ТЕМА 9. СТАТИСТИЧНІ МЕТОДИ ВИВЧЕННЯ ВЗАЄМОЗВ'ЯЗКІВ

9.1. Типи зв'язків між явищами та їх характеристика.
9.2. Метод порівняння паралельних рядів.
9.3. Метод аналітичного групування.
9.4. Парний кореляційно-регресійний аналіз.

9.1. Типи зв'язків між явищами та їх характеристика

Дослідження об'єктивно існуючих зв'язків між явищами – найважливіше завдання статистики. У процесі статистичного дослідження залежностей виявляються причинно-наслідкові зв'язки між явищами. Причинно-наслідкові зв'язки – це такий зв'язок явищ і процесів, коли зміна одного з них (*причини*) веде до зміни іншого (*наслідку*).

Ознаки явищ і процесів за їх значенням для вивчення взаємозв'язку поділяються на два класи. Ознаки, що зумовлюють зміни інших, пов'язаних з ними ознак, називають ***факторними***, або просто факторами. Ознаки, що змінюються під дією факторних ознак, називають ***результативними***. *Наприклад*, кіль кіс ть працівників – це факторна ознака, а обсяг виробленої продукції – результативна ознака.

У статистиці розрізняють ***функціональні (детерміновані)*** та ***стохастичні (ймовірнісні)*** зв'язки між явищами і процесами. ***Функціональним*** називають такий зв'язок, при якому певному значенню факторної ознаки відповідає одне строго визначене значення результативної ознаки. Такий зв'язок можливий за умови, що на поведінку однієї ознаки (результативної) впливає тільки друга ознака (факторна) і ніякі інші, тобто значення результативної ознаки на 100% залежить від факторної. Такі зв'язки є абстракціями, в реальному житті вони зустрічаються рідко, але знаходять широке застосування в точних науках і, в першу чергу, в математиці. Наприклад, залежність площі ~~кола~~ круга від радіуса: $S = \pi \cdot r^2$.

Функціональний зв'язок проявляється у всіх випадках спостереження і для кожної конкретної одиниці досліджуваної сукупності. У масових явищах проявляються ***статистичні зв'язки***, при яких строго певному значенню факторної ознаки ставиться у відповідність ~~безліч~~ множину значень результативної. Такі зв'язки мають місце, якщо на результативну ознаку діють декілька факторних, а для опису зв'язку використовується один або декілька визначальних (врахованих) факторів. Названі зв'язки є неповними, тому що завжди існують невраховані фактори, отже, значення результативної ознаки залежить від значень факторної менше, ніж на 100%. *Прикладом* статистичного зв'язку може служити залежність собівартості одиниці продукції від рівня продуктивності праці: чим вища продуктивність праці, тим нижча собівартість. Але на собівартість одиниці продукції, крім продуктивності праці, впливають й інші фактори: вартість сировини, матеріалів, палива, загальновиробничі і загальногосподарські витрати тощо. Тому не можна стверджувати, що підвищення продуктивності праці на 5% призведе до аналогічного зниження собівартості. Може спостерігатися і зворотна картина, якщо на собівартість впливатимуть більшою мірою інші фактори, – наприклад, різко зростуть ціни на сировину і матеріали.

Окремим випадком статистичного (стохастичного) зв'язку є ***кореляційний зв'язок***, при якому зміна середнього значення результативної ознаки обумовлена зміною факторних ознак.

Крім того, зв'язки між явищами та його ознаками класифікуються за ступенем тісноти, напрямком і аналітичним вираженням.

За напрямом виділяють зв'язок ***прямий*** і ***зворотній***. При прямому зв'язку результативна ознака зростає зі збільшенням факторної, при зворотному – зростання факторної ознаки призводить до зниження значень результативної ознаки. Наприклад, чим більше стаж роботи, тим вища продуктивність праці (*прямий зв'язок*), а чим вище продуктивність праці, тим нижча собівартість одиниці продукції (*зворотний зв'язок*).

За формою (аналітичним вираженням) зв'язки поділяються на ***лінійні (прямолінійні)*** і ***нелінійні (криволінійні)***. Лінійні зв'язки виражаються рівнянням прямої, а нелінійні – рівнянням параболи, гіперболи, степеневої тощо.

За кількістю взаємодіючих факторів зв'язки поділяються на ***парний (однофакторний)*** і ***множинний (багатофакторний)*** зв'язки. При парному зв'язку на результативну ознаку діє одна факторна ознака, а при множинному – кілька факторних ознак.

9.2. Метод порівняння паралельних рядів

Після того, як на підставі теоретичного аналізу буде виявлено, що між досліджуваними явищами існує взаємозв’язок, необхідно виявити *тісноту цього зв’язку* та визначення його напрямку.

Якщо досліджувана статистична сукупність представлена невеликою кількістю вихідних даних, то наявність або відсутність кореляції між двома ознаками можна визначити ***методом паралельних рядів***. З цією метою значення факторної ознаки розташовують по мірі зростання або зменшення, і потім ранжовані значення зіставляють з результативною ознакою. Шляхом співставлення розташованих таким чином рядів визначаються істотні зв’язки та їх напрями на основі розрахунку спеціальних коефіцієнтів.

Найпростішим показником є коефіцієнт Фехнера ($К_ф$), який розраховується за формулою:

$$К_ф = \frac{С - Н}{С + Н}, \tag{9.1}$$

де С – кількість співпадінь знаків відхилень від середньої;
Н – кількість неспівпадінь знаків відхилень від середньої.

