

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Запорізький національний технічний університет

МЕТОДИЧНІ ВКАЗІВКИ

до виконання лабораторних робіт з дисципліни
**“КОМП’ЮТЕРНА ТЕХНІКА В НАУКОВИХ
ДОСЛІДЖЕННЯХ”**

для студентів спеціальності

7(8).05010201 ”Комп’ютерні системи та мережі”,
7(8).05010203 ”Спеціалізовані комп’ютерні системи”
усіх форм навчання

2015 р.

Методичні вказівки до виконання лабораторних робіт з дисципліни "Комп'ютерна техніка в наукових дослідженнях" для студентів спеціальностей 7(8).05010201 "Комп'ютерні системи та мережі", 7(8).05010203 "Спеціалізовані комп'ютерні системи" усіх форм навчання/ Укл.: К.М. Касьян, М.М. Касьян.- Запоріжжя: ЗНТУ, 2015. - 38с.

Укладачі: К.М. Касьян, доцент, к.т.н.
М.М. Касьян, доцент, к.т.н.

Рецензент Р.К. Кудерметов, доцент, к.т.н.

Відповідальний
за випуск: К.М. Касьян, доцент, к.т.н.

Затверджено
на засіданні кафедри
"Комп'ютерні системи та мережі"
Протокол № 3 від 26 жовтня 2015 р.

ЗАГАЛЬНІ ПОЛОЖЕННЯ

Методичні вказівки призначені для вивчення і практичного освоєння студентами основних методів аналізу різного роду експериментальних даних при наявності випадкових та непередбачених впливів, умінні виявляти закономірності на тлі випадків, аналізувати їх та будувати достовірні прогнози. Лабораторний курс складається з ряду взаємопов'язаних робіт, що охоплює стадії генерації експериментальних даних у вигляді часових рядів, їх обробки та аналізу за допомогою статистичних пакетів, а також документування отриманих результатів за допомогою спеціалізованого редактора.

Лабораторні роботи виконуються студентами шляхом застосування сучасних засобів і прийомів комп'ютерного аналізу даних.

Перед виконанням лабораторної роботи студенти повинні ознайомитися з методичними вказівками по її виконанню.

Для одержання заліку по виконаній роботі кожен студент здає викладачу цілком оформлень звіт. Під час співбесіди студент повинний виявити знання про мету роботи, по теоретичному матеріалу і змісту комп'ютерного аналізу даних, з демонстрацією результатів на прикладі конкретних часових рядів.

Загальний залік студент одержує після виконання і здачі останньої лабораторної роботи.

1 ЗАГАЛЬНІ ТЕОРЕТИЧНІ ВІДОМОСТІ

У повсякденному житті і професійній діяльності, а також у наукових дослідженнях ми постійно зіштовхуємося з подіями і явищами з невизначеним результатом. При цьому постійно приходиться приймати в подібних невизначених, зв'язаних з багатьма випадками ситуаціях свої рішення, іноді дуже важливі. У побуті чи в нескладному бізнесі ми можемо приймати такі рішення на основі здорового глузду, інтуїції, попереднього досвіду. Однак у більш серйозному бізнесі, в умовах жорсткої конкуренції, рішення повинні прийматися на основі ретельного аналізу наявної інформації, бути обґрунтованими і доказовими.

Для рішення задач, зв'язаних з аналізом даних при наявності випадкових і непередбачених впливів, математиками та іншими дослідниками (біологами, психологами, економістами і т.д.) за останні двісті років був вироблений могутній і гнучкий арсенал методів, названих у сукупності математичною статистикою (а також прикладною статистикою чи аналізом даних). Ці методи дозволяють виявляти закономірності на тлі випадків, робити обґрунтовані висновки і прогнози, давати оцінки імовірностей їхнього виконання або невиконання.

Широкому впровадженню методів аналізу даних у 60-х і 70-х роках XX століття чимало сприяла поява комп'ютерів, а починаючи з 80-х років XX століття - персональних комп'ютерів. Статистичні програмні пакети зробили методи аналізу даних більш доступними і наочними: тепер уже не було потрібно вручну виконувати трудомісткі розрахунки по складних формулах, будувати таблиці і графіки - усю цю чорнову роботу взяв на себе комп'ютер, а людині залишилася головним чином творча робота: постановка задач, вибір методів їхнього рішення та інтерпретація результатів.

Результатом появи могутніх і зручних пакетів для аналізу даних на персональних комп'ютерах стало різке розширення і зміна кола споживачів методів аналізу даних. Якщо раніш ці методи розглядалися головним чином як інструмент наукових досліджень то, починаючи із середини 1980-х років основними покупцями статистичних пакетів стали вже не наукові, а комерційні організації, а також урядові і медичні установи. Таким чином, методи аналізу даних і статистичні пакети для комп'ютерів стали типовим і загальноновживаним інструментом планових, аналітичних, маркетингових відділів виробничих і торгових

корпорацій, банків і страхових компаній, урядових і медичних установ.

Приведемо кілька прикладів застосування методів статистичного аналізу даних у практичних задачах.

