В.Г.Лісняк

ВАРТІСТЬ ОБ'ЄКТІВ ЦИВІЛЬНИХ ПРАВ.

Частина ІІІ.

ПРОГНОЗИ ДОХОДІВ ДІЯЛЬНОСТІ ПІДПРИЄМСТВА.

м. Київ 2016

Лісняк Владислав Григорович
закінчив Військову космічну академію
ім. О.Ф. Можайського у м. Санкт-Петербурзі.

«ВАРТІСТЬ ОБ'ЄКТІВ ЦИВІЛЬНИХ ПРАВ»
Посібник з оцінки у трьох частинах. Частина ІІІ.
ПРОГНОЗИ ДОХОДІВ ДІЯЛЬНОСТІ ПІДПРИЄМСТВА.

- Свідоцтво про реєстрацію прав інтелектуальної власності на твір №69129 від 12 грудня 2016 року Державної служби інтелектуальної власності України.
- Заслужений експерт-оцінювач ВГО ВСЕО.

ЗМІСТ

Вартість підприємницької діяльності[1].

Підходи до визначення вартості підприємницької діяльності.

Визначити вартість підприємницької діяльності означає відповісти на питання, який за розміром дохід у часі отримає його власник. А звідси, вартість підприємства ототожнюється з розумінням грошового еквівалента **ресурсів підприємства**, що пов’язані з:

– **активами**[2] матеріальними, нематеріальними, фінансовими, інтелектуальними, праці, іншими ресурсами за витратами на їхнє створення (витратний підхід);
– **доходами** від продажу продукції виробленої на підприємстві (дохідний підхід).

В підприємництві, де метою діяльності є отримання прибутку, переважним при визначенні вартості підприємства є дохід від продажу його продукції, а не активів. За дохідним підходом вартість підприємства визначена **попитом** продукції, тобто доходами від її продажу, в той час, як за витратним підходом, з боку **пропозиції** – витратами на його створення.

Визначенню вартості підприємництва передує **аналіз його діяльності**, а саме:

1). **Аналіз фінансової звітності[3]** діяльності підприємства.

- Аналізу фінансової звітності передує її аудит, коригування за аудитом та розрахунки фінансових показників балансу підприємства з висновком щодо фінансового стану.

2). **Аналіз економічної діяльності** підприємства.

- Аналіз економічного стану підприємства окрім висновків про його стан передбачає рекомендації оцінювачів щодо прогнозних моделей обсягів виробництва продукції та доходів від її реалізації за можливостями підприємства та станом ринку.

За аналізом діяльності підприємства обґрунтовують такі показники, як:

- **обсяги економічного** (ефективного) або **оптимального** рівня виробництва[4] продукції підприємства та витрачених ресурсів з однієї сторони;
- **ефективність ринку** за ціною обсягів попиту чи пропозиції продукції з другої.

3). **Аналіз ринку** надає можливість встановити рівень його насичення та здійснити коригування прогнозів доходів підприємства за обсягами реалізації та цінами продукції.

Аналіз діяльності підприємства це визначення показників його діяльності за системою методів та процедур в цілому або окремих господарських одиниць за темпами зростання та тенденціями (трендом, сезонністю та циклічністю) змін обсягів продукції, виявлення суті та причин, що сприяють відхиленню прогнозованих показників від очікуваних.

Управління підприємством складається з двох етапів:

- **на першому** – здійснюють аналіз наявної інформації та розраховують показники, що характеризують фінансову та економічну діяльність підприємства;
- **на другому** – на підставі отриманих показників приймається управлінські рішення.

---

[1] Інколи вживається іноземний термін «**бізнес**», як його аналог, або термін підприємництво (справа, тощо).
[2] Активи не вважаються доходами від діяльності підприємства (це засоби забезпечення його діяльності), але у разі його ліквідації можуть ними бути.
[3] В деяких джерелах інформації аналіз фінансової звітності є складовою економічного аналізу. Зазвичай ми користуємося його результатами для проведення економічного аналізу та прогнозування доходів.
[4] **Економічний рівень виробництва** досягається за рівності **середніх витрат граничним**, тоді, як **оптимальний рівень виробництва** досягається за рівністю **граничних витрат та граничного доходу**.

Предметом економічного аналізу є фінансово-господарська діяльність підприємства. Існуючі методи аналізу дозволяють дослідити роботу не тільки підприємства чи процесів, але й окремих галузей, регіонів і всього господарства країни в цілому.

Об'єктом економічного аналізу є економічні явища та процеси діяльності економіки країни, як в цілому, так і окремих підприємств, галузі. Наприклад, об'єктами економічного аналізу можуть бути виробнича чи комерційна діяльність, наявність та рух ресурсів, прибуток чи дохід за обсягами реалізації виробленої продукції; ритмічність, циклічність та сезонність виробництва, а також ситуація, що склалася на ринку певних товарів.

Обов'язковою умовою товарного виробництва є забезпеченість ресурсами. Серед них природні, виробничі, інтелектуальні, фінансові, праці чи інші. Для виробництва продукції найбільш важливі ресурси **праці** (людські), **фінансові, матеріальні**, а також **основні засоби виробництва** підприємства. У виробничому процесі ресурси використовуються у певному співвідношенні, що визначено технологією. Технологія – це послідовність виробничих операцій з перетворення матеріальних та нематеріальних ресурсів у продукт підприємства, що при реалізації отримує статус товару.

Види економічного аналізу діяльності підприємства.

За часом розрізняють такі види економічного аналізу, як:

a) ретроспективний (звернений у минуле);

b) оперативний (поточний);

c) перспективний (прогнозний, звернений у майбутнє).

А). Ретроспективний аналіз включає:

1. **Фінансово-економічний** аналіз за звітний періодом (квартал, рік). Критеріями діяльності підприємства є його фінансові показники – прибутку, рентабельності, ліквідності.

2. **Управлінський** аналіз. Крім показників фінансово-економічного аналізу, значна увага надається вивченню ефективності праці, стану основних фондів, технологій і матеріальних ресурсів підприємства.

3. **Статистично-економічний** аналіз вивчає діяльність підприємства, а також регіонів, галузей і народного господарства країни в цілому за обсягами продукції упродовж років.

4. **Порівняльний** або **міжгосподарський** аналіз. Обмежується даними кількох споріднених підприємств. Орієнтуючись на показники кращих підприємств він надає можливість більш зваженої оцінки роботи підприємства, яке аналізується.

Б). Оперативний аналіз.

Оперативний аналіз здійснюють на підприємствах і його підрозділах. Інформація для аналізу береться з поточного бухгалтерського обліку. Завданням оперативного аналізу є виявлення негативних явищ у діяльності підприємства та їх вчасне усунення.

В). Перспективний аналіз (прогнозний, стратегічний).

Перспективний аналіз передує подіям виробничого процесу, прогнозує їх наслідки та ефективність. Застосовується для техніко-економічного обґрунтування (ТЕО) майбутньої діяльності підприємств, заміни засобів виробництва новітніми технікою і технологіями.

Особлива роль в економічному аналізі належить бухгалтерському обліку та статистиці. Вони забезпечують необхідною первісною інформацією та деякими методами її аналізу. В економічному аналізі застосовують сучасні засоби обробки інформації статистичними методами та економетричними моделями процесів виробництва.

Визначенню вартості підприємства передує «прогнозний аналіз» доходів за результатами економічного аналізу його діяльності. Аналіз надає можливість прогнозувати грошові надходження від виробництва продукції підприємства у часі, упродовж майбутніх періодів за моделями **виробничих** функцій, **кривих** зростання, **ковзких** середніх. Зазначені моделі будуються на статистичних даних обсягів виготовленої продукції або реалізованих товарів підприємства минулих часових періодів (кварталів чи років).

Економетричні підходи моделювання доходів підприємницької діяльності.

І. Поняття економетрики.

Економетрика це наука, що надає кількісну та якісну характеристику економічних моделей виробництва, використовуючи математичні та статистичні методи.

Вона є наслідком поєднання таких наук, як:

- економіка;
- статистика (її різновидів математичної та економічної статистики); та
- вищої математики.

Предметом дослідження в економетриці є процеси підприємницької діяльності (виробництва), які економічна теорія характеризує лише загалом. Метою такого дослідження є надання кількісної та якісної характеристики масивів статистичних значень обсягів товарів чи інших в економічних процесах, як закономірностей виробництва. Аналіз динаміки змін певних величин в економічних процесах здійснюється за моделями, що побудовані на емпіричних статистичних даних. Більшість методів, що застосовуються для вивчення закономірностей економічних процесів запозичені з математичної статистики.

Специфіка застосування методів математичної статистики в економіці в тому, що економічні показники є випадковими величинами змінними у часі. Тому в процесі аналізу економічних показників застосовують удосконалення, які в математичній статистиці не використовують. Переважна більшість статистичних моделей економічних процесів містять помилки, вплив яких можливо усунути або зменшити методами економетрики. Таким чином, **економетрика** за допомогою математичних та статистичних методів, описує **економічні** закономірності **процесів виробництва і надає їхню кількісну та якісну оцінку**.

Моделями економетрики вирішується коло завдань, які класифікують за:

1) прикладною метою; 2) рівнем ієрархії; 3) областю їх виявлення.

За прикладною метою:

а). прогноз соціально-економічних показників, що визначають стан та розвиток системи;

б). моделювання можливих варіантів соціально-економічного розвитку системи для встановлення найбільш важливих показників, що впливають на стан цієї системи.

За рівнем ієрархії:

а). задачі макрорівня (для країни в цілому);

б). задачі мезорівня (для галузей та регіонів);

в). задачі мікрорівня (для підприємств, корпорацій, тощо).

За областю виявлення проблем економічної системи:

а). ринок;

б). інвестиційна, соціальна та фінансова політика;

в). ціноутворення;

г). розподільчі відношення;

д). попит та споживання;

е). відокремлений комплекс деяких проблем;

ІІ. Економетричні моделі.

Існують такі класи економетричних **моделей**:

1) часових рядів;
2) регресійні з одним рівнянням;
3) системи одночасних рівнянь.

Розглянемо їх більш детально.

1. Моделі часових рядів.

Це моделі залежності результативної ознаки від факторної ознаки, як правило, **часу** чи інших факторних ознак, інформацію про які розглядають за плином часу.

До **моделей часових рядів**, у яких результативна ознака залежить від факторної ознаки за **часом**, відносять:

а). Модель тренда (зростаюча або спадна тенденція у часі);

б). Модель сезонності (залежність від сезонних коливань);

в). Модель тренду та сезонності (загальний вплив тренду та сезонності).

До **моделей часових рядів** в яких результативна ознака залежить від **факторних** або **результативних** ознак датованих іншими інтервалами часу відносять:

- Моделі з **розподіленим лагом**, що пояснюють варіацію результативної ознаки в залежності від попередніх значень у часі **факторних ознак**;
- Моделі **авторегресії**[5], що пояснюють варіацію результативної ознаки в залежності від попередніх значень у часі **результативних ознак**;
- Моделі **очікування**, що пояснюють варіацію результативної ознаки в залежності від майбутніх значень змінних у часі **факторних чи результативних ознак**.

[5] Модель, в якій факторні змінні утримують лагові значення результативної ознаки або **модель авторегресії ковзкої середньої** (англ. *autoregressive moving-average model, ARMA*) — одна з математичних моделей, що використовують для аналізу та прогнозування стаціонарних часових рядів в статистиці. Модель ARMA узагальнює дві більш прості моделі часових рядів — модель авторегресії (AR) та модель ковзкої середнього (MA).

Моделі часових рядів будуються на стаціонарних та нестаціонарних часових рядах. **Стаціонарні ряди** мають постійні значення у часі **середньої, дисперсії та автокореляції**, тобто ряди, що не утримують трендову чи(та) сезонну компоненти. Якщо часовий ряд містить ці складові, то він стає нестаціонарним часовим рядом.

2. Регресійні моделі з одним рівнянням.

Моделі в яких залежна результативна змінна ознака (*У*) подана як функція незалежних факторних ознак ($X_1,..X_n$): $Y = f(\beta_0, \beta_1 x_1, \ \beta_2 x_2, \dots \beta_n x_n)$, де $\beta_0, \ \beta_1, \ \beta_2, \dots \beta_n$ – параметри регресійного рівняння генеральної сукупності. Регресійні рівняння розподілені на **парні**, з одною факторною ознакою та **множинні**, з двома та більше ознаками. В залежності від аналітичного виразу рівняння поділяються на адитивні (лінійні[6] або нелінійні) та мультиплікативні. Прикладами регресійних моделей з одним рівнянням є функції **попиту** чи **пропозиції** за **ціною** або **виробничі функції**.

3. Системи одночасних рівнянь.

Системи одночасних рівнянь утримують взаємозалежні регресійні рівняння та рівняння тотожності, кожне з яких може включати факторні й результативні ознаки з інших рівнянь цієї системи. Вид та значення параметрів тотожності відомі. Регресійні рівняння, що входять в систему одночасних рівнянь мають ще назву «**поведінкових рівнянь**». Значення коефіцієнтів цих рівнянь невідомі і підлягають визначенню. Прикладом системи одночасних, так званих поведінкових рівнянь, є модель ринкової ціни. Наприклад, модель ринкової ціни будується за уявленням несхильних до ризиків індивідуумів[7] про ціни, за умовою ринкової ситуації на ринку на вільному та конкурентному ринку, тобто рівноваги обсягів товару за ціною попиту-пропозиції[8], ринковими нормою дохідності та активністю на ринку.

Недоліком **моделі визначення обсягів товарів** за ціною наведеної далі є те, що в ній не враховують вплив економічних важелів на ціну, а саме **норми дохідності** на **товари** та **коефіцієнта активності ринку**:

$$\begin{cases} Q_t^S = a_0 + a_1 \times P_t + a_2 \times P_{t-1} & \text{– обсяг пропозиції товарів за ціною;} \\ Q_t^D = b_0 + b_1 \times P_t + b_2 \times I_t & \text{– обсяг попиту товарів за ціною;} \\ Q_t^D = Q_t^S & \text{– тотожність рівноваги обсягів попиту-пропозиції товарів на ринку за ціною.} \end{cases}$$

Q_t^S – обсяг пропозиції товару на час *t*;
P_t – ціна товару на час *t*;
I_t – доход споживачів на момент часу *t*.
Q_t^D – обсяг попиту товару на час *t*;
P_{t-1} – ціна товару на час (*t-1*);

Узагальнена економічна модель ціни за обсягами попиту чи пропозицією товарів за ціною, нормою[9] дохідності та коефіцієнтом активності ринку, розроблена В.Г. Лісняком та приведена у монографії «Вартість нерухомого майна», **частина І**, розділ І-й, п.1.2.

---

[6] Моделі в яких степінь факторних ознак (многочлена) не перевищує одиниці.
[7] Тобто за моделями їхніх функцій корисності та ринкової норми дохідності.
[8] Саме умови рівноваги за ринковими обсягами попиту чи пропозиції та уявленням несхильних до ризиків індивідуумів формують **ринкові ціни**. Модель ціни потребує заміни рівнянь на **рівняння економічного походження**.
[9] Наприклад, для моделі визначення ринкової вартості – відповідно **ринкова норма дохідності**.

III. Економетричне моделювання.

Метою економетричного моделювання є уникнення обставин, що вносять помилки в результати показників статистичних методів, таких як **наявність** у рівнянні регресії:

1) асиметричності[10] зв'язків;
2) мультиколінеарності зв'язків;
3) ефекту гетероскедастичності[11] дисперсії;
4) автокореляції[12] помилок;
5) удаваній кореляції;
6) лагів[13] між двома економічними величинами (явищами).

Для описання суті процесу моделювання виділяють такі його етапи:

1. Визначення.

На етапі визначають:
1) мету моделювання;
2) набір факторних[14] ознак, що включають у модель;
3) значення параметрів (коефіцієнтів) при ознаках, їх суть та роль.

Метою економетричного дослідження є:

- аналіз економічного процесу, явища чи об'єкту;
- прогноз економічних показників, що характеризують цей процес;
- моделювання розвитку процесу за значеннями факторних ознак;
- підготовка управлінського рішення.

Включення факторних ознак в модель має бути обґрунтовано умовами:

- факторні ознаки мають бути **суттєвими**, а їхні значення переважно кількісними[15];
- на одну факторну ознаку рекомендована кількість значень від 5 до 7;
- вибірка має бути однорідною, для забезпечення якості моделі;
- значення змінних ознак мають бути розподілені симетрично.

2. Апріорний.

На етапі підготовки до побудови моделі здійснюється попередній теоретичний аналіз економічної суті процесу чи явища, формування і формалізація отриманої апріорної інформації.

3. Параметризація (або етап моделювання).

На цьому етапі **визначаються** з аналітичним виразом моделі, складом та формою зв'язків, що в неї входять, тобто здійснюється безпосереднє моделювання.

Головна задача етапу моделювання полягає у виборі найбільш оптимального виду функції залежності результативної ознаки від факторних. **Перевага** віддається формам залежності функцій **лінійних** перед нелінійними, як більш надійних.

---

[10] Сутність термінів буде розкрита далі.
[11] Гетероскедастичність припускає змінність дисперсії помилок за значеннями ознак кожного з факторів x_i.
[12] Кореляційна залежність між послідовними рівнями часового ряду. Статистичний взаємозв'язок між послідовностями величин одного ряду, взятими зі зсувом за часом, для випадкового процесу.
[13] Інтервалу часу між двома величинами, що періодично повторюються, одне є причиною, друге – наслідком.
[14] В деяких випадках, як факторні ознаки можуть розглядатися результативні ознаки або їхні похідні.
[15] Включення в модель регресії **якісних** факторних ознак, окрім кількісних, вимагає застосування непараметричних методів проведення розрахунків рівняння регресії за ранговими коефіцієнтами Спірмена.

Вирішується задача **специфікації** моделі шляхом:

- апроксимації аналітичною формою виявлених зв'язків та співвідношень між ознаками;
- відбору[16] у модель суттєвих факторних ознак залежних та/або незалежних;
- формулювання припущень та обмежень у моделі.

Результат моделювання залежить від правильної специфікації моделі.

4. Інформаційний.

Етап збору необхідної статистичної інформації полягає у реєстрації емпіричних значень ознак моделі, показників економічних процесів та аналізу якості отриманої інформації.

5. Ідентифікація моделі.

Здійснюється статистичний аналіз моделі і визначаються її невідомі параметри.

6. Оцінка якості моделі (верифікація).

Якість моделі перевіряють на достовірність і адекватність, тобто наскільки якісно вирішені задачі специфікації та ідентифікації моделі, і яка точність розрахунків за нею. Модель має бути адекватною реальному економічному процесу. Якщо якість моделі не задовольняє переліченим за якістю умовам, то повертаються до другого етапу моделювання.

7. Інтерпретація (тлумачення) результатів моделювання.

Серед найбільш відомих **моделей[17] економічних процесів**:

- моделі споживання;
- моделі ризику і дохідності цінних паперів;
- моделі пропозиції праці;
- моделі зростання (макроекономічні);
- моделі інвестування;
- моделі маркетингу;
- моделі валютних курсів і валютних криз.

IV. Класифікація економетричних змінних та вихідних даних вибірки.

В економетричних дослідженнях використовують вибірки з інформації щодо **вихідних даних за часом**, що мають назву:

- просторових даних (cross-sectional data);
- часових даних (time-series data).

Просторові дані – це сукупність значень економічної вихідної інформації вибірки, що всебічно або частково характеризує об'єкт за певний період. Наприклад, сукупність різної інформації підприємства за місяць, квартал, рік про чисельність його робітників, обсяг продукції, розмір основних фондів, обсяги споживання продукції на ринку відповідного виду, тощо.

Часові дані – це сукупність значень економічної вихідної інформації з вибірки за їх динамікою у різних періодах часового інтервалу, що характеризують певний об'єкт. Це може бути інформація про **динаміку** значень річного чи квартального індексу

[16] В сенсі, встановлення **суттєвих** факторних ознак, що підлягають включенню у аналітичний вираз функції.
[17] Економетричних моделей.

споживчих цін, щоденних курсів валют, річних обсягів виробництва чи реалізації продукту, іншої, що природньо упорядкована у довготривалому проміжку часу.

Існують певні відмінності **часових даних** від **просторових**, а саме:

- на відміну від просторових, часові дані **статистично залежні та схильні до автокореляції** (тобто, існує залежність часових даних минулих та поточного періодів);
- часові дані ряду є величини, що можуть мати різні розподілення.

Ознаки, що характеризують економічний процес, явище або об'єкт – є сукупністю економічної інформації. Ознаки, в економетричних моделях, пов'язані між собою і можуть бути:

- **результативною** або залежною змінною ознакою, що пояснюється;
- **факторною** або незалежною змінною ознакою, що пояснює результативну ознаку.

Економічні змінні. До економетричних моделей включені чотири їх види:

Екзогенні (незалежні) змінні – такі, значення яких задаються ззовні (*X*);

- Ендогенні (залежні) змінні – значення яких визначаються у моделі (*У*);
- Лагові екзогенні чи ендогенні змінні, що віднесені до попередніх проміжків часу, які знаходяться у економетричних рівняннях із змінними, що віднесені до поточного проміжку часу. Наприклад, X_{t-1} – лагова екзогенна змінна, Y_{t-1} – лагова ендогенна змінна;
- Зумовлені[18] (змінні, що пояснюють) – лагові X_{t-1} та поточні X_t екзогенні змінні, а також лагові ендогенні змінні Y_{t-1}, що пояснюють результативну змінну.

Для пояснення однієї або декількох поточних ендогенних змінних економетрічної моделі використовують значення зумовлених змінних, що є заздалегідь заданими (зумовленими) факторами і знаходяться ззовні цієї моделі.

V. Модель парної регресії.

Якщо в ході дослідження був виявлений зв'язок між факторними та результативною ознаками, то постає задача виявити вид цієї залежності. Регресія надає аналітичний вираз зв'язку в якому значення результативної ознаки обумовлені впливом факторних ознак. Кількісну ступінь їхньої залежності можливо встановити за рівнянням регресії.

Базисною економетричною моделлю є модель парної (або однофакторної) регресії. Для опису економічних процесів, які рівномірно розвиваються у часі, використовують модель, що, є поліномом першого ступеня, функція якого має аналітичний вираз:

$Y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$, рівняння регресії, де

- Y_i – результативна, залежна змінна, а x_i – факторна, незалежна змінна;
- β_0, β_1 – позначення параметрів **генеральної сукупності**; або
- b_0, b_1 – позначення коефіцієнтів[19] **вибіркової сукупності**;
- ε_i – помилка моделі регресії, що обумовлена передумовами:

[18] Предопределенные (объясняющие переменные), російською.
[19] В деяких джерелах коефіцієнти рівняння регресії вибіркової сукупності мають назву «статистик».

- нерепрезентативності[20] вибірки, коли сукупність за факторними ознаками не повною мірою пояснює у регресії варіацію результативної ознаки, оскільки у рівнянні неврахований вплив інших ознак;
- ймовірністю помилки виміру значення факторної ознаки у моделі та інші.

Методи визначення аналітичної форми залежності між ознаками функції такі:

- на підставі візуальної оцінки характеру зв'язку ознак.
  За формою лінії на графіку побудованого на емпіричних значеннях факторних та результативної ознак можливо стверджувати про наявність та форму залежності між ознаками. Площина з системою координат для нанесення значень результативної ознаки за відповідними факторними ознаками має назву **кореляційного поля**;
- на підставі теоретичного та логічного аналізу процесу і його соціально-економічної сутності.

Показник факторної ознаки рівняння парної регресії **генеральної** сукупності має назву параметра регресії «β_1», а **вибіркової** сукупності – коефіцієнта регресії «b_1». Його значення вказує наскільки в середньому зміниться результативна ознака за зміною факторної ознаки на одну одиницю свого значення. **Знак коефіцієнту** регресії вказує на напрямок зміни цього зв'язку. Якщо знак показника β_1 (b_1) **додатний** (**+**), зв'язок між ознаками прямий, тобто за зростанням факторної ознаки зростає і результативна. І навпаки, коли знак β_1 (b_1) **від'ємний** (**–**), зв'язок між ознаками зворотній, із зростанням факторної ознаки результативна ознака зменшується. За умови, коли факторні ознаки дорівнюють нулю, результативна ознака дорівнює значенню параметра (коефіцієнта) перетину β_0 (b_0) з віссю ординат, якщо така її інтерпретація має економічний сенс.

VI. Нормальна лінійна модель парної регресії.

Нормальна або класична лінійна[21] модель парної регресії будується на таких **припущеннях**:

1. Факторна ознака (x_i) є величина детермінована (тобто визначена), що не залежить від розподілення помилок (ε_i) рівняння регресії;
2. Математичне очікування помилок (ε_i) моделі регресії дорівнює нулю, або $E(\varepsilon_i) = 0$;
3. Дисперсія помилок (ε_i) рівняння регресії для всіх ознак є величина незмінна $D(\varepsilon_i) = G^2 = const$ тут, як умова сталості (гомоскедастичності)[22] дисперсії помилок випадкових відхилень на тій же множині;
4. Помилки (ε_i) рівняння регресії некорелюються між собою, тобто коваріація помилок двох різних спостережень дорівнює нулю або відсутня їх автокореляція. Це припущення справедливо для випадку коли вихідні дані не є «часовими рядами».

[20] Вибірка (непредставницька), чисельні характеристики якої відмітні від їх аналогів за генеральною сукупністю.
[21] Функція, графіком якої є пряма лінія, з чим пов'язана її назва, лінійність передбачає зміну результативної ознаки пропорційно зміні факторним ознакам; **многочлен першого ступеня** $f(x) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$.
[22] Гомоскедастичність припускає однакову дисперсію помилок за значеннями ознак кожного з факторів x_i.

5. Помилки рівняння це випадкові величини, що розподілені за нормальним законом, за яким їх математичне очікування та дисперсія дорівнюють відповідно «**0**» та $G^2(\varepsilon_i)$, або $N\left(0; G^2(\varepsilon_i)\right)$ що випливає з припущень 2, 3 та 4.

За цих припущень рівняння нормальної моделі лінійної парної регресії матиме вид:

$y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$, де

- y_i – значення залежної змінної;
- x_i – значення незалежної змінної;
- β_0, β_1 – позначення параметрів генеральної сукупності регресії;
- ε_i – помилка моделі (інколи вживають термін випадкова величина, у зв'язку з тим, що саме вона формує випадкові значення результативної ознаки).

Вибіркова лінійна регресія.

Парні лінійні моделі надають залежність між факторною та результативною змінними, наприклад:

- витратами на відпустку та складом родини;
- витратами на рекламу та обсягом продукції, що виробляється;
- витратами на споживання та валовим національним продуктом.

При цьому результативна ознака «*y*» вважається залежною змінною і розглядається як функція незалежної змінної «***x***» – факторної ознаки.

У загальному вигляді проста регресійна модель вибіркової сукупності має вираз:

$y_i = b_0 + b_1 x_i + e_i$, де

- y_i – вектор спостережень за залежною змінною $y_i = (y_1\ y_2, \ldots, y_n)$,
- x_i – вектор спостережень за незалежною змінною $x_i = (x_1, x_2, \ldots, x_n)$;
- b_0, b_1 – невідомі коефіцієнти вибіркової сукупності;
- e_i – вектор відхилень (помилок), $e_i = (e_1, e_2, \ldots, e_n)$;

Регресійна модель називається лінійною, якщо її функція лінійна за аналітичним виразом.

Наведене вище рівняння є лінійною моделлю за функцією прямої лінії на площині з коефіцієнтом b_0 у точці перетину вісі ординат та нахилом b_1 до вісі абсцис. Для визначення з аналітичною формою лінії регресії, необхідно визначити її невідомі коефіцієнти b_0, b_1.

Як приклад, розглянемо рівняння прямої лінії $\tilde{y}_i = b_0 + b_1 x_i$ на координатній площині. Найкращим критерієм проведення цієї прямої лінії через множину точок її значень є критерій мінімізації відхилень фактичних значень результативної ознаки від розрахованих, що виражені формулою: $e_i = y_i - \tilde{y}_i = y_i - (b_0 + b_1 x_i)$, i=1,...n, де $\tilde{y}_i$ – значення теоретично розрахованої результативної ознаки рівняння $\boldsymbol{i}$-ої точки на прямій, яка відповідає значенню ознаки $\boldsymbol{x_i}$ на рис.1.

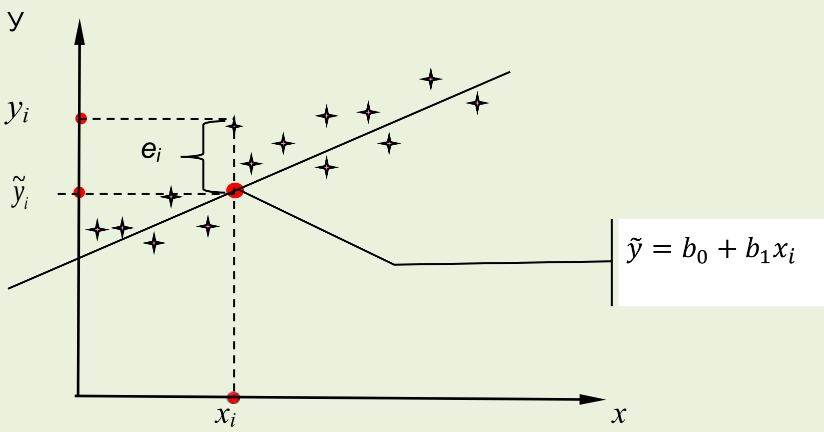


Рис.1

Пряма проводиться таким чином, щоб сума квадратів помилок була мінімальною.

Метод найменших квадратів полягає у визначені невідомих коефіцієнтів рівняння регресії шляхом мінімізації суми квадратів помилок $\sum_{i=1}^{n} e_i^2 \to min$, тобто маємо:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 = \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2 \to min.$$

VII. Методи визначення невідомих коефіцієнтів рівняння регресії.

На етапі проведення регресійного аналізу та моделювання рівняння регресії вибирається функція, що описує залежність результативної (*У*) від факторної ознаки (*X*). В подальшому потрібно визначити невідомі коефіцієнти b_0, b_1 рівняння. Методи їх розрахунку побудовані на врахуванні відхилень емпіричних[23] значень результативних ознак від їх теоретично розрахованих такі:

1. **Метод найменших квадратів** (МНК[24]) полягає у знаходженні такого рівняння регресії за яким отримане значення суми квадратів відхилень емпіричних значень результативної ознаки від її розрахованих теоретичних значень мінімальна: $\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 = min.$

   Метод має такі **переваги**:

   а) розрахунки зведені до механічної процедури знаходження коефіцієнтів;

   б) доступність отриманих математичних висновків.

   Недоліком методу МНК є велика чутливість до викидів[25].

2. **Метод суми модулів** полягає у визначенні за модулем суми відхилень емпіричних значень результативної ознаки від її теоретичних розрахованих за рівнянням регресії: $F = \sum_{i=1}^{n}|y_i - \tilde{y}_i|$.

   Перевагою метода суми модулів є його нечутливість до викидів оцінок вихідних даних.

   Серед **недоліків** такі:

- потреба врахування деякої нелогічності чи невідповідності у розрахункових процедурах;
- більшим відхиленням потрібна надавати більшу вагу для їх врівноваження в загальній сумі відхилень даних спостережень;

[23] Значення, що спостерігаються, тобто фактичні.
[24] Вперше був використаний А.М.Лежандром в 1805 році.
[25] Значень ознак, що мають значні відхилення з масиву значень вихідних даних.

- різним за значенням коефіцієнтів $b_0, \dots b_n$ можуть відповідати однакові суми модулів відхилень: $F = \sum_{i=1}^{n} g_i \times |y_i - \tilde{y}_i|$, де g_i – вага з якою конкретне відхилення входить в функцію *F*.

3. Поєднання переваг обох методів.

Для знаходження оптимальних значень коефіцієнтів $b_0, \dots b_n$ необхідно мінімізувати функцію за цими параметрами. Процес мінімізації функції здійснюється за умовами таких значень його коефіцієнтів за яких сума квадратів (модулів з врахуванням їх ваги «g») відхилень емпіричних значень результативної ознаки від її розрахованих теоретичних була б мінімальна. Метод найменших квадратів найбільш поширений при визначенні коефіцієнтів рівняння регресії $b_0, \dots b_n$.

VIII. Метод найменших квадратів для рівняння парної регресії.

Для знаходження невідомих коефіцієнтів рівняння лінійної парної регресії розглянемо метод найменших квадратів (МНК).

Нехай підібрана емпірична лінія, що задана рівнянням $Y_i = a + bx_i + \varepsilon_i$, форма якої дозволяє судження про лінійний зв'язок між результативною та факторною ознаками. Для того щоб підібране рівняння якісно описувало генеральну сукупність потрібно мінімізувати функцію суми квадратів відхилень фактичних (емпіричних) значень результативної ознаки від їх розрахованих за рівнянням. Тобто, теоретичні значення результативної ознаки розраховуємо з рівняння $\tilde{y}_i = \tilde{a} + \tilde{b}x_i$, після знаходження коефіцієнтів рівняння за методом найменших квадратів, що і наведено далі. Невідомими у функції є значення коефіцієнтів «*a,b*». Значення результативної і факторної ознак відомі зі спостережень. Для мінімізації $\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$, потрібно взяти часткові похідні функції за коефіцієнтами і прирівняти їх нулю. Отриману систему рівнянь вирішити відносно коефіцієнтів функції.

Оскільки $\tilde{y}_i = \tilde{a} + \tilde{b}x_i$, маємо таке рівняння для часткової похідної за параметром «a»:

$F'_{\tilde{a}} = \frac{\partial F}{\partial \tilde{a}} = \sum_{i=1}^{n}((y_i - \tilde{y}_i)^2)' = \sum_{i=1}^{n}(y_i^2 - 2y_i\tilde{y}_i + \tilde{y}_i^2)' = \sum_{i=1}^{n}\left(y_i^2 - 2y_i(\tilde{a} + \tilde{b}x_i) + (\tilde{a} + \tilde{b}x_i)^2\right)' = \sum_{i=1}^{n}\left(y_i^2 - 2y_i(\tilde{a} + \tilde{b}x_i) + \left(\tilde{a}^2 + 2\tilde{a}\tilde{b}x_i + (\tilde{b}x_i)\right)^2\right)' = \sum_{i=1}^{n}(0 - 2y_i + 2\tilde{a} + 2\tilde{b}x_i)' = \sum_{i=1}^{n}(-y_i + \tilde{a} + \tilde{b}x_i)' = 0.$

Звідки $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \tilde{a} + \sum_{i=1}^{n} \tilde{b}x_i$ або $\sum_{i=1}^{n} y_i = n \times \tilde{a} + \tilde{b}\sum_{i=1}^{n} x_i$

Для часткової похідної за параметром «*b*» маємо: $F'_{\tilde{b}} = \frac{\partial F}{\partial \tilde{b}} = \sum_{i=1}^{n}((y_i - \tilde{y}_i)^2)'$, звідки часткова похідна функції має рівняння: $\sum_{i=1}^{n} y_i x_i = \tilde{a}\sum_{i=1}^{n} x_i + \tilde{b}\sum_{i=1}^{n} x_i^2$.