Якщо відхилення індивідуальних значень кожної ознаки від їх середніх величин більше нуля, то значенню присвоюється знак «+», в протилежному випадку – знак «-». В тому випадку, коли за обома показниками знаки однакові, має місце їх співпадіння, а коли вони різні – неспівпадіння. Коефіцієнт Фехнера знаходиться в межах від -1 до +1. Якщо коефіцієнт Фехнера прагне до 0, то зв’язок між показниками слабкий, а якщо коефіцієнт прагне до 1, то зв’язок тісний. Цей коефіцієнт має додатне значення при наявності прямого зв’язку, а від’ємне – зворотного.

Більш досконалим показником вважається коефіцієнт кореляції рангів Спірмена:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)}, \tag{9.2}$$

де d – різниця рангів факторного та результативного показників;
n – кількість одиниць в сукупності.

При цьому під рангом розуміють порядковий номер значення показника у порядку зростання або зменшення. Коефіцієнт кореляції рангів також змінюється від -1 до +1. При позитивному значенні коефіцієнта зв'язок між показниками прямий, а при негативному – зворотній. Якщо коефіцієнт наближається до 1, між показниками існує тісний (сильний) зв'язок, якщо його значення менше 0,3, вважається, що взаємозв'язок практично відсутній.

Таким чином, наведені коефіцієнти дають можливість не тільки оцінити тісноту взаємозв'язку між факторною та результативною ознаками, але й визначити його напрямок (прямий чи зворотній).

Розглянемо приклад розрахунку коефіцієнта Фехнера та коефіцієнта кореляції рангів Спірмена за даними про ціну та обсяг продажу товару (табл. 9.1).

Таблиця 9.1 – **Вихідні дані для розрахунку коефіцієнта Фехнера та коефіцієнта кореляції рангів Спірмена**

Ціна, грн. (*X*)	Обсяг продажу, шт. (*Y*)	Знаки відхилень		Ранги		*d*	d^2
		за *X*	за *Y*	за *X*	за *Y*		
1	2	3	4	5	6	7	8
350	120	-	+	2	6	-4	16
460	104	+	-	5	2	3	9
630	76	+	-	8	1	7	49
380	111	-	-	3	4	-1	1
490	123	+	+	6	7	-1	1
520	105	+	-	7	3	4	16
260	140	-	+	1	8	-7	49
430	116	-	+	4	5	-1	1
3520	**895**	–	–	–	–	–	**142**

Розраховуємо середні значення показників:

$$\overline{X} = \frac{3520}{8} = 440 \text{ грн.}$$

$$\overline{Y} = \frac{895}{8} = 112 \text{ шт.}$$

З граф 3 і 4 визначаємо, що знаки співпали 2 рази (С=2), а не співпали 6 разів (Н=6). Отже, коефіцієнт Фехнера становить:

$$К_ф = \frac{2-6}{2+6} = -0,5$$

Таким чином, можна зробити висновок, що між ціною та обсягом продажу існує зворотній середній зв’язок.
Розрахуємо коефіцієнт кореляції рангів Спірмена:

$$\rho = 1 - \frac{6 \cdot 142}{8(64-1)} = -0,69$$

Одержане значення коефіцієнта також підтверджує наявність зворотного середнього зв’язку між досліджуваними показниками.

9.3. Метод аналітичного групування

Метод аналітичного групування полягає в тому, що сукупність розбивається на групи за факторною ознакою, далі за кожною групою та за сукупністю визначаються середні значення факторної та результативної ознаки. Порівняння середніх значень факторної та результативної ознак дозволяє зробити певні висновки про наявність та напрямок взаємозв’язку між ними.

Порядок проведення аналітичного групування проілюструємо на конкретному прикладі визначення залежності між чистим прибутком та фондовіддачею.

Базові вихідні дані представлено в табл. 9.2.

Таблиця 9.2 – **Базові вихідні дані для здійснення аналітичного групування підприємств за чистим прибутком та фондовіддачею**

№ з.п.	Фондовіддача, грн./грн.	Чистий прибуток, млн. грн.
1	2	3
1	6,9	5,1
2	12,9	9,4
3	8,3	5,8
4	11,4	9,6
5	8,7	6,8
6	9,0	4,5
7	13,1	8,9
8	5,6	3,9

Продовження табл. 9.2

1	2	3
9	8,5	4,5
10	5,9	3,6
11	6,9	5,6
12	9,0	6,3
13	5,9	3,8
14	6,9	5,2
15	9,2	7,7
16	5,9	4,4
17	7,6	6,3
18	10,5	8,5
19	8,8	7,5
20	5,7	4,7
21	8,0	6,5
22	8,9	7,0
23	5,9	4,5
24	3,5	3,1
25	4,8	3,3
26	3,3	2,4

Групування розпочинається з того, що за факторною ознакою (фондовіддача) окремі одиниці сукупності поєднуються в однорідні групи. Для цього визначається кількість груп факторної ознаки:

$$K = 1 + 3{,}322 \cdot \lg 26 = 5{,}7 \approx 6$$

Розраховуємо довжину інтервалу факторної ознаки:

$$\Delta_x = \frac{13{,}1 - 3{,}3}{6} = 1{,}6 \approx 2$$

Довжина інтервалу округлюється у бік збільшення на один знак більше, ніж число знаків у значенні ознаки.

Визначаються значення результативної ознаки в кожному інтервалі зміни факторної ознаки (табл. 9.3).