Розглянемо досить просту, але часто зустрічається задачу. Припустимо, що Ви увели важливе нововведення: змінили систему оплати праці, перейшли на випуск нової продукції, використовували нову технологію і т.д. Вам здається, що це дало позитивний ефект, але чи дійсно це так? А може бути цей удаваний ефект визначений зовсім не Вашим нововведенням, а природною випадковістю, і вже завтра Ви можете одержати прямо протилежний, але настільки ж випадковий ефект? Для рішення цієї задачі треба сформулювати два набори чисел, кожний з яких містить значення цікавлячого Вас показника ефективності до і після нововведення. Статистичні критерії порівняння двох вибірок покажуть Вам, випадкові чи не випадкові розходження цих двох рядів чисел.

Інша важлива задача — прогнозування майбутнього поведіння деякого часового ряду: зміни курсу долара, цін і попиту на продукцію чи сировину і т.д. Для такого часового ряду за допомогою статистичного пакета програм підбирають деяке аналітичне рівняння - будують регресійну модель. Якщо ми припускаємо, що на цікавлячий нас показник впливають деякі інші фактори, їх теж можна включити в модель, попередньо (за допомогою того ж статистичного пакета) перевіривши істотність (значимість) цього впливу. Потім на основі побудованої моделі можна зробити прогноз і вказати його точність.

У багатьох технологічних процесах необхідно систематично контролювати стан процесу, щоб вчасно втрутитися при відхиленнях його від нормального режиму і запобігти тим самим втрати від випуску неякісної продукції. Для цього використовуються статистичні методи контролю якості, повсюдне і неухильне застосування яких багато в чому визначило разучі успіхи японської промисловості.

Ще часто зустрічається задача зв'язана з класифікацією об'єктів. Наприклад, Ви є начальником кредитного відділу банку. Зштовхнувшись з неповерненням кредитів, Ви вирішите надалі видавати кредити лише фірмам, що «схожі» з тими, котрі себе добре зарекомендували, і не видавати тим, що «схожі» з неплатниками чи шахраями. Для класифікації фірм можна зібрати показники їхньої діяльності (наприклад, розмір основних фондів, валютного балансу, вид діяльності,

обсяг реалізації і т.д.), і провести кластерний аналіз. У багатьох випадках наявні об'єкти вдається згрупувати в декілька груп (кластерів), і Ви зможете побачити, чи не належить запитуюча кредит фірма до групи неплатників.

Можна привести ще безліч інших цікавих прикладів застосування методів аналізу даних у самих різних областях: торгівлі й охороні здоров'я, утворенні і керуванні і т.д.

Методи статистичного аналізу є універсальними і можуть застосовуватися в самих різних галузях людської діяльності. Скажемо, прококування курсу долара і прогноз попиту на автомобілі робляться за допомогою тих самих процедур. У переважній більшості випадків усі потрібні користувачам задачі по аналізу даних можуть бути вирішені за допомогою універсальних статистичних пакетів. Зрозуміло, є і такі області людської діяльності, у яких виникають специфічні, що не зустрічаються в інших областях, задачі по аналізу даних, і тому вимагаються специфічні статистичні засоби. Однак таких областей дуже мало.

2 ЗАВДАННЯ НА ВИКОНАННЯ КУРСУ ЛАБОРАТОРНИХ РОБІТ

Лабораторний курс складається з виконання наступного комплексу лабораторних робіт:

- Лабораторна робота №1 “Побудова часових рядів і вивчення їхніх основних характеристик”;
- Лабораторна робота №2 “Аналіз часових рядів за допомогою статистичного пакета TableCurve”;
- Лабораторна робота №3 “Аналіз часових рядів за допомогою статистичного пакета SPSS”;
- Лабораторна робота №4 “Документування результатів аналізу часових рядів за допомогою спеціалізованого пакету LaTeX”.

2.1 Лабораторна робота №1. “Побудова часових рядів і вивчення їхніх основних характеристик”

2.1.1 Мета роботи

Метою роботи є побудова і вивчення основних характеристик часових рядів, використовуючи стандартні математичні методи опису.

2.1.2 Теоретичні відомості

Часовим рядом називається набір значень $x_1, x_2, \dots, x_n$ досліджуваної величини, зареєстрованих у послідовні моменти часу:

$$t_1, t_2, \dots, t_n.$$

Часовий ряд прийнято описувати за допомогою закономірної не випадкової і випадкової складових.

Невипадкова (детермінована) складова часового ряду являє собою функцію від часу, що обчислюється в дискретні моменти часу.

Випадкова складова – набір випадкових величин, розподіл яких невідомий.

Форми розкладання (декомпозиції) часового ряду на детерміновану і випадкову компоненти можуть розрізнятися.

