МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»
імені ІГОРЯ СІКОРСЬКОГО

В.І. Сущук-Слюсаренко, Р.А. Гадиняк

МАТЕМАТИЧНА СТАТИСТИКА

НАВЧАЛЬНИЙ ПОСІБНИК

для самостійної роботи студентів з дисципліни

«Теорія ймовірностей та математична статистика»

Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського

як навчальний посібник для студентів, які навчаються

за спеціальністю 121 «Інженерія програмного забезпечення»

Київ
КПІ ім. Ігоря Сікорського
2019

Рецензенти: Галицька І.Є., к.т.н., доц.

Відповідальний редактор: Заболотня Т.М., к.т.н., доц.

Гриф надано Методичною радою КПІ ім. Ігоря Сікорського (протокол № 6 від 31.01.2020 р.)

За поданням Вченої ради ФПМ (*протокол № 8 від 27.01.2020 р.*)

Електронне мережне навчальне видання

В.І. Сущук-Слюсаренко, ст.викладач

Р.А. Гадиняк, ст.викладач

МАТЕМАТИЧНА СТАТИСТИКА

НАВЧАЛЬНИЙ ПОСІБНИК

для самостійної роботи студентів з дисципліни

«Теорія ймовірностей та математична статистика»

Математична статистика. Навч.посіб. для самостійної роботи студентів з дисципліни «Теорія ймовірностей та математична статистика» [Електронний ресурс]: навч. посіб. для студ. спеціальності 121 «Інженерія програмного забезпечення»/ В. І. Сущук-Слюсаренко, Р.А.Гадиняк ; : КПІ ім.Ігоря Сікорського.-Електронні текстові данні (1 файл: 941Кбайт). – Київ : КПІ ім.Ігоря Сікорського, 2019. – 60 с.

Навчальний посібник розроблено для більш детального ознайомлення студентів з теоретичними відомостями та практичними прийомами математичної статистики, а також як посібник для використання на практичних заняттях і для самостійної роботи студентів в процесі підготовки дипломних, курсових робіт тощо. Навчально-методичне видання містить приклади застосування методів дісперсного аналізу в різних галузях знань, подані деякі теоретичні матеріали, які стосуються фінансового ринку і обробки часових рядів для використання цих тем під час роботи над різними проектами, наведено приклади розв’язання типових задач, завдання для самостійної роботи студентів можна використовувати, як домашню роботу. Посібник призначений для студентів, які навчаються за спеціальністю 121 «Інженерія програмного забезпечення» факультету прикладної математики «КПІ» ім.Ігоря Сікорського.

Зміст

ВСТУП

Навчальний посібник розроблено для допомоги студентам спеціальності 121 «Інженерія програмного забезпечення» в опануванні тем, які за програмою виносяться на самостійне опрацювання. Методичний посібник містить короткі теоретичні відомості та практичні завдання математичної статистики. Посібник можна використовувати на практичних заняттях і для самостійної роботи студентів в процесі підготовки дипломних, курсових робіт тощо. У виданні містяться приклади застосування методів дисперсного аналізу в різних галузях знань, подані деякі теоретичні матеріали, які стосуються фінансового ринку і обробки часових рядів для використання цих тем під час роботи над різними проектами, наведено приклади розв'язання типових задач, завдання для самостійної роботи студентів можна використовувати, як домашню роботу. Теми 2 і 3 часто використовують при аналізі процесів, які відбуваються на фондових біржах, в банківській сфері тощо. Наведені практичні прийоми використання інформації можуть бути застосовані студентами при розробці наукових проектів та при підготовці магістерської дисертації. Навчально-методичне видання призначене для студентів, які навчаються за спеціальністю 121 «Інженерія програмного забезпечення» факультету прикладної математики «КПІ» ім.Ігоря Сікорського.

ТЕМА 1. ДИСПЕРСІЙНИЙ АНАЛІЗ

1.1. Короткі теоретичні відомості

На практиці часто виникає необхідність узагальнення завдання, тобто перевірки відмінності вибіркових середніх m сукупностей ($m > 2$). Наприклад, потрібно оцінити вплив різних типів алгоритмів стиснення інформації на якість стиснення, тобто у скільки разів заданий алгоритм стискує інформацію, кількості кодів від швидкості роботи алгоритма і таке інше. Для ефективного вирішення такого завдання потрібен новий підхід, який і реалізується в дисперсійному аналізі.

Дисперсійний аналіз — це статистичний метод, призначений для оцінки впливу різних факторів на результат експерименту, а також для наступного планування аналогічних експериментів.

Дисперсійний аналіз був розроблений англійським математиком — статистиком Р.А. Фішером (1918 р.) для обробки результатів агрономічних дослідів по виявленню умов отримання максимального врожаю різних сортів сільськогосподарських культур. Термін «дисперсійний аналіз» Фішер вжив пізніше. За кількістю факторів, вплив яких досліджується, розрізняють однофакторний та багатофакторний дисперсійний аналіз.

Однофакторний дисперсійний аналіз

Однофакторна дисперсійна модель має вигляд: $x_y = \mu + F_i + \varepsilon_{ij}$,

де x_{ij} — значення досліджуваної змінної, отриманої на *i*-му рівні чинника

$(i = 1, 2, ..., m)$ з *j*-им порядковим номером $(j = 1, 2, ..., n)$;

F_i — ефект, обумовлений впливом *i*-го рівня чинника;

ε_{ij} — випадкова компонента, або збурення, викликане впливом неконтрольованих чинників, тобто варіацією змінної в середині окремого рівня.

Під **рівнем чинника** будемо розуміти деяку його міру або стан, наприклад, кількість годин, які було витрачено на дешифрування зображення, потужність алгоритма стиснення або границі довірчого інтервалу і таке інше.

Основні передумови дисперсійного аналізу :

1. Математичне сподівання збурення ε_{ij} дорівнює нулю для будь-яких *i*, тобто: $M(\varepsilon_{ij}) = 0$. (1.1)
2. Збурення ε_{ij} взаємно незалежні.
3. Дисперсія збурення ε_{ij} (чи змінної x_{ij}) постійна для будь-яких *i* і *j*, тобто: $D(\varepsilon_{ij}) = \sigma^2$ (1.2)
4. Збурення ε_{ij} (чи змінна x_{ij}) має нормальний закон розподілу $N(0; \sigma^2)$.

Вплив рівнів чинника може бути як фіксованим (або систематичним) (модель I), так і випадковим (модель II).

Нехай, наприклад, необхідно з'ясувати, чи є суттєві відмінності між програмними продуктами по деякому показнику якості, тобто перевірити вплив на якість одного чинника, наприклад, кількості кодів в програмі.

Якщо включити в дослідження усі програми, то вплив рівня такого чинника систематичний (модель I), а отримані висновки можуть бути застосовані тільки до тих окремих програм, які розглядалися при дослідженні; якщо ж включити тільки відібрану випадково частину програм, то вплив чинника випадковий (модель II). У багатофакторних комплексах можлива змішана модель III, в якій окремі чинники будуть випадковими, а інші — фіксованими. Нехай є m партій програмних продуктів, які бажано продати. З кожної партії відібрано відповідно $n_1, n_2, ..., n_m$ програм (вважатимемо, що $n_1 = n_2 = ... = n_m = n$). Значення показника якості цих програм представимо у вигляді матриці спостережень

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix} = (x_{ij}), (i = 1,2,\dots,m;\ j = 1,2,...,n).$$

Необхідно перевірити істотність впливу партії програмних продуктів на їх якість. Якщо вважати, що елементи рядків матриці спостережень — це чисельні значення (реалізації) випадкових величин $X_1, X_2 \dots, X_m$, які виражають якість програм, і мають нормальний закон розподілу з математичними сподіванням відповідно $a_1, a_2 \dots, a_m$ і однаковими дисперсіями σ^2, то це завдання зводиться до перевірки нульової гіпотези $H_0 : a_1 = a_2 = ... = a_m$. Позначимо усереднення по якому-небудь індексу зірочкою замість індексу, тоді середній показник якості виробів *i*-ї партії, або групове середнє для *i*-го рівня фактора, набере вигляду:

$$\overline{x_{i*}} = \frac{\sum_{j=1}^{n} x_{ij}}{n}, \qquad (1.3)$$

а загальне середнє – $\bar{x}_{**} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}}{mn} = \frac{\sum_{i=1}^{n} \bar{x}_{i*}}{m}.$ (1.4)

Розглянемо суму квадратів відхилень спостережень x_{ij} від загального середнього $\bar{x}_{**}$:

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{**})^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(\bar{x}_{i*} - \bar{x}_{**})^2 + \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{i*})^2 +$$

$$+2\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{**})(\bar{x}_{i*} - \bar{x}_{**})^2, \qquad (1.5)$$

або $Q = Q_1 + Q_2 + Q_3$. Останній доданок

$$Q_3 = 2\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{i*})(\bar{x}_{i*} - \bar{x}_{**}) = 2\sum_{i=1}^{m}(\bar{x}_{i*} - \bar{x}_{**})(\bar{x}_{i*} - \bar{x}_{**})\sum_{j=1}^{n}(x_{ij} - \bar{x}_{i*}) = 0$$

оскільки сума відхилень значень змінної від її середнього, тобто $\sum_{j=1}^{n}(x_{ij} - \bar{x}_{i*})$ дорівнює нулю. Перший доданок можна записати у вигляді:

$$Q_1 = \sum_{i=1}^{m}\sum_{j=1}^{n}(\bar{x}_{i*} - \bar{x}_{**})^2 = n\sum_{i=1}^{m}(\bar{x}_{i*} - \bar{x}_{**})^2. \qquad (1.6)$$

В результаті отримаємо наступну тотожність: $Q = Q_1 + Q_2$, де (1.7)

$Q = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{**})^2$ — *загальна*, або *повна*, сума квадратів відхилень;

$Q_1 = n\sum_{i=1}^{m}(\bar{x}_{i*} - \bar{x}_{**})^2$ — сума квадратів відхилень групових середніх від

загального середнього, або *міжгрупова* (*факторна*) сума квадратів відхилень;

$Q_2 = n\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{i*})^2$ — сума квадратів відхилень спостережень від групових середніх, або *внутрішньогрупова* (*залишкова*) сума квадратів відхилень.

Формула (1.7) містить основну ідею дисперсійного аналізу. Якщо поділити обидві частини цієї рівності на обсяг спостережень, то отримаємо правило додавання дисперсій. Рівність (1.7) показує, що загальна варіація показника якості (Q), складається з двох компонент — Q_1 і Q_2, які характеризують мінливість досліджуваного показника між партіями (Q_1) і мінливість "в середині" партії (Q_2).

У дисперсійному аналізі аналізуються не самі суми квадратів відхилень, а так звані *середні квадрати*, які є незміщеними оцінками відповідних дисперсій, що знаходяться діленням сум квадратів відхилень на відповідне число ступенів свободи. Число ступенів свободи визначається як загальне число спостережень мінус число рівнянь, які їх зв'язують. Тому для середнього квадрата s_1^2, що є незміщеною оцінкою міжгрупової дисперсії, число ступенів свободи $k_1 = m - 1$, оскільки при його розрахунку використовуються m групових середніх, пов'язаних між собою одним рівнянням (1.4). А для середнього квадрата s_2^2, який є незміщеною оцінкою внутрішньогрупової дисперсії, число ступенів свободи $k_2 = mn - m$, бо при її розрахунку використовуються усі mn спостережень, пов'язаних між собою m рівняннями (1.3). Таким чином, $s_1^2 = Q_1 / (m-1);\ \ s_2^2 = Q_2 / (mn - m).$

Знайдемо математичні сподівання середніх квадратів s_1^2 і s_2^2, підставивши в їх формули вираз x_{ij} (1.1) через параметри моделі.

$$M(s_1^2) = \frac{n}{m-1} M\left(\sum_{i=1}^{m} (\mu + F_i + \varepsilon_{i*} - \mu - F_* - \varepsilon_{**})^2 \right) =$$

$$= \frac{n}{m-1} M\left(2\sum_{i=1}^{m} ((F_i - F_*) + (\varepsilon_{i*} - \varepsilon_{**}))^2 \right) = \frac{n}{m-n} M\left(\sum_{i=1}^{m} (F_i - F_*)^2 \right) +$$

$$+ \frac{n}{m-1} M\left(2\sum_{i=1}^{m} (F_i - F_*)(\varepsilon_{i*} - \varepsilon_{**}) \right) + \frac{n}{m-n} M\left(\sum_{i=1}^{m} (\varepsilon_{i*} - \varepsilon_{**})^2 \right) =$$

$$= \frac{n}{m-1} M\left(2\sum_{i=1}^{m} (F_i - F_*) \right) + \sigma^2 \qquad (1.8)$$

бо $M\left(2\sum_{i=1}^{m} (F_i - F_*)(\varepsilon_{i*} - \varepsilon_{**}) \right) = 0$ з урахуванням властивостей математичного сподівання, а

$$M(s_2^2) = \frac{1}{m(n-1)} M\left(\sum_{i=1}^{m} \sum_{j=1}^{n} (\mu + F_i + \varepsilon_{ij} - \mu - F_i - \varepsilon_{i*})^2 \right) =$$

$$= \frac{1}{m} M\left(\frac{\sum_{i=1}^{m} \sum_{j=1}^{n} (\varepsilon_{ij} - \varepsilon_{**})^2}{n-1} \right) = \frac{1}{m} \sum_{i=1}^{m} \sigma_{\varepsilon_{ij}}^2 = \frac{1}{m} \cdot m\sigma^2 = \sigma^2. \qquad (1.9)$$

Схему дисперсійного аналізу представимо у вигляді таблиці 1.1. Для моделі I (із фіксованими рівнями фактора) $F_i\ (i = 1,2,...,m)$ — величини невипадкові, тому $M(s_1^2) = n\sum_{i=1}^{n} (F_i - F_*)^2 / (m-1) + \sigma^2.$ Гіпотеза H_0 набуде вигляду $F_i = F_*\ (i = 1,2,...,m)$, тобто вплив усіх рівнів чинника

однаковий. У разі справедливості цієї гіпотези $M(s_1^2) = M(s_2^2) = \sigma^2$.