Отримали систему нормальних рівнянь відносно параметрів «$\tilde{a}$» та «$\tilde{b}$»:

$$\sum_{i=1}^{n} y_i x_i = \tilde{a}\sum_{i=1}^{n} x_i + \tilde{b}\sum_{i=1}^{n} x_i^2 \text{ та } \sum_{i=1}^{n} y_i = n \times \tilde{a} + \tilde{b}\sum_{i=1}^{n} x_i$$

Звідки маємо формули для визначення коефіцієнтів вибіркової сукупності:

$$\tilde{a} = \bar{y} - \tilde{b}\bar{x} \qquad \text{та } \tilde{b} = \frac{\frac{1}{n}\left(\sum_{i=1}^{n} y_i x_i\right) - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2}, \quad \text{або} \quad \tilde{b} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{cov(x,y)}{var(x)},$$

де $\bar{x} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)$; $\bar{y} = \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)$; $\overline{xy} = \frac{1}{n}\left(\sum_{i=1}^{n} y_i x_i\right)$ – середні значення;

$var(x)$ – дисперсія незалежної ознаки; а $cov(x, y)$ – є коваріація змінних результативної та факторної ознаки за абсолютним показником тісноти їх зв'язку, що розраховується за формулою: $cov(x, y) = \overline{xy} - \bar{x}\bar{y}$.

IX. Коефіцієнти кореляції та детермінації.

Як визначити щільність зв'язку між залежною змінною «*У*» та незалежною «*Х*».

Критерієм впливу незалежної «*Х*» на залежну «*У*» є ***коефіцієнт кореляції***. Коефіцієнт кореляції (r_{xy}) є відносною мірою зв'язку між змінними факторною та результативною ознаками і розраховується за формулою: $r_{yx} = \dfrac{cov(x, y)}{\sqrt{Var(x)Var(y)}}$.

*Значення коефіцієнта кореляції знаходиться в інтервалі: [**–1 до +1**];*

Додатне його значення свідчить про прямий зв'язок, а від'ємне – про зворотній зв'язок між ознаками. Наближення коефіцієнта кореляції до значення «±1» свідчить про наявність сильного кореляційного зв'язку, якщо відбувається його наближення до «0» – то зв'язок відсутній.

Коефіцієнт детермінації.

Розглянемо відхилення фактичних значень від розрахованих теоретичних.

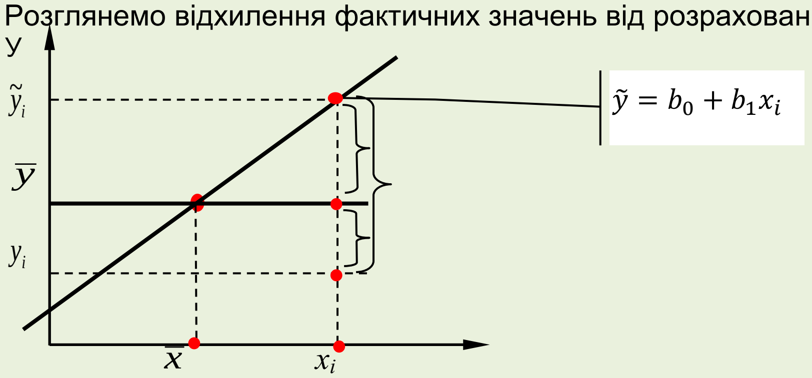


Рис.2.

Рівняння відхилення теоретичних значень ознак від фактичних має вираз: $(\tilde{y} - y_i) = (\tilde{y}_i - \bar{y}) + (\bar{y} - y_i)$, перепишемо його: $(y_i - \bar{y}) = (\tilde{y}_i - \bar{y}) + (y_i - \tilde{y})$ та піднесемо обидві до квадрата і просумуємо за всіма індексами: $\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = \sum_{i=1}^{n}\left(\tilde{y}_i - \bar{y}\right)^2 + \sum_{i=1}^{n}\left(y_i - \tilde{y}\right)^2$, або як *SST=SSR+SSE*, що теж саме, але за абревіатурою позначень сум квадратів[26].

де: $\tilde{y}_i$ – теоретичне розрахункове значення результативної ознаки i-ої точки;

y_i – фактичне значення i-ої точки; $\bar{y}$ – середньоарифметичне значення сукупності;

[26] SS – скорочено від sum of squares або сума квадратів;

$\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2$ – сума квадратів відхилень[27] <u>загальна</u>, позначається *SST*;

$\sum_{i=1}^{n}\left(\tilde{y}_i-\bar{y}\right)^2$ – сума квадратів відхилень[28], що <u>пояснює регресію</u> позначається *SSR*;

$\sum_{i=1}^{n}\left(y_i-\tilde{y}\right)^2$ – сума квадратів відхилень[29] <u>помилок</u>, що позначається *SSE*.

Поділивши обидві частини рівняння на кількість значень «*n*» отримаємо дисперсії:

$\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2$ загальна дисперсія ($\sigma^2_{заг.}$);

$\frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i-\bar{y})^2$ дисперсія, що пояснює регресію ($\sigma^2_{регр.}$);

$\frac{1}{n}\sum_{i=1}^{n}(y_i-\tilde{y})^2$ дисперсія помилок, ($\sigma^2_{пом.}$).

Загальна дисперсія лінійної регресії складається з двох частин: $\sigma^2_{заг.}=\sigma^2_{пом.}+\sigma^2_{регр.}$, тобто дисперсії, що пояснюється регресією $\sigma^2_{регр.}$ та дисперсії помилок $\sigma^2_{пом.}$.

Звідси: $1=\frac{\sigma^2_{\text{пом}}}{\sigma^2_{\text{заг}}}+\frac{\sigma^2_{\text{регр}}}{\sigma^2_{\text{заг}}}$, перша частина є пропорція дисперсії помилок у загальній дисперсії, яку не можна пояснити регресійним зв’язком, а друга частина її дисперсії пояснюється регресією. Відношення дисперсії, що пояснює регресією до загальної дисперсії, має назву **коефіцієнта детермінації**, позначається як: $R^2=\frac{\sigma^2_{\text{регр}}}{\sigma^2_{\text{заг}}}$, а його значення знаходиться в межах від <u>**0 *до* 1**</u> .

Зв’язок між коефіцієнтом кореляції (r_{xy}) та нахилом b_1.

Визначимо аналітичний зв’язок між <u>коефіцієнтом кореляції (r_{xy}) та нахилом b_1</u>.

Для коефіцієнта кореляції: $$r_{yx}=\frac{cov(x,y)}{\sqrt{Var(x)Var(y)}}=\frac{cov(x,y)}{\sigma_x\sigma_y}$$

Для нахилу b_1: $$b_1=\frac{cov(x,y)}{Var(x)}=\frac{cov(x,y)}{\sigma_x^2}$$

$$r_{yx}=\frac{cov(x,y)}{(\sigma_x)^2}\times\left(\frac{\sigma_x}{\sigma_y}\right)=b_1\frac{\sigma_x}{\sigma_y}$$

Оскільки значення середньоквадратичних відхилень σ_x, σ_y додатні, знак коефіцієнта кореляції (КК) збігається зі знаком коефіцієнта b_i, а значення i-го КК пов’язане з нахилом b_i, та середньоквадратичним відхиленням i-факторної та результативної ознак.

<u>Зв’язок між коефіцієнтом детермінації (R^2) та коефіцієнтом кореляції</u> (r_{xy}).

Коефіцієнт детермінації визначається за формулою: $R^2=b_1^2\frac{\sigma_x^2}{\sigma_y^2}$. <u>Множинний лінійний коефіцієнт кореляції</u> визначається за формулою: $r_{yx}=b_1\frac{\sigma_x}{\sigma_y}$, <u>звідки</u> $R^2 = r^2$ або коефіцієнт детермінації дорівнює квадрату множинного лінійного коефіцієнта кореляції.

[27] Що обумовлена різницею **фактичного** та її **середньоарифметичного** (існують й інші середні) значення ознаки;
[28] Що обумовлена різницею **розрахованого** значення будь-якої ознаки та її **середньоарифметичного** значення.
[29] Що обумовлена різницею **фактичного** значення будь-якої ознаки та **розрахованого** її значення.

X. Поняття про ступені вільності.

Кожна сума формули квадратів відхилень *SST=SSR+SSE* пов'язана з числом, яке називають її «**ступенем вільності»**.

Ступень вільності[30] (*degrees of freedom, англ.*) це **число**, за кількістю значень певної ознаки, що можуть вільно варіюватися (змінюватися) після вирахування значення параметра рівняння. Або число **незалежних значень** сукупності з ***n***-значень факторної $x_1, x_2,...x_n$ чи результативної ознаки $y_1, y_2,...y_n$, яке потрібно для розрахунку її суми квадратів.

1. Загальна сума квадратів відхилень (*SST*[31]) має $df = (n-1)$ ступенів вільності, оскільки з чисел $(y_1 - \bar{y}), (y_2 - \bar{y}), \ldots, (y_n - \bar{y})$ утворених для її розрахунку з «*n*» – значень результативної ознаки $y_1, y_2,...y_n$, незалежні тільки (*n–1*) в силу властивості середньої арифметичної: $\sum_{i=1}^{n}(y_i - \bar{y})^2 = 0$.
2. Сума квадратів відхилень, що пояснюються регресією (*SSR*) має $df = (k-1)$ ступінь вільності, де «*k*» – загальна кількість коефіцієнтів регресії, включаючи перетин[32]. Для розрахунку суми квадратів, що пояснює парну регресію, використовують лише одну незалежну одиницю інформації, коефіцієнт b_1, який впливає на значення результативної ознаки $y_1, y_2,...,y_n$, та, як наслідок, на суму квадратів відхилень регресії. Запишемо відхилення суми квадратів $\sum_{i=1}^{n}(\tilde{y}_i - \bar{y})^2 = b_1^2 \sum_{i=1}^{n}(\tilde{x}_i - \bar{x})^2$, що пояснює парну регресію. Звідси, суму квадратів відхилень, що пояснює **парну регресію**, можливо утворити використовуючи одну одиницю незалежної інформації, а саме b_1, а ступенів вільності $df = 2 - 1 = 1$.
3. Сума квадратів відхилень помилок (SSE)[33] має $df = (n-k)$ ступенів вільності.

   Визначається ця ступінь вільності, як різниця між загальною кількістю значень варіант сукупності і загальною кількістю коефіцієнтів «*k*» багатофакторної лінійної регресії. У разі парної лінійної регресії оцінюються два коефіцієнти «*k*», тобто b_0 та b_1. Для розрахунку суми квадратів помилок (*SSE*) парної регресії маємо $(n-2)$ ступенів вільності.

   Для багатофакторної лінійної регресії суми квадратів відхилень по ступенях вільності можливо розкласти за формулою $n-1 = (k-1) + (n-k)$, а для парної регресії за їхніми значеннями, як $n-1 = (2-1) + (n-2)$.

   Загальна сума квадратів відхилень має вираз: $SST = SSR + SSE$.

   Використовуючи приведені вище рівняння побудуємо таблицю дисперсійного аналізу.

[30] Позначається як df.
[31] Total Sum Square, англ.
[32] Коефіцієнт перетину враховується у загальній кількості коефіцієнтів, навіть за його значення нулю.
[33] Error Sum Square (SSE), англ.

Дисперсійний аналіз рівняння парної лінійної регресії.

Найменування	Кількість ступенів вільності	Сума квадратів відхилень	Середні суми квадратів відхилень (дисперсії) на одну ступінь вільності
Факторна варіація (що пояснена регресією)	$df_1 = k - 1$	$SSR = \sum_{i=1}^{n}(\tilde{y}_i - \bar{y}_i)^2$	$MSR = \frac{\sum_{i=1}^{n}(\tilde{y}_i - \bar{y}_i)^2}{k-1}$
Варіація помилок[34] (залишкова, що не пояснена регресією)	$df_2 = (n - k)$[35]	$SSE = \sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$	$MSE = \frac{\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}{n-k}$
Загальна варіація	$df_{\text{загальна}} = (n-1)$	$SST = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2$	$MST = \frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{n-1}$

Визначення дисперсії на одну ступінь вільності призводить дисперсії до порівняних їх значень. Зіставляючи середні суми квадратів відхилень факторну і помилок (або залишкову) тобто, суми їх квадратів відхилень в розрахунку на одну ступінь вільності, отримаємо величину *F*-критерію Фішера: $F = \frac{MSR}{MSE}$.

Перевірка гіпотези про значущість рівняння регресії за критерієм Фішера.

Для проведення аналізу якості рівняння регресії, необхідно розрахувати його **числові характеристики** – **математичне** очікування, **середньоквадратичне** відхилення, **дисперсія**, інші числа, що у стислій формі виражають найбільш суттєві риси розподілу. Кореляційний та регресійний аналіз, як правило, проводяться для обмеженої за кількістю значень вибірки за певною ознакою сукупності. За цих умов такі числові характеристики рівняння регресії, як коефіцієнти кореляції та детермінації можуть бути спотворені. Щоб перевірити значущість числових характеристик рівняння, тобто на скільки показники вибіркової сукупності наближені до показників генеральної і чи не є вони результатом збігу випадкових обставин, потрібно перевірити на **значущість** (**адекватність**) побудовану модель рівняння.

Значущість лінійної моделі рівняння регресії можливо перевірити за допомогою коефіцієнта детермінації. Якщо його значення наближене до **одиниці**, то вважають, що модель рівняння регресії значуща, якщо до **нуля**, то незначуща, тобто кореляційний зв'язок між залежною та незалежною ознаками відсутній.

Проте, який висновок можемо зробити, якщо значення **коефіцієнта детермінації** не має вираженого граничного його значення, наприклад, 0,5 чи 0,45. Потрібний інший критерій, що відповідав би за **адекватність** збудованої моделі, тобто її значущість. Найбільш поширеним **є критерій Фішера**.

[34] Незміщена оцінка дисперсії помилок лінійного рівняння регресії.
[35] Для рівняння парної регресії загальна кількість коефіцієнтів регресії «*k*» два b_0 (перетин) та b_1 (нахил).

При аналізі рівняння регресії на **значущість** можливі такі варіанти:

1. Модель за критерієм Фішера **значуща** у цілому, а також всі коефіцієнти значущі:
- модель може бути використана для прийняття рішень і здійснення прогнозів.
2. Модель за *F*-критерієм Фішера **значуща**, але частина коефіцієнтів незначущі:
- модель придатна для прийняття деяких рішень, але не для прогнозів.
3. Модель за *F*-критерієм **значуща**, але всі коефіцієнти регресії незначущі:
- модель вважається неадекватною.

Перевірити на значущість рівняння регресії це встановити, чи порівнянні розраховані значення результативної ознаки з їх фактичним значенням, чи достатньо включених у модель рівняння **факторних суттєвих ознак**, що пояснюють результативну ознаку.

Критерій Фішера встановлює, наскільки якісно регресійна модель пояснює загальну дисперсію залежної змінної. **Гіпотеза H_0** про незначущість коефіцієнта детермінації є припущенням незначущості рівняння регресії. **Тест** за критерієм **Фішера** важливий у регресійному аналізі та є перевіркою обмежень прийнятих у моделі регресії.

Нульова гіпотеза «H_0» – це припущення про рівність нулю всіх коефіцієнтів рівняння регресії (окрім постійного коефіцієнта перетину b_0). За такого припущення коефіцієнт детермінації дорівнює нулю.

Для багатофакторної лінійної регресії значення *F*-критерію Фішера розраховується як:

$F = \frac{R^2}{1-R^2} \times \frac{df_2}{df_1}$; чи $F = \frac{\left(\left(\sum_i^n (\tilde{y}_i - \bar{y})^2\right)/n\right)/(k-1)}{\left(\left(\sum_i^n (y_i - \tilde{y}_i)^2\right)/n\right)/(n-k)}$; або табличне $F(df_1;\ df_2)$ його значення де R^2 – коефіцієнт детермінації, а df_1 та df_2 кількість ступенів вільності,

$df_1 = k-1$, та $df_2 = n-k$, пов'язані з певною сумою квадратів відхилень, а «k» – кількість коефіцієнтів у рівнянні регресії, включаючи коефіцієнт перетину (в деяких джерелах за значення «k» приймається кількість коефіцієнтів лише при факторних ознаках).

Чисельник рівняння є дисперсія, що пояснюється регресією, а знаменник дисперсія, що не пояснюється регресією, кожна з яких віднесена до свого числа ступеня вільності. Якщо значення ***F*-критерію Фішера** більше критичного табличного значення при заданому рівні значущості та відповідних ступенях вільності, то нульова гіпотеза відкидається, що означає статистичну значущість регресії. У протилежному випадку модель визнається незначущою.

Перевірка на значущість лінійного рівняння регресії зводиться до перевірки гіпотези на значущість її коефіцієнтів «b_i» при факторних ознаках рівняння регресії або про значущість коефіцієнту детермінації. Для **судження про значущість** моделі за вибіркою, визначають середню помилку апроксимації.

Середня помилка апроксимації – це середня відносна помилка результативної ознаки, що розрахована за середнім значення відносних по модулю помилок результативної ознаки вибірки: $\bar{A} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \tilde{y}_i}{y_i}\right| \times 100\%$.

Перевірка на **значущість** рівняння регресії здійснюється за розрахованою середньою помилкою апроксимації, значення якої має не перевищувати **12-15%**. Для перевірки рівняння регресії мають право на існування такі гіпотези:

Гіпотеза *H₀* за якою коефіцієнт детермінації R^2 незначущий і, як наслідок, незначуще рівняння лінійної регресії. Для перевірки значущості моделі за значущістю коефіцієнта детермінації здійснюють такі дії:

- **На першому етапі** розраховуємо значення *F*-критерію: $F = \frac{MSR}{MSE}$ де:

MSR	– середня суми квадратів відхилень, що пояснює регресивну модель;
MSE	– середня суми квадратів відхилень помилок у моделі регресії;
df_1 та df_2	– ступені вільності, що пов'язані з зазначеними сумами відповідно.

- **На другому етапі** задаємо рівень значущості гіпотези: $\alpha = 1 - \gamma$; де γ – ймовірність наближена за значенням в межах 0,95÷0,99.

  Рівнем значущості гіпотези називають **ймовірність** $\alpha = 0{,}01 \div 0{,}05$ вчинити помилку першого роду, тобто прийняти гіпотезу H_0 коли вона невірна. Значення **коефіцієнту детермінації** за *F*-критерієм у цьому випадку значуще. Якщо ми вважаємо, що можлива помилка складає 5%, то це означає, що ми можемо помилитися у 5% випадків, а у 95% наші висновки будуть правильними.

- **На третьому етапі** за таблицями *F*-критерію Фішера з ступенями вільності df_1 та df_2, а також рівнем значущості (α) знаходимо **критичне** значення критерію ($F_{кр}$).

- **Якщо** розраховане значення *F*-критерію Фішера більше критичного його значення $F_{кр}$, то відкидаємо гіпотезу H_0, що коефіцієнти рівняння регресії $b_i=0$ або що $\tilde{y}_i = \bar{y}$, тобто того, що графік рівняння лінійної регресії є лінія, яка паралельна вісі абсцис, з ризиком помилитися не більше ніж 5%, а побудована модель адекватна дійсності.

Для багатофакторної лінійної регресії виду $Y = b_0 + b_1 \times X_1 + b_2 \times X_2$ збудованої за 20ма значеннями ознаки, маємо $df_1 = 3 - 1 = 2$ ступенів вільності пов'язаних з дисперсію регресії (кількість коефіцієнтів три) та $df_2 = 20 - 3 = 17$ ступенів вільності пов'язаних з дисперсією помилок.

Перевірка гіпотези значущості коефіцієнтів b_i регресії за t – критерієм Стьюдента.

Перевіркою статистичної гіпотези про значущість окремих коефіцієнтів моделі є перевірка припущення, що коефіцієнти рівняння регресії **значуще відрізняються** від нуля.

Необхідність перевірки гіпотези на значущість коефіцієнтів моделі викликана необхідністю отримання значущих результатів розрахунків побудованої моделі. Припустимо, що за вихідними даними вибіркової сукупності була збудована лінійна модель регресії. Завдання полягає у перевірці значущості її коефіцієнтів, що отримані методом найменших квадратів. Знайдені МНК коефіцієнти розподілені за нормальним законом і мають вираз $b_0 \sim N\left(\beta_0;\ \sigma^2_{\beta_0} = \sigma^2_\varepsilon \times \frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i-\bar{x})^2}\right)$ та $b_1 \sim N\left(\beta_1;\ \sigma^2_{\beta_1} = \sigma^2_\varepsilon \times \frac{1}{\sum_{i=1}^n (x_i-\bar{x})^2}\right)$, де β_0, β_1 – математичні очікування параметрів, а $\sigma^2_{\beta_0}$ та $\sigma^2_{\beta_1}$ їх невідомі дисперсії, що не можливо обчислити за неможливості обчислення дисперсії помилки генеральної сукупності $\sigma^2_{e_y}$, але можливо розрахувати їх дисперсії або середньоквадратичні відхилення $\tilde{\sigma}_{b_0}$ та $\tilde{\sigma}_{b_1}$ за вибірковою сукупністю. ***T*-статистика Стьюдента** будується за принципом: у чисельнику випадкова величина з нульовим математичним очікуванням (при виконанні нульової гіпотези), а в знаменнику – вибіркове стандартне відхилення цієї випадкової величини за її незміщеною оцінкою дисперсії.

Фактичні значення t-критерію розраховують за формулою: $t_{b_i} = \frac{b_i - \beta_i^*}{\tilde{\sigma}_{b_i}}$, як відношення за відхиленням коефіцієнту регресії від його гіпотетичного значення до його середньоквадратичного (стандартного) відхилення, де n – загальна кількість вибірки;

b_i – значення коефіцієнта факторної ознаки регресії;

β_i^* – гіпотетичне значення, яке набуває коефіцієнт β_i генеральної сукупності;

i – номер коефіцієнта факторної ознаки;

$\tilde{\sigma}_{b_i}$ – середньоквадратичні відхилення коефіцієнтів факторних ознак, що розраховані для:

$$\tilde{\sigma}^2_{b_0} = \frac{\sum_{i=1}^n e_i^2}{(n-k)} \times \frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i-\bar{x})^2};\quad \tilde{\sigma}_{b_0} = \tilde{\sigma}_\varepsilon \times \sqrt{\frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i-\bar{x})^2}} = \tilde{\sigma}_\varepsilon \times \frac{\sqrt{\sum_{i=1}^n x_i^2}}{n\sigma_x}.$$

$$\tilde{\sigma}^2_{b_1} = \frac{\sum_{i=1}^n e_i^2}{(n-k)} \times \frac{1}{\sum_{i=1}^n (x_i-\bar{x})^2};\quad \tilde{\sigma}_{b_1} = \tilde{\sigma}_\varepsilon \times \sqrt{\frac{1}{\sum_{i=1}^n (x_i-\bar{x})^2}} = \tilde{\sigma}_\varepsilon \times \frac{1}{\sigma_x\sqrt{n}};$$ та

$\tilde{\sigma}_\varepsilon = \tilde{\sigma}_{\text{помилок}} = \sqrt{\frac{\sum_{i=1}^n e_i^2}{(n-k)}}$, а k – загальна кількість коефіцієнтів, для парної регресії 2;

Основна гіпотеза «H_0» – полягає у припущенні незначущості коефіцієнтів факторних ознак рівняння регресії, або $\rightarrow b_1 = b_2 = b_3 = \cdots = b_n = \beta_i^* = 0$.

Гіпотеза H_0 – або нуль-гіпотеза за якою коефіцієнти факторних ознак b_i незначущі, тобто дорівнюють нулю, або якщо для парної регресії значення $b_1=0$, то, як наслідок, результат рівняння парної регресії незначущий.

Альтернативна гіпотеза «H_1» (зворотна) – про припущення значущості коефіцієнтів регресії, тобто про нерівність нулю коефіцієнтів вибіркової (для генеральної сукупності параметрів) регресії або: $b_1 \neq b_2 \neq b_3 \neq \cdots \neq b_n \neq \beta_i^* \neq 0$.

Для перевірки гіпотези статистичної значущості коефіцієнтів **регресії та кореляції** використовується *t*-критерій Стьюдента. Розраховані за вибіркою значення *t*-критерію будь-якого з коефіцієнтів рівняння регресії чи коефіцієнта кореляції порівнюється з критичними їх значеннями таблиць розподілення Стьюдента.

Табличне значення *t*-критерію розраховується в залежності від обраного рівня значущості «$\alpha = 1- \gamma$» і числа ступенів вільності для квадратів відхилень помилок (*SSE*), що не пояснені регресією, яке у випадку лінійної регресії дорівнює (*n* – *k*), де «*n*» – кількість значень спостереження, а «*k*» – кількість коефіцієнтів. Якщо по модулю фактичне значення *t*-критерію більше табличного, то вважають, що коефіцієнт факторної ознаки або коефіцієнт кореляції з імовірністю $\gamma = 1- \alpha$ значущо відрізняється від нуля. Якщо фактичне значення *t*-критерію менше табличного (по модулю), то немає підстав відкидати основну гіпотезу, тобто того, що коефіцієнти регресії або кореляції незначуще відрізняється від нуля при заданому рівні значущості.

Середнє квадратичне відхилення помилок залежної змінної може бути розраховане за формулою: $\sigma_{e_y} = \sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 = n \times G^2(y) \times \left(1 - r_{yx}^2\right)$, де $G^2(y)$ – загальна дисперсія залежної змінної; а r_{yx}^2 парний коефіцієнт детермінації.

Значущість коефіцієнтів рівняння регресії за *t*-критерієм Стьюдента розраховують за формулами: $t_{b_0} = \frac{b_0}{\tilde{\sigma}_{b_0}} = \frac{|b_0| \times n\sigma_x}{\sigma_e \sqrt{\sum_{i=1}^{n} x_i^2}}$;   та $t_{b_1} = \frac{|b_1| \times \sqrt{n}}{\sigma_e}\ \sigma_x$.

Для перевірки гіпотези про значущість (відмітність від нуля) коефіцієнта кореляції лінійної парної регресії розраховують критерій Стьюдента за формулою: $t_{r_{yx}} = \frac{r_{yx}}{\sigma_{r_{yx}}} = \frac{r_{yx}}{\sqrt{\frac{1 - r_{yx}{}^2}{n-k}}}$,

де r_{yx} – коефіцієнт кореляції між змінними «*y*» та «x», а $\sigma_{r_{yx}} = \sqrt{\frac{1 - \left(r_{yx}\right)^2}{n-k}}$ – значення його середньоквадратичної помилки, що визначене за малою вибіркою.

Розраховані значення *t*-критерію порівнюємо із критичними значеннями *t*-критерію таблиці Стьюдента за $df = (n - k)$ ступенями вільності та рівнем значущості «$\alpha = 1- \gamma$». Якщо значення більше табличного, коефіцієнти регресії значущі.

Знаходження інтервалу довіри числових характеристик[36] за t-розподілом Стьюдента.

На практиці ми завжди маємо справу з обмеженою кількістю значень спостережень, і завдання полягає в оцінці їх точності вимірювань, тобто на підставі певної кількості цих значень вибіркової сукупності потрібно знайти міру їх наближення до значення за генеральною сукупністю (так би мовити «істинного» значення).

Фактичне значення будь-якого з коефіцієнтів рівняння регресії, розраховане за даними вибірки, називають оцінкою коефіцієнта рівняння регресії сукупності. Оцінки бувають точкові та інтервальні.

Після визначення точкової оцінки[37] коефіцієнтів рівняння потрібно мати впевненість у надійності цієї оцінки. Особливо важлива така впевненість про точність оцінок коефіцієнтів для невеликих вибірок (оскільки для великої вибірки незміщенність, переконливість та надійність їх оцінок забезпечується твердженнями теорії статистики). Тому точкова оцінка малої вибірки має бути доповнена її інтервальною оцінкою – тобто інтервалом $(x_i \pm E)$ у середині якого, з наперед заданою ймовірністю «γ»[38], знаходиться точне значення числової характеристики x_i. Задачу розрахунку такого інтервалу називають інтервальним оцінюванням, а сам інтервал – довірчим інтервалом. При цьому довірчою ймовірністю або надійністю числової характеристики називають ймовірність «γ» з якою його значення потрапляє в довірчий інтервал.

Помилкою оцінки (*estimation error*) називають різницю між значенням параметру генеральної сукупності і його оцінкою (значенням коефіцієнту розрахованого за даними вибірки).

Довірча ймовірність (рівень довіри, *confidence level*) – це ймовірність знаходження значення коефіцієнта у довірчому інтервалі.

Існує кілька різних форм запису довірчих інтервалів, розуміючи, що вони еквівалентні і застосовуються залежно від зручності і контексту. Наприклад, форма запису текстом або формулою: $P(x_1 < \bar{x} < x_2) = 0{,}95$, де $\bar{x} = \mu$ – середнє значення, що дорівнює істинному, а *P*[39] – ймовірність з якою це значення попадає у довірчий інтервал. Довірчий інтервал залежить від вибірки. Оскільки вибірка випадкова, для кожної вибірки ми будуємо свій довірчий інтервал. Для довірчої ймовірності 95% довірчий інтервал буде покривати значення істинного параметру в 95 випадках зі 100.

Побудуємо довірчий інтервал **середнього** (математично очікуваного) для сукупності, що має нормальний закон розподілу. Припустимо, у нас є парна випадкова вибірка з генеральної сукупності, що за кількістю дорівнює «*n*».

---

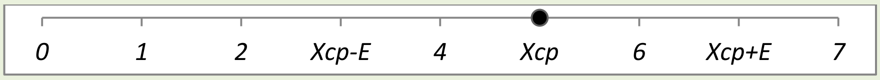
[36] **Числових характеристики регресії** таких, як математичне очікування, дисперсія, середнє квадратичне відхилення, коефіцієнти регресії, інші, що в стислій формі характеризують суттєві риси розподілу.

[37] Точкова оцінка (point estimate) це числова характеристика будь-якого коефіцієнта рівняння регресії чи інших числових характеристик, наприклад, середнє значення вибірки $\bar{x}$ є точковою оцінкою середнього значення «μ» генеральної сукупності.

[38] Літера гама (грецькою) – позначення ймовірності.

[39] Літера Р (французькою) probabilité – також позначення ймовірності.

Потрібно побудувати довірчий інтервал, який з довірчою ймовірністю «γ», що задана рівнянням $\gamma = 1 - \alpha$, буде містити середнє $\overline{x} = \mu$ генеральної сукупності: $\overline{x} - E < \mu < \overline{x} + E$.



За вибіркою можемо обчислити **середнє** значення, що є її **точковою оцінкою**. Надалі постає завдання знайти **інтервальну оцінку**, тобто обчислити межі довірчого інтервалу для середнього вибірки. При побудові довірчого інтервалу будемо посилатися на відомі нам властивості нормального закону розподілу. Існує два випадки, які ми і розглянемо.

Перший випадок. Середньоквадратичне відхилення «σ» відомо або $n \geq 30$.

Припустимо, що стандартне відхилення «σ» генеральної сукупності нам відомо або обсяг вибірки $n \geq 30$. Тоді, середнє генеральної сукупності, що має **нормальний закон розподілу**, з довірчою ймовірністю (*1-α*) знаходиться в довірчому інтервалі: $\bar{x} - E < \mu < \bar{x} + E$, де *E* – діапазон інтервальної оцінки, що знаходиться за формулою: $E = t_{\text{кр}} \times \frac{\sigma}{\sqrt{n}}$.

Послідовність дій для знаходження довірчого інтервалу наступна:

1. За вибіркою обчислюємо вибіркове середнє.

2. За таблицею нормального розподілу знаходимо $t_{\text{кр}}$ для довірчої ймовірності *(1 – α).*

3. Обчислюємо точність інтервальної оцінки за формулою: $E = t_{\text{кр}}^{0,05;\, n-2} \times \frac{\sigma}{\sqrt{n}}$.

Якщо значення «σ» невідомо при $n \geq 30$, замість «σ» у формулу підставляється її вибіркова оцінка.

4. Підставляємо отримані значення у формулу довірчого інтервалу: $\overline{x} - E < \mu < \overline{x} + E$.

__**Приклад №1**__.

Вибірка за віком з 50 студентів має середнє точкове значення 20,3 роки за умови стандартного відхилення 2^{-x} років.

Потрібно визначити 95% довірчий інтервал для середнього сукупності.

Рішення.

1. За вибіркою середнє за віком точкове значення вибіркової сукупності у 20,3 років.

2. Критичне значення t-критерію Стьюдента за заданою довірчою ймовірністю 95% та ступенями вільності $df = 50 - 2 = 48$[40], відповідає табличному його значенню $t_{\text{кр}}^{0,05;\, 48} = 2{,}01$.

3. Обчислимо значення довірчого інтервалу середнього сукупності (інколи вживають термін «точність», що не зовсім коректно), яке складає: $E = t_{\text{кр}}^{0,05;\, 48} \times \frac{\sigma}{\sqrt{n}} = 2{,}01 \times \frac{2}{\sqrt{50}} = 0{,}57$.

4. Підставимо отримані значення довірчого інтервалу: $20{,}3 - 0{,}57 < \mu < 20{,}3 + 0{,}57$.

[40] Для парної регресії *k*=2 (за кількістю коефіцієнтів перетину та факторного).

5. Середній вік студентів з ймовірністю 0,95 знаходиться в межах $19{,}73 < \mu < 20{,}87$.

Важливо, що при одному і тому ж обсягу вибірки збільшення довірчої ймовірності збільшує діапазон інтервальної оцінки (тобто зменшує «точність»), і навпаки. Крім цього, при постійній ймовірності, збільшення обсягу вибірки за кількістю «*n*» призводить до звуження довірчого інтервалу значення оцінки, тобто збільшує «точність».

Формула $E = t_{\text{кр}} \times \frac{\sigma}{\sqrt{n}}$ дозволяє визначити мінімальний обсяг вибірки «*n*», необхідний для отримання інтервальної оцінки з заданою довірчою ймовірністю що потрапляє в інтервал заданого розміру.

Приклад №2.

Потрібно оцінити середній вік студентів.

Оцінка має бути з точністю до **1 року** та імовірністю 99%.

Стандартне відхилення віку студентів – 2 роки.

Який розмір вибірки необхідний у цьому випадку?

Рішення.

