

$$\begin{array}{ll} \text{а) } \frac{\sum p_1 q_1}{\sum q_1} \div \frac{\sum p_0 q_1}{\sum q_1} & \text{б) } \frac{\sum p_0 q_1}{\sum q_1} \div \frac{\sum p_0 q_0}{\sum q_0} \\ \text{в) } \frac{\sum p_1 q_1}{\sum q_1} \div \frac{\sum p_0 q_0}{\sum q_0} & \text{г) } \frac{\sum p_1 q_0}{\sum q_0} \div \frac{\sum p_0 q_1}{\sum q_1} \quad \text{д) } \frac{\sum p_0 q_0}{\sum q_0} \div \frac{\sum p_1 q_1}{\sum q_1} \end{array}$$

10. Орташа индекстердің мынандай түрлері болады:

- а) интенсивті
- б) құрылымдық
- в) арифметикалық
- г) аналитикалық
- д) хронологиялық

11. Өзіндік құнның тұрақты құрамды индексінде мына көрсеткіш тұрақты болады:

- а) вариант
- б) өзіндік құн
- в) баға
- г) шығын
- д) өнім көлемі

12. Есепті мерзімдегі нақты тауар айналымы және тауарлар бағасының дербес индекстері туралы мәліметтер белгілі болған жағдайда тауарлар бағасының жалпы индексін анықтау үшін қандай индексті қолданады?

- а) арифметикалық орташа индексті
- б) гармониялық орташа индексті
- в) тұрақты құрамды индексті
- г) өзгермелі құрамды индексті
- д) құрылымдық өзгеру индекссін

9-тарау. ҚОҒАМДЫҚ ҚҰБЫЛЫСТАРДЫҢ ӨЗАРА БАЙЛАНЫСЫН ЗЕРТТЕУ

9.1. Құбылыстардың өзара байланысы туралы жалпы түсінік

Қоғамдық құбылыстар бір-бірімен байланысты болады. Бір құбылыстың өзгеруі екінші құбылысты тудырады немесе өзгертеді. Қоғамдық құбылыстардың өзара байланысын, тәуелділігін айқындау арқылы олардың даму заңдылығын анықтауға болады. Сондықтан әлеуметтік-экономикалық құбылыстардың өзара байланысын зерттеу статистиканың негізгі міндеттерінің бірі болып саналады. Құбылыстар байланысын статистикалық зерттеу нәтижесінде құбылыстар арасындағы қатынастың себебі, салдары, құбылысқа, үдеріске әсер ететін факторлар ықпалы анықталады. Мұндай зерттеулер бірнеше кезеңнен тұрады. Байланысты статистикалық зерттеудің алғашқы кезеңінде экономикалық немесе әлеуметтік құбылыстың табиғаты экономикалық теория, әлеуметтану тұрғысынан талданады, яғни сапалық

талдау жасалады. Бұл құбылыстар арасындағы байланысты анықтау ең алдымен теориялық талдауға негізделеді дегенді білдіреді. Екінші кезеңде құбылыстар байланысының үлгісі құрылады., ал үшінші кезеңде құбылыстың сапалық ерекшеліктері ескеріле отырып, зерттеу нәтижелері түсіндіріледі.

Құбылыстардың өзара байланысын зерттеу үшін белгілердің екі түрін қолданады:

- факторлық белгі;
- нәтижелік белгі.

Факторлық белгі деп өзімен байланысты белгілерді өзгертетін белгіні айтады.

Нәтижелік белгі деп факторлық белгілер әсерінен өзгертетін белгіні айтады. Бұл белгілерді себеп-салдар белгілері деп те атайды.

Құбылыстар арасындағы өзара байланыстың екі түрі болады:

- функционалды байланыс;
- стохастикалық байланыс.

Функционалды байланыс деп факторлық белгінің бір мәніне нәтижелік белгінің бір немесе бірнеше айқындалған мәні сәйкес келетін байланысты айтады. Мысалы жүрілген жол жылдамдыққа және уақытқа, квадраттың ауданы оның қабырғасының ұзындығына байланысты болады, Функционалды байланысты мынадай теңдеу түрінде көрсетуге болады:

$$y_i = f(x_i),$$

мұндағы y_i – нәтижелік белгі ($i = 1, \dots, n$);

x_i – факторлық белгі;

$f(x_i)$ – факторлық және нәтижелік белгі байланысын көрсететін функция.

Функционалды байланыс көбінесе математика, физика, т.б. нақты ғылымдар сипаттайтын құбылыстарда байқалады. Егер белгілер арасында функционалды байланыс болса, ол байланыс бақылау жүргізілген барлық жағдайла және жиынтықтың барлық бірлігінде байқалуы тиіс.

Егер белгілер арасындағы тәуелділік әрбір жеке жағдайда емес, жалпы алғанда, бақылау саны өте үлкен болған жағдайда ғана байқалса, ондай байланысты **стохастикалық байланыс** дейді. Стохастикалық байланыста факторлық белгінің бір мәніне нәтижелік белгінің көптеген мәндері сәйкес келеді. Мысалы, еңбек өтілі (стажы) бірдей жұмысшылардың тарфтік дәрежелері әр түрлі болуы мүмкін, себебі тарифтік дәреже жұмысшының біліктілігіне, қабілетіне, т.б. факторларға байланысты болады. Стохастикалық байланыста нәтижелік белгіге есепке алынған факторлық белгілерден басқа кездейсоқ факторлар да әсер етеді. Стохастикалық байланыс жиынтықтың жеке бірлігі үшін байқалмайды, ондай байланысты тек жалпы жиынтықта байқауға болады. Стохастикалық байланысты мынадай теңдеу түрінде көрсетуге болады:

$$\hat{y}_i = f(x_i) + \varepsilon_i,$$

мұндағы \hat{y}_i – нәтижелік белгінің есептелген мәні ($i = 1, \dots, n$);

x_i – факторлық белгі;

$f(x_i)$ – есепке алынған факторларға байланысты анықталған нәтижелік белгі мәні ;

ε_i – есепке алынбаған факторларға байланысты анықталған нәтижелік белгі мәні.

Стохастикалық байланыстың жеке бір жағдайы болып корреляциялық байланыс саналады. Факторлық белгі өзгергенде нәтижелік белгінің орташасы өзгеретін байланысты **корреляциялық байланыс** дейді. Корреляция сөзін ғылымға ағылшын биологы және статистигі Френсис Гальтон XIX ғасырдың аяғында енгізген. Ол кезде корреляция (correlation) сөзі «сәйкес келу», «байланыс сияқты» деген мағынада қолданылды. Көптеген қоғамдық құбылыстарға корреляциялық байланыс тән.

Функционалдық және стохастикалық байланыстарды бағытына байланысты тура және кері байланыс деп бөледі. Егер факторлық белгі өскенде (кемігенде) нәтижелік белгі де өссе (кемісе), онда осы екі белгінің арасында **тура** байланыс болады, яғни көрсеткіштер арасында тура байланыс болғанда факторлық белгі мен нәтижелік белгі бір бағытта дамиды. Мысалы, тарифтік дәреже мен еңбек өнімділігі көрсеткіштерінің арасында тура байланыс бар, себебі жұмысшының тарифтік дәрежесі өскен сайын еңбек өнімділігі өседі. Егер факторлық белгі өскенде (кемігенде) нәтижелік белгі кемісе (өссе), онда аталған белгілер арасында **кері** байланыс болады. Мысалы өнімнің өзіндік құны мен еңбек өнімділігі арасында кері байланыс бар, себебі еңбек өнімділігі өскен сайын, өнімнің өзіндік құны кемиді.

Аналитикалық формасына немесе қалай өрнектелгеніне қарай байланыстарды түзу сызықты және қисық сызықты байланыстар деп бөледі. Құбылыстар арасында **түзу сызықты байланыс** болғанда факторлық белгі өскен сайын нәтижелік белгі де үздіксіз өсіп (кеміп) отырады. Математикалық тұрғыдан түзу сызықты байланысты түзудің теңдеуімен өрнектейді.

$$\bar{y}_x = a_0 + a_1 \cdot x ,$$

мұндағы a_0, a_1 – түзу параметрлері;

x – факторлық белгі;

\bar{y}_x – нәтижелік белгінің орташа шамасы

Зерттелетін құбылыстар арасында **қисық сызықты байланыс** болғанда факторлық белгі өскенде нәтижелік белгі бірқалыпты өзгермейді. Қисық сызықты байланыс парабола, гипербола, т.б. қисықтардың теңдеулері арқылы өрнектеледі. Төменде парабола теңдеуі көрсетілген:

$$\bar{y}_x = a_0 + a_1 \cdot x + a_2 \cdot x^2 ,$$

мұндағы a_0, a_1, a_2 – парабола параметрлері.

Ал гипербола теңдеуі мынадай түрде беріледі:

$$\bar{y}_x = a_0 + \frac{a_1}{x} .$$

Нәтижелік белгіге әсер ететін факторлар санына қарай құбылыстар арасындағы байланыстар бір факторлы және көп факторлы болып бөлінеді.

Бір факторлы байланыс факторлық бір белгі мен нәтижелік белгі арасындағы байланысты білдіреді. Мұндай қарапайым байланысты әдетте жұптық байланыс, ал корреляция теориясында жұптық корреляция деп атайды. Мысалы, шығарылған өнім көлемі мен кәсіпорынның негізгі капиталының арасында корреляциялық байланыс болады.