Для випадкової моделі II доданок F_i у виразі (1.1) — величина випадкова. Позначивши її дисперсію $\sigma_F^2 = M\left(\sum_{i=1}^{m}(F_i - F_*)^2 / (m-1)\right)$, отримаємо з (1.8):

$$M(s_1^2) = n\sigma_F^2 + \sigma^2 \qquad (1.10)$$

і, як і в моделі I, $M(s_2^2) = \sigma^2$.

Таблиця 1.1. Схема дисперсійного аналізу

Компоненти дисперсії	Сума квадратів	Число ступенів свободи	Середній квадрат	Математичне сподівання середнього квадрату
Міжгрупова	$Q_1 = n\sum_{i=1}^{m}(\bar{x}_{i*} - \bar{x}_{**})^2$	$m-1$	$s_1^2 = \frac{Q_1}{m-1}$	$M(s_1^2) = \begin{cases} \frac{n}{m-1}\sum_{i=1}^{m}(F_i - F_*)^2 + \\ +\sigma^2 \text{ (I модель)} \\ n\sigma_F^2 + \sigma^2 \text{ (II модель)} \end{cases}$
Внутрішньо групова	$Q_2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{i*})^2$	$mn-m$	$s_2^2 = \frac{Q_2}{mn-m}$	$M(s_2^2) = \sigma^2$
Загальна	$Q = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - x_{**})^2$	$mn-1$		

У разі справедливості нульової гіпотези H_0, яка для моделі II набирає вигляду $\sigma_F^2 = 0$, маємо: $M(s_1^2) = M(s_2^2) = \sigma^2$. Отже, для однофакторного аналіза як для моделі I, так і моделі II *середні квадрати* s_1^2 і s_2^2 є

незміщеними і незалежними оцінками однієї і тієї ж дисперсії σ^2.

Перевірка нульової гіпотези H_0 — це перевірки відмінності незміщених вибіркових оцінок S_1^2 і S_2^2 дисперсії σ^2. Гіпотеза H_0 відкидається, якщо фактично обчислене значення статистики $F = \frac{S_1^2}{S_2^2}$ більше критичного $F_{\alpha;k_1;k_2}$, визначеного на рівні значущості α при числі ступенів свободи $k = mn - m$, і приймається, якщо $F \le F_{\alpha;k_1;k_2}$. Спростування гіпотези H_0 означає наявність істотних відмінностей в якості факторів, які перевіряються, на даному рівні значущості.

◄**Приклад 1.1.** Є чотири партії ІС для виробництва гаджетів. З кожної партії відібрано по п'ять зразків і проведені випробування на визначення часу функціонування в умовах екстремального навантаження. Результати випробувань приведені в таблиці 1.2. Необхідно з'ясувати, чи істотний вплив різних партій ІС на час функціонування. Прийняти $\alpha = 0{,}05.$

Таблиця 1.2. Данні до задачі 1.1

Номер партії	Час функціонування в екстремальних умовах (год)				
1	200	140	170	145	165
2	190	150	210	150	150
3	230	190	200	190	200
4	150	170	150	170	180

Розв'язання. Маємо $m = 4,\ n = 5.$ Знайдемо середні значення часу функціонування для кожної партії по формулі (1.3) :

$\bar{x}_{1*} = (200+140+170+145+165)/5 = 164(год)$ і, аналогічно

$\bar{x}_{2*} = 170;\ \bar{x}_{3*} = 202;\ \bar{x}_{4*} = 164(год)$.

Середнє значення часу функціонування всіх відібраних зразків за формулою (1.4): $\bar{x}_{**} = (200+140+...+170+180)/20 = 175(год)$.

Або, інакше, через групові середні:

$$\bar{x}_{**} = (164+170+202+164)/4 = 175(год).$$

Обчислимо суми квадратів відхилень за формулами (1.5), (1.6):

$$Q_1 = 5\sum_{i=1}^{4}(\bar{x}_{i*} - \bar{x}_{**})^2 = 5\left((164-175)^2 + (170-175)^2 + (202-175)^2 + \right.$$
$$\left. + (164-175)^2\right) =$$
$$= 5 \cdot 996 = 4980;$$
$$Q_2 = \sum_{i=1}^{4}\sum_{j=1}^{5}(\bar{x}_{ij} - \bar{x}_{i*})^2 = (200-164)^2 + ... + (165-164)^2 +$$
$$(190-170)^2 + ... + (150-170)^2 + ... + (230-202)^2 + (200-202)^2 +$$
$$+ (150-164)^2 + ... + (180-164)^2 = 7270;$$

$$Q = \sum_{i=1}^{4}\sum_{j=1}^{5}(\bar{x}_{ij} - \bar{x}_{**})^2 = (200-275)^2 + (140-175)^2 + ... +$$
$$+ (170-175)^2 + (180-175)^2 = 12250.$$

Відповідне число ступенів свободи для цих сум $m-1=3;\ mn-m=5\cdot 4-4=16;\ mn-1=5\cdot 4-1=19.$ Результати зведемо в табл. 1.3.

Фактичне значення статистики $F = \frac{s_1^2}{s_2^2} = \frac{1660}{454,4} = 3,65.$ За таблицею критичне значення F – критерію Фішера –Снедекора на рівні значущості $\alpha = 0,05$ при $k_1 = 3$ і $k_2 = 16$ ступенях свободи $F_{0,05;3;16} = 3,24$.

Оскільки $F > F_{0,05;3;16}$, то нульова гіпотеза відкидається, тобто на рівні значущості $\alpha = 0,05$ (з надійністю 0,95) відмінність між партіями

Таблиця 1.3.Результати обчислень до задачі 1.1

Компоненти дисперсії	Суми квадратів	Число ступенів свободи	Середні квадрати
Міжгрупова	4980	3	1660,0
Внутрішньогрупова	7270	16	454,4
Загальна	12250	19	

ІС істотно впливає на час функціонування гаджетів в умовах екстремальних навантажень.►

Поняття про двофакторний дисперсійний аналіз

Припустимо, що в розглянутій задачі про час функціонування ІС з різних (*m*) партій вироби виготовлялися різними (*l*) виробниками. Потрібно з'ясувати, чи є суттєві відмінності в якості виробів по кожному фактору:

A — партія виробів, B — виробник. В результаті приходимо до задачі *двофакторного дисперсійного аналізу.*

Усі наявні дані представимо у вигляді табл. 1.4, в якій по рядках — рівні A_i фактора A, по стовпцях — рівні B_j чинника B, а у відповідних комірках таблиці знаходяться значення показника якості виробів $x_{ijk}\left(i=1,2,...,m;\ j=1,2,...,l;\ k=1,2,...,n\right)$.

Двофакторна дисперсійна модель має вигляд:

$$x_{ijk} = \mu + F_i + G_j + I_{ij} + \varepsilon_{ijk}, \qquad (1.14)$$

де x_{ijk} — значення спостереження в комірці *ij* з номером *k*;

μ - загальна середня;

F_i - ефект, обумовлений впливом *i*-го рівня фактора *A*;

G_j - ефект, обумовлений впливом *j*-го рівня фактора *B*;

I_{ij} - ефект, обумовлений взаємодією двох факторів, тобто відхилення від середнього за спостереженнями у комірці *ij* від суми перших трьох доданків в моделі (1.14);

ε_{ijk} - збурення, обумовлене варіацією змінної в середині окремої комірки.

Таблиця 1.4. Двофакторний дисперсійний аналіз

A \ B	B_1	B_2	…	B_j	…	B_l
A_1	$x_{111},...,x_{11k}$	$x_{121},...,x_{12k}$	…	$x_{1j1},...,x_{1jk}$	…	$x_{1l1},...,x_{1lk}$
A_2	$x_{211},...,x_{21k}$	$x_{221},...,x_{22k}$	…	$x_{2j1},...,x_{2jk}$	…	$x_{2l1},...,x_{2lk}$
⋮	…	…	…	…	…	…
A_i	$x_{i11},...,x_{i1k}$	$x_{i21},...,x_{i2k}$	…	$x_{ij1},...,x_{ijk}$	…	$x_{il1},...,x_{ilk}$
⋮	…	…	…	…	…	…
A_m	$x_{m11},...,x_{m1k}$	$x_{m21},...,x_{m2k}$	…	$x_{mj1},...,x_{mjk}$	…	$x_{ml1},...,x_{mlk}$

Будемо вважати, що ε_{ijk} має нормальний закон розподілу $N\left(0;\sigma^2\right)$, а всі математичні сподівання F_*, G_*, I_{i*}, I_{*j} дорівнюють нулю. Групові середні знаходяться за формулами:

по комірці – $$\overline{x}_{ij*} = \frac{\sum_{k=1}^{n} x_{ijk}}{n}, \qquad (1.15)$$

по рядку – $\overline{x}_{i**} = \dfrac{\sum_{j=1}^{l} \overline{x}_{ij*}}{l}$, (1.16)

по стовпцю – $\overline{x}_{*j*} = \dfrac{\sum_{i=1}^{m} \overline{x}_{ij*}}{m}$. (1.17)

загальне середнє – $\overline{x}_{***} = \dfrac{\sum_{i=1}^{m}\sum_{j=1}^{l} \overline{x}_{ij*}}{ml}$. (1.18)

Таблиця дисперсійного аналіза має вид:

Таблиця 1.5. Формули для двофакторного дисперсійного аналіза

Компоненти дисперсії	Сума квадратів	Число ступенів свободи	Середні квадрати
Міжгрупова (фактор *A*)	$Q_1 = ln\sum_{i=1}^{m}(\overline{x}_{i**} - \overline{x}_{***})^2$	*m*-1	$s_1^2 = \dfrac{Q_1}{m-1}$
Міжгрупова (фактор *B*)	$Q_2 = mn\sum_{j=1}^{l}(\overline{x}_{*j*} - \overline{x}_{***})^2$	*l*-1	$s_2^2 = \dfrac{Q_2}{l-1}$
Взаємодія (*AB*)	$Q_3 = n\sum_{i=1}^{m}\sum_{j=1}^{l}(\overline{x}_{ij*} - \overline{x}_{i**} - \overline{x}_{*j*} + \overline{x}_{***})^2$	(*m*-1)(*l*-1)	$s_3^2 = \dfrac{Q_3}{(m-1)(l-1)}$
Залишкова	$Q_4 = \sum_{i=1}^{m}\sum_{j=1}^{l}\sum_{k=1}^{n}(x_{ijk} - \overline{x}_{ij*})^2$	*mln-ml*	$s_4^2 = \dfrac{Q_4}{(mnl)(ml)}$
Загальна	$Q = \sum_{i=1}^{m}\sum_{j=1}^{l}\sum_{k=1}^{n}(x_{ijk} - \overline{x}_{***})^2$	*mln*-1	

Можна показати, що перевірка нульових гіпотез H_A, H_B, H_{AB} про відсутність впливу на розглянуту змінну факторів *A*, *B* та їх взаємодії *AB*

здійснюється порівнянням відношень $\frac{s_1^2}{s_4^2}, \frac{s_2^2}{s_4^2}, \frac{s_3^2}{s_4^2}$ (для моделі I з фіксованими рівнями факторів) або відношень $\frac{s_1^2}{s_3^2}, \frac{s_2^2}{s_3^2}, \frac{s_3^2}{s_4^2}$ (для випадкової моделі II) з відповідними табличними значеннями F— критерію Фішера-Снедекора. Для змішаної моделі III перевірка гіпотез щодо факторів із фіксованими рівнями проводиться так само, як у моделі II, а факторів із випадковими рівнями — як в моделі I. Якщо n = 1 (при одному спостереженні в комірці) не всі нульові гіпотези можуть бути перевірені, оскільки випадає компонента Q_3 із загальної суми квадратів відхилень, а з нею і середній квадрат s_3^2, бо в цьому випадку не може бути мови про взаємодію чинників.