Для заданої ймовірності 99% значення $\alpha = 1 - 0{,}99 = 0{,}01$. А вибірка за кількістю у 30 студентів, що задана ітерацією та $df = 30 - 2 = 28$, для зворотного двостороннього розподілення має табличне значення критерію Стьюдента $t_{\text{кр}}^{0{,}01;\,28} = 2{,}76$. За умовою діапазон відхилення середнього віку студентів $\mathrm{E} = 1$, при його стандартному $\sigma = 2$. Звідки формула розрахунку необхідної за кількістю вибірки матиме вираз: $n = \left(\frac{t_{\text{кр}}\,\sigma}{E}\right)^2 = \left(\frac{2{,}76 \times 2}{1}\right)^2 = 30.$

Наведеними вище формулами неможливо користуватися, якщо стандартне відхилення генеральної сукупності невідомо, чи обсяг вибірки малий ($n \leq 30$). Обидві ситуації цілком типові тому розглянемо другий випадок.

Другий випадок. Середньоквадратичне відхилення «σ» невідомо або $n \leq 30$.

Будуємо довірчий інтервал для середнього генеральної сукупності, що має нормальний закон розподілу: $\bar{x} - E < \mu < \bar{x} + E$.

Маємо просту випадкову вибірку кількістю «*n*» спостережень з генеральної сукупності. Дисперсія невідома і обсяг вибірки невеликий ($n \leq 30$), що не дозволяє скористатися наведеними вище формулами.

Відмінність побудови довірчого інтервалу для середнього полягає у заміні *t*-критерію стандартного нормального розподілу *t*-критерієм за **розподілом Стьюдента**. Розподіл був введений **В.С.Госсетом**, під псевдонімом Стьюдент (англ. «Student» – Студент). Розподіл схожий на нормальний розподіл, оскільки крива розподілу має форму дзвону, симетрична за середнім значенням і не стикається з віссю абсцис та являє сімейство кривих, що розрізняються за числом ступенів вільності. **Математичне очікування** помилок цього розподілу «**нуль**», а **дисперсія** помилок $D = \frac{n}{n-2}$, де $n > 2$.

При збільшенні обсягу вибірки за кількістю значень розподіл наближається до нормального. Для знаходження «*t*-критерію» використовуємо таблицю розподілу Стьюдента.

Побудова довірчих інтервалів для середнього генеральної сукупності має нормальний закон розподілу. Його точковою оцінкою є середнє значення за вибіркою. Оцінку інтервалу за точковим значенням будуємо за значенням *E*, що є її «точністю». Якщо вибірка достатньо велика або нам відомо стандартне відхилення генеральної сукупності, спираючись на властивості нормального закону обчислюємо інтервал розкиду середнього значення за визначеним *t*-критерієм таблиці нормального розподілу.

Якщо обсяг вибірки невеликий $n \le 30$ і стандартне відхилення генеральної сукупності не відомо, побудова довірчого інтервалу відбувається за допомогою розподілу Стьюдента, а *t*-критерій визначаємо за таблицею розподілу Стьюдента. Формули дозволяють оцінювати мінімальний обсяг вибірки, необхідний для отримання інтервальної оцінки із заданими діапазоном (точністю) та довірчою ймовірністю. Знаходження інтервалів довіри для коефіцієнтів b_i сукупності за *t*-розподілом Стьюдента аналогічне за процедурою знаходження для середнього сукупності наведеної вище.

XI. Визначення дисперсії помилок регресії.

Дисперсію помилок генеральної сукупності у більшості випадків визначити неможливо. Тому виникає потреба у розрахунку (оцінці) її незміщеного значення вибіркової сукупності.

Незміщену оцінку дисперсії помилок рівняння лінійної регресії визначають за формулою: $\tilde{G}^2(\varepsilon) = \tilde{S}^2(\varepsilon) = \frac{\sum_{i=1}^{n} e_i^2}{n-k}$, де ***n*** – об'єм вибіркової сукупності; ***k***=2 для парної регресії, а e_i – помилка[41] регресії; $e_i = y_i - \tilde{y}_i = \tilde{\beta}_0 - \tilde{\beta}_1 \times x_i$.

Для моделі лінійної множинної регресії незміщена оцінка дисперсії помилки розраховується за формулою: $\tilde{S}^2(\varepsilon\ ) = \frac{SSE}{n-k} = \frac{\sum_{i=1}^{n} e_i^2}{n-k}$, де *SSE* – сума квадратів помилок, ***k*** – кількість коефіцієнтів моделі регресії, включаючи коефіцієнт перетину. Сума квадратів помилок базується на кількості ступенів вільності, яка визначається різницею між кількістю спостережень ознаки і кількістю коефіцієнтів, що визначаються. Для рівняння парної лінійної регресії це коефіцієнти β_0та β_1, тобто *k=2*, а ступенів вільності $(n-2)$. Ступені вільності позначають, як df. Для утворення загальної суми квадратів відхилень (*SST*) значень результативної ознаки від їх середнього з чисел $(y_1 - \bar{y}), (y_2 - \bar{y}), \dots (y_{n-1} - \bar{y})$ потрібно $(n-1)$ елементів, оскільки за властивості середньої арифметичної $\sum_{i=1}^{n}(y_i - \bar{y}) = 0$ лише $(n-1)$ їх значень незалежні. Суму квадратів відхилень (*SSR*), що пояснює **парну регресію** отримують використовуючи тільки одну одиницю інформації, оскільки саме нахил «β_1» рівняння регресії пояснює значення результативної ознаки, звідси маємо один ступінь вільності.

[41] Інколи вживають термін «залишки», що вносить неоднозначність визначення одного і того ж значення.

Таким чином, значення дисперсії помилок **вибіркової сукупності** $\tilde{S}^2(\varepsilon)$ є незміщеною оцінкою значення дисперсії помилок **генеральної сукупності** $\tilde{G}^2(\varepsilon)$. В теорії припускають, що розрахунок (оцінка) точкового значення будь-якої <u>числової характеристики</u> регресії, що отримана методом найменших квадратів має складові:

1. **Істинне** значення, що визначене за <u>генеральною сукупністю</u>;
2. **Випадкове** значення або помилка (ε_i), в межах якої змінюється істинне значення.

На практиці розклад на складові неможливий, оскільки невідомо істинні точкові значення числових характеристик рівняння регресії генеральної сукупності, але теорія дозволяє надати судження про якість <u>точкових оцінок</u> за методом найменших квадратів, а саме щодо:

- <u>незміщення оцінки</u> – коли **математичне очікування** точкової оцінки числових характеристик сукупності вибірки (середньої, дисперсії, коефіцієнтів) **дорівнює їх значенню** у генеральній сукупності;
- <u>ефективність (надійність) оцінки</u> – за якою статистика точкової оцінки характеристик генеральної сукупності, має **мінімальну стандартну помилку**, тобто функція розподілення наближена до істинної або теоретичної, за якої дисперсія у виборці найменша;
- <u>спроможність оцінки</u> – коли за зростанням обсягу вибірки <u>набуття точкової оцінки</u> будь-якої з числових характеристик вибірки значення її параметра у генеральній сукупності стає **достовірною за ймовірністю подією**.

<u>Вибіркове середнє</u>, що задовольняє всім переліченим критеріям є найкращою числовою характеристикою середнього значення генеральної сукупності.

Параметр β_1 за МНК-оцінкою розраховується за формулою: $\tilde{\beta}_1 = \frac{cov(x,y)}{G^2(x)}$.

Коваріація між залежною змінною «*У*» та незалежною змінною «*Х*» може бути подана як $Cov(x;y) = Cov(x;\ \beta_0 + \beta_1 x + \varepsilon) = Cov(x;\beta_0) + Cov(x;\beta_1 x) + Cov(x;\varepsilon)$, або як: $Cov(x;y) = \beta_1 G^2(x) + Cov(x,\varepsilon)$, звідки значення параметра β_1 за МНК-оцінкою приймає вид: $\tilde{\beta}_1 = \frac{cov(x,y)}{G^2(x)} = \frac{\beta_1 G^2(x) + Cov(x,\varepsilon)}{G^2(x)} = \beta_1 + \frac{Cov(x,\varepsilon)}{G^2(x)}$, тобто подана сумою постійного значення параметра рівняння та значення його помилки (випадкової величини).

Припущення про нормальне розподілення залишків дозволяє нам здійснювати перевірку параметрів регресії та кореляції за допомогою критеріїв Фішера та Стьюдента (Госсета), а також значущість та адекватність моделі в цілому. Разом с тим оцінки регресії, що знайдені за методом МНК володіють прийнятними властивостями навіть за відсутності нормального розподілення залишків при цьому важливо лише щоб відбувалася гомоскедастичність дисперсії та відсутність автокореляції помилок.

XII. Переконливість та незміщення МНК-оцінок.

У побудові нормальної лінійної регресійної моделі **прийняті такі припущення**:

- факторна ознака (*X*) рівняння регресії є детермінованою величиною;
- математичне очікування помилок (ε_i) рівняння регресії, тобто середнє їх значення дорівнює нулю $\mathrm{E}(\varepsilon_i) = 0$;
- дисперсія помилок (ε_i) величина постійна $D(\varepsilon_i) = G^2(\varepsilon_i) = const$, тобто виконується умова гомоскедастичності (однаковості) дисперсії помилок;
- помилка (ε_i) та незалежна ознака рівняння регресії некорелюють між собою тобто їх коваріація дорівнює нулю $Cov(x, \varepsilon) = 0$.
- помилки $(\varepsilon_1, \varepsilon_2)$ різних значень незалежної ознаки рівняння регресії некорелюють між собою (відсутня їх автокореляція), тобто $Cov(\varepsilon_1, \varepsilon_2) = 0$.

Тоді оцінка коефіцієнтів, що отримана за МНК вважається дійсною оцінкою параметрів, якщо вона задовольняє властивостям таким, як незміщення, переконливості, ефективності.

1. Оцінка коефіцієнта має назву «**незміщеної** його оцінки», якщо його **математичне очікування** за вибірковою сукупністю дорівнює його значенню за генеральною сукупністю, тобто $E(\tilde{\theta}) = \theta,\ E(\tilde{\theta}) - \theta = \varphi = 0$, де φ – зміщення оцінки.

<u>Виходячи з визначення властивості незміщення</u> маємо $E(\tilde{\beta}_1) = \beta_1$. Математичне очікування значення коефіцієнта вибіркової сукупності дорівнює **сумі** його **істинного значення** за <u>генеральною сукупністю</u> і **помилки**. Враховуюче, що його коваріація дорівнює нулю маємо:

$$E(\tilde{\beta}_1) = E\left(\beta_1 + \frac{cov(x,\varepsilon)}{G^2(x)}\right) = \beta_1 + \frac{0}{G^2(x)} = \beta_1.$$

Таким чином МНК-оцінка коефіцієнта $\tilde{\beta}_1$ є його незміщеною оцінкою параметра β_1.

2. Оцінка коефіцієнта має назву «<u>**переконливої**</u>», якщо вона задовольняє закону великих чисел, за яким зі збільшенням вибірки сукупності оцінка точкового значення коефіцієнту вибіркової сукупності прямує до значення його параметру у генеральній сукупності.

<u>Для визначення переконливості оцінки</u> достатньо виконання **двох умов**:

1. Зміщення оцінки параметра дорівнює нулю, $\varphi = 0$ за кількості «*n*», що прямує до нескінченності, $n \to \infty$.
2. Дисперсія параметра прямує до нуля $G^2(\tilde{\theta}) \to 0$ за кількості «*n*» спостережень у вибірці, що прямує до нескінченності $n \to \infty$.

XIII. Розрахунки числових характеристик рівняння регресії.

Розглянемо вирішення стандартної задачі. Метою є відповідь на питання, чи є отримані коефіцієнти регресії «*a*» та «*b*» статистично значущими? Будемо досліджувати залежність величини дивідендів «*Y*» акцій від прибутковості акцій «*X*».

Масив вихідних даних приведений в табл. №1.

Таблиця №1

№	x_i	y_i	x^2	y^2	$x_i \times y_i$
1	8,19	42,08	67,14	1770,46	344,77
2	7,65	36,51	58,53	1333,01	279,32
3	4,56	23,76	20,76	564,40	108,24
4	5,92	27,23	35,02	741,68	161,17
5	8,15	25,44	66,45	647,33	207,41
6	6,51	35,44	42,37	1256,17	230,71
7	6,23	34,18	38,82	1168,53	212,99
8	6,42	27,33	41,27	747,19	175,6
9	6,90	32,69	47,55	1068,57	225,42
10	7,45	34,74	55,55	1207,12	258,96
11	7,58	32,47	57,38	1054,09	245,94
12	6,60	30,18	43,61	910,77	199,29
13	3,28	28,01	10,77	784,56	91,91
14	7,42	39,48	54,99	1558,45	292,74
15	8,91	47,81	79,32	2285,38	425,77
16	8,42	45,70	70,87	2088,63	384,74
17	6,11	29,39	37,39	864,01	179,73
18	9,48	38,19	89,96	1458,76	362,25
19	8,97	30,78	80,43	947,36	276,03
20	5,79	37,86	33,48	1433,68	219,09
21	4,77	24,29	22,77	590,16	115,92
22	7,79	37,28	60,64	1389,55	290,29
23	6,99	27,78	48,89	771,49	194,21
24	4,43	31,61	19,60	999,26	139,95
25	6,04	31,09	36,50	966,62	187,84
26	8,84	40,95	78,14	1677,22	362,02
27	8,06	41,38	64,93	1712,12	333,41
28	8,57	45,16	73,42	2039,53	386,98
29	4,78	21,62	22,83	467,50	103,30
30	7,01	39,76	49,12	1581,10	278,67
31	3,25	24,20	10,57	585,61	78,67
32	6,41	32,40	41,05	1050,04	207,60
33	6,91	29,79	47,80	887,69	206,00
34	5,84	35,63	34,14	1269,63	208,2
35	6,38	32,35	40,75	1046,49	206,49
36	8,37	37,57	70,00	1411,56	314,34
37	8,02	34,01	64,33	1156,46	272,74
38	11,7	43,06	136,92	1853,84	503,82
39	5,92	38,63	35,05	1492,23	228,71
40	7,69	34,49	59,18	1189,37	265,30
41	9,32	35,13	86,86	1233,89	327,37
42	8,63	37,22	74,46	1385,56	321,21
43	6,14	27,77	37,68	771,31	170,49
44	8,22	39,84	67,55	1587,14	327,43
45	6,2	34,03	38,50	1158,06	211,15
46	2,35	22,38	5,55	501,04	52,71
47	5,34	30,51	28,48	930,89	162,82
48	6,36	33,92	40,47	1150,52	215,78
49	5,16	27,16	26,63	737,70	140,16
50	7,49	28,94	56,12	837,46	216,78
Разом ∑	**343,52**	**1679,22**	**2510,59**	**58321,19**	**11912,44**
Середнє	**6,87**	**33,58**			**238,25**

Рівняння регресії має аналітичний вираз:

$$y_i = b_0 + b_1 \times x_i + \varepsilon_i$$

де y_i – вектор спостережень за залежною змінною $y_i = (y_1, y_2, \ldots y_n)$

x_i – вектор спостережень за незалежною змінною $x_i = (x_1, x_2, \ldots x_n)$;

b_0, b_1 – невідомі параметри регресійної моделі вибіркової сукупності;

e_i – вектор помилок, $e_i = (e_1, e_2, \ldots e_n)$.

Система нормальних рівнянь має наступний вид	
в загальному вигляді:	за вихідними даними:
$\sum_{i=1}^{n} y_i x_i = \tilde{a}\sum_{i=1}^{n} x_i + \tilde{b}\sum_{i=1}^{n} x_i^2$	$11912{,}22 = 50a + 343{,}52b$
$\sum_{i=1}^{n} y_i = n \times \tilde{a} + \tilde{b}\sum_{i=1}^{n} x_i$	$1679{,}22 = 343{,}52a + 2510{,}59b$

Звідси маємо таке рівняння парної лінійної регресії: $Y = 16{,}437 + 2{,}496x$.

Дисперсійний аналіз програмою EXCEL введеного масиву значень у директорії «Дані» та розділу «аналіз даних» пакету аналізу надав такі результати:

Регресійна статистика		
Множинний коефіцієнт кореляції (R)	0,698	$=\sqrt{0{,}487}$
Коефіцієнт детермінації R^2	0,487	≈ 937,03 / 1924,34
Нормований коефіцієнт детермінації R^2	0,476	≈ 1–(1–0,487) × 49 / 48
Стандартна помилка[42]	4,535	$\approx \sqrt{987{,}31/48} = \sqrt{20{,}57}$
Кількість значень ознаки	50	

Дисперсійний аналіз, за розрахунками програми EXCEL

Найменування		***d f***	***SS***	***MS***	***F***	*Значущість F*	
a		***b***	***c***	***d***	***e***	***f***	***g***
				(c1) / (b1)	***(d1) / (d2)***		
1	Регресія	1	937,03	**937,03**	**45,555**	1,78106E–08	
2	Залишок	48	987,31	20,57			
3	Разом	49	1924,34				
Коефіцієнти			*Стандартна*	***(b4) / (c4)***		**(b4) +(c4)×(d6)**	**(b4)–(c4)×(d6)**
Назва		*Значення*	*помилка*	*t-статист*	***P****-значення*	*Верхні 95%*	*Нижні 95%*
4	**Перетин**	**16,437**	2,62	6,273	9,58E-08	**21,71**	**11,17**
Значення перетину (функції Y) за довірчим інтервалом = 16,437±2,62*2,011							
				(b5) / (c5)		**(b5)+(c5)×(d6)**	**(b5)–(c5)×(d6)**
5	*b*	**2,496**	0,37	6,750	1,78E-08	**3,24**	**1,75**
	Значення змінної X_1 за довірчим інтервалом = 2,496±0,37*2,011						
6	(n-m=48; α=0,05) *t-критичне (48; 0,05)* =**2,011**						

Числові характеристики рівняння регресії.

Вибіркові середні:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{343{,}52}{50} = 6{,}87; \quad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{1679{,}22}{50} = 33{,}58;$$

$$\overline{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n} = \frac{11912{,}44}{50} = 238{,}25.$$

Вибіркові дисперсії:

$$S^2(x) = \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2 = \frac{2510{,}59}{50} - 6{,}87^2 = 3{,}01;$$

$$S^2(y) = \frac{\sum_{i=1}^{n} y_i^2}{n} - \bar{y}^2 = \frac{58321{,}2}{50} - 33{,}58^2 = 38{,}78.$$

Середньоквадратичні відхилення

$$S(x) = \sqrt{S^2(x)} = \sqrt{3{,}01} = 1{,}73$$

$$S(y) = \sqrt{S^2(y)} = \sqrt{38{,}78} = 6{,}22$$

[42] Термін введений **Одні Юлом** (англ. *George* ***Udny Yule***) в 1897 році.

Для оцінки якості коефіцієнтів регресії розрахуємо значення за формулами табл. №2.

Таблиця №2

№	Ознаки функції		Аналітичний	Значення квадратів відхилень[43]		
	факторна	результативна	вираз функції	загальну	помилок	регресію
п/п	x_i	y_i	$\tilde{y}_i = a + bx_i$	$(y_i - \bar{y})^2$	$(y_i - \tilde{y})^2$	$(\tilde{y}_i - \bar{y})^2$
1	8,19	42,08	36,88	72,25	27,04	10,89
2	7,65	36,51	35,53	8,58	0,96	3,80
3	4,56	23,76	27,82	96,43	16,48	33,18
4	5,92	27,23	31,21	40,32	15,84	5,62
5	8,15	25,44	36,78	66,26	128,60	10,24
6	6,51	35,44	32,68	3,46	7,62	0,81
7	6,23	34,18	31,99	0,36	4,80	2,53
8	6,42	27,33	32,46	39,06	26,32	1,25
9	6,90	32,69	33,66	0,79	0,94	0,01
10	7,45	34,74	35,03	1,35	0,08	2,10
11	7,58	32,47	35,36	1,23	8,35	3,17
12	6,60	30,18	32,91	11,56	7,45	0,45
13	3,28	28,01	24,62	31,02	11,49	80,28
14	7,42	39,48	34,96	34,81	20,43	1,90
15	8,91	47,81	38,67	202,49	83,54	25,91
16	8,42	45,70	37,45	146,89	68,06	14,98
17	6,11	29,39	31,69	17,56	5,29	3,57
18	9,48	38,19	40,10	21,25	3,65	42,51
19	8,97	30,78	38,82	7,84	64,64	27,46
20	5,79	37,86	30,89	18,32	48,58	7,24
21	4,77	24,29	28,34	86,30	16,40	27,46
22	7,79	37,28	35,88	13,69	1,96	5,29
23	6,99	27,78	33,88	33,64	37,21	0,09
24	4,43	31,61	27,49	3,88	16,97	37,09
25	6,04	31,09	31,51	6,20	0,18	4,28
26	8,84	40,95	38,50	54,32	6,00	24,21
27	8,06	41,38	36,55	60,84	23,33	8,82
28	8,57	45,16	37,83	134,10	53,73	18,06
29	4,78	21,62	28,37	143,04	45,56	27,14
30	7,01	39,76	33,93	38,19	33,99	0,12
31	3,25	24,20	24,55	87,98	0,12	81,54
32	6,41	32,40	32,44	1,39	0,00	1,30
33	6,91	29,79	33,68	14,36	15,13	0,01
34	5,84	35,63	31,01	4,20	21,34	6,60
35	6,38	32,35	32,36	1,51	0,00	1,49
36	8,37	37,57	37,33	15,92	0,06	14,06
37	8,02	34,01	36,45	0,18	5,95	8,24
38	11,70	43,06	45,64	89,87	6,66	145,44
39	5,92	38,63	31,21	25,50	55,06	5,62
40	7,69	34,49	35,63	0,83	1,30	4,20
41	9,32	35,13	39,70	2,40	20,88	37,45
42	8,63	37,22	37,98	13,25	0,58	19,36
43	6,14	27,77	31,76	33,76	15,92	3,31
44	8,22	39,84	36,95	39,19	8,35	11,36
45	6,20	34,03	31,91	0,20	4,49	2,79
46	2,35	22,38	22,30	125,44	0,01	127,24
47	5,34	30,51	29,76	9,42	0,56	14,59
48	6,36	33,92	32,31	0,12	2,59	1,61
49	5,16	27,16	29,32	41,22	4,67	18,15
50	7,49	28,94	35,13	21,53	38,32	2,40
Разом	$\sum$=**343,52**	$\sum$=**1679,22**	$\sum$=**1679,21**	Значення сум квадратів відхилень		
Середнє	$\bar{x}$	$\bar{y}$		*SST*	≈ *SSE*	+ *SSR*
значення	**6,87**	**33,58**		**1924,30**[44]	**≈ 987,48**	**+ 937,22**

[43] Що пояснюють суму.

Аналіз точності коефіцієнтів регресії.

З табл. 2 загальна дисперсія результативної ознаки складатиме: $S_y^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1} = \frac{1924{,}3^{45}}{50-1} = 39{,}27$, факторної ознаки $S_x^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n} = \frac{150{,}4}{50} = 3{,}01$, звідки оцінка їх середніх квадратичних відхилень (СКВ) $S_y = \sqrt{S_y^2} = \sqrt{39{,}27} = 6{,}27$ та $S_x = \sqrt{S_x^2} = \sqrt{3{,}01} = 1{,}73$, що порівняна з їх оцінками за іншими формулами, див. вище. Незміщеною оцінкою дисперсії помилок результативної ознаки: $S_{y(SSE)}^2 = \frac{\sum_{i=1}^{n}(y_i-\tilde{y})^2}{n-k} = \frac{987{,}3}{50-2} = 20{,}57$, де $S_{y(SSE)}^2$ – непояснена дисперсія помилок (SSE), величина розкиду результативної ознаки, а

$S_{y(SSE)} = \sqrt{S_{y(SSE)}^2} = \sqrt{20{,}57} = 4{,}54$ – оцінка її стандартної помилки.

Стандартні помилки коефіцієнтів «a» та «b» складатимуть:

$$S_a = S_{y(SSE)} \times \frac{\sqrt{\sum_{i=1}^{n} x_i^2}}{n \times S_x} = 4{,}54 \times \frac{\sqrt{2510{,}59}}{50 \times 1{,}73} = 2{,}63;$$

$S_b = \frac{S_{y(SSE)}}{\sqrt{n} \times S_x} = \frac{4{,}54}{\sqrt{50} \times 1{,}73} = 0{,}37$; де S_x – стандартне відхилення факторної змінної;

Оцінимо значущість показників рівняння регресії, яке найбільш адекватно описує залежність між показниками з імовірністю 95%.

За допомогою МНК ми отримали оцінки коефіцієнтів рівняння регресії, які характерні для конкретного статистичного спостереження (конкретного набору значень «*x*» і «*y*»). <u>Для оцінки статистичної значущості коефіцієнтів регресії і кореляції розраховують *t*-критерій Стьюдента та довірчі інтервали кожного з показників</u>. Висувається гіпотеза H_0 про випадкову природу показників, тобто про незначущу їх відмітність від нуля. Щоб перевірити значущість параметрів, тобто, чи значуще вони відрізняються від нуля, для генеральної сукупності використовують статистичні методи перевірки гіпотез.

В якості основної (нульової) гіпотези висувають гіпотезу про незначущу відмінність від нуля параметра в генеральній сукупності. Поряд з основною гіпотезою висувають альтернативну (конкуруючу) гіпотезу про відмітність нулю параметра в генеральній сукупності.

Перевіримо гіпотезу H_0 про рівність окремих коефіцієнтів регресії нулю (при альтернативній H_1 коли значення коефіцієнтів відмітні від нуля) на рівні значущості α=0,05. У разі, якщо основна гіпотеза виявиться невірною, приймаємо альтернативну. Для перевірки цієї гіпотези використовується *t*-критерій Стьюдента.

Знайдене за даними спостережень значення *t*-критерію (його ще називають спостережуваним або фактичним) порівнюється з табличним (критичним) його значенням, визначеним за таблицями розподілу Стьюдента (які наводяться в кінці підручників).

Табличне значення визначається залежно від рівня значущості (α) і числа ступенів вільності, яке в разі лінійної парної регресії дорівнює (*n*-2), де *n*-кількість значень спостереження.

---

[44] Ця приведена таблиця розрахована з точністю, як на екрані монітора комп'ютера.
[45] Тут і далі використані дані з табл. 2.

Якщо по модулю фактичне значення t-критерію більше табличного, то основну гіпотезу відкидають і вважають, що з імовірністю «*1–α*» параметр або статистична характеристика у генеральній сукупності значущо відрізняється від нуля.

Якщо фактичне значення t-критерію менше табличного (по модулю), то немає підстав відкидати основну гіпотезу, тобто параметр або статистична характеристика в генеральній сукупності незначуще відрізняється від нуля при заданому рівні значущості «α».

Значення t-критерію коефіцієнтів рівняння визначимо за формулами: $t_b = \frac{b}{S_b} = \frac{2{,}50}{0{,}37} = 6{,}76$ та $t_\alpha = \frac{a}{S_a} = \frac{16{,}44}{2{,}63} = 6{,}25$, а табличне критичне значення $t_{крит}$ (n-k; α)=(48; 0,05)=2,011.

Оскільки значення розрахованого t-критерію більше критичного, то статистична значущість коефіцієнтів «a» та «b» регресії підтверджується, тобто гіпотезу про рівність нулю цих коефіцієнтів відкидаємо.

Висновок: отримані числові характеристики коефіцієнтів регресії статистично значущі, і можуть бути використані у рівнянні лінійної регресії для розрахунків прогнозів та аналізів.

Довірчі інтервали для коефіцієнтів рівняння регресії.

Розрахуємо довірчі інтервали коефіцієнтів перетину та факторної ознаки рівняння регресії, які з надійністю 95% будуть мати значення для коефіцієнту:

- перетину $a \pm t_{кр}^{0,05;\, n-2} \times S_a = 16{,}44 \pm 2{,}011 \times 2{,}63 = 16{,}4 \pm 5{,}29;$
- факторної ознаки $b \pm t_{кр}^{0,05;\, n-2} \times S_b = 2{,}496 \pm 2{,}011 \times 0{,}37 = 2{,}50 \pm 0{,}74.$

XIV. Мультиколінеарність[46] причини виникнення та наслідки.

Мультиколінеарність у множинній регресії **виникає внаслідок** порушення припущення незалежності факторних ознак, тобто виникнення у них тісної залежності, наближеної до функціональної. Причиною мультиколінеарності є неякісний відбір незалежних факторних ознак рівняння, серед яких є повністю колінеарні, тобто такі, що дублюють одна одну.

Основні наслідки, до яких може привести мультиколінеарність[47] наступні:

1) гіпотеза щодо незначущості коефіцієнтів рівняння регресії за t–критерієм приймається, але рівняння регресії за F–критерієм виявляється значущим, що свідчить про завищення значення коефіцієнта множинної кореляції;

2) розраховані оцінки значень коефіцієнтів рівняння множинної регресії безпідставно завищені або мають не очікувані знаки їхнього зв’язку;

3) додаток або позбавлення одного-двох значень факторної ознаки сильно впливає на оцінку коефіцієнтів рівняння регресії;

4) наявність мультиколінеарності робить модель непридатною для застосування;

[46] Наявність сильної залежності між факторними ознаками моделі коли вони незалежні. Розрізняють **коллінеарність повну**, що означає наявність **функціональної** (між результативною та факторними ознаками) залежності і **часткову** або **просту** мультиколінеарність – наявність сильної кореляції між факторними ознаками.
[47] **Мультиколінеарність** факторних ознак спотворює результати розрахунків коефіцієнтів регресії.

Для виявлення мультиколінеарності розраховується коефіцієнт парної кореляції між двома незалежними факторними ознаками. Якщо значення коефіцієнту парної кореляції між факторними ознаками більше 0,8, тобто $r_{x_i x_j} > 0{,}8$ (в деяких джерелах >0,7), то існує мультиколінеарність.

Висновок про наявність мультиколінеарності двох факторних ознак можливо зробити за значенням **визначника їх матриці**. Якщо його значення мале, то в моделі регресії між цими ознаками існує мультиколінеарність.

Усунення мультиколінеарності.

Якщо модель використовується для вивчення зв'язків факторних та результативної ознак рівняння регресії, то мультиколінеарність факторних ознак підлягає усуненню. При цьому звертається увага на знаки при коефіцієнтах ознак у рівнянні та їх значення.

Існують такі способи усунення мультиколінеарності:

- збір додаткових вихідних даних (що має труднощі на практиці);
- перетворення значень факторних та результативної ознак у рівнянні, наприклад, замість фактичних значень беруть значення їхніх логарифмів: $lny = b_0 + b_1 lnx_1 + b_2 lnx_2 + \varepsilon$;
- у випадку неефективності наведених вище способів, використовують методи зміщеної оцінки невідомих коефіцієнтів або виключення з моделі одного з факторів, що дублюють одне одного.

Алгоритм включення факторних ознак у рівняння регресії представляє такі кроки:

- включають лише ті ознаки, які з результативною ознакою мають найбільшу парну кореляцію.

Середні та точкові коефіцієнти еластичності.

Для вивчення залежності між ознаками рівняння регресії, окрім коефіцієнтів **кореляції** та **детермінації** використовують коефіцієнти **еластичності**, які надають значення ступеня їхнього зв'язку.

Коефіцієнт еластичності показує, на скільки відсотків зміниться результативна ознака при зміні значення факторної ознаки на 1%.

Загальна формула еластичності має вигляд: $E_x(y) = \lim_{\Delta x \to 0}\left(\frac{\Delta y}{y} \div \frac{\Delta x}{x}\right) = \frac{x}{y} \times y'$, де y' – часткова похідна по факторній ознаці. Коефіцієнт еластичності для середнього значення «$\bar{i}$ої» факторної ознаки розраховується за формулою: $E_{x_i}(y) = \frac{\bar{x}_i}{y(\bar{x}_i)} \times y'_{\bar{x}_i}$.

Середній коефіцієнт еластичності показує наскільки відсотків зміниться результативна ознака відносно свого середнього значення при зміні значення факторної ознаки на один відсоток. Коефіцієнти еластичності для різних функцій розраховуються за індивідуальним формулами, наприклад, точкові коефіцієнти еластичності для функцій наведених нижче розраховуються за формулами табл. 3.

Таблиця 3

Функція	Вираз функції	Похідна функції	Коефіцієнти еластичності
Лінійна	$y = b_0 + b_1 \times x + \varepsilon$	$y' = b_1$	$E_x(y) = \frac{b_1 \times x}{b_0 + b_1 \times x + \varepsilon}$
Показникова	$y = b_0 \times b_1^x \times \varepsilon$	$y' = b_0 \times b_1^x \times lnb_1$	$E_x(y) = x \times Lnb_1$
Степенева	$y = b_0 \times x^{b_1} \times \varepsilon$	$y' = b_0 \times b_1 \times x^{b-1}$	$E_x(y) = b_1$
Обернена	$y = \frac{1}{(b_0 + b_1 \times x + \varepsilon)}$	$y' = \frac{-b_1}{(b_0 + b_1 \times x + \varepsilon)^2}$	$E_x(y) = \frac{b_1 \times x}{b_0 + b_1 \times x + \varepsilon}$
Логарифмічна	$y = b_0 + b_1 \times Lnx + \varepsilon$	$y' = \frac{-b_1}{x}$	$E_x(y) = \frac{-b_1}{b_0 + b_1 \times Lnx}$
Експоненціальна	$y = e^{a+bx}$	$y' = b \times e^{a+bx}$	$E_x(y) = b \times x$

Коефіцієнти еластичності використовують для аналізу виробничих функцій, але в **деяких випадках** вони не мають сенсу, тобто їх інтерпретація щодо зміни результативної ознаки у відсотках за зміною факторної ознаки неможлива або безглузда.

XV. Криві зростання.

Прогнозування моделей економічних процесів.

Криві зростання відображають закономірності розвитку процесів у часі та отримують шляхом аналітичного вирівнювання часових рядів. Вони, здебільш, однофакторні прогнозні моделі в яких фактором, що впливає на результативну ознаку є **час**. Вирівнювання ряду за допомогою функцій у більшості випадків виявляється зручним засобом опису емпіричних даних, що характеризують розвиток процесу у часі. Використанню кривих зростання передує ґрунтовний аналіз процесу для з'ясування можливості екстраполяції[48] його тенденцій.

Криві зростання часто використовуються у дослідженні динаміки реальних процесів різного походження, наприклад: «цикл виробництва і просування на ринок товару», «накопичення капіталу», «розповсюдження інформації маркетингових послуг», тощо. Їх застосовують при аналізі міграційних та інших процесів зростання у людському суспільстві чи інших.

Аналітичне вирівнювання складається з наступних етапів:

- **вибір типу кривої** форма якої відповідає характеру зміни часового ряду;
- **визначення чисельних значень** коефіцієнтів (параметрів) цієї кривої.

Знайдена функція дозволяє отримати вирівняні рівні ряду. Вибір типу кривої передбачає знайомство з основними видами кривих та вивчення їх властивостей.

Інтерес представляють **перетворення приростів**, які можна представити у вигляді лінійної функції. Ці характеристики використовуються при виборі виду кривої зростання.

Економічна практика вже має певні типи кривих, які найчастіше використовують в подібних дослідженнях. До них належать криві, функції яких мають аналітичний вираз:

a) поліномів; **b)** експонент; **c)** логістичних;
d) обернених; **e)** степеневих.

