**ЧАСОВІ РЯДИ: АНАЛІЗ ТА ПРОГНОЗУВАННЯ**

*Мета*: закріплення знань про структуру часового ряду, його аналіз та основні етапи й методи прогнозування процесів, представлених часовими рядами. Придбання навичок проведення аналізу часового ряду та прогнозування.

**Теоретичні знання**: задача прогнозування в Data Mining. Часові ряди, їх види. Аналіз часових рядів. Структура часового ряду. Побудова трендової моделі. Аналіз структури часового ряду, що містить сезонні та циклічні коливання. Адитивна та мультиплікативні моделі часового ряду. Етапи побудови моделі часового ряду методом ковзного середнього. Використання інструментальних засобів MS Excel, MatLab для побудови прогностичної моделі часового ряду.

## І. Стислі теоретичні відомості

**1.1. Прогнозуванння в Data Mining**

Однією з найбільш актуальних, затребуваних та найбільш складних задач ІАД є ***прогнозування*** (*Forecasting*) – оцінка пропущених або майбутніх значення показників на основі аналізу особливостей даних.

Складність прогнозування пов’язана з необхідністю аналізу і оцінювання великих обсягів даних, ускладненням методів, появою нових підходів до прогнозування процесів різної природи.

Тому на сьогодні стан розвитку методів прогнозування тісно пов’язаний із розвитком ІТ.

Для вирішення задачі прогнозування також ***застосовуються методи*** математичної статистики, нейронні мережі, дерева рішень, генетичні алгоритми.

Основою для прогнозування слугує інформація, яка зберігається у вигляді ***часових рядів***.

**1.2. Часові ряди: основні поняття**

**Часовий ряд** – це послідовність спостережуваних значень досліджуваної ознаки, упорядкованих у хронологічному порядку: .

Окремі елементи часових рядів, що є значеннями спостережуваного показника, називаються ***рівнями ряду***,  – кількість рівнів ряду, яку називають ***довжиною часового ряду***.

Часові ряди складаються із двох елементів:

* ***періоду часу***, за який або за станом на який приводяться числові значення;
* *числових значень* показника ***yt*** – ***рівнів ряду***.

Часові ряди, як правило, виникають у результаті виміру деякого показника: показники (характеристики) технічних, природних, соціальних, економічних й інших систем (наприклад, погодні дані).

В залежності від характеру рівнів ряду розрізняють наступні ***види часових рядів***:

1) *моментний ряд* – часовий ряд, значення рівнів якого визначаються на певний момент часу;

2) *інтервальний ряд* – часовий ряд, значення рівнів якого визначаються за певні періоди часу (день, місяць, квартал);

3) *похідний ряд* – часовий ряд, утворений середніми або відносними значеннями показників.

При аналізі часових рядів намагаються визначити основний напрямок розвитку, тенденцію – ***тренд***.

**1.3. Характеристики часового ряду**

***Математичне сподівання*** розраховують як середнє арифметичне рівнів ряду:



***Дисперсія***:

а) зміщена оцінка: 

б) незміщена оцінка: 

У залежності від структури часового ряду значення кожного наступного рівня може залежати від попередніх значень. Кількісно таку залежність вимірюють, розраховуючи *автоковаріацію* та коефіцієнти автокореляції між рівнями аналізованого часового ряду та його рівнями, зсунутими на декілька періодів у часі.

***Часовий лаг*** – це кількість періодів часу між значеннями тих рівнів часового ряду, залежність між якими досліджується.

***Автоковаріація порядку***  – це коваріація між значеннями рівнів часового ряду, зсунутими один відносно одного на  періодів часу:

, (10.1)

де , ,  – часовий лаг.

Для достатньо великих  різницею між  та  можна знехтувати, припустивши, що . Тоді ***автоковаріацію порядку***  визначають за формулою:

а) ***зміщена оцінка***:

 (10.2)

б) ***незміщена оцінка***:

 (10.3)

Автоковаріація порядку  дорівнює дисперсії часового ряду:

 (10.4)

Значення автоковаріації залежить від одиниць вимірювання. Тому для аналізу часових рядів зручніше користуватися безрозмірною величиною – коефіцієнтом автокореляції, який може приймати значення в інтервалі  
[-1, 1].

***Автокореляція рівнів ряду*** – кореляційна залежність між послідовними рівнями часового ряду.

***Коефіцієнт автокореляції***  порядку – характеризує тісноту зв’язку рівнів часового ряду, зсунутих один відносно одного на  періодів часу:

, (10.5)

Для великих  припускаємо, що . Тоді коефіцієнт автокореляції визначають як *відношення автоковаріацій порядку*  та  за формулою:

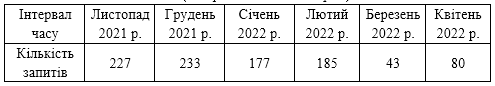
 (10.6)

Чим ближчим до 1 або до -1 буде значення коефіцієнта автокореляції, тим тіснішим буде зв’язок між рівнями часового ряду. Значення коефіцієнта, близькі до нуля, говорять про відсутність зв’язку між рівнями ряду.

***Стаціонарний часовий ряд*** – це часовий ряд, який має сталі й скінчені математичне сподівання і дисперсію для кожного періоду часу та сталу й скінчену автоковаріацію будь-якого порядку в усі моменти часу.

*Таблиця 10.1*

Динаміка зміни пошукових запитів „*купити нетбук*” по Україні за даними Google Trends



***Приклад 1***. Для часового ряду, представленого у таблиці 10.1, розрахувати автоковаріацію нульового і першого порядків та коефіцієнт автокореляції першого порядку.

Автоковаріацію часового ряду порядку  розраховуємо за формулою 10.4:





Автоковаріацію часового ряду порядку  будемо розраховувати за формулою 10.1, оскільки часовий ряд має невелику довжину . Маємо:

,

,





 .

Коефіцієнт автокореляції порядку  розраховуємо за формулою 10.5. Маємо:



,



,





,



**1.4. Аналіз часових рядів**

***Аналіз часових рядів*** (Time Series Analysis) – сукупність методів аналізу, призначених для виявлення структури часових рядів та їх прогнозування.

Виявлення структури часового ряду є необхідним для побудови моделі явища, яке є джерелом аналізованого часового ряду. Побудована модель використовується для визначення майбутніх або невідомих пропущених значень часового ряду.

У структурі часового ряду  можна виділити такі ***складові компоненти***:

* тренд 
* циклічну компоненту 
* сезонну компоненту 
* випадкову компоненту 

Основним структурним компонентом часового ряду є тренд. Він характеризує наявність загального напряму зміни ознаки, що досліджується, протягом тривалого часу. Тому при здійсненні аналізу часових рядів перш за все намагаються визначити тренд як основний, характерний напрям зміни  за деякий проміжок часу.

***Трендом*** (англ. Trend) називають невипадкову функцію , що формується під дією загальних або довгострокових тенденцій, які впливають на часовий ряд.

***Коливання*** – складова часового ряду, яка містить значення, близькі до тих, що повторюються, відносно основної тенденції – тренду. Коливання й відповідні їм компоненти часового ряду бувають сезонними та циклічними.

***Сезонна компонента*** (англ. Seasonal Component) часового ряду містить коливання з невеликим періодом (день, місяць, рік), обумовлені впливом природно-кліматичних умов на досліджувану ознаку. Такі коливання проявляються у добувних галузях, сільському господарстві, споживанні енергоносіїв, води, певних груп товарів тощо.

***Циклічна компонента*** (англ. Cyclical Component) часового ряду містить коливання з достатньо великим періодом (декілька років).Наприклад: часовий ряд інтенсивності сонячної активності, що має повторювані цикли з періодом близько 11 років.

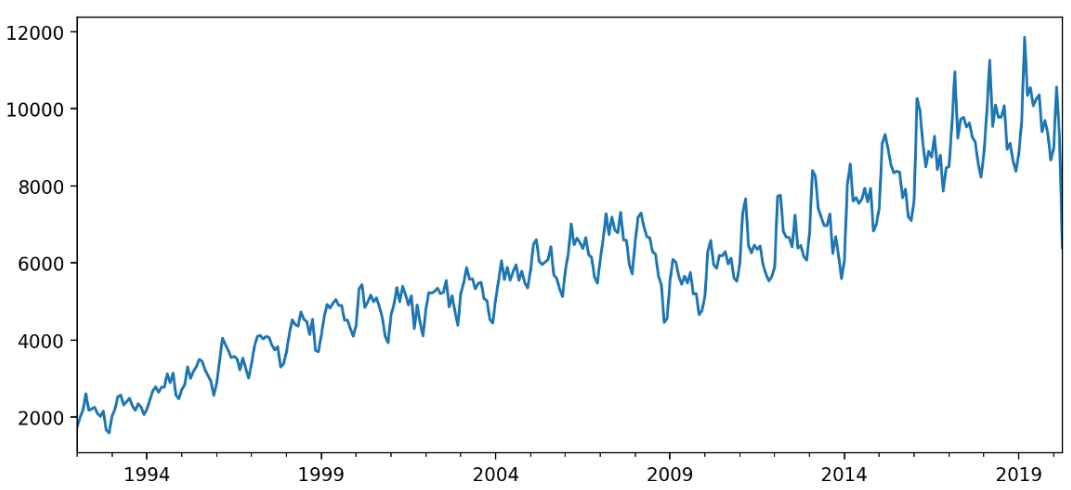
***Випадкова (залишкова) компонента*** (англ. Random/Remainder Component) – складова часового ряду, обумовлена слабкими впливами великого числа випадкових або неврахованих невипадкових факторів.

Успішність аналізу часових рядів залежить від правильного вибору інтервалу між сусідніми рівнями. Вибір надмірно великого чи замалого інтервалу може призвести до втрати інформації про особливості ряду. Наприклад, можна не виявити наявність сезонної чи циклічної компонент часового ряду.

У структурі часового ряду можуть бути представлені усі перераховані вище компоненти, або тільки деякі з них. Проводячи аналіз часового ряду з метою дослідження його структури та природи застосовують спеціальні методи для виявлення наявних компонент та здійснення декомпозиції часового ряду.

***Декомпозиція часового ряду*** (англ. Time Series Decomposition) – виділення його компонент та дослідження кожної компоненти окремо.