◄ **Приклад 1.2**. У табл. 1.6 наведений час роботи в звичайних умовах 18 відібраних для дослідження ІС (діб) в залежності від виробника (фактор *A*) та технологічності процесу виготовлення (фактор *B*). Необхідно на рівні значущості $\alpha = 0,05$ оцінити суттєвість (достовірність) впливу кожного фактора і їх взаємодії на час роботи ІС.

Таблиця 1.6. Данні до задачі 1.2

Добовий обсяг виробництва ІС (фактор *A*)	Час, який витрачається на виробництво однієї ІС, (хв) (фактор *B*)	
	B_1=80	B_2=100
A_1 = 30	530, 540, 550	600, 620, 580
A_2 = 100	490, 510, 520	550, 540, 560
A_3 = 300	430, 420, 450	470, 460, 430

Розв'язання. Маємо $m=3;\ l=2;\ n=3.$ Визначимо середні значення часу роботи:

в комірках – за (1.15): $\bar{x}_{11*}=\dfrac{530+540+550}{3}=540;$

і аналогічно,

$\bar{x}_{12*}=600;\ \bar{x}_{21*}=506,7;\ \bar{x}_{22*}=550;\ \bar{x}_{31*}=433,3;\ \bar{x}_{32*}=453,3;$

по рядках – за (1.16): $\bar{x}_{1**}=\dfrac{540+600}{2}=570$

і, аналогічно, $\bar{x}_{2**}=528,4;\ \bar{x}_{3**}=443,2;$

по стовпцях — за (1.17): $\bar{x}_{*1*}=\dfrac{540+506,7+433,3}{3}=493,3$

і, аналогічно, $\bar{x}_{*2*}=534,4.$

Загальний час роботи — за (1.18):

$$\bar{x}_{***}=\frac{540+600+506,7+550+433,3+453,3}{6}=513,9\left(\partial i\acute{o}\right)$$

Всі середні значення часу роботи (діб) відобразимо у табл. 1.7

Таблиця 1.7. Результати обчислень задачі 1.2

Добовий обсяг виробництва ІС (фактор *A*)	Час, який витрачається на виробництво однієї ІС, (хв) (фактор *B*)		
	$B_1=80$	$B_2=100$	x_{i**}
$A_1=30$	$\bar{x}_{11*}=540,0$	$\bar{x}_{12*}=600,0$	$\bar{x}_{1**}=570,0$
$A_2=100$	$\bar{x}_{21*}=506,7$	$\bar{x}_{22*}=550,0$	$\bar{x}_{2**}=528,4$
$A_3=300$	$\bar{x}_{31*}=433,3$	$\bar{x}_{32*}=453,3$	$\bar{x}_{3**}=443,3$
$\bar{x}_{*j*}$	$\bar{x}_{*1*}=493,3$	$\bar{x}_{*2*}=534,4$	$\bar{x}_{***}=513,9$

З таблиці 1.7 випливає, що зі збільшенням обсягу виробництва у групі середній час роботи ІС в середньому зменшується, а при збільшенні часу, який витрачається на виробництво однієї ІС — в середньому збільшується. Але чи є ця тенденція достовірною або пояснюється випадковими причинами? Для відповіді на це питання за формулами табл. 1.5 обчислимо необхідні суми квадратів відхилень:

$$Q_1 = 2 \cdot 3\left((570-513{,}9)^2 + (528-513{,}9)^2 + (443{,}2-513{,}9)^2\right) = 50011{,}1;$$

$$Q_2 = 3 \cdot 3\left((493{,}3-513{,}9)^2 + (534{,}4-513{,}9)^2\right) = 7605{,}6;$$

$$Q_3 = 3\left((540-570-493{,}3+513{,}9)^2 + \ldots + (453{,}3-443{,}3-534{,}4+513{,}9)^2 + (443{,}2-513{,}9)^2\right) = 1211{,}1;$$

$$Q_4 = (530-540)^2 + \ldots + (550-540)^2 + (600-600)^2 + \ldots +$$

$$+ (580-600)^2 + \ldots + (470-453{,}3)^2 + \ldots + (430-453{,}3)^2 = 3000{,}0;$$

$$Q = (530-513{,}9)^2 + (540-513{,}9)^2 + (430-513{,}9)^2 = 61827{,}8$$

Середні квадрати знаходимо діленням отриманих сум на відповідну їм кількість ступенів свободи $m-1=2$; $l-1=1$; $(m-1)(l-1)=2$; $mln-ml=18-6=12$; $mln-1=18-1=17$.

Результати зведемо в табл. 1.8.

Очевидно, що дані фактори мають фіксовані рівні, тобто ми знаходимося в рамках моделі I. Тому для перевірки істотності впливу факторів A, B та їх взаємодії AB необхідно знайти відношення:

$$F_A = \frac{s_1^2}{s_4^2} = \frac{25005{,}5}{250{,}0} = 100{,}0;\ F_B = \frac{s_2^2}{s_4^2} = \frac{7605{,}6}{250{,}0} = 30{,}4;\ F_{AB} = \frac{s_3^2}{s_4^2} = \frac{605{,}6}{250{,}0} = 2{,}42;$$

і порівняти їх з табличними значеннями відповідно

$$F_{0{,}05;2;12} = 3{,}88;\ F_{0{,}05;1;12} = 4{,}75;\ F_{0{,}05;2;12} = 3{,}88.$$

Таблиця 1.8. Аналіз результатів обчислень

Компонента дисперсії	Суми квадратів	Число ступенів свободи	Середні квадрати
Міжгрупова (фактор A)	$Q_1 = 50011,1$	2	$s_1^2 = 25005,5$
Міжгрупова (фактор B)	$Q_2 = 7605,6$	1	$s_2^2 = 7605,6$
Взаємодія (AB)	$Q_3 = 1211,1$	2	$s_3^2 = 605,6$
Залишкова	$Q_4 = 3000,0$	12	$s_4^2 = 250,0$
Загальна	$Q = 61827,8$	17	

Оскільки $F_A > F_{0,05;2;12}$ і $F_B > F_{0,05;1;12}$, то вплив обсягу виробництва (фактору A) та часу, що витрачений на виробництво однієї ІС (фактора B) є істотним. В силу того що $F_{AB} > F_{0,05;2;12}$, взаємодія зазначених факторів незначна. ►

При вирішенні реальних завдань методом дисперсійного аналізу використовуються *статистичні програмні пакети.*

Відхилення від основних передумов дисперсійного аналізу — нормальності розподілу досліджуваної змінної і рівності дисперсій в групах (якщо воно не надмірне) — не позначається істотно на результатах при рівному числі спостережень в групах, але може бути дуже чутливим при нерівному їх числі. Крім того, при нерівному числі спостережень різко

зростає складність апарату дисперсійного аналізу. Тому рекомендується планувати схему з рівним числом спостережень в групах, а якщо зустрічаються відсутні дані, то відшкодовувати їх середніми значеннями інших спостережень. При цьому, проте, штучно введені відсутні дані не слід враховувати при обчисленні ступенів свободи.

Контрольні питання

1. В чому полягає сутність дисперсійного аналізу?
2. За яких умов застосовують однофакторний дисперсійний аналіз?
3. За яких умов застосовують двофакторний дисперсійний аналіз?
4. Як інтерпретувати результати дисперсійного аналізу?

Вправи для самостійної роботи

1. На протязі шести років використовувалися п’ять різних технологій по виготовленню чипів для маркування товарів (год$\times 10^5$). Данні, які було отримано в процесі дослідження, наведено в таблиці.

Рік	Технологія (фактор А)				
	А1	А2	А3	А4	А5
1	1,2	0,6	0,9	1,7	1,0
2	1,1	1,1	0,6	1,4	1,4
3	1,0	0,8	0,8	1,3	1,1
4	1,3	0,7	1,0	1,5	0,9
5	1,1	0,7	1,0	1,2	1,2
6	0,8	0,9	1,1	1,3	1,5
Всього	6,5	4,8	5,4	8,4	7,1

На рівні значущості $\alpha = 0,05$ дослідити вплив різних технологій на час роботи чипів.

ТЕМА 2. АНАЛІЗ ЧАСОВИХ РЯДІВ

2.1. Загальні відомості про часові ряди та їх аналіз

Під **часовим рядом** (**динамічним рядом**, або **рядом динаміки**) розуміють ряд спостережень деякої ознаки (випадкової величини) X в послідовні рівновіддалені моменти часу. Окремі спостереження називаються **рівнями** ряду, які позначатимемо $x_t \ (t=1,2,...,n)$, де n — число рівнів.

У таблиці 2.1 наведені дані, які відображають ціну і попит на певний товар за восьмирічний період (ум. од.), тобто два часових ряди — ціни товару x_t, і попиту y_t на нього.

Таблиця 2.1. Приклад часового ряда

Рік, t	1	2	3	4	5	6	7	8
Ціна, x_t	492	462	350	317	340	351	368	381
Попит, y_t	213	171	291	309	317	362	351	361

На рис. 2.1 часовий ряд y_t, зображений графічно. У загальному вигляді при дослідженні часового ряду x_t виділяються декілька складових: $x_t = u_t + \nu_t + c_t + \varepsilon_t$, $(t=1,2,...,n)$, де

u_t — **тренд**, компонента, яка змінюється повільно і така, що описує вплив довготривалих чинників, тобто тривалу тенденцію зміни ознаки (наприклад, зростання населення, економічний розвиток, зміна структури споживання і таке інше);

ν_t — **сезонна компонента**, яка відображає повторюваність процесів протягом не дуже тривалого періоду (року, іноді місяця, тижня і т.і., наприклад, обсяг продажів товарів або перевезень пасажирів у різні пори року);

c_t — **циклічна компонента**, яка відображає повторюваність процесів протягом тривалих періодів (наприклад, вплив хвиль економічної активності, демографічних «ям», циклів сонячної активності, тощо);

ε_t — **випадкова компонента**, яка відображає вплив випадкових чинників, які не підлягають обліку і реєстрації.

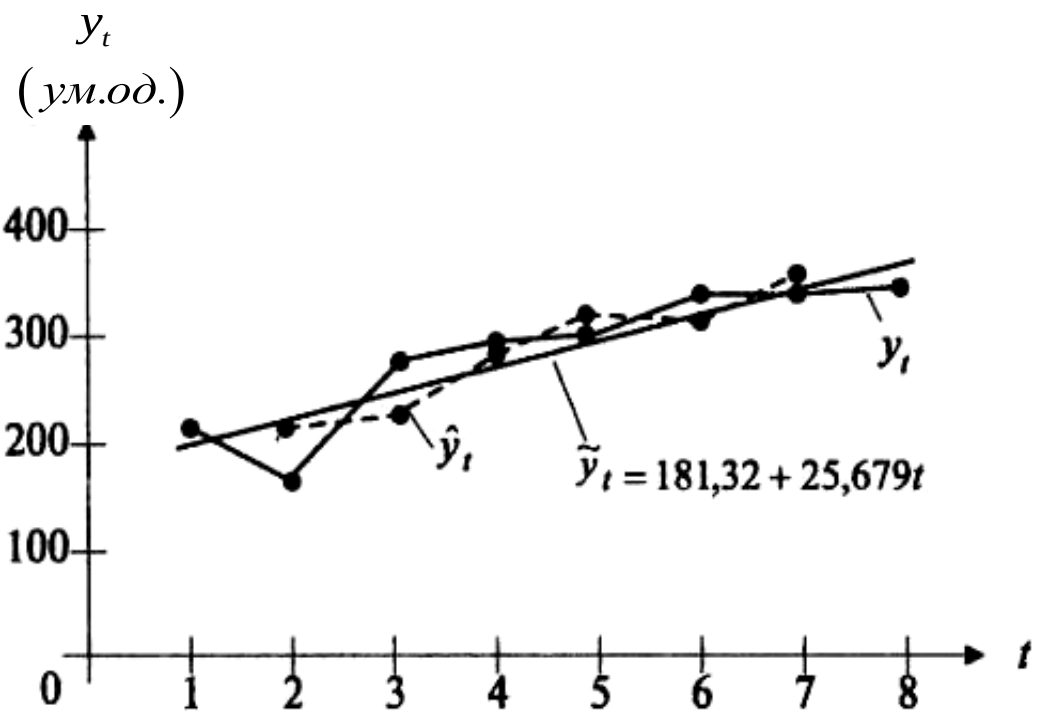

Рис. 2.1. Часовий ряд попиту y_t на товар

На відміну від ε_t, перші три складові (компоненти) u_t, ν_t, c_t є закономірними, невипадковими.

Найважливішим класичним завданням при дослідженні часових рядів є виявлення та статистична оцінка основної тенденції розвитку досліджуваного процесу і відхилень від неї.

Основні етапи аналізу часових рядів:

1. Графічне представлення та опис поведінки часового ряду.
2. Виділення і видалення закономірних (невипадкових) складових часового ряду (тренду, сезонних і циклічних складових).
3. Згладжування і фільтрація (видалення низько- або високочастотних складових часового ряду).
4. Дослідження випадкової складової часового ряду, побудова і перевірка адекватності математичної моделі для її описування.
5. Прогнозування розвитку досліджуваного процесу на основі наявного часового ряду.
6. Дослідження взаємозв'язку між різними часовими рядами.