Көп факторлы байланыс екі немесе одан да көп факторлық белгілер мен нәтижелік белгі арасындағы байланысты білдіреді. Мұндай байланысты көптік корреляция деп те атайды. Көптік корреляцияда факторлық белгілер нәтижелік белгіге бір мезгілде, бірге әсер етеді. Мысалы кәсіпорынның шығарған өнімінің көлеміне негізгі капитал, еңбек өнімділігі, мамандардың біліктілігі, еңбекті ұйымдастыру деңгейі, т.б. көптеген факторлар әсер етеді.

9.2. Құбылыстар байланысын зерттеудің статистикалық әдістері

Зерттелетін әлеуметтік-экономикалық құбылыстар арасындағы байланысты, олардың сипатын, даму бағытын айқындау үшін мынадай әдістер қолданылады:

- параллель қатарларды салыстыру әдісі;
- аналитикалық топтастыру әдісі,
- баланс әдісі;
- график әдісі;
- индекстік талдау әдісі;
- корреляциялық-регрессиялық әдіс.

Параллель қатарларды салыстыру әдісі құбылыстар арасында стохастикалық байланыстың бар немесе жоқ екенін, байланыс болған жағдайда, оның сипаты, даму бағыты қандай екенін анықтауға мүмкіндік береді. Бұл әдіс екі немесе одан да көп көрсеткіштер қатарларын салыстыруға негізделген. Қатарларды салыстыру үшін алдымен нәтижелік белгіге әсер ететін факторлық белгілерді зерттеу мақсатына байланысты не өсу, не кему ретімен орналастырады, содан кейін нәтижелік белгінің өзгерісін бақылайды. Параллель қатарларды салыстыру әдісінің негізгі кемшілігі – құбылыстар арасындағы байланысты сандық жағынан бағалау мүмкіндігінің болмауы. Соған қарамастан бұл әдіс өте қарапайым әрі ыңғайлы болғандықтан, экономикалық-статистикалық талдауларда жиі қолданылады. Енді осы әдісті қолдануға мысал келтірейік.

Мысал. Мектеп оқушыларының апталық жүктемесі туралы мынадай мәліметтер берілген:

9.1-кесте

Мектеп оқушыларының апталық жүктемесі

Сынып	1–2	3–4	5	6	7	8–9	10–11
Апталық жүктеме, сағат	20	25	31	32	34	35	36

Кестеде берілген мәліметтерден оқушының қай сыныпта оқитыны мен апталық жүктеме көрсеткіші арасында тура байланыс бар екенін байқау қиын емес.

Аналитикалық топтау әдісін стохастикалық байланысты анықтау үшін қолданғанда, ең алдымен зерттелетін жиынтықтың барлық элементтерін факторлық белгі бойынша топтайды. Содан кейін әр топ үшін нәтижелік белгінің орташа мәнін есептейді. Факторлық белгі өзгерген сайын нәтижелік белгі қалай өзгередінін салыстыра отырып, эмпириялық корреляциялық қатынас көрсеткішін (бұл көрсеткіш туралы 6.3-тарауда айтылған) пайдаланып, байланыс бағытын, тығыздығын анықтауға болады. Эмпириялық корреляциялық қатынас көрсеткіші мына формуламен есептеледі:

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}},$$

мұндағы η – эмпириялық корреляциялық қатынас көрсеткіші;

δ^2 – топаралық дисперсия;

σ^2 – жалпы дисперсия.

Эмпириялық корреляциялық қатынас көрсеткіші факторлық және нәтижелік белгілер арасындағы тығыздықты бағалау үшін қолданылады. Оның абсолютті мәні 0 мен 1 аралығында өзгереді. Егер топтастыру белгісі мен нәтижелік белгі арасында байланыс болмаса, онда корреляциялық қатынас көрсеткіші нольге тең болады ($\eta = 0$). Ал топтастыру белгісі мен нәтижелік белгі арасында функционалдық байланыс болған жағдайда $\eta = 1$ болады. Бұл нәтижелік белгіге тек топтастыру белгісі ғана әсер ететінін көрсетеді. Корреляциялық қатынас көрсеткішінің мәні 1-ге жақындаған сайын корреляциялық байланыс функционалдық байланысқа жақындайды. Құбылыстар арасындағы байланыс тығыздығын дұрыс бағалау үшін мынадай кестені пайдалануға болады (9.2-кесте):

9.2-кесте

Байланыс тығыздығын анықтау кестесі

Корреляциялық қатынас көрсеткішінің мәні (η)	0,1–0,3	0,3–0,5	0,5–0,7	0,7–0,9	0,9–0,99
Байланыс тығыздығы	әлсіз	орташа	айқын	тығыз	өте тығыз

Аналитикалық топтастыруда топ санын айқындау, интервал шегін белгілеу топтастыру мақсатына сәйкес анықталады. Егер топ саны аз болса, байланыс сипаты айқындалмайды. Ал топ санын көбейту әр топтағы бірлік санының азаюына себеп болады, яғни бұл жағдайда топтық орташаның сенімділігі төмендейді. Сондықтан топ санын дұрыс анықтаудың маңызы зор. Аналитикалық топтау әдісін қолдану мысалы ретінде зауыттарды негізгі капитал мөлшері бойынша топтау нәтижесінде шығарылған өнім мен негізгі капитал арасындағы байланысты анықтауды келтіруге болады.

Баланс әдісі статистикада құбылыстар байланыстарын талдау, экономикалық пропорцияларды анықтау, салалар, аймақтар арасындағы байланысты сипаттау үшін қолданылады. Материалдық ресурстар қозғалысын мынадай баланс түрінде көрсетуге болады:

$$a + b = c + d,$$

мұндағы a – мерзім басындағы ресурстар қоры;

- b* – мерзім ішінде қосылған ресурс;
- c* – мерзім ішінде жұмсалған ресурс;
- d* – мерзім соңындағы ресурстар қоры.

Баланс құру арқылы қатар жұмыс күшінің, негізгі капиталдың, қаржы құралдарының, т.б. көрсеткіштердің қозғалыстарын да зерттеуге болады. Баланс әдісі көрсеткіштер байланысын талдау үшін ғана қолданылмайды, ол мәліметтерді тексеруге, белгісіз көрсеткіштерді анықтауға мүмкіндік береді. Ел экономикасын макро деңгейде сипаттау, талдау үшін қолданылатын нарықтық экономиканың халықаралық статистикалық үлгісі – ұлттық шоттар жүйесінде (ҰШЖ) баланс әдісі кеңінен қолданылады. Бұл жүйеде шоттар екі бөліктен тұрады және баланстық кесте түрінде беріледі, яғни шоттың ресурстар бөлігінің жалпы сомасы сол шоттың пайдалану бөлігінің жалпы сомасына тең болуы керек. Төменде Қазақстанның ұлттық шоттар жүйесіндегі негізгі шоттардың бірі – табыстарды қайта бөлу шотының 2009 жылғы баланстық кестесі мысал ретінде келтірілген (9.3-кесте).

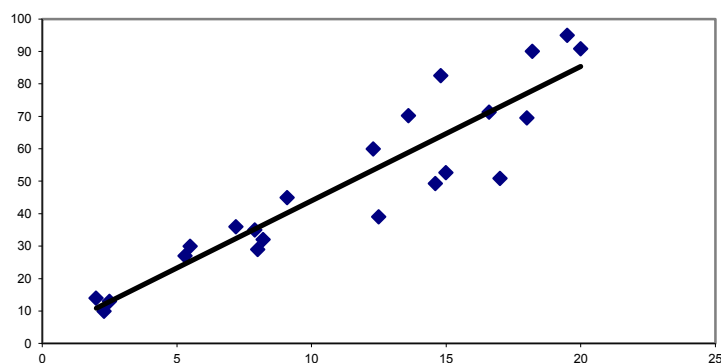
9.3-кесте

Табыстарды қайта бөлу шоты, (ағымдағы бағамен), млрд.тг.

Ресурстарды пайдалану	Σ	Ресурстар	Σ
3. «Қалған әлемге» берілген ағымдағы трансферттер	259,6	1. Жалпы ұлттық табыс	15115,7
4. Қолда бар жалпы ұлттық табыс	14981,9	2. «Қалған әлемнен» қабылданған ағымдағы трансферттер	125,8
Барлығы	15241,5	Барлығы	15241,5

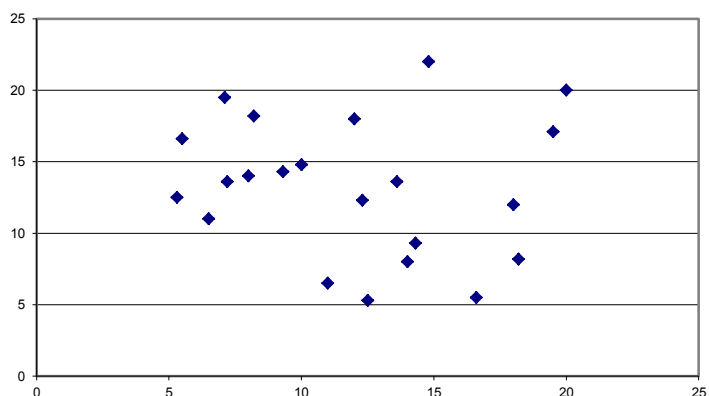
Экономикалық баланстардың ішінде ең күрделі баланс - ұлттық шоттар жүйесіндегі салааралық баланс. Теориялық тұрғыдан қарағанда салааралық баланс ұдайы өндіріс үдерісінің экономикалық-математикалық үлгісі болып саналады. Бұл үлгіде өнімді өндіру, бөлу, тұтыну, қор жинау арасындағы байланыс толық қамтылып көрсетіледі, сондықтан салааралық баланс экономика салаларының арасындағы байланысты талдауға, негізгі экономикалық пропорцияларды айқындауға мүмкіндік береді. Қазіргі кезде құбылыстар арасындағы байланыстарды талдау үшін әр түрлі баланстар жүйесі қолданылады.

