Міністерство освіти і науки України
Університет митної справи та фінансів

ЛАБОРАТОРНИЙ ПРАКТИКУМ ІЗ ДИСЦИПЛІНИ “МИТНА СТАТИСТИКА”

ДЛЯ СТУДЕНТІВ ЕКОНОМІЧНОГО ФІНАНСОВОГО ТА ФАКУЛЬТЕТУ УПРАВЛІННЯ

Дніпро
2017

ББК 65.428
УДК 311:339.543

Рецензенти:
Говоруха В. Б. – д. фіз.-мат. н., проф., завідувач кафедри вищої математики Дніпропетровського аграрно-економічного університету.

Рекомендовано до друку кафедрою вищої математики та інформатики (протокол засідання кафедри №15 від 15.03.16)

Лабораторний практикум із дисципліни "Митна статистика" для студентів економічного, фінансового та факультету управління/ уклад. : В. М. Россочинський, О. Ф. Булгакова. – Дніпро: Університет митної справи та фінансів, 2016. – с. 46.

ISBN

Укладачі: Россочинський В. М., старший викладач кафедри прикладної математики та інформатики Університету митної справи та фінансів;
Булгакова О.Ф., старший викладач кафедри прикладної математики та інформатики Університету митної справи та фінансів

Зміст

ВСТУП

Метою виконання лабораторних робіт (далі – л. р.) є закріплення теоретичних знань і набуття практичних навичок проведення статистичних досліджень, які можуть виникати в реальній діяльності митних установ.

Лабораторний практикум містить основну теоретичну інформацію, постановку задачі, вихідні дані та приклад розв'язування типових задач до кожної л. р., а також вимоги до оформлення звітів.

Тематика л. р.: виявлення наявності та періоду коливань часових рядів; визначення типу коливань і прогнозування; аналіз концентрації та диференціації розподілів.

Виконувати кожну л. р. треба за таким загальним планом.

1. Ознайомитися з основною теоретичною інформацією.
2. Самостійно сформувати свій варіант вихідних даних, номер якого збігається з порядковим номером виконавця в журналі його академічної групи за поточний семестр.
3. Сформулювати задачу та виконати л. р. відповідно до наведеного прикладу.
4. Оформити звіт.
5. Для захисту л. р. підготувати відповіді на питання для самоконтролю.

До кожної л. р. окремо оформлюється звіт про її виконання за наведеними вимогами, у разі порушення яких звіт потрібно переробити.

Звіт має містити:

1. Номер, тему і мету л. р., а також номер варіанта вихідних даних.
2. Основні теоретичні відомості, необхідні для виконання л. р.
3. Відповідно до наведеного прикладу: а) постановку задачі та вихідні дані; б) хід виконання л. р., тобто розв'язання поставленої задачі; в) висновки.

Для виконання та оформлення л. р. можна використовувати будь-яку обчислювальну техніку, бажано – комп'ютери з пакетами стандартних програм.

Під час захисту кожної л. р. виконавець повинен: знати основні терміни, визначення, поняття, формули, й вміти оперувати ними, пояснювати й обґрунтовувати всі свої дії щодо виконання роботи та зроблених висновків; знати відповіді на питання для самоконтролю.

Основна теоретична інформація, наведена в кожній л. р., не містить усіх необхідних відомостей із теми, що вивчається, тому виконання і захист л. р. передбачає попереднє детальне вивчення даної теми за підручниками, посібниками, конспектами лекцій, а також знання відповідної тематики з навчальної дисципліни "Статистика".

Результати всіх обчислень повинні мати не менше чотирьох десяткових знаків і не менше чотирьох значущих цифр.

ВИМОГИ ДО ОФОРМЛЕННЯ ЗВІТУ

Звіт до кожної л. р. оформлюється на одному боці аркушів білого паперу формату А4 (210х297 мм) тільки у рукописному вигляді. Допускається наявність у звіті аркушів з друкованими результатами обчислень, таблицями, графіками тощо.

Усі таблиці, графіки, схеми, діаграми й інші графічні ілюстрації повинні бути пронумеровані та підписані. Таблиці оформлюються за відповідними вимогами до оформлення статистичних таблиць.

Аркуші звіту нумеруються з титульного. Номер записується в правому верхньому куті кожного аркуша, крім титульного. Титульний аркуш повинен мати такий вигляд, як наведено на рис. 1.

Кожна сторінка звіту має такі параметри: верхнє, нижнє та ліве поля – 20 мм, праве – 10 мм.

Текст звіту треба писати чорним, синім або фіолетовим кольором, одним почерком, охайно та розбірливо, без будь-яких скорочень слів, крім загальноприйнятих скорочень та абревіатур. Для оформлення таблиць, графіків, діаграм та інших ілюстрацій дозволяється використання будь-яких інших кольорів, крім червоного та його відтінків.

Будь-які помилки можна виправляти довільно, але чітко й охайно.

Заголовки структурних елементів звіту розташовуються посередині рядка без переносів слів і крапки в кінці, але підкреслюються. Між заголовком, попереднім і наступним текстом повинен бути один порожній рядок. Після заголовка на цій сторінці має бути не менше одного рядка подальшого тексту.

Абзацний відступ – однаковий у всьому тексті звіту.

Формули, рівняння та інші математичні вирази розташовуються посередині окремого рядка. Безпосередньо під ними необхідно наводити пояснення до всіх позначень, які належать до цього виразу.

Переноси математичних виразів на наступний рядок небажані, але за необхідності це можна робити тільки на знаках арифметичних операцій або знака рівності, повторюючи цей знак на початку наступного рядка. При цьому для позначення операції множення застосовується знак “×”.

Обчислення за формулами слід оформляти так:

$$\begin{pmatrix}\text{Позначення}\\ \text{величини, що}\\ \text{обчислюється}\end{pmatrix} = (\text{Формула}) = \begin{pmatrix}\text{Формула з}\\ \text{підставленими числовими}\\ \text{значеннями змінних}\end{pmatrix} = \begin{pmatrix}\text{Результат}\\ \text{обчислення}\end{pmatrix}$$

Звіт оформлюється за таким планом (починаючи з початку другої сторінки):

Тема: (тема роботи).

Мета роботи: (мета роботи).

Основна теоретична інформація

(Наводиться інформація, необхідна безпосередньо для виконання даної роботи).

Постановка задачі

(Зазначаються: постановка задачі (з типового прикладу до кожної л. р.); вихідні дані для відповідного варіанта, що визначені в кожній л. р.).

Розв'язання задачі

(Відповідно до наведеного в кожній л. р. типового прикладу).

Висновки

(Наводяться детальні висновки, аналіз отриманих результатів, необхідні на погляд виконавця інші відомості).

У кожному звіті має бути чистий останній аркуш для рецензії та зауважень.

Міністерство освіти і науки України
Університет митної справи та фінансів

Кафедра прикладної математики
та інформатики

Лабораторна робота № (номер роботи)
із дисципліни "Митна статистика"
варіант № (номер варіанта
вихідних даних)

Роботу виконав
студент групи (номер групи)
Прізвище, ім'я, по батькові
(повністю)

Роботу перевірив
(посада, прізвище та ініціали
викладача)

Дніпро
Рік

Рис. 1. Титульний аркуш

ЛАБОРАТОРНА РОБОТА № 1

Тема: Коливання рядів динаміки

Мета роботи: навчитись виявляти наявність коливань рівнів динамічного ряду та визначати період коливань.

Вихідні дані: виконавець формує самостійно відповідно до п. 1.4 основної теоретичної інформації.

Основна теоретична інформація

1.1.Основні поняття та аналітичне вирівнювання

Будемо вирізняти два типи коливань рядів динаміки: періодичні та квазіперіодичні (від лат. *quasi* – майже, начебто, не зовсім).

Періодичні (квазіперіодичні) коливання часового ряду розуміємо як коливання його рівнів, коли **значення** (**знаки** "+" або"–") відхилень рівнів ряду від загального тренду хоча б наближено повторюються через певні однакові (або такі, що вважаються однаковими) проміжки часу *T*. При цьому величина *T* називається **періодом** коливань.

Якщо період коливань дорівнює одному року, то такі коливання прийнято називати **сезонними**.

Під час дослідження коливань часового ряду проміжки часу між моментами для моментного ряду і часові інтервали для інтервального ряду повинні бути однаковими або такими, що вважаються однаковими. Нагадаємо, що в деяких випадках нерівномірний інтервальний ряд можна рівномірезувати.

Якщо для даного часового ряду виявлено наявність коливань (див. п. 1.3), то досліджувати такий ряд можна за допомогою його аналітичного вирівнювання трендовою кривою виду

$$y = f(t) \equiv \varphi(t) + \psi(t), \tag{1.1}$$

де $$\psi(t) = \sum_{m=1}^{k}(a_m \cdot \cos m\tau + b_m \cdot \sin m\tau), \tag{1.2}$$

$\tau = \pi t / l$, $2l = T$ – період коливань, який повинен бути відомим.

Функція $\varphi(t)$ являє собою **систематичну складову тренду** (або **загальний тренд**), яка виражає загальну тенденцію розвитку явища чи процесу. Часто $\varphi(t)$ обирається лінійною або квадратичною:

$\varphi(t)=a + b\cdot t$ або $\varphi(t)=p + q\cdot t + r\cdot t^2$.

Вираз $\psi(t)$ є **коливальною складовою тренду** і являє собою перші *k* гармонік тригонометричного ряду Фур'є та описує коливання рівнів ряду, які накладаються на загальний тренд $\varphi(t)$.

Зауважимо, що в деяких випадках значення *T* і, відповідно, наявність коливань однозначно витікає з економічної чи фізичної суті явища, що вивчається. В інших випадках виявляти наявність коливань і визначати величину їхнього періоду необхідно статистичними методами (див. п 1.3).

Обчислення параметрів a_m та b_m виразу $\psi(t)$ за методом найменших квадратів зводиться до побудови і розв'язування системи $2k$ рівнянь з $2k$ невідомими та коефіцієнтами, що мають достатньо складну і громіздку структуру. Це потребує проведення значної кількості обчислень, що не завжди доцільно. Тож на практиці часто зручніше користуватись спрощеним способом вивчення коливань за допомогою їх приростів, який не вимагає побудови тренду виду (1.1).

1.2. Вивчення коливань часових рядів за допомогою приростів коливань

Припустимо, що заданий часовий ряд має коливання своїх рівнів з певним періодом T.

Приростом коливань z_k^i для k-го рівня i-го періоду часового ряду називається різниця між фактичним значенням y_k^i рівня ряду і відповідним значенням систематичної складової тренду $\varphi(t_j)$:

$$z_k^i = y_k^i - \varphi(t_j), \tag{1.3}$$

де y_k^i – k-й рівень ряду в i-му періоді, $\varphi(t_j)$ – значення функції $\varphi(t)$ для такого значення t_j, яке відповідає даному рівню ряду y_k^i; $k = \overline{1, M}$; M – кількість рівнів ряду в кожному періоді. При цьому загальний тренд $\varphi(t)$повинен бути попередньо знайденим за даними максимально можливої кількості періодів.

Для більшої точності та хоча б часткового виключення впливу флуктуацій значень ознаки Y прирости коливань бажано обчислювати за даними всіх періодів, для яких знаходився загальний тренд $\varphi(t)$, як середню арифметичну відповідних приростів коливань для кожного з цих періодів:

$$z_k = {}_{i=1}^{m} z_k^i/m, \tag{1.4}$$

де z_k^i – приріст коливань для k-го рівня ряду в i-му періоді, m – кількість періодів спостереження, z_k – **загальний приріст коливань** для k-го рівня ряду в кожному періоді.

Якщо функція $\varphi\ \ t$ лінійна або квадратична (див. п. 1.1), то **перевірити правильність** обчислення приростів коливань z_k^i та z_k можна за умовами:

$$ {}_{i=1}^{m}\ {}_{k=1}^{M} z_k^i \approx 0, \tag{1.5}$$

$$ {}_{k=1}^{M} z_k \approx 0, \tag{1.6}$$

де m – кількість періодів спостереження, M – кількість рівнів ряду в кожному з m періодів.

Наглядно спостерігати динаміку коливань рівнів ряду на обраних m періодах можна за допомогою **лінії приростів коливань**, яка являє собою сукупність точок із координатами (t_k, z_k), побудованих у прямокутній системі координат $t0z$ і послідовно сполучених відрізками прямих. При цьому перша точка лінії може відповідати будь-якій парі (t_i, z_i) з обраних m періодів, але загальний проміжок часу, на якому будується лінія приростів коливань, повинен дорівнювати періоду коливань T, а кількість точок, що утворюють лінію приростів коливань – числу M рівнів ряду в кожному періоді.

Для сезонних коливань прирости коливань прийнято називати **приростами сезонності**, а їх графічне зображення – **лінією приростів сезонності**. Останню іноді називають **сезонною хвилею**.

1.3. Визначення виду загального тренду, наявності коливань та їх періоду

У практиці статистичних досліджень можлива ситуація, коли наявність коливань рівнів часового ряду не є очевидною або достатньо обґрунтованою. Виявити наявність або відсутність коливань можна таким чином.

Спочатку треба переконатись в тому, що на суб'єктивний погляд дослідника значення ознаки Y взагалі суттєво змінюються протягом часового проміжку, на якому розглядається ряд динаміки. Це можна з'ясувати шляхом візуального аналізу кореляційного поля або оцінки величини розмаху варіації $R = y_{\max} - y_{\min}$ для тих рівнів ряду, які належать вищезазначеному часовому проміжку.

Якщо, на погляд дослідника, значення R невелике, то всі рівні y_i ряду визнаються наближено рівними між собою, а часовий ряд не має будь-яких коливань, крім можливих незначних флуктуацій його рівнів. Очевидно, що в цьому разі загальний тренд слід вважати сталим:

$$\varphi(t) \equiv C = \sum_{i=0}^{N} y_i/(\mathrm{N}+1),$$

де N+1 – кількість рівнів ряду, якщо їх нумерація починається з нуля.

Якщо буде визнано, що y_i суттєво відмінні між собою, то виявити їх відхилення від попередньо знайденого загального тренду можна за допомогою значень модулів приростів коливань $z_i = y_i - \varphi(t_i)$, обчислених для всіх y_i, що належать проміжку спостереження $(i = \overline{0,\ N})$.