Адитивною моделлю часового ряду називається представлення ряду у вигляді суми детермінованої і випадкової компонент, а саме:

$$x_t = d_t + \varepsilon_t \quad \text{при} \quad t = 1, \dots, n.$$

Мультиплікативною моделлю часового ряду називається представлення ряду у вигляді перемноження детермінованої і випадкової компонент, а саме:

$$x_t = d_t \times \varepsilon_t \quad \text{при} \quad t = 1, \dots, n.$$

Способи опису детермінованих компонент часового ряду сильно залежать від області застосування. В економічних (і багатьох інших)

застосуваннях у детермінованій компоненті часового ряду d_t звичайно виділяють три складові частини: тренд tr_t , сезонну компоненту s_t і циклічну компоненту c_t . Для адитивної моделі часового ряду можна записати:

$$d_t = tr_t + s_t + c_t, \quad \text{при } t = 1, \dots, n.$$

Трендом часового ряду tr_t при $t = 1, \dots, n$ називають плавно змінну, не циклічну компоненту, що описує чистий вплив довгострокових факторів, ефект яких позначається поступово.

Сезонна компонента s_t часового ряду при $t = 1, \dots, n$ описує поведінку, що змінюється регулярно протягом заданого періоду (року, місяця, тижня, дня і т.і.). Вона складається з послідовності циклів, що майже повторюються.

Циклічна компонента c_t часового ряду описує тривалі періоди відносного підйому і спаду. Вона складається з циклів, що змінюються по амплітуді і довжині.

Математично найпростішою моделлю випадкової компоненти часового ряду є послідовність незалежних випадкових величин. *Білим шумом* називають часовий ряд (випадковий процес) з нульовим середнім, якщо його складові випадкові величини незалежні і розподілені однаково при всіх t .

Послідовності незалежних випадкових величин далеко не завжди адекватно описують випадкові компоненти часових рядів. Теорією і практикою для опису випадкових послідовностей вироблені і більш складні моделі.

Нехай $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n, \dots$ - незалежні однаково розподілені випадкові величини (білий шум).

Процесом ковзного середнього (першого порядку) із середнім μ називають процес $X(t)$:

$$X(t) = \varepsilon_t + \theta \varepsilon_{t-1} + \mu,$$

де θ - деякий числовий коефіцієнт, а μ - константа.

У процесі ковзного середнього статистично залежні тільки сусідні величини $X(t-1)$ і $X(t)$. Значення процесу, що розділені проміжком часу 2 і більше, статистично незалежні, тому що в їхньому формуванні беруть участь різні складові ε_t . З цієї причини процеси ковзного середнього є безпосереднім і найпростішим узагальненням процесів білого шуму.

Процесом авторегресії (першого порядку) із середнім значенням μ називають випадковий процес $X(t)$, що задовольняє співвідношенню:

$$X(t) - \mu = \phi \cdot (X(t-1) - \mu) + \varepsilon_t,$$

де ϕ і μ - деякі числа.

Члени процесу авторегресії, що розділені проміжком часу $h > 0$, не стають незалежними, яким би великим ні було h . Однак залежність між ними швидко убуває з ростом h , якщо $|\phi| < 1$. Саме такі процеси авторегресії звичайно зустрічаються в прикладних задачах.

Числові характеристики часових рядів вводяться в повній аналогії з числовими характеристиками випадкових величин.

Математичне очікування:
$$\bar{\varepsilon} = \frac{1}{n} \sum_{t=1}^n \varepsilon_t .$$

Дисперсія:
$$\sigma^2 = \frac{1}{n} \sum_{t=1}^n (\varepsilon_t - \bar{\varepsilon})^2 .$$

Коефіцієнт Пірсона (показує ступінь статистичної залежності часового ряду і тренда: чим ближче цей коефіцієнт до 1, тим краще підібраний тренд):

$$R = \frac{\sum_{t=1}^n (x_t - \bar{x})(tr_t - \overline{tr})}{\sqrt{\sum_{t=1}^n (x_t - \bar{x})^2} * \sqrt{\sum_{t=1}^n (tr_t - \overline{tr})^2}},$$

де $\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t$, $\overline{tr} = \frac{1}{n} \sum_{t=1}^n tr_t$.

Автокореляційна функція (показує залежність між елементами випадкової складової, якщо усі $\bar{r}(k)$ близькі до 0 ($|\bar{r}(k)| < 0.2$), то всі елементи випадкової складової можна вважати незалежними, тобто ми маємо справу з білим шумом):

$$\bar{r}(k) = \frac{\sum_{t=1}^{n-k} (\varepsilon_t - \bar{\varepsilon}) \cdot (\varepsilon_{t+k} - \bar{\varepsilon})}{\sum_{t=1}^n (\varepsilon_t - \bar{\varepsilon})^2}; \quad k = 1, 2, 3, \dots, n/4$$

2.1.3 Завдання на виконання роботи

Для роботи використовується математичний пакет MathCAD, що надає повний набір функціональних можливостей, і забезпечує необхідну точність обчислень.