Серед найбільш поширених методів аналізу часових рядів виділимо **кореляційний і спектральний аналізи, моделі авторегресії та ковзного середнього.**

Часовий ряд $x_1, x_2, ..., x_i, ..., x_n$ розглядається як одна з реалізацій (траєкторій) випадкового процесу $X(t)$. Слід мати на увазі принципові відмінності часового ряду $x_t \ (t = 1,2,...,n)$ від послідовності спостережень $x_1, x_2, ..., x_n$, які утворюють випадкову вибірку. По-перше, на відміну від елементів вибірки члени часового ряду, як правило, не є статистично незалежними. По-друге, члени часового ряду не є однаково розподіленими.

2.2. Стаціонарні часові ряди та їх характеристики. Автокореляційна функція

Важливе значення в аналізі часових рядів мають **стаціонарні** часові ряди, ймовірносні властивості яких не змінюються в часі. Стаціонарні часові ряди застосовуються, зокрема, при описуванні випадкових складових рядів.

Часовий ряд $x_t \ (t = 1,2,...,n)$ називається **строго стаціонарним** (або стаціонарним у вузькому сенсі), якщо спільний розподіл ймовірностей *n* спостережень $x_1, x_2, ..., x_n$ такий самий, як і n спостережень $x_{1+\tau}, x_{2+\tau}, ..., x_{n+\tau}$ при будь-яких *n, t* і τ. Властивості строго стаціонарних рядів x_t не залежать від моменту t, тобто закон розподілу і його числові характеристики не залежать від *t*. Отже, математичне сподівання $a_x(t) = a$ і середнє квадратичне відхилення $\sigma_x(t) = \sigma$ можуть бути оцінені за спостереженнями $x_t \ (t = 1,2,...,n)$ за формулами:

$$\overline{x}_t = \frac{\sum_{t=1}^{n} x_t}{n}, \qquad (2.1)$$

$$s_t^2 = \frac{\sum_{t=1}^{n} (x_t - \overline{x}_t)^2}{n}. \qquad (2.2)$$

Ступінь зв'язку між послідовностями спостережень часового ряду $x_1, x_2, ..., x_n$ і $x_{1+\tau}, x_{2+\tau}, ..., x_{n+\tau}$ (зсунутих один до одного на τ одиниць, або, як то кажуть, з *лагом* τ) може бути визначена за допомогою коефіцієнта кореляції

$$\rho(\tau)=\frac{M\left(\left(x_t-a\right)\left(x_{t+\tau}-a\right)\right)}{\sigma_x\left(t\right)\sigma_x\left(t+\tau\right)}=\frac{M\left(\left(x_t-a\right)\left(x_{t+\tau}-a\right)\right)}{\sigma^2} \qquad (2.3)$$

бо $M(x_t)=M(x_{t+\tau})=a,\ \sigma_x(t)=\sigma_x(t+\tau)=\sigma.$

Оскільки коефіцієнт ρ(τ) вимірює кореляцію між членами одного і того самого ряду, його називають **коефіцієнтом автокореляції**, а залежність ρ(τ) **— автокореляційною функцією**. Через стаціонарність часового ряду $x_t\ \left(t=1,2,...,n\right)$ автокореляційна функція ρ(τ) залежить тільки від лага *τ*, причому $\rho\left(-\tau\right)=\rho\left(\tau\right)$, тобто при вивченні ρ(τ) можна обмежитися розглядом тільки додатних значень τ.

Статистичною оцінкою $\rho(\tau)$ є **вибірковий коефіцієнт автокореляції** r_t, який визначається за формулою коефіцієнта кореляції, в якій $x_i=x_t, y_i=y_{t+\tau}$, а n замінюється на $n-\tau$:

$$r_\tau=\frac{(n-\tau)\sum\limits_{t=1}^{n-\tau}x_t x_{t+\tau}-\sum\limits_{t=1}^{n-\tau}x_t\sum\limits_{t=1}^{n-\tau}x_{t+\tau}}{\sqrt{(n-\tau)\sum\limits_{t=1}^{n-\tau}x_t^2-\left(\sum\limits_{t=1}^{n-\tau}x_t\right)^2}\sqrt{(n-\tau)\sum\limits_{t=1}^{n-\tau}x_{t+\tau}^2-\left(\sum\limits_{t=1}^{n-\tau}x_{t+\tau}\right)^2}} \qquad (2.4)$$

Функцію r_t, називають **вибірковою автокореляційною функцією**, а її графік — **корелограмою**.

При розрахунку r_t, слід пам'ятати, що із збільшенням *τ* число $n-\tau$ пар спостережень $x_t, x_{t+\tau}$ зменшується, тому лаг τ повинен бути таким, щоб число $n-\tau$ було достатнім для визначення r_t. Зазвичай орієнтуються на співвідношення $\tau<\frac{n}{4}$.

◄**Приклад 2.1.** За даними табл. 2.1 для часового ряду y_t, знайти середнє значення, середнє квадратичне відхилення і коефіцієнти автокореляції

(для лагів τ =1;2).

Розв'язання. Середнє значення часового ряду знаходимо за формулою (2.1): $y_t = \frac{213 + 171 + ... + 361}{8} = 296{,}88$ (од.).

Дисперсію і середнє квадратичне відхилення можна обчислити за формулою (2.2), але в даному випадку простіше використовувати співвідношення $s_t^2 = \overline{y_t^2} - \overline{y}_t^2 = 92\ 478{,}38 - 296{,}882 = 4343{,}61;$

$$s_t = \sqrt{4343{,}61} = 65{,}31 \text{ (од.)},$$

де $\overline{y_t^2} = \frac{\sum_{t=1}^{n} y_t^2}{n} = \frac{213^2 + 171^2 + ... + 361^2}{8} = 92\ 478{,}38.$

Знайдемо коефіцієнт автокореляції r_t часового ряду (для лага τ = 1), тобто коефіцієнт кореляції між послідовностями семи пар спостережень y_t і y_{t+1} (t = 1,2,...,7):

y_t	213	171	291	309	317	362	351
$y_{t+\tau}$	171	291	309	317	362	351	361

Обчислюємо необхідні суми:

$$\sum_{t=1}^{7} y_t = 213+171+...+351 = 2014;$$

$$\sum_{t=1}^{7} y_t^2 = 213^2+171^2+...+351^2 = 609\ 506;$$

$$\sum_{t=1}^{7} y_{t+\tau} = 171+291+...+361 = 2162;$$

$$\sum_{t=1}^{7} y_{t+\tau}^2 = 171^2+291^2+...+361^2 = 694\ 458;$$

$$\sum_{t=1}^{7} y_t y_{t+\tau} = 213 \cdot 171 + 171 \cdot 291 + ... + 351 \cdot 361 = 642\,583.$$

Тепер за формулою (2.4) коефіцієнт автокореляції

$$r_1 = \frac{7 \cdot 642\,583 - 2014 \cdot 2162}{\sqrt{7 \cdot 609\,506 - 2014^2}\sqrt{7 \cdot 694\,458 - 2162^2}} = 0,725.$$

(**Самостійно**: обчислити коефіцієнт автокореляції r_2 часового ряду $y(t)$ для лага τ=2, тобто коефіцієнт кореляції між послідовностями шести пар спостережень y_t і y_{t+2}).►

Знання автокореляційної функції r_t може надати істотну допомогу при підборі моделі часового ряду і статистичній оцінці її параметрів.

2.3. Аналітичне вирівнювання (згладжування) часового ряду (виділення невипадкової компоненти)

Одним із найважливіших завдань дослідження часового ряду є виявлення основної тенденції досліджуваного процесу, яка виражається невипадковою складовою $f(t)$ (трендом). Для вирішення цього завдання спочатку необхідно вибрати вид функції $f(t)$. Найбільш часто використовуються наступні функції:

лінійна: $f(t) = b_0 + b_1 t;$

поліноміальна: $f(t) = b_0 + b_1 t + b_2 t^2 + ... + b_n t^n;$

експоненціальна: $f(t) = e^{b_0 + b_1 t};$

логістична: $f(t) = \frac{a}{1 + be^{-ct}};$

Гомперца: $\log_c f(t) = a - br^t,$ де $0 < r < 1.$

Це дуже відповідальний етап дослідження. При виборі відповідної функції $f(t)$ використовують змістовний аналіз (який може встановити характер динаміки процесу), візуальні спостереження (на основі графічного зображення часового ряду). При виборі поліноміальної функції може бути застосований метод послідовних різниць, який полягає в обчисленні різниць першого порядку $\Delta_t = x_t - x_{t-1}$, другого порядку $\Delta_t^{(2)} = \Delta_t - \Delta_{t-1}$ і так далі, і порядок різниць, при якому вони будуть приблизно однаковими, береться за степінь полінома. З двох функцій перевага віддається тій, при якій менше сума квадратів відхилень фактичних даних від розрахованих на основі цих функцій. При інших рівних умовах перевагу слід віддавати простішим функціям.

Для виявлення основної тенденції найчастіше використовується метод **найменших квадратів.** Значення часового ряду x_t або y_t розглядаються як залежна змінна, а час *t* — як пояснююча:

$$y_t = f(t) + \varepsilon_t, \qquad (2.5)$$

де ε_t — збурення, яке задовольняє основним властивостям регресійного аналізу. За методом найменших квадратів параметри прямої $\tilde{y}_t = b_0 + b_1 t$ знаходяться з системи нормальних рівнянь, в якій в якості x_i беремо *t* , а

$n_i = 1$: $\begin{cases} \sum_{t=1}^{n} y_t = b_0 n + b_1 \sum_{t=1}^{n} t \\ \sum_{t=1}^{n} y_t t = b_0 \sum_{t=1}^{n} t + b_1 \sum_{t=1}^{n} t^2 \end{cases}$, а параметри параболи

$\tilde{y}_t = b_0 + b_1 t + b_2 t^2$ — з системи нормальних рівнянь:

$$\begin{cases} b_0 n + b_1 \sum_{t=1}^{n} t + b_2 \sum_{t=1}^{n} t^2 = \sum_{t=1}^{n} y_t \\ b_0 \sum_{t=1}^{n} t + b_1 \sum_{t=1}^{n} t^2 + b_2 \sum_{t=1}^{n} t^3 = \sum_{t=1}^{n} y_t t \\ b_0 \sum_{t=1}^{n} t^2 + b_1 \sum_{t=1}^{n} t^3 + b_2 \sum_{t=1}^{n} t^4 = \sum_{t=1}^{n} y_t t^2 \end{cases} \cdot$$

Враховуючи, що значення змінної $t = 1,2...,n$ утворюють натуральний ряд чисел від 1 до n, суми $\sum_{t=1}^{n} t$, $\sum_{t=1}^{n} t^2$, $\sum_{t=1}^{n} t^3$, $\sum_{t=1}^{n} t^4$ можна виразити через число членів ряду n по відомим в математиці формулам:

$$\sum_{t=1}^{n} t = \frac{n(n+1)}{2}; \quad \sum_{t=1}^{n} t^2 = \frac{n(n+1)(2n+1)}{6}; \tag{2.6}$$

$$\sum_{t=1}^{n} t^3 = \frac{n^2(n+1)^2}{4}; \sum_{t=1}^{n} t^4 = \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30}. \tag{2.7}$$

◄**Приклад 2.2** За даними табл. 2.1 знайти рівняння невипадкової складової (тренда) для часового ряду y_t, вважаючи тренд лінійним.