При цьому параметри лінійного $\varphi(t)=a+b\cdot t$ або квадратичного $\varphi(t)=p + q\cdot t + r\cdot t^2$ трендів знаходяться із систем лінійних алгебраїчних рівнянь відповідно

$$\begin{cases} (N+1) \times a + \left(\sum_{i=0}^{N} t_i\right) \times b = \sum_{i=0}^{N} y_i\,; \\ \left(\sum_{i=0}^{N} t_i\right) \times a + \left(\sum_{i=0}^{N} t_i^2\right) \times b = \sum_{i=0}^{N} t_i \times y_i \end{cases} \qquad (1.7)$$

або

$$
\begin{aligned}
(N+1)\times p + \left(\sum_{i=0}^{N} t_i\right)\times q + \left(\sum_{i=0}^{N} t_i^2\right)\times r &= \sum_{i=0}^{N} y_i\,;\\
\left(\sum_{i=0}^{N} t_i\right)\times p + \left(\sum_{i=0}^{N} t_i^2\right)\times q + \left(\sum_{i=0}^{N} t_i^3\right)\times r &= \sum_{i=0}^{N} t_i\times y_i\,;\\
\left(\sum_{i=0}^{N} t_i^2\right)\times p + \left(\sum_{i=0}^{N} t_i^3\right)\times q + \left(\sum_{i=0}^{N} t_i^4\right)\times r &= \sum_{i=0}^{N} t_i^2\times y_i\,,
\end{aligned}
\qquad (1.8)
$$

де N– номер останнього рівня ряду, а (N+1) – загальна кількість рівнів ряду, якщо нумерація рівнів починається з нуля.

Існують формули розв'язку системи (1.7) в загальному вигляді:

$$
b = \frac{\sum_{i=0}^{N} t_i\times y_i - \frac{1}{N+1}\cdot\left(\sum_{i=0}^{N} t_i\right)\times\left(\sum_{i=0}^{N} y_i\right)}{\sum_{i=0}^{N} t_i^2 - \frac{1}{N+1}\cdot\left(\sum_{i=0}^{N} t_i\right)^2}, \qquad (1.9)
$$

$$
a = \frac{1}{N+1}\times\left(\sum_{i=0}^{N} y_i - b\times\sum_{i=0}^{N} t_i\right), \qquad (1.10)
$$

за якими можна обчислювати параметри лінійного тренду.

Якщо значення z_i будуть визнані дослідником невеликими, то можна вважати, що часовий ряд не має суттєвих відхилень значень ознаки Y від загального тренду $\varphi(t)$. Очевидно, що в цьому разі коливання рівнів ряду слід вважати відсутніми.

Якщо значення z_i, на думку дослідника, будуть достатньо великими, то можна вважати, що рівні ряду мають значні відхилення від загального тренду, які можуть виявитись періодичними або квазіперіодичними коливаннями.

В останньому випадку формально підтвердити наявність або відсутність коливань із певною заданою **надійністю** (або **надійною імовірністю)** γ можна за **методом знаків**.

Для застосування даного методу необхідно попередньо обрати значення T періоду коливань, наявність яких припускається та перевіряється. На жаль, не існує формального способу його обчислення, тому на практиці вибір величини T зазвичай здійснюється такими можливими шляхами:

- візуальним аналізом графіка загального тренду на кореляційному полі;
- ретельним аналізом природи і суті явища чи процесу, динаміка якого вивчається;
- урахуванням результатів можливих аналогічних попередніх досліджень.

Зауважимо, що метод знаків передбачає можливість зміни обраного значення T, якщо в цьому виникне потреба.

Припустимо, що для ряду динаміки з рівнями y_j $(j=\overline{0,N})$ обрано певне значення періоду T з M рівнями ряду в кожному періоді. Якщо загальний тренд ряду має один і той же аналітичний вираз $\varphi(t)$ на часовому проміжку $t_0;\ t_N$, то для застосування методу знаків слід обрати максимально можливу кількість $m = \ (N+1)/M$ $\ \ ($ x – ціла частина числа x) останніх періодів. При цьому доцільно і зручно використати введену у п. 1.2 подвійну індексацію для позначення тих рівнів ряду, які увійшли до обраних m періодів: y_k^i, де i– номер періоду ($i = \overline{1, m}$), k – номер рівня ряду в кожному періоді ($k = \overline{1, M}$). Нижченаведена таблиця 1.1 ілюструє відповідність між початковими і новими позначеннями для, наприклад, N = 13, M = 4 і $m = [(13+1)/4] = 3$.

Таблиця 1.1

Відповідність між початковими і новими позначеннями рівнів ряду

Початкові позначення	y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}
Нові позначення	–	–	y_1^1	y_2^1	y_3^1	y_4^1	y_1^2	y_2^2	y_3^2	y_4^2	y_1^3	y_2^3	y_3^3	y_4^3
Номер періоду (i)			1				2				3			

Суть методу знаків полягає у порівнянні знаків значень z_k^i та z_k^j для кожної можливої комбінації неупорядкованих пар (i ; j), де $i \neq j$, z_k^l – приріст коливань для k-го рівня y_k^l ряду в l-му періоді. При цьому ознакою наявності коливань є виконання умови

$$z_k^i \times z_k^j > 0 \tag{1.11}$$

для кожної пари (i ; j) і для кожного значення k.

Однак унаслідок впливу на значення рівнів ряду випадкових факторів для деяких пар (i ; j) та для деяких k умова (1.11) може не виконуватись навіть за фактичної наявності коливань. Очевидно, що число S(i ; j) таких пар $(z_k^i; z_k^j)$ для всіх $k = \overline{1, M}$ з i-го та j-го періодів, у яких знаки z_k^i та z_k^j не збігаються, повинно бути невеликим порівняно з M. При цьому природно, виникає питання: наскільки великим має бути максимальне значення S величин S(i;j), щоб можна було вважати, що незбігання знаків z_k^i та z_k^j у такої кількості S пар $(z_k^i; z_k^j)$ викликане не випадковими факторами, а фактичною відсутністю коливань із заданим періодом T.

Формальну відповідь на поставлене питання можна отримати за допомогою **критерія знаків** (або ***s*-критерія**), відповідно до якого спостережене (тобто обчислене за наявними даними) максимальне значення S величин S(i ; j): $S = \max_{i,j} S(i\ ;\ j)$ необхідно порівняти з його критичним значенням $S_{\text{кр.}}$, яке залежить від числа M рівнів ряду в кожному періоді та рівня значущості α: $S_{\text{кр.}} = S(M;\ \alpha)$, де α – імовірність ризику визнати коливання існуючими при їх фактичної відсутності, тобто ймовірність зробити помилку. При цьому значення $S(M;\ \alpha)$ знаходиться за таблицею критичних значень s- критерію (див. додаток 1).

Після знаходження значень $S(i\ ;\ j)$ для кожної можливої пари $(i\ ;\ j)$, $S(M;\ \alpha)$ та $S = \max_{i,j} S(i;j)$ перевірка наявності коливань виконується за правилом:

- якщо $S\ \leq S(M;\ \alpha)$, (1.12)

то коливання з даним періодом T вважаються існуючими з **надійною імовірністю** (або **надійністю**) $\gamma = 1 - \alpha$;

- якщо $S > S(M;\ \alpha)$,

то коливання з даним періодом T вважаються неіснуючими з тією ж надійністю.

Необхідно підкреслити, що невиконання умови (1.12) ще не означає, що коливання насправді не існують: вони лише вважаються неіснуючими з надійністю γ. На практиці це означає, що при проведенні великої кількості таких досліджень для часових рядів, що **фактично** не мають коливань, у $\gamma \times 100\ \%$ досліджень коливання повинні бути **визнані** неіснуючими, а у $\alpha \times 100\ \%$– існуючими. Очевидно, що в останніх випадках буде зроблено помилку.

1.4. Формування вихідних даних

Вихідні дані формуються виконавцем самостійно за допомогою комп'ютерних програм. При цьому спочатку необхідно обчислити номер N останнього рівня ряду динаміки за формулою

N=[(310 – l)/7],

а також числа G та H:

G= 0,05 l + 5,
H= 0,1 l – 1,35,

де l – порядковий номер виконавця за списком у журналі академічної групи за поточний семестр, [x] – ціла частина числа x.

Після обчислення чисел N, G і H виконати такі операції.

1. Увімкнути комп'ютер.

2. За допомогою маніпулятора **мышь** запустити табличний процесор Microsoft Excel.

3. Встановити курсор у комірку *C*1.

4. Увійти в меню **Сервис** та вибрати рядок **Анализ данных** (для версії Microsoft Office 2003) або перейти на вкладку **Данные** та у групі **Анализ** натиснути кнопку **Анализ данных** (для версії Microsoft Office 2010).

Увага! Якщо пакет аналізу Microsoft Excel 2010 не встановлений на вашому комп'ютері, завантажити його можна так:

*1. Відкрийте вкладку **Файл** та оберіть пункт **Параметры**.*

*2. Оберіть команду **Надстройки**, після чого у полі **Управление** оберіть рядок **Надстройки Excel**.*

*3. Натисніть кнопку **Перейти**.*

*4. У вікні **Надстройки** встановіть прапорець **Пакет анализа** та натисніть кнопку **ОК**.*

*5. Після завантаження пакета аналіза на вкладці **Данные** у групі **Анализ** стає доступною команда **Анализ данных**.*

5. Після появи вікна **Анализ данных** обрати рядок **Генерация случайных чисел** і натиснути кнопку ОК. На екрані монітора має з'явитись вікно з назвою **Генерация случайных чисел**.

6. У полі **Распределение** з випадаючого списку обрати рядок **Равномерное**.

7. У полях **Число переменных** та **Число случайных чисел** записати числа відповідно 1 та $N + 1$.

8. В області **Параметры** у поля` **Между** та **и** записати числа відповідно –1 та 1.

9. В області **Параметры вывода** у полі **Выходной интервал** задати адресу комірки $*C*$1 та натиснути кнопку **ОК**.

10. На екрані монітора у стовпці *C* мають з'явитися $N + 1$ чисел.

11. Записати у комірку *A*1 число 0.

12. Встановити курсор у комірку *A*1.

13. Увійти в меню **Правка**, обрати рядок **Заполнить → Прогрессия** (для версії Microsoft Office 2003) або перейти на вкладку **Главная**, у групі **Редактирование** натиснути кнопку **Заполнить** та обрати команду **Прогрессия** (для версії Microsoft Office 2010). На екрані монітора має з'явитись вікно **Прогрессия**.

14. У полях **Шаг** та **Предельное значение** записати числа відповідно 1 та *N*.

15. У полі **Расположение** обрати перемикач **по столбцам** та натиснути кнопку **ОК**. На екрані монітора у стовпці *A* мають з'явитися цілі числа від 0 до *N*.

16. У комірки $G1$, $H1$ та $L1$ записати попередньо знайдені числа відповідно G, H та l.

17. Встановити курсор у комірку $D1$ та занести формулу

=$G$1*(SIN(0,5236*(A1+$L$1))+3).

18. Скопіювати цю формулу у стовпці D для рядків від 1 до $N + 1$.

19. Встановити курсор у комірку $E1$ та занести формулу

=$H$1*A1.

20. Скопіювати цю формулу у стовпці E для рядків від 1 до $N + 1$.

21. Встановити курсор у комірку $B1$ та занести формулу

=ОКРУГЛ(СУММ(C1:E1);4).

Таким чином, одержані числа будуть округлені до 4 десяткових знаків.

22. Скопіювати цю формулу у стовпці B для рядків від 1 до $N + 1$.

23. На екрані монітора у стовпцях A та B будуть записані відповідно час $t_k = k$ і рівні y_k вихідного часового ряду ($k = \overline{0,N}$).

1.5. Приклад постановки і розв'язування типової задачі

Постановка задачі

Динаміка абсолютного приросту фізичного обсягу експорту певного товару по деякій митниці наведена в табл. 1.2.

Для даного динамічного ряду необхідно:

1. Побудувати кореляційне поле в прямокутній системі координат $t0y$.

2. За видом кореляційного поля обрати лінійну або квадратичну модель загального тренду $\varphi(t)$. Вибір виду тренду обґрунтувати письмово у звіті до л. р. За необхідності для полегшення вибору виду тренду можна побудувати кореляційне поле з більшим масштабом по вісі $0y$.

3. За методом найменших квадратів обчислити параметри обраної функції $\varphi(t)$ як розв'язки систем (1.7) або (1.8). Якщо за результатом візуального аналізу кореляційного поля буде обрано лінійну функцію, то її параметри зручно обчислювати за формулами (1.9) та (1.10).

4. На кореляційному полі побудувати графік загального тренду і шляхом візуального аналізу переконатись у наявності значних відхилень фактичних рівнів y_i $(i = \overline{0,N})$ ряду від загального тренду $\varphi(t)$ та визначити період T можливих коливань і число M рівнів ряду, що належать одному періоду.

5. Обчислити максимально можливе число $m = (N + 1)/M$ останніх періодів і за їх даними формально перевірити наявність коливань із зада-

ним періодом T (числом M) за методом знаків з надійністю $\gamma = 0{,}95$. Очевидно, що при цьому $y_1^1 = y_{N-m\cdot M+1}; \dots; y_M^m = y_N$.

Якщо за результатами застосування методу знаків буде виявлена відсутність коливань, то повторити дослідження з іншим допустимим періодом T (числом M). Зробити висновки щодо наявності або відсутності коливань із даним періодом і числом M.

Таблиця 1.2

Динаміка абсолютного приросту фізичного обсягу експорту (дані умовні)

Час $t_k = k$	0	1	2	3	4	5	6	7	8
Обсяг експорту, тис. т, y_k	14,7211	15,7497	16,9836	15,7084	12,9701	10,9708	6,9571	3,0934	– 0,4266
Час $t_k = k$	9	10	11	12	13	14	15	16	17
Обсяг експорту, тис. т, y_k	– 1,5967	– 3,3762	– 1,7688	– 1,0231	0,0159	0,9748	– 1,1037	– 1,4699	– 5,7071
Час $t_k = k$	18	19	20	21	22	23	24	25	26
Обсяг експорту, тис. т, y_k	– 9,9391	– 13,4611	– 15,7381	– 17,4264	– 19,3300	– 18,1732	– 18,0289	– 16,0605	– 15,0614
Час $t_k = k$	27	28	29	30	31	32	33	34	35
Обсяг експорту, тис. т, y_k	– 16,3546	– 17,9664	– 21,283	– 25,3754	– 29,3811	– 31,9349	– 33,8429	– 35,233	– 35,2884
Час $t_k = k$	36	37	38	39	40	41	42	43	44
Обсяг експорту, тис. т, y_k	– 34,1639	– 32,1511	– 31,2873	– 33,0277	– 34,1604	– 38,0622	– 41,7533	– 46,5092	– 49,3452

Розв’язування задачі

1. Будуємо кореляційне поле заданого динамічного ряду (рис. 1.1).
2. За результатом візуального аналізу кореляційного поля обираємо лінійну модель загального тренду $\varphi(t) = a + b \cdot t$.
3. За формулами (1.9) та (1.10) обчислюємо параметри тренду. Проміжні обчислення зручно організувати в таблиці (графи 1–4 табл. 1.3).

$$b = \frac{\sum_{i=0}^{44} t_i \cdot y_i - \frac{1}{45} \cdot \left(\sum_{i=0}^{44} t_i\right) \cdot \left(\sum_{i=0}^{44} y_i\right)}{\sum_{i=0}^{44} t_i^{\,2} - \frac{1}{45} \cdot \left(\sum_{i=0}^{44} t_i\right)^2} =$$

$$= \frac{-24677{,}8171 - \frac{1}{45}\cdot 990\cdot(-648{,}6659)}{29370{,}0000 - \frac{1}{45}\cdot 980100} \approx -1{,}3712\,;$$

$$a = \frac{1}{45}\cdot\left(\sum_{i=0}^{44} y_i - b\cdot\sum_{i=0}^{44} t_i\right) =$$

$$= \frac{1}{45}\cdot\left(-648{,}6659 - (-1{,}3712)\cdot 990\right) \approx 15{,}7516.$$

Маємо загальний тренд $\varphi(t)$=15,7516 – 1,3712 · t.