Розраховується маса, кількість значень результативної ознаки та її середнє значення в кожному інтервалі зміни факторної ознаки.

Таблиця 9.3 – **Аналітичне групування**

Групи підприємств за рівнем фондовіддачі, грн.	Чистий прибуток, млн. грн.			
	окремі значення	маса	кількість	середнє значення
3,3-5,3	2,4; 3,1; 3,3	8,8	3	2,9
5,3-7,3	3,6; 3,8; 3,9; 4,4; 4,5; 4,7; 5,1; 5,2; 5,6	40,8	9	4,5
7,3-9,3	6,3; 6,5; 5,8; 4,5; 4,5; 6,8; 7,5; 7,0; 6,3; 7,7	62,9	10	6,3
9,3-11,3	8,5	8,5	1	8,5
11,3-13,3	9,6; 9,4; 8,9	27,9	3	9,3
13,3-15,3	–	–	–	–

За даними табл. 9.3 будується графік залежності чистого прибутку від фондовіддачі підприємства (рис. 9.1).

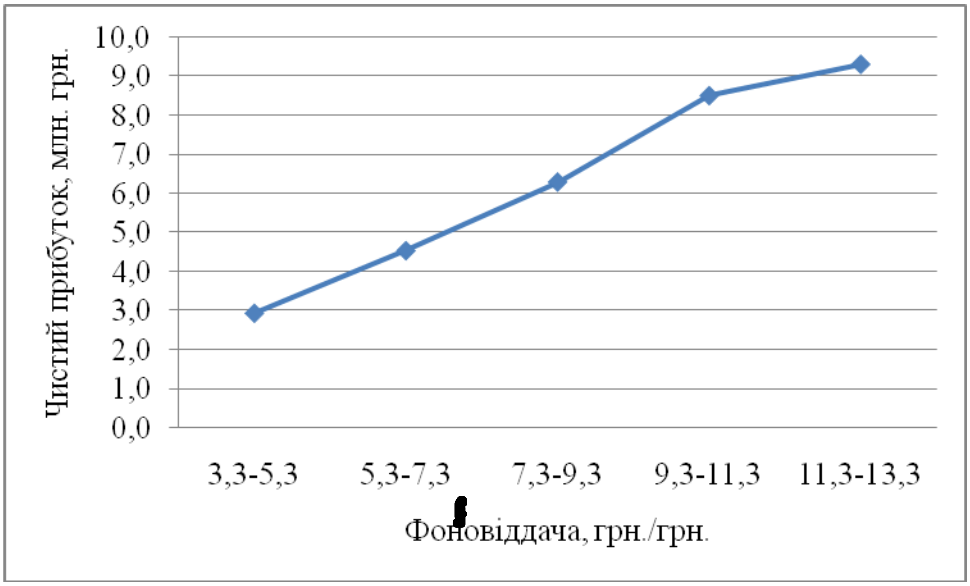


Рис. 9.1. Залежність чистого прибутку від фондовіддачі підприємства

Таким чином, зі зростанням фондовіддачі підприємства зростає середній чистий прибуток.

9.4. Парний кореляційно-регресійний аналіз

Кореляція і регресія відносяться до методів виявлення статистичної залежності між досліджуваними змінними.

Кореляція (від лат. **correlatio**), **кореляційна залежність** – взаємозалежність двох або кількох випадкових величин. Суть її полягає в тому, що при зміні значення однієї змінної відбувається закономірна зміна (зменшення або збільшення) іншої (-ших) змінної (-них).

При розрахунку кореляцій намагаються визначити, чи існує статистично достовірний зв'язок між двома або кількома змінними в одній або декількох вибірках. ***Наприклад***, взаємозв'язок між зростом і вагою дітей, взаємозв'язок між успішністю і результатами виконання тесту IQ, між стажем роботи і продуктивністю праці.

Важливо розуміти, що кореляційна залежність відображає **тільки взаємозв'язок між змінними** і не говорить про причинно-наслідкові зв'язки. Наприклад, якби в досліджуваній вибірці між зростом і вагою людини існувала кореляційна залежність то, це не означало б, що вага є причиною зросту людини, інакше, скидаючи зайві кілограми, зріст людини також би зменшувався. Кореляційний зв'язок лише говорить про взаємозв'язок даних параметрів, причому в даній конкретній вибірці, в іншій вибірці ми можемо не спостерігати отримані кореляції. ***Приклад 2:*** є позитивна кореляція між збільшенням зарплати менеджерів з продажу та якістю роботи з клієнтами (підвищення якості обслуговування, робота з запереченнями, знання позитивних якостей продукту в порівнянні з конкурентами) при відповідній мотивації персоналу. Збільшення обсягів продажу, а, отже, і зарплата менеджерів, зовсім не означають, що менеджери поліпшили якість роботи з клієнтами. Цілком ймовірно, що випадково надійшли великі замовлення або відділ маркетингу збільшив рекламний бюджет або сталося ще щось.

За формою кореляційний зв'язок може бути прямолінійним або криволінійним. *Прямолінійним* може бути, наприклад, зв'язок між кількістю відвіданих занять з курсу протягом семестру і кількістю правильно вирішених завдань на іспиті з цього курсу. *Криволінійним* може бути, наприклад, зв'язок між рівнем мотивації і ефективністю виконання завдання. При підвищенні мотивації ефективність виконання завдання спочатку зростає, потім досягається оптимальний рівень мотивації, якому відповідає максимальна ефективність виконання завдання. Подальше підвищення мотивації супроводжується зниженням ефективності.