Виконати наступні завдання:

- згенерувати часові ряди для приведених нижче моделей:

$$x_t = tr_t + \varepsilon_t - \text{ часовий ряд,}$$

де

$$tr_t = a + b \cdot t;$$

$$tr_t = a + b \cdot t + c \cdot t^2 + d \cdot t^3 + \dots;$$

$$tr_t = e^{a+bt};$$

$$tr_t = \frac{a}{1 + b \cdot e^{-c \cdot t}};$$

$$\log tr_t = a - b \cdot r^t, r \in (0,1);$$

$$tr_t = a \cdot \cos(\varpi \cdot t + \phi);$$

i

$\varepsilon_t \in N(0, \sigma^2)$ - білий шум (мат. очікування 0, дисперсія σ^2);

$\varepsilon_t = \alpha_t + \theta \cdot \alpha_{t-1}$, $\alpha_t \in N(0, \sigma^2)$ - ковзне середнє;

$\varepsilon_t = \mu \cdot \varepsilon_{t-1} + \alpha_t$, $\alpha_t \in N(0, \sigma^2)$ - процес авторегресії.

- побудувати графіки часового ряду x_t , тренда tr_t і випадкової складової ε_t .
- обчислити математичне очікування $\bar{\varepsilon}$ випадкової складової часового ряду.
- зробити розрахунок дисперсії $\bar{\sigma}^2$ випадкової складової часового ряду.
- обчислити коефіцієнт Пірсона R , автокореляційну функцію $\bar{r}(k)$ і побудувати її графік (корелограму).
- згенерувати текстовий файл з набором обчислених значень.

2.1.4 Зміст звіту

Звіт включає в себе:

- титульний лист;
- мета роботи;
- результати, отримані при генеруванні часових рядів;
- графіки для кожної моделі часового ряду;
- числові характеристики для кожної моделі часового ряду;
- висновки.

2.1.5 Контрольні питання

2.1.5.1 Що таке часовий ряд?

2.1.5.2 Що називається детермінованою складовою часового ряду, і з яких компонентів вона складається?

2.1.5.3 Які моделі найбільш часто використовують для опису випадкової складової часового ряду?

2.1.5.4 Перелічите основні числові характеристики часового ряду.

2.1.5.5 Як визначається і що показує коефіцієнт Пірсона?

2.1.5.6 Як визначається і що показує автокореляційна функція?

2.2 Лабораторна робота № 2. “Аналіз часових рядів за допомогою статистичного пакета TableCurve”

2.2.1 Мета роботи

Метою роботи є підбір моделей трендів для кожного часового ряду і порівняння результатів з моделями трендів, що використовувались для генерації часових рядів у лабораторній роботі №1.

2.2.2 Теоретичні відомості

Коротко розглянемо загальний порядок прикладного статистичного аналізу часових рядів. Звичайно метою такого аналізу є побудова математичної моделі ряду, за допомогою якої можна пояснити поведінку ряду і здійснити прогноз його подальшого поведінку.

Побудова і вивчення графіка. Аналіз часового ряду звичайно починається з побудови і вивчення його графіка. Якщо нестационарність часового ряду очевидна, то першою справою треба виділити і видалити нестационарну складову ряду. Процес видалення тренда й інших компонентів ряду, що приводять до порушення стаціонарності, може проходити в декілька етапів. На кожному з них розглядається ряд залишків, отриманий у результаті вирахування з вихідного ряду підбраної моделі тренда, або результат різницевих та інших перетворень ряду. Крім графіків, ознаками нестационарності часового ряду може бути не прагнуча до нуля автокореляційна функція (за винятком дуже великих значень лагів) та наявність яскраво виражених піків на низьких частотах у періодограмі.

Підбір моделі для часового ряду. Після того, як вихідний процес максимально наближений до стаціонарного, можна приступити до підбору різних моделей отриманого процесу. Мета цього етапу - опис і облік в подальшому аналізі кореляційної структури розглянутого процесу. При цьому на практиці частіше використовуються два типи моделей: параметричні моделі авторегресії - ковзного середнього (ARMA-моделі) і полігармонійні моделі.

Модель може вважатися підбраною, якщо залишковий компонент ряду є процесом типу білого шуму. Після підбору моделі звичайно виконуються:

- оцінка дисперсії залишків, що надалі може бути використана для побудови довірчих інтервалів прогнозу;

- аналіз залишків з метою перевірки адекватності моделі.

Прогнозування або інтерполяція. Останнім етапом аналізу часового ряду може бути прогнозування його майбутніх (екстраполяція) чи відновлення пропущених (інтерполяція) значень і вказівки точності цього прогнозу на базі підібраної моделі. Добре підібрати математичну модель вдається не для всякого часового ряду. Нерідко буває і так, що для опису підходять відразу декілька моделей. Неоднозначність підбору моделі може спостерігатися як на етапі виділення детермінованої компоненти ряду, так і при виборі структури ряду залишків. Тому дослідники досить часто даються до методу декількох прогнозів, зроблених за допомогою різних моделей.

Методи аналізу. Основні групи статистичних прийомів, що використовуються для аналізу часових рядів:

- графічні методи представлення часових рядів і їх супутніх числових характеристик;

- методи зведення до стаціонарних процесів;

- методи дослідження внутрішніх зв'язків між елементами часових рядів.

У вибіркових дослідженнях найпростіші числові характеристики описової статистики (середнє, медіана, дисперсія, стандартне відхилення, коефіцієнти асиметрії та ексцесу) звичайно дають досить інформативне представлення про вибірку. Графічні методи представлення й аналізу вибірок при цьому грають лише допоміжну роль, дозволяючи краще зрозуміти локалізацію і концентрацію даних, їхній закон розподілу.