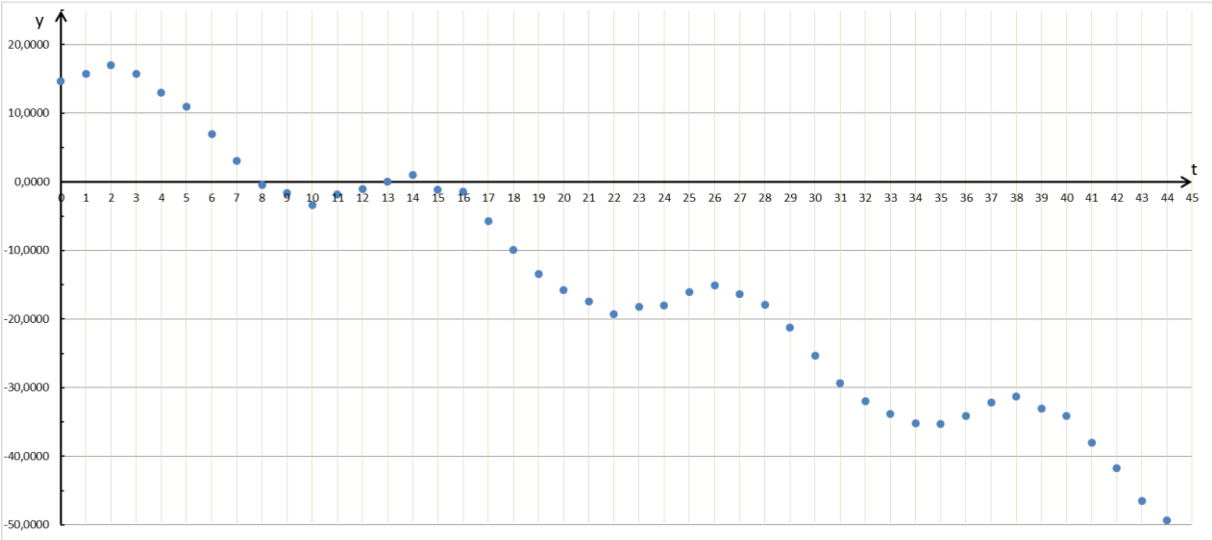

Рис. 1.1. Кореляційне поле

4. Із візуального аналізу графіка загального тренду (рис. 1.2) на кореляційному полі можна зробити висновки:

a) рівні y_i мають значні відхилення від загального тренду;

b) є підстави висунути припущення про наявність коливань з періодом T = 12, тобто одному періоду належить M = 12 рівнів ряду.

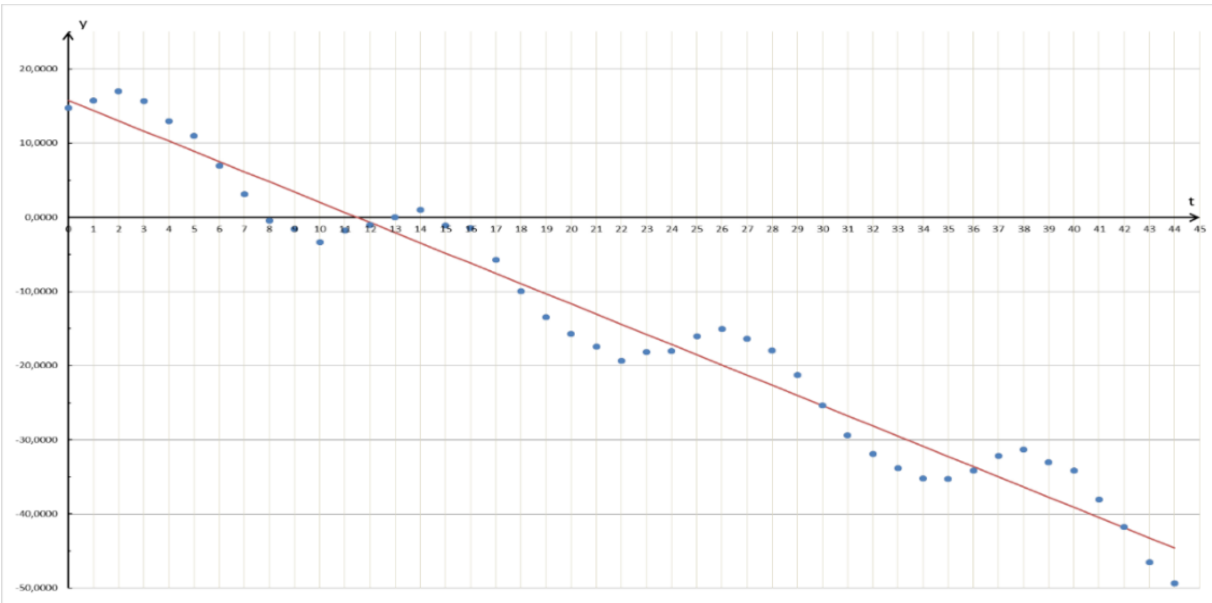

Рис. 1.2. Кореляційне поле та графік загального тренду

Таблиця 1.3

Розрахункова таблиця

	t_i	y_i	$t_i \cdot y_i$	t_i^2	$\varphi(t_i)$	$z_i = y_i - \varphi(t_i)$
	1	2	3	4	5	6
×	0	14,7211	0,0000	0,0000	–	–
	1	15,7497	15,7497	1,0000	–	–
	2	16,9836	33,9672	4,0000	–	–
	3	15,7084	47,1252	9,0000	–	–
	4	12,9701	51,8804	16,0000	–	–
	5	10,9708	54,8540	25,0000	–	–
	6	6,9571	41,7426	36,0000	–	–
	7	3,0934	21,6538	49,0000	–	–
	8	-0,4266	-3,4128	64,0000	–	–
	9	-1,5967	-14,3703	81,0000	3,4108	-5,0075
	10	-3,3762	-33,7620	100,0000	2,0396	-5,4158
	11	-1,7688	-19,4568	121,0000	0,6684	-2,4372
	12	-1,0231	-12,2772	144,0000	-0,7028	-0,3203
	13	0,0159	0,2067	169,0000	-2,0740	2,0899
	14	0,9748	13,6472	196,0000	-3,4452	4,4200
	15	-1,1037	-16,5555	225,0000	-4,8164	3,7127
	16	-1,4699	-23,5184	256,0000	-6,1876	4,7177
	17	-5,7071	-97,0207	289,0000	-7,5588	1,8517
	18	-9,9391	-178,9038	324,0000	-8,9300	-1,0091
	19	-13,4611	-255,7609	361,0000	-10,3012	-3,1599
	20	-15,7381	-314,7620	400,0000	-11,6724	-4,0657
	21	-17,4264	-365,9544	441,0000	-13,0436	-4,3828
	22	-19,3300	-425,2600	484,0000	-14,4148	-4,9152
	23	-18,1732	-417,9836	529,0000	-15,7860	-2,3872
	24	-18,0289	-432,6936	576,0000	-17,1572	-0,8717
	25	-16,0605	-401,5125	625,0000	-18,5284	2,4679
	26	-15,0614	-391,5964	676,0000	-19,8996	4,8382
	27	-16,3546	-441,5742	729,0000	-21,2708	4,9162
	28	-17,9664	-503,0592	784,0000	-22,6420	4,6756
	29	-21,2830	-617,2070	841,0000	-24,0132	2,7302
	30	-25,3754	-761,2620	900,0000	-25,3844	0,0090
	31	-29,3811	-910,8141	961,0000	-26,7556	-2,6255
	32	-31,9349	-1021,9168	1024,0000	-28,1268	-3,8081
	33	-33,8429	-1116,8157	1089,0000	-29,4980	-4,3449
	34	-35,2330	-1197,9220	1156,0000	-30,8692	-4,3638
	35	-35,2884	-1235,0940	1225,0000	-32,2404	-3,0480
	36	-34,1639	-1229,9004	1296,0000	-33,6116	-0,5523
	37	-32,1511	-1189,5907	1369,0000	-34,9828	2,8317
	38	-31,2873	-1188,9174	1444,0000	-36,3540	5,0667
	39	-33,0277	-1288,0803	1521,0000	-37,7252	4,6975
	40	-34,1604	-1366,4160	1600,0000	-39,0964	4,9360
	41	-38,0622	-1560,5502	1681,0000	-40,4676	2,4054
	42	-41,7533	-1753,6386	1764,0000	-41,8388	0,0855
	43	-46,5092	-1999,8956	1849,0000	-43,2100	-3,2992
	44	-49,3452	-2171,1888	1936,0000	-44,5812	-4,7640
Σ	**990,0000**	**-648,6659**	**-24677,8171**	**29370,0000**	×	×

5. Для формальної перевірки наявності коливань із даним періодом обчислимо максимально можливу кількість m періодів:

$$m = (44+1)/12 = 3,$$

після чого зробимо подвійну індексацію рівнів ряду, що належать останнім трьом періодам. При цьому $y_1^1 = y_{44-3\cdot12+1} = y_9,\ y_2^1 = y_{10}, \dots, y_{12}^1 = y_{20};$
$y_1^2 = y_{21}, \dots, y_{12}^2 = y_{32};\ y_1^3 = y_{33}, \dots, y_{12}^3 = y_{44}$.

Відкидаємо перші $N + 1 - m \cdot M = 44 + 1 - 3 \cdot 12 = 9$ рівнів ряду і обчислюємо прирости коливань $z_i = y_i - \varphi(t_i)$ для всіх рівнів ряду, що належать останнім трьом періодам ($i = \overline{9, 44}$). Обчислення зручно організувати в таблиці (графи 5, 6 табл. 1.3). Результати обчислень зручно оформити в табл. 1.4 (графи 2–4) з подвійною індексацією позначень приростів коливань.

Таблиця 1.4

Прирости коливань і їх знаки

k	Прирости коливань за один період			$sign\ z_k^i$		
	z_k^1	z_k^2	z_k^3	i= 1	i= 2	i= 3
1	2	3	4	5	6	7
1	–5,0075	–4,3828	–4,3449	–	–	–
2	–5,4158	–4,9152	–4,3638	–	–	–
3	–2,4372	–2,3872	–3,0480	–	–	–
4	–0,3203	–0,8717	–0,5523	–	–	–
5	2,0899	2,4679	2,8317	+	+	+
6	4,4200	4,8382	5,0667	+	+	+
7	3,7127	4,9162	4,6975	+	+	+
8	4,7177	4,6756	4,9360	+	+	+
9	1,8517	2,7302	2,4054	+	+	+
10	–1,0091	0,0090	0,0855	–	+	+
11	–3,1599	–2,6255	–3,2992	–	–	–
12	–4,0657	–3,8081	–4,7640	–	–	–

Знаки значень z_k^i наведено в графах 5–7 таблиці 1.4, звідки знаходимо:

$S(1;2) = 1;\ S(1;3) = 1;\ S(2;3) = 0;\ S = \max_{i,j} S(i;j) = 1.$

За таблицею критичних значень s-критерію для $\alpha = 1 - \gamma = 1 - 0{,}95 = 0{,}05$ (див. додаток 1) знаходимо $S_{кр.} = S(12; 0{,}05) = 2.$

Оскільки $S \leq S_{кр}$ і виконується умова (1.12), то з надійністю 95 % можна вважати, що часовий ряд має коливання з періодом T = 12.

1.6. Питання для самоконтролю

1. Для ряду динаміки дати означення:

а) періодичних коливань;

б) квазіперіодичних коливань;

в) періоду коливань;

г) сезонних коливань;

д) загального тренду;

е) приросту коливань;

ж) загального приросту коливань;

з) лінії приростів коливань.

2. Як можна перевірити правильність обчислення приростів коливань?

3. Якими способами можна виявити величину T періоду коливань?

4. Пояснити суть методу знаків.

5. Пояснити зміст критерію знаків.

ЛАБОРАТОРНА РОБОТА № 2

Тема: Коливання рядів динаміки

Мета роботи: навчитись визначати тип коливань і виконувати точковий та інтервальний прогнози.

Вихідні дані: вихідні дані для л. р.№ 1 та результати її виконання.

Основна теоретична інформація

2.1. Основна теоретична інформація для л. р. № 1

2.2. Уточнення параметрів загального тренду

Загальний тренд $\varphi(t)$, параметри якого обчислювались як розв'язки систем (1.7) або (1.8), може виявитись з певних причин викривленим. Тому після виявлення наявності коливань і визначення їх періоду T необхідно уточнити значення параметрів загального тренду.

Для цього необхідно побудувати і розв'язати систему лінійних алгебраїчних рівнянь виду (1.7) або (1.8). При цьому у рівняннях (1.7) та (1.8) й у формулах (1.9), (1.10):

а) додавання має виконуватись за всіма індексами i від $i = v$ до $i = u$ (${}_{i=v}^{u}$ замість ${}_{i=0}^{N}$), де $v = u - m \times M + 1$;

б) кількість рівнів нового часового ряду дорівнюватиме $m \times M$ замість $N + 1$: $i = \overline{v,\ u}$.

При цьому номер u останнього доданка і, відповідно, останнього рівня нового часового ряду визначається так. Необхідно побудувати три крайні праві точки A, B і C перетину графіка загального тренду $\varphi(t)$ з графіком динамічного ряду (рис. 2.1) і за цим рисунком знайти хоча б наближено графічно абсциси цих точок $t_A < t_B < t_C$.