Бастапқы жиналған мәліметтерді өңдеу нәтижесінде алынған көрсеткіштерді көрнекі түрде бейнелеп, яғни **график әдісін** пайдаланып, зерттелетін белгілер арасындағы байланыстың бағытын, формасын анықтауға болады. Екі көрсеткіштің байланысын график түрінде нүктелік диаграмма, яғни **корреляция өрісі** арқылы бейнелейді. Ол үшін координаттар жүйесінің абсцисса осіне факторлық белгі мәндерін, ордината осіне нәтижелік белгі мәндерін салады. Белгілер арасындағы байланыс тығыз болған сайын, нүктелер байланыс формасын анықтайтын сызық бойына көп шоғырланады (9.1-сурет).



9.1-сурет. Корреляция өрісі. Түзу сызықты байланыс.

Ал байланыс жоқ болса, нүктелер координата жүйесінде ретсіз орналасады (9.2-сурет).



9.2-сурет. Корреляция өрісі. Көрсеткіштер арасында байланыс жоқ.

Құбылыстар арасындағы байланысты зерттеу үшін **индекстер әдісі** де қолданылады. Индекстер күрделі құбылыстардың динамикасын ғана анықтамайды, олар сонымен қатар құбылысқа әсер ететін факторлардың ықпалын да айқындайды. Яғни құбылыстар байланысы индекстер арасында белгілі бір қатынастың болуына себепін тигізеді. Бір-бірімен байланысты экономикалық көрсеткіштер индекстері белгілі бір жүйені құрайды. Мысалы, тауар айналымының жалпы индексі тауарлар бағасының жалпы индексі мен физикалық көлемнің жалпы индексінің көбейтіндісіне тең болады.

$$I_p \times I_q = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} \times \frac{\sum P_0 Q_1}{\sum P_0 Q_0} = \frac{\sum P_1 Q_1}{\sum P_0 Q_0} = I_{pq} \text{ немесе: } I_p \times I_q = I_{pq},$$

мұндағы I_p – бағаның жалпы индексі;

I_q – тауарлардың физикалық көлемінің жалпы индексі;

I_{pq} – тауар айналымының жалпы индексі.

Индекстер әдісін пайдаланып, құбылыстар арасындағы өзара байланысты зерттеу үшін факторлық талдау жасағанда, статистикада мынадай ережеге сүйенеді: *егер нәтижелік белгі көлемдік және сапалық белгілердің көбейтіндісі түрінде анықталса, онда көлемдік фактордың нәтижелік белгіге тигізетін әсерін анықтау үшін сапалық көрсеткіштің базалық деңгейдегі, ал сапалық белгінің ықпалын анықтағанда, көлемдік фактордың ағымдағы деңгейдегі мәндерін алады.*

Жоғарыда келтірілген мысалда тауар айналымының өзгеруіне, яғни нәтижелік белгіге тауар бағаларының (сапалық фактордың) тигізген әсерін анықтау үшін тауарлардың физикалық көлемі (көлемдік фактор) ағымдағы деңгейде алынады, яғни бағаның агрегаттық индексі былай анықталады:

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1},$$

ал физикалық көлем индексі осы ережеге сәйкес мына формуламен есептеледі:

$$I_q = \frac{\sum p_0 q_1}{\sum p_0 q_0}.$$

Өндіріс шығындары, өнімнің өзіндік құны, шығарылған өнім көлемі индекстерінің арасында да тауар айналымының индекстеріне ұқсас байланыс болады:

$$I_z \times I_q = \frac{\sum z_1 q_1}{\sum z_0 q_1} \times \frac{\sum z_0 q_1}{\sum z_0 q_0} = \frac{\sum z_1 q_1}{\sum z_0 q_0} = I_{zq} \quad \text{немесе:} \quad I_z \times I_q = I_{zq}$$

Біз екі факторлы индекстер жүйесін, яғни нәтижелік белгіге екі фактор әсер ететін жағдайды қарастырдық. Іс жүзінде нәтижелік белгіге әсер ететін факторлар саны екіден көп болуы да мүмкін. Мұндай жағдайда жалпы индекс үш немесе одан да көп факторлық индекстерге жіктеледі, яғни құбылыстар байланысына факторлық талдау жасау үшін жүйелі тізбекті әдіс қолданылады.

Жүйелі тізбекті әдіс бойынша бірінші фактордың ықпалын анықтағанда басқа факторлардың барлығы ағымдағы деңгейде алынады. Екінші фактор индексін құрғанда бірінші фактор базалық деңгейде, ал қалған факторлар ағымдағы деңгейде болады. Үшінші фактордың индексін құру үшін бірінші және екінші факторлар базалық деңгейде, ал қалған факторлар ағымдағы деңгейде алынады. Мысал үшін нәтижелік белгі үш факторға байланысты болсын. Олардың байланысын шартты түрде былай белгілейік:

$$A = a \cdot b \cdot c$$

мұндағы A – нәтижелік белгі;

a, b, c – факторлық белгілер.

Мұндай жағдайда жүйелі тізбекті әдіс мынадай түрде болады:

$$I_A = \frac{a_1 \cdot b_1 \cdot c_1}{a_0 \cdot b_0 \cdot c_0} = I_a \cdot I_b \cdot I_c,$$

мұндағы $I_a = \frac{a_1 \cdot b_1 \cdot c_1}{a_0 \cdot b_1 \cdot c_1}; \quad I_b = \frac{a_0 \cdot b_1 \cdot c_1}{a_0 \cdot b_0 \cdot c_1}; \quad I_c = \frac{a_0 \cdot b_0 \cdot c_1}{a_0 \cdot b_0 \cdot c_0}$

Жоғарыдағы индекстердің өзара байланысын былай да жазуға болады:

$$I_A = I_a \cdot I_b \cdot I_c = \frac{a_1 \cdot b_1 \cdot c_1}{a_0 \cdot b_1 \cdot c_1} \times \frac{a_0 \cdot b_1 \cdot c_1}{a_0 \cdot b_0 \cdot c_1} \times \frac{a_0 \cdot b_0 \cdot c_1}{a_0 \cdot b_0 \cdot c_0} = \frac{a_1 \cdot b_1 \cdot c_1}{a_0 \cdot b_0 \cdot c_0}$$

Төрт немесе одан да көп фактор әсер ететін күрделі құбылыстар индекстері де осы үлгіге ұқсас құрылады.

Құбылыстар арасындағы байланысты корреляциялық-регрессиялық әдіспен талдау келесі параграфта қарастырылады.

9.3. Корреляциялық-регрессиялық талдау

Корреляциялық-регрессиялық әдіс – әлеуметтік-экономикалық құбылыстар арасындағы байланыстарды зерттейтін негізгі әдіс. Бұл әдістің көмегімен байланыс бағытын, тығыздығын, аналитикалық формасын, сонымен қатар нәтижелік белгіге факторлардың тигізетін тура ықпалын, басқа факторлар арқылы тигізетін жанама ықпалын анықтауға болады. Бірақ корреляциялық-регрессиялық әдісті кез-келген статистикалық мәліметтерге қолдана беруге болмайды. Корреляциялық-регрессиялық әдісті мынадай шарттар орындалғанда ғана қолданған жөн:

1. Негізгі жиынтықтан бақылауға алынатын бірліктер кездейсоқ іріктелуі қажет. Олай болмаған жағдайда бақылауға алынған бірліктер негізгі жиынтықты толық, дұрыс сипаттамайды, сондықтан корреляциялық-регрессиялық талдау нәтижесі де дұрыс қорытынды бермейді;
2. Бақылауға алынатын жиынтықтағы бірліктер саны факторлар санынан ең кемінде 5-6 рет артық болуы керек (8-10 есе артық болғаны тіптен жақсы), себебі корреляциялық байланыс үлкен сандар заңына сәйкес бақылау саны көп болғанда ғана байқалады;
3. Зерттелетін жиынтық біртекті болуы қажет. Егер жиынтық біртекті болмаса, регрессия теңдеуінің параметрлері дәл есептелмейді;
4. Жиынтықта факторлық және нәтижелік белгілердің таралуы қарапайым таралу заңына бағынуы керек. Бұл шарт регрессия теңдеуінің параметрлерін анықтау үшін ең кіші квадраттар әдісін қолдануға байланысты қойылады. Ең кіші квадраттар әдісі қарапайым таралу заңы сақталғанда жақсы нәтиже береді. Бірақ іс жүзінде бұл заң толық орындалмаса да ең кіші квадраттар әдісін қолдана береді;
5. Факторлық белгілер арасында функционалды байланыс болмау керек. Егер факторлық белгілер арасында айтарлықтай байланыс болса, регрессиялық үлгі сенімсіз болады, яғни талдау нәтижесі бұрмаланады.

Корреляциялық-регрессиялық талдау әдісі екі бағыттан тұрады:

- корреляциялық талдау;
- регрессиялық талдау .

Корреляциялық талдаудың міндеті болып факторлық белгілер мен нәтижелік белгі арасындағы байланыс тығыздығын, байланыс бағытын анықтау саналады.