Роль графічних методів при аналізі часових рядів зовсім інша. Справа в тому, що табличне представлення часового ряду та описові статистики найчастіше не дозволяють зрозуміти характер процесу, у той час як за графіком часового ряду можна зробити досить багато висновків. Надалі вони можуть бути перевірені й уточнені за допомогою розрахунків.

Людське око досить упевнено визначає за графіком часового ряду:

- наявність тренда і його характер;

- наявність сезонних і циклічних компонентів;

– ступінь плавності чи переривчастості змін послідовних значень ряду після усунення тренда. По цьому показнику можна судити про характер і ступінь кореляції між сусідніми елементами ряду.

При аналізі часових рядів часто використовуються допоміжні графіки для числових характеристик ряду:

- графік вибіркової автокореляційної функції (корелограми) з довірчою зоною (трубкою) для нульової автокореляційної функції;
- графік вибіркової часткової автокореляційної функції з довірчою зоною для нульової часткової автокореляційної функції;
- графік періодограми.

Перші два з цих графіків дозволяють судити про зв'язок (залежність) сусідніх значень часового ряду, вони використовуються при підборі параметричних моделей авторегресії - ковзного середнього. Графік періодограми дозволяє судити про наявність гармонійних складових у часовому ряді.

2.2.3 Завдання на виконання роботи

- Побудувати часові ряди за даними, отриманим при генерації рядів у лабораторній роботі №1.
- Підібрати лінії тренда для кожного ряду.
- Порівняти коефіцієнти в підібраних моделях тренда і коефіцієнт Пірсона з TableCurve і MathCAD.

2.2.4 Хід роботи

2.2.4.1 Побудова часового ряду.

У директорії Tcwin2 знайти і запустити файл Tc.exe. Командою File>Import і за допомогою діалогового вікна (рис. 2.1) відкрити текстовий файл з даними, отриманими при генерації часового ряду в лабораторній роботі №1. В процесі відкриття файлу ввести у відповідному діалоговому вікні (рис.2.2) початкове значення і крок для вісі X, а також задати назву для графіка і підписи для осей координат. У результаті на екран буде виведений імпортований часовий ряд (рис.2.3).

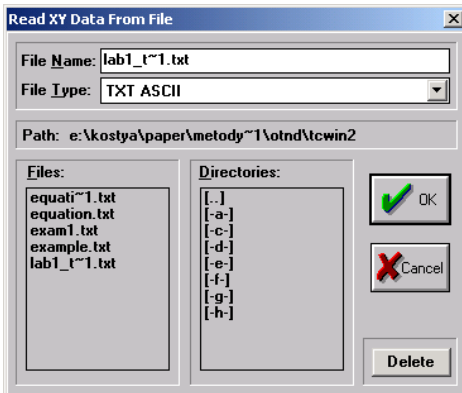


Рисунок 2.1 – Вікно імпорту текстового файлу з часовим рядом

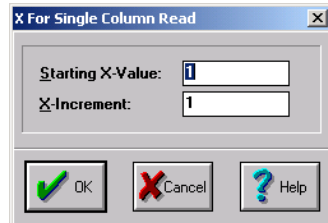


Рисунок 2.2 – Вікно завдання кроку по вісі X

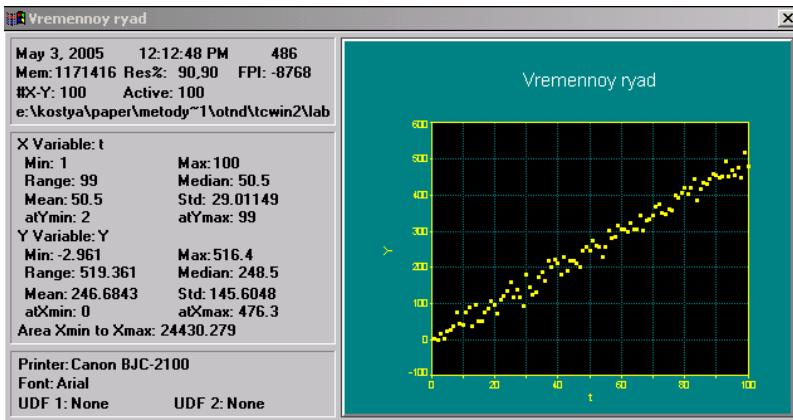


Рисунок 2.3 – Імпортований часовий ряд

2.2.4.2 Підбір тренда для часового ряду.

Вибрати в меню команду **Process>Curve-Fit Simple Equations**, в результаті програма автоматично підбере для ряду тренд і виведе його на екран (рис.2.4), також буде виведений список рівнянь для тренда, з яких можна підібрати тренд вручну, якщо підібраний автоматично не зовсім підходить.