Після цього значення u визначається за правилом:

- якщо $Q = t_C + (t_B - t_A)/2 < N$, то $u \approx Q$ за правилами округлення з точністю до цілих;
- якщо $N \leq Q \leq N + 0{,}5$, то $u = N$;
- якщо $N + 0{,}5 < Q$, то $u \approx \frac{t_C + t_B}{2} = P$ за правилами округлення з точністю до цілих.

Якщо обчислене значення $Q(P)$ виявиться точною серединою відрізка між двома сусідніми значеннями часу t на осі $0t$, то за u можна взяти $Q - 0{,}5$ або $Q + 0{,}5$ ($P - 0{,}5$ або $P + 0{,}5$) на суб'єктивний вибір дослідника. Наприклад, якщо $Q = 64{,}5$, то можна взяти $u = 64$ або $u = 65$ за умови $N \geq 65$.

Якщо після обчислення u і v виявиться, що $u = N$ і $v = 0$, то можна вважати, що попередньо знайдений загальний тренд $\varphi(t)$ не має викривлень, його параметри не потребують уточнень та він може використовуватись для визначення типу коливань і виконання прогнозу.

Для ілюстрації вищезазначеного розглянемо приклад на знаходження номера u. Нехай прикінцева частина графіка загального тренду на кореляційному полі має вигляд, зображений на рис. 2.1.

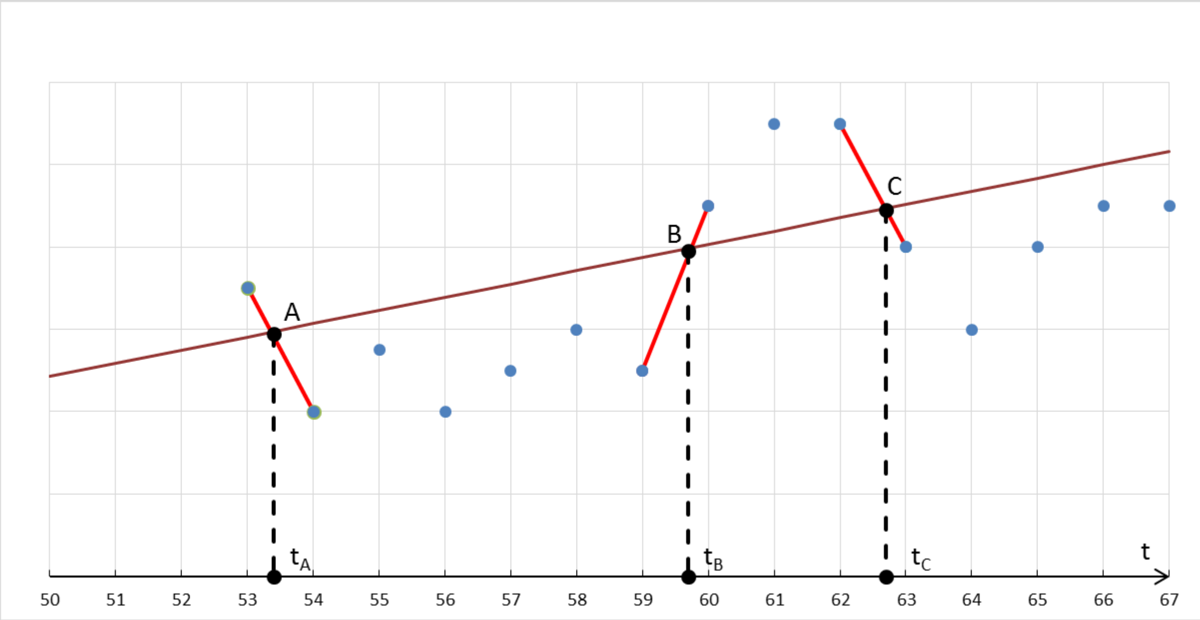

Рис. 2.1. Прикінцева частина графіка загального тренду
на кореляційному полі

За рис. 2.1 наближено графічно знаходимо $t_A \approx 53{,}4$; $t_B \approx 59{,}7$; $t_C \approx 62{,}6$. Будемо знаходити номер u для всіх можливих значень $N \geq 63$:

1. Нехай $N \geq 66$; $Q = t_C + \frac{t_B - t_A}{2} = 62{,}6 + \frac{59{,}7-53{,}4}{2} = 65{,}75$. Оскільки $Q < N$, то $u \approx Q = 65{,}75 \approx 66$.

2. Нехай $N = \overline{63,\ 65}$, тоді $Q = 65{,}75 > N + 0{,}5$ і $u \approx P = \frac{t_C + t_B}{2} = \frac{62.6+59.7}{2} = 61{,}15 \approx 61$.

Новий загальний тренд $\varphi(t)$, параметри якого обчислені за даними m періодів від y_v до y_u, не має викривлень і його можна використовувати для визначення типу коливань та виконання прогнозу.

2.3. Визначення типу коливань

Для визначення типу коливань необхідно спочатку обчислити для виправленого загального тренду $\varphi(t)$:

а) нові значення приростів коливань z_k^i за формулою (1.3) для y_i від y_v до y_u;

б) загальні прирости z_k за формулою (1.4);

в) число $z = \max_{i,k} z_k^i - z_k$ (2.1)

для всіх $i = \overline{1, m}$ та $k = \overline{1, M}$;

г) регресійне середнє квадратичне відхилення

$$S_{\text{рег}} = \sqrt{\frac{\sum_{i=1}^{m} \sum_{k=1}^{M} \left(z_k^i - z_k\right)^2}{m \cdot M - e}}, \tag{2.2}$$

де e – кількість параметрів загального тренду $\varphi(t)$, що знаходились за методом найменших квадратів. Після цього тип коливань визначається за правилом:

- якщо $z \leq 2 \times S_{\text{рег}}$, (2.3)

то коливання вважатимемо **періодичними**;

- якщо умова (2.3) не виконується, то коливання вважатимемо **квазіперіодичними**.

2.4. Точкове та інтервальне прогнозування ряду динаміки

Точковий прогноз, тобто точкове оцінювання рівня y_τ ряду для майбутнього часу $\tau > t_N$ виконується шляхом обчислення значення виправленого загального тренду $\varphi(t)$ для $t = \tau$ з урахуванням наявних коливань:

$$y_\tau = \varphi(\tau) + z_\tau, \tag{2.4}$$

де z_τ –загальний приріст коливань, що відповідає часу τ.

Інтервальний прогноз, тобто інтервальне оцінювання виконується тільки для рівномірних рядів і являє собою **надійний інтервал** $[y_л; y_п]$, який повинен із заданою **надійністю** (або **надійною імовірністю**) γ накрити майбутнє значення ознаки Y, що оцінюється.

Надійний інтервал визначається своїми межами, які обчислюються за формулами:

$$y_л = y_\tau - t(n;\alpha) \cdot S_{\text{рег}} \cdot \mu_\tau, \tag{2.5}$$

$$y_п = y_\tau + t(n;\alpha) \cdot S_{\text{рег}} \cdot \mu_\tau, \tag{2.6}$$

де $y_л$ та $y_п$ – відповідно ліва і права межі, $t(n;\alpha)$ – **коефіцієнт довіри** (або **довірче число**), який знаходиться за таблицею критичних значень розподілу Стьюдента для двосторонньої критичної області залежно від **числа ступенів вільності** $n = m \cdot M - e$ і **рівня значущості** $\alpha = 1 - \gamma$ (додаток 2); **коефіцієнт прогнозу**

$$\mu_\tau = \sqrt{\frac{m \cdot M \cdot (m \cdot M + 2)^2 + 3 \cdot (m \cdot M + 2 \cdot H_\tau)^2}{m \cdot M \cdot (m \cdot M + 1) \cdot (m \cdot M + 2)}}, \tag{2.7}$$

де $H_\tau = \tau - t_u$.

Таким чином, майбутнє значення рівня ряду для часу τ повинно з надійною імовірністю γ накриватись інтервалом

$$(y_\tau - t(n;\alpha) \cdot S_{\text{рег}} \cdot \mu_\tau \;;\; y_\tau + t(n;\alpha) \cdot S_{\text{рег}} \cdot \mu_\tau). \tag{2.8}$$

2.5. Приклад постановки та розв'язування типової задачі

Постановка задачі

1. За результатами виконання л. р. № 1 створити новий часовий ряд, що містить максимально можливу кількість m періодів і, відповідно $m \cdot M$ рівнів від y_v до y_u вихідного ряду, заданого таблицею 1.2.

2. Якщо $v = 0,\ u = N$ і новий ряд збігається з вихідним, то перейти до п. 4.

3. Для нового часового ряду обчислити нові уточнені значення параметрів нового не викривленого загального тренду і побудувати його графік на кореляційному полі.

4. Визначити тип коливань.

5. Виконати:

a) точкові прогнози на час $\tau = (N+1)$ та $\tau = (N+2)$;

b) інтервальний прогноз на час $\tau = (N+1)$ з надійністю $\gamma = 0{,}95$;

c) інтервальний прогноз на час $\tau = (N+2)$ з надійністю $\gamma = 0{,}90$.

Зобразити точкові й інтервальні прогнози на графіку.

6. Побудувати лінію приростів коливань.

7. Зробити висновки.

Розв'язування задачі

1. За прикінцевою частиною графіка попередньо знайденого тренду на кореляційному полі (рис. 2.2) наближено графічно знаходимо точки A, B, C і, відповідно, $t_A \approx 30$; $t_B \approx 36{,}2$; $t_C \approx 42$.

$Q = t_C + \frac{t_B - t_A}{2} = 42 + \frac{36{,}2-30}{2} = 45{,}1$.

Оскільки $N + 0{,}5 = 44 + 0{,}5 = 44{,}5 < 45{,}1 = Q$, то $u \approx \frac{t_C + t_B}{2} = \frac{42+36{,}2}{2} = 39{,}1 \approx 39$, тоді $v = u + 1 - m \cdot M = 39 + 1 - 3 \cdot 12 = 4$.

2. Оскільки $v \neq 0$ і $u \neq N$, то новий часовий ряд не збігається з вихідним і містить рівні від y_4 до y_{39}.

3. За видозміненими формулами (1.9) та (1.10) обчислюємо нові не викривлені параметри a і b загального тренду:

$$b = \frac{\sum_{i=4}^{39} t_i \cdot y_i - \frac{1}{3 \cdot 12} \cdot \left(\sum_{i=4}^{39} t_i\right) \cdot \left(\sum_{i=4}^{39} y_i\right)}{\sum_{i=4}^{39} (t_i^2) - \frac{1}{3 \cdot 12} \cdot \left(\sum_{i=4}^{39} t_i\right)^2} =$$

$$= \frac{(-15922{,}9700) - \frac{1}{36} \cdot 774 \cdot (-501{,}9984)}{20526{,}0000 - \frac{1}{36} \cdot 599076} \approx -1{,}3205;$$

$$a = \frac{1}{3 \cdot 12} \cdot \left(\sum_{i=4}^{39} y_i - b \cdot \sum_{i=4}^{39} t_i\right) =$$

$$= \frac{1}{36} \cdot \left(-501{,}9984 - (-1{,}3205) \cdot 774\right) \approx 14{,}4464.$$

Проміжні обчислення зручно організувати в табл. 2.1.

Маємо виправлений загальний тренд $\varphi(t)$ = 14,4464 – 1,3205·t (рис. 2.3).

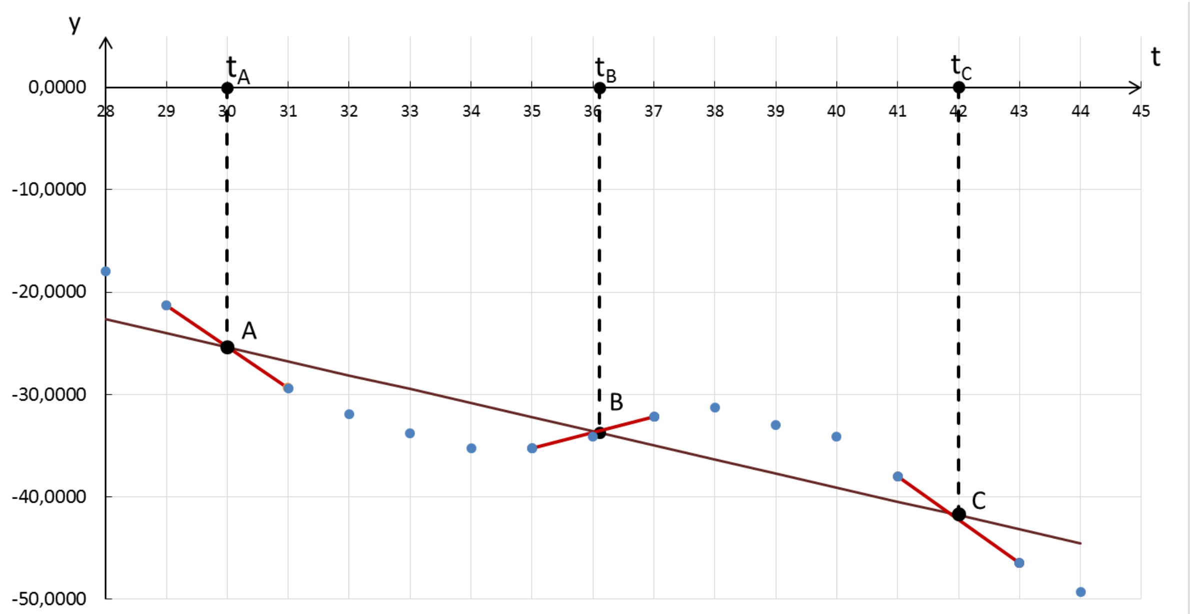

Рис. 2.2. Прикінцева частина графіка попередньо знайденого загального тренду на кореляційному полі

4. Для визначення типу коливань обчислимо для нового часового ряду від y_4 до y_{39} і виправленого загального тренду $\varphi(t)$ прирости коливань z_k^i за формулою (1.3), загальні прирости коливань за формулою (1.4) та перевіримо правильність обчислень за умовами (1.5) та (1.6).

Обчислимо регресійне середнє квадратичне відхилення $S_{\text{рег}}$ за формулою (2.2) та знаходимо значення z. Обчислення зручно організувати в табл. 2.2.