Регрессиялық талдаудың міндеті – корреляциялық байланысты неғұрлым дәл сипаттайтын байланыс формасын, яғни регрессия теңдеуі түріндегі статистикалық үлгіні анықтау. Регрессия теңдеуін құрғанда ең негізгі мәселе болып үлгіге қосатын факторлар санын анықтау саналады. Регрессиялық үлгі сапасы, талдау қорытындыларының дәлдігі факторлар санына тікелей байланысты болады. Егер жанама, аса маңызды емес факторларды үлгіге қоспаса, зерттеу нәтижесінің сапалы болуының ықтималдылығы жоғары болады, бірақ аз өлшемді үлгі құбылысты толық сипаттамауы да мүмкін. Сондықтан регрессия үлгісінде нәтижелік белгіге айтарлықтай әсер ететін факторлар толық болғаны жөн. Үлгіге қосатын факторларды анықтау кезінде ең алдымен көрсеткіштер арасындағы

байланыстың себеп-салдарына жеткілікті мөлшерде көңіл бөлу керек, өйткені олар зерттелетін құбылыстың мағынасын түсінуге мүмкіндік береді. Сол сияқты факторларды таңдауға аналитикалық топтастыру, параллель қатарларды салыстыру, график әдістері де көмектеседі. Олардың көмегімен құбылыстар арасында байланыстың бар-жоғын, байланыс бағытын, формасын анықтауға болады.

Жалпы жағдайда корреляциялық-регрессиялық талдау жұмыстары мынадай реттілікпен жүргізіледі:

- 1) нәтижелік белгіге әсер ететін көптеген факторлардың ішінен ең маңыздыларын іріктеп алу;
- 2) факторлық және нәтижелік белгілер арасындағы байланыс тығыздығын, байланыс бағытын анықтау;
- 3) нәтижелік белгінің факторларға тәуелділігін сипаттайтын математикалық функцияны табу;
- 4) анықталған регрессия үлгісін бағалау;
- 5) құбылыстар дамуына болжам жасау.

Осы аталған жұмыстардың алғашқы екеуі корреляциялық талдау, ал қалған үшеуі регрессиялық талдау міндеттері болып саналады.

Құбылыстар арасындағы байланысты зерттеу үшін алынған үлгідегі факторлар санына байланысты корреляциялық байланыс жұптық және көптік корреляция деп бөлінеді. Ал регрессия теңдеуінің аналитикалық формасына байланысты корреляциялық байланыс түзу сызықты және қисық сызықты болады.

Корреляциялық-регрессиялық талдау әдісі экономиканың әр түрлі саласында, әлеуметтік ғылымдарда (әсіресе психология, әлеуметтану), басқа да салаларда кеңінен қолданылады.

9.4. Жұптық корреляция

Жұптық корреляция бір факторлық белгі мен нәтижелік белгі байланысын қарастырады. Белгілер арасындағы байланыс түзу сызықты немесе қисық сызықты болуы мүмкін. Түзу сызықты байланыс түзудің теңдеуі, ал қисық сызықты байланыс парабола, гипербола, т.б. қисықтардың теңдеулері арқылы өрнектеледі және бұл теңдеулер регрессия сызығы деп аталады.

Регрессия сызығы – бұл факторлық белгі ($\tilde{\sigma}$) мәндерін нәтижелік белгінің орташа мәнімен ($\tilde{\sigma}$) байланыстыратын функция. Жұптық регрессия теңдеуін құрғанда көбінесе мына теңдеулерді қолданады:

1. Түзудің теңдеуі: $\tilde{\sigma}_x = a_0 + a_1 \cdot x$;
2. Парабола теңдеуі: $\tilde{\sigma}_x = a_0 + a_1 \cdot x + a_2 \cdot x^2$;
3. Гипербола теңдеуі: $\tilde{\sigma}_x = a_0 + \frac{a_1}{x}$;
4. Дәрежелік функция теңдеуі: $\tilde{\sigma}_x = a_0 \cdot x^{a_1}$;
5. Көрсеткіштік функция теңдеуі: $\tilde{\sigma}_x = a_0 \cdot a_1^{\tilde{\sigma}}$;

6. Логарифмдік функция теңдеуі: $\hat{\sigma}_x = a_0 + a_1 \cdot \lg x$,

мұндағы $\hat{\sigma}_x$ – нәтижелік белгінің теориялық мәні;

a_0 – регрессия теңдеуінің бос мүшесі;

a_1 – регрессия коэффициенті. Ол факторлық белгінің өзгеруіне байланысты нәтижелік белгінің орташа мәні қалай өзгергенін көрсетеді.

a_1 параметрін пайдаланып, икемділік коэффициентін мына формуламен анықтайды:

$$\hat{E} = a_1 \cdot \frac{\bar{\sigma}}{\bar{\sigma}},$$

мұндағы \hat{E} – икемділік коэффициенті;

$\bar{\sigma}$ – факторлық белгі мәндерінің орташасы;

$\bar{\sigma}$ – нәтижелік белгінің орташа шамасы.

Икемділік коэффициенті факторлық белгі 1%-ға өзгергенде нәтижелік белгі қалай өзгертінін көрсетеді.

Регрессия теңдеуі үшін қандай функцияны таңдау керектігі зеттеу мәліметтеріне байланысты болады. Егер факторлық белгі мен нәтижелік белгі жуықтап алғанда арифметикалық прогрессия бойынша өзгерсе, онда бұл олардың арасында сызықтық байланыс бар екенін көрсетеді. Ал белгілер арасында кері байланыс бар екені байқалса, регрессия сызығы гиперболоа теңдеуі арқылы өрнектеледі. Нәтижелік белгі арифметикалық прогрессиямен, ал факторлық белгі одан әлде қайда тез өзгерген жағдайда регрессия теңдеуі парабола немесе дәрежелік функция түрінде беріледі. Регрессия теңдеуінің параметрлерін анықтау үшін ең кіші квадраттар әдісі қолданылады. Бұл әдіс бойынша a_0 , a_1 параметрлері нәтижелік белгінің нақты мәндері мен теориялық мәндерінің ауытқуларының квадраттарының қосындысы ең аз шама болатындай, яғни мына шарт орындалатындай мәндерді қабылдауы қажет:

$$S = \sum (y_i - \hat{y}_x)^2 \rightarrow \min ,$$

мұндағы \hat{y}_i – нәтижелік белгінің нақты мәндері;

$\bar{\sigma}_x$ – нәтижелік белгінің теориялық мәндері.

S функциясының минимумын анықтау үшін осы функцияның a_0 , a_1 айнымалылары бойынша дербес туындыларын нольге теңестіру керек.

$$\frac{\partial S}{\partial a_0} = 0; \quad \frac{\partial S}{\partial a_1} = 0.$$

Соның нәтижесінде сызықтық жұп регрессияның параметрлерін анықтау үшін мынадай теңдеулер жүйесі алынады:

$$\begin{cases} na_0 + a_1 \sum \bar{\sigma} = \sum y \\ a_0 \sum \bar{\sigma} + a_1 \sum \bar{\sigma}^2 = \sum y\bar{\sigma} \end{cases}$$

мұндағы n – зерттелетін жиынтық көлемі.

Егер осы теңдеулер жүйесіндегі бірінші теңдеуді n -ге бөлсе, онда мынадай өрнекті аламыз:

$$\hat{a}_0 + \hat{a}_1 \cdot \bar{\delta} = \bar{\delta} \quad \text{немесе:} \quad \hat{a}_0 = \bar{\delta} - \hat{a}_1 \cdot \bar{\delta}$$

Әдетте регрессия теңдеуінің бос мүшесі \hat{a}_0 -ді соңғы формула бойынша есептейді. Жоғарыдағы теңдеулер жүйесін шешу және түрлендіру жасау нәтижесінде сызықтық жұп регрессия үшін \hat{a}_1 параметрін анықтайтын формуланы қорытып шығаруға болады:

$$\hat{a}_1 = \frac{\overline{\delta\delta} - \bar{\delta} \cdot \bar{\delta}}{\bar{\delta^2} - \bar{\delta}^2}$$

\hat{a}_0, \hat{a}_1 параметрлерінің мәндерін регрессия теңдеуіне қойып, нәтижелік белгі мәндерін анықтайды. Енді сызықтық регрессия теңдеуін құруға мысал келтірейік.

Мысал. 10 кәсіпорын көрсеткіштері туралы мынадай мәліметтер берілген:

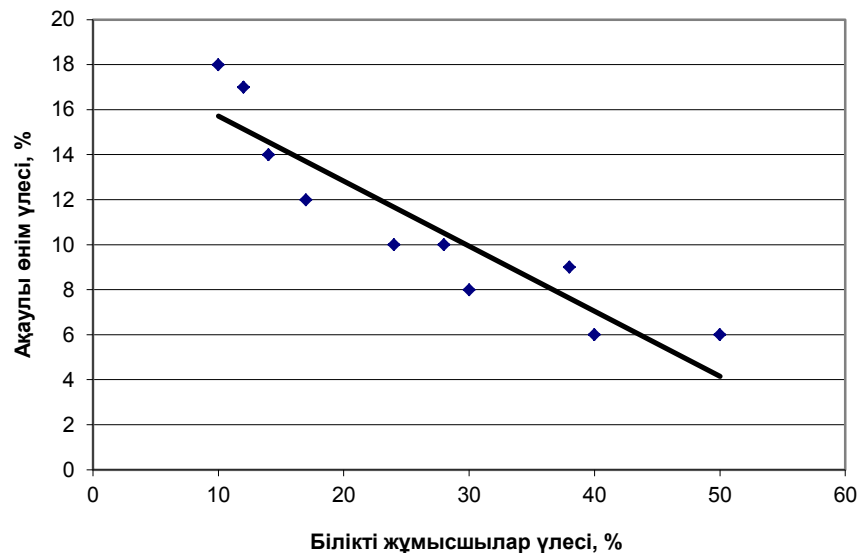
9.4-кесте

Кәсіпорындардың білікті жұмысшылар және ақаулы өнім үлесі бойынша бөлінуі

Бастапқы мәліметтер			Есептелген көрсеткіштер					
Кәсіпорын №	Білікті жұмысшылардың үлесі, % $\bar{\delta}$	Ақаулы өнім үлесі, % δ	$\bar{\delta\delta}$	$\bar{\delta^2}$	$\bar{\delta^2}$	$\bar{\delta}_x$	$\delta - \bar{\delta}_\delta$	$(\delta - \bar{\delta}_\delta)^2$
1	10	18	180	100	324	15,727	2,273	5,167
2	12	17	204	144	289	15,147	1,853	434
3	14	14	196	196	196	14,567	-0,567	0,321
4	17	12	204	289	144	13,697	-1,697	2,88
5	24	10	240	576	100	11,667	-1,667	2,779
6	28	10	280	784	100	10,507	-0,507	0,257
7	30	8	240	900	64	9,927	-1,927	3,713
8	38	9	342	1444	81	7,607	1,393	1,94
9	40	6	240	1600	36	7,027	-1,027	1,055
10	50	6	300	2500	36	4,127	1,873	3,508
Барлығы	263	110	2426	8533	1370	110	-	25,054

Бастапқы берілген мәліметтер бойынша шығарылған ақаулы өнім мен жұмысшы біліктілігі арасындағы байланыстың регрессиялық үлгісін құру керек.