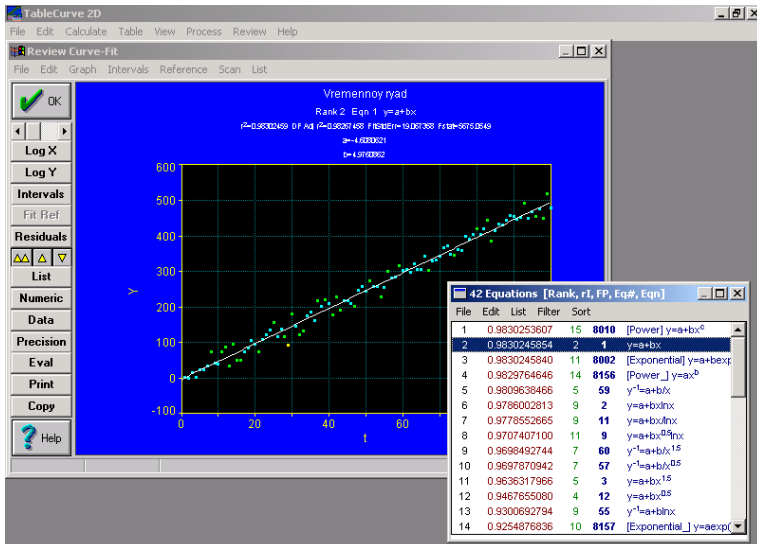


Рисунок 2.4 – Часовий ряд і підібраний до нього тренд

Коли тренд підібраний, потрібно оцінити його точність за допомогою квадрата коефіцієнта Пірсона, що разом з рівнянням тренда і його параметрами виводиться на екран над графіком (рис.2.5), та порівняти їх зі значеннями, використаними при генерації часового ряду в лабораторній роботі №1.

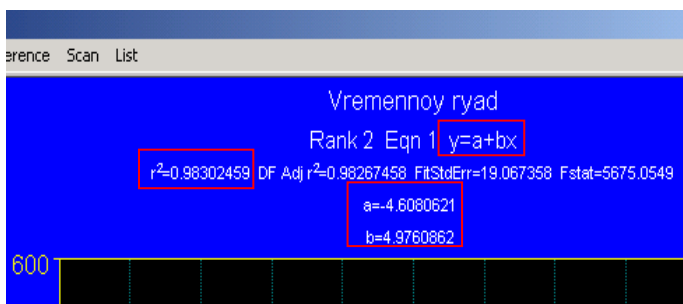


Рисунок 2.5 – Параметри підбраного тренда

2.2.5 Зміст звіту

Звіт включає в себе:

- титульний лист;
- мета роботи;
- графіки для часового ряду і підбраного тренда для кожного ряду, що досліджується;
- числові параметри підбраного тренда для кожного часового ряду, що досліджується, та результати їхнього порівняння з відповідними параметрами з лабораторної роботи №1;
- висновки.

2.2.6 Контрольні питання

2.2.6.1 Що є метою аналізу часових рядів.

2.2.6.2 Опишіть порядок аналізу часових рядів.

2.2.6.3 Перелічите основні групи статистичних прийомів, що використовуються для аналізу часових рядів.

2.2.6.4 У чому полягає роль графічних методів при аналізі часових рядів?

2.2.6.5 Яку інформацію несе графік часового ряду?

2.2.6.6 Які допоміжні графіки для числових характеристик ряду використовуються при аналізі часових рядів?

2.2.6.7 Яку інформацію несе графік автокореляційної функції?

2.3 Лабораторна робота №3. “Аналіз часових рядів за допомогою статистичного пакета SPSS”

2.3.1 Мета роботи

Метою роботи є підбір моделей трендів і випадкових компонент для кожного часового ряду і порівняння результатів з моделями трендів, отриманими в лабораторній роботі №2, а також з моделями трендів і випадкових компонент, що використовувались для генерації часових рядів в лабораторній роботі №1.

2.3.2 Теоретичні відомості

Після видалення детермінованої компоненти часовий ряд повинний звестися до стаціонарного процесу. Так що наступним кроком після виділення детермінованої компоненти повинний бути аналіз залишків, тобто вивчення ряду, отриманого з вихідного часового ряду після виключення детермінованої компоненти. При цьому можуть ставитися наступні цілі:

- Опис ряду за допомогою тієї чи іншої моделі, що відбиває залежність між його сусідніми елементами. На базі побудованої моделі можна здійснювати прогноз майбутнього поведіння ряду.

- Уточнення оцінки дисперсії часового ряду. Ця оцінка важлива для прогнозування, тому що виходячи з неї обчислюється ширина довірчої трубки прогнозу.

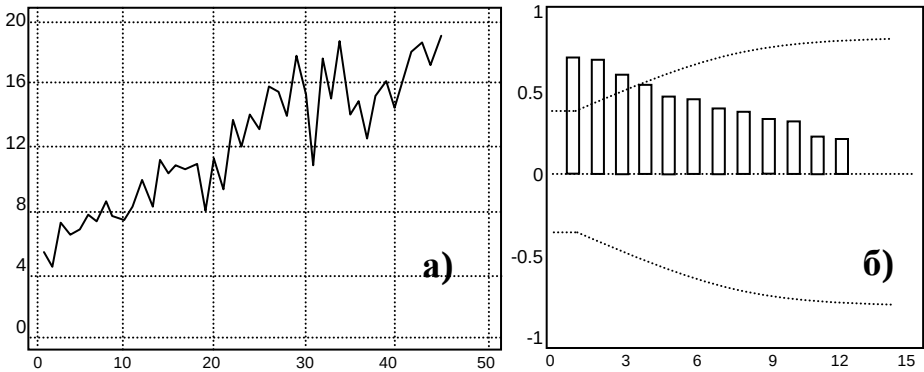
- Перевірка стаціонарності залишків (при нестаціонарності підбор детермінованої компоненти має потребу в уточненні).