Розв'язання. За формулами (2.6):

$$\sum_{t=1}^{8} t = \frac{8 \cdot 9}{1 \cdot 2} = 36; \sum_{t=1}^{8} t^2 = \frac{8 \cdot 9 \cdot 17}{6} = 204;$$

$$\sum_{t=1}^{8} y_t = 213 + 171 + ... + 361 = 2375;$$

$$\sum_{t=1}^{8} y_t^2 = 213^2 + 171^2 + ... + 361^2 = 739827;$$

$$\sum_{t=1}^{n} y_t t = 213 \cdot 1 + 171 \cdot 2 + ... + 361 \cdot 8 = 11766.$$

Система нормальних рівнянь має вигляд:

$$\begin{cases} 8b_0 + 36b_1 = 2375, \\ 36b_0 + 204b_1 = 11766, \end{cases}$$

Звідки $b_0 = 181{,}32;\ b_1 = 25{,}679$ і рівняння тренда $\tilde{y}_t = 181{,}32 + 25{,}679t$ (див. рис. 2.1), тобто попит щорічно збільшується в середньому на 25,7 од. При розв'язанні задачі можна було б не виписувати систему нормальних рівнянь, а представити рівняння регресії у вигляді $\tilde{y}_t - \bar{y} = b_1(t - \bar{t}),$ де $\bar{t} = \frac{\sum_{t=1}^{n} t}{n} = \frac{1+n}{2}, \bar{y} = \frac{\sum_{t=1}^{n} y_t}{n},$ а коефіцієнт регресії b_1 знайти по формулі:

$$b_1 = \frac{\overline{y_t t} - \bar{y}_t \bar{t}}{\overline{t^2} - \bar{t}^2},$$

де $\overline{y_t t} = \sum_{t=1}^{n} \frac{y_t t}{n};\ \bar{y}_t = \frac{\sum_{t=1}^{n} y_t}{n};\ \bar{t} = \frac{(1+n)}{2};\ \overline{t^2} = \frac{(n+1)(2n+1)}{6}.$

Перевіримо значущість отриманого рівняння тренда по F—критерию на 5%-му рівні значущості. Обчислимо суми квадратів:

а) обумовлену регресією:

$$Q_R = \sum_{t=1}^{n} \left(\tilde{y}_t - \bar{y}_t\right)^2 = \sum_{t=1}^{n} b_1^2 (t - \bar{t})^2 =$$

$$= b_1^2 \left(\sum_{t=1}^{n} t^2 - \frac{1}{n} \left(\sum_{t=1}^{n} t \right)^2 \right) = 25{,}679^2 \left(204 - \frac{36^2}{8} \right) = 27695{,}3;$$

б) загальну:

$$Q = \sum_{t=1}^{n} \left(y_t - \bar{y}_t\right)^2 = \sum_{t=1}^{n} y_t^2 - \frac{1}{n} \left(\sum_{t=1}^{n} y_t \right)^2 = 739827 - \frac{2375^2}{8} = 34748{,}9;$$

в) залишкову:

$$Q_e = Q - Q_R = 34748{,}9 - 27695{,}3 = 7053{,}6.$$

Знайдемо значення статистики: $F = \frac{Q_R (n-2)}{Q_e} = 23{,}56.$

Оскільки $F > F_{0,05;1;6} = 5,99$ (див. табл. для F - розподілу), то рівняння тренда значуще.►

Зауваження.

При застосуванні методу найменших квадратів для оцінки параметрів експоненціальною, логістичною функціями або функцією Гомперца виникають складності із розв'язанням системи нормальних рівнянь, тому заздалегідь, до отримання відповідної системи, роблять деякі перетворення цих функцій (наприклад, логарифмують та ін.).

Іншим методом згладжування часового ряду, тобто виділення невипадковій складовій, є **метод ковзних середніх**. Він заснований на переході від початкових значень членів ряду до їх середніх значень на інтервалі часу, довжина якого визначена заздалегідь. При цьому сам вибраний інтервал часу «ковзає» уздовж ряду.

Отриманий таким чином ряд ковзаючих середніх веде себе гладше, ніж початковий ряд, завдяки усереднюванню відхилень ряду. Дійсно, якщо індивідуальний розкид значень членів часового ряду x_t біля свого середнього (згладженого) значення a характеризується дисперсією σ^2, то розкид середніх з m членів часового ряду $\frac{(x_1 + x_2 + ... + x_m)}{m}$ біля того ж значення a характеризуватиметься істотно меншою величиною дисперсії, яка дорівнює σ^2/m. Для усереднювання можуть бути використані середнє арифметичне (просте і з деякими вагами), медіана та ін.

◄**Приклад 2.3** Провести згладжування часового ряду y_t за даними табл. 2.1 методом ковзаючих середніх, використовуючи просте середнє

арифметичне з інтервалом згладжування $m = 3$ роки.

Розв'язання. Ковзаючі середні знаходимо по формулі:

$$\hat{y}_t = \frac{\sum_{i=t-p}^{t+p} y_i}{m}, \qquad (2.8)$$

коли $m = (2p-1)$ — непарне число, то при $m = 3$ і $p = 1$.

Наприклад, при $t = 2$ за формулою (2.8):

$$\hat{y}_2 = \frac{1}{3}(y_1 + y_2 + y_3) = \frac{1}{3}(213 + 171 + 291) = 225 \text{ (од.)};$$

при t=3:

$$\hat{y}_3 = \frac{1}{3}(y_2 + y_3 + y_4) = \frac{1}{3}(171 + 291 + 309) = 257 \text{ (од.)};$$

В результаті отримаємо згладжений ряд:

t	1	2	3	4	5	6	7	8
y_t	—	225,0	257,0	305,7	329,3	343,3	358,0	—

На рис. 2.1 цей ряд зображений графічно у вигляді пунктирної лінії.►

2.4. Часові ряди і прогнозування. Автокореляція збурень

Одним із найважливіших етапів аналізу часового (динамічного) ряду є **прогнозування**. При цьому виходять із того, що тенденція розвитку, яка встановлена у минулому, може бути поширена (екстрапольована) на майбутній період. Завдання ставиться так: є часовий ряд $y_t = (t = 1, 2, ..., n)$ і потрібно дати прогноз рівня цього ряду в момент $n + \tau$. В основному курсі математичної статистики було розглянуто точковий та інтервальний прогнози значень залежної змінної Y, тобто

визначення точкових та інтервальних оцінок Y, отриманих для парної і множинної регресій для значень пояснюючих змінних X, які розташовані поза межами досліджуваного діапазону значень X.

Якщо розглядати часовий ряд як регресійну модель ознаки, яка вивчається, по змінній «час», то до нього можуть бути застосовані розглянуті вище методи аналізу. Одна з основних передумов регресійного аналізу полягає в тому, що збурення $\varepsilon_t (t = 1, 2, ..., n)$ є незалежною випадковою величиною з математичним сподіванням, рівним нулю. А при роботі з часовими рядами таке припущення виявляється у багатьох випадках невірним. Дійсно, якщо вид функції тренда вибраний невдало, то навряд чи можна говорити про те, що відхилення від неї (збурення ε_t) є незалежними. В цьому випадку спостерігається помітна концентрація позитивних і негативних збурень, і можна припускати їх взаємозв'язок. Якщо послідовні значення ε_t корелюють між собою, то кажуть про **автокореляцію збурень.**

Метод найменших квадратів і у випадку автокореляції збурень дає незміщені і грунтовні оцінки параметрів, проте їх інтервальні оцінки можуть містити грубі помилки. У разі виявлення автокореляції збурень доцільно знов повернутися до проблеми специфікації рівняння регресії (вибору функції тренда) переглянути набір включених в нього змінних і тому подібне .

Найбільш простим і достатньо надійним критерієм визначення автокореляції збурень є **критерій Дарбіна— Уотсона**. За допомогою цього критерію перевіряється гіпотеза про відсутності автокореляції між сусідніми залишковими членами ряду e_t і e_{t-1} (для лага τ=1), де e_t вибіркова оцінка ε_t. Статистика критерію має вигляд:

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}. \qquad (2.9)$$

При достатньо великому n можна вважати, що $\sum_{t=1}^{n} e_t^2 \approx \sum_{t=2}^{n} e_t^2 \approx \sum_{t=2}^{n} e_{t-1}^2$.

Тоді, після нескладних перетворень, отримаємо:

$$d \approx \frac{2\sum_{t=2}^{n} e_t^2 - 2\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2} \approx 2\left(1 - \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}\right). \qquad (2.10)$$

Статистика d знаходиться в межах від 0 до 4; за відсутності автокореляції $d \approx 2$ (оскільки при цьому $\sum_{t=2}^{n} e_t e_{t-1} = 0$); при повній додатній автокореляції $d \approx 0 \left(\sum_{t=2}^{n} e_t e_{t-1} \approx \sum_{t=2}^{n} e_t^2\right)$; при повній від'ємній — $d \approx 4$ $\left(\sum_{t=2}^{n} e_t e_{t-1} \approx -\sum_{t=2}^{n} e_t^2\right)$. Для d — статистики знайдені верхня d_B і нижня d_H критичні межі на рівнях значущості α = 0,01; 0,025 і 0,05.

Якщо фактично спостережене значення d:

1) $d_B < d < 4 - d_B$, то гіпотеза про відсутність автокореляції не відкидається (приймається);

2) $d_H \le d \le d_B$ або $4 - d_B \le d < 4 - d_H$, то питання про відкидання гіпотези залишається відкритим (область невизначеності критерію);

3) $0 \le d \le d_H$, то приймається альтернативна гіпотеза про позитивну автокореляцію;

4) $4 - d_H < d < 4$, то приймається альтернативна гіпотеза про негативну автокореляцію.

У табл. 2.2 наведений фрагмент таблиці значень статистик d_H і d_B критерію Дарбіна—Уотсона на рівні значущості α = 0,05.

Таблиця 2.2. Статистики Дарбіна - Уотсона

Кількість спостережень n	Кількість пояснюючих змінних					
	p=1		p=2		p=3	
	d_H	d_B	d_H	d_B	d_H	d_B
15	1,08	1,36	0,95	1,54	0,82	1,75
16	1,10	1,37	0,98	1,54	0,86	1,73
17	1,13	1,38	1,02	1,54	0,90	1,71
18	1,16	1,39	1,05	1,53	0,93	1,69
19	1,18	1,40	1,08	1,53	0,97	1,68
20	1,20	1,41	1,10	1,54	1,00	1,68
25	1,29	1,45	1,21	1,55	1,12	1,66
30	1,35	1,49	1,28	1,57	1,21	1,65
35	1,40	1,52	1,34	1,58	1,28	1,65
40	1,44	1,54	1,39	1,60	1,34	1,66
45	1,48	1,57	1,43	1,62	1,28	1,67
50	1,50	1,59	1,46	1,63	1,42	1,67

Недоліками критерію Дарбіна—Уотсона є наявність області невизначеності критерію, а також те, що критичні значення d — статистики визначені для об'ємів вибірки не менше 15.

◄**Приклад 2.4** Виявити на рівні значущості 0,05 наявність автокореляції збурень для часового ряду y_t по даним табл. 2.1.

Розв'язання. У прикладі 2.2 отримано рівняння тренда $\tilde{y}_t = 181{,}32 + 25{,}679t$ (од.). У табл. 2.3 наведений розрахунок сум необхідних для обчислення d — статистики.

Таблиця 2.3. Данні до задачі 2.4

t	y_t	$\tilde{y}_t$	$e_t = y_t - \tilde{y}_t$	e_{t-1}	$e_t e_{t-1}$	e_t^2
1	213	207,0	6,0	—	—	36,0
2	171	232,7	-61,7	6,0	-370,2	3806,9
3	291	258,4	32,6	-61,7	-2011,4	1062,8
4	309	284,0	25,0	32,6	815,0	625,0
5	317	309,7	7,3	25,0	182,5	53,3
6	362	335,4	26,6	7,3	194,2	707,6
7	351	361,1	-10,1	26,6	-268,7	102,0
8	361	386,8	-25,8	-10,1	260,6	665,6
$\sum_{t=1}^{8}$	—	—	—	—	-1198,0	7059,2

Тепер за формулою (2.10) статистика $d \approx 2(1+1198,0/7059,2)=2,34$. За табл. 2.2 при p=1, n=15 критичні значення d_H=1,08; d_B=1,36, тобто фактично знайдене d=2,34 знаходиться в межах від d_B до 4-d_B (1,36< d <2,64). Критичні значення d — статистики у табл. 2.2 відсутні, але можна допустити, що знайдене значення залишиться в інтервалі (d_B;4-d_B), тобто для даного часового ряду попиту на рівні значущості 0,05 гіпотеза про відсутність автокореляції збурень не відкидається (приймається).►

У разі відсутності значущої автокореляції збурень, методами регресійного аналізу може бути знайдена не тільки точкова, але й інтервальна оцінка рівнів ряду, тобто здійснений їх точковий та інтервальний прогнози.

◄**Приклад 2.5.** За даними табл. 2.1 дати точкову і, з надійністю 0,95, інтервальну оцінки прогнозу середнього та індивідуального значень попиту на деякий товар на момент t = 9 (дев'ятий рік).

Розв'язання. В прикладі 2.2, отримано рівняння регресії $\tilde{y}_t$= 181,32+25,679t, тобто щорічно попит на товар збільшувався в середньому на 25,7 од. Треба оцінити умовне математичне сподівання $M_{t=9}(Y) = \bar{y}(9)$. Оцінкою $\bar{y}(9)$ є групова середня

$$\tilde{y}_{t=9} = 181{,}32+25{,}679\cdot 9=412{,}4 \text{ од.}$$

Знайдемо оцінку s^2 дисперсії σ^2 (див. табл. 2.3):

$$s^2 = \frac{\sum_{t=1}^{8} e_t^2}{n-2} = \frac{7059{,}2}{8-2} = 1176{,}5.$$

Обчислимо оцінку дисперсії групової середньої :

$$s^2_{y_{t=9}} = 1176{,}5\left(\frac{1}{8} + \frac{(9-4{,}5)^2}{42}\right) = 714{,}3;$$

$$s_{y_{t=9}} = \sqrt{714{,}3} = 26{,}73 \text{ (од.)}$$

(використані дані, отримані у прикладі 2.2). $t_{0{,}95;6}$ =2,45. Інтервальна оцінка прогнозу середнього значення попиту:

$$412{,}4 - 2{,}45\cdot 26{,}73 \le \bar{y}(9) \le 412{,}4 + 2{,}45\cdot 26{,}73$$

або $346{,}9 \le \bar{y}(9) \le 477{,}9$ (од.).