Оскільки $\sum_{i=1}^{3}\sum_{k=1}^{12} z_k^i = -0{,}0018 \approx \sum_{k=1}^{12} z_k = -\,0{,}0006 \approx 0$, то немає підстав сумніватись у правильності обчислень приростів коливань. За даними табл. 2.2 (графи 6–8)

$z^1 = \max_k |z_k^1 - z_k| = 0{,}8695;$

$z^2 = \max_k |z_k^2 - z_k| = 0{,}6855;$

$z^3 = \max_k |z_k^3 - z_k| = 1{,}0323;$

$z = \max_i z^i = 1{,}0323.$

За даними табл. 2.2 (графи 9-11)

$S^1 = \sum_{k=1}^{12} (z_k^1 - z_k)^2 \approx 2{,}9200;$

$S^2 = \sum_{k=1}^{12} (z_k^2 - z_k)^2 \approx 1{,}3811;$

$S^3 = \sum_{k=1}^{12} (z_k^3 - z_k)^2 \approx 1{,}9309;$

$S_{\text{рег}} = \sqrt{\frac{\sum_{i=1}^{3} S^i}{3\cdot 12-2}} = \sqrt{\frac{6{,}2321}{34}} \approx 0{,}4281; \quad 2\cdot S_{\text{рег}} = 0{,}8562\,.$

Оскільки $z > 2 \cdot S_{\text{рег}}$ (умова 2.3 не виконується), то коливання вважатимемо квазіперіодичними.

5. а) Точкові прогнози на час $t_{45} = 45$ і $t_{46} = 46$ виконуємо шляхом екстраполяції виправленого загального тренду $\varphi(t)$ з урахуванням наявних коливань за формулою (2.4). Очевидно, що при цьому часу t_{45} відповідає шостий рівень у четвертому періоді y_6^4 нового часового ряду: $t_{45} - t_{39} = 45 - 39 = 6$ і відповідно загальний приріст коливань z_6, а часу t_{46} – z_7. Тоді

$$y_{45} = \varphi(45) + z_6 = -44{,}9761 + (-4{,}3379) = -49{,}3140;$$
$$y_{46} = \varphi(46) + z_7 = -46{,}2966 + (-4{,}7085) = -51{,}0051.$$

На кореляційному полі із графіком загального тренду будуємо точкові прогнози на час $t_{45} = 45$ і $t_{46} = 46$ (рис. 2.4).

б) Для виконання інтервального прогнозу на час $t_{45} = 45$ знаходимо за таблицею критичних точок розподілу Стьюдента для двосторонньої критичної області (додаток 2) значення коефіцієнта довіри $t(n;\alpha)$:

$$n = m \cdot M - e = 3 \cdot 12 - 2 = 34;$$
$$\alpha = 1 - \gamma = 1 - 0{,}95 = 0{,}05;$$

Таблиця додатку 2 не містить значень $t(n;\alpha)$ для n=34, тому значення $t(34; 0{,}05) = 2{,}035$ обчислюємо шляхом лінійної інтерполяції для заданих у таблиці додатку 2 значень $t(28; 0{,}05) = 2{,}05$ і $t(40; 0{,}05) = 2{,}02$.

Обчислюємо за формулою (2.7) коефіцієнт прогнозу μ_{45}:

$$H_{45} = t_{45} - t_{39} = 45 - 39 = 6;$$
$$\mu_{45} = \sqrt{\frac{3 \cdot 12 \cdot (3 \cdot 12 + 2)^2 + 3 \cdot (3 \cdot 12 + 2 \cdot 6)^2}{3 \cdot 12 \cdot (3 \cdot 12 + 1) \cdot (3 \cdot 12 + 2)}} = \sqrt{\frac{51984 + 6912}{50616}} \approx 1{,}0787.$$

За формулами (2.5), (2.6) обчислюємо ліву $y_{\text{л}}$ та праву $y_{\text{п}}$ межі надійного інтервалу:

$$y_{\text{л}} = -49{,}3140 - 2{,}035 \times 0{,}4281 \times 1{,}0787 \approx -50{,}2537;$$
$$y_{\text{п}} = -49{,}3140 + 2{,}035 \times 0{,}4281 \times 1{,}0787 \approx -48{,}3743\,.$$

в) Аналогічно виконуємо інтервальний прогноз на час $t_{46} = 46$:

$$\alpha = 1 - \gamma = 1 - 0{,}90 = 0{,}10;$$
$$t(34; 0{,}10) = 1{,}690;$$
$$H_{46} = t_{46} - t_{39} = 46 - 39 = 7;$$
$$\mu_{46} = \sqrt{\frac{3 \cdot 12 \cdot (3 \cdot 12 + 2)^2 + 3 \cdot (3 \cdot 12 + 2 \cdot 7)^2}{3 \cdot 12 \cdot (3 \cdot 12 + 1) \cdot (3 \cdot 12 + 2)}} = \sqrt{\frac{51984 + 7500}{50616}} \approx 1{,}0841.$$
$$y_{\text{л}} = -51{,}0051 - 1{,}69 \cdot 0{,}4281 \cdot 1{,}0841 \approx -51{,}7894\,;$$
$$y_{\text{п}} = -51{,}0051 + 1{,}69 \cdot 0{,}4281 \cdot 1{,}0841 \approx -50{,}2208\,.$$

Будуємо у збільшеному масштабі графіки точкового та інтервального прогнозів на час $t_{45} = 45$ і $t_{46} = 46$ (рис. 2.5).

Таблиця 2.1

Розрахункова таблиця

	t_i	y_i	$t_i \cdot y_i$	t_i^2
	1	2	3	4
×	4	12,9701	51,8804	16,0000
	5	10,9708	54,8540	25,0000
	6	6,9571	41,7426	36,0000
	7	3,0934	21,6538	49,0000
	8	−0,4266	−3,4128	64,0000
	9	−1,5967	−14,3703	81,0000
	10	−3,3762	−33,7620	100,0000
	11	−1,7688	−19,4568	121,0000
	12	−1,0231	−12,2772	144,0000
	13	0,0159	0,2067	169,0000
	14	0,9748	13,6472	196,0000
	15	−1,1037	−16,5555	225,0000
	16	−1,4699	−23,5184	256,0000
	17	−5,7071	−97,0207	289,0000
	18	−9,9391	−178,9038	324,0000
	19	−13,4611	−255,7609	361,0000
	20	−15,7381	−314,7620	400,0000
	21	−17,4264	−365,9544	441,0000
	22	−19,3300	−425,2600	484,0000
	23	−18,1732	−417,9836	529,0000
	24	−18,0289	−432,6936	576,0000
	25	−16,0605	−401,5125	625,0000
	26	−15,0614	−391,5964	676,0000
	27	−16,3546	−441,5742	729,0000
	28	−17,9664	−503,0592	784,0000
	29	−21,2830	−617,2070	841,0000
	30	−25,3754	−761,2620	900,0000
	31	−29,3811	−910,8141	961,0000
	32	−31,9349	−1021,9168	1024,0000
	33	−33,8429	−1116,8157	1089,0000
	34	−35,2330	−1197,9220	1156,0000
	35	−35,2884	−1235,0940	1225,0000
	36	−34,1639	−1229,9004	1296,0000
	37	−32,1511	−1189,5907	1369,0000
	38	−31,2873	−1188,9174	1444,0000
	39	−33,0277	−1288,0803	1521,0000
Σ	774,0000	−501,9984	−15922,9700	20526,0000

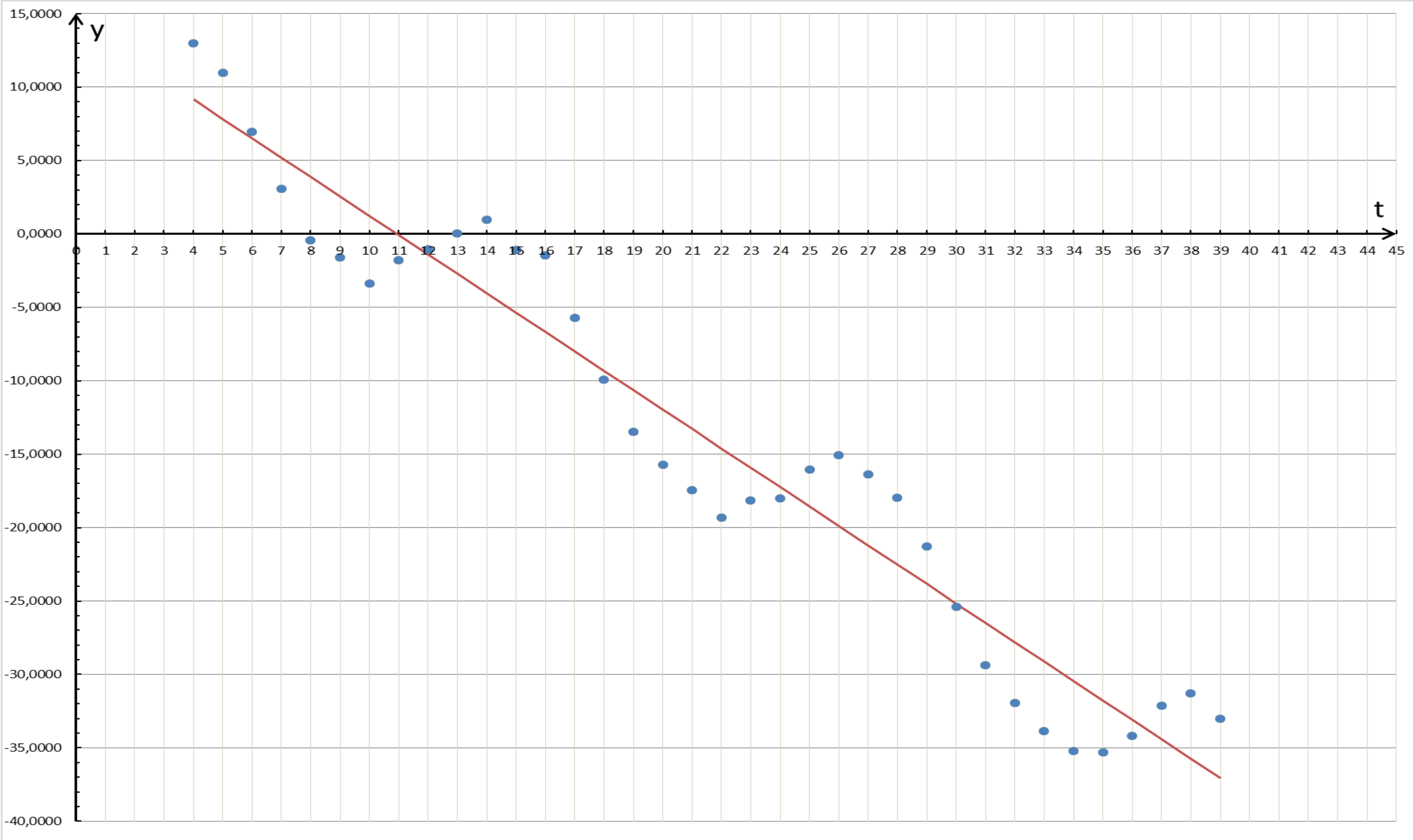

Рис. 2.3. Графік виправленого загального тренду на кореляційному полі нового ряду

Таблиця 2.2

Розрахункова таблиця

k	Прирости коливань				$z_k^i - z_k$			$z_k^i - z_k \; ^2$		
	за один період			загальні						
	z_k^1	z_k^2	z_k^3	z_k	$z_k^1 - z_k$	$z_k^2 - z_k$	$z_k^3 - z_k$	$z_k^1 - z_k \; ^2$	$z_k^2 - z_k \; ^2$	$z_k^3 - z_k \; ^2$
1	2	3	4	5	6	7	8	9	10	11
1	3,8057	5,2117	4,5612	4,5262	0,7205	0,6855	0,0350	0,5191	0,4699	0,0012
2	3,1269	2,2950	2,5651	2,6623	0,4646	0,3673	0,0972	0,2159	0,1349	0,0094
3	0,4337	–0,6165	–0,2068	–0,1299	0,5636	0,4866	0,0769	0,3176	0,2368	0,0059
4	–2,1095	–2,8180	–2,8920	–2,6065	0,4970	0,2115	0,2855	0,2470	0,0447	0,0815
5	–4,3090	–3,7745	–4,1253	–4,0696	0,2394	0,2951	0,0557	0,0573	0,0871	0,0031
6	–4,1586	–4,1423	–4,7128	–4,3379	0,1793	0,1956	0,3749	0,0321	0,0383	0,1406
7	–4,6176	–4,7254	–4,7824	–4,7085	0,0909	0,0169	0,0739	0,0083	0,0003	0,0055
8	–1,6897	–2,2481	–3,5173	–2,4850	0,7953	0,2369	1,0323	0,6325	0,0561	1,0656
9	0,3765	–0,7833	–1,0723	–0,4930	0,8695	0,2903	0,5793	0,7560	0,0843	0,3356
10	2,7360	2,5056	2,2610	2,5009	0,2351	0,0047	0,2399	0,0553	0,0000	0,0576
11	5,0154	4,8252	4,4453	4,7620	0,2534	0,0632	0,3167	0,0642	0,0040	0,1003
12	4,2574	4,8525	4,0254	4,3784	0,1210	0,4741	0,3530	0,0146	0,2248	0,1246
Σ	**2,8672**	**0,5819**	**–3,4509**	**–0,0006**	×	×	×	**2,9200**	**1,3811**	**1,9309**
$z^i = \max\limits_k z_k^i - z_k$					**0,8695**	**0,6855**	**1,0323**	×	×	×