За напрямом кореляційний зв'язок може бути *позитивним («прямим»)* і *негативним («зворотнім»)*. Позитивна кореляція (пряма) виникає при одночасній зміні двох змінних величин в однакових напрямах (в позитивному або негативному). Прикладами прямого зв’язку є зв’язок між внесенням добрив та урожайністю сільськогосподарських культур, рівнем годівлі та продуктивністю тварин, рівнем механізації виробничих процесів та продуктивністю праці. Прикладами зворотного зв’язку є зв’язок між урожайністю та собівартістю продукції, собівартістю продукції та рентабельністю виробництва, продуктивністю праці та собівартістю продукції.

Залежно від ***кількості досліджуваних ознак*** розрізняють *парну (просту)* та *множинну* кореляцію. При парній кореляції аналізують зв’язок між факторною та результативною ознаками; при множинній кореляції – залежність результативної ознаки від двох та більше факторних ознак.

Таким чином, кореляційний аналіз застосовується для знаходження *характеру і тісноти зв’язку* між випадковими величинами.

Регресійний аналіз має на меті *визначення (ідентифікацію) рівняння регресії*, включаючи статистичну оцінку його параметрів. Рівняння регресії дозволяє знайти значення залежної змінної, якщо величина незалежної або незалежних змінних відома.

Практично, мова йде про те, щоб, аналізуючи безліч точок на графіку (тобто безліч статистичних даних), знайти лінію, яка за можливістю, точно відображає укладену в цій множині закономірність (тренд, тенденцію) – лінію регресії.

За кількістю чинників розрізняють одно-, дво- та багатофакторні рівняння регресії.

Етапи кореляційно-регресійного аналізу:

– виявлення кореляційної залежності;
– підбір виду залежності;
– розрахунок параметрів рівняння регресії;
– визначення тісноти зв’язку.

9.4.1. Виявлення кореляційної залежності

Для виявлення кореляційної залежності здійснюється збір даних методом випадкової вибірки деякої кількості спостережуваних об’єктів з деякої однорідної сукупності, фіксації для кожного обраного об’єкта пари ознак (властивостей), взаємозв’язок яких буде предметом дослідження. Наявність зв’язку між ознаками можна виявити за допомогою графічного або аналітичного методів.

Сутність графічного методу полягає в побудові ***поля кореляції***, або сукупності точок, що відображають залежність між ознаками в певній системі координат.

Розглянемо конкретний приклад побудови кореляційного поля на основі даних табл. 9.3.

Розраховуємо довжину інтервалів зміни результативної та факторної ознак за формулою:

$$\Delta = \frac{X_{\max} - X_{\min}}{1 + 3{,}322 \cdot \lg n} \tag{9.3}$$

отримуємо:

– для результативної ознаки: $\Delta Y = \dfrac{9{,}6 - 2{,}4}{1 + 3{,}322 \cdot \lg 26} = 1{,}26 \approx 1{,}5$;

– для факторної ознаки: $\Delta X = \dfrac{13{,}1 - 3{,}3}{1 + 3{,}322 \cdot \lg 26} = 1{,}72 \approx 2$.

Після цього будуємо кореляційне поле (рис. 9.2).

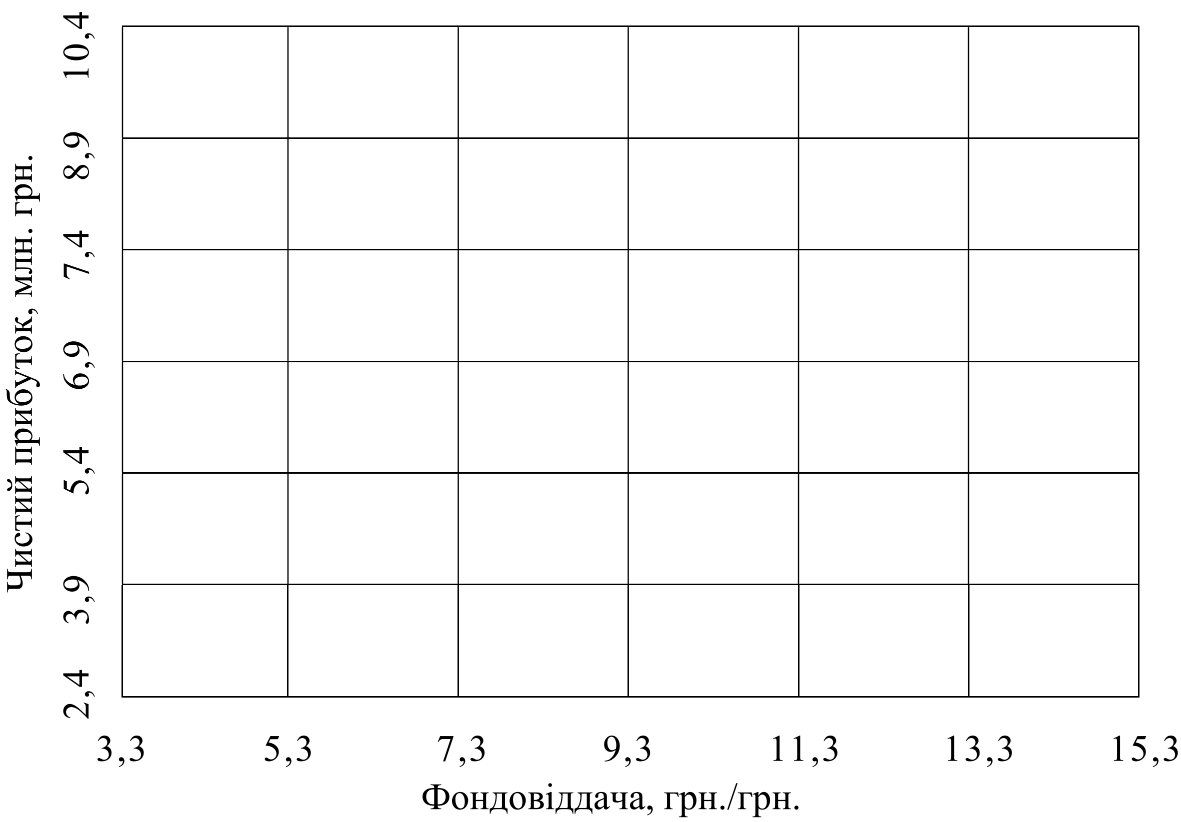
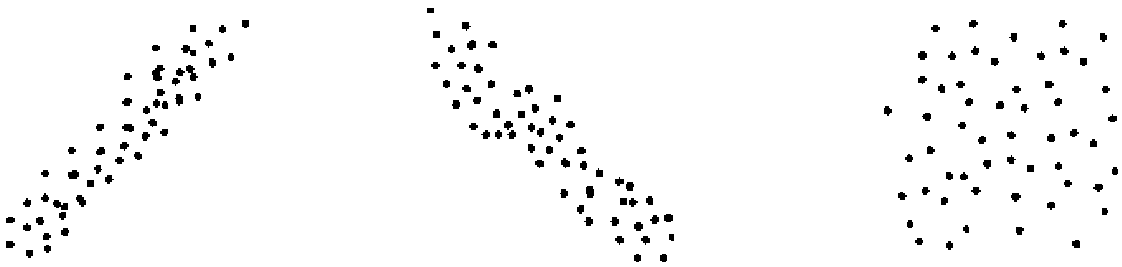


