін береміз.

3. Критерияның статистикасы:  $t_{\text{өйсе}} = r_0 \sqrt{\frac{n-2}{1-r_0^2}}$

4.  $(t_{\text{нәііе}}(\alpha, f))$  мұндағы  $t - f = n - 2$  бос дәрежесімен Стьюдент таралуы болатын статистика.

5.  $|t_{\text{өйсе}}| \geq t_{(n-2), \alpha}$  болғанда  $H_0$  жорамалы:  $r = 0$   $\alpha$  мәнділік деңгейімен теріске шығарылады.

Демек, параметрлердің арасында мәнді корреляция бар.

$|t_{\text{мәжс}}| < t_{(n-2), \alpha}$  болғанда  $H_0$  жорамалын теріске шығаруға негіз жоқ, яғни параметрлердің арасында мәнді корреляция жоқ.

**3 мысал.**  $x$  және  $y$  шамаларының өлшем нәтижелері кестеде берілген.

$x_i$	-2	0	1	2	4
$y_i$	0,5	1	1,5	2	3

1)  $y$ -н  $x$ -қа түзу регрессиясының таңдама теңдеуін;

2) Регрессия сызығын тұрғызу;

3) Корреляцияның таңдама коэффициентін есептеу;

4) Корреляциялық тәуелділіктің сипаты мен күшін анықтау.

**Шешуі.** Мұнда  $n=5$

1) а)  $X$  пен  $Y$ -тің арасында  $y = a + bx$  сызықты тәуелділігі бар деп ұйғарып, ең кіші квадраттар әдісімен  $a$  және  $b$  коэффициенттерін табамыз.

$$\sum_{i=1}^5 x_i = -2 + 0 + 1 + 2 + 4 = 5; \quad \sum_{i=1}^5 x_i^2 = 4 + 0 + 1 + 4 + 16 = 25$$

$$\sum_{i=1}^5 x_i y_i = -2 \cdot 0,5 + 0 \cdot 1 + 1 \cdot 1,5 + 2 \cdot 2 + 4 \cdot 3 = -1 + 1,5 + 4 + 12 = 16,5$$

$$\sum_{i=1}^5 y_i = 0,5 + 1 + 1,5 + 2 + 3 = 8$$

$$\text{калыпты жүйенің түрі} \begin{cases} 5a + 25b = 16,5 \\ 5a + 5b = 8 \end{cases}$$

Осы жүйені шеше отырып,  $b = 0,425$ ,  $a = 1,175$  аламыз. Сондықтан,  $y = 1,175 + 0,425x$ .

в) Немесе мына формаларды қолданып табуға болады  $b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$ ,  $a = \bar{y} - b\bar{x}$

$$\bar{x} = \frac{-2+0+1+2+4}{5} = \frac{5}{5} = 1, \quad \bar{y} = \frac{0.5+1+1.5+2+3}{5} = \frac{8}{5} = 1.6$$

$$\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y}) = (-2-1)(0.5-1.6) + (0-1)(1-1.6) + (1-1)(1.5-1.6) + (2-1)(2-1.6) + (4-1)(3-1.6) = 3.3 + 0.6 + 0 + 0.4 + 4.2 = 8.5$$

$$\sum_{i=1}^5 (x_i - \bar{x})^2 = (-2-1)^2 + (0-1)^2 + (1-1)^2 + (2-1)^2 + (4-1)^2 = 9 + 1 + 0 + 1 + 9 = 20$$

$$b = \frac{\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^5 (x_i - \bar{x})^2} = \frac{8.5}{20} = 0.425. \text{ Регрессия тура } b = 0.425 > 0.$$

$$a = \bar{y} - b\bar{x} = 1.6 - 0.425 \cdot 1 = 1.175.$$

Онда сызықтық регрессияның теңдеуі мына түрде жазылады  $y = a + bx \Rightarrow y = 1.175 + 0.425x$ .