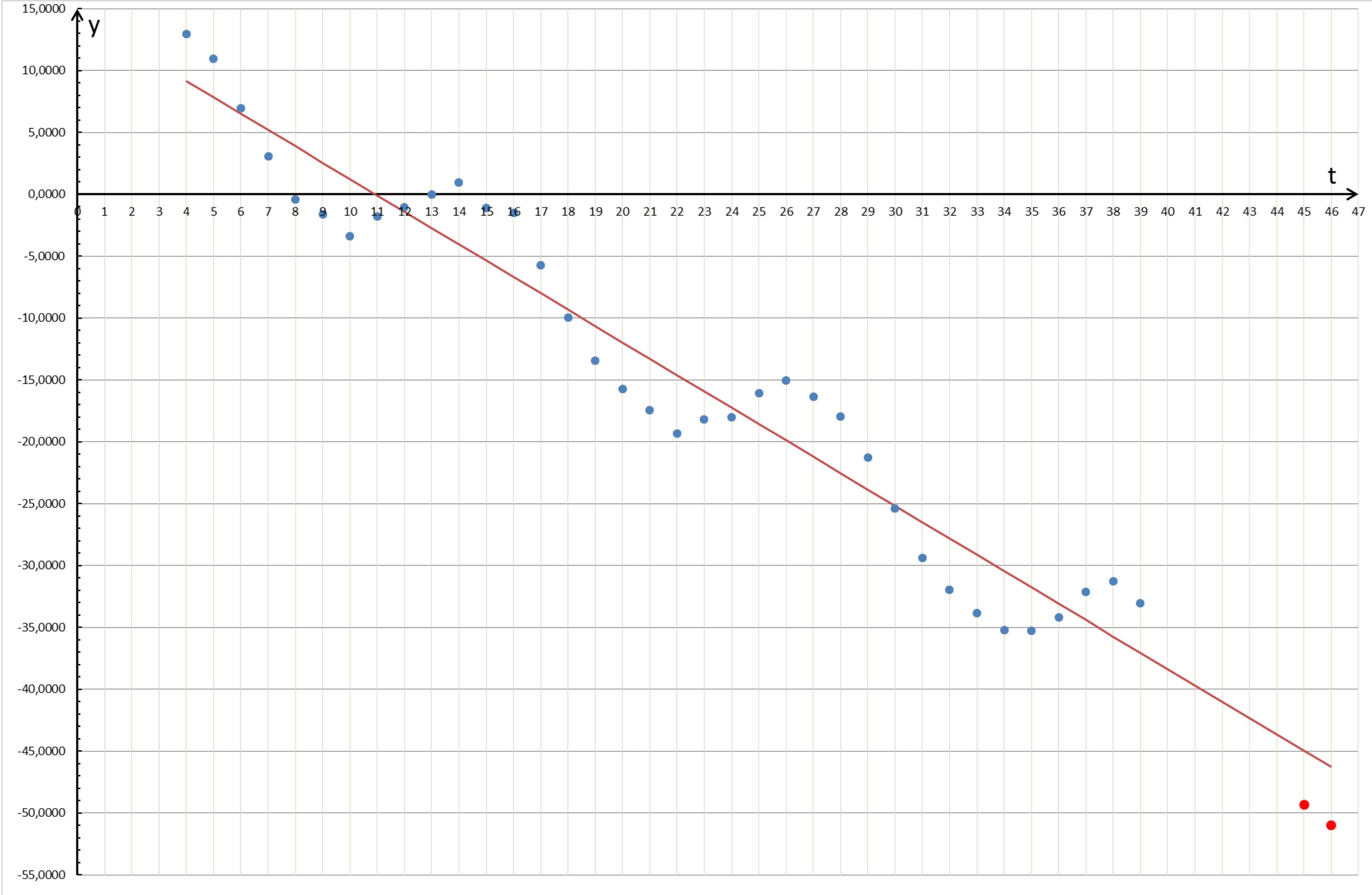

Рис. 2.4. Графік виправленого загального тренду на кореляційному полі нового ряду та точкові прогнози на час $t_{45} = 45$ і $t_{46} = 46$

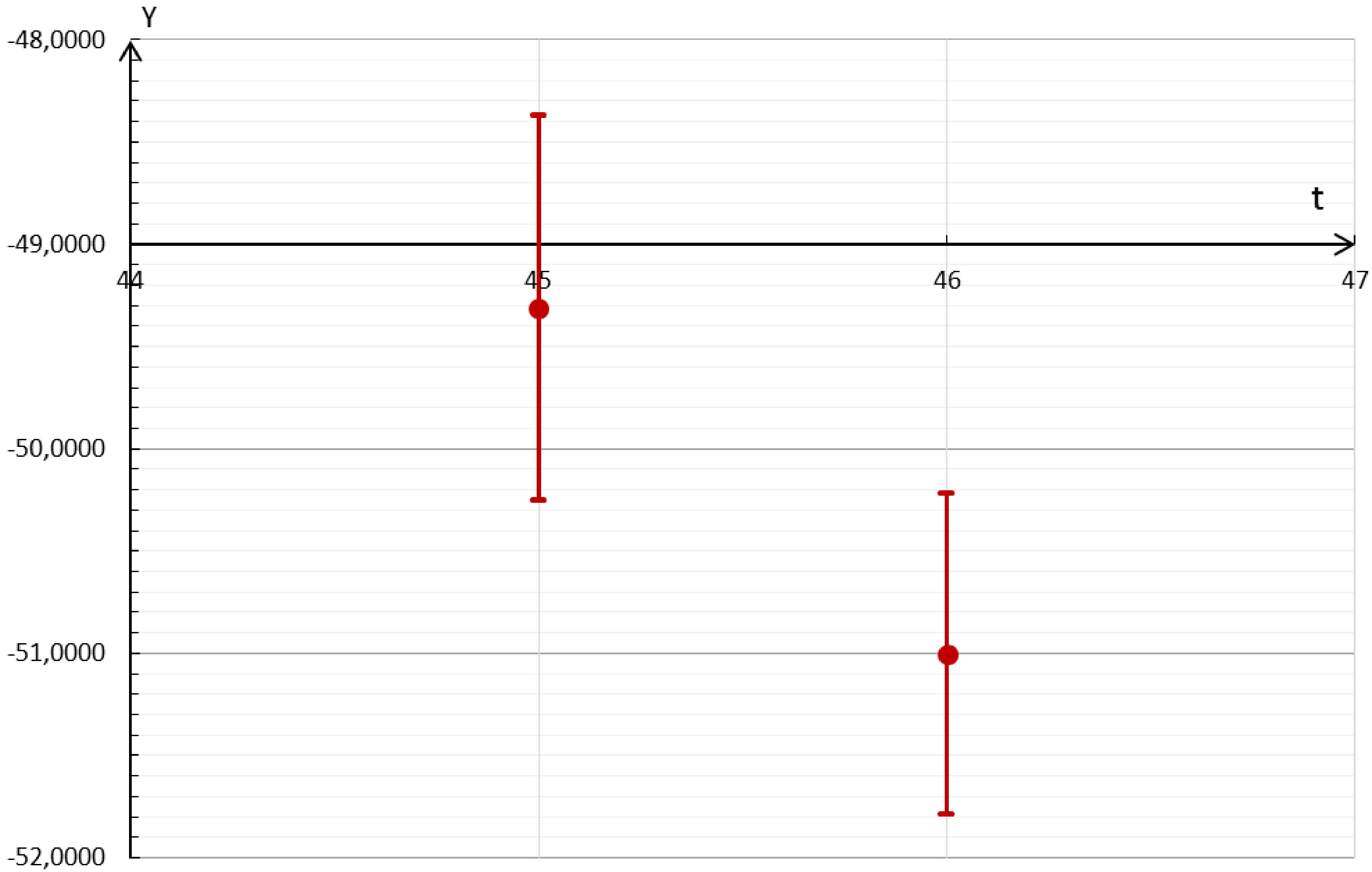

Рис. 2.5. Точкові та інтервальні прогнози на час $t_{45} = 45$ і $t_{46} = 46$

6. За даними графи 5 табл. 2.2 будуємо лінію приростів коливань:

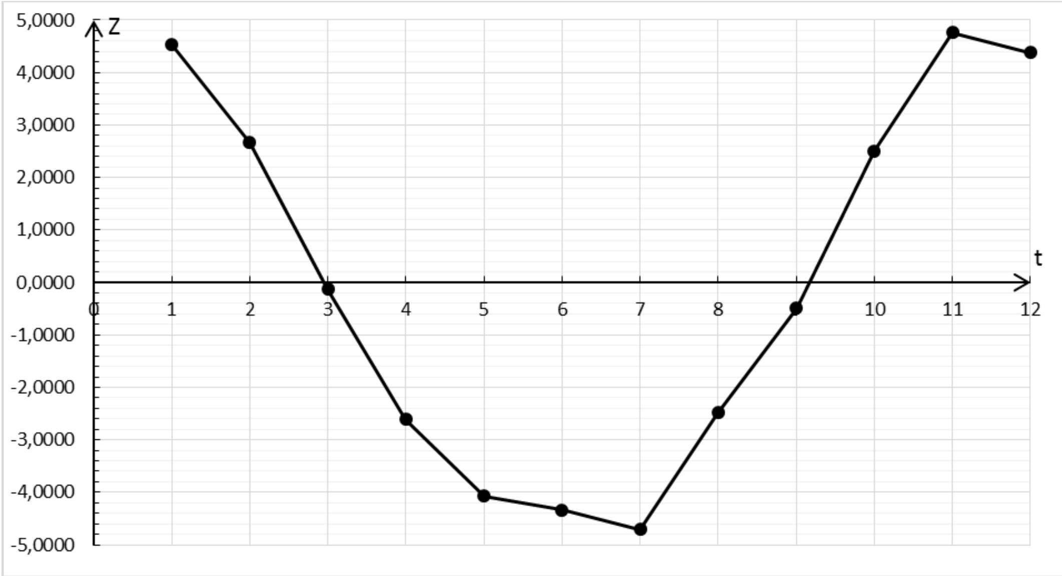

Рис. 2.6. Лінія приростів коливань абсолютного приросту фізичного обсягу експорту

7. За результатами проведеного дослідження можна зробити висновки:

- абсолютний приріст фізичного обсягу експорту має чітко виражені квазіперіодичні коливання з періодом $T = 12$ і загальним трендом $\varphi(t)$ = 14,4464 – 1,3205·t;
- якщо для часу $t > t_{44} = 44$ не зміниться вид і характер загального тренду та характер коливань, то можна очікувати, що абсолютний приріст фізичного обсягу експорту повинен знаходитись у межах:

а) від $-50{,}2537$ до $-48{,}3743$ з надійністю $\gamma = 0{,}95$ на час $t_{45} = 45$;

б) від $-51{,}7894$ до $-50{,}2208$ з надійністю $\gamma = 0{,}90$ на час $t_{46} = 46$;

- за результатами проведених обчислень і візуального аналізу лінії приростів коливань маємо: в кожному періоді максимальні відхилення абсолютного приросту фізичного обсягу експорту від загального тренду в середньому досягають +4,7620 та – 4,7085 (графа 5 табл. 2.2).

2.6. Питання для самоконтролю

1. Пояснити необхідність уточнення параметрів загального тренду для вихідного динамічного ряду.
2. Пояснити причину можливого викривлення загального тренду для вихідного часового ряду.
3. Навести процедуру:

а) визначення типу коливань;

б) виконання точкового прогнозу;

в) виконання інтервального прогнозу;

г) побудови лінії приростів коливань.

4. Пояснити статистичний зміст надійності γ.
5. Пояснити необхідність використання коефіцієнту прогнозу μ.

ЛАБОРАТОРНА РОБОТА № 3

Тема. Аналіз диференціації та концентрації розподілів

Мета роботи: навчитись визначати ступінь диференціації та концентрації розподілів.

Вихідні дані: формуються виконавцем самостійно відповідно до п.3.1 основної теоретичної інформації.

Основна теоретична інформація

3.1 Формування вихідної статистичної сукупності

Вихідна статистична сукупність формується виконавцем самостійно за допомогою комп'ютерної програми, що генерує псевдовипадкові числа для заданого закону та параметрів розподілу. При цьому обираємо обсяг сукупності n=60, а параметри розподілу обчислюємо за такими формулами:

$$m_i = 10+i+k;\ \sigma_i=0{,}25 \cdot m_i,$$

де i – порядковий номер виконавця за списком у журналі академічної групи за поточний семестр, число $k \in \mathrm{N}$ задає викладач. Після обчислення чисел m_i та σ_i виконати такі операції:

1. Увімкнути комп'ютер.
2. За допомогою маніпулятора **мышь** запустити табличний процесор Microsoft Excel.
3. Увійти в меню **Сервис** та обрати рядок **Анализ данных**.
4. Після появи вікна **Анализ данных** обрати рядок **Генерация случайных чисел** та натиснути кнопку ОК. На екрані монітора має з'явитись вікно з назвою **Генерация случайных чисел**.
5. У полі **Распределение** випадаючого меню обрати рядок **Нормальное**.
6. У полях **Число переменных**, **Число случайных чисел**, **Среднее** та **Стандартное отклонение** записати числа відповідно 1, n, m_i та σ_i.
7. У полі **Параметры вывода** обрати пункт **Новый рабочий лист** та натиснути кнопку ОК.
8. На екрані монітора у стовпці А має з'явитися 60 чисел, які слід округлити до одного десяткового знаку і розташувати у неспадному порядку. Одержані після округлення і упорядкування числа і будуть вихідною сукупністю.

3.2 Показники диференціації

Як відомо, структуру статистичної сукупності можна вивчати за допомогою побудови та подальшого аналізу дискретного чи інтервального варіаційного рядів, а також обчисленням середньої арифметичної та структурних середніх варіаційного ряду, до яких належать мода і медіана. Для більш детального вивчення структури статистичної сукупності та (або) відповідного варіаційного ряду, а також для кількісної оцінки ступеня нерівномірності розподілу ознаки в сукупності використовуються показники диференціації, до яких належать квантилі та коефіцієнти диференціації.

Квантилі – це такі числа, зокрема, значення варіаційної ознаки (варіанти), які ділять усю ранжировану (впорядковану) сукупність на рівні за обсягом частини. Найбільш поширеними квантилями є квартилі, квінтилі, децилі та процентилі, які відповідно ділять упорядковану сукупність на 4, 5, 10 та 100 рівних за обсягом частин. Очевидно, що кількість квартилів дорівнює 3, квінтилів – 4, децилів – 9, процентилів – 99.

Для дискретного варіаційного ряду розподілу квантилі відшукуються безпосередньо за визначенням.

Для інтервального варіаційного ряду розподілу частот (і.в.р. f) квантилі обчислюються за відповідними формулами. Зокрема, децилі обчислюються за формулою

$$De_i = x_{Dei} + h_{Dei}\frac{0{,}1i\sum f_i - S_{Dei-1}}{f_{Dei}}, \qquad (3.1)$$

де i – порядковий номер дециля;

$x_{De\,i}$ – нижня межа інтервалу i-го дециля, яким є перший з інтервалів, для яких накопичена частота S_j перевищує число $0{,}1i\sum f_i$;

j – номер інтервалу i-го дециля;

$h_{De\,i}$ – ширина інтервалу, де розташований i-й дециль;

$0{,}1\sum f_i$ – одна десята обсягу сукупності;

$S_{De\,i-1}$ – сума частот до інтервалу, де розташований i-й дециль;

$f_{De\,i}$ – частота інтервалу, де розташований i-й дециль.

Після знаходження квантилів за необхідності можна обчислити відповідні квантильні та фондові квантильні коефіцієнти диференціації.

Квантильний коефіцієнт диференціації – це відношення останнього квантиля до першого. Він показує, у скільки разів мінімальне значення відповідної частини (для квартилів ця частина дорівнює 25 %) найбільших значень ознаки перевищує максимальне значення такої ж частини найменших значень ознаки. Для квінтилів ця частина дорівнює 20 %, для децилів – 10 %, а для процентилів – 1 %. Очевидно, що квантильний коефіцієнт диференціації недостатньо адекватно відображує ступінь диференціації ознаки, оскільки враховує тільки мінімальне та максимальне значення певної частини відповідно найбільших і найменших значень ознаки, а не весь обсяг відповідних значень ознаки.

