

# Бравэ Пирсонның сызықтық корреляция коэффициенті

Орындағандар: Сраж Айнұр және Санақ Айгерім

# Корреляциялық талдау



«Корреляция» термині өзара байланыс, өзара тәуелділік мағынасын білдіреді. Корреляция туралы айтқанда "корреляциялық байланыс" термині қатар қолданылады.

**Корреляциялық байланыс** – бұл екі белгінің немесе бірнеше белгінің (көпмүшелі корреляциялық байланыс) бір – бірімен сәйкестене өзгеруі.

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  $X$  және  $Y$  регрессия функцияларының екеуі де сызықтық болса, онда  $X$  және  $Y$  шамаларының арасындағы корреляциялық тәуелділік **сызықтық корреляция** деп аталады.

Регрессия дегеніміз – бір тәуелді айнымалы (әдетте  $Y$  арқылы белгіленеді) және басқа айнымалылар қатары (тәуелсіз айнымалылар деп аталатын) арасындағы байланыстың күші мен сипатын анықтауға тырысатын қаржы, инвестициялау және басқа пәндерде қолданылатын статистикалық әдіс.

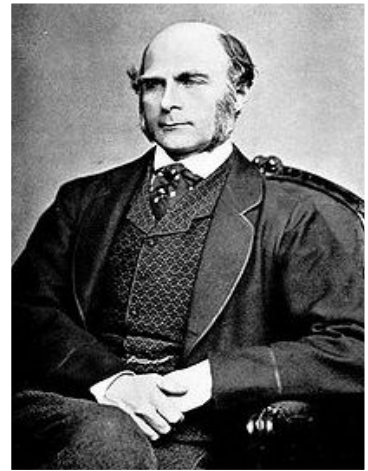
# Корреляциялық талдау

Корреляциялық талдау – бұл екі немесе одан көп кездейсоқ шамалардың арасындағы байланыстың тығыздығын және бағытын анықтайтын сандық әдіс.



Ж. Кювье  
(1769 - 1832)

«Корреляция» терминін алғаш рет француз палеонтологы Ж. Кювье енгізді, ал статистикада оны Ф. Гальтон алғаш қолданды.



Ф. Гальтон  
(1822 - 1911)

# Корреляция коэффициенті

**Корреляция коэффициенті** – байланыстың күшін және оның бағытын сипаттайтын,  $[-1, 1]$  аралығынды мәндерін қабылдайтын көрсеткіш.

Оны символмен  $r$  деп белгілейді.

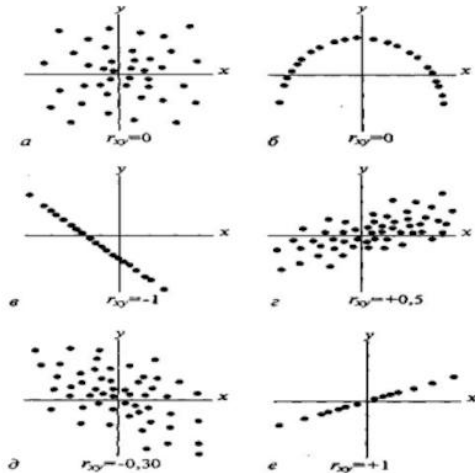
Байланыстың күшін бағалау үшін корреляция теориясында аңғылшын статистигі Чеддоқтың шкаласы қолданылады

Байланыстың сандық мөлшері	Күштің сапалық сипаттамасы
0,1 - 0,3	Әлсіз
0,3 - 0,5	Қалыпты
0,5 - 0,7	Айқын
0,7 - 0,9	Жоғары
0,9 - 1	Күшті

**Корреляция коэффициенті** - бір айнымалыны екіншісін біле отырып, қаншалықты жақсы болжауға болатынын бағалайтын статистика; Мысалы, адамдардың биіктігін білу арқылы олардың салмағын болжауға болатынын айтайық.

Корреляция оң (+) және теріс (-) болуы мүмкін. Корреляция белгісі екі айнымалының оң корреляциямен байланысты екенін көрсетеді (екі айнымалының мәні бір уақытта өседі немесе төмендейді) немесе теріс корреляция (бір айнымалының екіншісі кеміген сайын артады). Мысалы, студенттің сабаққа келмеуінің семестрдің соңындағы балмен  $-0.40$  корреляциясы бар делік (сабақтан қалу неғұрлым көп болса, балл соғұрлым төмен). Екінші жағынан, алынған балл мен қатысқан сабақтар саны арасындағы корреляция  $+0.40$  болады. Облигацияның беріктігі бірдей, бірақ оның белгісі сабақтарға қатыспағанымызды немесе сабаққа қатысқанымызға байланысты.