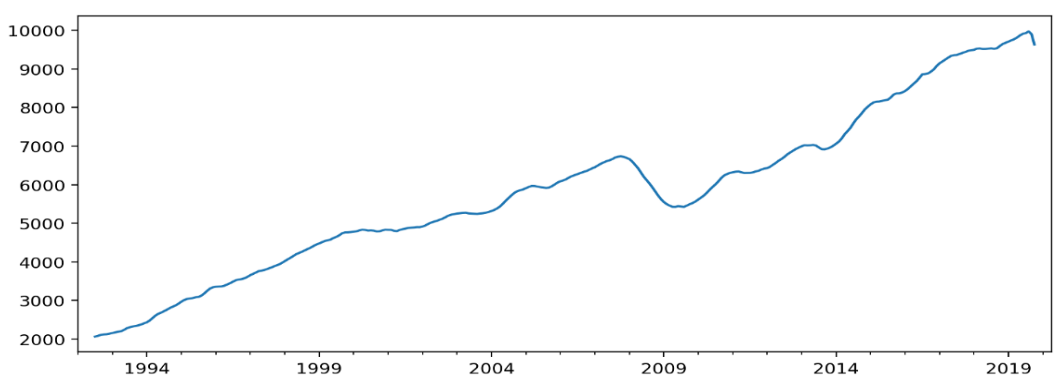
Графічне зображення часового ряду, який має у своєму складі тренд, сезонну та випадкову компоненти, наведено на рисунку 10.1. Після декомпозиції цього часового ряду було виділено окремі компоненти, зображені на рисунках 10.2-10.3.



Млн $

Рис. 10.1. Роздрібний продаж вживаних автомобілів у США

Млн $

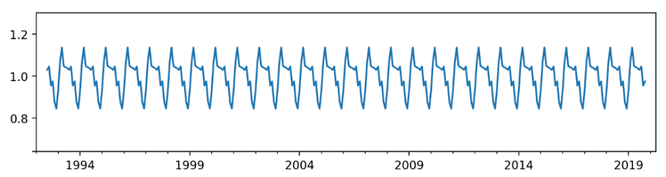


Рік

Рис. 10.2. Виділені у результаті декомпозиції тренд часового ряду

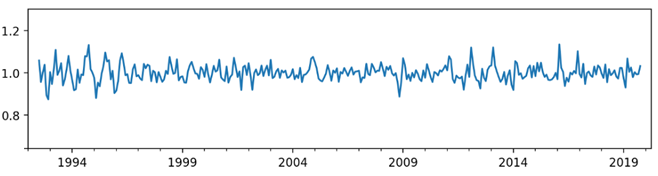
Тренд і циклічні та сезонні коливання є ***регулярними***, або ***систематичними*** ***компонентами***. Вони характеризують основні зміни досліджуваної ознаки з плином часу.

Випадкова компонента – ***випадковий шум*** (англ. Random Noise), є ***нерегулярною компонентою*** часового ряду, яка ускладнює виявлення основних регулярних компонент.



Млн $

б) сезонна компонента

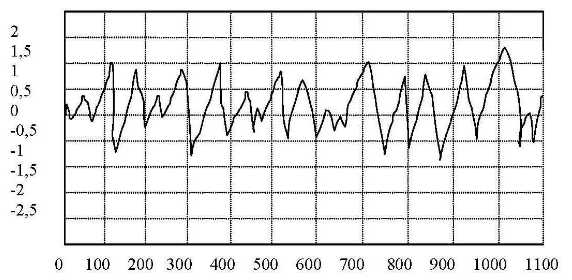


Млн $

в) випадкова компонента

Рис. 10.3. Виділені у результаті декомпозиції компоненти часового ряду

Особливе місце в аналізі даних займають стаціонарні часові ряди, у яких характеристики не змінюються з плином часу. Вони не мають у своєму складі тренду та коливань. Графічно стаціонарний часовий ряд виглядає так:



Значення кожного рівня стаціонарного ряду рівне сумі середнього значення рівнів ряду та випадкової компоненти.

**1.5. Побудова моделі часового ряду, оцінка її точності та адекватності**

Визначення структури й закономірностей часового ряду передбачає виявлення шумів, викидів, тренду, сезонної та циклічної компонент і є основою для побудови ***моделі часового ряду***.

Можна виділити два ***основних напрями моделювання*** часових рядів:

1. Моделювання регулярних компонент у сукупності.

2. Розкладання часового ряду на компоненти і моделювання кожної компоненти окремо.

***Виявлення структури часового ряду*** та типу трендової моделі здійснюють шляхом проведення:

1) ***аналізу графічного зображення часового ряду***: на етапі попереднього аналізу даних на координатну площину наносять точки з координатами  і по характеру їх розташування роблять висновок про наявність тренду і коливань, вид функції тренду, тип та період коливань часового ряду;

2) ***автокореляційного аналізу***: висновок про структуру часового ряду роблять на основі аналізу динаміки змін коефіцієнта автокореляції зі зростанням величини лагу.

Виокремлення виявлених у структурі часового ряду складових називають ***фільтрацією компонент***. Багато методів дослідження часових рядів включають різні способи фільтрації шуму для більш чіткого виділення регулярних складових. При здійсненні фільтрації важливою є адекватна декомпозиція часового ряду.

Залежно від виду зв’язку між складовими часового ряду може бути побудована адитивна, мультиплікативна та змішана моделі часового ряду.

***Адитивна модель*** (англ. Additive model) часового ряду передбачає, що кожен рівень ряду може бути представлений як сума його компонент: .

***Мультиплікативна модель*** (англ. Multiplicative model) часового ряду передбачає, що кожен рівень ряду може бути представлений як добуток його компонент: .

***Змішана модель*** часового ряду може бути представлена як сума та добуток його компонент. Наприклад: .

Для розв’язання задачі прогнозування важливим є ***моделювання тренду*** як основної тенденції часового ряду.

***Побудова трендової моделі*** передбачає правильне виділення усіх компонент часового ряду. Якщо часовий ряд містить сезонні чи циклічні коливання, їх виокремлюють зі структури часового ряду і далі будують трендову модель таким чином, щоб виділена випадкова компонента була близькою до нуля однаково розподіленою незалежною випадковою величиною – процесом типу ***білого шуму***.

***Білий шум*** – це стаціонарний часовий ряд, який має сталу дисперсію, середню, що дорівнює нулю та нульову автокореляцію.

***Моделювання тренду*** – основної тенденціїзміни досліджуваної ознаки, здійснюють відфільтровуючи випадкову компоненту часового ряду та коливання.

З цією метою використовують ***методи згладжування***, які можна розділити на наступні основні групи.

1. ***Механічне вирівнювання*** часового ряду: згладжування рівнів ряду проводиться з використанням значень сусідніх рівнів шляхом переходу до усереднених на деякому інтервалі значень.

2. ***Аналітичне вирівнювання*** часового ряду: реалізується за допомогою регресійних та адаптивних методів і базується на відомому загальному вигляді регулярних компонент часового ряду.

Використання для ***моделювання тренду*** регресійних методів передбачає ***побудову аналітичної функції***, яка характеризує залежність рівнів ряду від часу – тенденції часового ряду й є одним із поширених методів моделювання тренду.

На практиці побудова трендової моделі полягає у ***формуванні набору апроксимуючих функцій*** та оцінюванні їх параметрів і адекватності.

В якості тренду можуть бути обрані лінійна, показникова, параболічна, логарифмічна та інші функції. Для оцінки параметрів рівняння тренду застосовують метод найменших квадратів та інші методи.

Після оцінки параметрів трендових моделей критерієм відбору кращої із них є найбільше значення ***коефіцієнту детермінації***:

 (10.7)

де  – значення рівнів часового ряду, розрахованих за трендовою моделлю;

 – середнє фактичних значень рівнів часового ряду (з вилученими значеннями сезонної чи циклічної компонент у разі їх наявності у структурі ряду).

Коефіцієнт детермінації дозволяє оцінити ***ступінь апроксимації*** трендовою моделлю основної тенденціїзміни досліджуваної ознаки.

Із побудованих моделей обирають ту, яка має найбільше значення коефіцієнта детермінації, враховуючи наступні ***емпіричні правила***:

R2 > 0,95 – висока точність апроксимації

0,8 < R2 < 0,95 – задовільна точність апроксимації

R2 < 0,6 – незадовільна точність апроксимації.

Основою ***перевірки моделі*** ***на адекватність*** є дослідження випадкової компоненти часового ряду. Якщо випадкова компонента є ***білим шумом***, то фільтрація компонент часового ряду була проведена правильно, а побудована модель є адекватною.

Таким чином, ***основні етапи аналізу*** часових рядів із метою побудови моделі та здійснення прогнозування є наступними.

* + - 1. ***Попередній аналіз даних***: виявлення аномальних відхилень, перевірка на наявність тренду.
      2. Графічне зображення та ***попередній аналіз динаміки***, перевірка на стаціонарність, визначення структури часового ряду.
      3. ***Згладжування, фільтрація***, виокремлення компонент часового ряду.
      4. ***Побудова трендових моделей***: формування набору апроксимуючих функцій і чисельне оцінювання параметрів цих моделей.
      5. ***Перевірка адекватності трендових моделей***, оцінка точності апроксимації та вибір кращої моделі.
      6. ***Дослідження випадкової компоненти***, перевірка адекватності побудованої моделі часового ряду.
      7. ***Оцінка точності*** прогнозу, ***прогнозування*** з використанням побудованої моделі часового ряду.

***Оцінка точності прогнозу*** здійснюється з використанням наступних показників.

1. ***MFE*** (англ. Mean Forecast Error)– середня похибка прогнозу є мірою середнього відхилення прогнозних значень від фактичних, показує напрям зміщення прогнозу – заниженими чи завишеними є прогнозні значення:

, (10.8)

де  та  – прогнозоване та фактичне значення рівнів ряду.

2. ***MAE*** або ***MAD*** (англ. Mean Absolute Error/Mean Absolute Derivation)– середнє абсолютне відхилення прогнозованих значень від фактичних:

, (10.9)

3. ***MAPE*** (англ. Mean Absolute PercentageError)– середня абсолютна похибка у відсотках:

 (10.10)

Точність прогнозу є: дуже високою, якщо MAPE менше 10%; високою, якщо MAPE становить 10-20%; задовільною, якщо MAPE становить 20-50%; незадовільною, якщо MAPE більше 50%.

4. ***MPE*** (англ. Mean Percentage Error) – середня відсоткова похибка показує, заниженими чи завищеними є прогнозні значення:

 (10.11)

Якщо MPE < 0, прогноз є переоцінюючим, якщо MPE > 0 – недооцінюючим.

5. ***MSE*** (англ. Mean Squared Error)– середньоквадратична похибка дає загальне уявлення, чи є помилки при прогнозуванні:

 (10.12)

При підборі параметрів моделі та порівнянні декількох моделей обирають ті з них, для яких MSE буде мати менше значення.

Чим меншими є значення наведених показників, тим вищою є точність прогнозу.

**1.6. Виявлення аномальних відхилень**

На етапі підготовки даних до аналізу при здійсненні очищення набору даних для ***виявлення викидів*** – аномальних рівнів часового ряду, застосовують різні методи. Одним із них є метод Ірвіна, суть якого полягає у розрахунку критерію Ірвіна  для всіх рівнів часового ряду та порівняння їх із критичним (табличним) значенням критерію Ірвіна , визначеним на рівні значущості  (зазвичай  або ).