Більш точно ступінь диференціації ознаки можна оцінити, обчисливши **фондовий квантильний коефіцієнт диференціації**, який визначається як відношення середньої всіх значень ознаки останнього квантильного інтервалу до середньої всіх значень ознаки першого квантильного інтервалу.

У статистичній практиці найширше використовуються децильні коефіцієнти, які характеризують відношення 10 % найбільших значень ознаки до 10 % її найменших значень. Так, децильний і фондовий децильний коефіцієнти диференціації обчислюються за такими формулами:

$$K_{De} = \frac{De_9}{De_1}, \tag{3.2}$$

де K_{De} – децильний коефіцієнт диференціації;

De_9 – значення дев'ятого дециля;

De_1 – значення першого дециля.

$$K_f = \frac{\bar{x}_{10}}{\bar{x}_1}, \tag{3.3}$$

де K_f – фондовий децильний коефіцієнт диференціації;

$\bar{x}_{10}$ – середнє значення 10% найбільших значень ознаки;

$\bar{x}_1$ – середнє значення 10% найменших значень ознаки.

Необхідно зауважити, що якщо вихідні дані наведено у вигляді звичайного варіаційного ряду (з.в.р.), то ступінь диференціації ознаки слід оцінювати за допомогою фондового децильного коефіцієнта диференціації K_f. Децильний коефіцієнт диференціації K_{De} варто використовувати у випадку, коли вихідні дані наведено у вигляді і.в.р. f.

3.3 Показники концентрації

Вивчення структури сукупностей та (або) відповідних варіаційних рядів, а також кількісна оцінка ступеня нерівномірності розподілу ознаки в сукупності може здійснюватись також за допомогою коефіцієнту локалізації та коефіцієнтів концентрації.

Коефіцієнт концентрації – це узагальнююча для сукупності характеристика відхилення розподілу від рівномірного. Існують декілька коефіцієнтів концентрації, кожен з яких обчислюється за своєю формулою. Розглянемо найбільш поширені з них.

Коефіцієнт концентрації Джини обчислюється за формулою:

$$G = \frac{1}{m-1}\sum_{i=2}^{m} B_i, \tag{3.4}$$

де m – число груп, на які розподілені всі елементи сукупності, що вивчається (число інтервалів відповідного інтервального варіаційного ряду). Значення B_i обчислюються за формулою:

$$B_i = \left|(i-1)q_i - i\, q_{i-1}\right|, \tag{3.5}$$

де q_i – накопичена частка величини Y, концентрація якої вимірюється, для i -ї групи (i - го інтервалу).

Значення q_i обчислюється за формулою:

$$q_i = \sum_{k=1}^{i} d_k = \sum_{k=1}^{i} \frac{y_k}{\sum\limits_{k=1}^{m} y_k} = \frac{\sum\limits_{k=1}^{i} y_k}{\sum\limits_{k=1}^{m} y_k}, \tag{3.6}$$

де d_k (y_k) – частка (значення) величини Y для k-ї групи (k - го інтервалу).

Якщо припустити, що в кожному інтервалі значення величини Y розподілені хоча б наближено рівномірно (інтервали кількісно однорідні), то очевидно, що значення y_i величини Y для кожного i - го інтервалу являє собою добуток середини x_i i – го інтервалу значень величини Y на число цих значень f_i : $y_i = x_i \cdot f_i$, де x_i – середина i - го інтервалу; f_i – частота i - го інтервалу (число значень величини Y у i - му інтервалі).

Тоді

$$d_i = \frac{x_i \ f_i}{\sum\limits_{k=1}^{m} x_k \ f_k}, \quad q_i = \frac{\sum\limits_{k=1}^{i} x_k \ f_k}{\sum\limits_{k=1}^{m} x_k \ f_k}.$$

Зауважимо, що деякі або всі групи можуть містити в собі тільки одну одиницю сукупності (одне значення величини Y). В останньому випадку $m = n$.

Можна показати, що $G \in [0;1]$. При цьому $G = 0$, якщо величина Y розподілена по групам (по інтервалам) абсолютно рівномірно, тобто, кожна група (інтервал) містить рівні між собою значення y_i величини Y . Якщо величина Y має максимально нерівномірний розподіл по групам (інтервалам), тобто, весь обсяг Y сконцентрований в одній (байдуже якій) групі, то $G = 1$.

Очевидно, що чим більше значення G, тим вище ступінь концентрації величини Y і навпаки.

Будемо вважати концентрацію:

- високою (значною), якщо $G \in (0{,}2 ; 1]$;
- помірною (середньою), якщо $G \in (0{,}1; 0{,}2]$;
- низькою (слабкою), якщо ; $G \in (0 ; 0{,}1]$
- відсутньою, якщо G=0.

Зауважимо, що значення G не залежить від числа m груп (інтервалів). Цей факт в залежності від обставин і мети дослідження може вважатись як перевагою, так і недоліком коефіцієнта G.

Так, наприклад, якщо треба порівняти ступінь концентрації *Y* в сукупностях з різними значеннями числа *m*, то незалежність *G* від *m* є його перевагою.

Якщо ж, наприклад, треба порівняти ступінь монополізації певного ринку в різних регіонах, то незалежність *G* від *m* слід вважати його недоліком, оскільки навіть при $G = 0$ або $G \approx 0$ ступінь монополізації вважається тим вищим, чим менше число *m*.

Коефіцієнт концентрації Герфендаля вважається найбільш поширеним показником, що характеризує ступінь концентрації величини *Y* і, відповідно, нерівномірність її розподілу по групам (інтервалам). Він обчислюється за формулами, в яких всі позначення і їх зміст ті ж самі, що і для коефіцієнта *G*:

$$H = \sum_{i=1}^{m} d_i^2 \qquad (3.7)$$

або, що теж саме,

$$H = \sum_{i=1}^{m} \left(\frac{x_i \ f_i}{\sum_{i=1}^{m} x_i \ f_i} \right)^2 \qquad (3.8)$$

або, що теж саме,

$$H = \frac{\sum_{i=1}^{m} (x_i \times f_i)^2}{\left(\sum_{i=1}^{m} x_i \times f_i \right)^2}. \qquad (3.9)$$

Зауважимо, що деякі або всі групи можуть мати тільки одну одиницю сукупності. В останньому випадку $m = n$, $f_k = 1$.

Можна показати, що $H \in [A; 1]$, де $A = \frac{1}{m}$. При цьому $H = A$, якщо величина *Y* розподілена порівну між групами (інтервалами) і $H = 1$, якщо весь обсяг величини *Y* зосереджений в одній (довільній) групі.

Як і для коефіцієнта *G* значення *H* збільшується із зростанням ступеня концентрації *Y* і навпаки.

Будемо вважати концентрацію:

- високою (значною), якщо $H \in (0{,}8A + 0{,}2; 1]$;
- помірною (середньою), якщо $H \in (0{,}9A + 0{,}1; 0{,}8A + 0{,}2]$;
- низькою (слабкою), якщо $H \in (A; 0{,}9A + 0{,}1]$;
- відсутньою, якщо $H = A$.

Зауважимо, що мінімальне значення *H* залежить від числа *m* груп (інтервалів): чим більше *m*, тим менше $H_{\min} = A$. Це означає, що при порівняні ступеня концентрації однієї і тієї ж величини Y для сукупностей з різною кількістю груп (інтервалів) навіть при абсолютно рівномірному розподілі значень y_i величини Y між групами в кожній сукупності більш висококонцентрованою буде вважатись сукупність з меншим числом m. Цей факт в залежності від ситуації і мети дослідження може вважатись як недоліком, так і перевагою коефіцієнта H.

Коефіцієнт концентрації Лоренца обчислюється за формулою:

$$K = \frac{1}{2}\sum_{i=1}^{m}\left|w_i - d_i\right|, \tag{3.10}$$

де $d_i = \dfrac{f_i}{\sum\limits_{i=1}^{m} f_i} = \dfrac{f_i}{n}$ – частка кількості одиниць сукупності, що вивчається, які належать i-й групі (i-му інтервалу); m – число груп (інтервалів); f_i – частота i - ї групи (i-го інтервалу), тобто, кількість одиниць сукупності, що належать i - й групі (i - му інтервалу); $n = \sum\limits_{i=1}^{m} f_i$ – загальна кількість одиниць сукупності (обсяг сукупності); $w_i = \dfrac{y_i}{\sum\limits_{i=1}^{m} y_i}$ – частка величини Y, концентрація якої вимірюється, що належить i - й групі (i - му інтервалу) у загальному обсязі Y; y_i – обсяг величини Y, що належить i - й групі (i - му інтервалу); $\sum\limits_{i=1}^{m} y_i$ – загальний обсяг величини Y у всій сукупності.

Зауважимо, що деякі або всі групи можуть мати тільки один елемент сукупності. В останньому випадку $m = n$; $f_k = 1$; $d_k = 1/n$.

Можна показати, що $K \in [0;1]$. Зокрема, $K = 0$, якщо величина Y розподілена по групам (інтервалам) рівномірно в тому розумінні, що на кожну одиницю сукупності припадає одне і те ж значення Y. Очевидно, що при цьому $w_i = d_i$ для всіх $i = \overline{1,m}$. Якщо величина Y розподілена по групам (інтервалам) максимально нерівномірно, тобто, весь загальний обсяг Y сконцентрований в одній l - й групі (інтервалі), то

$$K_{\max} = \frac{1 + \sum\limits_{k=1, k \neq l}^{m} d_k - d_l}{2}.$$

При цьому, оскільки $\sum\limits_{k=1}^{m} d_k = 1$, то з останньої рівності витікає, що значення $K_{\max}$ тим ближче до 1, чим менше значення d_l. Зокрема, якщо $m = n$, то

$$K_{\max} = \frac{n-1}{n}$$

і тим ближче до 1, чим більше *n*.

Як і для вищерозглянутих коефіцієнтів *G* і *H* значення *K* зростає зі збільшенням ступеня концентрації *Y* і навпаки.

Будемо вважати концентрацію:

- високою (значною), якщо $K \in [0{,}2\,;1]$;
- помірною (середньою), якщо $K \in [0{,}1;\,0{,}2)$;
- низькою (слабкою), якщо $K \in (0\,;\,0{,}1)$;
- відсутньою, якщо $K = 0$.

Коефіцієнт концентрації CR_m

У деяких країнах за показник концентрації береться відношення, виражене у відсотках, обсягу *m* найбільших значень величини *Y* , концентрація якої оцінюється, до загального обсягу *Y* у всій досліджуваній сукупності. При цьому в різних країнах множина можливих значень *m* може бути різною. Так, у США і Франції число *m* береться одним з такого ряду: 4, 8, 20, 50, 100. У Канаді, Німеччині, Англії число *m* береться одним з такого ряду: 3, 6, 10.

Даний показник називається коефіцієнтом концентрації CR_m і обчислюється за формулою:

$$CR_m = \frac{\sum_{k=1}^{m} y_k}{\sum_{i=1}^{n} y_i} \times 100\%, \tag{3.11}$$

де y_i – *i*-те значення величини *Y*; *n* – обсяг сукупності (кількість всіх значень y_i); y_k – найбільші *m* значень величини *Y*.

В залежності від значень коефіцієнтів CR_3 та CR_4 прийнято виділяти три ступеня концентрації:

- високий, якщо $CR_3 \in (70\%\,;100\%]$ або $CR_4 \in (80\%\,;100\%]$;
- помірний, якщо $CR_3 \in (45\%\,;70\%]$ або $CR_4 \in (45\%\,;80\%]$;
- низький, якщо $CR_3 \in [0\%\,;\,45\%]$ або $CR_4 \in [0\%\,;\,45\%]$.

Всі вищеназвані і деякі інші узагальнюючі показники концентрації з 1977 р. регулярно обчислюються і публікуються західнонімецьким Федеральним бюро статистики. З 1968 р. коефіцієнт CR_m використовується Департаментом юстиції США для оцінки ступеня концентрації ринків. Зокрема, якщо CR_4 перевищував 75%, то вводились певні антимонопольні заходи (наприклад, обмеження на злиття підприємств), оскільки такий ринок розглядався як об'єкт монопольної практики. Регулярна публікація значень коефіцієнта *H* по основним галузевим ринкам США проводиться Бюро перепису з 1982 р.

Коефіцієнт локалізації також може слугувати кількісною оцінкою нерівномірності розподілу значень величини Y між окремими групами (інтервалами) і обчислюється за формулою:

$$L_i = \frac{w_i}{d_i}, \qquad (3.12)$$

де всі позначення і їх зміст такі ж, як і для коефіцієнта концентрації Лоренца K.

Якщо розподіл значень величини Y по групам (інтервалам) рівномірний, то $d_i = w_i$ і відповідно всі значення $L_i = 1$. У разі підвищеної концентрації обсягу y_i величини Y в i-й групі $L_i > 1$. Очевидно, що при цьому ступінь концентрації зростає із збільшенням значення L_i. Навпаки, якщо $L_i < 1$, то на дану групу припадає незначний обсяг величини Y.

3.4. Приклад постановки і розв’язування типової задачі

Постановка задачі

Нижче наведено умовні дані про обсяги митних платежів від імпорту однотипних товарів різними суб’єктами ЗЕД, тис. грн.:

10,9; 11,2; 11,3; 11,4; 11,5; 11,6; 12,3; 13,0; 13,3; 13,4; 14,0; 14,1; 14,7; 14,9; 15,1; 15,7; 15,8; 17,0; 17,1; 17,2; 17,4; 17,5; 18,1; 18,8; 20,9; 21,5; 21,6; 21,8; 22,2; 23,0; 26,9; 27,2; 30,1; 30,3; 32,1; 32,4; 38,6; 45,5; 51,8; 54,3; 61,0; 62,1; 73,0; 81,3; 108,0; 110,9; 140,1; 180,7; 187,3; 228,7.

За наведеними даними необхідно:

1. Побудувати інтервальний варіаційний ряд частот (і.в.р. f), утворивши не менше 5-ти можливо нерівних бажано (за можливістю) хоча б наближено кількісно однорідних інтервалів (груп). При цьому кількісно однорідною вважається група (інтервал), в якій варіанти розподілені хоча б наближено рівномірно. Варіанти всередині інтервалу (групи) розподілені рівномірно, якщо відстані між кожною парою сусідніх варіант в даному i-му інтервалі рівні: $\left|x_{i\,j} - x_{i\,j+1}\right| = \text{const}$, i – номер інтервалу, $j = \overline{1, f_i - 1}$, f_i – частота i-го інтервалу.