Рис. 9.2. Кореляційне поле

При побудові графіка на горизонтальній осі відкладаємо значення факторної ознаки (*x*), а на вертикальній – значення результативної ознаки (*y*). Відклавши на перетині відповідних значень *x* та *y* точки, отримуємо кореляційне поле.

Після відображення даних на полі можливі наступні три варіанти (рис. 9.3).



а) б) в)

Рис. 9.3. Схема прямолінійних кореляційних зв’язків: а) позитивний (прямий) зв’язок; б) негативний (зворотний) зв’язок; в) відсутній зв’язок

На основі графіку може бути зроблена гіпотеза про наявність лінійного кореляційного зв’язку, про нелінійний кореляційний зв’язок або про відсутність кореляційного зв’язку.

Відправним пунктом виявлення кореляційної залежності за допомогою аналітичного методу є кореляційна таблиця, яка систематизує результати спостереження над двома досліджуваними явищами або їх ознаками. В нашому випадку вона має такий вигляд (табл. 9.4).

Таблиця 9.4 – **Кореляційна таблиця залежності чистого прибутку від фондовіддачі**

Інтервали зміни чистого прибутку, млн. грн.	Середина інтервалу	Інтервали зміни фондовіддачі, грн./грн.						Всього
		3,3-5,3	5,3-7,3	7,3-9,3	9,3-11,3	11,3-13,3	13,3-15,3	
2,4-3,9	3,2	3	3					6
3,9-5,4	4,7		5	2				7
5,4-6,9	6,2		1	5				6
6,9-7,4	7,2			1				1
7,4-8,9	8,0			2	1			3
8,9-10,4	9,7					3		3
Всього		3	9	10	1	3		26

Розраховуємо середні значення результативної ознаки в кожному інтервалі зміни факторної ознаки:

$$\overline{Y}_{\bar{x}1} = \frac{3 \cdot 3{,}2}{3} = 3{,}2$$

$$\overline{Y}_{\bar{x}2} = \frac{3 \cdot 3{,}2 + 5 \cdot 4{,}7 + 1 \cdot 6{,}2}{9} = 4{,}4$$

$$\overline{Y}_{\bar{x}3} = \frac{4{,}7 \cdot 2 + 6{,}2 \cdot 5 + 7{,}2 \cdot 1 + 8{,}0 \cdot 2}{10} = 6{,}4$$

$$\overline{Y}_{\bar{x}4} = \frac{8{,}0 \cdot 1}{1} = 8{,}0$$

$$\overline{Y}_{\bar{x}5} = \frac{9{,}7 \cdot 3}{3} = 9{,}7$$

Після цього зіставляємо середні значення факторної та результативної ознак (табл. 9.5).

Таблиця 9.5 – **Зіставлення середніх значень факторної та результативної ознак**

Ознаки	Інтервали зміни ознак					
	1	2	3	4	5	6
Факторна	4,3	6,3	8,3	10,3	12,3	14,3
Результативна	3,2	4,4	6,4	8,0	9,7	–

Як бачимо, зі зміною факторної ознаки (фондовіддачі) змінюється середнє значення результативної ознаки (чистого прибутку). Отже, між одержуваними ознаками є взаємозв'язок.

9.4.2. Підбір виду залежності

Підбір виду залежності починається з аналізу теоретичного напрямку досліджуваної залежності, тобто, виходячи з економічного змісту, визначається передбачуваний напрямок дії факторної ознаки.