Для знаходження інтервальної оцінки прогнозу індивідуального значення $y^*(9)$ обчислимо дисперсію його оцінки:

$$s^2_{y_{t_0=9}} = 1176{,}5\left(1+\frac{1}{8} + \frac{(9-4{,}5)^2}{42}\right) = 1890{,}8;\ s_{y_{t_0=9}} = 43{,}48 \text{ (од.)},$$

а потім — саму інтервальну оцінку для $y^*(9)$:

$$412{,}4 - 2{,}45\text{-}43{,}48 \le y^*(9) \le 412{,}4 + 2{,}45\text{-}43{,}48$$

або $305{,}9 \le y^*(9) \le 518{,}9$ (од.).

Отже, з надійністю 0,95 середнє значення попиту на товар на дев'ятий рік знаходитиметься в межах від 346,9 до 477,9 (од.), а його індивідуальне значення — від 305,9 до 518,9 (од.). ►

Прогноз розвитку процесу, який вивчається, на основі екстраполяції часових рядів може виявитися ефективним, як правило, в рамках короткострокового, в крайньому випадку середньострокового періоду прогнозування. Якщо в даній регресійній моделі автокореляція збурень існує, то необхідні заходи по її усуненню (або зниженню).

Методи усунення або зниження автокореляції збурень

1. Метод послідовних різниць полягає в переході від рівнів рядів y_t, x_t до їх перших різниць $\Delta y_t = y_{t+1} - y_t$, $\Delta x_t = x_{t+1} - x_t$ $(t = 1,2,...,n)$ і розгляду рівняння регресії $\Delta y_t = b_0 + b_1 \Delta x_t$, в якому коефіцієнт b_1 інтерпретується як середній приріст змінної y_t при зміні приросту x_t на одну одиницю. Метод ефективний, коли невипадкова складова тимчасового ряду представляє пряму лінію.

2. Метод включення в модель регресії часу *t* в якості додаткової пояснюючою змінною: $y_t = b_0 + b_1 x_t + b_2 t$. Метод виправданий, якщо він не приводить до мультиколінеарності.

◄**Приклад 2.6** За даними табл. 2.1, яка відображає динаміку цін x_t і попиту y_t деякого товару за восьмирічний період, з'ясувати на рівні значущості 0,05, чи впливає ціна на попит.

Розв'язання. Якщо абстрагуватися від того, що змінні x_t і y_t є часовими рядами, то в припущенні існування лінійної регресії можна отримати рівняння регресії у вигляді: $\tilde{y}_t$= 635,2 — 0,8843x_t. По F - критерию рівняння регресії значуще на рівні 0,05, оскільки обчислене значення статистики $F = 9,00 > F_{0,05;1;6} = 5,99.$ Проте, такий висновок був би правомірний принаймні за відсутності автокореляції збурень ε_t часового ряду залежної змінної y_t. Використання критерію Дарбіна — Уотсона показує наявність істотної автокореляції залишкового часового ряду $e_t = y_t - \tilde{y}_t$. Таким чином, для обгрунтованої відповіді на питання завдання необхідно виключити автокореляцію збурень.

Перший спосіб. Перейдемо від рівнів ряду y_t і x_t до їх перших різниць $\Delta y_t = y_{t+1} - y_t$, $\Delta x_t = x_{t+1} - x_t$. Значення змінних Δy_t, Δx_t представимо в табл. 2.4.

Таблиця 2.4. Данні до задачі 2.6

Δx_t	-30	-112	-33	23	11	17	13
Δy_t	-42	120	18	8	45	-11	10

Рівняння регресії: $\Delta\tilde{y}_t = 9,94 - 0,7064\Delta x_t$. Перевірка рівняння по F — критерию на рівні 0,05 показує, що воно незначуще, оскільки F = 3,96< $F_{0,05;1;5}$ = 6,61. Отже, немає підстав вважати, що ціна на даний товар має істотний вплив на попит.

Другий спосіб. Включимо в модель регресії час *t* в якості додаткової пояснюючої змінної. Методом найменших квадратів отримаємо рівняння y_t = 380,4 - 0,4443 x_t + 19,22t, причому коефіцієнт при змінній *t* виявився значущим по *t*-критерію (t_{b_2} = 3,82 > $t_{0,95;5}$ = 2,57), а при змінній x_t — незначущим (t_{b_1} = 2,22 < $t_{0,95;5}$ = 2,57), Отже, підтверджується висновок про відсутність істотного впливу ціни x_t на попит y_t .►

3. Авторегресійна модель. Одним із методів усунення автокореляції є використання авторегресійної моделі. Для даного часового ряду далеко не завжди вдається підібрати *адекватну* модель, для якої ряд збурень ε_t задовольнятиме основним передумовам регресійного аналізу, зокрема, не буде мати автокореляції. Поширення набули регресійні моделі, в яких регресорами виступають **лагові змінні**, тобто змінні, вплив яких в регресійній моделі характеризується деяким запізненням. Ще одна відмінність: представлені в моделях пояснюючі змінні є величинами випадковими.

Авторегресійна модель *p*-го порядку має вигляд:

$$x_t = b_0 + b_1 x_{t-1} + b_2 x_{t-2} + ... + b_p x_{t-p} + \varepsilon_t (t = 1, 2, ..., n), \qquad (2.11)$$

де $b_0, b_1, ..., b_p$ —деякі константи.

Вона описує процес, який вивчається, в момент *t* в залежності від його значень в попередні моменти $(t-1), (t-2), ..., (t-p)$. Якщо досліджуваний процес x_t в момент *t* визначається лише його значеннями в попередній період $(t-1)$, то розглядають авторегресійну модель 1-го порядку (марківський випадковий процес):

$$x_t = b_0 + b_1 x_{t-1} + \varepsilon_t (t = 1, 2, ..., n). \qquad (2.12)$$

◄**Приклад 2.7** В таблиці 2.5 представлені дані, які відображають динаміку курсу акцій деякої компанії (грош. од.). Використовуючи авторегресійну модель 1-го порядку, дати точковий та інтервальний прогнози середнього та індивідуального значень курсу акцій в момент t= 23, тобто на глибину один інтервал.

Таблиця 2.5. Данні до задачі 2.7

t	1	2	3	4	5	6	7	8	9	10	11
y_t	971	1166	1044	907	957	727	752	1019	972	815	823
t	12	13	14	15	16	17	18	19	20	21	22
y_t	1112	1386	1428	1364	1241	1145	1351	1325	1226	1189	1213

Розв'язання. Спроба підібрати до даного часового ряду адекватну модель виявляється даремною. Відповідно до умови застосуємо авторегресійну модель вигляду (2.12). Отримаємо: $\tilde{y}_t = 284,0 + 0,7503 y_{t-1}$. Знайдене рівняння регресії значуще на 5%-вому рівні по F - критерию, оскільки фактично спостережене значення статистики F = 24,32 > $F_{0,05;1;19}$ = 4,35. Застосування критерію Дарбіна—Уотсона свідчить про незначущу автокореляцію збурень $e_t = y_t - \tilde{y}_t$. Обчислення дають точковий прогноз по рівнянню $\tilde{y}_t = 284,0 + 0,7503 y_{t-1}$.: $\tilde{y}_{t=23}$=284,0+0,7503·1213=1194,1 та інтервальний на рівні значущості 0,05 для середнього та індивідуального значень — $1046,6 \le \overline{y}_{t=23} \le 1341,6$; $879,1 \le y^*_{t_0=23} \le 1509,1$. Отже, з надійністю 0,95 середнє значення курсу акцій компанії на момент t = 23 буде знаходитись в межах від 1046,6 до 1341,6 (грош. од.), а його індивідуальне значення — від 879,1 до 1509,1 (грош. од.).►

Контрольні питання

1. В чому полягає відмінність часового ряду від вибіркового?
2. Основні характеристики стаціонарних часових рядів.
3. Для оцінки яких параметрів застосовується кореляційний аналіз?
4. Застосування авторегресійних моделей.

Вправи для самостійної роботи

1. В таблиці наведені данні, які відображають динаміку зростання середніх витрат на електронні гаджети (грн) за вісім останніх років.

t	1	2	3	4	5	6	7	8
y_t	11330	12220	13540	13890	13420	13770	14910	16840

В припущенні, що тренд є лінійним:

1) Знайти рівняння тренду і оцінити його значущисть на рівні 0,05.
2) За допомогою критерія Дарбіна-Уотсона з'ясувати, чи є залишковий ряд e_t автокорельованим на 5% - му рівні значущості.
3) За умови відсутності автокореляції збурень, знайти точковий і з надійністю 0,95 інтервальний прогноз середнього і індивідуального значень зростання витрат на гаджети на дев'ятий рік.

ТЕМА 3. ЛІНІЙНІ РЕГРЕСІЙНІ МОДЕЛІ ФІНАНСОВОГО РИНКУ

3.1. Загальні відомості

Кожен цінний папір — акція, облігація, контракт та інші — в кожен момент часу має вартість, яка називається **курсом** і встановлюється ринком (зазвичай як результат біржових котирувань). Навіть маючи усю повноту інформації про емітента, який випустив папір, однозначно визначити його курс в деякий момент часу в майбутньому, як правило, неможливо. В цьому випадку найприродніше розглядати курс цінного паперу, як значення випадкової величини *X*.

Нехай X_t — курс цінного паперу в момент часу *t*, а X_{t+1} — в момент часу $t+1$ (зазвичай одиниця часу — це проміжок між котируваннями). Звернемося до випадку, коли момент часу *t* вже настав, а моменту $t+1$ ще ні. Розглянемо величину

$$r_t = \frac{X_{t+1} - X_t}{X_t}.$$

Оскільки X_{t+1} — випадкова величина, то і r_t — теж випадкова величина. Вона називається **прибутковістю цінного паперу**. Очевидно, що значення саме цієї величини визначає привабливість цінного паперу для інвестора. І одне з головних завдань фінансового аналізу полягає в можливо точнішому передбаченні значення величини r_t.

3.1. Регресійні моделі

Моделі, які розглядаються у фінансовому аналізі, зв'язують випадкову величину r_t з величинами, які об'єктивно характеризують фінансовий ринок у цілому. Такі величини називаються **чинниками**. Залежно від постановки завдання чинники можуть вважатися як випадковими, так і детермінованими, тобто точно відомими величинами. У найпростішому випадку виділяється один чинник. Тоді статистична модель має вигляд:

$$r_t = \alpha + \beta F + \varepsilon. \qquad (3.1)$$

α і β – постійні (невідомі параметри), ε– випадкова величина, яка задовольняє умові: $M_F(\varepsilon) = 0$, де $M_F(\varepsilon)$ – умовне математичне сподівання випадкової величини ε відносно F. З цього припущення виходить, що і безумовне математичне сподівання величини ε також дорівнює нулю. Насправді: $M(\varepsilon) = M\left(M_F(\varepsilon)\right) = 0.$

Звідси слідує, що якщо чинник *F* розглядається як випадкова величина, то його коваріація з ε дорівнює нулю. Дійсно, використовуючи властивості умовного математичного сподівання, отримуємо:

$$cov(F,\varepsilon) = M(F\varepsilon) - M(F)\cdot M(\varepsilon) = M(F\varepsilon) =$$
$$= M\left(M_F(F\varepsilon)\right) = M\left(FM_F(\varepsilon)\right) = 0$$

Значення коефіцієнтів α і β неважко визначити через числові характеристики r_t і F:

$cov(r_t,F) = \beta\, cov(F,F) + cov(\varepsilon,F)$, або $cov(r_t,F) = \beta\, cov(F,F).$

Звідки $\beta = \dfrac{cov(r_t,F)}{cov(F,F)} = \dfrac{cov(r_t,F)}{D(F)}$. Перейшовши в рівнянні моделі (3.1)

до математичних сподівань, отримаємо:

$$M(r_t) = \alpha + \beta M(F) + M(\varepsilon).$$

Але $M(\varepsilon) = 0$, тому $\alpha = M(r_t) - \beta M(F) = M(r_t) - \dfrac{cov(r_t, F)}{D(F)} M(F)$

Коефіцієнт β називається **чутливістю прибутковості** цінного паперу до чинника F. Коефіцієнт α називається **зсувом**. У класичному регресійному аналізі значення чинників F вважаються детермінованими величинами, тобто модель (3.1) має вигляд: $r_t = \alpha + \beta F_t + \varepsilon_t$. $t = 1,2,..,n$ – моменти часу, які інтерпретуються як номер спостережень; $F_1,..,F_n$ – відомі значення чинників; r_t – спостережувані вибіркові значення випадкової величини r; α і β – невідомі параметри. Їх оцінки можна побудувати методом найменших квадратів:

$$\hat{\beta} = \frac{\hat{\text{cov}}(r, F)}{\hat{D}(F)} = \frac{n\sum_{t=1}^{n} F_t r_t - \left(\sum_{t=1}^{n} F_t\right)\left(\sum_{t=1}^{n} r_t\right)}{n\sum_{t=1}^{n} F_t^2 - \left(\sum_{t=1}^{n} F_t\right)},$$

$$\hat{\alpha} = \hat{M}(r) - \frac{\hat{\text{cov}}(r, F)}{\hat{D}(F)} \hat{M}(F) = \frac{1}{n}\sum_{t=1}^{n} r_t - \frac{1}{n}\left(\sum_{t=1}^{n} F_t\right)\hat{\beta}.$$

Різні моделі фінансового ринку розглядають різні величини в якості чинника *F*. Розглянемо далі основні з цих моделей.