***Критерій Ірвіна*** для часового ряду ряду , розраховують за формулою:

 (10.13)

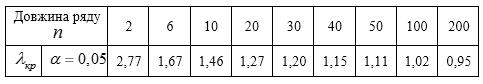
де  – середньоквадратичне відхилення часового ряду.

***Аномальним рівнем*** – значенням часового ряду  довжиною , яке сильно вибивається із загальної тенденції, з ймовірністю  буде вважатися таке, для якого розраховане значення критерію Ірвіна буде більшим за критичне значення, визначене на рівні значущості  для даного : .

Критичні значення критерію Ірвіна наведено у таблиці 10.2. Зазначимо, що з ростом довжини часового ряду критичне значення критерію зменшується. Цю властивість використовують у розрахунках по виявленню аномальних відхилень.

*Таблиця 10.2*

Критичні значення критерію Ірвіна 



Після виявлення аномальних значень виявляють їх природу. Якщо це технічні помилки, що виникають при зборі та передачі даних, агрегуванні значень (помилки першого роду), роблять заміну аномального значення значенням, яке рівне середній арифметичній сусідніх з аномальним рівнів ряду. Якщо аномальні значення відображають вплив об’єктивних факторів, які проявляють себе рідко та епізодично (помилки другого роду), аномальні значення заміні та вилученню не підлягають.

***Приклад 2***. Для часового ряду, представленого у таблиці 10.1, на рівні значущості  виявити наявність викидів із використанням критерію Ірвіна.

Незміщена оцінка дисперсії часового ряду буде рівна:





Тоді незміщена оцінка середньоквадратичного відхилення для даного часового ряду рівна: .

За формулою 10.13 розрахуємо критерій Ірвіна для кожного з рівнів:

,

,

,

,

.

У результаті на рівні значущості  одне із розрахованих значень критерію Ірвіна є більшим за критичне значення  для : . Отже, , тому з ймовірністю 95% ми можемо стверджувати, що на 5-му рівні часового ряду є викид.

**1.7. Перевірка наявності тренду**

Для ***достовірного підтвердження*** наявності чи відсутність тренду у структурі часового ряду використовують спеціальні критерії перевірки гіпотези про наявність тренду. З цією метою застосовують різні методи: метод перевірки різниць середніх, критерій серій та інші.

***1. Метод перевірки різниць середніх рівнів*** зводиться до розбиття часового ряду на дві майже однакові по числу рівнів частини, кожна з яких розглядається як нормально розподілена сукупність значень.

Якщо часовий ряд має тренд, то середні, розраховані для кожної частини ряду, повинні істотно (значимо) розрізнятися між собою.

Якщо ж розбіжність несуттєва, то часовий ряд не має тренду.

Тому перевірка наявності тренду зводиться до перевірки гіпотези про рівність середніх двох нормально розподілених сукупностей, що перевіряється за допомогою t-критерію Стьюдента. Однак цей метод можна застосовувати тільки у тих випадках, коли обидві частини ряду будуть мати однакову дисперсію. Для перевірки гіпотези про рівність дисперсій використовують F-критерій Фішера.

***Послідовність етапів*** методу перевірки різниць середніх рівнів є наступною.

1. Часовий ряд довжиною  розбивається на дві майже однакові частини  та  такі, що .

2. Розраховують середні обох частин:

 та .

3. Розраховують дисперсії обох частин ряду:

 та .

4. Розраховують емпіричне значення F-критерію Фішера за формулою:

. (10.14)

5. Якщо емпіричне значення критерію Фішера  менше за критичне на заданому рівні значущості , взяте зі ступенями свободи  та  ( – індекс тієї частини ряду, яка має більшу дисперсію), то гіпотеза про однорідність дисперсій приймається і необхідно переходити до наступного пункту.

6. Якщо , то гіпотеза про однорідність дисперсій відхиляється, метод не дає відповіді на питання про наявність тренду.

7. Розраховують емпіричне значення t-критерію Стьюдента за формулою:

, (10.15)

де – середньоквадратичне відхилення різниці середніх:

.(10.16)

8. Якщо емпіричне значення критерію Стьюдента менше за табличне  на заданому рівні значущості  із числом ступенів свободи , то гіпотеза про рівність середніх приймається – тренду немає.

9. Якщо , то гіпотеза про рівність середніх відхиляється, тренд є.

Недоліком методу перевірки різниць середніх рівнів є неможливість установити наявність тренду у разі зміни тенденції в середині ряду. Тоді для підтвердження наявності тренду застосовують інші методи.

При здійсненні описаних вище розрахунків для достовірного підтвердження наявності чи відсутності тренду можна скористатися наступними ***функціями MS Excel та MatLab***:

1) середнє значення: *MS Excel*: 

*MatLab*: 

2) незміщена оцінка дисперсії: *MS Excel*: 

*MatLab*: 

3) критичне (табличне) значення F-критерію Фішера:

*MS Excel*: 



*MatLab*: 

4) критичне (табличне) значення t-критерію Стьюдента:

*MS Excel*: 



*MatLab*: 

5) функція MS Excel, яка повертає результат F-тесту: двосторонню ймовірність того, що різниця між дисперсіями рядів в діапазонах  та  є несуттєвою:



6) функція MatLab, яка повертає результат t-тесту: : якщо функція повертає значення, рівне нулю, то гіпотеза про рівність середніх рядів приймається. Якщо функція повертає значення, рівне одиниці, то гіпотеза про рівність середніх рядів відхиляється на рівні значущості .

Функції для розрахунку критичних значень критеріїв Фішера та Стьюдента в MS Excel як аргумент містять рівень значущості , а в MatLab – рівень достовірності .

***2. Критерій серій***.

Ще одним методом, за допомогою якого можна установити наявність тренду у структурі часового ряду, є використання ***критерію серій***.

Послідовність його ***етапів*** є наступною.

1. З часового ряду , утворюють ранжований ряд та визначають його медіану .

2. Для часового ряду  формують послідовність „+” та „-” за правилом:

, (10.19)

де . Якщо значення рівне медіані, це значення пропускають.

3. Підраховують  – число серій в сукупності , де серія – це послідовність плюсів або мінусів, які ідуть підряд (один „+” та „-” також є серією).

4. Підраховують  – довжину самої більшої серії, яка має найбільшу кількість „+” або „-”.

5. Розраховують критичне значення кількості серій  для ряду довжиною на рівні значущості 0,05:

, (10.20)

(квадратні дужки у правій частині нерівності означають цілу частину числа).

6. Розраховують критичне значення довжини найдовшої серії  для ряду довжиною на рівні значущості 0,05:

. (10.21)

7. Перевіряють, чи виконуються нерівності:

. (10.22)

Для підтвердження гіпотези про відсутність у структурі ряду тренду довжина самої більшої серії не повинна бути дуже великою, а кількість серій дуже малою.

8. Якщо обидві нерівності виконуються – тренд відсутній. Якщо хоча б одна з нерівностей не виконується – гіпотеза про відсутність тренду відхиляється: тренд є.

***Приклад 3***. За даними про споживання електроенергії мешканцями регіону за 4 роки виявити структуру часового ряду (табл. 10.3). З цією метою:

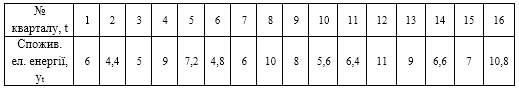
1) здійснити аналіз зміни ознаки за графічним зображення ряду;

2) виявити наявність тренду за методом перевірки середніх різниць;

3) виявити наявність чи відсутність тренду за допомогою критерію серій.

*Таблиця 10.3*

Дані про споживання електроенергії за 4 роки (млн. кВт∙г)

****

1. ***Аналіз характерної зміни ознаки за графічним зображення ряду***.

1.1. В MS Excel формуємо електронну таблицю з вхідними даними часового ряду та будуємо точкову діаграму  (рис. 10.4).

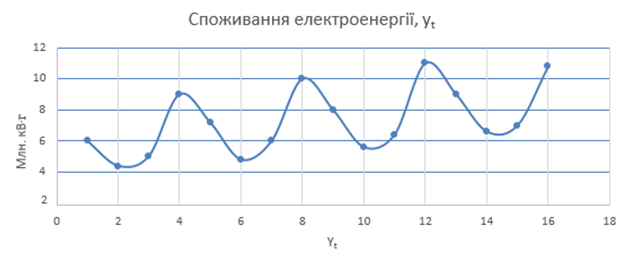


Рис. 10.4. Споживання електроенергії мешканцями регіону (*млн. кВт∙г*)

1.2. Аналіз графічного зображення часового ряду показує, що загальна тенденція зміни значень рівнів ряду є зростаючою. Також ряд містить сезонні коливання з періодичністю 4 квартали та приблизно рівною амплітудою коливань. Це дає підстави припустити, що для моделювання даного часового ряду може бути обрана адитивна модель, яка містить тренд, сезонну та випадкову компоненти: .

2. ***Виявлення наявності тренду з використанням методу перевірки середніх різниць***.

2.1. Установимо наявність тренду з використанням методу перевірки середніх різниць. Для цього часовий ряд в MS Excel розділимо на дві рівні частини  та , роміщені в діапазонах комірокк С7:С14 і С15:С22 та перевіримо ймовірність того, що різниця між дисперсіями двох частин є несуттєвою, ввівши формулу: . У комірці з формулою отримаємо значення 0,997, що свідчить про те, що з ймовірністю 0,997% можемо стверджувати, що дисперсії двох частин часового ряду є однорідними.

2.2. Для перевірки гіпотези про рівність середніх розрахуємо середні та дисперсії двох частин часового ряду за формулами:

для першої частини:  та 

для другої частини:  та 

Для першої частини ряду середнє рівне 6,55, дисперсія – 4,146. Для другої частини ряду середнє рівне 8,05, дисперсія – 4,157.

2.3. Розрахуємо  за формулою 9.16, отримаємо .

2.4. Далі за формулою 10.15 обчислимо емпіричне значення t-критерію Стьюдента, ввівши формулу:  (у комірках D16 і H16 знаходяться дисперсії двох частин ряду, а у комірці F17 – обчислене значення ). Отримаємо емпіричне значення критерію Стьюдента, рівне .

2.6. Для визначення критичного значення критерію Стьюдента на рівні значущості 0,05 введемо формулу: . У комірці з формулою отримаємо значення .

2.7. Оскільки розрахункове значення критерію Стьюдента  є меншим за критичне , визначене на рівні значущості 0,05, гіпотеза про рівність середніх приймається. Тому з ймовірністю 95% можемо стверджувати що у структурі часового ряду тренду немає.

3. ***Виявлення наявності тренду з використанням критерію серій***.