Потім будується емпірична лінія регресії. ***Емпірична лінія регресії*** характеризує зміну середнього значення функції під впливом чинника аргументу. Середні інтервальні значення результативної ознаки $\overline{Y}_{\text{сер}}$ відображаються на кореляційному полі у вигляді точок із середини інтервалів зміни факторної ознаки. Потім ці точки з’єднуються, і отримана ламана лінія має назву *емпіричної*. Її зиґзаґи сильніше виявляються в інтервалах з малою кількістю спостережень. За законом великих чисел можна стверджувати, що емпірична лінія регресії все більше згладжуватиметься при зростанні числа спостережень.

Граничне положення емпіричної лінії регресії, до якого вона прагне при необмеженому збільшенні числа спостережень, має назву ***теоретичної лінії регресії***, а процес її знаходження – ***вирівнюванням емпіричної лінії регресії***.

Суттєву роль в кореляційному аналізі відіграє підбір форми математичного рівняння, що найкращим чином описує досліджуваний процес.

9.4.3. Розрахунок параметрів рівняння регресії

Рівняння, що відображає зміну середньої величини однієї ознаки (*Y*) в залежності від другої (*X*), називається ***рівнянням регресії*** або ***рівнянням кореляційного зв’язку***.

Прямолінійну форму зв’язку визначають за рівнянням прямої лінії:

$$y_x = a_o + a_1 \cdot x, \tag{9.4}$$

де y_x – теоретичні (обчислені за рівнянням регресії) значення результативної ознаки;

a_0 – початок відліку, або значення y_x при умові, що x=0;

a_1 – коефіцієнт регресії (коефіцієнт пропорційності), який показує, як змінюється y_x при кожній зміні x на одиницю;

x – значення факторної ознаки.

При прямому зв'язку між корелюючими ознаками коефіцієнт регресії a_1 матиме додатне значення, при зворотному – від'ємне.

Невідомі параметри a_0 та a_1 знаходять ***способом найменших квадратів***. Сутність цього способу полягає в знаходженні таких параметрів рівняння зв'язку, при яких залишкова сума квадратів відхилень фактичних значень результативної ознаки (y) від її теоретичних (обчислених за рівнянням зв'язку) значень (y_x) буде мінімальною:

$$\sum\left(y-y_x\right)^2=\min. \tag{9.5}$$

Спосіб найменших квадратів зводиться до складання та розв'язання системи двох рівнянь з двома невідомими:

$$\begin{cases} a_o\cdot n+a_1\cdot\sum x_i=\sum y_i, \\ a_o\cdot\sum x_i+a_1\cdot\sum x_i^2=\sum y_i\cdot x_i. \end{cases} \tag{9.6}$$

Ця система рівнянь називається системою нормальних рівнянь. Вирішуючи її, отримуємо величини коефіцієнтів a_0 та a_1, а, отже, і аналітичний вираз залежності $y=a_o+a_1\cdot x$.

Для вихідних даних, представлених в табл. 9.2, ця система рівнянь має вигляд:

$$\begin{cases} 26a_0+\ \ 201{,}1a_1=\ \ 148{,}9, \\ 201{,}1a_0+1712{,}21a_1=1267{,}4 \end{cases}$$

$$\begin{cases} a_0+\ \ 7{,}7a_1=\ \ 5{,}7, \\ a_0+8{,}5a_1=6{,}3 \end{cases}$$

$$a_0=5{,}7-7{,}7a_1$$

$$5{,}7-7{,}7a_1+8{,}5a_1=6{,}3$$

$$5{,}7-7{,}7a_1+8{,}5a_1=6{,}3$$

$$0{,}8a_1=0{,}6$$

$$a_1=0{,}75$$

Тоді $a_0 = 5{,}7 - 7{,}7 \cdot 0{,}75 = -0{,}075$.

Отже, шукане рівняння має вигляд:

$$y_x = -0{,}075 + 0{,}75 \cdot x$$

Економічний зміст цього рівняння такий: коефіцієнт регресії показує, що в досліджуваній сукупності підприємств зі збільшенням фондовіддачі на 1 грн./грн. чистий прибуток зростає на 0,75 млн. грн. Параметр a_0 (у нашому прикладі -0,075) як вільний член рівняння має тільки розрахункове значення і не інтерпретується.

Підставивши у рівняння регресії значення *x*, дістанемо теоретичні рівні чистого прибутку в кожному підприємстві. Якщо сума теоретичних значень дорівнює сумі емпіричних значень $\left(\sum y_x = \sum y\right)$, то параметри рівняння визначені правильно.

Як зазначалося, при прямолінійній залежності спостерігається рівномірне збільшення (зменшення) результативної ознаки під впливом відповідної зміни факторної ознаки. У статистичній практиці трапляються і більш складні зв'язки, коли зі зміною аргументу змінюється не тільки функція, а й її приріст. ***Нелінійні (криволінійні) форми зв'язку*** різні. У статистичному аналізі найчастіше використовують *параболічну* та *гіперболічну* форми зв'язку.

Якщо криволінійна залежність має форму параболи другого порядку, зв'язок виражають таким рівнянням:

$$y = a_o + a_1 \cdot x + a_2 \cdot x^2. \tag{9.7}$$

Однією з особливостей цього типу кривої є те, що вона завжди має точку перетину (критичну точку), яка характеризує оптимальний варіант розміру величини результативної ознаки, і змінює напрямок свого руху лише один раз. Якщо в рівнянні величина a_1 виражена від'ємним числом, a_2 – додатним, то крива змінюватиме напрямок спаду на зростання.