[48] особливий тип наближення, при якому функція апроксимується поза заданого інтервалу.

А). Поліноми:

- $y(t) = a + a_1 t$ – першого степеня;
- $y(t) = a + a_1 t + a_2 t^2$ – другого степеня;
- $y(t) = a + a_1 t + a_2 t^2 + a_3 t^3$ – третього степеня.

Коефіцієнт «a_1» називають лінійним приростом, коефіцієнт «a_2» – прискоренням зростання, коефіцієнт «a_3» – зміною у прискоренні зростання.

Для полінома першого степеня характерний **постійний закон зростання**. Якщо розрахувати перші прирости за формулою: $U_t^{(1)} = y(t_i) - y(t_{i-1})$ де i = 2, 3,…*n*, то вони матимуть **постійні значення**, що дорівнюватимуть значенню «a_1».

Для полінома другого степеня перші прирости (різниці) мають лінійну залежність у часі, а ряд з перших приростів $U_t^{(1)}(t_{i=2})$, $U_t^{(1)}(t_{i=3})$, ... на графіку представляє пряму лінію. **Постійними** для нього будуть другі прирости $U_t^{(2)} = U_t^{(1)} - U_{t-1}^{(1)}$.

Для полінома третього степеня перші прирости будуть поліноми другого ступеня, другі прирости будуть лінійною функцією часу, а **треті прирости**, що розраховані за формулою, $U_t^{(3)} = U_t^{(2)} - U_{t-1}^{(2)}$ будуть **постійною** величиною.

За викладеним маємо наступні **властивості** поліноміальних **кривих зростання**:

- шляхом розрахунку послідовних різниць (приростів) можна перейти від полінома вищого порядку до поліному нижчого порядку;
- значення приростів для поліномів будь-якого порядку не залежать від значень самої функції.

Висновок. **Поліноміальні криві зростання** використовують для апроксимації (наближення) і прогнозування економічних процесів в яких їхній подальший розвиток не залежить від досягнутого рівня.

Експоненційні криві зростання використовують коли подальший розвиток процесів залежить від вже досягнутого рівня, наприклад, приріст залежить від значення функції. В економіці найчастіше застосовуються два різновиди показникових кривих:

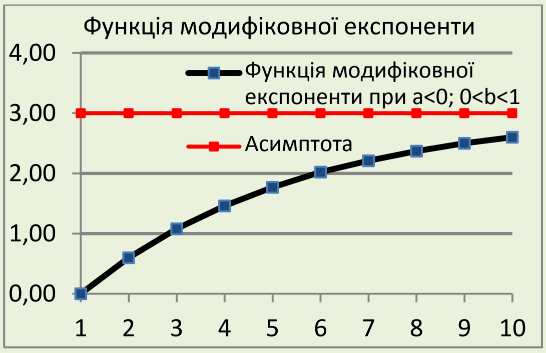
- проста експонента і модифікована експонента.

В). Експоненти.

Проста експонента має вираз: $y(t) = \alpha \times b^t$, де «$a>0$», «*b>0*». Коли «*b>1*» функція зростає за часом «*t*», якщо «*b<1*» спадає. Важливо також, що ордината функції змінюється з постійним темпом приросту. Якщо взяти відношення приросту до самої ординати, то воно буде постійним: $\frac{U_t}{y_t} = \frac{y_t - y_{t-1}}{y_t} = \frac{1}{b}$. Логарифмуючи функцію $y(t) = \alpha \times b^t$ за натуральним логарифмом отримаємо: $ln\, y(t) = ln\, \alpha + t \times ln\, b$, звідки логарифм результативної ознаки (ординати простої експоненти) має лінійну залежність у часі.

Модифікована експонента має вираз: $y(t) = \alpha_0 + \beta \times b^t$ і представляє сімейство кривих логарифмічної параболи $y(t) = \alpha_0 b^t c^{t^2}$, що найпоширена коли коефіцієнти мають значення: «$\boldsymbol{\beta}$<0», 0<*b*<1, а «α_0» є асимптота функції до якого наближаються її значення.

Можуть бути інші варіанти модифікованої експоненти, але на практиці найбільш часто зустрічається зазначені функції. Якщо логарифмувати перші її прирости функції, то отримаємо функцію, що лінійно залежить від часу, а якщо взяти відношення двох послідовних приростів $\frac{U_t}{U_{(t-1)}} = \frac{y_t - y_{(t-1)}}{y_{t-1} - y_{(t-2)}} = b$, то воно буде постійною величиною.



Експоненціальні криві використовують для опису процесів, що швидко зростають або спадають. Для розрахунку невідомих коефіцієнтів експоненціальної кривої, функція шляхом логарифмування перетворюється за формою до лінійного її виду, що надає змогу застосувати метод найменших квадратів для аналізу моделі, як в моделі лінійної регресії.

Модель природного зростання представляє собою диференційне рівняння з виразом $y'(t) = k \times y(t)$, рішенням якого є **експоненціальна функція**, що показує швидкість зростання обсягів реалізованої продукції за постійних інвестицій у виробництво та умови ненасичення ринку. Процеси, в яких результативна величина зростає за рівні інтервали часу в одну і ту ж кількість разів, мають назву «процесів природного зростання».

Ця модель є **функцією корисності** для індивідуумів несхильних до ризиків, що пояснює їх ставлення до корисності майна за еквівалентом його ринкової вартості упродовж строку експлуатації за ресурсом часу предмета оцінки.

1). Моделі природного зростання[49] обсягів в умовах ненасичення ринку (або модель **монопольного ринку**).

Надамо методологію побудови рівняння за моделями процесів природного зростання.

2. Нехай $y(t)$ – обсяг продукції галузі реалізований на момент часу «*t*».
3. Вважатимемо, що продукція негайно реалізується по фіксованій ціні «*p*», тобто існує ненасичення ринку.
4. Тоді на момент часу «*t*» дохід складатиме:

$$Y(t) = p \times y(t) \qquad (1)$$

5. Оскільки підприємство отримує прибуток тривалий час, тому йому вигідно розширювати обсяги виробництва, направляючи інвестиції $I(t)$ на розширення виробництва.
6. Оскільки величина інвестиції складає фіксовану частину доходу, маємо рівняння:

$$I(t) = m \times Y(t) = m \times p \times y(t), \quad 0 < m < 1 \qquad (2),$$

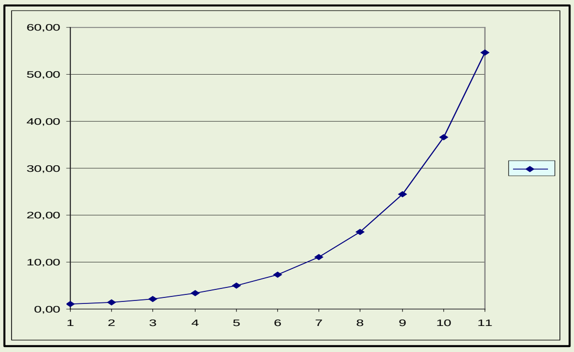
де «0<*m*<1» – є частка доходу за нормою інвестиції, величина якої постійна.

[49] В нашому випадку обсягу реалізованої продукції.

Інвестиція сприяє зростанню обсягу випуску продукції, а швидкість цього зростання буде пропорційна зростанню інвестиції.

У моделі природного зростання вважають, що швидкість зростання обсягу продукції, чи акселерація[50], пропорційна розміру інвестицій з коефіцієнтом «l», $y'(t) = l \times I(t)$ (3), тут ми нехтуємо періодом часу між її виготовленням та реалізацією. Підставляючи з (2) вираз для $I(t)$ у (3) отримуємо рівняння $y'(t) = m \times p \times l \times y(t) = k \times y(t)$[51], із змінними, що розділяються, де $k = m \times p \times l$, та розв'язуючи його маємо: $y(t) = y(t_0) \times e^{k\times(t-t_0)}$, де $y(t_0)$ початкове значення. Рівняння описує динаміку зростання цін при постійній інфляції.

Умова ненасичення ринку може бути прийнятна лише для вузького часового інтервалу. Сенс рівняння $y'(t) = k \times y(t)$ полягає в тому, що швидкість зміни функції в точці у часі (t) пропорційна значенню самій функції в цій точці.



Графік функції $y = C \times e^{kx}$

C). **Логістичні криві**.

2). Моделі природного зростання обсягів в умовах насичення ринку.

Кьютелет припустив, що за умови насичення ринку ціна «**p**» залежить від обсягу реалізованої продукції $y(t)$. З падінням ціни на продукцію $\frac{dP}{dY} < 0$, обсяг реалізації збільшується, а коефіцієнт k у рівнянні $y'(t) = k \times y(t)$ є спадною функцією. Його учень Ферхюльст припустив, що $k = l \times \left(1 - \frac{y(t)}{b}\right)$, а обсяг реалізованої продукції $y(t)$ в умовах насичення ринку описується рівнянням $y'(t) = l \times \left(1 - \frac{y(t)}{b}\right) \times y(t)$, та з розділенням змінних $\frac{dy}{y(t)\times\left(1-\frac{y(t)}{b}\right)} = l \times dt$.

[50] **Акселерація** – термін у макроекономічній теорії, що означає пришвидшення; принцип випереджуючого (прискореного) впливу зростання сукупного попиту на динаміку зростання капіталовкладень (перш за все в області основного капіталу). Механізм акселерації випливає з своєрідності основного капіталу, який представлений виробничими цінностями тривалого користування. При незмінності сукупного попиту не існувала б і потреба у вільних інвестиціях, але при зростанні споживчого попиту ці інвестиції, необхідні для розширення виробництва, що має зростати швидше, ніж попит, оскільки вони виступають у вигляді довгострокових інвестиційних товарів, що створюють базу для задоволення підвищеного попиту. Цей принцип дійсний і в «зворотному» напрямку: якщо темп зростання споживчого попиту сповільниться, то відносне падіння інвестиційного попиту буде стрімкішим. Таким чином, з точки зору принципу акселерації вільні інвестиції є тільки функцією «чистого» приросту доходу, що стимулює зростання сукупного попиту. Але це означає, що принцип акселерації реальний лише для тієї частини інвестицій, яка викликана дійсним (або очікуваним) зростанням споживчого попиту, не охоплюючи так звані «автономні» інвестиції.

[51] Математична модель Мальтуса (модель Бернуллі при кредитуванні) є моделлю природного зростання.

Зробивши незначні перетворення рівняння до табличного інтегралу маємо: $\frac{bdy}{-\left[\left(y^2-by+\left(\frac{b}{2}\right)^2-\left(\frac{b}{2}\right)^2\right)\right]}=l\times dt$; та $\frac{bdy}{-\left(\left(y-\frac{b}{2}\right)-\left(\frac{b}{2}\right)^2\right)}=l\times dt$, інтегруючи яке отримаємо

$$\frac{b}{2\times\frac{b}{2}}ln\left|\frac{\left(y-\frac{b}{2}\right)-\frac{b}{2}}{\left(y-\frac{b}{2}\right)+\frac{b}{2}}\right|=-l\times t+C_1,$$ або остаточно $ln\left|\frac{y-b}{y}\right|=-l\times t+C_1$.

Логарифмуючи отримане рівняння маємо його рішення $y(t)=\frac{b}{1+C\times e^{-(l\times t)}}$, яке описує зростання реалізованої продукції на ринку, що насичується, де $C = \pm EXP(C_1)$.

В економіці поширені процеси, що спочатку зростають повільно, а потім прискорюються, і нарешті уповільнюють своє зростання, прагнучи дістати певної межі. Як приклад, можна привести цикл реалізації обсягів товару з початку його виробництва до завершення, динаміку зміни цін попиту, обсягу товарів, що сягають рівня насичення, та інші. Для моделювання використовуються функції, що мають назву кривих зростання, серед яких виділяють ***S*–образні**, а саме логістичні криві **Гомперца** та **Перла-Ріда**, а також інші, що описують моделі динамічних процесів конкурентного ринку.

В умовах конкурентного ринку модель зростання, залишаючись диференційним рівнянням із змінними, що розділяються, набирає вираз:

$$y'=m\times l\times p(y)\times y\ldots \qquad (4).$$

Оскільки співмножники у правій частині рівняння позитивні, то перша похідна $y'>0$ описує зростаючу функцію. При дослідженні на опуклість відшукується також еластичність функції. Отже з (4) випливає:

$$y''=m\times l\times[p(y)\times y]'\times\left[\frac{dp(y)}{dy}\times\frac{y}{p(y)}+1\right] \qquad (5),$$ [52]

та оскільки $E_p(y)=\frac{dy}{y}\div\frac{dp}{p}=\frac{p\times dy}{y\times dp(y)}$ еластичність (5) має вираз:

$$y''=m\times l\times y'\times p(y)\times\left(\frac{1}{E_p(y)}+1\right),$$ а умова $y''=0$ надає $E_p(y)=-1$.

Таким чином, якщо попит еластичний $\left|E_p(y)\right|>1$, то $y''>0$, а функція опукла[53] до низу (увігнута), а якщо попит не еластичний $\left|E_p(y)\right|<1$, то $y''<0$, а функція опукла вгору або просто «опукла».

---

[52] $y''=m\times l\times[p(y)\times y]'=m\times l\times[p'(y)\times y'\times y+p(y)\times y']=m\times l\times y'\times p(y)\times\left[\frac{dp(y)}{dy}\times\frac{y}{p(y)}+1\right]$.

[53] Опукла крива має форму пагорба, увігнута крива форму западини («впадины» – рос.), окрім того для другої похідної виконується «правило дощу» – коли похідна позитивна вода наповнюється (увігнута крива), а для негативної похідної розтікається (опукла крива).

П р и к л а д.

1. Крива попиту задана рівнянням $p(y) = (4 - y)$.
2. Норма акселерації (мультиплікатор) $l = 2$.
3. Норма інвестицій $m = 0{,}5$.
4. Обсяг продукції $y(t_0) = 0{,}5$.

Запитання: Знайти вираз функції для обсягу реалізованої продукції?

Рішення.

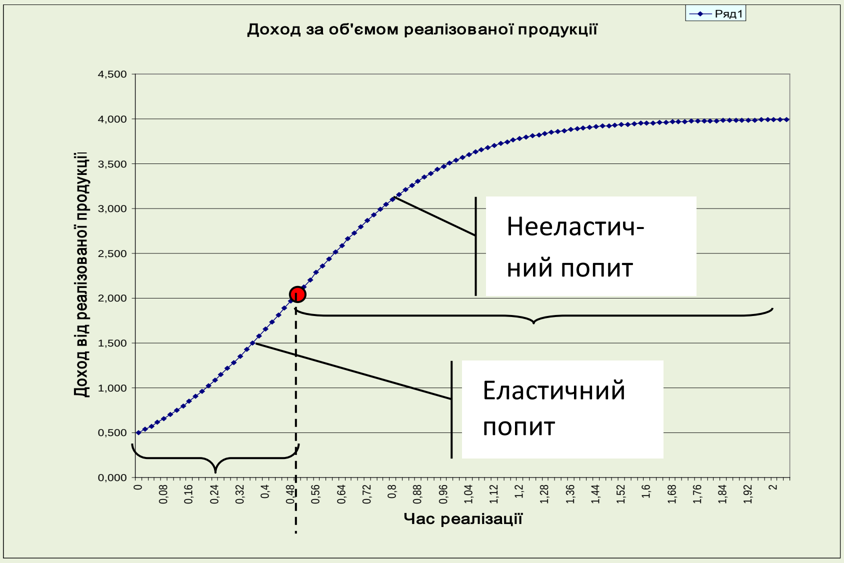
Рівняння швидкості зростання випуску продукції в моделі природного зростання набирає вигляду $y' = m \times l \times p(y) \times y = 0{,}5 \times 2 \times (4 - y) \times y$, або з розподіленням змінних $\frac{dy}{(4-y)\times y} = dt$. Виконуючи інтегрування обох частин маємо: $\int \frac{dy}{(4-y)\times y} = \int dt$; $\Longrightarrow$

$$\int \frac{dy}{(y^2-4y+4)-4} = \int dt, \Longrightarrow -\int \frac{dy}{(y-2)^2-2^2} = \int dt \Longrightarrow -\frac{1}{2\times 2} ln\left|\frac{(y-2)-2}{(y-2)+2}\right| = t + C_1$$

$$ln\left|\frac{y-4}{y}\right| = -4t + C_1 \Longrightarrow \frac{y-4}{y} = C \times e^{-4t}$$, де C = $\pm e^{C_1}$=±EXP(C_1).

Враховуючи, що, $y(t_0) = 0{,}5$, отримуємо С= –7, і остаточно маємо вираз для обсягу реалізованої продукції: $y(t) = \frac{4}{1+7\times EXP(-4t)}$. Графік доходу за обсягом реалізованої продукції, див. нижче.

Логістична крива, є зростаюча функція, що має вид: $y(t) = \frac{k}{1+\alpha\times e^{-(b\times t)}}$.



Якщо взяти похідну цієї функції, то можна побачити, що швидкість зростання логістичної кривої в кожний момент часу пропорційна досягнутому рівню функції, різниці між граничним значенням «*k*» і досягнутим рівнем. Логарифм відношення першого приросту функції до квадрату її значення (ординати) є лінійною функцією від часу.

Конфігурація графіка логістичної кривої близька до графіку кривої Гомперца, але логістична крива має точку симетрії, що збігається з точкою перегину.

Невідомі коефіцієнти експоненціальної кривої розраховують за методом найменших квадратів за лінійним аналогом цієї функції, що попередньо перетворена шляхом логарифмування, наприклад:

$$y = ab^x \qquad \ln y = \ln a + x \ln b$$
$$y = ae^{bx} \qquad \ln y = \ln a + bx$$
$$y = ae^{b_0 + b_1 x} \qquad \ln y = \ln a + b_0 + b_1 x$$

D). <u>Обернена крива</u>.

Узагальнена вибіркова модель оберненої кривої має вид: $y_i = a + b\frac{1}{x_i} + \varepsilon_i$, вона нелінійна за факторною змінною, але лінійна за коефіцієнтами рівняння «a» та «b», а тому є лінійною регресійною моделлю, і дійсно, позначивши $\frac{1}{x_i} = z_i$ отримаємо рівняння:

$$y_i = a + bz_i + \varepsilon_i.$$

Модель має особливості коли $x \to \infty$, значення результативної змінної прямує до граничного його значення. Графік моделі в значній мірі залежить від знака коефіцієнтів рівняння «a» та «b»[54]. Нахил моделі визначається за формулою: $\frac{dy}{dx} = -b\frac{1}{x^2}$. Він є додатним коли $b < 0$, і від'ємним коли $b > 0$.

E). <u>Степенева крива</u>.

Степенева функція є однією з найпоширених функцій кривих зростання, що описує широкий спектр динамічних процесів економіки. Загальний вид степеневої кривої $y = ax^b$. Найчастіше розглядають випадки коли $a \geq 0$, що типові для динамічних процесів економіки. Якщо значення коефіцієнта «b» не ціле число, то розглядають випадки коли $x \geq 0$, при цьому, в залежності від знака цього коефіцієнта, степенева функція описуватиме різні динамічні процеси, такі як: прискорене зростання, уповільнене зростання, а також спад, тощо. При значенні коефіцієнта $b = 1$, функція перетворюється в лінійну.

Питання полягає у розрахунку невідомих коефіцієнтів мультиплікативної (степеневої) кривої, що вирішується шляхом логарифмічного перетворення, наприклад:

$$\ln y = \ln ax^b = \ln a + b \times \ln x.$$

Найбільш відомими степеневими функціями – є виробничі функції Кобба-Дугласа або Соллоу. Крім того, їх використовують, як криві байдужості чи криві попиту на товари різних категорій (так звані криві Торнквіста, тощо).

[54] Пам'ятаємо, що для рівняння побудованого на даних вибіркової сукупності, числа «a» та «b», мають назву «коефіцієнтів регресії» на відміну від «параметрів регресії» за їхньою назвою у генеральній сукупності

Задача. Розглянемо динаміку зростання грошової маси за 25 місяців, табл. 4

Таблиця 4.

Місяць (X)	Грошова маса (У)	Місяць (X)	Грошова маса (У)	Місяць (X)	Грошова маса (У)
1	1863,03	10	5314,33	19	7086,18
2	2244,71	11	5931,01	20	7186,88
3	2810,83	12	6212,53	21	7339,33
4	3154,65	13	6386,32	22	7522,33
5	3218,28	14	6232,02	23	7861,00
6	3643,18	15	6234,57	24	7540,00
7	3935,35	16	6846,33	25	7592,00
8	4244,73	17	6514,33	26	Прогнозне значення?
9	4702,86	18	6805,48	?	

Для визначення прогнозу на 26-й місяць треба надати мультиплікативну модель зростання грошової маси виду $y = ax^b$, де результативна ознака грошова маса, а факторна ознака місяці. Для розрахунку невідомих коефіцієнтів функції перетворимо степеневу функцію, шляхом логарифмування, на лінійну виду: $ln\, y = ln\, a + b\, ln\, x$, а значення лінеаризації за логарифмами результативної та факторної ознаки зведемо в табл.5, де *Ln*(Y) логарифм функції грошової маси.

Таблиця 5.

Ln(X) місяців де X=1,2,…,9	*Ln*(Y) грошової маси	*Ln*(X) місяців, X=10,11…18)	*Ln*(Y) грошової маси	*Ln*(X) місяців, X=19,20…25	*Ln*(Y) грошової маси
Ln(1)=0,00	*Ln*(1863,03)=7,53	*Ln*(10)=2,30	8,58	*Ln*(19)=2,94	8,87
Ln(2)=0,69	*Ln*(2244,71)=7,72	2,40	8,69	3,00	8,88
Ln(3)=1,10	*Ln*(2810,83)=7,94	2,48	8,73	3,04	8,90
Ln(4)=1,39	*Ln*(3154,65)=8,06	2,56	8,76	3,09	8,93
Ln(5)=1,61	*Ln*(3218,28)=8,08	2,64	8,74	3,14	8,97
Ln(6)=1,79	*Ln*(3643,18)=8,20	2,71	8,74	3,18	8,93
Ln(7)=1,95	*Ln*(3935,35)=8,28	2,77	8,83	3,22	8,93
Ln(8)=2,08	*Ln*(4244,73)=8,35	2,83	8,78		
Ln(9)=2,20	*Ln*(4702,86)=8,46	2,89	8,83		

В директорії «Дані» програми EXCEL за піктограмою «аналіз даних» графа «регресія» здійснюємо аналіз за вхідним масивом логарифмів вихідних даних грошової маси та місяців, результат дисперсійного аналізу масиву програмою EXCEL виведемо в табл. №6.

Таблиця 6.

Регресійна статистика						
Множинний коефіцієнт кореляції (R)			0,98836			
Коефіцієнт детермінації R^2			0,97672			
Нормований коефіцієнт детермінації R^2			0,97574			
Стандартна помилка			**0,06464**			
Спостереження			25			
Дисперсійний аналіз	*df*	*SS*	*MS*	***F**-стат.*	$\boldsymbol{F}_{критичне}$	$t_{кр}$(0,05;23)
Регресія	1	4,0391	4,0391	966,59	**4,279**	**2,0686**
Залишок	23	0,0961	0,0042			
Разом	24	4,1352				
Коефіцієнти:	*Значення коефіцієнт*	*Стандартна помилка*	*t-статистика*	*Р-значення*	*Нижні 95%*	*Верхні 95%*
Y-перетину $(ln\, a)$	7,407	0,039	190,369	0,000	7,326	7,487
Факторної ознаки X_1	0,492	0,016	31,090	0,000	0,459	0,524

У розділі «формули» програми EXCEL підгрупа інші формули знаходимо критичне значення $\boldsymbol{F}_{кр.}$ Фішера для ступенів вільності n=1; n=23, ймовірності P=95%, $\boldsymbol{F}_{кр.}$(1;23; 95%)=4,279.

Знаходимо критичне значення критерію Стьюдента $t_{кр.(0,05;\ 23)}$=2,0686 та значення коефіцієнтів b=0,492 та $ln\, a = 7{,}407$, звідки $a = e^{7{,}407}$, при значенні коефіцієнту Фішера F=966,6, що значно більше $F_{кр}$=4,279 його критичного значення. Звідси модель рівняння регресії адекватна. Значення t-критерію Стьюдента коефіцієнтів рівняння також перевищують їхні критичні значення, що вказує на значущість коефіцієнтів рівняння регресії. Прогнозне точкове значення грошової маси на 26-й місяць матимемо з рівняння:

$$y = ax^b = e^{7{,}407} \times 26^{0{,}492} = 8184{,}33\,.$$

При стандартній помилці $\sigma_\varepsilon = 0{,}06464$ остаточне значення в межах довірчого інтервалу становитиме: $y = ax^b \pm t_{\frac{\alpha}{2};n-2} \times \sigma_\varepsilon \times \sqrt{\frac{n+1}{n}} = 8184{,}33 \pm 2{,}069 \times 0{,}06464 \times \sqrt{\frac{24+1}{24}}$, з деяким обмеженням точності формули розрахунку, що не вплинуло на довірчий інтервал, або $8184{,}33 \pm 0{,}14$.

XVI. Ковзкі середні.

Поняття про ковзкі середні.

Найпоширений спосіб виявлення тенденції виробництва є згладжування часового ряду. Ідея полягає у заміні фактичних значень часового ряду розрахунковими значеннями тренду. Найпростіших спосіб отримання значень тренду ряду – використання ковзких середніх.

Ковзка середня це термін усереднених значень часового ряду. Іншими словами, ковзка середня є лінія середніх значень, що розраховані за певний інтервал часу. **Ковзка середня** є **трендовим індикатором**. За допомогою якого можна відстежити початок нового тренду (зростання чи спаду) або завершення поточного, а по куту його нахилу можна судити про стрімкість тренду.

Ковзка середня[55] – назва сімейства функцій, значення яких в точці їх визначення дорівнюють середньому значенню вихідної функції за попередній період. Ковзкі середні зазвичай розраховують за даними часових рядів для згладжування короткострокових коливань значень ряду і виділення основних тенденцій чи циклів. Математично ковзке середнє це вид згорнутої функції, що використовується для визначення загального напряму процесу.

Надамо пояснення суті ковзких середніх. Нехай маємо часовий ряд: $y_1, y_2, \ldots y_k, y_{k+1}, \ldots y_n$. Для кожних «$k$» значень $y_1, y_2, \ldots., y_k$ цього ряду, де «$k<n$», можна розрахувати середнє. Обчисливши перше середнє значення з «k» значень $y_1, y_2, \ldots., y_k$, переходимо до розрахунку середньої з наступних «k» значень: $y_2, y_3, \ldots., y_{k+1}$, а далі послідовно із значень $y_3, y_4, \ldots., y_{k+2}$, і так до кінця ряду. Таким чином інтервал згладжування, тобто інтервал для якого розраховується середнє значення ніби ковзає вздовж значень часового ряду. Звідси і назва.

Незручність полягає у розміщені розрахункових даних у центрі інтервалу згладжування.

[55] англ. moving average

Якщо кількість «*k*» значень *для розрахунку* ковзкої середньої є непарною, то дати ковзкої середньої співпадають з датами значень часового ряду, якщо парними, то дати середньої розташовані між датами фактичних значень.

Кількість значень, які входять в інтервал згладжування «*k*» це **параметр довжини згладжування ковзкого середнього**, тобто кількість періодів, наприклад, годин, днів, місяців, тощо, іноколи його називають **порядком** (або рівнем) ковзкої середньої, наприклад, якщо інтервал згладжування складається з «*k*» значень часового ряду, то маємо ковзку середню «*k*»-порядку.

За методами побудови існує кілька видів **ковзкої середньої**:

- Проста (Simple moving average – SMA).
- Лінійно-зважена (Linear-Weighted moving average – WMA).
- Експонентна (Exponential moving average – EMA).
- Згладжена (Smoothed).

Розглянемо аналітичний вираз простої ковзкої середньої.

Проста ковзка середня хоча і є примітивним індикатором, але вона є базовим індикатором технічного аналізу, основа для багатьох торгових стратегій і інших індикаторів, тому знати принцип роботи цього індикатора зобов'язаний кожен трейдер[56].

Просте ковзке середнє або арифметичне ковзке середнє[57] «$k^{го}$-порядку» часового ряду у точці $t = t_1;\ t_2;\ t_3 \ldots$ чисельно дорівнює середньому арифметичному із значень результативної ознаки за встановлений інтервал згладжування за *k*–рівнів значень і обчислюється як:

$\bar{y}_{SMA(t)} = \frac{y_1+y_2+\cdots+y_{t-2}+y_{t-1}+y_t+y_{t+1}+y_{t+2}+\cdots+y_k}{k}$, див. графік нижче.

Графік простої ковзкої кривої.

З підвищенням порядку згладжування ковзка крива стає гладкішою.

[56] Торговець, спекулянт.
[57] англ. Simple moving average, абревіатура SMA.

Проте може бути ситуація коли ковзка середня 12-порядку є гладкішою ніж ковзка середня вищого, наприклад, 15 порядку, що пов'язано з ефектом сезонності часового ряду, причому період сезонних коливань збігається з інтервалом згладжування.

Якщо «*k*» не парне то його можна подати як «*k=2p+1*», де *p*=1,2,3,…, а ковзка середня розраховуватиметься за формулою: $\bar{y}_{SMA(t)} = \frac{y_{t-p}+y_{t-p+1}+\cdots+y_t+y_{t+1}+\cdots+y_{t+p}}{k}$, і дати ковзкої середньої співпадають з датами спостережень. Коли ж «*k*» парне, то дати ковзкої середньої розташовані між датами фактичних значень.

Властивості ковзкої середньої.

1. Ковзку середню можливо обчислити за рекурентною формулою[58] із співвідношенням рекурентності. Нехай «*k*» не парне, тоді маємо:

$$\bar{y}_{SMA(t)} = \frac{y_{t-p}+y_{t-p+1}+\cdots+y_t+y_{t+1}+\cdots+y_{t+p}}{k}; \text{ та}$$

$$\bar{y}_{SMA(t+1)} = \frac{y_{t-p+1}+y_{t-p+2}+\cdots+y_{t+1}+y_{t+2}+\cdots+y_{t+p+1}}{k} =$$

$$= \frac{(y_{t-p}+y_{t-p+1}+y_{t-p+2}+\cdots+y_{t+1}+y_{t+2}+\ldots+y_{t+p}) - y_{t-p}+y_{t+p+1}}{k};$$

Звідки маємо $\bar{y}_{SMA(t+1)}=\bar{y}_{SMA(t)} - \frac{y_{t-p}-y_{t+p+1}}{k}$.

2. При згладжуванні часового ряду ковзкою середньою «*k*-порядку» відбувається втрата «*k-1*» її значень.
3. Якщо «*k*» не парне, то дати ковзкої середньої збігаються з датами спостережень, коли парне, то дати ковзкої середньої розташовані проміж дат реальних спостережень.

Розглянемо приклад загладжуваного часового ряду ковзкою середньою.

У табл. 7 наведено щомісячні дані про продаж м'ясної продукції за цінами тони упродовж року. Розрахуємо середні ціни за ковзкою середньою 4-го порядку, коли «*k*» парне.

Таблиця 7.

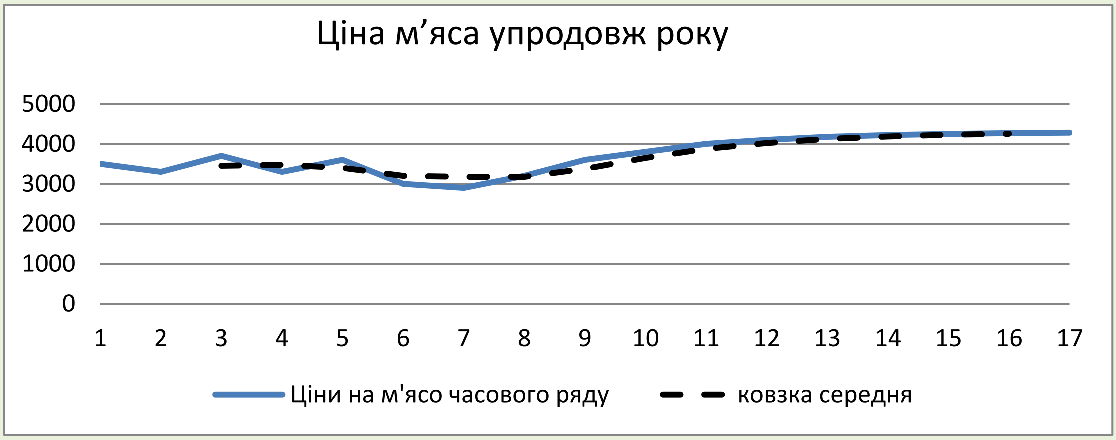
Дата	15 Січ	15 Лют	15 Бер	15 Кві	15 Тра	15 Чер	15 Лип	15 Сер	15 Вер	15 Жов	15 Лис	15 Гру
Ціна	**3500**	**3300**	**3700**	**3300**	**3600**	**3000**	**2900**	**3200**	**3600**	**3800**	**4000**	**4100**

Дата SMA	01 Бер	01 Кві	15 Тра	01 Чер	01 Лип	01 Сер	01 Вер	01 Жов	01 Лис	-
Значення SMA	***3450***	*3475*	*3400*	*3200*	*3175*	*3175*	*3375*	*3650*	*3875*	-

Суму перших чотирьох значень січнем-квітнем (3500+3300+3700+3300) ділимо на 4 та отримуємо перше значення ковзкої середньої – 3450, яке розташовуємо за датою у середині цих чотирьох дат, тобто 1 березня. Довжина згладженого ряду ковзкої середньої зменшена на «*k* – 1 =4 – 1=3» і складає (*n – k+1)*=12 – 4+1=9 значень.

Проілюструємо просту ковзку середню розраховану у табл.7 графіком нижче.

[58] Формула виду, що виражає кожен член послідовності через «*p*» попередніх членів.



Ковзкі середні використовують для прогнозування, при цьому «*k*» останніх значень згладжених ковзкою середньою часового ряду і є прогнозом для наступної дати, див. табл. 8.

Таблиця 8.

Дата ряду часового	Фактичні дані	Дата ковзкої середньої	Ковзка середня SMA 4-го порядку	Центрова ковзка SMA 4-го порядку	Відносна помилка
15.Січ.2014	3500				
15.Лют.2014	3300				
		01.Бер.2014	3450		
15.Бер.2014	3700			3463	6,41%
		01.Кві.2014	3475		
15.Кві.2014	**3300**			**3438**	**4,18%**[59]
		01.Тра.2014	3400		
15.Тра.2014	3600			3300	8,33%
		01.Чер.2014	3200		
15.Чер.2014	3000			3188	6,27%
		01.Лип.2014	3175		
15.Лип.2014	2900			3175	9,48%
		01.Сер.2014	3175		
15.Сер.2014	3200			3275	2,34%
		01.Вер.2014	3375		
15.Вер.2014	3600			3513	2,42%
		01.Жов.2014	3650		
15.Жов.2014	3800			3763	0,97%
		01.Лис.2014	3875		
15.Лис.2014	4000			**3947**	1,33%
		01.Гру.2014	**4016**		
15.Гру.2014	4100			**4072**	0,68%
		01.Січ.2014	**4124**		
15.Січ.2015	**4175**[60]			**4155**	0,48%
		01.Лют.2014	**4186**		
15.Лют.2015	**4219**			**4207**	0,28%
		01.Бер.2014	**4228**		
15.Бер.2015	**4249**			**4241**	0,19%
		01.Кві.2014	**4254**		**Середнє**
15.Кві.2015	**4268**				**3,3%**
15.Тра.2015	**4280**				

[59] $|3300 - 3438|/3300 = 4{,}18\%$;
[60] $4100 + 1/4 \times (4100 - 3800)$.