3.1. Установимо наявність тренду у структурі часового ряду, представленого у таблиці 10.5, із використанням критерію серій. Розрахунки будемо проводити в MS Excel. У комірки В5:В20 введемо значення рівнів часового ряду та визначимо медіану, ввівши у комірці G4 формулу: . Розраховане значення медіани буде рівне 6,8 (рис. 10.5).

3.2. У ставпці *Серії* у діапазоні комірок С5:С20 проставляємо: „+” – якщо значення рівня ряду ліворуч більше за медіану, „-”– якщо менше за медіану та пропускаємо, якщо рівне медіані (рис. 10.5).

3.3. За формулами 10.20 та 10.21 розраховуємо критичні значення кількості серій  та довжини найдовшої серії  для ряду довжиною  на рівні значущості 0,05:

1. для розрахунку  у комірку І11 вводимо формулу: , отримуємо ;
2. для розрахунку  у комірку І7 вводимо формулу: , отримуємо ;

(ТОБР() – функція MS Excel, яка бере цілу частину від числа, відкидаючи дробову).

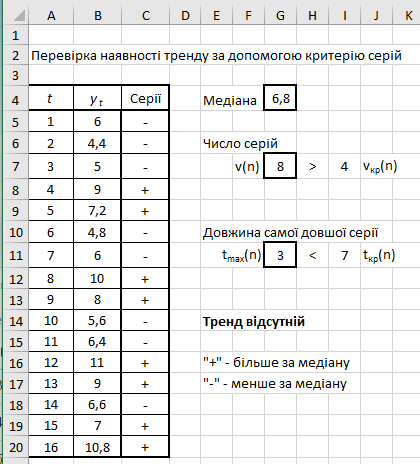


Рис. 10.5. Перевірка наявності тренду за допомогою критерію серій

3.4. Рахуємо у стовпці *Серії* (рис. 10.5) кількість серій  – послідовність плюсів або мінусів, які ідуть підряд і довжину найдовшої серії  – яка має найбільшу кількість „+” або „-”, та порівнюємо їх із критичними. Маємо:

.

Оскільки система нерівностей 10.22 виконується, з ймовірністю 95% можна стверджувати, що тренд у структурі часового ряду відсутній.

Таким чином, аналіз графічного зображення часового ряду, представленого у таблиці 10.3, дав підстави для припущення про наявність у структурі ряду зростаючої тенденції. Однак застосування методу перевірки середніх різниць та критерію серій дозволило установити, що з ймовірністю 95% можна стверджувати про відсутність у структурі ряду тренду.

**1.8. Автокореляційний аналіз часового ряду**

При наявності у структурі часового ряду тренду і циклічних та сезонних коливань обов’язково буде мати місце автокореляція, оскільки значення кожного наступного рівня ряду буде залежати від значень попередніх рівнів.

***Автокореляційний аналіз*** дозволяє виявити структуру часового ряду та визначити, які невипадкові чинники беруть участь у формування його значень.

Для його проведення розраховують коефіцієнти автокореляції з лагом , значення якого послідовно зростає, починаючи від . Розраховані значення дозволяють побудувати графік автокореляційної функції.

***Автокореляційна функція*** часового ряду відображає залежність значень коефіцієнта автокореляції від величини лагу й є послідовністю коефіцієнтів автокореляції першого, другого і т.д. порядків.

***Корелограма*** – графік автокореляційної функції.

Побудова корелограми полегшує проведення аналізу, оскільки вона дозволяє візуально відобразити динаміку зміни коефіцієнта автокореляції зі зростанням величини лагу.

При здійсненні ***аналізу значень коефіцієнта автокореляції*** необхідно враховувати наступне.

1. Коефіцієнт автокореляції дозволяє характеризувати тісноту лійнійного зв’язку між рівнями часового ряду.

2. Близькість значень коефіцієнта автокореляції до 1 або до -1 говорить про тісний лінійний зв’язок між рівнями часового ряду.

3. Для деяких часових рядів із сильною нелінійною тенденцією коефіцієнт автокореляції може бути близьким до нуля.

4. Знак коефіцієнта автокореляції не дозволяє робити висновок про напрям зміни тенденції. При додатному чи від’ємному значенні коефіцієнта автокореляції часовий ряд може мати як спадну так і зростаючу тенденцію.

***Аналіз корелограми*** дозволяє зробити висновок про структуру часового ряду, виходячи із наступних положень.

1. Для ***стаціонарного часового ряду*** характерним є чергування затухаючих додатніх та від’ємних статистично незначущих значень коефіцієнтів автокореляції.

2. Якщо найбільшим по модулю виявилося значуще значення коефіцієнта автокореляції 1-го порядку, досліджуваний ряд ***містить тільки тренд***.

3. Якщо корелограма містить багато максимальних і мінімальних значень коефіцієнтів автокореляцій, а найбільшим по модулю виявилося значуще значення коефіцієнта автокореляції -го порядку, досліджуваний ряд містить ***коливання з періодом .***

4. Якщо жоден із коефіцієнтів автокореляції не є значущим, то можна зробити припущення про те, що ряд не містить тренду та коливань або ряд має сильний нелінійний тренд і потребує додаткового аналізу.

За наявності тренду *коефіцієнти* автокореляції є *значущими*. Для ***перевірки на значущість*** коефіцієнта автокореляції у разі відсутності тренду можна скористатися ***критерієм стандартної похибки***. Якщо значення коефіцієнту автокореляції  порядку  не виходить на межі інтервалу:

 (10.23)

де  – довжина часового ряду, то коефіцієнт автокореляції  не буде значимим і автокореляція порядку  відсутня.

А якщо значення  виходить за межі цього інтервалу, то з ймовірністю 95% можемо стверджувати, що коефіцієнт автокореляції  є значимим.

Якщо усі коефіцієнти автокореляції часового ряду не є значимими – не виходять за межі довірчого інтервалу, заданого нерівністю 9.23, ***часовий ряд є стаціонарним***.

***Автокореляційний аналіз випадкової компоненти*** дозволяє оцінити правильність побудованої моделі часового ряду.

1. Якщо при побудові моделі часового ряду тренд і сезонну компоненту було виокремлено та описано правильно, корелограма випадкової компоненти не виходить за межі довірчого інтервалу й згасає при , де – порядок автокореляції (лаг).

2. Якщо при при побудові моделі часового ряду тренд було визначено та виокремлено не правильно, автокореляційна функція випадкової компоненти не згасає при .

3. Якщо при при побудові моделі часового ряду не правильно визначено та виокремлено сезонну компоненту, автокореляційна функція випадкової компоненти зростає скачками кратно періоду коливань.

Для здійснення ***розрахунку*** ***коефіцієнтів автокореляції*** можна скористатися наступними формулами MS Excel та MatLab.

1. В MS Excel:

, де *масив1* – діапазон комірок з рівнями аналізованого часового ряду, *масив2* - діапазон комірок з рівнями часового ряду, зсунутими відносно заданого на  періодів.

2. В MatLab: , де  - вектор зі значеннями рівнів часового ряду.

Після знаходження коефіцієнтів автокореляції здійснюють побудову корелограми та проводять її аналіз для виявлення структури часового ряду.

**1.9. Побудова адитивної моделі часового ряду**

Відомо декілька підходів до ***аналізу структури часових рядів***, що містять сезонні або циклічні коливання.

Найпростіший підхід – розрахунок значень сезонної компоненти методом ковзної середньої та побудова адитивної  або мультиплікативної  моделі часового ряду (де  – тренд,  – сезонна компонента,  – випадкова компонента). Етапи побудови адитивної моделі можуть бути застосовані також і для побудови мультиплікативної моделі, оскільки вона зводиться до адитивної моделі логарифмуванням

Вибір однієї з двох моделей відбувається на основі ***аналізу структури сезонних коливань***:

* якщо амплітуда коливань приблизно постійна, будують адитивну модель часового ряду, в якій значення сезонної компоненти передбачаються постійними для різних циклів (рис. 10.6);
* якщо амплітуда сезонних коливань зростає або зменшується, будуть мультиплікативну модель часового ряду, яка ставить рівні ряду в залежність від значень сезонної компоненти (рис. 10.7).

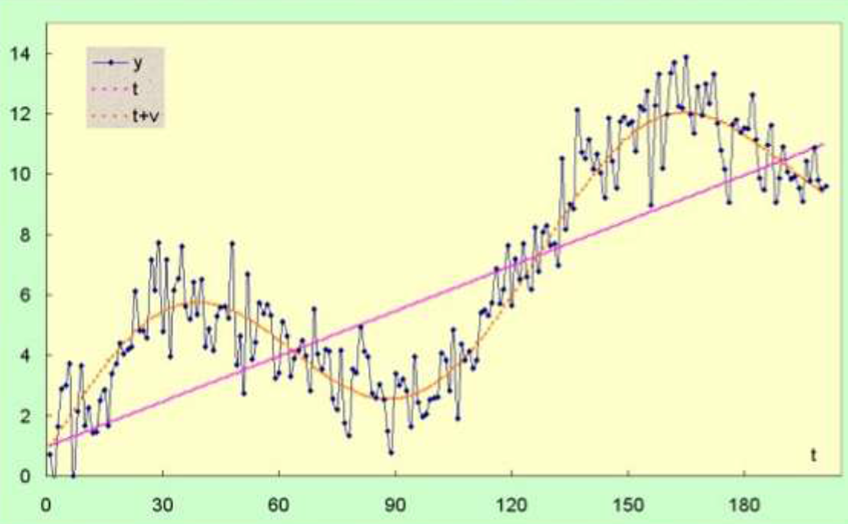


Рис. 10.6. Адитивна модель часового ряду

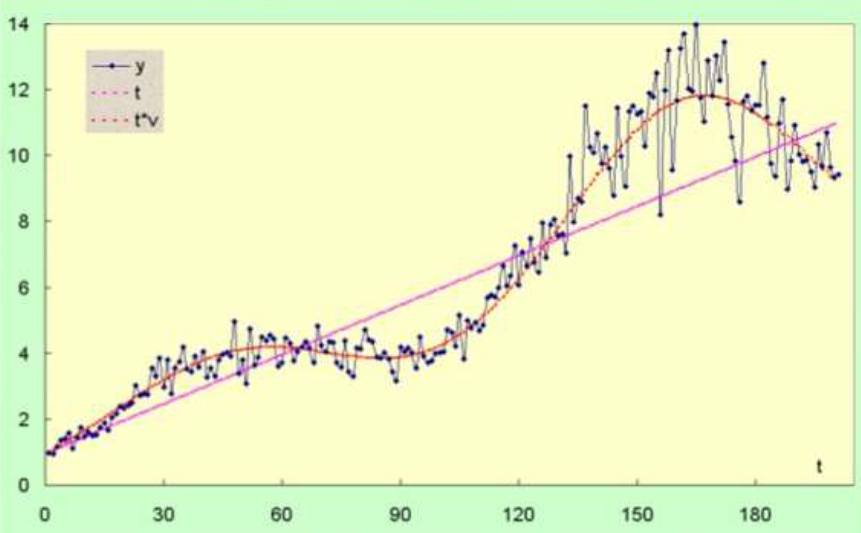


Рис. 10.7. Мультиплікативна модель часового ряду

***Побудова моделей*** (адитивної і мультиплікативної) зводиться до розрахунку значень Т, S та Е для кожного рівня ряду. Процес побудови моделей включає ***наступні етапи***.