Шешуі: Байланыс формасын анықтау үшін график әдісін қолданамыз. Ол үшін δ, δ мәндеріне сәйкес келетін нүктелерді графикте белгілеп, корреляция өрісін бейнелейміз. Бұл мысалда корреляция өрісі ақаулы өнімдер мен жұмысшы біліктілігі арасында сызықтық кері байланыс бар екенін айқын көрсетеді.



9.3-сурет. Ақаулы өнімдер мен жұмысшы біліктілігі арасындағы байланыс.

Сондықтан регрессия теңдеуі түзу сызықты болады.

$$\hat{\sigma}_x = a_0 + a_1 \cdot x.$$

Түзу параметрлерін жоғарыда келтірілген формулалар және 10.4-кестедегі есептеулер бойынша анықтаймыз.

$$\hat{a}_1 = \frac{\bar{\sigma\sigma} - \bar{\sigma} \cdot \bar{\sigma}}{\bar{\sigma^2} - \bar{\sigma}^2} = \frac{\frac{2426}{10} - \frac{263}{10} \cdot \frac{110}{10}}{\frac{8533}{10} - \left(\frac{263}{10}\right)^2} = \frac{-46,7}{161,6} = -0,29.$$

$$\hat{a}_0 = \bar{\sigma} - \hat{a}_1 \cdot \bar{\sigma} = \frac{110}{10} + 0,29 \cdot \frac{263}{10} = 18,627,$$

яғни, регрессия теңдеуі мынадай түрде болады:

$$\hat{\sigma}_x = 18,627 - 0,29 \cdot x.$$

Қарастырылған мысалда $\hat{a}_1 < 0$ болды, бұл параметр зерттеліп отырған көрсеткіштер арасында кері байланыс бар екенін және білікті жұмысшылардың үлесі 1 пайызға артқанда ақаулы өнімдердің үлесі 0,29 пайызға кемитінін көрсетеді.

Регрессиялық үлгіні іс жүзінде қолдану үшін оның нақты статистикалық мәліметтермен барабар болуының маңызы зор. Әдетте корреляциялық-регрессиялық талдау көлемі бойынша шектеулі жиынтықтарда жүргізіледі.

Сондықтан регрессия теңдеуінің параметрлері, корреляция коэффициенті кездейсоқ факторларды әсерінен дұрыс анықталмауы мүмкін. Негізгі жиынтыққа бұл көрсеткіштердің қаншалықты тән екенін білу үшін **статистикалық үлгі барабарлығы** тексеріледі. Егер регрессия теңдеуінің параметрлері немесе байланыс тығыздығын анықтайтын көрсеткіш мәнді болса, онда регрессия үлгісі де мәнді деп саналады. Сондықтан статистикалық үлгінің барабарлығын тексеру үшін теңдеу параметрлерінің немесе байланыс тығыздығы көрсеткішінің мәнділігін анықтайды. Бақылау объектісінің

бірліктерінің саны 30-дан аспаса, онда регрессия теңдеуінің әр коэффициентінің **мәнділігі** Стьюденттің t -критерийі бойынша тексеріледі. Ол үшін t -критерийдің нақты (есептелетін) мәні мына формуламен анықталады:

$$\hat{a}_0 \text{ параметрі үшін: } t_{a_0} = |a_0| \cdot \frac{\sqrt{n-2}}{\sigma_{\hat{a}_0}};$$

$$\hat{a}_1 \text{ параметрі үшін: } t_{a_1} = |a_1| \cdot \frac{\sqrt{n-2}}{\sigma_{\hat{a}_1}} \cdot \sigma_{\hat{\delta}}.$$

мұндағы n – бақылау көлемі;

t_{a_0}, t_{a_1} – \hat{a}_0, \hat{a}_1 параметрлері үшін есептелген t -критерийдің мәндері;

$\sigma_{\hat{a}_0, \hat{a}_1}$ – \hat{a}_0, \hat{a}_1 нәтижелік белгісінің нақты мәндерінің $\hat{\delta}_x$ теориялық мәндерден ауытқуының квадраттық орташасы. Бұл көрсеткішті мына формуламен есептейді:

$$\sigma_{\hat{a}_0, \hat{a}_1} = \sqrt{\frac{\sum (y - \hat{y}_x)^2}{n}};$$

$\sigma_{\hat{\delta}}$ – $\hat{\delta}$ факторлық белгісінің жалпы орташадан ауытқуының квадраттық орташасы (орташа квадраттық ауытқу).

$$\sigma_{\hat{\delta}} = \sqrt{\frac{\sum (\delta - \bar{\delta})^2}{n}} \quad \text{немесе:} \quad \sigma_{\hat{\delta}} = \sqrt{\frac{\sum \delta^2}{n} - \left(\frac{\sum \delta}{n}\right)^2}.$$

Жоғарыдағы формулалар бойынша есептелген t_{a_0}, t_{a_1} мәндері Стьюдент кестесіндегі мәнмен салыстырылады. Егер t -критерийінің нақты, яғни есептелген мәні кестедегі мәнінен үлкен болса ($t_{\text{іәәәәә}} > t_{\text{әәәәәә}}$), онда регрессия теңдеуінің параметрі мәнді деп саналады. Енді 10.4-кесте мәліметтері бойынша анықталған $\hat{\delta}_x = 18,627 - 0,29 \cdot x$ регрессия теңдеуі параметрлерінің мәнділігін тексерейік. Ол үшін алдымен $\sigma_{\hat{a}_0, \hat{a}_1}$ көрсеткішін есептейміз:

$$\sigma_{\hat{a}_0, \hat{a}_1} = \sqrt{\frac{\sum (y - \hat{y}_x)^2}{n}} = \sqrt{\frac{25,054}{10}} = 1,58.$$

Енді $\sigma_{\hat{\delta}}$ көрсеткішін анықтаймыз:

$$\sigma_{\hat{\delta}} = \sqrt{\frac{\sum \delta^2}{n} - \left(\frac{\sum \delta}{n}\right)^2} = \sqrt{\frac{8533}{10} - \left(\frac{263}{10}\right)^2} = \sqrt{161,61} = 12,71.$$

Осыдан кейін t -критерийдің нақты мәндерін әр параметр үшін есептеуге болады:

$$t_{a_0} = |a_0| \cdot \frac{\sqrt{n-2}}{\sigma_{\hat{a}_0, \hat{a}_1}} = 18,627 \cdot \frac{2,828}{1,58} = 33,3;$$

$$t_{a_1} = |a_1| \cdot \frac{\sqrt{n-2}}{\sigma_{\hat{a}_0, \hat{a}_1}} \cdot \sigma_{\hat{\delta}} = 0,29 \cdot \frac{2,828}{1,582} \cdot 12,71 = 6,59.$$

Еркіндік дәрежесі $\hat{e} = n - 2 = 8$, ал мәнділік деңгейі $\alpha = 0,05$ болғанда Стьюдент таралуының кестесі бойынша t -критерийдің сыни мәні 2,31-ге тең,

яғни, регрессия теңдеуінің екі параметрі үшін де $t_{\text{іәәә}} > t_{\text{әәәә}}$ шарты орындалады. Бұл жоғарыда анықталған регрессия теңдеуінің параметрлері кездейсоқ шамалар емес екендігін, олардың статистикалық мәні бар екендігін көрсетеді.

Егер құбылыстар арасындағы байланыс қисық сызықты болса және ол парабола теңдеуімен өрнектелсе, онда теңдеудің параметрлерін анықтау үшін мынадай теңдеулер жүйесі құрылады:

$$\begin{cases} na_0 + a_1 \sum \tilde{\sigma} + a_2 \sum \tilde{\sigma}^2 = \sum y \\ a_0 \sum \tilde{\sigma} + a_1 \sum \tilde{\sigma}^2 + a_2 \sum \tilde{\sigma}^3 = \sum y\tilde{\sigma} \\ a_0 \sum \tilde{\sigma}^2 + a_1 \sum \tilde{\sigma}^3 + a_2 \sum \tilde{\sigma}^4 = \sum y\tilde{\sigma}^2 \end{cases}$$

Ал гипербола теңдеуінің параметрлері төмендегі жүйе бойынша анықталады:

$$\begin{cases} na_0 + a_1 \sum \frac{1}{\tilde{\sigma}} = \sum y \\ a_0 \sum \frac{1}{\tilde{\sigma}} + a_1 \sum \frac{1}{\tilde{\sigma}^2} = \sum \frac{y}{\tilde{\sigma}} \end{cases}$$

9.5 Құбылыстар байланысының тығыздығын анықтау

Құбылыстар арасындағы байланысты зерттегенде байланыс бағытын, тығыздығын анықтаудың маңызы зор. Байланыс тығыздығы нәтижелік белгінің факторлық белгіге қаншалықты тәуелді екенін сипаттайды. Әр түрлі факторлық белгілер мен нәтижелік белгі арасындағы байланыс тығыздығының көрсеткіштерін салыстыру арқылы нақты бір жағдайда нәтижелік белгіге қандай факторлық белгілердің ықпалы басым болатынын айқындауға болады. Байланыс тығыздығын анықтау үшін статистикада әр түрлі көрсеткіштер қолданылады. Енді сол көрсеткіштерге тоқталайық.