Зважені ковзкі середні.

Недолік простих ковзких середніх полягає в тому, що рівні, які осереднені однаково впливають на отриманий результат незалежно від їх віддалення у минуле. Якщо в інтервал осереднення потрапить будь-який аномальний рівень значень, що значно відрізняється від інших у будь-яку сторону, змінна середня відреагує на нього двічі: **перший раз** – коли рівень увійде в інтервал осереднення, і **другий** раз – коли вийде з нього. Проте, тільки в першому випадку висновки про можливу зміну тренда будуть мати під собою підставу. Цього **недоліку позбавлені** зважені ковзкі середні (weighted moving average – WMA(k), англ.), а також експоненціальні середні (exponential moving average – EMA(k), англ.) при розрахунку значень яких більші вагові коефіцієнти надаються значенням, що наближені до теперішнього часу. Порівняємо значення простої та зваженої ковзкої середньої, табл.9, 10.

Таблиця 9.

Дата часового ряду	Ціна одиниць	∑5рівня за 26/04-07/05	∑5 рівня за 29/04-12/05	∑5 рівня за 05/05-13/05	∑5 рівня за 06/05-17/05	Проста ковзка середня 5 рівня
26.Кві.2014	70,2					(SMA)
29.Кві.2014	70,3					
05.Тра.2014	69,0	345,1				69,0
06.Тра.2014	67,3		344,9			69,0
07.Тра.2014	68,3			347,3		69,5
12.Тра.2014	70,0				354,4	70,9
13.Тра.2014	72,7				365,7	73,1
14.Тра.2014	76,1				371,4	74,3
16.Тра.2014	78,6					
17.Тра.2014	74,0					

Найпростіший варіант системи зважування полягає у використанні чисел натурального ряду, тобто значення рівня інтервалу осереднення приймають вагу: перше – 1, друге – 2, третє – 3, і так далі, останнє значення приймає вагу «*k*». Тоді формула розрахунку зваженої ковзкої середньої матиме наступний вигляд: $\bar{y}_{WMA(t)} = \frac{\sum_{\substack{i=1\\j=1}}^{k} y_i \times w_{ij}}{\sum_{j=1}^{k} w}$.

Вага значень може надаватися за лінійною, логарифмічною чи іншою функцією. При цьому значенням цін, що за датами наближені до поточних, тобто на тепер, надається більша вага. Розрахунок здійснюється за середньою арифметичною зваженою.

Таблиця 10.

Дата	Ціна	Ваги значень рівнів w_{ij}						∑ Сума 5 рівня	Ковзка зважена середня[61] 5 рівня
		w_{1j}	w_{2j}	w_{3j}	w_{4j}	w_{5j}	w_{6j}		
26.Кві	70,2	j=1							
29.Кві	70,3	j=2	1						
05.Тра	69,0	j=3	2	1				1028,5	68,6
06.Тра	67,3	j=4	3	2	1			1033,4	68,9
07.Тра	68,3	j=5	4	3	2	1		1052,0	70,1
12.Тра	70,0		5	4	3	2	1	1085,2	72,3
13.Тра	72,7			5	4	3	2	1123,8	74,9
14.Тра	76,1				5	4	3	1128,1	75,2
16.Тра	78,6					5	4		
17.Тра	74,0						5		
Разом ∑		15							

[61] WMA

Використання ваг робить середню зважену більш чутливою до намічених змін у тенденції і прискорює подачу сигналів змін. За даними табл. 9 та 10 порівняємо дві перші прості та зважені ковзкі середні. За ковзкими зваженими 5-6 травня з'явилася тенденція зростання ціни, але прості ковзкі середні однакові. Зважена ж середня "відчула" зародження зростаючого тренду та зросла на 0,3 одиниці.

Експоненціальна ковзка середня[62]

Методи експоненціального згладжування виникли для одночасного прогнозування руху великої кількості складських товарів. Простота та ефективність зробили ці методи згладжування найбільш поширеними. Серед них є методи простого, подвійного та потрійного згладжування такі, як методи Холта та Холта-Вінтера, тощо.

Метод простого <u>експоненціального згладжування застосовується</u> за відсутності **тенденції та сезонності** часового ряду.

Метод подвійного <u>експоненціального згладжування та метод Холта</u> застосовують за **наявності лінійної тенденції та сезонності** часового ряду.

Метод потрійного <u>експоненціального згладжування</u> застосовується за наявності **нелінійної тенденції**. <u>За наявності, як нелінійної тенденції так і сезонності</u> застосовують метод Холта-Вінтера.

Згладжування ряду здійснюється ковзкою середньою за експоненціальними вагами, які надають більшу перевагу значенням у кінці інтервалу, ніж значенням на його початку, але останні не виключаються з аналізу оскільки мають певну інформацію. Назва методу внаслідок того, що згладжування відбувається зваженою середньою, в якій ваги змінюються за експоненціальним законом. Експоненціальна середня ***EMA*** першого порядку розраховується як: $S_{y_n^1} = \sum_{i=1}^{n} \alpha(1-\alpha)^{n-i} y_i$, де $S_{y_n^1}$ – експоненціальна середня першого порядку для *n*-го періоду, «α» – коефіцієнт згладжування знаходиться в межах $0 < \alpha < 1$; звідки $\alpha(1-\alpha)^{n-i} \geq 0$ та $\sum_{i=1}^{n} \alpha(1-\alpha)^{n-i} \to 1$. Експоненціальна середня першого порядку є прогнозним значенням показника в *(n+1)* періоді, тобто: $S_{y_n^1} = y_{n+1}$. Сенс експоненціальної середньої в тому, що вага, яка надається останній, тобто найближчий до поточного моменту ціні, залежить від довжини інтервалу згладжування ковзкої середньої, чим він коротше тим більша вага. Для експоненціальної середньої дійсна така рекурентна формула: $S_{y_n^1} = \alpha \times y_n + (1-\alpha) S_{y_{n-1}^1}$.

<u>Постає питання, щодо початкового значення експоненціальної ковзкої середньої.</u>

Єдиного підходу вибору початкового значення не існує, найчастіше приймають його фактичне початкове значення $S_{y_0^1} = \mathrm{y}_1$, з масиву значень з флуктуаціями[63] та тенденціями (курси валют, акцій,тощо). Інколи, коли значення цін на товари стабільні у часі та коливаються навколо середнього, обирають середнє їх значення за формулою: $S_{y_0^1} = \frac{\sum_{i=1}^{n} \mathrm{y}_1}{n}$.

[62] Exponential moving average – *EMA*(k), англ.
[63] Коливання з відхиленням від середніх значень.

Розглянемо, як приклад, прогноз прибутку за експоненціальної ковзкою середньою на 1997рік за масивом вихідних даних див. табл.11.

Таблиця 11.

Роки	Прибуток Y_1 млн. грн.	EMA[64] $(\alpha = 0{,}6)$ $S_{y_n^1} = \alpha(1-\alpha)^{n-i} y_i$ млн. грн.	Кількість спостережень: загальна	Кількість спостережень: i
1989	Y_1=36,3	0,04	*n*=8	1
1990	Y_2=75,0	0,18		2
1991	Y_3=57,2	0,35		3
1992	Y_4=69,0	1,06		4
1993	Y_5=55,5	2,13		5
1994	Y_6=73,3	7,04		6
1995	Y_7=64,1	15,38		7
1996	Y_8=60,0	36,00		8
1997	**Y_9=?**	**∑62,18**		

Прогнозне значенням прибутку на 1997рік дорівнює $y_{1997} = S_{y^1_{1996}} = 62{,}18$ млн. грн.

<u>На відміну від зваженої ковзкої</u> середньої, експоненціальна середня **відображає не тільки динаміку рівнів в межах періоду осереднення**, але і враховує попередній розвиток ринку в межах часу, що розглядається. Ступінь врахування попередньої інформації визначається коефіцієнтом «*α*», який може бути обраний в інтервалі від 0 до 1. Чим більше значення цього коефіцієнта, тим більша вага додається поточним значенням і зменшена значущість значень всієї попередньої інформації. Величина «*α*» визначається наступною формулою: $\alpha = \frac{2}{k+1}$, де *k* – кількість значень спостережень, які входять в інтервал згладжування.

З врахуванням зазначеного коефіцієнту, ковзка експоненціальна середня розраховується за такою рекуррентною[65] формулою: $EMA(k)_i = \alpha \times y_i + (1-\alpha) \times EMA(k)_{i-1}$. Початковим 1$^{-м}$ рівнем в ланцюжку ковзких середніх є проста змінна середня $k^{-го}$ порядку:

$$EMA(k_i) = SMA(k)_1 = \frac{\sum_{i=1}^{k} y_i}{k}.$$

Таким чином, для отримання експоненційної ковзкої середньої необхідно виконати наступні етапи розрахунку:

1. Вибрати період осереднення (*k*).
2. Розрахувати коефіцієнт (*α*).
3. Обчислити просту змінну середню з перших (*k*) рівнів.
4. Наступний, за періодом осереднення, рівень ціни множиться на «*α*» і додається до попередньої ковзкої середньої помноженої на (*1–α*).
5. Крок 4 у подальшому повторюється до закінчення розрахунків ковзких середніх (EMA) за даними всіх значень ряду.

Розглянемо розрахунок за 5-рівневою експоненційною ковзкою середньою. Для цього спочатку визначимо значення коефіцієнта «*α*» у формулі: $\alpha = \frac{2}{k+1} = \frac{2}{5+1} = 0{,}333$.

---

[64] Експоненціальна середня першого порядку

[65] Повторюваність, англ., тут у сенсі дублювання формули іншим, альтернативним аналітичним її виразом.

Розрахунки експоненційної ковзкої середньої зведемо в табл. 12.

Таблиця 12.

№	Дата	Ціна	$\alpha \times y_i$	$(1-\alpha) \times EMA(k)_{i-1}$	$EMA(k)_i$	
	a	*b*	*c*	*d*	*e*	*f*
(i)			(Кол.*b*)[66]×α	Кол.*b* ($\sum y_{i \div k}$)/5×(1-*α*)	Кол.(*c*+*d*)	$\left(\sum_{i=1}^{5} \text{Кол.}\, b_i\right)/5.$
1	26.Кві	74,6				
2	29.Кві	72,0				
3	05.Тра	72,1				
4	06.Тра	69,7				
5	07.Тра	69,5			→	(71,6)
6	12.Тра	70,6	23,5	47,7[67]	**71,2**	
7	13.Тра	71,8	23,9	47,2	**71,1**	
8	14.Тра	75,0	25,0	47,2	**72,2**	
9	16.Тра	77,7	25,9	47,6	**73,5**	
10	17.Тра	79,3	26,4	48,6	**75,0**	
	Разом					

Оскільки перше значення «**71,6**» представляє значення простої ковзкої середньої ($SMA(k)_i$), воно не вважається значенням експоненціальної ковзкої середньої і не використовується у подальших розрахунках. Вважається, що експоненціальна ковзка середня краще відображає ринкові ціни, тому що вплив кожної попередньої ціни зменшується експоненціально (тобто стрімкіше) з її віддаленістю від значення поточної ціни.

XVII. Прогнозування попиту продажів.

Прогнозування попиту.

Вартість підприємницької діяльності спирається на прогноз доходів за обсягом реалізації продукції. Обсяг реалізації продукції залежить від попиту, незалежного чи залежного.

Незалежний попит відповідає рівновазі обсягів попиту та пропозиції, за яким ринок достатньо насичений, не має дефіциту товару та його надлишку, а споживачі не піддаються зовнішньому впливу у ціноутворенні, типу рекламних акцій, доступних, але неринкових кредитів, низьких цін у конкурентів чи форс-мажорних (повені, тощо) обставин;

Залежний попит обумовлений саме переліченим вище зовнішнім впливом.

Всі подальші припущення щодо прогнозу будуть пов'язані з **незалежним попитом**, якщо не обумовлено інше. Крім того, оскільки математичне прогнозування засноване на статистичних методах, знадобляться фактичні дані минулих періодів у часі. Припускаємо, що минулий попит (за фактом) складався також, як незалежний. **Залежний попит**» має різні значення у часі, на противагу йому «попит незалежний» обумовлений лише умовами рівноваги за обсягами, майже незмінний, як у минулому, так і в майбутньому, і може мати лише незначні коливання в межах обсягів рівноваги.

Припущення щодо незалежності попиту ґрунтується на інформації про продаж товару за логікою, що спирається на розумінні ринкової ціни товару, розміру ринку та частки товару компанії на ньому. Якщо це припущення хибне, якісний прогноз не збудувати.

[66] «колонка *b*» $70{,}6 \times 0{,}333 = 23{,}5.$

[67] $(74{,}6 + 72{,}0 + 72{,}1 + 69{,}7 + 69{,}5)/5 \times 0{,}667 = 47{,}7.$

Втім, можливе ще прогнозування за експертною оцінкою, що у ряді випадків може бути єдиним методом, навіть при достатній кількості вихідних даних. Тут неможливо обійтися без участі фахівця з прогнозування, який здатний узгодити результати прогнозів декількох методів математичних розрахунків вартості та надати узагальнений висновок.

Розглянемо математичні методи прогнозування.

Постановка задачі.

Є інформація щодо цін за роками продажів товарів, що приведена у табл.13.

Таблиця 13.

Продаж за роками та цінами

Дата	Ціна	**Дата**	Ціна	**Дата**	Ціна	**Дата**	Ціна	**Дата**	Ціна
0	925	**8**	1187	**16**	1428	**24**	1271	**32**	1448
1	930	**9**	1217	**17**	1099	**25**	1052	**33**	1553
2	1200	**10**	1138	**18**	1299	**26**	1769	**34**	1329
3	1142	**11**	1371	**19**	1257	**27**	1678	**35**	**1279**
4	1086	**12**	1005	**20**	1212	**28**	1481		
5	1551	**13**	973	**21**	1599	**29**	1100		
6	1239	**14**	1302	**22**	1830	**30**	1200		
7	1318	**15**	1756	**23**	1447	**31**	1430		

Часовий ряд володіє двома критичними властивостями:

- значення ряду мають бути обов'язково послідовно впорядковані.

  Переставте будь-які значення місцями, і отримаєте інший ряд.

- значення ряду мають бути розподілені за рівними фіксованими інтервалами часу.

  Прогнозування ряду за формою графіку, означає «продовження» низки його значень через ті ж самі інтервали на заданий рівень прогнозування.

Звідси випливає вимога до точності вихідних даних. Якщо треба отримати місячний прогноз, вихідні дані мають бути вибрані за місячними значеннями. Дані помісячних значень не можна використовувати безпосередньо, оскільки кількість днів, протягом яких здійснювалися операції з обсягів виробництва чи реалізації продукції у кожному місяці – різні, що вносить помилку, оскільки обсяг пропорційний часу у днях. Проблема вирішується зведенням вихідних даних до середньоденних.

Для того, щоб зробити якісь припущення щодо процесу, маємо, зменшити ступінь нашого незнання. Припускаємо, що процес має внутрішні закономірності за плином часу, абсолютно об'єктивні у змінюваному оточенні. Математично це має вираз $Y(t) = f(t) + \varepsilon(t)$,

де $Y(t)$ – значення ряду, наприклад, обсяг продажів у момент часу «t_i»;

- $f(t)$ – функція, що описує обсяг продаж, назвемо її прогнозною моделлю;
- $\varepsilon(t)$ – помилка, що пов'язана з випадковістю обсягів продаж або з незнанням, невмінням чи неможливістю врахувати інші несуттєві фактори моделі $f(t)$.

Завдання полягає в пропозиції такої моделі, величина помилки якої була б менше тієї, що спостерігається. Якщо ми матимемо таку модель, то зможемо здійснювати прогнозування за нею. Більш того, чим точніше модель описуватиме процес у минулому, тим більше буде впевненості, що вона відповідатиме цьому процесу і в майбутньому.

Динамічний процес, як правило, є повторюваним (ітеративним)[68] у часі.

Виходячи з простого погляду на графік вибираємо просту модель, а її параметри таким чином, щоб величина $\varepsilon(t) = Y(t) - f(t)$ була мінімально можливою. Цю величину називають «залишками» (residuals), що не вдалося описати моделлю, оскільки це те, що залишилося після вирахування з фактичних значень моделі її теоретично розрахованих значень. Для оцінки того, наскільки якісно модель описує процес, необхідно розрахувати значення інтегральної характеристики залишків (або помилок). Для обчислення сумарного значення помилок використовують середнє абсолютне або середньоквадратичне значення залишків всього ряду значень. Якщо сума помилок велика, то модель намагаються «поліпшити», тобто вибрати більш складний або простіший її вид чи змінити кількість факторних ознак.

У **специфікації моделі** треба дотримуватися, як мінімум двох правил:

- **використання моделі має бути однозначно роз'яснено за її змістом**.

  Якщо у модель обсягів продажу холодильників у Ризі ввести, в якості фактора, погоду в Києві, то не виключено, що отримаємо адекватну модель, можливо навіть з меншою помилкою. Поки не буде сформульовано причинно-наслідковий зв'язок, між результативною та факторною ознаками місцезнаходження, вводити цей фактор не має підстав.
- **поліпшення якості прогнозу моделі коштує грошових витрат**.

Тому завжди необхідно розуміти, що починаючи з якогось моменту витрати за результатом уточнення моделі перестають окупатися, тому треба вчасно зупинитися. Існують й інші перестороги щодо складних моделей. Справа в тому, що модель для кожного ряду створюється один раз, потім лише іноді перевіряється і коригується, тоді як прогнози по цій моделі створюються на регулярній основі. Який сенс у тому, що при кожному новому складанні прогнозу продажів доведеться спочатку досліджувати поведінку продажів товару за обсягом та створювати для нього модель? Системний підхід полягає в тому, щоб автоматично контролювати значення помилки прогнозування та за певними правилами коригувати деякі моделі. Наприклад, тільки ті, у яких значення відносної помилки вище певної межі і одночасно обсяг продажів товарів вище певного рівня. Чи просто візуально оцінювати діаграму, де по вертикалі відкладена величина помилки, а по горизонталі – ціна або обсяги продажів товару. Оскільки ціни на ринках постійно змінюються з часом, розроблені одного разу моделі перестають відображати реалії.

Більш прості моделі, як правило, стійкіші до змін незважаючи на більшу помилковість. Є одна особливість моделей. Чим простіше модель, тим зрозуміліша вона управлінцям підприємства, відповідальними за прийняття рішення, і тим вище їх довіра до отриманого прогнозу. Тому віддавати перевагу слід не складнішій моделі прогнозування, що забезпечить можливо більшу точність, а більш простій, зрозумілішій керівникам компанії. Коли відібрана модель отримує підтримку керівників підприємства, то і результати прогнозування активно використовуються[69].

---

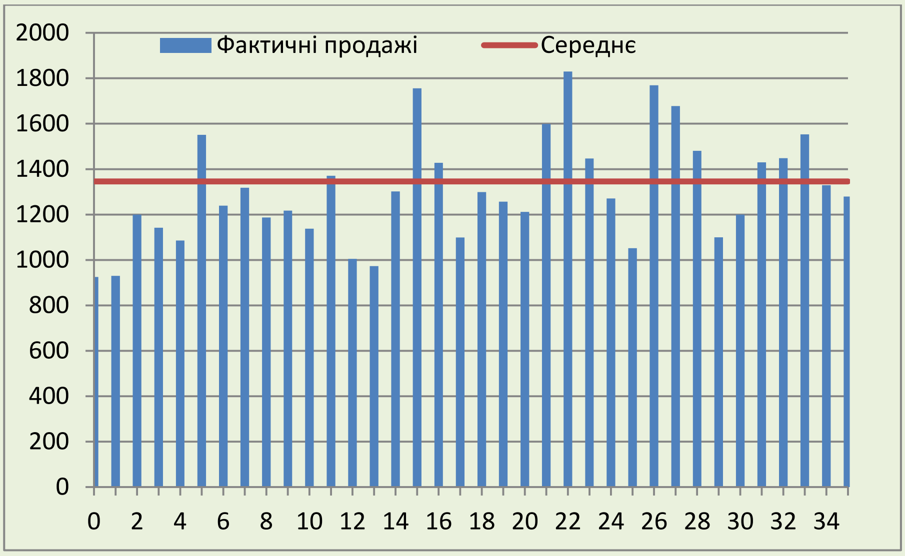
[68] (санскрит «itera» лат. – інший, повторюваний. Повторне застосування математичної операції (із зміненими даними), яке надає можливість поступово наблизитися до потрібного результату.

[69] Книга «Бізнес-прогнозування» «Business Forecasting», Hanke, Reitsch, Wichern».

Методи прогнозування.

1. Просте середнє.

У простому випадку, коли значення коливаються навколо деякого рівня, очевидним є оцінка середнього значення та припущення, що реальні продажі будуть коливатися навколо цього значення, $Y(t) = f(t) + \varepsilon(t)$, де середнє розраховується, як $\bar{y}_{MA}(t) = \frac{\sum_{i=1}^{n} x_i}{n} = 1360{,}9$, за інформацією табл. 13.

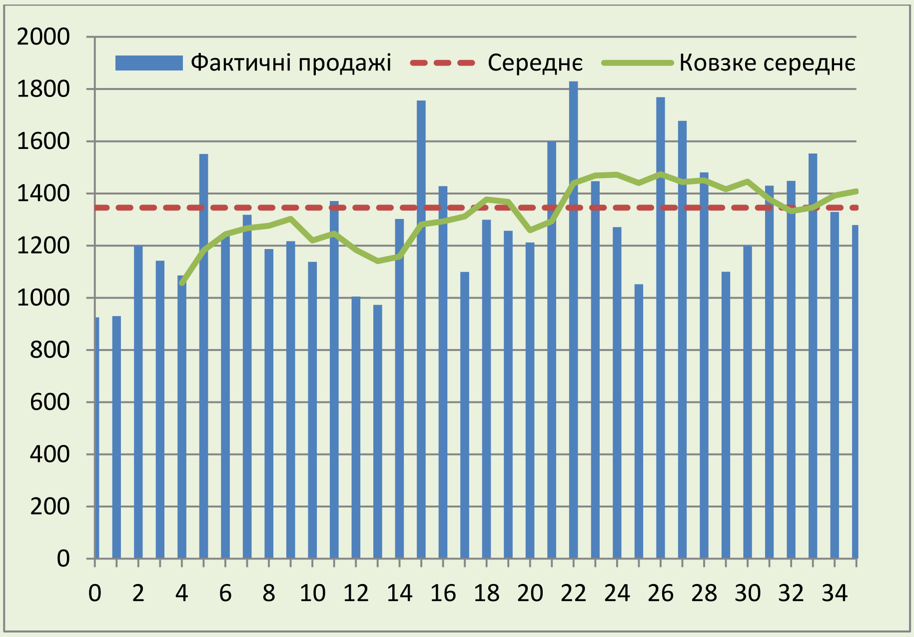


2. Ковзке середнє.

У реальності графік середньої дещо інший. Виробництво зростає, оборот збільшується. Одна з моделей середньої, що враховує тенденцію зростання та зменшує вплив спрямованих у минуле даних значень, використовує для обчислення середнього значення декілька останніх періодів «*k*» значень.

Метод отримав назву «ковзкого середнього», що розраховується за формулою:

$$\bar{y}_{SMA}(\tilde{t})_{(\text{від } k=5 \text{ до } n-k)} = \frac{\sum_{i=1}^{k} x_i}{k},$$ див. вихідні дані у табл. 13, вище.



3. **Зважена ковзка середня.**

Наступним кроком у модифікації моделі є припущення про те, що наближені до певної дати значення ряду більш адекватно відображають ситуацію. Тому, кожному наближеному наступному значенню присвоюється вага більша ніж попередньому, віддаленому за датою значенню. Для зручності можна вибрати коефіцієнти таким чином, щоб сума їх становила одиницю, тоді не доведеться у формулі виконувати ділення на суму їх ваг. Будемо вважати, що такі коефіцієнти нормовані за одиницею.

$$\tilde{y}_{WMA(\text{від } k=5 \text{ до } n-k)}(t) = \frac{\sum_{i=1}^{k} \alpha_i x_i}{\sum_{i=1}^{k} \alpha_i}.$$

Графіки ковзкої середньої та ковзкої середньої зваженої наведені нижче.

4. **Просте експоненціальне згладжування[70].**

Одним з різновидів усереднення є метод експоненціального згладжування. Коефіцієнти згладжування за цим методом вибирається певним чином, а їх значення встановлюється за експоненціальним законом. Зупинимося на ньому трохи докладніше, оскільки метод отримав поширення завдяки простоті і легкості обчислень усереднених згладжених значень.

Нехай нам потрібний прогноз на момент часу t+1 (на наступний період). Позначимо його як $\hat{y}_{t+1} = \hat{y}_t + \alpha(y_{\,t} - \hat{y}_t)$ або, як $\hat{y}_{EMA(t+1)} = \alpha \times y_{(t+1)} + (1-\alpha) \times \hat{y}_t$, де $0 < \alpha < 1$.

Ми отримали так зване рекурентне співвідношення[71], коли наступний член виражається через попередній. Тепер прогноз минулого періоду виражаємо тим же способом через позаминуле значення ряду і так далі. У підсумку вдається отримати формулу прогнозу.

$$\hat{Y}_{t+1} = \alpha \sum_{t=1}^{n} (1-\alpha)^{t-1} \times Y_{t-1} + (1-\alpha)^{t} \times Y_{\,0}$$

Продемонструємо на графіку згладжування при різних значеннях коефіцієнта згладжування. Якщо оборот монотонно зростатиме, то при такому підході ми будемо систематично отримувати занижені цифри прогнозів, і навпаки. Методика згладжування припускає для першого значення прогнозу використання його фактичного першого значення, а далі за формулою рекурсії: $\hat{y}_{EMA(t+1)} = \alpha \times y_{(t+1)} + (1-\alpha) \times \hat{y}_t$.

---

[70] *SES – Simple Exponential Smoothing, англ.*

[71] Однотипні формули, які пов'язують між собою один за одним елементи деякої послідовності (це може бути послідовність чисел, функцій і т. д.). Залежно від природи об'єктів, пов'язаних рекурентними співвідношеннями, ці співвідношення можуть бути алгебраїчними, функціональними, диференціальними, інтегральними і т.п.

Нижче наданий графік[72] та табл. №14 розрахунку значень ковзких середніх та ковзкої експоненціального згладжування за різними значеннями коефіцієнтів згладжування «α».

Дата місяців	Обсяг факт од.	Вага α од.	Середня проста од.	Ковзка середня (одиниць – обсяг) **Таблиця 14**			
				Проста	Зважена	Експоненціальна	
				k=5 рівня		$\alpha = 0{,}1$	$\alpha = 0{,}3$
0	925	0,10				925,00	925,0
1	930	0,15				925,50	926,50
2	1200	0,20				952,95	1008,55[73]
3	1142	0,25				971,86[74]	1048,59
4	1086	0,30		1056,6[75]	1083,3[76]	983,27	1059,81
5	1551			1181,8	1238,2	1040,04	1207,17
6	1239		1345,7	1243,6	1268,0	1059,94	1216,72
7	1318			1267,2	1292,5	1085,74	1247,10
8	1187			1276,2	1274,7	1095,87	1229,07
9	1217			1302,4	1266,4	1107,98	1225,45
10	1138			1219,8	1204,7	1110,98	1199,21
11	1371			1246,2	1249,1	1136,99	1250,75
12	1005			1183,6	1173,1	1123,79	1177,03
13	973			1140,8	1109,8	1108,71	1115,82
14	1302			1157,8	1154,3	1128,04	1171,67
15	1756			1281,4	1334,8	1190,83	1346,97
16	1428			1292,8	1374,3	1214,55	1371,28
17	1099			1311,6	1330,5	1203,00	1289,60
18	1299			1376,8	1343,7	1212,60	1292,42
19	1257			1367,8	1311,5	1217,04	1281,79
20	1212			1259,0	1245,3	1216,54	1260,85
21	1599			1293,2	1338,9	1254,79	1362,3
22	1830			1439,4	1509,6	1312,31	1502,61
23	1447			1469,0	1518,9	1325,78	1485,93
24	1271			1471,8	1470,1	1320,30	1421,45
25	1052			1439,8	1357,2	1293,47	1310,62
26	1769			1473,8	1448,0	1341,02	1448,13
27	1678			1443,4	1491,4	1374,72	1517,09
28	1481			1450,2	1502,5	1385,35	1506,26
29	1100			1416,0	1406,4	1356,82	1384,38
30	1200			1445,6	1359,8	1341,14	1329,07
31	1430			1377,8	1339,0	1350,03	1359,35
32	1448			1331,8	1345,0	1359,83	1385,95
33	1553			1346,2	1403,9	1379,15	1436,07
34	1329			1392,0	1411,1	1374,14	1403,95
35	1279			1407,8	1386,8	1364,63	1366,47
Разом	**47101**						

[72] За даними табл. 13, див. вище;
[73] 0,3×1200+(1-0,3)×926,50=1008,55;
[74] 0,1×1142+(1-0,1)×952,95=971,86;
[75] (925+930+1200+1142+1086)/5=1056,6;
[76] 925×0,1+930×0,15+1200×0,2+1142×0,25+1086×0,3=1083,3;

Щоб встановити тенденцію зростання чи падіння отриманої моделі вводиться поняття «тренда», тобто лінії, яка системно усереднює спрямованість значень часового ряду.

Загальна модель часового ряду, як об'єкта статистичного аналізу.

Враховуємо, що при побудові економетрічної моделі використовуються:

1) дані, що характеризують об'єкт за різними показниками на певний період часу;

2) дані, що характеризують об’єкт за ряд послідовних періодів часу.

Моделі, побудовані за даними першого типу, називають просторовими моделями. Моделі, що побудовані за даними другого типу, називають моделями часових рядів.

Часовий ряд (або ряд динаміки) – це сукупність значень будь-якої ознаки за кілька послідовних періодів часу. Кожен рівень часового ряду формується під впливом певної кількості факторів, які умовно можна поділити на три групи:

1) фактори, що формують тенденцію ряду;
2) фактори, що формують циклічні коливання (сезонні та глобальні) ряду;
3) випадкові значення (помилки) ряду.

Більшість економічних даних часових рядів мають тенденцію, що характеризує сукупний довгостроковий вплив певних груп факторів на динаміку досліджуваної ознаки. Всі фактори, окремо, чи певною групою, можуть надавати різноспрямований вплив на результативну ознаку. Проте у сукупності вони можуть формувати її зростаючу або спадну тенденцію. На рис. 3 показаний часовий ряд, що має **тенденцію** зростання.

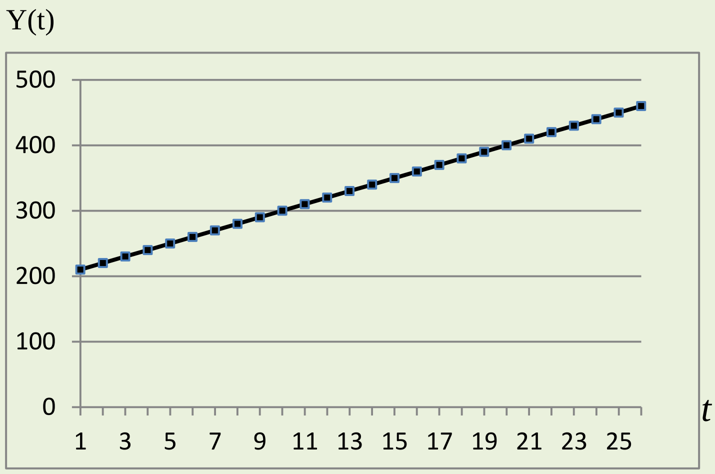


Рис. 3.

Досліджувана результативна ознака може мати **циклічні коливання**. Ці коливання можуть носити **сезонний** характер, оскільки діяльність деяких галузей економіки залежить від пори року, наприклад, обсяги або ціни певної сільськогосподарської продукції в літній період, можуть бути меншими, ніж в зимовий, чи навпаки; або рівень доходів курортних міст у зимовий період менші порівняно з літнім. При наявності масивів даних за більш тривалі періоди можливо виявити **глобальні циклічні** коливання, пов'язані із загальною динамікою кон’юнктури ринку.

На рис. 4 поданий часовий ряд, що має **циклічну сезонну складову** (компоненту).

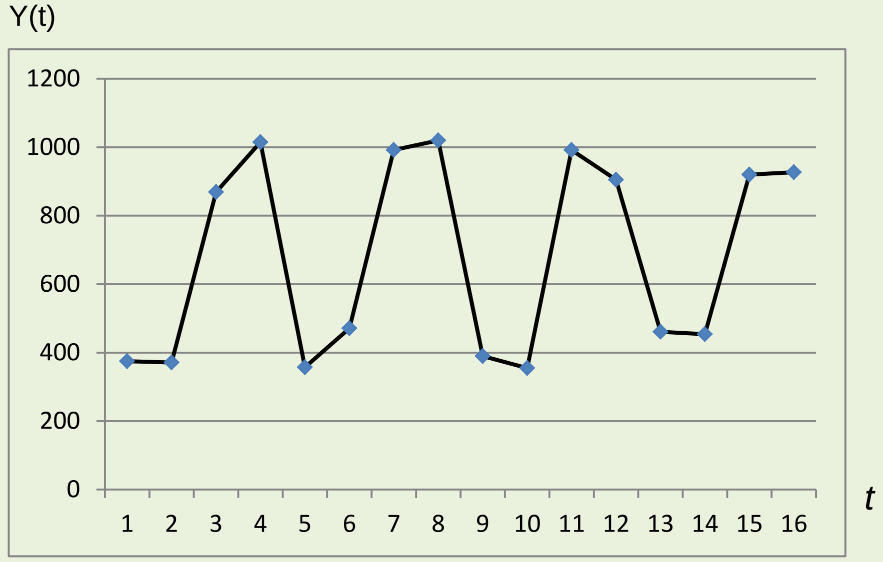


Рис. 4.

Деякі часові ряди не містять тенденції і циклічної складової, а кожен наступний їх рівень утворюється як сума середнього рівня ряду і деякої (позитивної або негативної) випадкової компоненти.

Вочевидь, що реальні дані не відповідають повністю будь-яким моделям. Найчастіше вони містять всі три складові. Кожен їх рівень формується під впливом тенденції, сезонних чи глобальних коливань і випадкової компоненти.