1. Ринкова модель.

Одна з найпоширеніших моделей використовує як чинник F прибутковість ринкового індексу r_M. **Ринковим індексом** називається зважена сума курсів акцій найбільш значних емітентів фінансового ринку. Наприклад, в США найбільш поширені наступні індекси:

1) *DJ* (**індекс Доу-Джонса**) −розраховується за 30 найбільш значущими корпораціям, таким як Microsoft, Coca Cola, General Motors і так далі.
2) **Індекс S&P 500** (*Standard and Poor's*) − розраховується по 500 найбільш великим компаніям.
3) **Зведений індекс** *NYSE* − для його розрахунку використовуються курси акцій, зареєстрованих на Нью-Йоркській фондовій біржі.

Очевидно, ринковий індекс певною мірою відображує стан економіки в цілому. Отже ринкова модель показує, наскільки прибутковість цінного паперу відповідає економічній динаміці країни (або навіть співтовариства країн). Випадкова величина ε характеризує залежність прибутковості цінного паперу від обставин, специфічних саме для її емітента.

Прибутковість ринкового індексу по суті є усередненою прибутковістю різних цінних паперів. Якщо розглядати безліч усіх цінних паперів, що фігурують на ринку, то коефіцієнт β навмання вибраного цінного паперу є значенням випадкової величини (позначимо її тією ж буквою β). Якщо цінний папір включений в індекс, то за самим визначенням індексу має місце рівність: $M(\beta)=1$. Якщо конкретне спостережуване значення цінного паперу більше одиниці, то його прибутковість росте в середньому швидше, ніж ринок в цілому. Такі папери називаються "агресивними"; папери з коефіцієнтом β, меншим одиниці, називаються "оборонними".

◄ **Приклад 3.1.** Значення прибутковості акцій *Widget Manufacturing* і прибутковості індексу дані в таблиці 3.1. Оцінити залежність прибутковості акцій від прибутковості індексу.

Розв'язання. Метод найменших квадратів дає наступні результати:

$\beta=0,63$; $\alpha=0,79$. Стандартні помилки $\sigma_\beta=0,28$; $\sigma_\alpha=2,03$ коефіцієнт детермінації $R^2=0,27$. Отримані характеристики свідчать про те, що прибутковість акцій Widget Manufacturing істотно залежить від ринкового

індексу. Як видно, акції Widget Manufacturing є "оборонними".►

На фінансовому ринку присутні і папери з нульовим коефіцієнтом β. В цьому випадку має місце співвідношення $r_M = r_F$, звідки випливає, що очікувана прибутковість цінного паперу фіксована і не залежить від стану ринку в цілому. Така ситуація характерна, наприклад, для облігацій.

2. Модель дотичного портфеля

Іншим чинником, який часто використовується в лінійних регресійних моделях, є прибутковість деякого виділеного портфеля цінних паперів, який називається *дотичним*. Це поняття було введене Г. Марковичем в 1952 р. Опишемо це поняття.

Таблиця 3.1. Приклад моделі дотичного портфеля

Рік	Квартал	Прибутковість WM	Прибутковість індексу
1	1	-13,38	2,52
	2	16,79	5,45
	3	-1,67	0,76
	4	-3,46	2,36
2	5	10,22	8,56
	6	7,13	8,67
	7	6,71	10,8
	8	7,84	3,33
3	9	2,15	-5,07
	10	7,95	7,1
	11	-8,05	-11,57
	12	7,68	4,65
4	13	4,75	14,59
	14	7,55	2,66
	15	-2,36	3,81
	16	4,98	7,99

Нехай на фінансовому ринку обертається *n* цінних паперів і капітал, рівний одиниці, інвестується в ці папери так, що x_i − капітал, який інвестується в *i*-й папір. Набір чисел $p = x_1, x_2, ..., x_n$, який задовольняє умові $x_1 + x_2 + ... + x_n = 1$, назвемо **портфелем цінних паперів**. Зрозуміло, деякі числа x_i можуть бути нульовими. (Насправді, деякі x_i можуть бути і від'ємними: відповідна ситуація називається **продажем цінного паперу без покриття**).

Кожному портфелю *p* відповідає випадкова величина r_p − прибутковість, яка визначається аналогічно прибутковості одного цінного паперу. Очевидно, $r_p = \sum_{i=1}^{n} x_i r_i$. Розглянемо координатну площину, на якій по осі ординат відкладається математичне сподівання прибутковості (очікувана прибутковість $\bar{r}_p$), а по осі абсцис — стандартне відхилення прибутковості $\sigma_p = \sqrt{D(r_p)}$. Величина σ_p називається **ризиком портфеля**. Тоді кожному портфелю може бути поставлена у відповідність точка на такій координатній площині, а вся множина допустимих портфелів відображається в деяку двовимірну фігуру, яка називається **допустимою множиною** (рис. 3.1). Між множиною всіх портфелів і допустимою множиною немає взаємно−однозначної відповідності. Звичайно, два різні портфелі можуть мати рівні значення $\bar{r}_p$ та σ_p. Природно припустити, що інвестор вважає за краще отримати велику прибутковість із найменшим ризиком, тобто з двох портфелів з однаковим значенням $\bar{r}_p$ він вибере той, значення σ_p якого менше. Це означає, що найбільш прийнятному портфелю відповідає точка на границі *AB* (див. рис. 3.1). Лінія *AB* називається **ефективною множиною**.

Проблема вибору точки ефективної множини вирішується кожним інвестором індивідуально і, здавалося б, залежить від його схильності до ризику (чи, навпаки — до уникнення ризику). Але виявляється, що ефективній множині належить точка, яка є виділеною для всіх інвесторів. Припустимо, що окрім придбання цінних паперів, інвестор має можливість безризикового надання і отримання позик. Таке припущення

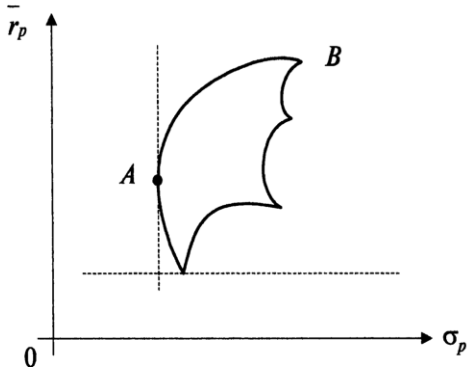

Рис. 3.1. Множина всіх портфелів

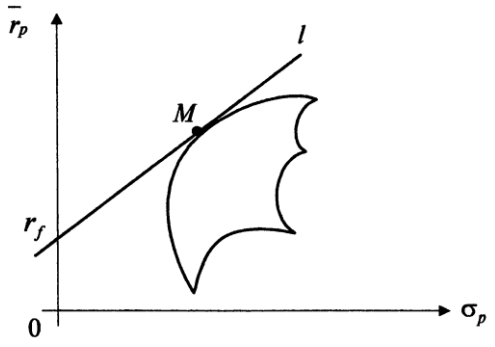

Рис. 3.2. Безризикова ставка

цілком відповідає дійсності, якщо інвестор має можливість купувати державні облігації і брати кредит. Зробимо ще одне припущення (вже зовсім не таке беззаперечне), що безризикове надання і отримання позик відбувається з однією і тією ж процентною ставкою r_F, яка називається **безризиковою ставкою**. Розглянемо пряму *l*, яка перетинає вісь ординат в точці r_F і дотичну до ефективної множини (рис. 3.2). Рівняння прямої *l* має вигляд $\bar{r} = r_F + \frac{\sigma}{\sigma_M}\left(\bar{r}_M - r_F\right)$. Розглянемо портфель $p = \left(x_F, x_M\right)$, де x_F — безризикові вкладання (позитивні у разі придбання облігацій і від'ємні при позиці засобів) з фіксованою ставкою r_F, x_M — вкладання в

портфель, яке відповідає точці *M*. Тоді

$$\bar{r}_p = x_f r_f + x_M \bar{r}_M = (1 - x_M) r_f + x_M \bar{r}_M =$$

$$= r_f + x_M (\bar{r}_M - r_f); \quad \sigma_p = x_M \sigma_M, \quad \text{звідки} \quad \bar{r} = r_f + \frac{\sigma}{\sigma_M} (\bar{r}_м - r_f), \text{ тобто}$$

точка (σ_p, $\bar{r}_p$) належить прямій *l*. Очевидно також, що будь-яка точка напівпрямої *l*, яка лежить в першій чверті, досяжна за допомогою комбінації (x_F, x_M). Таким чином, за наявності можливості безризикового надання і отримання позик, допустима множина розширюється, а ефективною множиною стає пряма *l*.

Портфель, який відповідає точці дотику *M* (рис. 3.2), називається **дотичним портфелем**. Отже, оптимальною для будь-якого інвестора стратегією є інвестування частини коштів в дотичний портфель, а частини — у безризикові облігації. Або навпаки: отримання позики для додаткового інвестування в дотичний портфель.

На практиці точне знаходження дотичного портфеля неможливе. Але для багатьох практичних цілей виявляється корисною модель, у якій в якості чинника вибрана прибутковість дотичного портфеля, а точніше — різниця між r_M і безризиковою ставкою r_F. Таким чином, $F = r_M - r_F$ і модель набуває вигляду: $r_i = \alpha_i + \beta_i (r_M - r_F) + \varepsilon_i$, де *i* — номер цінного паперу.

3. Неврівноважені і врівноважені моделі

Прибутковість цінних паперів, які є на ринку, можна розглядати залежно від часу. При цьому будуть залежати від часу числові характеристики випадкової величини r_p. Так само залежитимуть від часу і значення параметрів α і β.

Модель фінансового ринку називається **врівноваженою**, якщо числові характеристики випадкових величин, які входять до неї, постійні в часі. Економічний сенс подібного припущення очевидний: ринок вважається "сталим", збалансованим. В цьому випадку можна отримати деякі конкретні результати, які істотно спрощують ситуацію.

Розглянемо модель залежності прибутковості цінного паперу від прибутковості дотичного портфеля (передбачається, що безризикова ставка отримання і надання позик для усіх учасників ринку одна і та сама і дорівнює r_F). Якщо модель врівноважена, тобто ринок збалансований, то дотичний портфель задовольняє наступній властивості: **доля кожного цінного паперу в ньому відповідає її відносній ринковій вартості**. Такий портфель називається **ринковим** і визначається однозначно. Таким чином, розглядаючи врівноважені моделі, ми ототожнюємо поняття дотичного і ринкового портфелів, прибутковість якого r_M.

В цьому випадку регресійна модель для *i* - го цінного паперу має вигляд:

$$r_i = \alpha_i + \beta_{iM}\left(r_M - r_F\right) + \varepsilon_i .$$

В разі врівноваженому випадку має місце наступна теорема.

Теорема. Для усіх цінних паперів, які обертаються на ринку, коефіцієнт α_i, один і той самий і дорівнює безризиковій ставці.

Маємо $\bar{r}_i = \alpha_i + \beta_{iM}\left(\bar{r}_M - r_F\right)$. Розглянемо портфель *p*, який складається з *i*-го цінного паперу і ринкового портфеля *M* в пропорції x_i та $\left(1-x_i\right)$ відповідно. Очікувана прибутковість такого портфеля складе

$$\bar{r}_p = x_i \bar{r}_i + (1-x_i)\bar{r}_M \qquad (3.2)$$

а стандартне відхилення буде

$$\sigma_p = \sqrt{x_i^2\sigma_i^2 + (1-x_i)^2\sigma_M^2 + 2x_i(1-x_i)\sigma_{iM}} . \qquad (3.3)$$

Всі такі портфелі відображаються на криву, яка сполучає точки *i* та *M* (рис. 3.3).