1. Фехнер коэффициенті – байланыс тығыздығын сипаттайтын ең қарапайым көрсеткіш. Оны неміс ғалымы Г. Фехнер ұсынған. Фехнер коэффициенті факторлық және нәтижелік белгілердің орташа шамадан ауытқуларының таңбасы негізінде анықталады. Оны есептеу мынадай реттілікпен жүргізіледі:

1. Факторлық және нәтижелік белгілердің орташа шамалары есептеледі;
2. Белгі варианттары үшін мынадай ауытқулар $(x_i - \bar{x}), (y_i - \bar{y})$ есептеледі;
3. $(x_i - \bar{x}), (y_i - \bar{y})$ ауытқуларының таңбаларының сәйкес келу, келмеуі анықталады. Мұндай жағдайда таңбалары сәйкес келген ауытқуларды \hat{a} әрпімен, ал сәйкес келмеген ауытқуларды \hat{a} әрпімен белгілейді;
4. Фехнер коэффициенті мына формула бойынша есептеледі:

$$\hat{E}_{\sigma} = \frac{\sum \hat{a} - \sum \hat{a}}{\sum \hat{a} + \sum \hat{a}}$$

Бұл көрсеткіш мына аралықтағы мәндерді қабылдайды:

$$-1 \leq \hat{E}_{\sigma} \leq +1$$

Фехнер коэффициентінің мәні 1-ге жақын болса, ол көрсеткіштер бір-бірімен тығыз байланысты дегенді білдіреді. Енді 10.4-кесте мәліметтері бойынша Фехнер коэффициентін есептейік. Ол үшін $\bar{o} = 26,3$, $\bar{\delta} = 11,0$ ескеріп, мынадай көмекші кестені құрамыз (9.5-кесте):

9.5-кесте

Фехнер коэффициентін анықтауға қажетті көмекші кесте

№	δ	o	$(x_i - \bar{x})$ ауытқуының таңбасы	$(o_i - \bar{o})$ ауытқуының таңбасы	Ауытқу таңбасының сәйкестігін белгілеу
1	10	18	-	+	\hat{a}
2	12	17	-	+	\hat{a}
3	14	14	-	+	\hat{a}
4	17	12	-	+	\hat{a}
5	24	10	-	-	\hat{a}
6	28	10	+	-	\hat{a}
7	30	8	+	-	\hat{a}
8	38	9	+	-	\hat{a}
9	40	6	+	-	\hat{a}
10	50	6	+	-	\hat{a}
Барлығы	263	110	-	-	-

Көмекші кесте мәліметтерін пайдаланып, Фехнер коэффициентін есептейміз:

$$\hat{E}_o = \frac{\sum \hat{a} - \sum \hat{a}}{\sum \hat{a} + \sum \hat{a}} = \frac{1-9}{1+9} = -0,8.$$

Есептеу нәтижесі ақаулы өнімдер мен жұмысшы біліктілігі көрсеткіштерінің арасында кері байланыс бар екенін ($\hat{E}_o < 0$) және ол байланыстың тығыз екенін көрсетеді.

2. Сызықтық корреляция коэффициенті құбылыстар арасында сызықтық байланыс болғанда байланыс тығыздығын анықтау үшін қолданылады. Оны мына формуламен есептейді:

$$r = \frac{\overline{\delta o} - \bar{\delta} \cdot \bar{o}}{\sigma_x \cdot \sigma_y},$$

мұндағы r – сызықтық корреляция коэффициенті;

$\sigma_\delta - \bar{\delta}$ факторлық белгісінің орташа квадраттық ауытқуы;

$\sigma_o - \bar{o}$ белгісінің орташа квадраттық ауытқуы.

$$\sigma_\delta = \sqrt{\frac{\sum (\delta - \bar{\delta})^2}{n}} \quad \text{немесе:} \quad \sigma_\delta = \sqrt{\frac{\sum \delta^2}{n} - \left(\frac{\sum \delta}{n}\right)^2}$$

Сызықтық корреляция коэффициентінің жоғарыдағы формуласын өзгертіп, мынадай түрде жазуға болады:

$$r = \frac{\sum (x - \bar{x}) \cdot (y - \bar{y})}{n \cdot \sigma_x \cdot \sigma_y},$$

мұндағы n – бақылау көлемі.

Сызықтық корреляция коэффициентін ағылшын математигі Пирсонның атымен **Пирсон коэффициенті** деп те атайды. Регрессия теңдеуінің a_1 параметрі мен Пирсон коэффициентінің арасындағы байланысты мына формула арқылы көрсетуге болады:

$$r = a_1 \cdot \frac{\sigma_x}{\sigma_y}$$

Сызықтық корреляция коэффициенті мына аралықта өзгереді:

$$-1 \leq r \leq +1.$$

Егер r теріс мәнді қабылдаса, онда зерттелетін құбылыстар арасында кері байланыс, ал оң мәнді қабылдаса – тура байланыс болады. $r = 0$ болғанда, құбылыстар арасында байланыс болмайды. Сызықтық корреляция коэффициентінің мәні 1-ге жақындаған сайын көрсеткіштер байланысының тығыздығы артады.

Әдетте құбылыстар байланысының тығыздығын анықтағанда Чеддок шкаласы деп аталатын мынадай шкаланы пайдаланады:

9.6-кесте

Байланыс тығыздығын анықтау шкаласы

Тығыздықты анықтайтын көрсеткіш мәні	$r = 0$	$0 < r < 0.3$	$0.3 \leq r < 0.5$	$0.5 \leq r < 0.7$	$0.7 \leq r < 0.9$	$0.9 \leq r < 1$	$r = \pm 1$
Байланыс сипаты	байланыс жоқ	әлсіз байланыс	орташа байланыс	айқын байланыс	тығыз байланыс	өте тығыз байланыс	функционалдық байланыс

Енді 9.4-кесте мәліметтері бойынша сызықтық корреляция коэффициентін есептейік. Алдыңғы есептеулер бойынша: $\bar{\delta\delta} = 242,6$; $\bar{\delta} = 26,3$; $\bar{\delta} = 11,0$;

$\sum \delta^2 = 1370$; $\sigma_{\delta} = 12,71$ екені белгілі. Ал сызықтық корреляция коэффициенті формуласының бөліміндегі δ белгісінің орташа квадраттық ауытқуын мына формуламен анықтаймыз:

$$\sigma_{\delta} = \sqrt{\frac{\sum \delta^2}{n} - \left(\frac{\sum \delta}{n}\right)^2} = \sqrt{\frac{1370}{10} - 121} = 4,0.$$

Осы есептеу нәтижелерін төмендегі формулаға қоямыз:

$$r = \frac{\bar{\delta\delta} - \bar{\delta} \cdot \bar{\delta}}{\sigma_x \cdot \sigma_y} = \frac{242,6 - 26,3 \cdot 11,0}{12,71 \cdot 4,0} = \frac{-46,7}{50,84} = -0,919.$$

Бұл ақаулы өнім үлесі мен жұмысшы біліктілігі арасында кері байланыс бар екенін және ол байланыстың өте тығыз екенін көрсетеді.

Сызықтық корреляция коэффициентінің квадраты **детерминация коэффициенті** деп аталады.

$$R = r^2;$$

мұндағы R – детерминация коэффициенті. Бұл көрсеткіш нәтижелік белгі вариациясына факторлық белгі вариациясы қалай әсер ететінін көрсетеді. Біздің мысалда: $R = r^2 = (-0,919^2) = 0,845$, яғни, ақаулы өнім вариациясының 84,5 пайызы жұмысшы біліктігінің өзгеруінің әсерінен болады.

Сызықтық регрессияның мәнділігін сызықтық корреляция коэффициенті бойынша да анықтауға болады. Сызықтық корреляция коэффициенті үшін t -критерийдің есептелген, яғни нақты мәні кестедегі мәннен артық болса ($t_{r\hat{\epsilon}\hat{\delta}} > t_{r\hat{\alpha}\hat{\beta}}$), онда корреляция коэффициенті мәнді деп саналады. Мұндай жағдайда t -критерийдің нақты мәні мына формуламен анықталады:

$$t_{r\hat{\epsilon}\hat{\delta}} = |r| \cdot \sqrt{\frac{n-2}{1-r^2}},$$

мұндағы $t_{r\hat{\epsilon}\hat{\delta}}$ – корреляция коэффициенті үшін t -критерийдің есептелген нақты мәні;

n – бақылау көлемі.

Енді жоғарыдағы мысал үшін бұл көрсеткіштің мәнін анықтайық.

$$t_{r\hat{\epsilon}\hat{\delta}} = |r| \cdot \sqrt{\frac{n-2}{1-r^2}} = 0,919 \cdot \sqrt{\frac{10-2}{1-0,845}} = 6,6.$$

Еркіндік дәрежесі $\hat{\epsilon} = n - 2 = 8$, ал мәнділік деңгейі $\alpha = 0,05$ болғанда Стьюдент таралуының кестесі бойынша t -критерийдің сыни мәні 2,31-ге тең, яғни $t_{r\hat{\epsilon}\hat{\delta}} > t_{r\hat{\alpha}\hat{\beta}}$ шарты орындалады. Сондықтан сызықтық корреляция коэффициенті кездейсоқ шама емес, ал $\hat{\delta}_x = 18,627 - 0,29 \cdot x$ теңдеуі статистикалық тұрғыдан мәнді деп 0,95 ықтималдылығымен айтуға болады.