1. **Попередній аналіз даних**, визначення структури часового ряду.

2. **Вирівнювання вхідного ряду методом ковзного середнього**. Для цього необхідно:

1) просумувати рівні ряду послідовно за кожні ***і*** періодів із зсувом на один момент часу і визначити умовні значення – ***укрупнити дані***;

2) розділити отримані суми на число, яке дорівнює кількості періодів, що укрупнювалися;

3) привести ці значення відповідно з фактичними моментами часу: знайти середні значення з двох послідовних ковзних середніх – ***центровані ковзні середні***.

3. **Розрахунок значень сезонної компоненти *S****.* Оцінку сезонної компоненти знаходимо як різницю між фактичними рівнями ряду і центрованими ковзними середніми.

Після розрахунку оцінки сезонної компоненти для рівнів ряду, необхідно знайти середнє значення цієї компоненти для однойменних періодів-сезонів (наприклад, кварталів, місяців).

Скореговане значення сезонної компоненти знаходять як різницю між її середньою оцінкою і коригуючим коефіцієнтом. Коригуючий коефіцієнт розраховують як середнє значення середніх сезонної компоненти для однойменних періодів-сезонів.

4. **Усунення сезонної компоненти** з вихідних рівнів ряду і **отримання вирівняних даних**. Ці значення розраховують для кожного моменту часу і містять тільки тенденцію (тренд) і випадкову компоненту: .. Для усунення сезонної компоненти з рівнів ряду необхідно вирахувати її значення з кожного рівня вхідного ряду за формулою .

5. ***Побудова трендових моделей***: формування набору апроксимуючих функцій, чисельне оцінювання їх параметрів, оцінка точності апроксимації, вибір кращої моделі.

6. ***Аналітичне вирівнювання рівнів***, розрахунок значень  з використанням обраної трендової моделі.

7. ***Оцінка значень випадкової компоненти***  шляхом розрахунку значень для кожного рівня ряду за формулою: .

8. ***Перевірка адекватності побудованої моделі***, ***оцінка точності прогнозу***.

9. ***Прогнозування*** з використанням побудованої моделі часового ряду. Прогнозне значення рівнів часового ряду є сумою трендової та сезонної компонент.

**1.10. Створення прогнозу в Excel**

***Інструментальні засоби прогнозування*** в MS Excel:

1. Есктраполяція, виконана ***побудовою лінії тренда***. Для цього необхідно:
   1. побудувати діаграму (краще точкову) по даним часового ряду;
   2. визвати контекстне меню побудованої діаграми та обрати пункт ***додати лінію тренду***ю.
2. Функція ***ЛИНЕЙН/LINEST (****y;x****)***: дозволяє розрахувати коефіцієнти лінійного наближення y = ax + b. За отриманим рівнянням можна здійснювати точковий прогноз даних на певний період.
3. Функція ***ПРЕДСКАЗ/FORECAST(****new\_x;y;x****)***: дозволяє здійснювати точковий прогноз даних на певний період на основі лінійної залежності.
4. Функція ***ТЕНДЕНЦИЯ/TREND(****y;x;new\_x****)***: дозволяє отримувати прогноз для декількох наступних періодів на основі лінійної залежності.
5. Функція ***РОСТ/GROWTH(****y;x;new\_x****)***: дозволяє отримувати прогноз для декількох наступних періодів на основі експоненційної залежності.

***Аргументи функцій***: ***y*** - дані часового ряду, ***x*** - періоди часу, ***new\_x*** - наступний період (періоди) часу, для якого (яких) здійснюється прогноз.

***Спосіб введення функцій массиву***: описані вище функції є функціями масиву, тому для їх правильного введення необхідно:

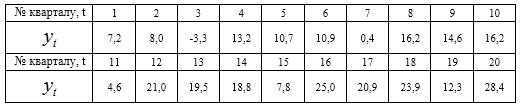
1. ввести у комірку потрібну функцію з аргументами та натиснути ***Enter***;
2. виділити комірку з формулою та необхідну кількість суміжних комірок, натиснути клавішу F2 а потім одночасно клавіши ***Shift*** + ***Ctrl*** + ***Enter***.
3. ***Лист прогнозу***: дозволяє миттєво створювати прогноз по даним часового ряду на основі заданих параметрів: довірчого інтервалу, початку та кінця прогнозного періоду, врахування сезонності (вручну та автоматично), способів заповнення відсутніх точок та обробки дублікатів.

**ІІ. Виконання лабораторної роботи**

**Завдання 1**. Здійснити побудову моделі часового ряду за даними щоквартальних значень змінної  за 5 років, представленими у таблиці 10.4 та здійснити прогнозування значень з горизонтом прогнозу – 1 рік.

*Таблиця 10.4*.

Дані щоквартальних значень змінної  за 5 років



**2.1. Попередній аналіз даних, визначення структури часового ряду**

Для виявлення структури часового ряду застосуємо ***автокореляційний аналіз***, для чого необхідно дослідити динаміку зміни коефіцієнта автокореляції та автокореляційну функцію – корелограму і виявитити значущість коефіцієнтів автокореляції.

1. В MatLab сворюємо скрипт, який містить команди для побудови графіку часового ряду, розрахунку коефіцієнтів автокореляції для кожного лагу від 0 до 19 та побудови корелограми – функції, яка містить послідовність коефіцієнтів автокореляції:

% y - вектор з рівнями часового ряду

y=[ 7.2 8.0 -3.3 13.2 10.7 10.9 0.4 16.2 14.6 16.2 4.6 21.0 19.5 18.8 7.8 …

25.0 20.9 23.9 12.3 28.4];

[acf, lag]=autocorr(y); % розрахунок коефіцієнтів автокореляції acf

acf=acf'; lag=lag';

[acf lag] % виведення коефіцієнтів автокореляції, lag – лаг

subplot(2,1,1)

plot(y) % графік часового ряду

subplot(2,1,2)

autocorr(y)% побудова корелограми

У результаті виконання даного скрипта у вікні Command Window буде виведено значення коефіцієнтів автокореляції й у окремому вікні – графіки часового ряду та корелограми. Графік часового ряду зображено у верхній частині рисунку, а графік корелограми – у нижній (рис. 10.8).

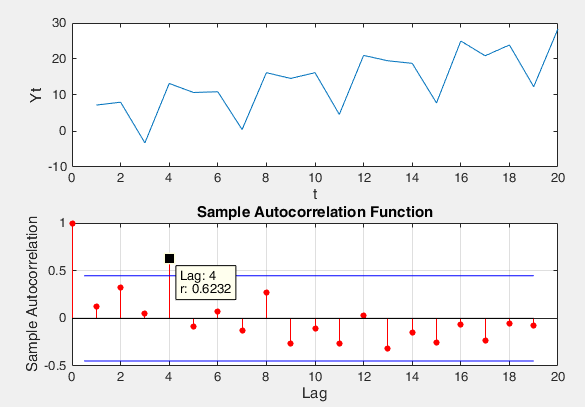


Рис. 10.8. Графік та корелограма часового ряду

2. Аналіз графічного зображення часового ряду показує загальну зростаючу тенденцію зміни значень рівнів ряду та наявність коливань із періодичністю 4 квартали та приблизно рівною амплітудою коливань. Це дає підстави припустити, що часовий ряд може бути описаний за допомогою адитивної моделі, яка містить тренд, сезонну та випадкову компоненти.

3. Аналіз графіку автокореляційної функції показує, що корелограма містить максимальні та мінімальні значення коефіцієнтів автокореляції, які відповідають максимумам та мінімумам часового ряду (рис. 10.7). Найбільшим по модулю виявився коефіцієнт автокореляції 4-го порядку.

4. Аналіз значень коефіцієнтів автокореляції та їх значущості показав, що усі коефіцієнти автокореляції, крім коефіцієнта 4-го порядку, не є значимими. Усі коефіцієнти, крім коефіцієнта 4-го порядку, не виходять за межі довірчого інтервалу , розрахованого за формулою 10.24 для ряду довжиною , який на графіку корелограми находиться між двома синіми лініями. Це дає підстави стверджувати, що досліджуваний часовий ряд містить сезонні коливання з періодом 4 квартали.

5. Для достовірного підтвердження наявності тренду застосуємо ***метод перевірки середніх різниць***. Розрахунки будемо проводити в MatLab. Для розв’язання поставленої задачі створюємо скрипт:

% часовий ряд задаємо як вектор y

y=[7.2 8.0 -3.3 13.2 10.7 10.9 0.4 16.2 14.6 16.2 4.6 21.0 19.5 18.8 7.8 …

25.0 20.9 23.9 12.3 28.4];

n=length(y); % визначаємо довжину часового ряду

% розбиття часового ряду на дві майже рівні частини

n1=floor(n/2); n2=n-n1;

y1=y(1:n1); y2=y((n1+1):n);

% розрахунок середніх значень, дисперсій кожної частини

mean1=mean(y1);

mean2=mean(y2);

var1=var(y1);

var2=var(y2);

% обчислення емпіричного значення критерію Фішера

if var1>var2

Fr=var1/var2

else

Fr=var2/var1

end

% визначення критичного значення критерію Фішера

Fcr=finv(0.95,n1,n2) % на рівні значущості 0,05

% перевірка гіпотези про рівність дисперсій

if Fr>=Fcr

disp('метод не може визначити наявність тренду');

else

% обчислення розрахункового значення критерію Стьюдента

sr=sqrt(((n1-1)\*var1+(n2-1)\*var2)/(n1+n2-2));

t\_r=abs(mean1-mean2)/(sr\*sqrt(1/n1+1/n2))

% визначення критичного значення критерію Стьюдента

t\_cr=tinv(0.95,n1+n2-2) % на рівні значущості 0,05

% перевірка гіпотези про рівність середніх

if t\_r<t\_cr

disp('тренд відсутній')

else

disp('тренд є')

end

end

% ймовірність того, різниця середніх двох частин є несуттєвою

virt\_cr=ttest2(y1,y2, 0.05)

Результатом виконання даного скрипта буде підтвердження з ймовірністю 95% наявності або відсутності тренду в структурі часового ряду або повідомлення про те, що даний метод не дає можливості установити наявність тренду.

6. Здійснимо аналіз отриманих результатів (рис. 10.9). Емпіричне значення критерію Фішера  на рівні значущості  виявилося меншим за критичне , тому гіпотеза про однорідність дисперсій приймається, можна переходити до розрахунку критерію Стьюдента.