У більшості випадків фактичний рівень часового ряду можна представити як **суму** або **добуток** трендової, циклічної та випадкової складової.

Модель, в якій часовий ряд поданий як сума зазначених складових, має назву **адитивної моделі** часового ряду.

Модель, в якій часовий ряд представлений як добуток перерахованих складових, має назву **мультиплікативної моделі** часового ряду.

Основне завдання економетричного дослідження окремого часового ряду полягає у виявленні та наданні кількісного значення кожної з перерахованих вище компонент, для того щоб використовувати отриману інформацію при прогнозуванні майбутніх значень ряду або побудові моделей взаємозв'язку двох або більше часових рядів.

Автокореляція рівнів часового ряду.

При наявності в часовому ряді **тенденції та циклічних** коливань значення кожного наступного рівня ряду залежать від попередніх. Кореляційну залежність між послідовними рівнями часового ряду називають **автокореляцією рівнів ряду**.

Кількісно її можна виміряти за допомогою лінійного коефіцієнта кореляції між рівнями вихідного часового ряду і рівнями цього ряду, зміщеними на кілька інтервалів у часі.

Оскільки коефіцієнт автокореляції порядку першого рівня вимірює залежність між рівнями ряду y_t та y_{t-1}, то формула для його розрахунку має вираз:

$$r_1 = \frac{\sum_{t=2}^{n}(y_t-\bar{y}_1)\times(y_{t-1}-\bar{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t-\bar{y}_1)^2\times\sum_{t=2}^{n}(y_{t-1}-\bar{y}_2)^2}},$$

де $\bar{y}_1 = \frac{1}{n-1}\times\sum_{t=2}^{n} y_t$; а $\bar{y}_2 = \frac{1}{n-1}\times\sum_{t=2}^{n} y_{t-1}$.

Аналогічно можливо визначити коефіцієнти автокореляції ряду рівнів другого та більших порядків. Так, коефіцієнт автокореляції ряду рівня другого порядку характеризує тісноту зв'язку між рівнями порядків y_t та y_{t-2} та визначається за формулою:

$$r_2 = \frac{\sum_{t=3}^{n}(y_t-\bar{y}_3)\times(y_{t-2}-\bar{y}_4)}{\sqrt{\sum_{t=3}^{n}(y_t-\bar{y}_3)^2\times\sum_{t=3}^{n}(y_{t-2}-\bar{y}_4)^2}},$$

де $\bar{y}_3 = \frac{1}{n-2}\times\sum_{t=3}^{n} y_t$; а $\bar{y}_4 = \frac{1}{n-2}\times\sum_{t=3}^{n} y_{t-2}$.

Кількість періодів, за якими розраховується коефіцієнт автокореляції, має назву **лагу**. Зі збільшенням лага кількість пар значень, за якими розраховується коефіцієнт автокореляції, зменшується. Для забезпечення статистичної достовірності коефіцієнтів автокореляції вважається за доцільне використовувати правило – **максимальний лаг** має бути не більше $n/_4$.

Властивості коефіцієнта автокореляції.

1. Він будується за аналогією з лінійним коефіцієнтом кореляції і таким чином характеризує тісноту тільки лінійного зв'язку поточного і попереднього рівнів ряду. Тому за коефіцієнтом автокореляції можна судити про наявність лінійної (або близькою до лінійної) тенденції. Для деяких часових рядів, що мають сильну нелінійну тенденцію (наприклад, параболу або експоненту), коефіцієнт автокореляції рівнів вихідного ряду може наближатися до нуля.

2. За знаком коефіцієнта автокореляції не можна зробити висновок про зростаючу або спадну тенденції у рівнях ряду. Більшість даних часових рядів економічних процесів містять позитивну автокореляції рівнів, однак при цьому можуть мати спадну тенденцію.

Послідовність коефіцієнтів автокореляції рівнів першого, другого та вищих порядків називають **функцією автокореляції часового ряду**. Графік залежності значень цих коефіцієнтів від номера лага (порядку коефіцієнта автокореляції) має назву **коррелограми**.

Аналіз функції автокореляції та коррелограми дозволяє визначити порядок (лаг) коефіцієнта автокореляції, при якому автокореляція найбільша, а отже, і порядок (лаг) при якому зв'язок між поточним і попередніми рівнями ряду найбільш тісний, тобто за допомогою аналізу функції автокореляції і коррелограми можна **виявити структуру ряду**.

Якщо найбільшим виявився коефіцієнт автокореляції рівня **першого порядку**, досліджуваний ряд містить тільки **тенденцію**. Якщо найбільшим виявився коефіцієнт автокореляції рівня **порядку** «*k*», то ряд містить **циклічні коливання з періодичністю** в «*k*» періодів часу. Якщо жоден з коефіцієнтів автокореляції не є значущим, можна зробити одне з двох припущень щодо структури цього ряду: або ряд **не містить тенденції** і **циклічних** коливань, або ряд **містить** сильну нелінійну тенденцію, для виявлення якої потрібно провести додатковий аналіз.

Звідси коефіцієнт автокореляції рівнів і функцію автокореляції доцільно використовувати для виявлення наявності або відсутності **трендової та циклічної** (сезонної) компоненти у часовому ряді.

Розглянемо приклад.

Нехай маємо дані про загальну кількість правопорушень на митниці, див. табл.15

Таблиця 15.

Рік	Квартал	*t*	Кількість правопорушень
2000	I	1	375
	II	2	371
	II	3	869
	IV	4	1015
2001	I	5	357
	II	6	471
	II	7	992
	IV	8	1020
2002	I	9	390
	II	10	355
	II	11	992
	IV	12	905
2003	I	13	461
	II	14	454
	II	15	920
	IV	16	927

Побудуємо графік поля кореляції за даними табл.16 (див. далі).

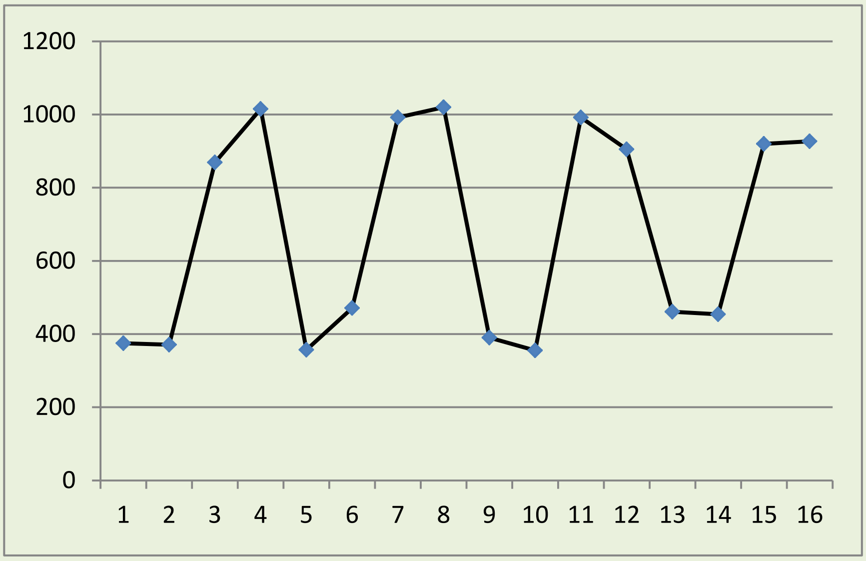


Рис.5.

За графіком видно, що функція за формою нагадує синусоїду.

Для розрахунку коефіцієнта **автокореляції першого порядку** складаємо допоміжну таблицю, див. табл. №16.

Таблиця 16.

№ кварталу	Кількість порушень y_t	Кількість порушень y_{t-1}	$y_t - \bar{y}_1$	$y_{t-1} - \bar{y}_2$	$(y_t - \bar{y}_1) \times (y_{t-1} - \bar{y}_2)$	$(y_t - \bar{y}_1)^2$	$(y_{t-1} - \bar{y}_2)^2$
1	**(375)**	–	–	–	–	–	–
2	371	375	-328,93	-288,13	94774,60	108194,94	83018,90
3	869	371	169,07	-292,13	-49390,42	28584,66	85339,94
4	1015	869	315,07	205,87	64863,46	99269,10	42382,46
5	357	1015	-342,93	351,87	-120666,78	117600,98	123812,50
6	471	357	-228,93	-306,13	70082,34	52408,94	93715,58
7	992	471	292,07	-192,13	-56115,41	85304,88	36913,94
8	1020	992	320,07	328,87	105261,42	102444,80	108155,48
9	390	1020	-309,93	356,87	-110604,72	96056,60	127356,20
10	355	390	-344,93	-273,13	94210,73	118976,70	74600,00
11	992	355	292,07	-308,13	-89995,53	85304,88	94944,10
12	905	992	205,07	328,87	67441,37	42053,70	108155,48
13	461	905	-238,93	241,87	-57790,00	57087,54	58501,10
14	454	461	-245,93	-202,13	49709,83	60481,56	40856,54
15	920	454	220,07	-209,13	-46023,24	48430,80	43735,36
16	927	920	227,07	256,87	58327,47	51560,78	65982,20
Разом	∑**10499**	∑**9947**	∑0,05	∑0,05	∑74085,12	∑1153760,86	∑1187469,78
Середнє	$\bar{y}_1$=**699,93**[77]	$\bar{y}_2$=**663,13**					

Слід зауважити, що середнє значення першого рівня розраховано діленням на кількість з **15** порушень, оскільки тепер маємо на одне порушення менше (перше не враховується). Кількість порушень кожного наступного за рівнем середнього зменшується ще на одну одиницю, наприклад, для середнього рівня другого порядку, їх кількість вже менше на два.

Тепер обчислимо коефіцієнт **автокореляції рівня першого порядку за формулою**:

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y}_1)\times(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t - \bar{y}_1)^2 \times \sum_{t=2}^{n}(y_{t-1} - \bar{y}_2)^2}} = \frac{74085{,}12}{\sqrt{1153760{,}86 \times 1187469{,}78}} = 0{,}063924.$$

[77] Середнє для першого стовпця розраховано без врахування значення у 375 першого кварталу.

Для розрахунку коефіцієнта автокореляції **другого порядку** складемо допоміжну табл.17.

Таблиця 17.

№ кварталу	Кількість порушень y_t	Кількість порушень y_{t-2}	$y_t - \bar{y}_3$	$y_{t-2} - \bar{y}_4$	$(y_t - \bar{y}_3) \times (y_{t-2} - \bar{y}_4)$	$(y_t - \bar{y}_3)^2$	$(y_{t-2} - \bar{y}_4)^2$
1	**(375)**	–	–	–	–	–	–
2	**(371)**	–	–	–	–	–	–
3	869	375	145,57	-269,79	-39273,33	21190,62	72786,64
4	1015	371	291,57	-273,79	-79828,95	85013,06	74960,96
5	357	869	-366,43	224,21	-82157,27	134270,94	50270,12
6	471	1015	-252,43	370,21	-93452,11	63720,90	137055,44
7	992	357	268,57	-287,79	-77291,76	72129,84	82823,08
8	1020	471	296,57	-173,79	-51540,90	87953,76	30202,96
9	390	992	-333,43	347,21	-115770,23	111175,56	120554,78
10	355	1020	-368,43	375,21	-138238,62	135740,66	140782,54
11	992	390	268,57	-254,79	-68428,95	72129,84	64917,94
12	905	355	181,57	-289,79	-52617,17	32967,66	83978,24
13	461	992	-262,43	347,21	-91118,32	68869,50	120554,78
14	454	905	-269,43	260,21	-70108,38	72592,52	67709,24
15	920	461	196,57	-183,79	-36127,60	38639,76	33778,76
16	927	454	203,57	-190,79	-38839,12	41440,74	36400,82
Разом ∑	**∑10128**	**∑9027**			**-1034792,71**	1037835,36	**1116776,3**
Середнє	$\bar{y}_3$=**723,43**[78]	$\bar{y}_4$=**644,79**					

Звідки $r_2 = \frac{\sum_{t=3}^{n}(y_t - \bar{y}_3)\times(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^{n}(y_t - \bar{y}_3)^2 \times \sum_{t=3}^{n}(y_{t-2} - \bar{y}_4)^2}} = \frac{\mathbf{-1034792{,}71}}{\sqrt{1037835{,}36 \times \mathbf{1116776{,}3}}} =$ **– 0,961183**.

Аналогічно розраховуємо коефіцієнти автокореляції рівнів вищих порядків, див. табл. №18.

Таблиця 18.

порядок коеф-нта	автокореляція	порядок коеф-нта	автокореляція
1-го	**+ 0,063294**	7-го	– 0,069444
2-го	**– 0,961183**	8-го	+ 0,964629
3-го	– 0,036290	9-го	+ 0,162064
4-го	+ 0,964735	10-го	– 0,972918
5-го	+ 0,050594	11-го	– 0,065323
6-го	– 0,976516	12-го	+ 0,985761

За значеннями коефіцієнтів автокореляції рівнів порядків будуємо **коррелограму** рис 6.

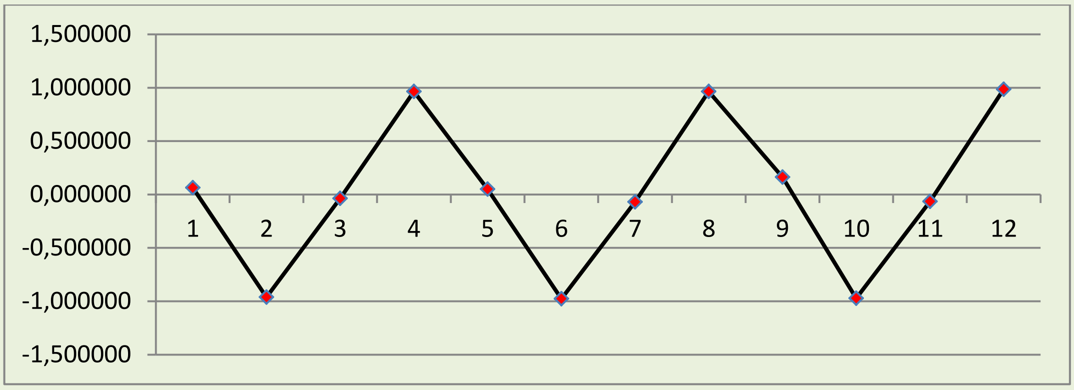


Рис.6

Аналіз коррелограми рівнів часового ряду дозволяє зробити висновок про наявність у досліджуваному ряді сезонних коливань періодичністю у 4-и квартали.

[78] Середнє для першого стовпця розраховано без врахування значень перших двох кварталів.

Моделювання сезонних коливань.

Найпростіший підхід до моделювання сезонних коливань це розрахунок значень сезонної компоненти методом ковзкої середньої і побудова адитивної або мультиплікативної моделі часового ряду. Загальний вид адитивної моделі наступний:

$$Y = T + S + E.$$

Модель припускає, що кожен рівень часового ряду може бути представлений, як сума компонент – трендової (*T*), сезонної (*S*) та випадкової (*E*).

Загальний вид мультиплікативної моделі має вигляд:

$$Y = T \times S \times E.$$

Ця модель припускає, що кожен рівень часового ряду може бути представлений як добуток трендової, сезонної і випадкової компонент.

Вибір однієї з двох моделей здійснюється за аналізом амплітуди сезонних коливань, коли вона:

- **постійна**	вибирають **адитивну** модель часового ряду, в якій значення сезонної компоненти передбачаються постійними для різних рівнів;
- **зростає або зменшується**	вибирають **мультиплікативну** модель часового ряду, яка ставить рівні ряду в залежність від значень сезонної компоненти.

Побудова адитивної і мультиплікативної моделей зводиться до розрахунку значень трендової (*T*), сезонної (*S*) та випадкової (Е) складових для кожного рівня часового ряду.

Процес побудови моделі включає в себе наступні кроки.

1) Вирівнювання вихідного ряду методом ковзкої середньої.
2) Розрахунок значень сезонної компоненти *S*.
3) Усунення сезонної компоненти з вихідних рівнів ряду та отримання вирівняних даних (*T*+*E*) в адитивної або (*T*×*E*) в мультиплікативної моделі.
4) Аналітичне вирівнювання рівнів (*T*+*E*) або (*T*×*E*) і розрахунок значень «*T*» з використанням отриманого рівняння тренда.
5) Розрахунок отриманих за моделлю значень (*T*+*E*) або (*T*×*E*).
6) Прогноз майбутніх значень рівнів часового ряду на основі побудованої моделі.

Методику побудови кожної з моделей розглянемо на прикладах.

Побудова адитивної моделі часового ряду.

Звернемося до даних про обсяг правопорушень на митниці за чотири роки, наведених в табл. 1. Було показано, що даний часовий ряд містить сезонні коливання періодичністю 4, оскільки кількість правопорушень в перший-другий квартали нижче, ніж в третій-четвертий. Розрахуємо складові адитивної моделі часового ряду.

Крок 1.

Проведемо вирівнювання вихідних рівнів ряду методом ковзкої середньої. Для цього:

1.1. Просумуємо рівні ряду послідовно за кожні чотири квартали із зміщенням на один квартал і визначимо річну кількість правопорушень (графа 3 табл. 19).

1.2. Розділивши отримані суми на 4, знайдемо ковзкі середні (графа 4 табл. 19). Отримані таким чином вирівняні значення вже не містять сезонної компоненти.

1.3. Приведемо ці значення у відповідність з фактичними моментами часу, для чого знайдемо середні значення з двох послідовних ковзких середніх – **центровані ковзкі середні** (графа 5 табл. 19).

Таблиця 19.

№ кварталу	Кількість порушень Y_t	Разом ∑ за 4 квартали	Ковзка середня за 4 квартали	Ковзка середня центрована	Оцінка сезонної компоненти
1	**2**	**3**	**4**	**5**	**6**
		∑1,2,3,4	граф.3 (ряд.2)**/4**	граф.4 (∑ряд.2,3)**/2**	графи (2) – (5)
1	375	–	–	–	–
2	371	2630	657,50	–	–
3	869	2612	653,00	655,25	213,75
4	1015	2712	678,00	665,50	349,50
5	357	2835	708,75	693,38	-336,38
6	471	2840	710,00	709,38	-238,38
7	992	2873	718,25	714,13	277,87
8	1020	2757	689,25	703,75	316,25
9	390	2757	689,25	689,25	-299,25
10	355	2642	660,50	674,88	-319,88
11	992	2713	678,25	669,38	322,62
12	905	2812	703,00	690,63	214,37
13	461	2740	685,00	694,00	-233,00
14	454	2762	690,50	687,75	-233,75
15	920	–	–	–	–
16	927	–	–	–	–

Крок 2. Знайдемо оцінки сезонної компоненти як різницю між фактичними рівнями ряду і центрованими ковзкими середніми (графа 6 табл. 19). Використовуємо ці оцінки для розрахунку значень сезонної компоненти (табл.20). Для цього знайдемо середні за кожен квартал (по всіх роках) оцінки сезонної компоненти S_i.

Таблиця 20.

Рік	Найменування	i–й квартал I	II	III	IV	Сума (∑)
2000	показники сезонної компоненти	–	–	213,75	349,50	
2001		-336,38	-238,38	277,88	316,25	
2002		-299,25	-319,88	322,63	214,38	
2003		-233,00	-233,75	0	0	
Разом за i–й квартал сума ∑		**-868,63**	**-792,01**	**814,26**	**880,13**	
Середня оцінка сезонної S_i за i–й квартал		-289,54	-264,00	271,42	293,38	**11,26**
Коефіцієнт коригування k (див. далі)		2,815	2,815	2,815	2,815	
Коригована сезонна компонента S_i		**-292,36**	**-266,82**	**268,61**	**290,57**	0

В моделях з сезонною компонентою зазвичай передбачається, що сезонні впливи за період взаємно гасяться. У адитивної моделі це виражається в тому, що вплив коливань за сумою значень сезонної компоненти по всіх кварталах повинен дорівнювати нулю.

Для моделі маємо: $\sum 289{,}54 - 264 + 271{,}42 + 293{,}38 = 11{,}26$, звідки коефіцієнт коригування $k = 11{,}26/4 = 2{,}815$. Відкориговані значення сезонної компоненти $S_i = \bar{S}_i - k$ заносимо у табл. 20 та перевіряємо рівність нулю суми значень сезонної компоненти: $\sum -292{,}36 - 266{,}82 + 268{,}61 + 290{,}57 = 0{,}00$.

Крок 3. Виключимо вплив сезонної компоненти, віднімаючи її значення з кожного рівня вихідного часового ряду і отримаємо $T + E = Y - S$ (гр. 4 табл. 21). Ці значення розраховуються за кожен квартал і містять тільки тенденцію і випадкову компоненту.

Крок 4. Визначимо значення тренду (*T)* моделі. Для цього визначимо **лінійний тренд** часового ряду за **графою 4** (Y_t-S_i), та отримаємо формулу тренду:$T = 671{,}76 + 0{,}9254 \times t$. За цією формулою знайдемо значення тренду підставляючи у рівняння t = 1,2,...,16, для кожного моменту часу та занесемо до графи 5, табл. 21. Далі визначимо значення випадкової складової та занесемо у графу 7.

Таблиця 21.

№ кварталу	Кількість порушень Y_t	S_i	$Y_t - S_i$	T	$T+S$	$E=Y_t-(T+S)$	E^2	$(y_t - \bar{y})^2$
1	**(2)**	**3**	**4**	**5**	**6**	**7**	**(8)**	**(9)**
1	375	-292,36	667,36	672,69	380,33	-5,33	28,41	92799,44
2	371	-266,82	637,82	673,61	406,79	-35,79	1280,92	95252,48
3	869	268,61	600,39	674,54	943,15	-74,15	5498,22	35861,00
4	1015	290,57	724,43	675,46	966,03	48,97	2398,06	112473,04
5	357	-292,36	649,36	676,39	384,03	-27,03	730,62	104090,12
6	471	-266,82	737,82	677,31	410,49	60,51	3661,46	43526,48
7	992	268,61	723,39	678,24	946,85	45,15	2038,52	97575,02
8	1020	290,57	729,43	679,16	969,73	50,27	2527,07	115851,74
9	390	-292,36	682,36	680,09	387,73	2,27	5,15	83885,54
10	355	-266,82	621,82	681,01	414,19	-59,19	3503,46	105384,64
11	992	268,61	723,39	681,94	950,55	41,45	1718,10	97575,02
12	905	290,57	614,43	682,86	973,43	-68,43	4682,66	50791,64
13	461	-292,36	753,36	683,79	391,43	69,57	4839,98	47799,08
14	454	-266,82	720,82	684,72	417,90	36,10	1303,21	50908,90
15	920	268,61	651,39	685,64	954,25	-34,25	1173,06	57777,74
16	927	290,57	636,43	686,57	977,14	-50,14	2514,02	61191,92
Середнє	$\bar{y} = 679{,}63$					Разом ($\sum$)=	**37902,92**	**1252743,8**

Крок 5. Знайдемо значення рівнів ряду за адитивною моделлю. Для цього додамо до рівнів T значення сезонної компоненти відповідних кварталів (графа 6 табл. 21). Графіки фактичних значень часового ряду і теоретичних, отриманих по адитивній моделі на рис. 7.

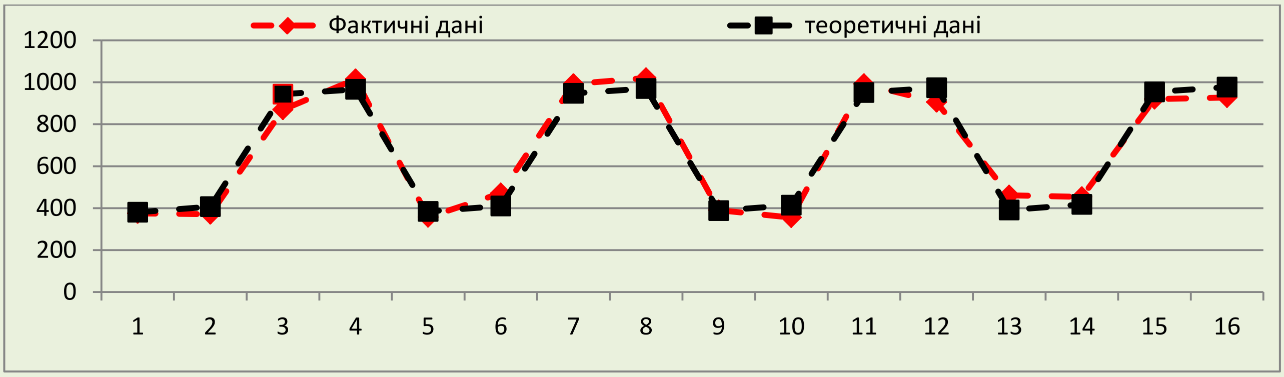


Рис. 7.

Для оцінки якості моделі визначимо коефіцієнт детермінації:

$$R^2 = 1 - \frac{\sum E}{\sum_1^n (y_t - \bar{y})^2} = 1 - \frac{37902{,}92}{1252743{,}8} = 0{,}97$$ [79]

Отже, можна сказати, що адитивна модель пояснює 97% загальної варіації рівнів часового ряду кількості правопорушень по кварталам за 4 роки.

<u>Крок 6</u>. **Прогнозування за адитивною моделлю**. Припустимо, що за нашим прикладом необхідно дати прогноз про загальний обсяг правопорушень на I-й та II-й квартали 2004 року.

Прогнозне значення *F(t)* рівня часового ряду в адитивної моделі є сума трендової і сезонної компонент. Для визначення трендової компоненти скористаємося рівнянням тренда $T = 671{,}776 + 0{,}9254 \times t.$

$$\mathrm{T}(17) = 671{,}776 + 0{,}9254 \times 17 = 687{,}49.$$

$$\mathrm{T}(18) = 671{,}776 + 0{,}9254 \times 18 = 688{,}42.$$

Значення сезонних компонент за 1-й та 2-й квартали дорівнюють:

S_1 = –292,448 та S_2 = – 266,781.

Таким чином: $F(17) = T(17) + S1 = 687{,}49 + (-292{,}357) \approx 395.$

$$F(18) = T(18) + S2 = 688{,}42 + (-266{,}815) \approx 422.$$

Тобто у перші два квартали 2004 року слід було очікувати відповідно 395 та 422 правопорушень.

1. Побудова мультиплікативної моделі.

<u>Крок 1</u>. Методика, що застосовується на цьому кроці, майже повністю збігається з методикою побудови адитивної моделі за винятком оцінки сезонної компоненти.

Таблиця 22.

№ кварталу	Кількість порушень Y_t	Разом за 4 квартали	Ковзка середня за 4 квартали	Центрована ковзка середня	Оцінки сезонної компоненти
1	**2**	**3**	**4**	**5**	**6**
1	375	–	–	–	–
2	371	2630	657,50	–	–
3	869	2612	653,00	655,25	1,3262
4	1015	2712	678,00	665,50	1,5252
5	357	2835	708,75	693,38	0,5149
6	471	2840	710,00	709,38	0,6640
7	992	2873	718,25	714,13	1,3891
8	1020	2757	689,25	703,75	1,4494
9	390	2757	689,25	689,25	0,5658
10	355	2642	660,50	674,88	0,5260
11	992	2713	678,25	669,38	1,4820
12	905	2812	703,00	690,63	1,3104
13	461	2740	685,00	694,00	0,6643
14	454	2762	690,50	687,75	0,6601
15	920	–	–	–	–
16	927	–	–	–	–

[79] Всі розрахунки попередніх таблиць проведені з точністю як на екрані монітора комп'ютера.

Крок 2. Оцінки сезонної компоненти знаходимо, як частку від ділення фактичних значень ряду на центровані ковзкі середні (графа 6 табл. 22). Ці оцінки використовуються для розрахунку сезонної компоненти *«S»* (табл. 23). Для цього знайдемо середні за кожний квартал оцінки сезонної компоненти *«S_i»*. Так само, як і в адитивній моделі, вважається, що сезонні впливи за період взаємно гасяться.

В мультиплікативній моделі це виражається в тому, що сума значень сезонної компоненти по всіх кварталах повинна дорівнювати числу періодів в циклі. У нашому випадку кількість періодів одного циклу дорівнює 4.

Таблиця 23.

№	Показники	Рік	квартал				Разом
			I	**II**	**III**	**IV**	
1	Оцінки сезонної складової	2000	–	–	1,3262	1,5252	
2		2001	0,5149	0,6640	1,3891	1,4494	
3		2002	0,5658	0,5260	1,4820	1,3104	
4		2003	0,6643	0,6601	–	–	
5	Разом за $\boldsymbol{i}$–й квартал сума ∑		**1,745**	**1,8501**	**4,1973**	**4,2850**	
6	Середня оцінка сезонної складової (ряд.5) /3 = $\boldsymbol{S_i}$		0,5817	0,6167	1,3991	1,4283	∑=4,0258
7	Коефіцієнт коригування «***k***»		4/4,0258=0,9936				
8	Коригована сезонна складова $S_i(k)$ = (ряд.6)×0,9936		0,5780	0,6128	1,3901	1,4192	4,0

Маємо $0{,}5817 + 0{,}6167 + 1{,}3991 + 1{,}4283 = 4{,}0258$, звідки коефіцієнт коригування $k = \frac{4}{4{,}0258} = 0{,}9936$. Скориговані значення сезонної компоненти $S_i(k)$ отримуємо добутком її середньої оцінки на коефіцієнт коригування k. Перевіряємо умову рівняння 4-х сум значень сезонної компоненти:

$$0{,}5780 + 0{,}6128 + 1{,}3901 + 1{,}4192 = 4{,}0.$$

Крок 3. Розділимо кожен рівень вихідного ряду на відповідні значення сезонної компоненти. В результаті отримаємо величини $T \times E = Y/S$ (графа 4 табл. 24), які містять тільки тенденцію і випадкову компоненту.

Крок 4. Визначимо тренд (*T)* в мультиплікативній моделі. Для цього розрахуємо параметри лінійного тренду, використовуючи рівні графи 4, *Y / S або T × E*.

В результаті отримаємо рівняння її тренда: $T = 651{,}59 + 3{,}2825 \times t$. Підставляючи в рівняння t = 1, 2,...,16, знайдемо значення *T* для кожного моменту часу (графа 5 табл. 24).

Таблиця 24.

№ кварталу	Кількість порушень Y_t	S_i	Y_t / S_i	T	T × S	E= Y_t / (T×S)
1	**2**	**3**	**4**	**5**	**6**	**7**
1	375	0,5780	648,789	654,873	378,517	0,991
2	371	0,6128	605,418	658,155	403,317	0,920
3	869	1,3901	625,135	661,438	919,465	0,945
4	1015	1,4192	715,192	664,720	943,371	1,076
5	357	0,5780	617,647	668,003	386,106	0,925
6	471	0,6128	768,603	671,285	411,363	1,145
7	992	1,3901	713,618	674,568	937,717	1,058
8	1020	1,4192	718,715	677,850	962,005	1,060
9	390	0,5780	674,740	681,133	393,695	0,991
10	355	0,6128	579,308	684,415	419,410	0,846
11	992	1,3901	713,618	687,698	955,969	1,038
12	905	1,4192	637,683	690,980	980,639	0,923
13	461	0,5780	797,578	694,263	401,284	1,149
14	454	0,6128	740,862	697,545	427,456	1,062
15	920	1,3901	661,823	700,828	974,221	0,944
16	927	1,4192	653,185	704,110	999,273	0,928

Крок 5. Знайдемо рівні ряду, помноживши значення T на відповідні значення сезонної компоненти S, (графа 6 табл. 24). На графіку відкладаємо фактичні значення рівнів часового ряду і теоретичні, отримані за мультиплікативною моделлю, рис.8.

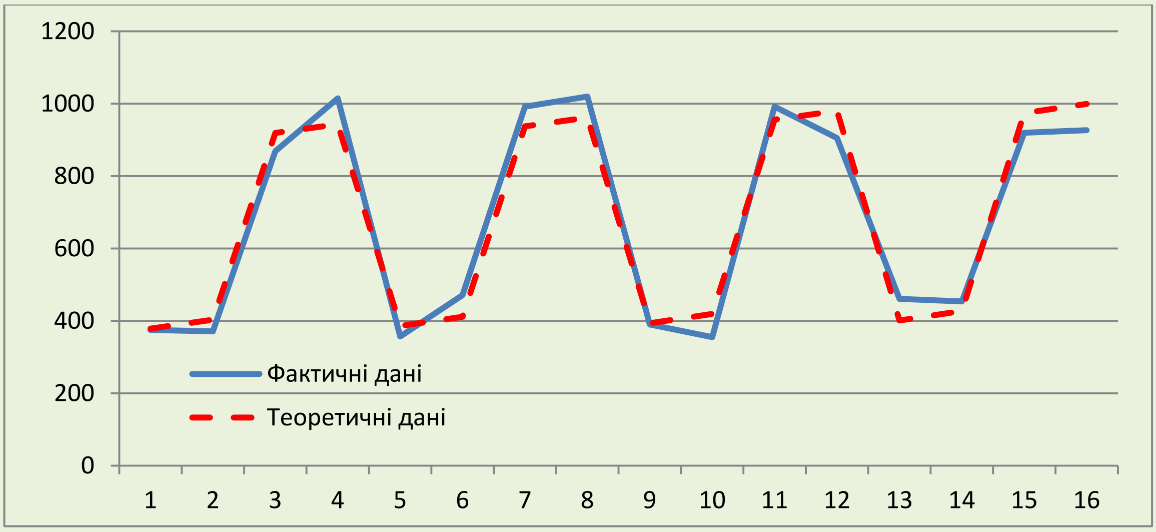


Рис. 8.

Розрахунок помилки в мультиплікативній моделі здійснюється за формулою: $E = \frac{Y}{T \times S}$

Для порівняння мультиплікативної моделі та інших моделей часового ряду можна, за аналогією з адитивною моделлю, знаходимо коефіцієнт детермінації використовуючи суму квадратів абсолютних помилок $(y_t - T \times S)^2$ та загальну дисперсію:

$$R^2 = 1 - \frac{\sum_1^n (y_t - T \times S)^2}{\sum_1^n (y_t - \bar{y})^2} = 1 - \frac{43071{,}66}{1252743{,}96} = 0{,}966.$$

Порівняння коефіцієнтів детермінації адитивної та мультиплікативної моделей дозволяє зробити висновок про те, що обидві моделі однаково апроксимують вихідні дані.

Крок 6. Прогнозування за мультиплікативною моделлю за прикладом правопорушень на I-й та II-й квартали 2004 року, прогнозне їх значення *F(t)* рівня часового ряду в мультиплікативної моделі є добуток трендової і сезонної компонент. для визначення трендової компоненти скористаємося рівнянням тренда $T = 651{,}59 + 3{,}2825 \times t$.

Отримаємо: $T(17) = 651{,}59 + 3{,}2825 \times 17 = 707{,}39.$

$T(18) = 651{,}59 + 3{,}2825 \times 18 = 710{,}68.$

Значення сезонних компонент за 1-й та 2-й квартали дорівнюють:

S_1 = 0,578 та S_2 = 0,6128.