$$c) b = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} \quad a = \frac{\overline{x^2y} - \bar{x}x\bar{y}}{\overline{x^2} - \bar{x}^2}$$

$$\bar{x} = \frac{-2+0+1+2+4}{5} = \frac{5}{5} = 1, \quad \bar{y} = \frac{0.5+1+1.5+2+3}{5} = \frac{8}{5} = 1.6$$

$$\overline{xy} = \frac{-2 \cdot 0.5 + 0 \cdot 1 + 1 \cdot 1.5 + 2 \cdot 2 + 4 \cdot 3}{5} = \frac{16.5}{5} = 3.3$$

$$\overline{x^2} = \frac{(-2)^2 + 0^2 + 1^2 + 2^2 + 4^2}{5} = \frac{25}{5} = 5$$

$$b = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{3.3 - 1 \cdot 1.6}{5 - 1^2} = \frac{3.3 - 1.6}{4} = 0.425, \quad a = \frac{\overline{x^2y} - \bar{x}x\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{5 \cdot 1.6 - 1 \cdot 3.3}{5 - 1^2} = 1.175$$

$$y = a + bx \Rightarrow y = 1.175 + 0.425x.$$

### Регрессия коэффициентінің мәнділігі туралы жорамалды тексеру

1.  $H_0$  және  $H_1$  жорамалдарын анықтаймыз.

$H_0: b=0$  (айнымалылар арасында сызықтық тәуелділік жоқ)

$H_1: b \neq 0$ .

2.  $\alpha=0,05$  мәнділік деңгейін береміз.

3. Критерийдің статистикасы

$$F = \frac{b^2 \sum_{i=1}^n (x_i - \bar{x})^2}{S^2},$$

мұндағы  $S^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-2}$ ,  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$  және  $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$ .

$$\sum_{i=1}^5 (y_i - \bar{y})^2 = (0.5-1.6)^2 + (1-1.6)^2 + (1.5-1.6)^2 + (2-1.6)^2 + (3-1.6)^2 = 3.7$$

$$S^2 = \frac{\sum_{i=1}^5 (y_i - \bar{y})^2}{n-2} = \frac{3.7}{5-2} = 1.23$$

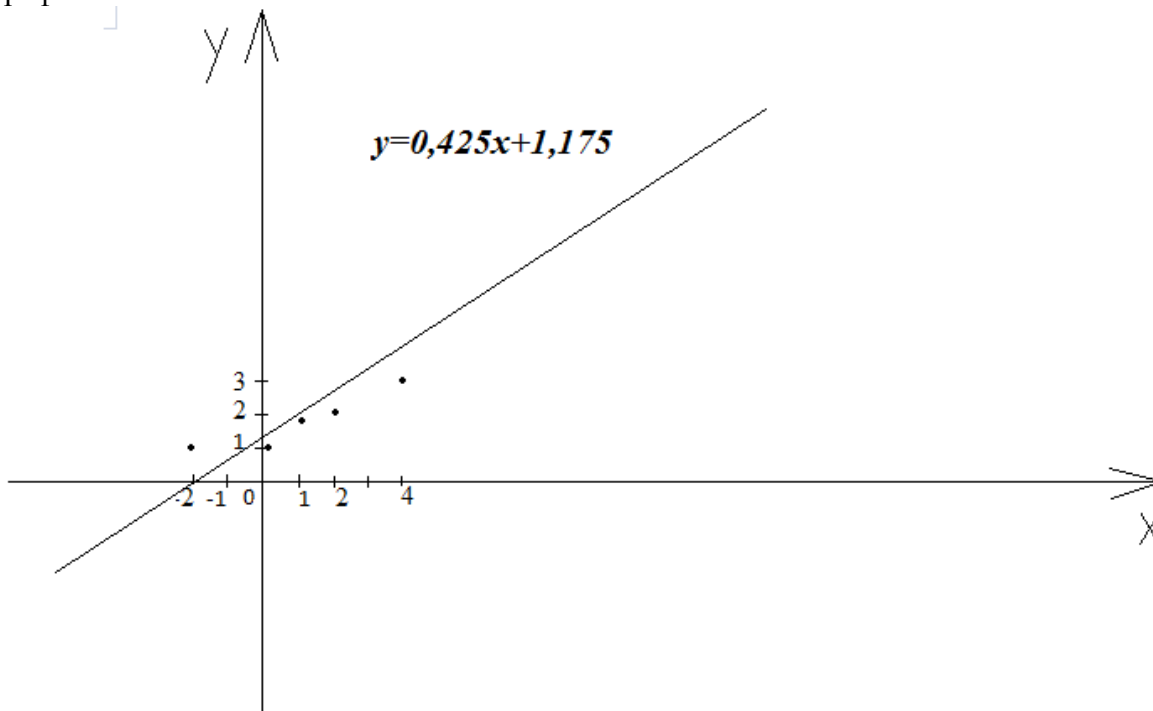
$$F_{bak} = \frac{b^2 \sum_{i=1}^n (x_i - \bar{x})^2}{S^2} = \frac{0.425^2 \cdot 20}{1.23} = 2.94$$