Як модель стаціонарних часових рядів найчастіше використовуються процеси авторегресії, ковзного середнього та їхні комбінації.

Для перевірки стаціонарності ряду залишків і оцінки його дисперсії на практиці найчастіше використовуються вибіркова автокореляційна (корелограма) і часткова автокореляційна функції.

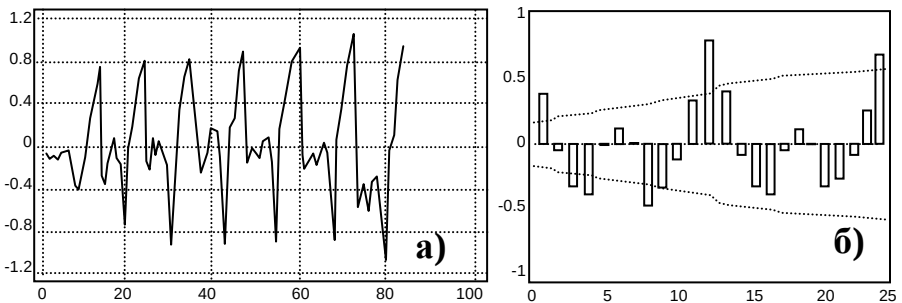
Аналіз корелограми - це порой досить непроста задача. Розглянемо поведінку корелограми для деяких нестаціонарних рядів. У цьому випадку варто пам'ятати, що корелограма практично не несе ніякої інформації про статистичну залежність чи незалежність членів часового ряду, однак вона може відбивати причини порушення стаціонарності.

Для часового ряду, що містить *тренд*, корелограма не прагне до нуля з ростом значення лага. Її характерне поведіння зображене на рис. 2.6.



а) вихідний часовий ряд; б) його корелограма
Рисунок 2.6 - Корелограма ряду, що містить тренд:

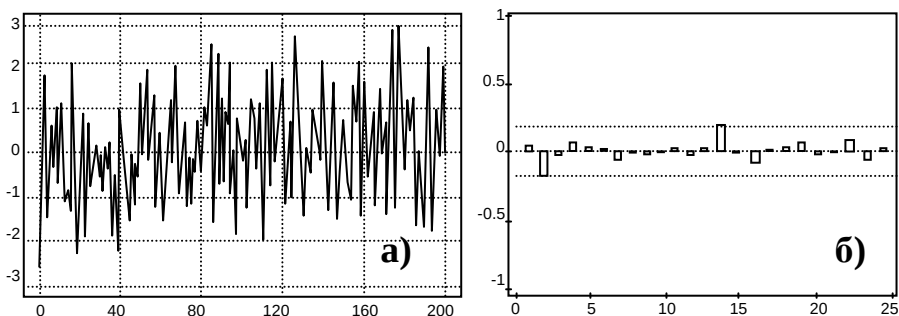
Для ряду з *сезонними коливаннями* корелограма також буде містити періодичні сплески, що відповідають періоду сезонних коливань. Це дозволяє встановлювати передбачуваний період сезонності. Типове поведіння корелограми для ряду з сезонними коливаннями приведене на рис. 2.7.



а) вихідний часовий ряд; б) його корелограма
Рисунок 2.7 - Корелограма ряду з сезонними коливаннями:

У випадку стаціонарних випадкових процесів корелограма показує корельованість значень часового ряду при різних відстанях між ними.

Автокореляційна функція r_k білого шуму дорівнює нулю для всіх $k \neq 0$. На рис. 2.8 зображена типова корелограма білого шуму. Для гаусовського білого шуму можна вказати 95% довірчий інтервал для кожного конкретного значення $\bar{r}_k$ у вигляді $-1/\sqrt{n} \pm 2/\sqrt{n}$. Він зображений на графіку корелограми пунктирними лініями. Якщо вибіркові оцінки кореляційної функції попадають у зазначені довірчі інтервали, то можна припустити, що значення процесу є білим шумом. Однак досить часто одне чи декілька значень вибіркової автокореляційної функції білого шуму можуть виходити з зазначених меж. Особливо часто цей ефект можна спостерігати при наявності невеликого числа спостережень.