Для розрахунку параметрів рівняння параболи другого порядку використовується така система нормальних рівнянь:

$$\begin{cases} n \cdot a_o + a_1 \cdot \sum x + a_2 \cdot \sum x^2 = \sum y, \\ a_o \cdot \sum x + a_1 \cdot \sum x^2 + a_2 \cdot \sum x^3 = \sum y \cdot x, \\ a_o \cdot \sum x^2 + a_1 \cdot \sum x^3 + a_2 \cdot \sum x^4 = \sum y \cdot x^2. \end{cases} \quad (9.8)$$

Якщо емпірична лінія регресії досліджуваної залежності має вигляд гіперболи, то для визначення зв'язку між ознаками використовують рівняння:

$$y = a_o + \frac{a_1}{x}. \quad (9.9)$$

Для знаходження параметрів рівняння необхідно вирішити систему нормальних рівнянь такого вигляду:

$$\begin{cases} n \cdot a_0 + a_1 \cdot \sum \frac{1}{x} = \sum y, \\ a_0 \cdot \sum \frac{1}{x} + a_1 \cdot \sum \frac{1}{x^2} = \sum \frac{y}{x}. \end{cases} \quad (9.10)$$

Розглянуті залежності (пряма, парабола і гіпербола) не вичерпують усього різноманіття залежностей, що зустрічаються в реальних дослідженнях.

9.4.4. Розрахунок тісноти зв'язку

Важливим завданням кореляційного аналізу є визначення тісноти зв'язку між корелюючими величинами. Кількісним показником тісноти прямолінійного зв'язку результату з одним фактором є ***коефіцієнт парної кореляції***, який обчислюють за формулою:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_y \cdot \sigma_x} = \frac{M}{\sigma_y \cdot \sigma_x}, \quad (9.11)$$

де r – лінійний коефіцієнт кореляції;

σ_x – середнє квадратичне відхилення факторної ознаки;

σ_y – середнє квадратичне відхилення результативної ознаки.

Якщо врахувати, що $\sigma_x = \sqrt{\overline{x^2} - \bar{x}^2}$, а $\sigma_y = \sqrt{\overline{y^2} - \bar{y}^2}$, то найбільш зручною формулою для визначення лінійного коефіцієнта кореляції є:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\left(\overline{x^2} - \bar{x}^2\right) \cdot \left(\overline{y^2} - \bar{y}^2\right)}}. \tag{9.12}$$

При парній залежності коефіцієнт кореляції коливається від 0 до +1 при прямому зв'язку та від 0 до -1 – при зворотному зв'язку. Чим ближче коефіцієнт кореляції до ±1, тим тісніший зв'язок між *y* та *x* і, навпаки, чим ближче коефіцієнт кореляції до 0, тим слабший зв'язок між результативною та факторною ознаками.

Якщо $r<0{,}3$, то зв'язку немає; якщо $r=0{,}3$-$0{,}5$, зв'язок слабкий; якщо $r=0{,}5$-$0{,}7$ – зв'язок середній і якщо $r>0{,}7$ – зв'язок тісний.

Тісноту зв'язку при криволінійних формах залежності визначають за ***індексом кореляції (кореляційного відношення)***:

$$\eta = \sqrt{\frac{\sigma_{y/x}^2}{\sigma_y^2}}, \tag{9.13}$$

де $\sigma_{y/x}^2$ – міжгрупова дисперсія;

σ_y^2 – загальна дисперсія.

Індекс кореляції змінюється в межах від 0 до +1, тобто *завжди є додатною величиною*. Він показує, яку частку у загальному середньоквадратичному відхиленні результативної ознаки займає середньоквадратичне відхилення факторної ознаки.

У статистичній практиці найчастіше використовують такі робочі формули для визначення індексу кореляції:

$$\eta = \sqrt{\frac{\frac{\sum\left(y_x - \bar{y}\right)^2}{n}}{\frac{\sum\left(y - \bar{y}\right)^2}{n}}}; \; \eta = \sqrt{\frac{\sum\left(y_x - \bar{y}\right)^2}{\sum\left(y - \bar{y}\right)^2}}. \tag{9.14}$$

Індекс кореляції можна використати і для визначення тісноти зв'язку при прямолінійній залежності. У цьому разі абсолютна величина індексу кореляції збігається з лінійним коефіцієнтом кореляції. Якщо зв'язок криволінійний, то $i>r$. Математично встановлено, що коли різниця між індексом кореляції і коефіцієнтом кореляції не перевищує 0,1, то гіпотезу про прямолінійність зв'язку можна вважати доведеною.

При використанні для аналізу показників коефіцієнтів кореляції і кореляційних відношень слід мати на увазі таку особливість коефіцієнта кореляції і кореляційного відношення. Коефіцієнт кореляції визначає однакову міру зв'язку між першою і другою ознакою, тобто міра зв'язку між *X* і *Y* така сама, як і між *Y* і *X*, тобто $r_{xy} = r_{yx}$. Кореляційне відношення свідчить про наявність дещо іншого характеру зв'язку між *X* і *Y*, що має прояв у тому, що показник кореляційного відношення між *Y* і *X* не такий самий, як кореляційне відношення між *Y* і *X*. Це ніби парадоксальне ствердження беззаперечно підтверджується результатами досліджень з будь-якими біологічними об'єктами. Дане положення підкреслює принципову і важливу особливість біологічних об'єктів, яка полягає в тому, що зворотні зв'язки в біологічних об'єктах мають дещо різну інтерпретацію. Існують такі корелюючі ознаки, природа яких ставить неможливим зрівноважувати вагомість їх значень. ***Наприклад***, середня маса колоска пшениці безпосередньо залежить від кількості опадів. Але кількість опадів і характер їх випадання зовсім не залежить від того, наскільки збільшується вага колоску. Подібна нерівність зворотних зв'язків між екологічними умовами і реакцією на них живих організмів обумовлює нерівність зворотних зв'язків. Ця нерівність може досягати критичних рівнів, коли кореляційне відношення першої і другої ознаки буде мати значну величину, а кореляційне відношення другої ознаки за першою − наближатись до нуля. Отже, показник кореляційного відношення визначається для оцінки криволінійної залежності між змінними величинами *X* і *Y*.