4. Критикалық нүктелер және критикалық аймақ  $F_{кр} = F_{кр}(\alpha, 1, n-2)$ . F статистикасы 1 және (n-2) бос дәрежелерімен Фишер таралуына тең.

$$F_{сыни} = F_{сыни}(0.05, 1, 3) = 10.13$$

5. Егер  $|F_{бак}| = 2.94 < F_{\alpha, 1, n-2} = 10.13$ , онда  $H_0$  теріске шығарылады, яғни сызықтық тәуелділік мәнді деген қорытынды жасауға болады.

2) график



3) Сызықтық корреляцияның таңдама коэффициенті

$$r_T = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y} = \frac{3.3 - 1 \cdot 1.6}{2 \cdot 0.86} = \frac{1.7}{1.72} = 0.99$$

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{5 - 1^2} = 2 \quad \sigma_y = \sqrt{\sigma_y^2} = \sqrt{\overline{y^2} - (\bar{y})^2} = \sqrt{3.3 - 1.6^2} = \sqrt{0.74} \approx 0.86$$

$$\text{Немесе } r_T = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{8.5}{\sqrt{20 \cdot 3.7}} = 0.99$$

4) Сонымен, қарастырылып отырған x пен y-тің арасындағы корреляциялық тәуелділік сипаты бойынша – тура, күші бойынша – күшті.

### Корреляцияның таңдама коэффициентінің мәнділігі туралы жорамалды тексеру.

1.  $H_0$  және  $H_1$  жорамалдарын тағайындайық:

$H_0$  :  $r = 0$  (яғни, корреляция жоқ),

$H_1$  :  $r \neq 0$ .

2.  $\alpha = 0.05$  мәнділік деңгейін береміз.

3. Критерияның статистикасы:

$$t_{bak} = r_T \sqrt{\frac{n-2}{1-r_T^2}} = 0.99 \sqrt{\frac{5-2}{1-0.99^2}} = 0.99 \sqrt{\frac{3}{0.0199}} = 12.16$$



4.  $t_{\text{крит}}(\alpha, df) = t_{\text{крит}}(0.05, 3) = 3.18$ , мұндағы  $df = n - 2$  бос дәрежесімен Стьюдент таралуы болатын статистика.

5.  $|t_{\text{бак}}| \geq t_{(n-2), \alpha} \Rightarrow |t_{\text{бак}}| \geq t_{(5-2), 0.05} \Rightarrow 12.16 > 3.18$  болғанда  $H_0$  жорамалы  $\alpha=0,05$  мәнділік деңгейімен жоққа шығарылады. Демек, параметрлердің арасында мәнді корреляция бар.

### **Тапсырмалар.**

1. Ер кісілердің кеудесінің көлемі  $Y$  (см) мен бойларының ұзындықтарының  $X$  (см) арасындағы тәуелділік зерттелген. Бақылау нәтижелері көлемі 7 болатын екі өлшемді таңдама түрінде кестеде келтірілген:

X	162	164	179	172	182	188	168
Y	88	94	98	100	102	108	112

Табу керек:

- 1)  $y$ -н  $x$ -қа түзу регрессиясының таңдама теңдеуін;
- 2) Регрессия сызығын тұрғызу;
- 3) Корреляцияның таңдама коэффициентін есептеу;
- 4) Корреляциялық тәуелділіктің сипаты мен күшін анықтау.

2. Жүректің бір минут көлемінде соғуы ( $Y$ ) мен жүректің сол жақ құлақшасындағы орта қысымның ( $X$ ) арасындағы тәуелділік зерттелген. Бақылау нәтижелері көлемі 5 болатын екі өлшемді таңдама түрінде кестеде берілген:

X	4,8	6,4	9,3	11,2	17,7
Y	0,4	0,69	1,29	1,64	2,4

Табу керек:

- 1)  $y$ -н  $x$ -қа түзу регрессиясының таңдама теңдеуін;
- 2) Регрессия сызығын тұрғызу;
- 3) Корреляцияның таңдама коэффициентін есептеу;
- 4) Корреляциялық тәуелділіктің сипаты мен күшін анықтау.