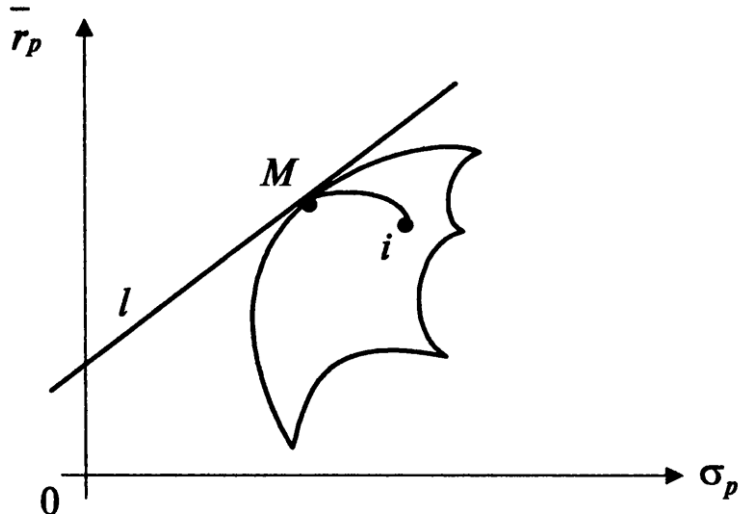

Рис. 3.3. Крива ризиків

З рівності (3.2) отримуємо: $\frac{d\bar{r}_p}{dx_i} = \bar{r}_i - \bar{r}_M$. А з рівності (3.3):

$$\frac{d\sigma_p}{dx_i} = \frac{x_i\sigma_i^2 - \sigma_M^2 + x_i\sigma_M^2 + \sigma_{iM} - 2x_i\sigma_{iM}}{\sqrt{x_i^2\sigma_i^2 + (1-x_i)^2\sigma_M^2 + 2x_i(1-x_i)\sigma_{iM}}},$$

звідки

$$\frac{d\bar{r}_p}{d\sigma_p} = \frac{(\bar{r}_i - \bar{r}_M)\sqrt{x_i^2\sigma_i^2 + (1-x_i)^2\sigma_M^2 + 2x_i(1-x_i)\sigma_{iM}}}{x_i\sigma_i^2 - \sigma_M^2 + \sigma_{iM} - 2x_i\sigma_{iM}}.$$

В точці M $x_i = 0$, звідси нахил кривої в точці M :

$$\frac{d\bar{r}_p}{d\sigma_p} = \frac{(\bar{r}_i - \bar{r}_M)\sigma_M}{\sigma_{iM} - \sigma_M^2}. \qquad (3.4)$$

Але, крива дотикається прямої *l*, тому

$$\frac{d\overline{r}_p}{d\sigma_p} = \frac{\overline{r}_M - r_F}{\sigma_M} \ . \qquad (3.5)$$

Прирівнявши праві частини рівностей (3.4) та (3.5), отримаємо

$$\overline{r}_i = r_F + \left(\frac{\overline{r}_M - r_F}{\sigma_M^2} \right) \sigma_{iM} \ .$$

Таким чином, єдиним параметром, який характеризує цей цінний папір, є його чутливість "бета" до ринкового портфеля.

4. Модель оцінки фінансових активів (CAPM)

Рівняння $\overline{r}_i = r_F + \beta_{iM} \left(\overline{r}_M - r_F \right)$ називається **ринковою лінією** цінного паперу. Воно визначає залежність очікуваної прибутковості цінного паперу від її чутливості "бета" $\beta_{iM} = \frac{\sigma_{iM}}{\sigma_M^2}$.

Розглянемо портфель $p = x_1, x_2, ..., x_n, \ \sum_{i=1}^{n} x_i = 1$. Прибутковість портфеля $r_p = \sum_{i=1}^{n} x_i r_i$. Звідси маємо, що очікувана прибутковість

$$\overline{r}_p = \sum_{i=1}^{n} x_i \overline{r}_i = \sum_{i=1}^{n} x_i r_F + \left(\overline{r}_M - r_F \right) \sum_{i=1}^{n} x_i \beta_{iM} = r_F + \left(\overline{r}_M - r_p \right) \beta_{iM} \ .$$

Тут $\beta_{pM} = \frac{\mathrm{cov}(r_p, r_M)}{\sigma_M^2} = \frac{\sum_{i=1}^{n} x_i \, \mathrm{cov}(r_i, r_M)}{\sigma_M^2}$. Рівняння $\overline{r}_p = r_F + \left(\overline{r}_M - r_p \right) \beta_{iM}$ називається **рівнянням моделі оцінки фінансових активів**. Для її використання необхідно отримати оцінки параметрів дотичного портфеля — очікуваної прибутковості та ризику, а також коваріацій прибутковостей цінних паперів, які входять в *p*, з прибутковістю ринкового портфеля.

Практичне значення моделі оцінки фінансових активів полягає в тому, що вона може використовуватись для виявлення невірно оцінених паперів в неврівноваженій ситуації. Так, якщо прибутковість цінного паперу вища за ту, яка задається рівнянням $\bar{r}_p = r_F + \left(\bar{r}_M - r_p\right)\beta_{iM}$, то папір є **переоціненим**, в протилежному випадку – **недооціненим**.

3.2. Зв'язок між очікуваною прибутковістю і ризиком оптимального портфеля

Нехай для усіх учасників ринку є єдина безризикова ставка r_F отримання і надання позик. Тоді **оптимальним** буде портфель, який є комбінацією дотичного і безризикового. Оптимальний портфель має вигляд: $p_{опт} = x_0, x_1, x_2, ..., x_n$. Тут x_0 — частина коштів, вкладених у безризикові папери (казначейські векселі), якщо $x_0 > 0$, або x_0 — позикові кошти, використані для вкладення в дотичний портфель, якщо $x_0 < 0$. Числа $x_1, x_2, ..., x_n$ знаходяться в тому ж співвідношенні, в якому знаходяться компоненти дотичного портфеля. Очевидно, що прибутковості оптимального і дотичного портфелів пов'язані лінійним співвідношенням:

$$r_p = x_0 r_F + \lambda r_M .$$

Звідси слідує, що коефіцієнт кореляції $\rho(r_p, r_M) = 1$, тобто $\dfrac{\mathrm{cov}(r_p, r_M)}{\sigma_p \sigma_M} = 1$ та $\beta_{pM} = \dfrac{\sigma_p}{\sigma_M}$. Таким чином, рівняння CAPM має вигляд:

$$\bar{r}_p = r_F + \frac{\left(\bar{r}_M - r_p\right)}{\sigma_M}\sigma_p \qquad (3.6)$$

Воно задає залежність очікуваної прибутковості оптимального портфеля від його ризику. Розглянемо геометричний зміст рівняння (3.6). При зроблених припущеннях про безризикову ставку r_F — ефективна множина є прямою *l* (рис. 3.4).

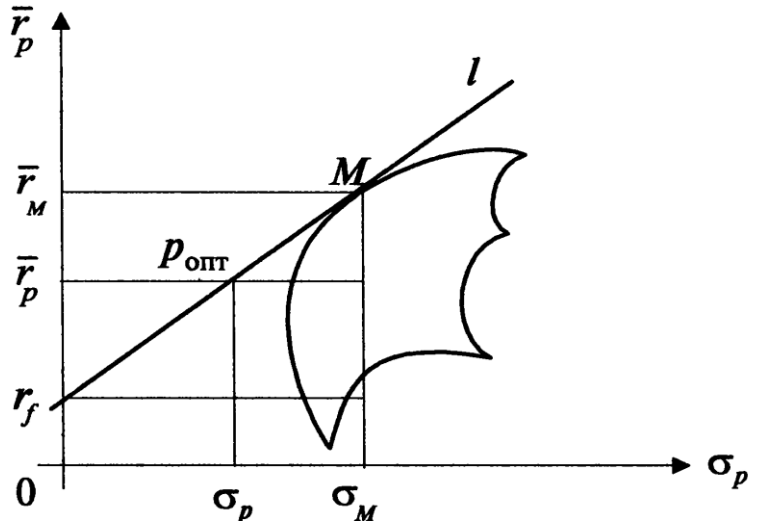

Рис. 3.4. Геометрична ілюстрація оптимальної множини

З подібності трикутників виходить: $\frac{\bar{r}_p - r_F}{\sigma_p} = \frac{\bar{r}_M - r_F}{\sigma_M}$, а це еквівалентно рівності (3.6).

3.3. Багатофакторні моделі

Однофакторні моделі у багатьох випадках є цілком адекватними, проте найчастіше вони виявляються занадто спрощеними, і тоді доводиться розглядати залежність прибутковості цінного паперу від декількох (*m*) чинників, тобто лінійні регресійні моделі виду : $r_i = \alpha + \sum_{k=1}^{m} \beta_k F_i^{(k)} + \varepsilon_i$.

α, β_k — параметри, $F_i^{(k)}$ — чинники, які визначають стан ринку (i - номер спостереження). Такими чинниками можуть бути, наприклад, рівень інфляції, темпи приросту валового внутрішнього продукту та ін. Якщо цей цінний папір відноситься до деякого сектора економіки, то безумовно слід розглядати чинники, специфічні для цього сектора.

◄ **Приклад 3.2**. У таблиці 3.2 приведені дані за шість років про темп росту, рівень інфляції та прибутковості акцій компаній *Widget Manufacturing*. Розглядається факторна модель вигляду : $Widget = \alpha + \beta_1 temp + \beta_2 inf$. За допомогою методу найменших квадратів виходить рівняння вгляду : $\overline{Widget} = 5{,}77 + 2{,}17 temp - 0{,}67 inf$

При цьому середні квадратичні помилки оцінок параметрів β_1 і β_2 дорівнюють 1,125 і 1,162.

Таблиця 3.2. Данні до прикладу 3.2

Рік	Темп росту ВВП, % (*temp*)	Рівень інфляції, % (*inf*)	Прибутковість акцій, % (*Widget*)
1	5,7	1,1	14,3
2	6,4	4,4	19,2
3	7,9	4,4	23,4
4	7,0	4,6	15,6
5	5,1	6,1	9,2
6	2,9	3,1	13

Слід прагнути до можливо меншої кількості пояснюючих змінних (чинників), оскільки окрім ускладнення моделі "зайві" чинники призводять до збільшення помилок оцінок. Так, в розглянутому прикладі стандартна

помилка оцінки параметра p_2 виявилася більше її значення за абсолютною величиною. Чіткої залежності прибутковості акцій компанії *Widget* від інфляції немає. В цьому випадку природно видалити чинник інфляції і розглядати залежність прибутковості *Widget* тільки від темпу росту ВВП. Відповідне рівняння регресії $\overline{Widget} = \alpha + \beta_1 temp$, оцінюване за допомогою методу найменших квадратів, має вид: $\overline{Widget} = 4 + 2temp$.

При цьому помилка параметра β_1 зменшується і стає рівною 1,002.

Проте в реальних ситуаціях іноді доводиться розглядати моделі залежності від десятків і навіть сотень чинників.

Контрольні питання

1. Що відрізняє статистичний аналіз фінансового ринку від статистичних аналізів в інших галузях.
2. Дати порівняльну характеристику для індексів, які використовуються на біржі.
3. Що таке «портфель цінних паперів»?
4. Для чого використовують метод найменших квадратів при аналізі прибутковості цінних паперів.

СПИСОК ЛІТЕРАТУРИ

Основна література:

1. В. І. Сущук-Слюсаренко, Є.С. Сулема, А.С. Герасимов. Інформаційні технології та математична статистика: навчально- методичний посібник для студентів спеціальностей 121 «Інженерія програмного забезпечення, Програмне забезпечення комп'ютерних та інформаційно-пошукових систем» – К. : НТУУ КПІ, 2018. – 133 с.
2. І.Є.Галицька, В.І.Сущук-Слюсаренко. Теорія ймовірностей та математична статистика. Розділ 1 «Математична статистика». Для студентів напрямів підготовки 6.050103 «Системна інженерія», 6.170101 «Безпека інформаційних і комунікаційних систем» / К.: ІВЦ "Видавництво «Політехніка»", 2011. - 216с.
3. Сапсай Т.Г., Сущук-Слюсаренко В. І.. Основи теорії ймовірностей: навчально-методичний посібник до вивчення дисципліни «Теоріія ймовірностей та математична статистика» для студентів напрямів підготовки 6.050103 «Програмна інженерія»,6.050102 «Комп'ютерна інженерія» [Електронне видання] /– К. : НТУУ «КПІ», 2014. – 201 с.

Додаткова література:

4. Кремер Н.Ш. Теория вероятностей и математическая статистика. – М.: Юнити, 2004. – 574с.
5. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Основы моделирования и первичная обработка данных. – М.: Финансы и статистика, 1983. – 250с.
6. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Исследование зависимостей. – М.: Финансы и статистика, 1985. – 220с.

7. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: Юнити, 1998. – 374с.
8. Бочаров П.П., Печенкин А.В. Теория вероятностей и математическая статистика. – М.: ЮНИТИ, 1998. – 245 с.
9. Вентцель Е.С. Исследование операций. Задачи, принципы, методология. – М.: Наука, 1990. – 296 с.
10. Крамер Г. Математические методы статистики: пер. с англ. – М.: Мир, 1975. – 145с.
11. Кремер Н.Ш. Математическая статистика. – М.: Экономическое образование, 1992. – 195с.
12. Справочник по прикладной статистике / Под ред. Э. Ллойда, У. Лидермана: пер. с англ. – М.: Финансы и статистика, 1989. – 411с.
13. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютерах / Под ред. В.Э. Фигурнова. – М.: ИНФРА- М, 1998. – 245с.
14. Ферстер Э., Ренц Б. Методы корреляционного и регрессионного анализа: пер. с нем. – М.: Финансы и статистика, 1983. – 371с.
15. Шарп У., Гордон Дж. А., Бейли Д. Инвестиции: пер. с англ. – М.: ИНФРА-М, 1997. – 178с.
16. Экономико – математические методы и прикладные модели / Под ред. В.В. Федосеева. – М.: ЮНИТИ, 1999.