Після розрахунку показника тісноти зв'язку проводиться перевірка значущості коефіцієнта регресії і коефіцієнта кореляції за допомогою *t*-критерію Стьюдента або дисперсійного критерію Фішера.

Розраховані значення *t*-критерію порівнюються з критичними їхніми значеннями при прийнятому рівні значущості і числі ступенів свободи. У соціально-економічних дослідженнях рівень значущості зазвичай приймається рівним 0,05. Якщо розрахункове значення *t*-критерію більше критичного, параметр визнається значущим, тобто відхиляється гіпотеза про те, що параметр насправді дорівнює нулю і лише в силу випадкових обставин він виявився рівним величині, що перевіряється.

При криволінійній залежності для оцінки значущості коефіцієнтів регресії слід застосовувати *t*-критерій, а для кореляційного відношення – критерій Фішера.

Розраховане значення *F*-критерію потрібно порівняти з його критичним значенням для прийнятого рівня значущості і числі ступенів свободи. Якщо обчислене значення *t*-критерію або *F*-критерію перевищує критичне, то отримане рівняння регресії можна використовувати в практичних розрахунках.

ТЕСТОВІ ЗАВДАННЯ

1. При функціональному зв'язку кожному конкретному значенню факторної ознаки відповідає:

а) певна множина результативної ознаки;

б) конкретне значення результативної ознаки;

в) середнє значення результативної ознаки;

г) обмежена кількість результативних ознак.

2. Прямий – це такий зв'язок між факторною та результативною ознаками, при якому зі:

а) збільшенням факторної ознаки результативна збільшується;

б) збільшенням факторної ознаки результативна зменшується;

в) зменшенням факторної ознаки результативна збільшується;

г) збільшенням факторної ознаки результативна не змінюється.

3. Зворотний – це такий зв'язок між факторною та результативною ознаками, при якому зі:

а) збільшенням факторної ознаки результативна збільшується;

б) збільшенням факторної ознаки результативна зменшується;

в) зменшенням факторної ознаки результативна зменшується;

г) збільшенням факторної ознаки результативна не змінюється.

4. За формою (аналітичним вираженням) зв'язки поділяють на:

а) лінійні та нелінійні;

б) прямі та зворотні;
в) однофакторні та багатофакторні;
г) слабкі та сильні.

5. За напрямом зв'язки поділяють на:
а) лінійні та нелінійні;
б) прямі та зворотні;
в) однофакторні та багатофакторні;
г) слабкі та сильні.

6. За кількістю взаємодіючих факторів зв'язки поділяють на:
а) лінійні та нелінійні;
б) прямі та зворотні;
в) однофакторні та багатофакторні;
г) слабкі та сильні.

7. Лінійний коефіцієнт кореляції змінюється в межах:
а) від -1 до +1;
б) від 0 до +1;
в) від -1 до 0;
г) від -∞до +∞.

8. Кореляційне відношення змінюється в межах:
а) від -1 до +1;
б) від 0 до +1;
в) від -1 до 0;
г) від -∞до +∞.

9. Якщо кореляційне відношення дорівнює нулю, це свідчить про:
а) зворотний функціональний зв'язок;
б) наявність лінійного зв'язку;
в) наявність квадратичного зв'язку;
г) відсутність зв'язку.

10. Якщо лінійний коефіцієнт кореляції дорівнює +1, це свідчить про:
а) щільний лінійний зв'язок;
б) прямий функціональний зв'язок;

в) відсутність зв'язку;

г) зворотний функціональний зв'язок.

11. Якщо лінійний коефіцієнт кореляції дорівнює (-1), це свідчить про:

а) щільний лінійний зв'язок;

б) прямий функціональний зв'язок;

в) відсутність зв'язку;

г) зворотний функціональний зв'язок.

12. Якщо зі зміною факторної ознаки результативна ознака змінюється більш-менш рівномірно, то такий зв'язок описується функцією:

а) лінійною;

б) степеневою;

в) гіперболічною;

г) параболічною.

13. Рівняння, за допомогою яких визначають статистичний зв'язок між корелюючими величинами, називають рівнянням:

а) тренду;

б) кореляції;

в) регресії;

г) детермінації.

14. Параметр a_0 у рівнянні регресії $y_x = a_o + a_1 \cdot x$ характеризує:

а) середній щорічний абсолютний приріст рівнів вирівняного ряду динаміки;

б) зміну результативної ознаки при кожній зміні факторної на одиницю;

в) вирівняний рівень ряду динаміки;

г) не інтерпретується.

15. Параметр a_1 у рівнянні регресії $y_x = a_o + a_1 \cdot x$ характеризує:

а) середній щорічний абсолютний приріст рівнів вирівняного ряду динаміки;

б) зміну результативної ознаки при кожній зміні факторної на одиницю;