## ГРАФИКАЛЫҚ КОРРЕЛЯЦИЯ



# Корреляциялық байланыс



**Корреляция коэффициенті** – салыстырылатын екі статистикалық белгілердің арасындағы байланыс күшінің математикалық көрсеткіші. Корреляция қандай формуламен есептелінсе де, оның мәні ылғи да  $-1$  және  $+1$  шегінде болады. Коэффициенттің шеткі мәнінің мағынасын былай түсіну керек:

- егерде коэффициенттің мәні  $-1$ -ге тең болса, онда байланыс **кері пропорционалдық** тип бойынша функционалды деп саналады;
- егерде ол  $+1$ -ге тең болса, онда белгілер арасындағы байланыс **тура (тікелей)** пропорционалдылық тәуелділік типі бойынша біржақты (функционалды, статистикалы емес);
- егерде оның мәні нольге тең болса, онда салыстырылатын белгілердің арасындағы байланыс нолдік шамаға тең немесе **байланыс жоқ** дегенді білдіреді.

# Жұптасқан Пирсонның корреляция коэффициенті

**Сызықты (жұптасқан Пирсонның) корреляция коэффициенті** –

байланыстың күшін және бағытын сипаттайтын көрсеткіш:

$$r_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}},$$

мұндағы  $r_{xy}$  – корреляция коэффициенті,

$x, y$  – корреляцияланатын қатар,

$\bar{x}, \bar{y}$  – орташа шамалар.

Жұптасқан корреляция коэффициенті **параметрлік** коэффициент болып табылады.

# Корреляция коэффициентінің есептеуге мысал



Келесі мәндер үшін Пирсонның сызықты корреляция коэффициентін есептеу керек:

1000 тұрғынға шаққанда ауыру, $x$	ОРЗ-мен	352	228	340	300	196	258	237
1000 тұрғынға пневмониямен ауыру, $y$	шаққанда	64	60	52	48	46	41	32



## Шешуі.

1) Есептеу кестесін құру.

№	$x$	$y$	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x}) \cdot (y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
1	352	64	79	15	1185	6241	225
2	228	60	-45	11	-495	2025	121
3	340	52	67	3	201	4489	9
4	300	48	27	-1	-27	729	1
5	196	46	-77	-3	231	5929	9
6	258	41	-15	-8	120	225	64
7	237	32	-36	-17	612	1296	289
<b>Қосын- дысы</b>	<b>1911</b>	<b>343</b>	<b>0</b>	<b>0</b>	<b>1827</b>	<b>20934</b>	<b>718</b>
<b>Орташа</b>	<b>273</b>	<b>49</b>					

2) Корреляция коэффициентін есептеу керек:

$$r_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} = \frac{1827}{\sqrt{20934 \cdot 718}} = 0,47.$$

3) Алынған нәтижені талдау: қарастырылған белгілердің арасындағы байланыс **тура, қалыпты**.

4) Корреляция коэффициентінің орташа қателігін есептеу:

$$m_r = \pm \frac{1 - r_{xy}^2}{\sqrt{n}} = \pm \frac{1 - 0,47^2}{\sqrt{7}} = 0,3.$$

Корреляция коэффициенті өзінің орташа қателігінен 3 есе артық **болмағандықтан сенімді емес**.

# КОРРЕЛЯЦИЯЛЫҚ Қ ТАЛДАУ

Орындаған: Санақ Айгерім





1. X және Y мәндерінің қосындысын есептеңіз

$$X_{agr} = 344$$

$$Y_{IQ} = 1549$$

2. X} және Y үшін орташа арифметикалық мәнді есепте

$$\bar{X}_{agr} = 24,6$$

$$\bar{Y}_{IQ} = 110,5$$

3. X және Y үшін орташа арифметикалық мәннен тексерілген әрбір ауытқуды есептеңіз.

№	$\bar{X}_{agr} - X_{agr}$	$\bar{Y}_{IQ} - Y_{IQ}$
1	0,6	10,6
2	-2,4	-4,4
3	-1,4	-6,4
4	3,6	-8,4
5	4,6	-23,4
6	-6,4	16,6
7	-1,4	5,6
8	2,6	7,6
9	4,6	-0,4
10	6,6	-13,4
11	-5,4	-11,4
12	-4,4	1,6
13	0,6	0,6
14	-1,4	24,6

4. Содан кейін әрбір ауытқуды квадраттаймыз: Содан кейін ауытқулардың квадраттарының қосындысын есептейміз:

№	$(\bar{X}_{agr} - X_{agr})^2$	$(\bar{Y}_{IQ} - Y_{IQ})^2$
1	0,36	112,36
2	5,76	19,36
3	1,96	40,96
4	12,96	70,56
5	21,16	547,56
6	40,96	275,56
7	1,96	31,36
8	6,76	57,79
9	21,16	0,16
10	43,56	179,56
11	29,16	129,96
12	19,36	2,56
13	0,36	0,36
14	1,96	605,16

$$\Sigma(\bar{X}_{agr} - X_{agr})^2 \text{ и } \Sigma(\bar{Y}_{IQ} - Y_{IQ})^2$$

$$\Sigma(\bar{X}_{agr} - X_{agr})^2 = 207,44$$

$$\Sigma(\bar{Y}_{IQ} - Y_{IQ})^2 = 2073,24$$

6. Әрбір бақылау үшін орташа арифметикалық шама мен шаманың айырмасының көбейтіндісі есе

$N_i$	$(\bar{X}_{agr} - X_{agr}) * (\bar{Y}_{IQ} - Y_{IQ})$
1	6,36
2	10,56
3	8,96
4	-30,24
5	-107,64
6	-106,24
7	-7,84
8	19,76
9	-1,84
10	-88,44
11	61,56
12	-7,04
13	0,36
14	-34,44

7. Сумманы есепте  $(\bar{X}_{agr} - X_{agr}) * (\bar{Y}_{IQ} - Y_{IQ})$

$$\Sigma(\bar{X}_{agr} - X_{agr}) * (\bar{Y}_{IQ} - Y_{IQ}) = -276,16$$

8. Алынған мәндерд  $\sigma X_{agr}, \sigma Y_{IQ},$

$$\Sigma(\bar{X}_{agr} - X_{agr}) * (\bar{Y}_{IQ} - Y_{IQ})$$

Пирсон корреляция коэффициентінің формуласына қоямыз:

$$r_{X_{agr}Y_{IQ}} = \frac{-276,16}{\sqrt{207,44 * 2073,24}} = \frac{-276,16}{\sqrt{430072,9056}} = \frac{-276,16}{655,799} = -0,4211$$

9. Қорытынды: Корреляция коэффициентінің мәндер кестесіне сәйк  $r_{X_{agr}Y_{IQ}} = -0,4211$

**әлсіз теріс корреляция** деген қорытындыға келеміз.

# Талқылауға арналған сұрақтар:

1. Функционалдық және корреляциялық тәуелділіктер.
2. Параметрлік (Браве-Пирсон корреляция коэффициенттері және корреляциялық қатынас) және параметрлік емес (Спирман, Кендалл) байланысты бағалау әдістері.
3. Корреляциялық қатынастар. Жалған корреляция мәселесі. Корреляциялық плейад әдісі.
4. Сапалық белгілерді бағалау кезінде корреляция коэффициенттерін анықтау. Атау шкаласы. Ассоциация коэффициентін есептеу. Дұрыстығын тексеру.
5. Корреляциялық қатынастар. Фишер критерийі бойынша корреляция формасын анықтау. Үш немесе одан да көп белгілердің өзара байланысы.
6. Бравэ Пирсонның сызықтық корреляция коэффициенті.
7. Корреляцияның дәрежелік коэффициенті. Корреляция коэффициентінің сенімді шекаралары, корреляция коэффициентінің статистикалық маңыздылығын бағалау, екі корреляция коэффициентін салыстыру.



Корреляциялық талдау 2 негізгі міндетті шешуден тұрады:

1. Байланыс формасын анықтау, яғни функция түрін табу;
2. Байланыс күшін(тығыздығын) анықтау, яғни  $x$  әртүрлі мәндер үшін  $y$  дәрежесін бағалау.

Белгілер арасындасындағы статистикалық байланысты белгілердің тәжірбиелік мәндерінен ең төмен ауытқып, эксперименттік материалда байқалатын негізгі заңдылықты білдіретін математикалық функцияның көмегімен беруге тырысады.

Байланыс теңдеулері(немесе регрессия теңдеулері ) болатын функциялар байқалу формасы бойынша мынадай болады:

1. түзу сызықты;
2. қисық сызықты(параболалық, гиперболалық, дәрежелік және т.б.).

Байланыс формасын таңдауда, бірінші кезекте, қисықтың сол немесе басқа типі құбылыстың немесе процестің шынайы табиғатын, физикалық мәнін белгілейтіндігін ескеру қажет. Байланыс формасын графикалық анықтау үшін тәжірбиелік деректерді арнаулы корреляциялық кестеге немесе корреляциялық торға енгізеді.