3. Корреляциялық қатынас – құбылыстар арасындағы байланыс тығыздығын анықтау үшін жиі қолданатын көрсеткіштердің бірі. Корреляциялық қатынасты сызықтық байланыс үшін де, сызықтық емес байланыс үшін де қолдануға болады. Корреляциялық қатынасты эмпириялық және теориялық корреляциялық қатынас деп екіге бөледі.

Эмпириялық корреляциялық қатынас топтастырылған мәліметтер бойынша анықталады. Оны есептеу үшін мына формула қолданылады:

$$\eta_{\hat{y}} = \sqrt{\frac{\delta^2}{\sigma^2}},$$

мұндағы $\eta_{\hat{y}}$ – эмпириялық корреляциялық қатынас;

δ^2 – топаралық дисперсия (бұл көрсеткіш нәтижелік белгінің топтық орташаларының жалпы орташадан ауытқуын сипаттайды);

σ^2 – жалпы дисперсия.

Теориялық корреляциялық қатынас нәтижелік белгінің тегістелген, яғни регрессия теңдеуі бойынша анықталған мәндері арқылы есептеледі:

$$\eta_{\hat{\delta}} = \sqrt{\frac{\delta^2}{\sigma^2}},$$

мұндағы $\eta_{\hat{\delta}}$ – теориялық корреляциялық қатынас;

δ^2 – нәтижелік белгінің тегістелген мәндерінің (δ) дисперсиясы. Бұл көрсеткішті мына формуламен анықтайды:

$$\delta^2 = \frac{\sum(\delta - \bar{\delta})^2}{n}$$

σ^2 – нәтижелік белгінің эмпириялық (нақты) мәндерінің дисперсиясы, яғни:

$$\sigma^2 = \frac{\sum(\delta - \bar{\delta})^2}{n}.$$

Осы формулаларды ескеріп, теориялық корреляциялық қатынасты былай да анықтауға болады:

$$\eta_{\delta} = \sqrt{\frac{\sum(\delta - \bar{\delta})^2}{\sum(\delta - \bar{\delta})^2}}.$$

Корреляциялық қатынасты есептеу дисперсияларды қосу ережесіне негізделген. Ол ереже бойынша жалпы дисперсия топтық дисперсиялардың орташасы мен топаралық дисперсияның қосындысына тең болады:

$$\sigma^2 = \bar{\sigma}_i^2 + \delta^2$$

Топаралық дисперсия топтастыру негізі болып саналатын фактордың әсерінен болған вариацияны, топтық дисперсия топтастыру негізі болған белгіден басқа факторлардың, яғни есепке алынбаған факторлардың әсерінен болған белгі вариациясын, ал жалпы дисперсия барлық факторлардың әсерінен жиынтықтықта болған вариацияны сипаттайды.

Топтық дисперсиялардың орташасын қалдық дисперсия деп те атайды және оны былай анықтайды:

$$\bar{\sigma}_i^2 = \sigma_{\text{қалдық}}^2 = \frac{\sum(\delta - \bar{\delta})^2}{n}$$

Олай болса теориялық корреляциялық қатынастың формуласын мынадай түрде жазуға болады:

$$\eta_{\delta} = \sqrt{\frac{\delta^2}{\sigma^2}} = \sqrt{\frac{\sigma^2 - \sigma_{\text{қалдық}}^2}{\sigma^2}} = \sqrt{1 - \frac{\sigma_{\text{қалдық}}^2}{\sigma^2}},$$

немесе:

$$\eta_{\delta} = \sqrt{1 - \frac{\sum(\delta - \bar{\delta})^2}{\sum(\delta - \bar{\delta})^2}}.$$

Корреляциялық қатынастың абсолютті мәні 0 мен 1 аралығында өзгереді. Бұл көрсеткіш бойынша құбылыстар байланысының тығыздығын анықтағанда да Чеддок шкаласын пайдалануға болады.

Енді 9.4-кесте мәліметтері бойынша теориялық корреляциялық қатынасты анықтайық. Ол үшін ең алдымен қосымша есептеулерді көрсету үшін көмекші кесте құрамыз.

9.7-кесте

Теориялық корреляциялық қатынасты анықтауға қажетті есептеулер

№	δ	$\bar{\delta}$	$\delta - \bar{\delta}$	$(\delta - \bar{\delta})^2$	$\delta - \bar{\delta}$	$(\delta - \bar{\delta})^2$
1	18	15,757	7	49	4,727	22,629

2	17	15,147	6	36	4,147	17,198
3	14	14,567	3	9	3,567	12,723
4	12	13,697	1	1	2,697	7,274
5	10	11,667	-1	1	0,667	0,445
6	10	10,507	-1	1	-0,493	0,243
7	8	9,927	-3	9	-1,073	1,151
8	9	7,607	-2	4	-3,393	11,512
9	6	7,027	-5	25	-3,973	15,785
10	6	4,127	-5	25	-6,873	47,238
Барлығы	110	110	-	160	-	135,914

Теориялық корреляциялық қатынасты осы есептеулерді пайдаланып анықтаймыз:

$$\eta_{\delta} = \sqrt{\frac{\sum(\delta - \bar{\delta})^2}{\sum(\delta - \bar{\delta})^2}} = \sqrt{\frac{135,914}{160}} = \sqrt{0,849} = 0,921,$$

Осы көрсеткішті былай да есептеуге болады:

$$\eta_{\delta} = \sqrt{1 - \frac{\sum(\delta - \bar{\delta})^2}{\sum(\delta - \bar{\delta})^2}} = \sqrt{1 - \frac{25,054}{160}} = \sqrt{0,843} = 0,918.$$

Бұл зерттеліп отырған көрсеткіштер арасында өте тығыз бар екенін білдіреді.

4. Рангтер корреляциясының коэффициенті факторлық және нәтижелік белгі рангтерінің айырмасы арқылы анықталады. Ранг деп өсу немесе кему ретімен орналасқан көрсеткіш мәндерінің реттік нөмірін айтады. Зерттелетін белгілердің мәндерін нөмірлеу принципі құбылыстар байланыстарын анықтаудың параметрлік емес әдістерінің негізі болып саналады. Параметрлік емес әдістерді қолдану арқылы сандық көрсеткіштердің де, сапалық көрсеткіштердің де арасындағы байланыс тығыздығын анықтауға болады. Олай болса рангтер корреляциясының коэффициенті сандық көрсеткіштер үшін де, сапалық көрсеткіштер үшін де қолданылады. Бұл көрсеткішті **Спирмен коэффициенті** деп те атайды және ол мына формуламен есептеледі:

$$\rho = 1 - \frac{6 \cdot \sum d^2}{n \cdot (n^2 - 1)},$$

мұндағы ρ – рангтер корреляциясының коэффициенті (Спирмен коэффициенті);

n – бақылау саны ;

d^2 – рангтер айырмасының квадраты.

$$d^2 = (R_x - R_y)^2;$$

R_x – факторлық белгі рангі;

R_y – нәтижелік белгі рангі.

Рангтер корреляциясы коэффициентін есептеу алгоритмі төмендегідей:

1. Ең алдымен факторлық белгі рангтерін (R_x), содан кейін нәтижелік белгі рангтерін (R_y) анықтайды;

2. Рангтер айырмасы есептеледі ($d = R_x - R_y$);
3. Рангтер айырмасы квадраттарының қосындысы анықталады ($\sum d^2$);
4. Формула бойынша ρ коэффициенті есептеледі.

Бұл көрсеткіш $[-1; +1]$ интервалындағы кез-келген мәнді қабылдайды. Рангтер корреляциясының коэффициентінің мәнділігі Стьюденттің t -критерийі арқылы тексеріледі. Ол үшін t -критерийдің нақты мәнін мына формуламен анықтайды:

$$t_{\rho \text{ i} \grave{\text{a}} \text{e} \grave{\text{o}} \grave{\text{u}}} = |\rho| \cdot \sqrt{\frac{n-2}{1-\rho^2}},$$

мұндағы $t_{\rho \text{ i} \grave{\text{a}} \text{e} \grave{\text{o}} \grave{\text{u}}}$ – рангтер корреляциясы коэффициенті үшін t -критерийдің есептелген нақты мәні;
 n – бақылау көлемі.

$t_{\rho \text{ i} \grave{\text{a}} \text{e} \grave{\text{o}} \grave{\text{u}}} > t_{\rho \text{ e} \grave{\text{a}} \text{n} \grave{\text{o}} \grave{\text{a}}}$ шарты орындалған жағдайда рангтер корреляциясы коэффициенті мәнді деп саналады.

Рангтер корреляциясының коэффициенті медицинада жиі қолданады. Мысалы, әр түрлі факторлардың адам денсаулығына әсерін анықтау үшін Спирмен коэффициентін пайдалануға болады. Енді осы көрсеткішті есептеуге мысал келтірейік.