А прогнозні значення: $F(17) = T(17) \times S1 = 687{,}49 \times 0{,}578 \approx 409.$

$F(18) = T(18) \times S2 = 688{,}42 \times 0{,}6128 \approx 436$

Тобто у перші два квартали 2004 року слід було очікувати відповідно 409 і 436 правопорушень.

Як бачимо адитивна та мультиплікативна моделі надають порівнянні результати.

Тестові завдання.

Парна регресія і кореляція

1. Найбільш наочним вибором рівняння парної регресії є:

а) аналітичний;

б). графічний**;**

в) експериментальний (табличний).

2. Розраховувати параметри парної лінійної регресії можливо за умов, якщо на кожну факторну ознаку у моделі регресії маємо:

а). не менше 5 спостережень**;**

б) не менше 7 спостережень;

в) не менше 10 спостережень.

3. **Суть методу** найменших квадратів полягає у:

а) мінімізації суми залишкових величин;

б) мінімізації дисперсії результативної ознаки;

в). мінімізації суми квадратів залишкових величин**.**

4. **Коефіцієнт лінійного** парного рівняння регресії:

а). показує середню зміну результату зі зміною фактора на одну одиницю;

б) оцінює статистичну значущість рівняння регресії;

в) показує, на скільки відсотків зміниться в середньому результативна ознака, якщо факторна зміниться на **1%.**

5. На підставі спостережень за 50 сім'ями побудовано рівняння регресії:

$y = 284{,}56 + 0{,}672x$ де у – споживання, х – дохід.

Чи відповідають знаки і значення коефіцієнтів регресії теоретичним уявленням?

а) так;

б) ні;

в). нічого певного сказати не можна**.**

6. **Суть коефіцієнта детермінації** R_{xy}^2 полягає у наступному:

а) оцінює якість моделі з відносних відхилень по кожному спостереженню;

б). характеризує частку дисперсії результативної ознаки, що пояснюється регресією у загальній дисперсії результативної ознаки**;**

в) характеризує частку дисперсії результативної ознаки викликану впливом не врахованих у моделі факторних ознак.

7. **Якість моделі з відносних відхилень** за кожним спостереженням оцінює:

а) коефіцієнт детермінації $\mathrm{R}_{\mathrm{xy}}^2$

б) F-критерій Фішера;

в). середня помилка апроксимації $\bar{A}$**.**

8. **Значущість** рівняння регресії в цілому оцінює:

а). F-критерій Фішера;

б) t-критерій Стьюдента;

в) коефіцієнт детермінації R_{xy}^2**.**

9. **Класичний метод** визначення коефіцієнтів регресії заснований на:

а). методі найменших квадратів**;**

б) методі максимальної правдоподібності;

в) регресійному аналізі.

10. **Залишкова сума** квадратів дорівнює нулю:

а). коли правильно підібрана регресійна модель;

б) коли між ознаками існує точний функціональний зв'язок**;**

в) ніколи.

11. **Сума квадратів** відхилень, що пояснює **парну модель регресії** має число ступенів свободи:

а) n–1;

б). 1**;**

в) n–2.

12. **Сума квадратів** відхилень **залишків** у лінійній парній моделі має число ступенів свободи:

а) n–1;

б) 1;

в). n–2**.**

13. **Загальна** сума квадратів відхилень у лінійній парній моделі має число ступенів свободи:

а). n–1**;**

б) 1;

в) n–2.

14. Для оцінки **значущості коефіцієнтів** регресії розраховують:

а). F-критерій Фішера;

б) *t*-критерій Стьюдента**;**

в) коефіцієнт детермінації R_{xy}^2

15. Яке рівняння регресії не можна звести до **лінійного вигляду**:

а) $y = a + b \times \ln x$**;**

б) $y = ax^b$;

в) $y = a + b \times x^c$;

г) жодне;

16. Яке з рівнянь є **степеневим**:

а) $y = a + b \times \ln x$;

б). $y = ax^b$**;**

в). $y = a + b \times x^c$**.**

17. **Параметр** «*b*» в степеневій моделі є:

а) коефіцієнтом детермінації;

б). коефіцієнтом еластичності**;**

в) коефіцієнтом кореляції.

18. **Коефіцієнт кореляції** r_{xy} може приймати значення:

а). від [-1 до +1]**;**

б) від 0 до 1;

в) будь-які.

19. Яке з наступних рівнянь **нелінійне** за параметрами, що оцінюються:

а) $y = a + b \times x + \varepsilon$;

б) $y = a + b \times \ln x + \varepsilon$;

в). $y = ax^b \times \varepsilon$**.**

XVIII. Методи декомпозиції[80].

Ефективність управління визначається адекватним розумінням сутності керованого об'єкта. Складні процеси, що відбуваються в економіці, впливають на перспективи її розвитку та стан соціальної сфери і викликають необхідність дослідження їхньої спрямованості, природи, а також факторів, що на них впливають.

Мета декомпозиція є вирішення глобальної задачі процесу через серію більш простих задач підсистем спрощених за структурою процесу. Декомпозиція (або децентралізація) складається із спрощення процесу, у цілому, до його простих складових, що містять окремі його властивості. Метод застосовується для обґрунтування структури побудови складних систем, зокрема, для обґрунтування схем взаємодії між рівнями економічних систем при вирішенні задач планування та управління.

Декомпозиція дозволяє аналізувати будь-яку складну систему за окремими взаємопов'язаними з нею її підсистемами, що, в свою чергу, також можуть бути спрощені на більш прості їх частини. Дослідженню підлягають матеріальні об'єкти, динамічні процеси, явища. Початкова композиційна складна система характеризується градацією нульового рівня. Після її спрощення (декомпозиції) отримують підсистеми першого рівня. Подальше спрощення підсистем першого рівня призводить до появи підсистем другого рівня і т.д.

Система розкладається тільки по одній ознаці, постійній для всіх рівнів.

Ознаками декомпозиції (спрощення) можуть виступати такі фактори як:

- складові частини за функціональним призначенням;
- складові за конструктивними елементами (чи за їхніми матеріалами та ін.);
- структурні ознаки (вид схеми, способи та ін.);
- складові частини етапів чи процесів (життєвий цикл, фізичний стан тощо);
- предметні характеристики (економічні, інформаційні, технологічні та ін.);
- інші.

Ступінь подробиць опису та кількість рівнів отриманої ієрархічної структури визначають вимогами потреби та зручності працюючого з нею фахівця.

На практиці зміна певних характеристик багатьох економічних процесів є результатом комбінації їхніх різних компонент, основними з яких є:

1. **Тенденція «Т»** (тренд), що визначає загальну тенденцію розвитку на довгострокову перспективу;
2. **Циклічність «С»**, яка пояснює чергування глобальних періодів; фази розширення та стискання; етапи підйому та падіння економіки. Ця компонента часового ряду називається тренд-циклічною, тобто включає в себе коливання відносно тривалого періоду часу. Взагалі, це глобальна циклічна складова з періодом десятка років, як наприклад для світової економіки чи сонячної активності, тощо.

[80] Децентралізації (спрощення) системи.

3. **Сезонні цикли «S»**, що пов'язані з сезонністю року або з економічним та соціальним життям суспільства, наприклад, свята, канікули, розпродажі і таке інше;
4. **Помилки «E»** – складові значення певної характеристики, що обумовлені випадковими факторами.

В залежності від комбінування окремих компонент у моделі розрізняють адитивні, мультиплікативні та змішані моделі декомпозиції.

1. Модель **адитивної** декомпозиції – є модель у якій вплив окремих компонент на результативну ознаку підсумовується: $y(t) = T(t) + C(t) + S(t) + E$.
2. Модель **мультиплікативної** декомпозиції – є моделлю у якій вплив окремих компонент на результативну ознаку обчислюються, як їх добуток: $y(t) = T(t) \times C(t) \times S(t) \times E$.
3. У **змішаній формі** моделі її складові частини можуть як додаватися, так і перемножуватись.

Тенденція (тренд).

На графіку нижче, прямою лінію показаний вихідний часовий ряд за його зростанням, що описаний формулою $f(t) = T(t) = 9{,}51t + 1132{,}5$.

Така тенденція (тренд) називається лінійною оскільки згладжений вид графіку нагадує пряму лінію. Це найбільш вживаний вид тренду, рідше зустрічаються поліноміальні, експоненціальні, логарифмічні та інші тренди. Вибравши лінійний вид тренду, його конкретні параметри підбирають методом найменших квадратів.

Циклічна глобальна складова характерна для моделі з періодом десятка років, а оскільки глобальних проблем не вирішуємо, то і циклічну компоненту не розглядаємо.

Сезонні цикли (декомпозиція).

Однак на практиці виявляється недостатньо моделювати поведінку за монотонним характером зростання чи спаду ряду. Справа в тому, що розгляд конкретних даних продажів надає підстави твердження про наявність періодичного повторювання значень за шаблоном сезонності, як закономірності ряду. Наприклад, розглядаючи продаж морозива, вочевидь, що взимку обсяг його продажів нижче середнього їх значення. Таке повторювання даних продажів за сезоном абсолютно логічне, тому виникає питання, як можна використати цю інформацію для зменшення прогнозування невизначеності, тобто ризиків?

Значна кількість економічних процесів мають сезонний характер, так виникає поняття «сезонності» у прогнозуванні – тобто повторюваність характеристик процесів за величиною їх значень через певні проміжки часу. Наприклад, зростання обсягів продажів ялинкових іграшок в останні тижні року можна розглядати як сезонність. Зростання обсягів продажів супермаркету в кінці тижня можна розглядати як сезонність з тижневою періодичністю.

Складова моделі «сезонності» необов'язково пов'язана саме з сезоном у побутовому розумінні (пори року). У моделюванні, будь-яка періодичність може називатися сезонністю. З цієї точки зору сезонність характеризується насамперед періодом або **лагом сезонності** – тобто кількістю періодів, через яке відбувається поновлення повторення. Наприклад, якщо маємо ряд місячних продажів, то можемо припускати, що період становить 12 місяців. Важливо виявити та відокремити вплив сезонних факторів за їх періодичністю.

Розрізняють моделі з адитивною і мультиплікативною сезонністю.

В адитивній моделі сезонна поправка додається або вираховується за формулою $y(t) = T(t) + S(t)$, тобто у лютому місяці продаємо на 350 одиниць (більше) менше ніж в середньому, у другому здійснюється множення на коефіцієнт сезонності за формулою $y(t) = T(t) \times S(t)$, у лютому місяці продаємо в 1,15 менше (більше) ніж в середньому.

Зауважимо, що сама наявність сезонності має бути пояснена з точки зору доцільності. Сезонність є проявом властивості товару (особливостей його споживання за певним місцем). Якщо ми зможемо ідентифікувати сезонність товару, то ми будемо впевнені, що такі коливання триватимуть і в майбутньому. При цьому один і той же товар може мати характеристики сезонності, що різнитимуться за місцем споживання. Якщо пояснити сезонність з точки зору доцільності неможливо, немає і підстав для застосування такого шаблону. Можливо є інші фактори, зовнішні по відношенню до товару, що пояснюватимуть сенс сезонності.

При виборі тренду ми маємо вибирати просту аналітичну функцію (парної чи більшої за кількістю ознак регресії), тоді, як сезонність виражається табличною функцією.

Найпоширеніший випадок – річна сезонність з 12 періодами за кількістю місяців – це таблиця з 11 мультиплікативних коефіцієнтів, що представляють поправку щодо базового місяця.

Або 12 коефіцієнтів щодо середньомісячного значення, важливо тільки, що незалежними залишаються лише 11, оскільки 12й визначається з вимоги $\sum_{i=1}^{12} S_i = 12$, в цьому випадку маємо 11 ступенів вільності. Якщо маємо модель з лінійним трендом та періодом 12 місяців то загальна кількість ступенів вільності визначиться за сумою складових ступенів вільності[81] тренду та сезонності.

Класична сезонна декомпозиція.

Декомпозиція ряду продажів.

Ми часто спостерігаємо спрямованість графіку продажів за рядом значень, в якому присутні компоненти тренда і сезонності. Якщо відома послідовність значень ряду у минулому, то можливо поліпшити якість його прогнозованих значень у майбутньому. Для здійснення такого прогнозу з фактичних значень ознак виключимо тренд і сезонність і тим самим значно зменшимо невизначеність майбутнього.

Процедура видалення невипадкових складових значень моделі з її фактичних даних має назву **декомпозиції**.

Сезонна декомпозиція це визначення значень числових сезонних коригувань. Для їх визначення візьмемо найбільш поширений випадок:

- дані про продажі згруповані помісячно (оскільки потрібний прогноз з точністю до місяця), передбачається лінійний тренд і сезонність з лагом 12 за мультиплікативною моделлю.

Згладжування ряду.

Згладжування це процедура заміни вихідного ряду іншим, більш плавним, що розрахований на вихідному. Метою такої процедури є встановлення загальних тенденцій.

Методів згладжування існує багато, найбільш поширені з них такі:

- **укрупнення часових інтервалів**. Очевидно, що ряд продажів, агрегований помісячно, веде себе більш гладко, ніж ряд, заснований на денних продажах;
- **ковзке середнє**;
- **аналітичне вирівнювання**. У цьому випадку вихідний ряд замінюється деякою аналітичною функцією, вид і параметри якої підбираються за мінімумом помилок.

Ми будемо використовувати згладжування методом змінного середнього. Ідея полягає в тому, що набір з декількох точок ми замінюємо однією за принципом «центру мас» – значення якої дорівнює середньому цих точок, а сам «центр мас» знаходиться у центрі відрізка утвореного крайніми точками. Таким чином ми встановлюємо якийсь «середній» рівень для цих точок.

[81] Кількість ступенів вільності тренду залежить від аналітичного виразу за яким цей тренд описується.

Для того, щоб провести сезонну декомпозицію, класичний підхід пропонує спочатку провести згладжування ряду за датою, що в **точності збігається з лагом сезонності**.

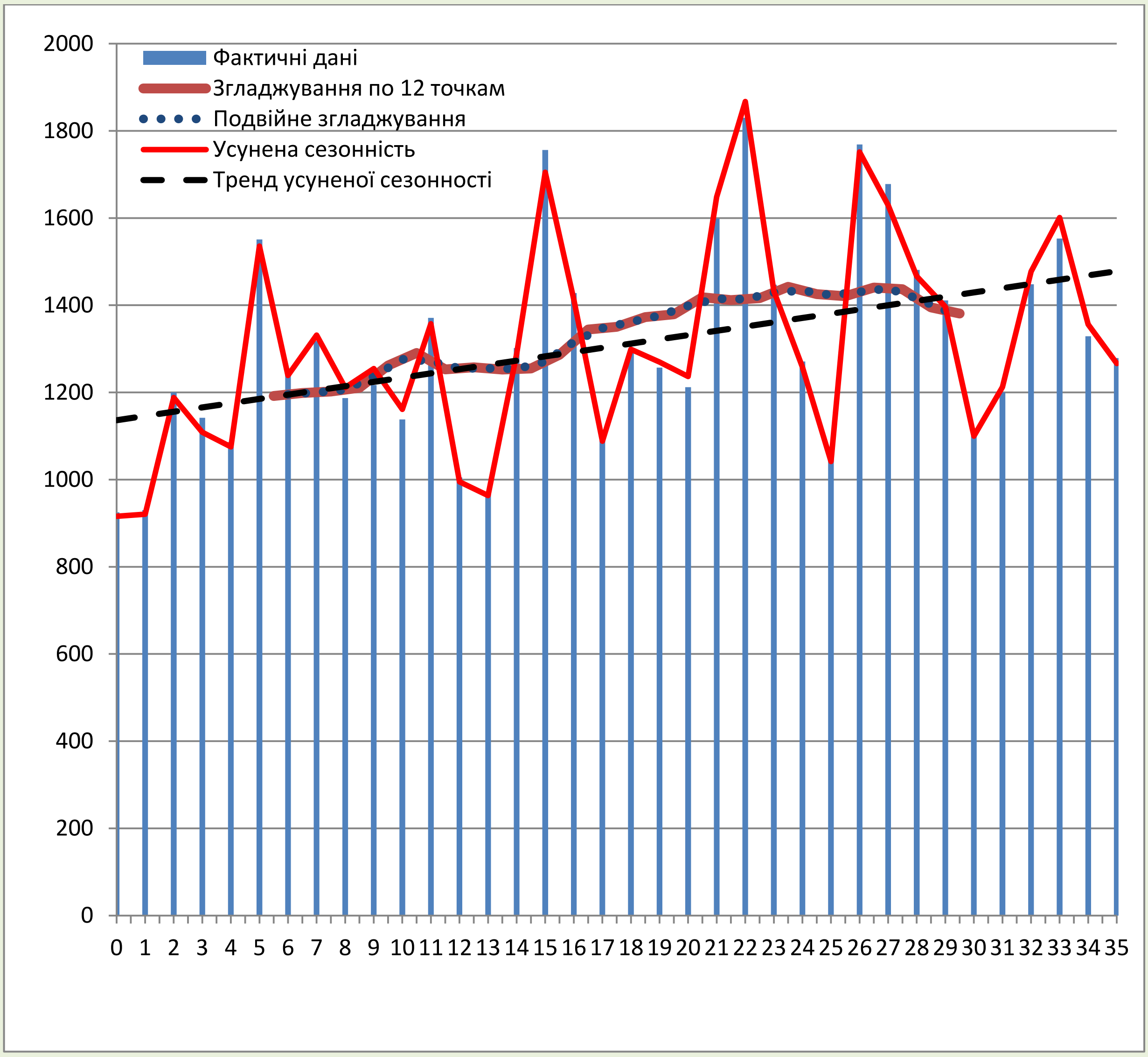
У нашому випадку лаг дорівнює 12, і якщо згладжування буде по 12 точках, то збурення, що пов'язані з сезонністю, нівелюються і ми отримаємо загальний середній рівень.

Порівнювання фактичних продажів зі згладженими відбуватиметься**:**

- для адитивної моделі – відніманням (додаванням) з фактичних значень значення згладженого ряду;
- для мультиплікативної – діленням (множенням).

В результаті отримаємо набір коефіцієнтів, для кожного місяця (залежно від довжини ряду). Якщо згладжування пройшло успішно, ці коефіцієнти будуть мати не надто великий розкид. На графіку наш ряд, згладжений по 12 точках та подвійним згладжуванням.

Графік згладжування та усунення сезонності, а також тренда часового ряду позбавленого сезонності.



Розрахунки згладжування наведені в табл. 25 (мультиплікативна модель).

Таблиця 25.

№	Дата мі-сяць	Обсяг факт	Згладже-но за **12 позиція-ми**	Подвійне згладжу-вання	**Сезон-ність**, коефіцієнт	**Сезонність усереднена** коефіцієнтом не нормованим	нормова-ним на одиницю	відновленим за аналогом нормованого	Значення без се-зонності	Тренд по усуненій сезонності[82] 9,73x+1137,05	Прогноз одиниць
		Од.	одиниць	одиниць	одиниць	одиниць	одиниць	одиниць	грн.	грн.	грн.
	A	B	C	D	E	F	G		H	I	J
1	**0**	925						(G25)=1,01	915,84[83]	1137,05[84]	
2											
3	**1**	930						(G27)=1,01	920,79	1146,78	
4											
5	**2**	1200						(G29)=1,01	1188,12	1156,52	
6											
7	**3**	1142						(G31)=1,03	1108,74	1166,25	
8											
9	**4**	1086						(G33)=1,01	1075,25	1175,98	
10											
11	**5**	1551						(G35)=1,01	1535,64	1185,71	
12			1192,00[85]								
13	**6**	1239		1195,34[86]	1,04[87]	1,00[88]	1,00[89]		1239,00	1195,44	
14			1198,67								
15	**7**	1318		1200,46	1,10	0,99	0,99		1331,31	1205,17	
16			1202,25								
17	**8**	1187		1206,50	0,98	0,98	0,98		1211,22	1214,90	
18			1210,75								
19	**9**	1217		1236,34	0,98	0,97	0,97		1254,64	1224,63	
20			1261,92								
21	**10**	1138		1276,17	0,89	0,98	0,98		1161,22	1234,37	
22			1290,42								
23	**11**	1371		1271,59	1,08	1,01	1,01		1357,43	1244,10	
24			1252,75								
25	12	1005		1255,25	0,80	1,00	1,01		995,05	1253,83	
26			1257,75								
27	13	973		1255,21	0,78	1,01	1,01		963,37	1263,56	
28			1252,67								
29	14	1302		1253,71	1,04	1,01	1,01		1289,11	1273,29	
30			1254,75								
31	15	1756		1270,67	1,38	1,03	1,03		1704,85	1283,02	
32			1286,58								
33	16	1428		1315,42	1,09	1,01	1,01		1413,86	1292,75	
34			1344,25								
35	17	1099		1347,42	0,82	1,01	1,01		1088,12	1302,48	
36			1350,58								
37	18	1299		1361,67	**0,95**	1,02		(G13)=1,00	1299,00	1312,22	
38			1372,75								
39	**19**	1257		1376,04	0,91			(G15)=0,99	1269,70	1321,95	
40			1379,33								
41	20	1212		1398,79	0,87			(G17)=0,98	1236,73	1331,68	
42			1418,25								
43	21	1599		1415,00	1,13			(G19)=0,97	1648,45	1341,41	
44			1411,75								
45	22	1830		1413,96	1,29			(G21)=0,98	1867,35	1351,14	
46			1416,17								
47	23	1447		1429,17	1,01			(G23)=1,01	1432,67	1360,87	
48			1442,17								
49	24	1271		1433,88	0,89			(G25)=1,01	1258,42	1370,60	
50			1425,58								
51	25	1052		1423,21	0,74			(G27)=1,01	1041,58	1380,33	
52			1420,83								
53	26	1769		1430,67	1,24			(G29)=1,01	1751,49	1390,07	

[82] Див. дисперсійний аналіз далі;
[83] B1/G1;
[84] 1137,05+9,73×0=1137,05; (див. табл. 26).
[85] СУММ(C1:C23)/12=1192; або за функцією СРЗНАЧ(C1:C23) програми excel;
[86] СУММ(C12:C14)/2=1195,34;
[87] B13/D13;
[88] СУММ(E13:E35)/12;або за функцією СРЗНАЧ(E13:E35) програми excel;
[89] F13/СУММ($G$13:$G$35)*12;

№	Дата місяць	Обсяг факт	Згладжено по 12 позиціям	Подвійне згладжування	**Сезонність**, коефіцієнт	**Сезонність усереднена** коефіцієнтом не нормованим	нормованим на 1	відновленим за аналогом нормованого	Значення за усуненої сезонності	Тренд по усуненій сезонності: 9,73x+1137,05	Прогноз одиниць
		Од.	одиниць	одиниць	одиниць	одиниць	одиниць	грн.	грн.	грн.	грн.
	A	*B*	*C*	*D*	*E*	*F*		*G*	*H*	*I*	*J*
54			1440,50								
55	27	1678		1438,59	1,17			(G31)=1,03	1629,13	1399,80	
56			1436,67								
57	28	1481		1415,80	1,05			(G33)=1,01	1466,34	1409,53	
58			1394,92								
59	29	1411		1387,92	1,02			(G35)=1,01	1397,03	1419,26	
60			1380,92								
61	30	1100						(G13)=1,00	1100,00	1428,99	
62											
63	31	1200						(G15)=0,99	1212,12	1438,72	
64											
65	32	1448						(G17)=0,98	1477,55	1448,45	
66											
67	33	1553						(G19)=0,97	1601,03	1458,18	
68											
69	34	1329						(G21)=0,97	1356,12	1467,92	
70											
71	35	1279						(G23)=1,01	1266,34	1477,65	
72	36									1487,38	1502,3[90]
73	37									1497,11	1512,08
74	38									1506,84	1521,91
75	39									1516,57	1562,07
76	40									1526,30	1541,56
76	41									1536,03	1551,39
77	42									1545,77	1545,77
78	43									1555,50	1539,95
79	44									1565,23	1533,93
80	45									1574,96	1527,71
81	46									1584,69	1553,00
82	47									1594,42	1610,36

Дисперсійний аналіз для ряду з усуненою сезонністю наведено в табл. 26

Таблиця 26

Регресійна статистика	
Множинний коефіцієнт кореляції R	0,4324
Коефіцієнт детермінації R^2	0,1870
Нормований R^2	0,1631
Стандартна помилка	216,92
Спостереження	36

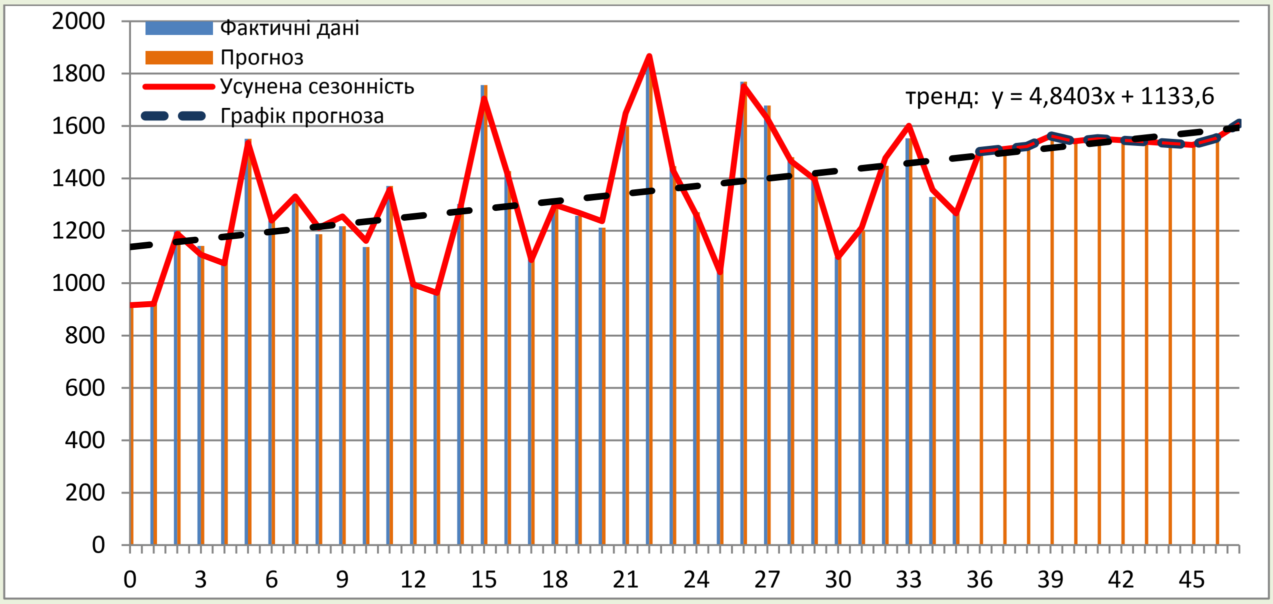
Дисперсійний аналіз

	df	*SS*	*MS*	*F*	*Значущ. F*
Регресія	1	367897,9	367897,9	7,81879	0,008445
Залишок	34	1599803,7	47053,05		
Разом	35	1967701,63			

Коефіцієнт	*значення*	*Станд помилка*	*t-стат.*	*Р-Значення*	*Нижні 95%*	*Верхні 95%*
Y-перетину	1137,05	70,82	16,05	0,00	993,12	1281,00
При факторній X_1	9,73	3,48	2,80	0,00	2,66	16,80

За даними таблиці з усуненою сезонністю (стовп. Н) розрахуємо лінійний тренд і складемо прогноз на наступних 12 місяців за трендом. Прогнозні значення на 12 місяців знайдемо, як добуток прогнозних значень тренда на його коефіцієнт сезонності. З врахуванням додаткових прогнозних значень новий тренд дещо відмітний від тренду за усуненої сезонності.

[90] $\boldsymbol{I}(72) \times \boldsymbol{H}(29) = 1487{,}38 \times 1{,}01 = 1502{,}3$



Висновки.

Як видно з графіку, очищені від сезонності дані не вкладаються у лінійну залежність маючи значні відхилення. Можливо, позбавившись від викидів у вихідних даних, все зміниться. Для більш точного визначення сезонності за класичною декомпозицією бажано мати не менше 4-5 повних циклів даних, оскільки один цикл не забезпечує точність в обчисленні коефіцієнтів. Розраховані прогнози значень, як бачимо на графіку, за тенденцією наближені до лінійної залежності.

Прогнозування різними методами.

Прогнозування методом ковзкої середньої.

Екстраполяція це підхід, що заснований на методології розповсюдженні минулих і поточних тенденцій, закономірностей, а також зв'язків досліджуваного об'єкта на прогноз його майбутнього розвитку. До методів екстраполяції відносять: методи ковзкої середньої, експоненціального згладжування, найменших квадратів. Метод ковзких середніх найбільш розповсюджений метод згладжування часових рядів. Застосовуючи його можливо елімінувати[91] випадкові коливання і отримати значення результату за впливом суттєвих чинників.

Згладжування за допомогою ковзких середніх засноване на тому, що в середніх величинах взаємно гасяться випадкові відхилення. Це відбувається внаслідок заміни первинних значень часового ряду середньої арифметичної значенням всередині обраного інтервалу часу. Отримане значення відповідне за його рівнем відноситься до середини обраного інтервалу часу (періоду).

Потім період зсувається на одне значення, і розрахунок середньої повторюється. При цьому **довжина (рівень) періоду часу** загладжуваної середньої залишається незмінним. Таким чином, в кожному розглянутому випадку середня центрована, тобто віднесена до серединної точці інтервалу згладжування і являє собою рівень для цієї точки.

[91] виключити

При згладжуванні часового ряду ковзкими середніми в розрахунках беруть участь всі рівні ряду. Чим ширше інтервал згладжування, тим більш плавним виходить тренд. Згладжений ряд коротше початкового на (*n-1*) спостережень, де «*n*» – рівень (довжина) інтервалу згладжування. При великих значеннях «*n*» коливання згладженого ряду значно знижується. Одночасно помітно скорочується кількість їх значень, що створює труднощі.

Вибір інтервалу згладжування залежить від цілей дослідження. При цьому слід дотримуватися періоду часу в якому відбувається подія, а отже, і усунення впливу випадкових факторів.

Метод прогнозування ковзкою середньою використовується при короткострокових за довжиною рівнів часових періодів. Його формула: $y_{t+1} = m_{t-1} + \frac{1}{n}(y_t - y_{t-1})$, де ($t$+1) – прогнозний період; t – період, що передує прогнозному періоду (рік, місяць і т.д.); $У_{t+1}$ – прогнозований показник; m_{t-1} – змінна середня за два періоди до прогнозного; «*n*» – число рівнів, що входять в інтервал згладжування; $У_t$ – фактичне значення досліджуваного явища за попередній період; $У_{t-1}$ – фактичне значення досліджуваного явища за два періоди, що передують прогнозному.

Приклад застосування методу ковзкої середньої для розробки прогнозу.

Задача 1.

Є дані, що характеризують рівень безробіття в регіоні у відсотках (%).

Січень	Лютий	Березень	Квітень	Травень	Червень	Липень	Серпень	Вересень	Жовтень
2,99	2,66	2,63	2,56	2,40	2,22	1,97	1,72	1,56	1,42

Побудувати прогноз безробіття на листопад, грудень і січень місяці, методами:

1) згладжування ковзкою середньою;
2) експоненціального згладжування;
2) найменших квадратів.

Розрахувати помилки отриманих прогнозів за кожним методом та надати висновки щодо отриманих результатів за методами.

Рішення 1. Розрахунок помилок методом ковзкої середньої.

Для розрахунку прогнозного значення методом ковзкої середньої необхідно:

1. Визначити інтервал за рівнем періоду згладжування, у нашому випадку *n* = 3.
2. Розрахувати ковзку середню для перших трьох місяців січня, лютого та березня:

$m_{\text{лютого}} = (У_{\text{січ}} + У_{\text{лют}} + У_{\text{бер}})/3 = (2{,}99 + 2{,}66 + 2{,}63)/3 = 2{,}76.$

Отримане значення заносимо в табл. 27, «стовп. ***с***» в середину періоду за рівнем.

Далі розраховуємо ковзку середню «*m*» для наступного рівня за місяцями лютий, березень, квітень $m_{\text{березня}} = (У_{\text{лют}} + У_{\text{бер}} + У_{\text{кві}})/3 = (2{,}66 + 2{,}63 + 2{,}56)/3 = 2{,}62.$

По аналогії розраховуємо ковзку середню «*m*» місяця для інших періодів.

3. Розрахувавши **ковзку середню** для всіх періодів для *n* = 3, будуємо прогноз

- на **листопад** місяць за формулою: $y_{t+1} = m_{t-1} + \frac{1}{n}(y_t - y_{t-1})$,

  та заносимо у **прогнозні показники** табл.27, «стовп. *b*».

  $У_{листопаду} = 1{,}57 + \frac{1}{3} \times (1{,}42 - 1{,}56) = 1{,}57 - 0{,}05 = 1{,}52.$

  Розраховуємо ковзку середню «*m*» *n* – рівнів для жовтня.

  $m_{жовтня} = (1{,}56 + 1{,}42 + 1{,}52)/3 = 1{,}5$, будуємо прогноз

- на **грудень**, та заносимо у табл.27, «стовп. *b*».

  $У_{грудень} = 1{,}5 + 1/3 \times (1{,}52 - 1{,}42) = 1{,}53.$

  Розраховуємо ковзку середню «*m*» *n* -рівнів для листопаду.

  $m_{листопаду} = (1{,}42 + 1{,}52 + 1{,}53)/3 = 1{,}49$, будуємо прогноз

- на **січень**, та заносимо у табл.27, «стовп. *b*».

  $У_{січень} = 1{,}49 + 1/3 \times (1{,}53 - 1{,}52) = 1{,}49.$

Таблиця 27

Місяць	Рівень безробіття **фактичний** $У_t$, (%)	розрахована **Ковзка** середня для рівня *n*=3	Відносна помилка
a	***b***	***c***	***d***
Січень	2,99%		
Лютий	2,66%	**2,76%**	**3,76%**[92]
Березень	2,63%	2,62%	0,38%
Квітень	2,56%	2,53%	**1,17%**
Травень	2,40%	2,39%	0,42%
Червень	2,22%	2,20%	0,90%
Липень	1,97%	1,97%	0,00%
Серпень	1,72%	1,75%	1,74%
Вересень	1,56%	1,57%	0,64%
Жовтень	1,42%	**1,50%**	
Прогнозні показники			**∑=9,01%**
Листопад	**1,52%**	**1,49%**	
Грудень	**1,53%**	**1,51%**	
Січень	**1,49%**		

Розрахуємо **середню відносну помилку** за формулою:

$\varepsilon = \frac{1}{n} \times \left[\frac{|y_{факт} - y_{ковзка}|}{y_{факт}}\right] = \frac{9{,}01}{8} = 1{,}13\%$, що < 10% - точність прогнозування велика.