## Функционалдык және корреляциялык тәуелділіктер

Статистикада қоғамдық құбылыстар пен процестердің арасындағы өзара байланысты анықтамай тұрып алдымен сол өзгеріске әсерін тигізетін факторлары мен нәтижелі белгілері арасындағы тәуелділікті анықтайды. Оның өзі құбылыстың ерекшелігіне қарай *функционалдык және корреляциялык байланыс* болып екі түрге бөлінеді.

**Функционалдык байланыс** деп бір белгі мәнінің өзгеруіне әсерін тигізетін екінші бір белгінің сәйкес келуін, яғни бір факторлы белгінің өзгерісі салдарынан нәтижелі белгі мәнінің өзгеруін айтады.

$y=f(x_i)$  мұнда **y-нәтижелі белгі**,  **$x_i$ -факторлық белгі**,  $f(x_i)$  осы екі белгі арасындағы *функционалдык байланыс*.

Қоғамдық құбылыстардың өзгеруіне әртүрлі кездейсоқ факторлардың әсер етуін **корреляциялык байланыс** деп айтады.

Статистикада корреляциялык байланыс деп нәтижелі және факторлы белгілер арасындағы сәйкестіктің болмауын айтады.

$$y_i = f(x_i) + E_i$$

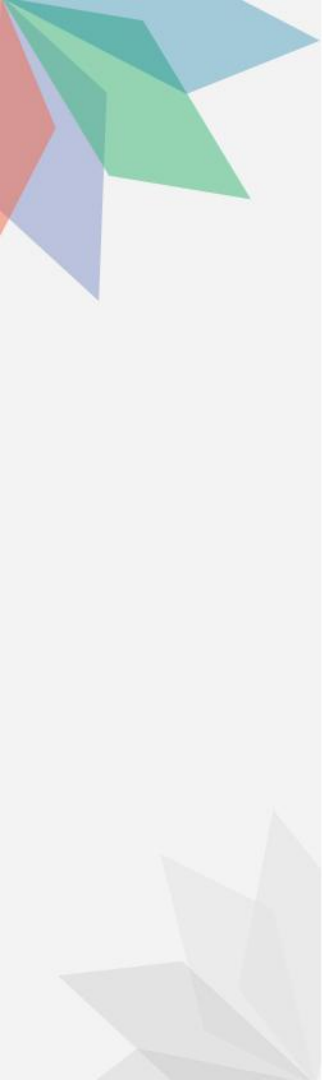
мұнда  $f(x_i)$  корреляциялык байланыста белгілі болған  $f$  пен  $y_i$  арасында және ерекше алынған бір немесе бірнеше факторлық белгінің әсерінен қалыптасқан нәтижелі белгі бөлігі.  $E_i$  - жанама және кездейсоқ факторлардың әсерінен пайда болған нәтижелі белгінің бөлігі.

Түзу сызықты байланыста себепті ықпалдар белгі мәндерінің өсуіне немесе кемуіне қарай нәтижелі белгі мәндері бірқалыпты әрі үздіксіз өседі немесе кемиді.

## **Параметрлік (Браве-Пирсон корреляция коэффициенттері және корреляциялық қатынас) және параметрлік емес (Спирман, Кендалл) байланысты бағалау әдістері.**

Статистикалық әдістердің немесе тестердің корреляция дәрежесін есептейтіні немесе жалпылауды мүмкін ететін 2 түрі бар.

- q Бірінші түрі ол өте кең қолданылатын параметрлік әдістер, оларға орташша мән, дисперсия сияқты параметрлер қолданылады.
  
  - q Екінші түрі ол параметрлік емес әдістер, олар мына жағдайда, зерттеуші өте аз таңдау көлемімен немесе сапалы мәліметтермен жұмыс жасағанда өте құнды жәрдемін береді, бұл әдістер өте қарапайым болып келеді себебі қолдануда, есептеуде өте ыңғайлы болады.
- Параметрлік емес әдістердің бір маңызды міндеттерінің бірі ол популяцияның бір бөлігінде алынған мәліметтерді анализдеу, соның арқасында жалпы барлық популяцияға қорындынды жасау мақсатын орындайды.

The image features a white background with decorative geometric shapes in the corners. The top-left corner has several overlapping triangles in shades of blue, green, and red. The bottom-left corner has a cluster of overlapping triangles in shades of grey and light blue.

**Назарларыңызға  
рахмет!!!**