Мысал. Фермерлік шаруашылықтардан түскен жалпы түсім мен пайдаланылған минералды тыңайтқыштар туралы мынадай мәліметтер берілген (9.8-кестенің 1,2-бағандары):

9.8-кесте

Спирмен коэффициентін есептеу

Шаруашылық №	1 гектар жерге жұмсалған тыңайтқыш, кг \bar{o}	Жалпы түсім, мың тонна \acute{o}	Рангтер		Рангтер айырмасы $d = R_x - R_y$	d^2
			R_x	R_o		
А	1	2	3	4	5	6
1	20	4,7	10	9	1	1
2	25	4,5	6	10	-4	16
3	29	5,5	5	4	1	1
4	22	5,1	8	7	1	1
5	21	4,8	9	8	1	1
6	24	5,2	7	6	1	1
7	34	7,0	2	3	-1	1
8	30	5,3	4	5	-1	1
9	35	7,5	1	2	-1	1
10	32	7,7	3	1	2	4
Барлығы	-	-	-	-	-	28

Осы мәліметтер бойынша жалпы түсім мен қолданылған тыңайтқыш көлемі арасындағы байланыс тығыздығын анықтау керек.

Шешуі: Алдымен 1 гектар жерге жұмсалған тыңайтқыш көлемі бойынша (жоғарыдағы кестенің 3-бағаны), содан кейін жалпы түсім бойынша (4-баған) рангтерді анықтаймыз. Қалған есептеулер рангтер корреляциясы коэффициентінің формуласына сәйкес жүргізіледі.

$$\rho = 1 - \frac{6 \cdot \sum d^2}{n \cdot (n^2 - 1)} = 1 - \frac{6 \cdot 28}{10 \cdot 99} = 0,83.$$

Бұл жалпы түсім мен қолданылған тыңайтқыш көлемі арасында тығыз байланыс бар екенін көрсетеді. Енді есептелген көрсеткіштің мәнділігін Стьюденттің t -критерийі арқылы тексереміз:

$$t_{\rho} = |\rho| \cdot \sqrt{\frac{n-2}{1-\rho^2}} = 0,83 \cdot \sqrt{\frac{8}{1-0,83^2}} = 4,22.$$

$t_{\rho} > t_{\rho}$ шарты (біздің мысалда $4,22 > 2,31$) орындалғандықтан, жоғарыда анықталған рангтер корреляциясы коэффициенті статистикалық тұрғыдан мәнді деп саналады.

Ранг бойынша орналасқан көрсеткіштер арасындағы байланысты **Кендалл коэффициенті** бойынша да анықтауға болады. Кендалл коэффициенті мына формула бойынша есептеледі:

$$\tau = \frac{2S}{n \cdot (n-1)} = \frac{2 \cdot (P-Q)}{n \cdot (n-1)},$$

мұндағы τ – Кендалл коэффициенті;

n – бақылау саны ;

P – ρ белгісінің әр рангі үшін берілген қатарда сол рангтен кейін орналасқан және мәні бойынша одан артық рангтердің қосындысы;

Q – ρ белгісінің әр рангі үшін берілген қатарда сол рангтен кейін орналасқан және мәні бойынша одан кем рангтердің қосындысы.

Бұл көрсеткішті есептеу мынадай реттілікпен жүргізіледі:

1. Факторлық белгі рангтері өсу немесе кему ретіне сәйкес анықталады.
2. Нәтижелік белгі мәндерін факторлық белгі мәндеріне сәйкес орналастырады.
3. Нәтижелік белгінің әр рангі үшін берілген қатарда сол рангтен кейін орналасқан және мәні бойынша одан артық рангтердің қосындысы анықталады.
4. Факторлық белгінің әр рангі үшін берілген қатарда сол рангтен кейін орналасқан және мәні бойынша одан кем рангтердің қосындысы анықталады.
5. τ коэффициенті формула бойынша есептеледі.

Мысал. 9.8-кесте мәліметтерін пайдаланып, Кендалл коэффициентін анықтау керек.

Шешуі: Ең алдымен шаруашылықтарды 1 гектар жерге жұмсалған тыңайтқыш көлемінің өсу реті бойынша орналастырамыз және Кендалл коэффициентін есептеуге қажетті мынадай кестені құрамыз:

9.9-кесте

Кендалл коэффициентін анықтауға қажетті есептеулер

\bar{o} бойынша ранг	\bar{o}	\bar{o}	P	Q
1	20	4,7	8	1
2	21	4,8	7	1
3	22	5,1	6	1
4	24	5,2	5	1
5	25	4,5	5	0
6	29	5,5	3	1
7	30	5,3	3	0
8	32	7,7	0	2
9	34	7,0	1	0
10	35	7,5		
Барлығы	-	-	38	7

Кестедегі есептеулерді формулаға қоямыз:

$$\tau = \frac{2 \cdot (P - Q)}{n \cdot (n - 1)} = \frac{2 \cdot (38 - 7)}{10 \cdot 9} = \frac{62}{90} = 0,69.$$

Есептеу нәтижелері жалпы түсім мен қолданылған тыңайтқыш көлемі арасында айқын байланыс бар екенін көрсетеді.

5. Ассоциация және контингенция коэффициенттері альтернативті белгі түрінде берілген екі сапалық белгі байланысының тығыздығын анықтау үшін қолданылады. Оларды есептеу үшін мынадай кесте құрылады:

9.10-кесте

Ассоциация және контингенция коэффициенттерін есептеу кестесі

Белгілер	\hat{A}	$\bar{\hat{A}}$	Барлығы
B	\hat{a}	b	$\hat{a} + b$
\bar{B}	c	d	$c + d$
Барлығы	$\hat{a} + c$	$b + d$	n

Кестеде \hat{a}, b, c, d белгілеулері арқылы альтернативті белгілердің өзара қиысуының жиіліктері, ал n арқылы жиіліктердің жалпы саны өрнектелген. Мысалы, \hat{a} – жұмыс істейтін ерлер саны; b – жұмыс істемейтін ерлер саны; c – жұмыс істейтін әйелдер саны; d – жұмыс істемейтін әйелдер саны. Осы кесте бойынша ассоциация коэффициенті былай анықталады:

$$\hat{E}_a = \frac{ad - bc}{ad + bc}.$$

Ал контингенция коэффициентін мына формуламен есептейді:

$$\hat{E}_k = \frac{ad - bc}{\sqrt{(a + b) \cdot (b + d) \cdot (a + c) \cdot (c + d)}}.$$

Контингенция коэффициенті әрқашан ассоциация коэффициентінен кіші болады. Енді осы коэффициенттерді есептеуге мысал келтірейік:

Мысал. Мынадай мәліметтер берілген:

9.11-кесте

Ассоциация және контингенция коэффициенттерін есептеу

Жұмысқа қанағаттану	Ерлер	Әйелдер	Барлығы
---------------------	-------	---------	---------

Өз жұмысына қанағаттанады	250	80	330
Өз жұмысына қанағаттанбайды	50	120	170
Барлығы	300	200	500

Осы мәліметтер бойынша ассоциация және контингенция коэффициенттерін анықтау керек.

$$\text{Шешуі: } \hat{E}_a = \frac{ad - bc}{ad + bc} = \frac{250 \cdot 120 - 80 \cdot 50}{250 \cdot 120 + 80 \cdot 50} = \frac{26000}{34000} = 0,87;$$

Ассоциация коэффициенті бұл мысалда адамның жынысы мен оның өз жұмысына қанағаттануының арасында тығыз байланыс бар екенін көрсетеді.

$$\hat{E}_k = \frac{250 \cdot 120 - 80 \cdot 50}{\sqrt{(250 + 80) \cdot (80 + 120) \cdot (250 + 50) \cdot (50 + 120)}} = \frac{26000}{58017,2} = 0,45.$$

Ал контингенция коэффициенті адамның өз жұмысына қанағаттануы мен оның жынысы арасында орташа ғана байланыс бар екенін көрсетеді. Іс жүзінде контингенция коэффициенті ассоциация коэффициентіне қарағанда құбылыстар байланысының тығыздығын дәлірек өлшейтіні дәлелденген.

9.6. Көптік корреляция

Көптік корреляция нәтижелік белгі мен екі немесе одан да көп факторлық белгілер арасындағы байланысты көрсетеді. Құбылыстар арасындағы байланысты көптік регрессия әдісімен зерттегенде нәтижелік және факторлық белгілер байланысы мынадай функция арқылы өрнектеледі:

$$\bar{y}_x = f(x_1, x_2, \dots, x_n)$$

мұндағы $\delta_1, \delta_2, \dots, \delta_n$ - факторлық белгілер;

y_x - нәтижелік белгі.

Көптік регрессия үлгісін құру мынадай кезеңдерді қамтиды:

- байланыс формасын (регрессия теңдеуін) таңдау;
- факторлық белгілерді іріктеу;
- бақылау санын анықтау.

Көптік корреляциялық байланысты сандық жағынан сипаттау регрессия теңдеуі арқылы жүзеге асырылады. Әлеуметтік-экономикалық құбылыстар арасындағы байланысты мынадай функциялар арқылы өрнектеуге болатыны практика жүзінде дәлелденген:

1) сызықтық функция: $\tilde{\delta} = a_0 + a_1 x_1 + a_2 \delta_2 + \dots + a_n \delta_n;$

2) парабола: $\tilde{\delta} = a_0 + a_1 x_1^2 + a_2 x_2^2 + \dots + a_n x_n^2;$

3) гипербола: $\tilde{\delta} = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + \dots + \frac{a_n}{x_n};$

4) дәрежелік функция: $\tilde{\delta} = a_0 \cdot x_1^{a_1} \cdot \delta_2^{a_2} \cdot \dots \cdot \delta_n^{a_n};$

5) көрсеткіштік функция: $\tilde{\delta} = a^{\delta_0 + a_1 \delta_1 + a_2 \delta_2 + \dots + a_n x_n}$

Осы функциялардың ішінде қолдануға ыңғайлы болғандықтан сызықтық функция жиі қолданылады. Ал сызықтық емес регрессия теңдеулерін математикалық түрлендірулер нәтижесінде сызықтық түрге келтіруге болады.