Точність прогнозу в межах 10-20% є доброю. Далі вирішимо задачу прогнозування методами експоненціального згладжування та найменших квадратів.

[92] $|2{,}66 - 2{,}76|/2{,}66 \times 100 = 3{,}76\%$

Прогнозування методом експоненціального згладжування.

Метод експоненціального згладжування найбільш ефективний при прогнозах середньострокових часових періодів за довжиною рівнів. Він надає прогноз тільки на один період вперед. Основні переваги – простота процедури обчислень і можливість обліку ваги значень вихідної інформації. Формула експоненціального згладжування: $U_{t+1} = \alpha \times y_t + (1 - \alpha) \times U_t$

де t – період, що передує прогнозному;
$t+1$ – прогнозний період;
U_{t+1} – прогнозований показник;
α – параметр згладжування;
y_t – фактичне значення показника за період, що передує прогнозному;
U_t – експоненціально зважена середня для періоду, що передує прогнозному.

При прогнозуванні цим методом виникають деякі ускладнення щодо

• вибору значення параметра експоненціального згладжування «α»; та
• встановлення початкового експоненціального значення ряду «U_0».

Від величини «α» залежить швидкість зниження ваги впливу попередніх значень. Чим більше «α», тим менший вплив значень попередніх років. Якщо значення «α» близьке до одиниці, то при прогнозуванні, більший вплив надають останні значення. Якщо значення «α» близьке до нуля, то ваги, якими зважуються рівні значень тимчасового ряду, спадають повільніше, тобто при прогнозі враховуються всі (або майже всі) минулі значення.

Тобто, якщо є впевненість, що початкові умови, на підставі яких розробляється прогноз, достовірні, слід використовувати невелике значення параметра згладжування ($\alpha \to 0$). Коли параметр згладжування малий, то досліджувана функція поводиться як середня великої кількості минулих значень. Якщо немає достатньої впевненості у початкових умовах прогнозування, то слід використовувати велике значення «α», що при прогнозуванні надасть перевагу значенням останніх спостережень. Точного методу для вибору оптимального значення параметра згладжування «α» немає. В окремих випадках пропонується[93] визначати величину «α», виходячи з довжини інтервалу згладжування. При цьому «α» обчислюється за формулою: $\alpha = \frac{2}{n+1}$, де n – кількість значень спостережень, що входять в інтервал згладжування.

Вибір U_0 (експоненціально зваженого середнього початкового) вирішується так:

- якщо існують вихідні дані про розвиток явища в минулому, то за «U_0» можна скористатися середньою арифметичною;
- якщо їх немає, то в якості прогнозу «U_0» використовують перше значення бази $У_1$.

Також можна скористатися і експертними її оцінками.

Відзначимо, що при вивченні часових рядів і прогнозуванні економічних процесів метод експоненціального згладжування не завжди «спрацьовує». Це обумовлено тим, що економічні часові ряди бувають занадто короткими (15-20 значень), і в разі, коли темпи зростання, тобто приросту великі, даний метод не «встигає» відобразити всі зміни.

[93] Автор методу проф. Браун.

Приклад застосування методу експоненціального згладжування для розробки прогнозу за умовами задачі 1, розглянемо далі.

Січень	Лютий	Березень	Квітень	Травень	Червень	Липень	Серпень	Вересень	Жовтень
2,99	2,66	2,63	2,56	2,40	2,22	1,97	1,72	1,56	1,42

- Потрібно побудувати прогноз рівня безробіття в регіоні на листопад, грудень та січень місяці, використовуючи метод експоненціального згладжування.
- Розрахувати помилку отриманого прогнозу цього методу.

Рішення 2. Розрахунок помилок методом експоненціального згладжування.

1). Визначаємо значення параметра згладжування за формулою: $\alpha = \frac{2}{n+1}$, де «*n*» – кількість значень, що входять в інтервал довжини згладжування $\alpha = 2/(10+1) = 0{,}18$.

2). Визначаємо початкове значення U_o двома способами:

- як середня арифметична $U\text{о} = (2{,}99 + 2{,}66 + 2{,}63 + 2{,}56 + 2{,}40 + 2{,}22 + 1{,}97 + 1{,}72 + 1{,}56 + 1{,}42)\ /\ 10\ =\ 22{,}13\ /\ 10\ =\ 2{,}21.$
- за прогнозне початкове приймаємо фактичне перше значення ряду *U*о=2,99.

3). Розраховуємо експоненціально зважену середню для кожного періоду, використовуючи формулу: $\mathrm{U_{t+1}} = \alpha \times \mathrm{y_t} + (1-\alpha) \times \mathrm{U_t}$, кожного зі способів і так далі, див. нижче,

Місяць	Перший спосіб	Другий спосіб
$U_{лют}$	$2{,}99 \times 0{,}18 + (1-0{,}18) \times 2{,}21 = 2{,}35$	$2{,}99 \times 0{,}18 + (1-0{,}18) \times 2{,}99 = 2{,}99$
$U_{бер}$	$2{,}66 \times 0{,}18 + (1-0{,}18) \times 2{,}35 = 2{,}41$	$2{,}66 \times 0{,}18 + (1-0{,}18) \times 2{,}99 = 2{,}93$
$U_{кві}$	$2{,}63 \times 0{,}18 + (1-0{,}18) \times 2{,}41 = 2{,}45$	$2{,}63\times 0{,}18 + (1-0{,}18) \times 2{,}93 = 2{,}88$

а результати зводимо в табл. 28.

4) Розраховуємо середню відносну помилку: $\varepsilon = \frac{1}{n} \times \sum_{\mathrm{i}=1}^{n} \frac{|y_{\text{факт}} - y_{\text{розрахов}}|}{y_{\text{факт}}}$.

Таблиця 28.

Найменування	Рівень ($U_ф$) безробіття (%)	Експоненціальна ($U_р$) зважена розрахована		Відносна помилка ($U_ф$-$U_р$)/$U_ф$	
Місяці	**фактичні**	**1-й спосіб**	**2-й спосіб**	**1-й спосіб**	**2-й спосіб**
Січень	2,99%	2,21%	2,99%	26,09%	0,00%
лютий	**2,66%**	**2,35%**	**2,99%**	**11,65%**	**12,41%**
березень	**2,63%**	**2,41%**	**2,93%**	**8,37%**	**11,41%**
квітень	**2,56%**	**2,45%**	**2,88%**	**4,30%**	**12,50%**
травень	2,40%	2,47%	2,82%	2,92%	17,50%
червень	2,22%	2,46%	2,74%	10,81%	23,42%
липень	1,97%	2,42%	2,65%	22,84%	34,52%
серпень	1,72%	2,34%	2,53%	36,05%	47,09%
вересень	1,56%	2,23%	2,38%	42,95%	52,56%
жовтень	1,42%	2,11%	2,23%	48,59%	57,04%
Разом				214,57%	268,45%
	Відносна середня помилка			**21,46%**	**26,85%**
	Прогноз				
листопад		1,99%	2,08%		

У кожному випадку ***точність прогнозу*** **є задовільною** оскільки середня відносна помилка потрапляє в межі 20-50%. Точність прогнозу, що перевищує 50% є незадовільною.

Прогнозування методом найменших квадратів.

Суть методу найменших квадратів полягає в мінімізації суми квадратів відхилень між фактичними, що спостерігаються і розрахунковими значеннями. Розрахункові значення знаходяться по підібраному рівнянню регресії. Чим менше відхилення між фактичними значеннями і розрахунковими, тим точніший прогноз, побудований за рівнянням регресії.

Теоретичний аналіз сутності досліджуваного процесу, зміна якого відображається тим-часовим часовим рядом, служить основою для вибору кривої. Іноді приймаються до уваги міркування про характер росту рівнів часового ряду. Так, якщо зростання обсягу випуску продукції очікується за арифметичною прогресією, то згладжування здійснюється прямою. Якщо ж виявляється, що зростання йде за геометрично прогресією, то згладжування треба здійснювати показниковою функцією.

Формула до якої застосовується метод найменших квадратів є рівняння прямої лінії:

$Y_{t+1} = b_0 + b_1 x_i$, де (t+1) – прогнозний період; $У_{t+1}$ – прогнозований показник; та «b_i» – коефіцієнти; x – умовне позначення часу. Розрахунок коефіцієнтів «b_i» здійснюється формулами: $b_1 = \frac{\frac{1}{n}\left(\sum_{i=1}^{n} y_i x_i\right) - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2}$; $b_0 = \bar{y} - b_1\bar{x}$, де У – фактичні значення часового ряду динаміки; «n» – довжина рівнів часового ряду;

Згладжування часових рядів методом найменших квадратів має на меті відображення закономірності розвитку досліджуваного явища. В аналітичному вираженні функції тренда час розглядається як незалежна змінна, а рівні ряду виступають як результативна ознака цієї незалежної змінної.

Розвиток процесу залежить не від віддаленості у минуле, з начального моменту, а від факторів, що впливали на його розвиток, в якому напрямку і з яким прискоренням. Звідси ясно, що розвиток процесу в часі виступає як результат дії цих факторів.

Правильно встановити тип кривої, тип аналітичної залежності від часу - одна з найскладніших задач попереднього прогнозного аналізу. Підбір виду функції, яка описує тренд, параметри якої визначаються методом найменших квадратів, проводиться в більшості випадків емпірично, шляхом побудови ряду функцій і порівняння їх між собою за величиною середньоквадратичної помилки, що обчислюється за формулою: $S(\varepsilon) = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}{n-k}}$, де y_i – фактичні значення часового ряду, $\tilde{y}_i$ – розрахункові (згладжені) значення, k – кількість коефіцієнтів факторних ознак включаючи перетин.

Недоліки методу найменших квадратів:

a. Прогноз за аналітичним рівнянням економічного процесу буде точним для нетривалого періоду часу, тому рівняння регресії по надходженню додаткової нової інформації слід перераховувати;
b. Значний обсяг математичних розрахунків з підбору рівняння регресії, яка в цілому нівелюється внаслідок використанням комп'ютерних програм.

Рішення 3. Методом найменших квадратів.

За умовами задачі 1 маємо:

Січень	Лютий	Березень	Квітень	Травень	Червень	Липень	Серпень	Вересень	Жовтень
2,99	2,66	2,63	2,56	2,40	2,22	1,97	1,72	1,56	1,42

- Потрібно побудувати прогноз рівня безробіття в регіоні на листопад, грудень, січень місяці, використовуючи метод найменших квадратів.
- Розрахувати помилку отриманого прогнозу цього методу.

Регресійна статистика за програмним продуктом Excel

Множинний R^2	0,988
R^2 коефіцієнт детермінації	0,977
Нормований R^2	0,974
Стандартна помилка	0,085
Кількість значень спостереження	10

Дисперсійний аналіз

Найменування	Df	**SSR;SSE;SST**		**MS** (середні)		**F(***MSR/MSE***)**	Значущість **F**
Регресія	1	*SSR*	2,432	*MSR*	2,432	334,7538	8,19E-08
Залишок	8	*SSE*	0,058	*MSE*	0,00726		
Загальна	9	*SST*	2,490	*MST*			

	а	**б**	**в**	**г**	**д**	**ж**
Коефіцієнти		*Стандартна помилка*	*t-статистика (а/б)*	*P-значення*	*Нижні 95%*	*Верхні 95%*
Y-перетин	3,160	0,060	54,220	–	3,020	3,290
при ознаці X_1	–0,170	0,010	–18,300	–	–0,190	–0,150

Для визначення прогнозу на листопад–січень місяці проведемо розрахунки рівня безробіття за аналітичним виразом $y = -0{,}17 \times x + 3{,}16$ та визначимо середню відносну та середньоквадратичну помилку, див. табл. 29.

Таблиця 29

Найменування		Рівень (%) безробіття фактичний Y	Умовне позначення часу X	Рівень (%) безробіття розрахований	Помилка відносна $\frac{Уфакт - Урозр}{Уфакт}$	Квадратичне відхилення $(y_i - \tilde{y}_i)^2$
№	***a***	***b***	***c***	***d***	***e***	***f***
1	січень	2,99	1	2,99[94]	0,00%	0,0000[95]
2	лютий	2,66	2	2,82	6,02%	0,0256
3	березень	2,63	3	2,65	0,76%	0,0004
4	квітень	2,56	4	2,48	3,13%	0,0064
5	травень	2,40	5	2,31	3,75%	0,0081
6	червень	2,22	6	2,14	3,60%	0,0064
7	липень	1,97	7	1,97	0,00%	0,0000
8	серпень	1,72	8	1,80	4,65%	0,0064
9	вересень	1,56	9	1,63	4,49%	0,0049
10	жовтень	1,42	10	1,46	2,82%	0,0016
				Разом ∑	**29,22%**	**0,0598**
				Середнє	**2,92%**	**8,15%**
				Прогноз на		
№	***a***	***b***	***c***	***d***	***e***	***f***
	листопад		11	**1,29**[96]		
	грудень		12	1,12		
	січень		13	**0,95**		

Середня відносна помилка $\varepsilon = 29{,}22/10 = 2{,}92\% < 10\%$ – точність прогнозу висока.

Середня квадратична помилка $\sqrt{\frac{\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}{n-k}} = \sqrt{\frac{0{,}0598}{10-1}} = 0{,}0815$,

де k=1, кількість коефіцієнтів при факторних ознаках.

Висновок:

Порівнюючи отримані методами ковзкої середньої, експоненціального згладжування і МНК результати можливо стверджувати, що середня відносна помилка методу експоненціального згладжування потрапляє в межі 20-50%. Це свідчить, що точність прогнозу цього методу є лише задовільною. Результати методів ковзкої середньої та МНК надають високу точність прогнозу, оскільки середня відносна помилка менш 10%. За методом ковзких середніх результати прогнозів більш значущі (на листопад – 1,52%, грудень – 1,53%, січень – 1,49%), оскільки середня відносна помилка за цим методом найменша – 1,13%.

---

[94] y$= -0{,}17 \times x + 3{,}16 = -0{,}17 \times 1 + 3{,}16 = 2{,}99$

[95] $(b_1 - d_1) \,/\, b_1$

[96] $Y = b + ax = 3{,}160 + (-0{,}17 \times 11) = 1{,}29.$

XIX. Виробничі функції

Виникнення теорії виробничих функцій пов’язують з появою статті «Теорія виробництва» економіста Пола Дугласа і математика Чарльза Кобба. У статті описана залежність між обсягом продукції і ресурсами на її виробництво, що побудована на емпіричних статистичних даних обробної промисловості США.

Виробнича функція (або функція випуску) відображає залежність між обсягами витрат за кількістю суттєвих ресурсів, що задіяні у виробництві, наприклад, витрат за працею, виробничими фондами (фондовим капіталом) і обсягом виготовленої продукції при існуючому технічному та організаційному забезпеченні.

Виробничі функції можуть бути **простими**, з лінійною залежністю обсягу продукції виробництва від суттєвих ресурсів (факторних ознак), до **систем рівнянь**, які включають рекурентні співвідношення[97], що описують обсяги випуску продукції в різні періоди за часом.

Найбільш поширені **мультиплікативні форми виробничих функцій**. Їх особливість в тому, що вони обертаються в нуль, якщо один з множників функції дорівнює нулю. Функція обсягу випуску продукції відображає те, що включені в неї за факторними ознаками ресурси приймають участь у процесі виробництва, а відсутність будь-якого фактора робить виробництво неможливим.

Обсяг виробництва продукції може бути змінений в силу зміни обсягу будь-якого з факторів, в той час, як кількість іншого залишається без зміни.

Обсяг виробництва продукції може залишатися без зміни при відповідних співвідношеннях обсягів ресурсів праці та фондового капіталу.

Виробнича функція (субституційна[98]) має загальний вираз:

$Y(x_i) = b\prod_{i=1}^{n} x_i^{a_i}$;

де b – значення коефіцієнту (параметру) функції;

x_i – значення i-го фактору виробничої функції;

a_i – показник степені i-го фактора.

Найчастіше використовують так звану функцію Кобба-Дугласа з аналітичним виразом: $Y = Y(K, L)$; $Y'_{x_i} > 0$; $Y''_{x_i} < 0$ у натуральному чи вартісному обчисленні за суттєвими факторами капіталу та праці, де:

K – ресурс виробничого фондового капіталу за кількістю грошей – фактор «x_1»; чи

L – ресурс витрат праці за кількістю людино-годин(або днів) чи грошей – фактор «x_2».

---

[97] В основі алгоритмічних прийомів накопичення суми і здобутку покладена ідея, що результат обчислень на кожному кроці циклу повинен залежати від результату обчислень на попередньому кроці. Узагальненим математичним виразом цієї ідеї є рекурентні співвідношення. Послідовність векторів задана рекурентним співвідношенням, якщо заданий початковий вектор і функціональна залежність наступного вектора від попереднього.

[98] Функція, що може бути заміщена за обсягами ресурсів (заміщена функція).

На підставі умовно введеної субстиційності (заміщення) ресурсів виробництва, при інших рівних умовах, роблять **два висновки** щодо функціонального їх взаємозв'язку:

1. Збільшення одного з факторів ресурсів виробництва призведе до збільшення обсягу виробництва продукції – тобто перша похідна **додатна**.
2. Гранична[99] ефективність по зростаючому ресурсу із збільшенням величини цього фактора зменшується – тобто друга похідна **від'ємна**.

Рівень організаційної та технічної забезпеченості відображається у відповідних формах взаємодій ресурсів. У розглянутому випадку цей рівень постійний, тобто передбачається відсутність технічного прогресу.

Таким чином, субституційна функція виробництва може бути представлена у вигляді, що відображає взаємозв'язок між кількістю праці і обсягом випуску продукції при незмінному фондовому капіталі, де капітал $k_1 > k_o$, див. рис. 9.

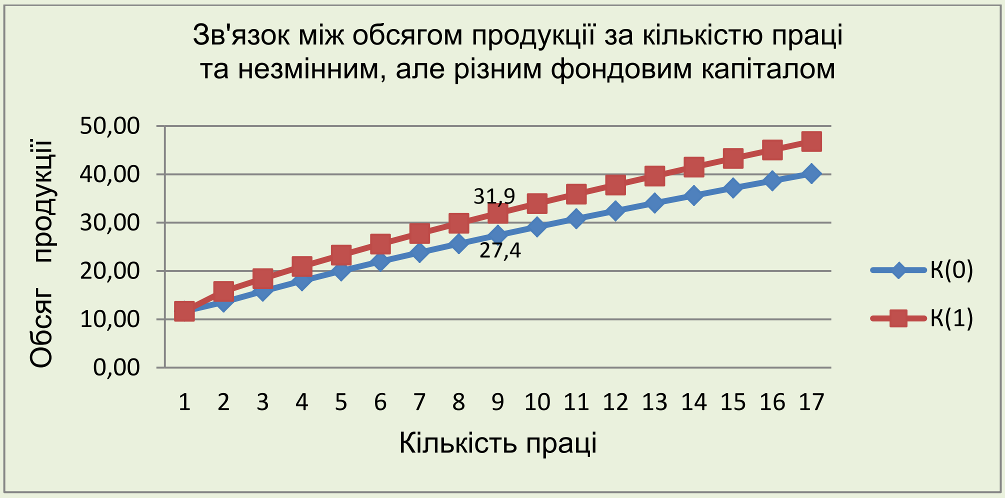


Рис. 9. Зв'язок між обсягом виробництва продукції та працею при незмінному, але різному фондовому капіталі.

Кожне збільшення фондового капіталу за кількістю означає збільшення виробництва та зміщення графіку функції випуску продукції вище попереднього, а також зростання граничної продуктивності праці за певної кількості робочої сили. Збільшення обсягу випуску продукції при збільшенні фактора «фондового капіталу», тобто при $k_1 > k_o$ означає, що крива капіталу k_1 має більший нахил порівняно з кривою капіталу k_o, див. рис. 9.

Із збільшенням фондового капіталу збільшується і середня продуктивності праці, яка є часткою від ділення обсягу продукції на величину витраченої праці. Однак, при цьому зменшується коефіцієнт праці, що означає середню кількість витраченої праці на кожну одиницю випуску і, таким чином, є зворотною величиною середньої продуктивності праці.

Величина фондового капіталу, у наведеній формулі є екзогенною[100] факторною змінною. Технічний прогрес, а також ефект збільшення фондового капіталу (виробничих потужностей) за рахунок інвестицій в цій моделі не враховуються.

[99] гранична ефективність праці (граничний продукт праця – MPL) – це приріст обсягу виробництва від кожної наступної одиниці праці (норма прибутку на додаткову одиницю, капіталу, праці, тощо).

П. Дуглас ще у 1927р. виявив, що відстані на графіках логарифмів значень залежності у часі **обсягів випуску продукції** *(Ln Y)*, **капітальних фондів** *(Ln K)* та **витрат праці** *(Ln L)* мають постійну пропорцію. Функцію, що описує математичну залежність за цією особливістю, запропонував Чарльз Кобб, а її аналітичний вираз: $Y = A \times K^{\alpha} \times L^{\beta}$, де

- *Y* – субституційна функція випуску продукції;
- *A* – технологічний коефіцієнт, що вказує на залежність випуску обсягу виробництва від технології виробництва (припускається строк повної зміни технології через 30-40 років);
- *K, L* – фондовий капітал (у грошах) і витрати праці людей (людино-днів, інше);
- *α, β* – коефіцієнти еластичності обсягу виробництва за витратами капіталу та праці.

При деяких умовах співвідношень еластичності «*α, β*» у функції, еластичність по *α*=0,25 забезпечує зростання обсягу виробництва на 0,25% при зростанні капіталу на 1%.

Функція запропонована Кнутом Уікселлом (Knut Wicksell)[101], та першими, хто використав емпіричні дані для її побудови, були П Дуглас та Ч Кобб. При зростанні значень фондового капіталу «*K*» та праці «*L*» виникає проблема **нескінченного росту обсягів продукції**, через що функція втрачає свою привабливість та економічну обґрунтованість.

Функція $Y = A \times K^{\alpha} \times L^{\beta} \times e^{vt}$ (де v – норма технічного прогресу, а *t*-час), що отримана добутком на коефіцієнт «e^{vt}» знімає проблему нескінченного росту обсягу виробництва продукції та робить функцію економічно привабливою.

Еластичність обсягу випуску продукції по капіталу та праці дорівнює відповідно за «*α*» , як $\frac{dY/dK}{Y/K} = \frac{A(\alpha|K^{\alpha-1}|)L^{\beta}}{AK^{\alpha-1}L^{\beta}} = \alpha$, та «*β*», як $\frac{dY/dL}{Y/L} = \beta$.

Тобто, при збільшенні витрат капіталу на 1% випуск продукції збільшується на «*α%*», а при збільшенні витрат праці на 1% відповідно на «*β%*». Можливо припустити, що обидва значення «*α*» та «*β*» мають бути позитивними, тобто більше **нуля**, оскільки збільшення витрат виробничих факторів має визивати зростання випуску продукції. Водночас вони мають бути менше **одиниці**, так, як розумно припустити, що зменшення ефекту від масштабу виробництва призводить до більш повільного зростання випуску продукції за кожним фактором. Рівень ефективності виробництва **не залежить від його масштабів**.

Масштаби виробництва.

1. Якщо в функції ($Y = AK^{0,6}L^{0,4}$) випуску обсягу продукції коефіцієнти *α+β=1,* то кажуть про **постійний ефект від масштабу виробництва** (обсяг продукції «*Y*» збільшується в тій самій пропорції, що «*K*» та «*L*»).

2. Якщо в функції ($Y = A \times K^{0,7} \times L^{0,5}$) сума *α+β>1,* то кажуть, що функція має **зростаючий ефект від масштабу виробництв**а (це означає, що якщо обсяги «*K*» та «*L*»

---

[100] факторна незалежна ознака зовнішнього походження;
[101] в деяких джерелах Філіп Уікстід (Philip Wicksteed).

збільшуються в деякій пропорції, то обсяг продукції «Y» зростає в більшій пропорції). Середні витрати зменшуються за умов розширення масштабів виробництва.

3. Якщо в функції $\left(Y = A \times K^{0,4} \times L^{0,3}\right)$ сума $\alpha+\beta<1$, то має місце **регресний ефект від масштабу виробництва** (обсяг «*Y*» збільшується в меншій пропорції, ніж «*K*» та «*L*»). Середні витрати, розраховані на одиницю продукції, зростають.

Відповідно з допущенням про конкурентність ринків факторів виробництва коефіцієнтів «*α*» і «*β*» мають подальшу інтерпретацію прогнозованої частки доходу, отриманого відповідно за рахунок капіталу і праці.

Якщо ринок праці має конкурентний характер, то ставка заробітної плати (*w*) дорівнюватиме граничному продукту праці, тобто $w = \frac{dY}{dL} = AK^{\alpha}\beta L^{\beta-1} = \beta\frac{Y}{L}$.

Отже, загальна сума заробітної плати ($w \times L$) дорівнюватиме βY, а частка праці в загальному випуску продукції $\frac{wL}{Y}$ складе постійну величину «*β*». Аналогічним чином норма прибутку виражається як: $\rho = \frac{dY}{dK} = A\alpha K^{\alpha\text{-}1}L^{\beta} = \alpha\frac{Y}{K}$, отже, загальний прибуток ($r \times K$) дорівнюватиме «***αY***», а частка прибутку буде постійною за значенням і дорівнюватиме «*α*».

Існує ряд проблем застосування функції, особливо у тих випадках коли вона застосовується для економіки в цілому. Зокрема, коли між випуском продукції, виробничих засобів і працею у виробничому процесі існує технологічна залежність, то зовсім необов'язково, що така залежність існуватиме коли зазначені фактори комбінуються у масштабах економіки в цілому. По-друге, навіть якщо така залежність для економіки і існує в цілому, то немає ніяких підстав вважати, що вона буде мати просту форму.

Як приклад, при побудові виробничої функції Кобба–Дугласа значення коефіцієнтів технологічного «*A*», еластичності «*α*» та «*β*» можливо розрахувати методом найменших квадратів, для чого:

- функцію Кобба-Дугласа призводять до лінійного її вигляду шляхом логарифмування

  $Ln(Y) = ln(A) + \alpha \times ln(K) + \beta \times ln(L)$.

Мета МНК полягає в мінімізації суми квадратів відхилень між логарифмами спостережуваних величин Y_i, K_i, L_i виробничої функції та розрахунку значень коефіцієнтів $ln(A); \propto; \beta$, де i=1,…*N*; а *N* – загальна кількість значень сукупності.

Збудуємо виробничу функцію ВВП (внутрішнього валового продукту) країни **Мексика**, на вихідних за 1955-1974 роки див. табл. 30, та критеріями ВВП – **результативної ознаки** в грошових одиницях (мільйонах песо) та **факторних ознак**:

– «фондового капіталу» у грошових одиницях (мільйонах песо), у цінах 1960р.
– кількості праці за чисельністю робочих (тис. чол.).

Таблиця 30

№ року	**Результативна** ознака	**Факторні** ознаки	
	ВВП (в цінах 1960г.) (млн. песо)	Капітал (млн. песо)	Робоча сила (тис. чол.)
	Y	***K***	***L***
1955	114043	182113	8310
1956	120410	193749	8529
1957	129187	205192	8738
1958	134705	215130	8952
1959	139960	225021	9171
1960	150511	237026	9569
1961	157897	248897	9527
1962	165286	260661	9662
1963	178491	275466	10334
1964	199457	295378	10981
1965	212323	315715	11746
1966	226977	337642	11521
1967	241194	363599	11540
1968	260881	391847	12066
1969	277498	422382	12297
1970	296530	455049	12955
1971	306712	484677	13338
1972	329030	520553	13738
1973	354057	561531	15924
1974	374977	609825	14154

Після логарифмування первісних даних маємо такі їх логарифми, табл. 31.

Таблиця 31.

	Результативна ознака	**Факторні** ознаки	
Рік	***Ln(Y)***	***Ln(K)***	***Ln(L)***
1955	11,6443	12,1124	9,0252
1956	11,6987	12,1743	9,0512
1957	11,7690	12,2317	9,0754
1958	11,8108	12,2790	9,0996
1959	11,8491	12,3239	9,1238
1960	11,9218	12,3759	9,1663
1961	11,9697	12,4248	9,1619
1962	12,0154	12,4710	9,1760
1963	12,0923	12,5262	9,2432
1964	12,2034	12,5960	9,3039
1965	12,2659	12,6626	9,3713
1966	12,3326	12,7297	9,3519
1967	12,3934	12,8038	9,3536
1968	12,4718	12,8786	9,3981
1969	12,5336	12,9537	9,4171
1970	12,5999	13,0282	9,4692
1971	12,6337	13,0912	9,4984
1972	12,7039	13,1626	9,5279
1973	12,7772	13,2384	9,6756
1974	12,8346	13,3209	9,5578

Регресійна статистика за програмним продуктом Excel

Множинний коефіцієнт *R*	0,99753	
R^2 коефіцієнт детермінації	0,99508	=2,75165 / 2,76525
Нормований R^2	0,99450	=1–(1–0,99508)×19/17
Стандартна помилка[102]	0,02829	$=\sqrt{0{,}0136/(20-3)}$
Кількість значень сукупності	20	

Дисперсійний аналіз

Найменування	*Df*	***SSR;SSE;SST***		***MS* (середні)**		***F***=*MSR/MSE*	Значу-щість ***F***
Регресія	2	*SSR*	2,75165	*MSR*	1,3758	1719	2,41E-20
Залишок	17	*SSE*	0,01360	*MSE*	0,0008		
Загальна	19	*SST*	2,76525	*MST*			

	а	**б**	**в**	**г**	**д**	**ж**
Коефіцієнти при		*Стандартна помилка*	*t-статистика колонка (а/б)*	*Р-значення*	*Нижні 95%*	*Верхні 95%*
Y-перетину	-1,652	0,606	-2,726	0,014	-2,931	-0,373
фактору X_1	0,846	0,093	9,062	0,000	0,649	1,043
фактору X_2	0,340	0,186	1,830	0,085	-0,052	0,732

Ці показники визначаються наступним чином. R^2 – коефіцієнт детермінації характеризує частку варіації залежної змінної, обумовленої регресією або мінливістю змінних, що пояснюють регресію і визначається за $R^2=\frac{SSR}{SST}=\frac{2{,}75165}{2{,}76525}=0{,}9951,$

де: SSR – сума квадратів, що обумовлена регресією;
SST – загальна сума квадратів відхилень фактичних значень змінної від їх середнього.

У нашому випадку коефіцієнт детермінації R^2=0,9951 близький до одиниці, що говорить про репрезентативність моделі. Побудована регресія, добре описує математичну залежність між результативною і факторними ознаками. Нормований коефіцієнт детермінації R^2 враховує кількість коефіцієнтів «*k*» рівняння регресії включаючи перетин:

$\tilde{R}^2=1-(1-R^2)\times\frac{N-1}{N-k-1}=1-(1-0{,}99508)\times\frac{20-1}{20-3}=0{,}99450,$ де: *N* – кількість значень змінної (20 одиниць), а k=2 – кількість факторів рівняння регресії.

Число ступенів вільності (df) для кожної суми квадратів відхилень рівняння регресії, що пояснюється:

- регресією $df=\mathrm{M}-1=3-1=2;$
- помилками (залишками) $df=N-\mathrm{M}=20-3=17;$
- загальною сумою $df=N-1=20-1=19,$

де: *N* – кількість значень ознаки; *M* – кількість коефіцієнтів регресії, включаючи перетин, навіть у випадку якщо його значення нуль.

[102] Стандартна помилка – стандартне відхилення вибіркового розподілу статистики; позначається ***SE***. Характеризує випадкову помилку вибірки; використовується при побудові відповідних довірчих інтервалів і статистичної перевірці гіпотез. Найбільш часто використовується ***SE*** середнього арифметичного. Вона обчислюється як $SE_{\bar{x}}=\frac{s}{\sqrt{n}}$, де ***s*** - стандартне відхилення змінної на підставі незміщеної оцінки її вибіркової дисперсії, а «***n***» – обсяг вибірки. Чим менше стандартне відхилення «***s***» та більше «***n***», тим менше ***SE***. Стандартна помилка середнього арифметичного застосовується при побудові довірчого інтервалу для математичного сподівання, інтервального оцінювання випадкової помилки вибірки, знаходження об'єму репрезентативної вибірки при заданих довірчій ймовірності і гранично допустимій помилці вибірки.

Суми квадратів відхилень визначаються наступним чином:

1. Сума квадратів відхилень результативної ознаки, <u>що зумовлена регресією</u>, розрахована за сумою квадратів відхилень теоретичних значень від її середнього (<u>*SSR*</u>):

$$SSR = \sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2 = 2{,}75165$$

де $\hat{y}_i$ – теоретичні значення змінної «*y*».

2. Сума квадратів відхилень <u>помилок</u>, що характеризує вплив неврахованих факторів (<u>*SSE*</u>), розрахована за сумою квадратів відхилень фактичних від теоретичних значень ознаки:

$$SSE = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 = 0{,}01360.$$

3. Сума квадратів відхилень <u>загальна</u> (<u>*SST*</u>), розрахована за сумою квадратів відхилень фактичних від середнього її значення: $SST = \sum_{i=1}^{N}(y_i - \bar{y})^2 = 2{,}76525$

Середні суми квадратів відхилень (*MS*) являють собою незміщені оцінки дисперсій залежної змінної, обумовлених відповідно регресією і впливом неврахованих випадкових значень помилок:

$$MSR = \frac{SSR}{k-1} = \frac{SSR}{df_R} = \frac{2{,}75165}{3-1} = 1{,}3758; \quad MSE = \frac{SSE}{n-k} = \frac{SSE}{df_e} = \frac{0{,}01360}{20-3} = 0{,}0008$$

F-критерій значущості рівняння регресії визначається за формулою:

$$F = \frac{R^2 \times (N-k)}{(1-R^2)\times(k-1)} = \frac{SSR \times (N-k)}{(k-1)\times SSE} = \frac{2{,}75165\times 17}{2\times 0{,}01360} \approx 1720 \quad \text{або}$$

$$F = \frac{MSR}{MSE} = \frac{1{,}3758}{0{,}0008} \approx 1720.$$

Значення критерію Фішера *F*=1720 більше його табличного значення $F_{0,05;2;17}$=3,59, тобто, рівняння регресії значуще, отже, і досліджувана залежна змінна «*Y*» дуже близько описується включеними в регресійну модель факторними змінними «*K*» та «*L*».

Стандартна помилка – це оцінка стандартного відхилення розподілу коефіцієнта регресії навколо його істинного значення.

«*t*-статистика» Стьюдента – це оцінка коефіцієнта, яка поділена на його стандартну помилку.

На підставі отриманих даних маємо аналітичну залежність за функцією Кобба-Дугласа:

$Ln(Y) = -1{,}652 + 0{,}846 ln(K) + 0{,}340 ln(L)$ або $Y = 0{,}1917 \times K^{0{,}846} L^{0{,}340}$.

Отримана аналітична модель може бути використана для прогнозування майбутніх значень обсягу виробництва на основі відомих або очікуваних ресурсів капіталу і робочої сили.