7. Емпіричне значення критерію Стьюдента  на рівні значущості  виявилося більшим за критчне  (рис. 10.9), тому гіпотеза про про рівність середніх відхиляється. З ймовірністю 95% можемо стверджувати, що у структурі аналізованого часового ряду тренд є.

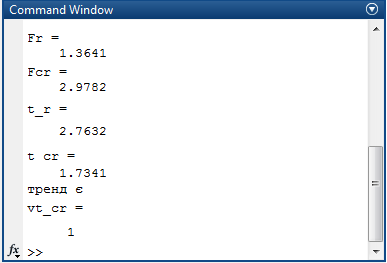


Рис. 10.9. Результат визначення наявності тренду методом перевірки різниць середніх рівнів

8. Про наявність тренду свідчить також результат t-тесту (рис. 10.9): на рівні значущості  гіпотезу про рівність середніх двох частин часового ряду було відхилено. Отже, з ймовірністю 95% було достовірно установлено, що аналізований часовий ряд має тренд.

9. Таким чином, проведений аналіз дозволив установити, що у структурі часового ряду є тренд, сезонна компонента з періодом 4 квартали і приблизно рівною амплітудою коливань та випадкова компонента. Тому для моделювання часового ряду доцільно обрати адитивну модель: .

**2.2. Вирівнювання рівнів ряду методом ковзної середньої**.

Розглянемо послідовність розрахунків для вирівнювання вихідних рівнів ряду ***методом ковзної середньої*** (рис. 10.10).

1. Розрахуємо суми послідовних членів ряду:

* у комірку D8 введемо формулу для обчислення суми перших 4-х членів ряду: =C7+C8+C9+C10;
* у комірку D9 – формулу для обчислення суми 2-го, 3-го, 4-го та 5-го членів ряду: =C8+C9+C10+C11;
* у наступних комірках стовпця D електронної таблиці вводимо формули для обчислення суми 3-го, 4-го, 5-го та 6-го членів ряду і т.д. до комірки D24, в яку необхідно ввести формулу для обчислення суми 17-го, 18-го, 19-го та 20-го членів ряду: =C23+C24+C25+C26.

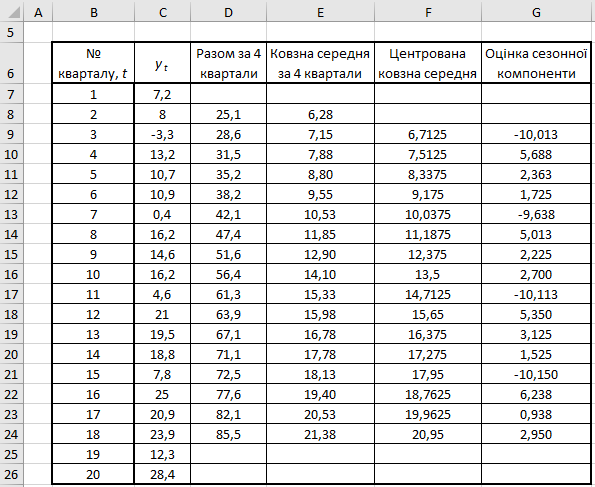


Рис. 10.10. Вирівнювання вхідних рівнів ряду методом ковзної середньої

2. У комірках стовпця Е електронної таблиці розраховуємо ковзну середню за 4 квартали:

* у комірку Е8 введемо формулу: =D8/4;
* у комірку Е9 – формулу: =D9/4;
* далі – аналогічно до комірки Е24, в яку необхідно ввести формулу: =D24/4.

3. У комірках стовпця F електронної таблиці розраховуємо центровану ковзну середню:

* у комірку F9 введемо формулу: =(E8+E9)/2;
* у комірку F10 введемо формулу: =(E9+E10)/2;
* далі – аналогічно до комірки F24, в яку необхідно ввести формулу: =(E23+E24)/2.

4. У комірках стовпця G електронної таблиці розраховуємо оцінку сезонної компоненти як різниці між значенням рівня ряду та значення відповідної рівню центрованої ковзної середньої:

* у комірку G9 введемо формулу: =C9-F9;
* далі – аналогічно до комірки G24, в яку необхідно ввести формулу: =C24-F24.

**2.3. Розрахунок середніх оцінок сезонної компоненти по кварталам**

1. Сформуємо електронну таблицю, яка буде містити значення оцінок сезонної компоненти по кварталам та рокам (рис. 10.11).

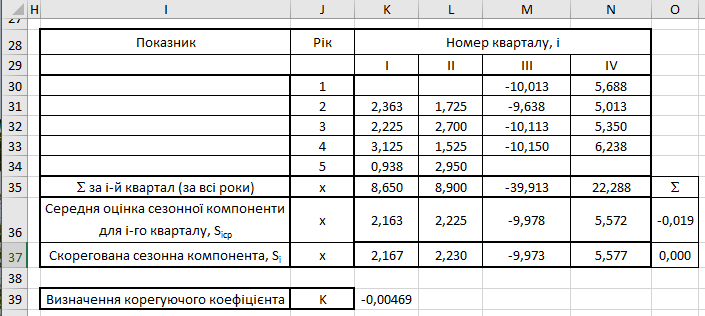


Рис. 10.11. Розрахунок середніх оцінок сезонної компоненти

2. Розраховуємо суму оцінок за і-й квартал за всі роки:

* за 1-й квартал – у комірку К35 введемо: =СУММ(K30:K34);
* за 2-й квартал – у комірку L35 введемо: =СУММ(L30:L34);
* за 3-й квартал – у комірку M35 введемо: =СУММ(M30:M34);
* за 4-й квартал – у комірку N35 введемо: =СУММ(N30:N34).

3. Розраховуємо середню оцінку сезонної компоненти за квартали:

* за 1-й квартал – у комірку К36 введемо: =K35/4;
* за 2-й квартал – у комірку L36 введемо: =L35/4;
* за 3-й квартал – у комірку M36 введемо: =M35/4;
* за 4-й квартал – у комірку N36 введемо: =N35/4.

4. Розрахуємо коригуючий коефіцієнт k. Для даної моделі маємо:

*2,163 + 2,225 – 9,978 + 5,572 = -0,019*

Тоді k = -*0,019/4* = -*0,00469*.

5. Розрахуємо скориговані значення сезонної компоненти як різницю між її середньою оцінкою і коригуючим коефіцієнтом .

6. Перевіримо рівність нулеві суми значень сезонної компоненти:

*2,167 + 2,230 – 9,973 + 5,577 = 0* та занесемо їх у таблицю для відповідних кварталів.

**2.4. Усунення сезонної компоненти з вихідних рівнів часового ряду**

1. Сформуємо електронну таблицю, яка буде містити розраховані значення компонент моделі (рис. 10.12).

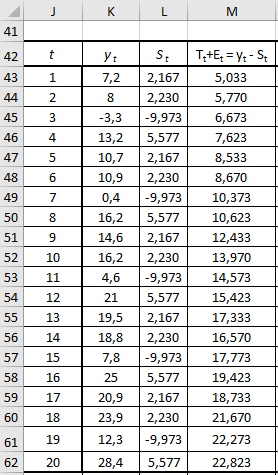


Рис. 10.12. Усунення сезонної компоненти з рівнівчасового ряду

2. Для усунення сезонної компоненти необхідно вирахувати її значення з кожного рівня вихідного ряду  (комірки М43:М62).

Ці значення розраховуються для кожного періоду часу і містять тільки тенденцію та випадкову компоненту.

**2.5. Побудова трендових моделей**

1. Для визначення типу моделі тренду побудуємо точкову діаграму для вирівняних даних  (рис. 10.13).

2.Визиваємо контекстне меню діаграми та обираємо пункт *Додати лінію тренду*.

3. У вікні *Формат лінії тренду*, яке з’явиться праворуч (рис. 10.14), налаштовуємо параметри та формат трендової моделі: тип ліній тренду – обираємо *Лінійна*, прогноз вказуємо на 4 періоди вперед, ставимо прапорці навпроти опцій – показувати рівняння на діаграмі та розмістити коефіцієнт детермінації .

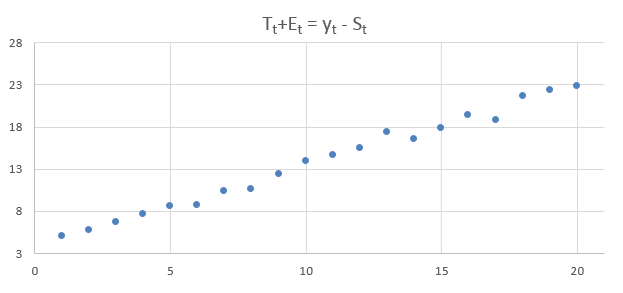


Рис. 10.13. Точкова діаграма вирівняних даних 

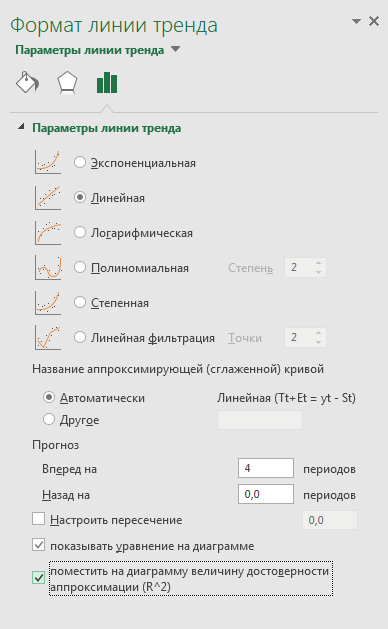


Рис. 10.14. Вікно Формат лінії тренду

4. На діаграмі з’явиться лінійна апроксимуюча функція з прогнозом на 4 періоди вперед, її рівняння та коефіцієнт детермінації, який показує точність апроксимації функцією вирівняних даних (рис. 10.15).

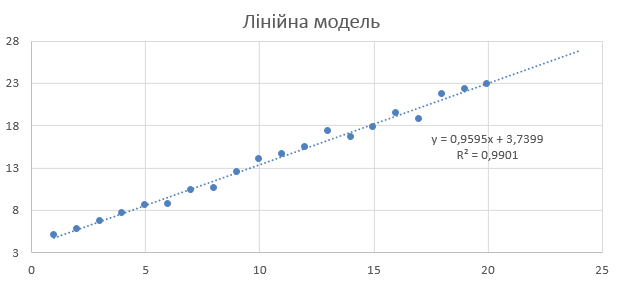


Рис. 10.15. Лінійна трендова модель

5. Для формування набору апроксимуючих функцій, чисельного оцінювання їх параметрів та оцінки точності апроксимації діаграму вирівняних даних (рис. 10.13) копіюємо та виконуємо декілька раз пункти 2, 3 і 4, обираючи для апроксимації по черзі експоненціальну, логарифмічну, степеневу та поліноміальну функції. На кожній із діаграм з’явиться апроксимуюча функція з прогнозом на 4 періоди вперед, рівнянням та коефіцієнтом детермінації (рис. 10.16-рис. 10.19).