2. Обчислити нижчезазначені показники диференціації та концентрації та зробити відповідні висновки: а) децильний та фондовий децильний коефіцієнти диференціації; б) коефіцієнти концентрації Джині, Герфендаля, Лоренца, CR_3; в) коефіцієнт локалізації.

Розв’язування задачі

1. За результатами аналізу вихідного звичайного варіаційного ряду (з. в. р.) доцільно побудувати інтервальний варіаційний ряд частот у вигляді, наведеному у перших трьох рядках табл. 3.1. При цьому будемо вважати, що кожна група (інтервал) табл. 3.1. хоча б наближено кількісно однорідна.

Таблиця 3.1

Розподіл митих платежів за їх обсягом

i	1	2	3	4	5	6	7	Разом
Обсяг митних платежів, тис. грн. $[x_i'; x_i'')$	10; 12	12;15	15; 20	20; 30	30; 50	50; 100	100; 250	
Кількість суб'-єктів ЗЕД (f_i)	6	8	10	8	6	6	6	50
Накопичена частота (S_i)	6	14	24	32	38	44	50	×

2. <u>Децильний та фондовий децильний коефіцієнти диференціації.</u>

Децильний коефіцієнт диференціації обчислюється за формулою (3.2). Значення першого De_1 і дев'ятого De_9 децилів обчислюються за формулою (3.1) за даними побудованого у табл. 3.1 і.в.р. *f*.

При цьому інтервалом 1-го дециля буде перший інтервал і.в.р. *f*, оскільки він є першим з інтервалів, для яких виконується умова $S_j > 0{,}1 \cdot i \cdot n$, де *j* – номер інтервалу, *i* – номер дециля, *n* – обсяг сукупності: $S_1 = 6 > 0{,}1 \cdot 1 \cdot 50 = 5$.

Інтервалом 9-го дециля буде, очевидно, сьомий інтервал і.в.р. *f*, оскільки він є першим з інтервалів, для яких виконується вищенаведена умова: $S_7 = 50 > 0{,}1 \cdot 9 \cdot 50 = 45$.

Тоді

$$De_1 = x_{De_1} + h_{De_1} \frac{0{,}1\sum f_i - S_{De_{1-1}}}{f_{De_1}} = 10 + 2 \times \frac{0{,}1 \times 50 - 0}{6} \approx 11{,}67 \text{ тис. грн.;}$$

$$De_9 = x_{De_9} + h_{De_9} \frac{0{,}1 \times 9\sum f_i - S_{De_{9-1}}}{f_{De_9}} = 100 + 150 \times \frac{0{,}1 \times 9 \times 50 - 44}{6} =$$

$$= 125{,}00 \text{ тис. грн.}$$

і децильний коефіцієнт диференціації дорівнює:

$$K_{De} = 125{,}00 \,/\, 11{,}67 \approx 11.$$

Фондовий децильний коефіцієнт диференціації обчислюється за формулою (3.3) за даними вихідного з.в.р. Очевидно, що 10% від обсягу сукупності *n*=50 дорівнюють $0{,}1 \cdot 50 = 5$.

Обчислюємо середню суму найменших (тобто, перших) п’яти митних платежів:

$$\bar{x}_1 = \frac{10{,}9 + 11{,}2 + 11{,}3 + 11{,}4 + 11{,}5}{5} = 56{,}3 \,/\, 5 = 11{,}26 \text{ (тис. грн.).}$$

Обчислюємо середню суму найбільших (тобто, останніх) п’яти митних платежів:

$$\bar{x}_{10} = \frac{110{,}9 + 140{,}1 + 180{,}7 + 187{,}3 + 228{,}7}{5} = 847{,}7 \,/\, 5 = 169{,}54 \text{ (тис. грн.).}$$

Тоді фондовий децильний коефіцієнт диференціації дорівнюватиме:

$$K_f = 169{,}54 \,/\, 11{,}26 \approx 15.$$

За результатами проведених обчислень можна зробити такі висновки.

Якщо вихідні дані наведено у вигляді і.в.р. *f*, то децильний коефіцієнт варіації показує, що:

а) найбільше значення 10 % найменших митних платежів не перевищує 11,67 тис. грн.;

б) найменше значення 10 % найбільших митних платежів не менше, ніж 125,00 тис. грн.;

в) рівень диференціації обсягів митних платежів можна вважати досить високим, оскільки децильний коефіцієнт диференціації дорівнює 11, тобто мінімальне значення 10% найбільших митних платежів в 11 разів більше за максимальне значення 10% найменших митних платежів.

Якщо вихідні дані наведено у вигляді з.в.р., то фондовий децильний коефіцієнт диференціації показує, що середнє значення 10% максимальних митних платежів у 15 разів більше за середнє значення 10% мінімальних митних платежів, що свідчить про високий ступінь диференціації митних платежів.

Очевидно, що фондові квантильні коефіцієнти диференціації можна обчислювати без обчислення середніх як відношення відповідних сум, якщо ці суми мають однакову кількість доданків. Так, для нашого прикладу

$$K_f = \frac{847{,}7}{56{,}3} \approx 15 \approx \frac{\bar{x}_{10}}{\bar{x}_1}.$$

<u>Коефіцієнт концентрації Джині</u> обчислюємо за даними табл. 3.1 за формулою (3.4). Проміжні обчислення зручно організувати в табл. 3.2.

За даними таблиці 3.2 $\sum_{i=2}^{7} B_i = 3{,}2336$ і $G = \frac{3{,}2336}{6} \approx 0{,}539.$

Оскільки $0{,}539 \in \left(0{,}2;1\right]$, то розподіл обсягів митних платежів слід вважати суттєво нерівномірним, а концентрацію цієї величини – високою. І дійсно, як видно с таблиці 3.2, обсяг митних платежів у 7-й групі ($x_7 \cdot f_7 = 1050$тис. грн.) становить майже половину загального обсягу всіх митних платежів (2289 тис. грн.).

Таблиця 3.2

Розрахункова таблиця

i	1	2	3	4	5	6	7	Σ
x_i	11,0	13,5	17,5	25,0	40,0	75,0	175,0	×
f_i	6	8	10	8	6	6	6	50
$x_i \times f_i$	66	108	175	200	240	450	1050	2289
$d_i = \dfrac{x_i \ f_i}{\sum_{k=1}^{7} x_k \ f_k}$	0,0288	0,0472	0,0765	0,0874	0,1048	0,1966	0,4587	1,0000
$q_i = \dfrac{\sum_{k=1}^{i} x_k \ f_k}{\sum_{k=1}^{7} x_k \ f_k}$	0,0288	0,0760	0, 1525	0,2399	0,3447	0,5413	1,0000	×
$(i-1)\, q_i$	×	0,0760	0,3050	0,7197	1,3788	2,7065	6,0000	×
$i\, q_{i-1}$	×	0,0576	0,2280	0,6100	1,1995	2,0682	3,7891	×
B_i	×	0,0184	0,0770	0,1097	0,1793	0,6383	2,2109	3,2336

Коефіцієнт концентрації Герфендаля обчислюємо за формулою (3.7) за даними таблиці 3.1, використовуючи значення d_i, вже обчислені в таблиці 3.2. Обчислення зручно організувати в табл. 3.3.

Для нашого прикладу A = 1/7 ≈ 0,143. Тоді $0{,}9A + 0{,}1 \approx 0{,}229$; $0{,}8A + 0{,}2 \approx 0{,}314$. Оскільки обчислене за даними таблиці 3.3 значення $H \approx 0{,}277 \in \left(0{,}9A + 0{,}1;\, 0{,}8A + 0{,}2\right]$, то розподіл обсягів митних платежів по групам (інтервалам) слід вважати помірно нерівномірним, а ступінь концентрації цієї величини – середньою.

Таблиця 3.3

Розрахункова таблиця

Обсяг митних платежів, тис. грн. $\left[x_i'; x_i''\right)$	10; 12	12; 15	15; 20	20; 30	30; 50	50; 100	100; 250	Разом
Кількість суб'єктів ЗЕД (f_i)	6	8	10	8	6	6	6	50
$d_i = \frac{x_i f_i}{\sum_{k=1}^{m} x_k f_k}$	0,0288	0,0472	0,0765	0,0874	0,1048	0,1966	0,4587	1,0000
d_i^2	0,0008	0,0022	0,0059	0,0076	0,0110	0,0387	0,2104	0,2766

Коефіцієнт концентрації Лоренца обчислюємо за даними таблиці 3.1 за формулою (3.10). Для зручної організації проміжних обчислень побудуємо допоміжну таблицю 3.4.

Оскільки $y_i = x_i f_i$, то $w_i = \frac{x_i f_i}{\sum_{i=1}^{m} x_i f_i}$. Якщо в правій частині останньої рівності чисельник і знаменник поділити на $n = \sum_{i=1}^{m} f_i$, то $w_i = \frac{x_i d_i}{\sum_{i=1}^{m} x_i d_i}$.

Таблиця 3.4

Розрахункова таблиця

Обсяг митних платежів, тис. грн. $\left[x_i'; x_i''\right)$	Кількість суб'єктів ЗЕД (f_i)	Частка (d_i)	$x_i \times d_i$	$\frac{x_i d_i}{\sum_{i=1}^{m} x_i d_i}$	$\left\lvert d_i - \frac{x_i d_i}{\sum_{i=1}^{m} x_i d_i} \right\rvert$
1	**2**	**3**	**4**	**5**	**6**
10; 12	6	0,12	1,32	0,0288	0,0912
12; 15	8	0,16	2,16	0,0472	0,1128
15; 20	10	0,20	3,50	0,0765	0,1235
20; 30	8	0,16	4,00	0,0874	0,0726
30; 50	6	0,12	4,80	0,1048	0,0152
50; 100	6	0,12	9,00	0,1966	0,0766
100; 250	6	0,12	21,00	0,4587	0,3387
Разом	50	1,00	45,78	1,0000	0,8306

Використовуючи формулу (3.10) та підсумок графи 6 табл. 3.4, отримаємо значення коефіцієнта концентрації Лоренца:

$$K = 0{,}5 \times 0{,}8306 = 0{,}4153.$$

Таке значення коефіцієнта концентрації Лоренца свідчить про значну концентрацію обсягів митних платежів в окремих групах інтервального варіаційного ряду, оскільки $0.4153 \in [0{,}2\,;1]$.

<u>Коефіцієнт концентрації CR_3</u> обчислюємо за вихідним з.в.р. за формулою (3.11).

$\sum_{i=1}^{n} y_i$ =10,9 + 11,2 + 11,3 + 11,4 + 11,5 + 11,6 + 12,3 + 13,0 + 13,3 + 13,4 + + 14,0 + 14,1 + 14,7 + 14,9 + 15,1 + 15,7 + 15,8 + 17,0 + 17,1 + 17,2 + 17,4 + + 17,5 + 18,1 + 18,8 + 20,9 + 21,5 + 21,6 + 21,8 + 22,2 + 23,0 + 26,9 + 27,2 + + 30,1 + 30,3 + 32,1 + 32,4 + 38,6 + 45,5 + 51,8 + 54,3 + 61,0 + 62,1 + 73,0 + + 81,3 + 108,0 + 110,9 + 140,1 + 180,7 + 187,3 + 228,7 = 2081,3.

Тоді частка трьох найбільших значень митних платежів у загальному їх обсязі становить:

$$CR_3 = \frac{228{,}7 + 187{,}3 + 180{,}7}{2081{,}3} 100\% \approx 0{,}287 \times 100\% = 28{,}7\%.$$

Оскільки сукупність складається з n = 50 значень, то частка d кількості m найбільших значень y_k від загальної кількості всіх значень y_i становить:

$$d = (m / n) \times 100\ \% = (3 / 50) \times 100\ \% = 6\ \%.$$

Таким чином, коефіцієнт концентрації CR_3 свідчить, що 28,7 % загального обсягу митних платежів припадає на 6 % суб'єктів ЗЕД, що свідчить про значну концентрацію обсягів митних платежів.

<u>Коефіцієнт локалізації</u> L_i для кожного інтервалу і.в.р. f обчислюємо за формулою (3.12). При цьому вже обчислені значення w_i та d_i беремо із табл. 3.4.

Обчислення зручно оформити в таблиці 3.5.

Таблиця 3.5

Розрахункова таблиця

i	1	2	3	4	5	6	7
w_i	0,0288	0,0472	0,0765	0,0874	0,1048	0,1966	0,4587
d_i	0,12	0,16	0,20	0,16	0,12	0,12	0,12
L_i	0,39	0,30	0,38	0,55	0,87	1,64	3,82

За даними таблиці 3.5 видно, що частка обсягу митних платежів в другій групі становить 30 % від частки кількості суб'єктів ЗЕД другої групи. В сьомій групі відповідне співвідношення становить близько 380 %.

Результати дослідження свідчать про значну нерівномірність розподілу обсягів митних платежів по суб'єктам ЗЕД.

Запитання для самоконтролю

1. Дати означення а) квантилів, б) квартилів, в) квінтилів, г) децилів, д) процентилів. Записати формули для їх обчислення та пояснити зміст позначень.

2. Як обчислюються та що характеризують
а) коефіцієнти диференціації: 1) децильний, 2) фондовий децильний;
б) коефіцієнти концентрації: 1) Джині, 2) Герфендаля, 3) Лоренца, 4) CR_m;
в) коефіцієнт локалізації.

Записати формули для їх обчислення та пояснити зміст позначень.

Додаток 1

Критичні значення $S\;M;\;\alpha$ критерія знаків; M – кількість рівнів ряду в одному періоді, $\alpha = 0{,}05$ – рівень значущості.

M	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
$S\;M;\;0{,}05$	0	0	0	1	1	1	2	2	3	3	3	4	4	5	5	5	6	6

Додаток 2

Критичні точки розподілу Стьюдента $t\;n;\;\alpha$ для двосторонньої критичної області; n – число ступенів вільності; α – рівень значущості.

n \ α	0,10	0,05
3	2,35	3,18
4	2,13	2,78
5	2,01	2,57
6	1,94	2,45
7	1,89	2,36
8	1,86	2,31
9	1,83	2,26
10	1,81	2,23
11	1,80	2,20
12	1,78	2,18
14	1,76	2,14
16	1,75	2,12
18	1,73	2,10
20	1,73	2,09
22	1,72	2,07
24	1,71	2,06
28	1,70	2,05
40	1,68	2,02
60	1,67	2,00
120	1,66	1,98
∞	1,64	1,96