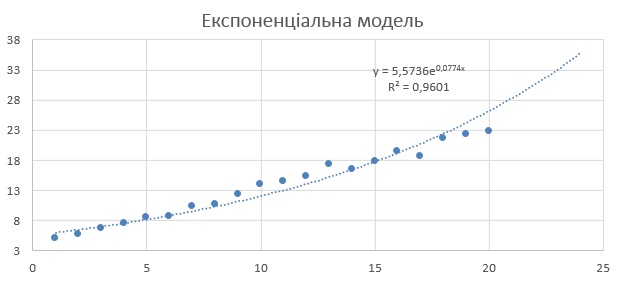


Рис. 10.16. Експоненціальна трендова модель

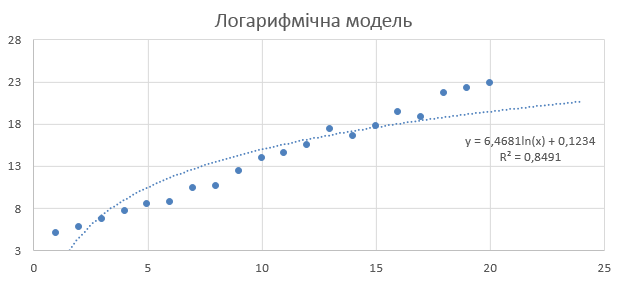


Рис. 10.17. Логарифчічна трендова модель

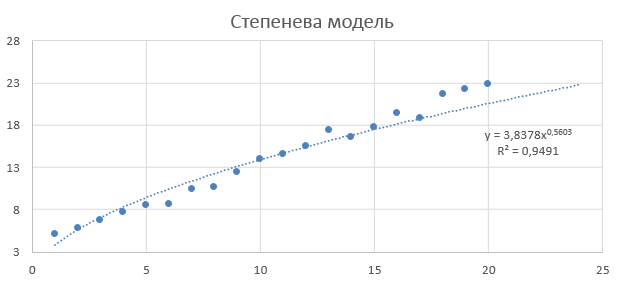


Рис. 10.18. Степенева трендова модель

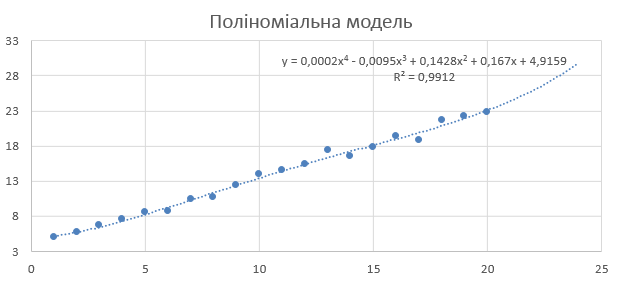


Рис. 10.19. Поліноміальна трендова модель

6. Здійснюємо порівняння коефіцієнтів детермінації та робимо висновок, що кращими моделями є лінійна та поліноміальна, оскільки саме для них рівень апроксимації вирівняного ряду є найвищим. Для лінійної моделі коефіцієнт детермінації , для поліноміальної моделі  Різниця між коефіцієнтами є незначною, тому для подальшої побудови моделі часового ряду обираємо лінійну трендову модель, задану рівнянням:

.

**2.6. Аналітичне вирівнювання рівнів ряду, оцінка значень випадкової компоненти**

1. Визначимо компоненту  даної моделі. Для цього проведемо аналітичне вирівнювання ряду  за допомогою лінійного тренду 

2. У комірки V64 і W64 введемо коефіцієнти лінійної функції  і  (рис. 10.20). Продовжимо формування таблиці для розрахунку компонент часового ряду (рис. 10.21), доповнивши таблицю на рисунку 10.12 стовпцем розрахунку трендової компоненти для кожного рівня ряду (комірки N43:N62).

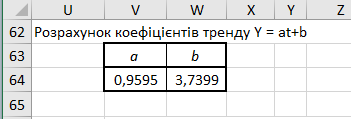


Рис. 10.20. Коефіцієнти лінійного рівняння тренду

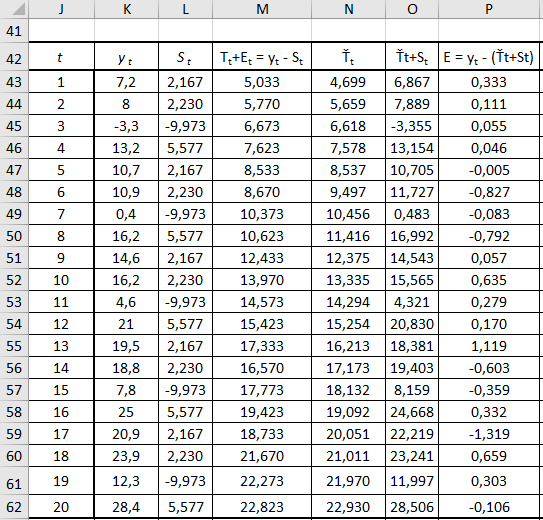


Рис. 10.21. Розрахунок компонент адитивної моделі часового ряду

4. Знайдемо значення рівнів ряду, отримані за адитивною моделлю. Для цього додаємо до рівнів  значення сезонної компоненти для відповідних кварталів  (комірки О43:О62) та оцінюємо значення випадкової компоненти – розраховуємо абсолютну похибку за формулою  (комірки Р43:Р62).

5. Проведена декомпозиція часового ряду (виділення компонент) дозволяє побудувати діаграми часового ряду та його компонент (рис 10.22): лінії тренду, з якої вилучено сезонну компоненту, сезонної та випадкової компонент, (користуючись даними, розрахованими у таблиці на рисунку 10.20).

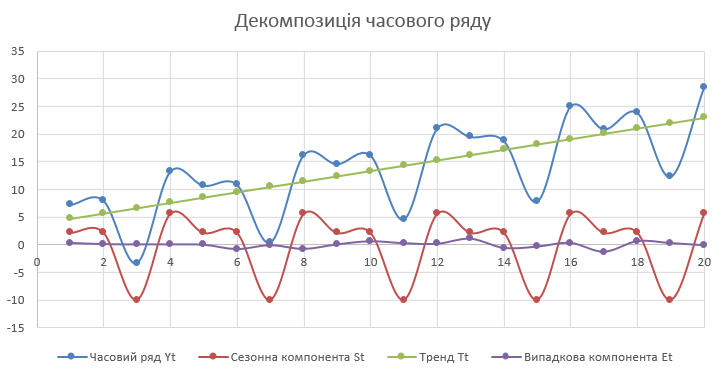


Рис. 10.22. Графіки часового ряду та його компонент

**2.7. Перевірка адекватності побудованої моделі, оцінка точності прогнозу**

1. Для оцінки адекватності побудованої моделі часового ряду імпортуємо в MatLab значення випадкової компоненти – стопця Е з таблиці MS Excel (рис. 10.21, комірки Р43:Р62) до MatLab та з допомогою функції  побудуємо її корелограму, ввівши команду:

autoсorr(E) % побудова корелограми випадкової компоненти

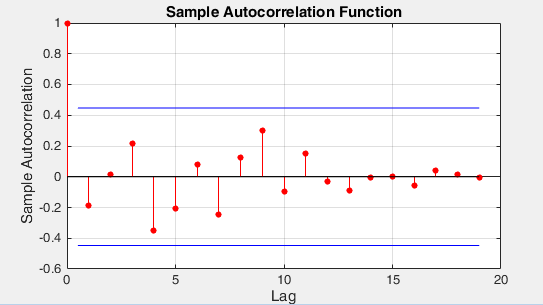


Рис. 10.23. Розрахунок оцінок якості моделі часового ряду

2. Аналіз графіку автокореляційної функції показує, що корелограма містить чергування затухаючих додатніх та від’ємних значень коефіцієнтів автокореляції (рис. 10.23). Усі коефіцієнти автокореляції не є значимими – не виходять за межі довірчого інтервалу, який на графіку корелограми знаходиться між двома синіми лініями. Отже, випадкова компонента є стаціонарним рядом, а побудована модель часового ряду є адекватною.

3. Для оцінки точності прогнозу за побудованою моделлю часового ряду розраховуємо  (комірки Q43:Q62) і знаходимо суму квадратів похибок та розрахуємо значення  та  (рис. 10.24). Аналізуємо отримані результати оцінки прогнозної якості побудованої моделі:

* ***MAD*** – середнє абсолютне відхилення розраховане за формулою 10.9: в середньому прогнозовані значення будуть відхилятися від фактичних у більшу або меншу сторону на 0,4 (формула у комірці R64: ).
* ***MSE*** – середня квадратична помилка становить 0,3046;
* ***MAPE*** – середня похибка апроксимації, розрахована за формулою 10.10, становить 3,96%, що свідчить про дуже високу точність прогнозу побудованої моделі (формула у комірці S66:  ).
* ***MPE*** – середня процентна похибка, розрахована за формулою 10.11, також близька до 0 та має додатній знак, що означає незначне завищення показника, в цілому прогноз близький до незміщеного (формула у комірці P65: ).

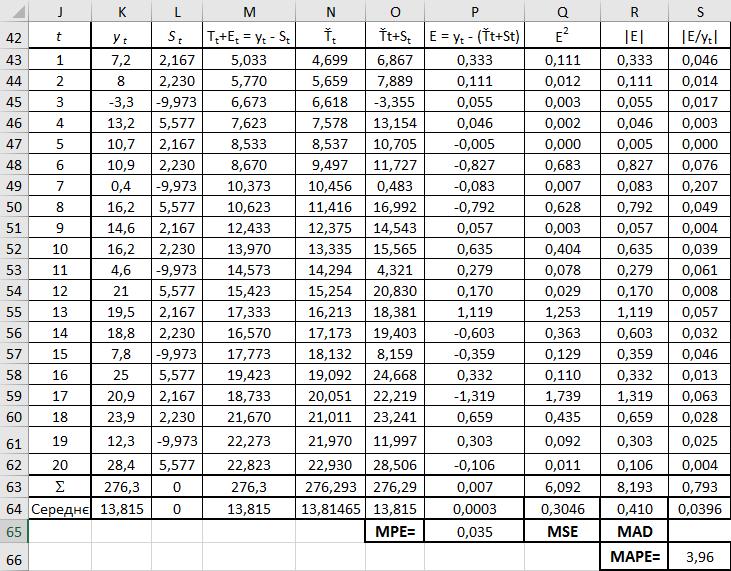


Рис. 10.24. Розрахунок оцінок точності прогнозу

Отримані оцінки свідчать про те, що побудована модель має високу точність, а отриманий на її основі прогноз близький до незміщеного.

**2.8. Прогнозування з використанням побудованої моделі часового ряду**

1. За допомогою побудованої адитивної тренд-сезонної моделі часового ряду здійснимо прогнозування значень з горизонтом прогнозу в 1 рік – на 4 квартали вперед.

2. Сформуємо таблицю MS Excel для розрахунку значень часового ряду для наступних після  чотирьох періодів, ввівши у стовпці  значення: 21, 22, 23 та 24. Значення у стовці  розраховуємо за формулою  У стовпці  вводимо відповідні їм значення сезонної компоненти (рис. 10.25).

3. Прогнозовані значення рівнів часового ряду  є сумою компоненти , розрахованої за побудованою лінійною трендовою моделлю та відповідного значення сезонної компоненти: . Ці значення можеть відхилятися від фактичних у більшу або меншу сторону не більше ніж на 0,4, оскільки розраховане середнє абсолютне відхилення MAD =0,4.

4. Побудуємо графік часового ряду , де перші 20 періодів будуть містити фактичні значення рівнів часового ряду, а наступні 4 після них – значення, спрогнозовані за побудованою моделлю (рис. 10.26).

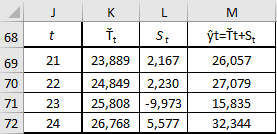


Рис. 9.25. Розрахунок прогнозованих значень часового ряду

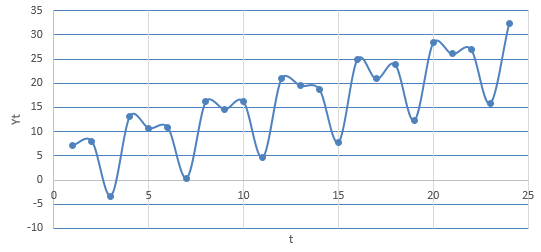


Рис. 9.26. Графічне зображення часового ряду та прогнозованих значень

**ІІІ. Прогнозування з використанням Листа прогнозу MS Excel**

У MS Excel є можливість для швидкого створення прогнозу з використанням інструментального засобу ***Лист прогнозу***.

**Завдання 2**. За даними часового ряду, що містить щоквартальні значення змінної  за 5 років, представлені у таблиці 10.4, здійснити побудову прогнозу на 3 наступні роки.

1. Для побудови прогнозу необхідно виділити дані з періодами та значеннями часового ряду: B7:C26 та обрати вкладку *Дані* – групу *Прогноз* – *Лист прогнозу*. Відкриється вікно *Создание листа прогноза* (рис. 10.27).

2. У вікні *Создание листа прогноза* необхідно натиснути кнопку *Параметри* та задати потрібні параметри прогнозу:

* + - початок прогнозу: *20*;
    - довірчий інтервал: *95%*;
    - сезонність: *установити вручну 4*;
    - заповнити відсутні точки: *інтерполяція*;
    - об’єднати дублікати даних: *середнє*;
    - завершення прогнозу: *12*.

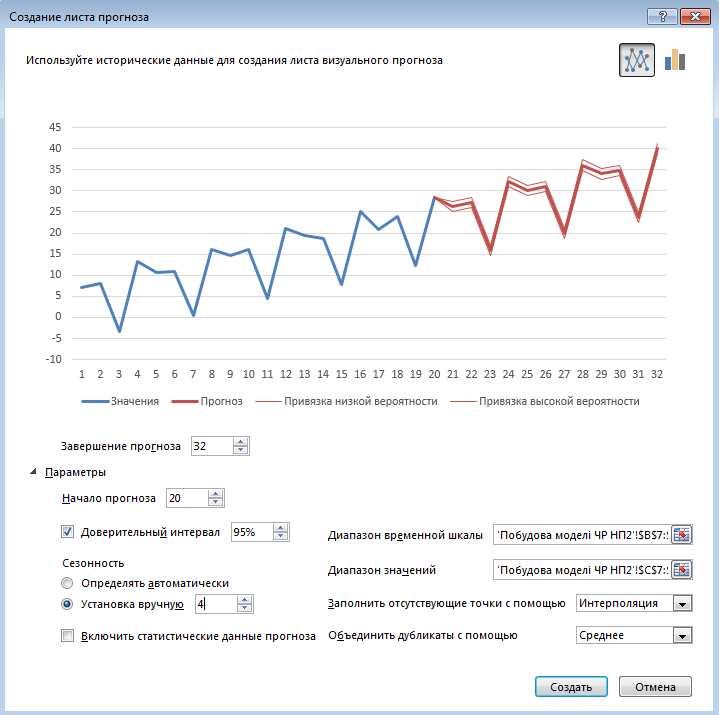


Рис. 10.27. Вікно Створення листа прогнозу

3. Натиснувши кнопку  у правому верхньому куті вікна *Создание листа прогноза*, можна переглянути діаграму з прогнозованими значеннями (рис. 10.28).

4. Натиснути кнопку *Создать*, відкриється новий робочий аркуш з відображенням результату прогнозу часового ряду у графічному вигляді та у вигляді таблиці (рис. 10.29).

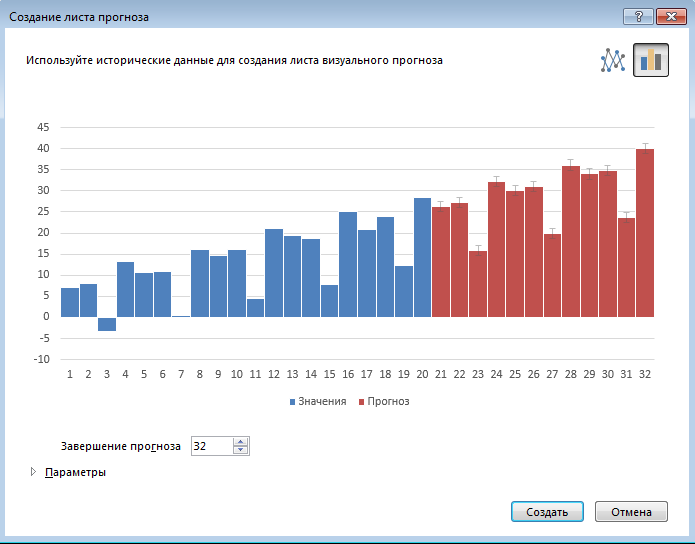


Рис. 10.28. Діаграма часового ряду з прогнозованими значеннями

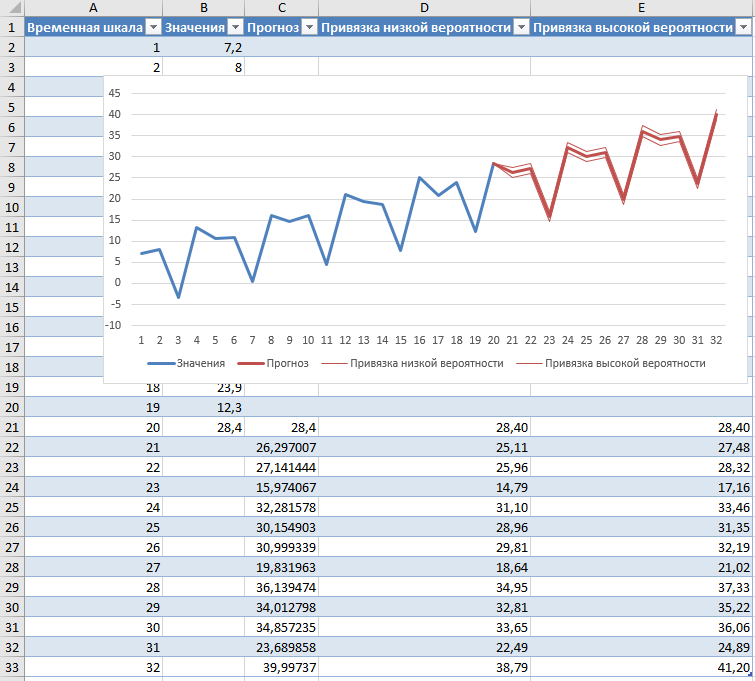


Рис. 10.29. Робочий аркуш із результатами прогнозу

## ІІІ. ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ

**Завдання 3.** За даними часового ряду, що містить щоквартальні значення змінної  за 4 роки, представлені у таблиці 10.5 (по варіантам), необхідно здійснити аналіз даних з метою побудови адитивної (чи мультиплікативної) моделі часового ряду:

1) побудувати графік і корелограму часового ряду та здійснити аналіз динаміки зміни значень рівнів ряду і провести автокореляційний аналіз часового ряду для уточнення його структури;

2) установити наявність тренду з використанням методу перевірки середніх різниць або за допомогою критерію серій;

3) здійснити згладжування рівнів ряду: вирівнювання вихідного ряду методом ковзного середнього;

4) розрахувати значення сезонної компоненти та усунути її з вихідних рівнів ряду, провести декомпозицію часового ряду;

5) здійснити побудову трендових моделей: формування набору апроксимуючих функцій, чисельне оцінювання їх параметрів, оцінку точності апроксимації, вибір кращої моделі;

6) провести аналітичне вирівнювання, розрахунок значень  з використанням обраної трендової моделі;

7) дослідити випадкову компоненту на стаціонарність, перевірити адекватність побудованої моделі часового ряду;

8) оцінити точності прогнозу з використанням побудованої моделі часового ряду.

9) здійснити побудову прогнозу на 4 наступні квартали;

10) скористатися листом прогнозу MS Excel для побудови прогнозу на наступний рік.

## КОНТРОЛЬНІ ПИТАННЯ

1. Задача прогнозування в Data Mining.
2. Часові ряди: основні поняття та характеристики.
3. Види часових рядів у залежності від характеру рівнів ряду та динаміки змін основних характеристик.
4. Автоковаріація та автокореляція рівнів часового ряду.
5. Структура часового ряду та основні методи її виявлення.
6. У чому полягає аналіз часових рядів? Декомпозиція часового ряду.
7. Види моделей часового ряду, основні підходи до побудови моделей.
8. Оцінка точності та адекватності побудованої моделі часового ряду.
9. Основні етапи прогнозування процесів, представлених одномірними часовими рядами.
10. Виявлення аномальних відхилень, критерій Ірвіна.
11. Перевірка наявності тренду методом перевірки різниць середніх рівнів та з використання критерію серій.
12. Автокореляційних аналіз часового ряду.
13. Моделювання часового ряду на основі механічного згладжування.
14. Авторегресійний аналіз часового ряду.
15. Яким чином здійснюється розрахунок значень сезонної компоненти методом ковзного середнього?
16. Основні етапи побудови адитивної/мультиплікативної моделей часового ряду.
17. Прогнозування в MS Excel з використанням побудови лінії тренду, листа прогнозу.

***Таблиця 10.5***

**Індивідуальні варіанти до завдання 3 лабораторної роботи**