а) вихідний ряд; б) корелограма
Рисунок 2.8 – Корелограма білого шуму

Траєкторії багатьох стаціонарних випадкових процесів виглядають набагато більш гладко, чим траєкторії білого шуму. Це зв'язано з наявністю позитивної кореляції між двома чи декількома сусідніми членами подібних рядів. Якщо ж кореляція між сусідніми членами ряду негативна, то траєкторії подібних процесів будуть більш зламаними, чим траєкторії білого шуму. Найпростішим прикладом процесів, у яких залежні одне чи декілька сусідніх значень, є процеси *ковзного середнього*. На рис. 2.9 приведені графіки ста значень реалізації

процесу ковзного середнього з коефіцієнтом $\theta = 0.75$ і його корелограмами. На рис. 2.10 приведені аналогічні графіки при $\theta = -0.75$.

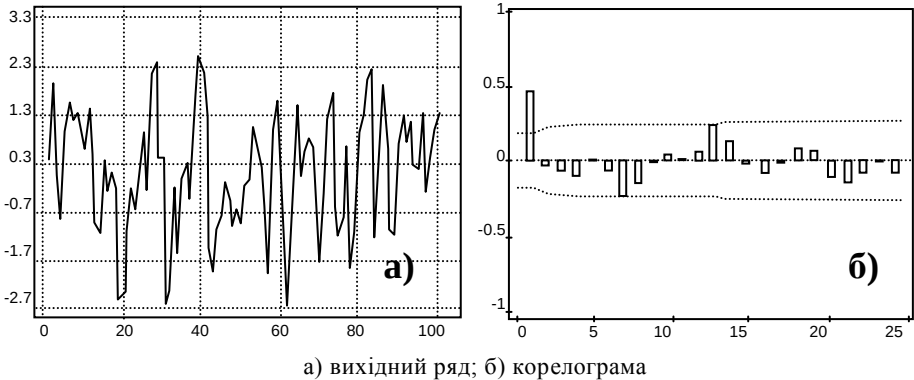


Рисунок 2.9 - Корелограма процесу ковзного середнього при $\theta = 0.75$

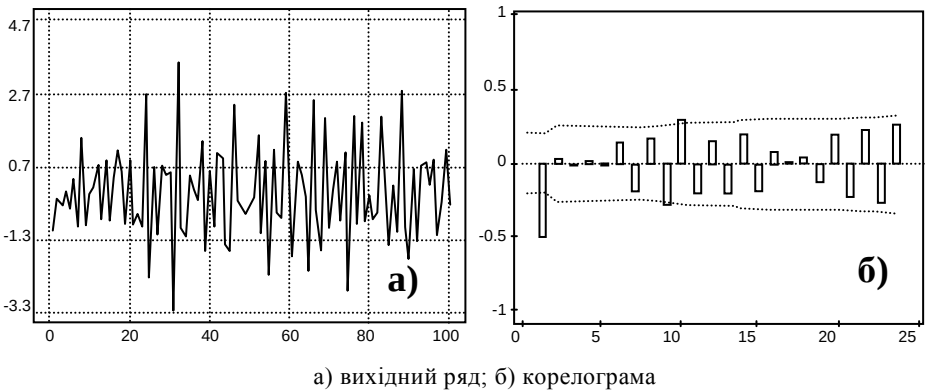


Рисунок 2.10 - Корелограма процесу ковзного середнього при $\theta = -0.75$

На графіках видно, що хоча отримані оцінки значень r_k при $k = 2, 3, \dots$ не дорівнюють нулю, вони значиме не відрізняються від нульових значень, тому що попадають у 95% довірчий інтервал, що побудований у припущенні рівності нулю відповідних значень автокореляційної функції.

Для процесів ковзного середнього другого порядку відрізняються від нуля тільки значення r_1 і r_2 , а всі наступні значення r_k при $k = 3, 4, \dots$ дорівнюють нулю. Нарешті, для процесів ковзного середнього порядку q відмінні від нуля тільки перші q значень автокореляційної функції. Строячи графіки корелограм для подібних процесів, можна на підставі зазначеної властивості зробити попередній висновок про можливий порядок процесу ковзного середнього, котрий може бути використаний для опису спостереженого ряду.

Зазначене правило гарне, якщо підібраний порядок моделі ковзного середнього невеликий, скажемо від одного до чотирьох-п'яти. Однак на практиці часто зустрічаються стаціонарні процеси з автокореляційною функцією помітно відмінною від нуля навіть при великих затримках. Згідно сформульованому правилу, їх можна намагатися описати процесами ковзного середнього високих порядків. Це приводить до великого числа коефіцієнтів процесу ковзного середнього, котрі підлягають подальшій оцінці. При цьому точність цих оцінок помітно знижується. Практична цінність таких багатопараметричних моделей ковзного середнього невелика. У цій ситуації краще спробувати описати часовий ряд за допомогою моделі авторегресії. Якщо і ці спроби не закінчаться успіхом - перейти до комбінованих моделей авторегресії - ковзного середнього.

Нагадаємо, що найпростіший процес авторегресії першого порядку $X(t)$ з нульовим середнім задається співвідношенням:

$$X(t) = \phi X(t - 1) + \varepsilon_t,$$

де ε_t не залежить від $X(t - 1)$. Члени навіть цього найпростішого процесу не стають незалежними з ростом проміжку часу між ними. Однак за певних умов на коефіцієнти ця залежність швидко убиває.

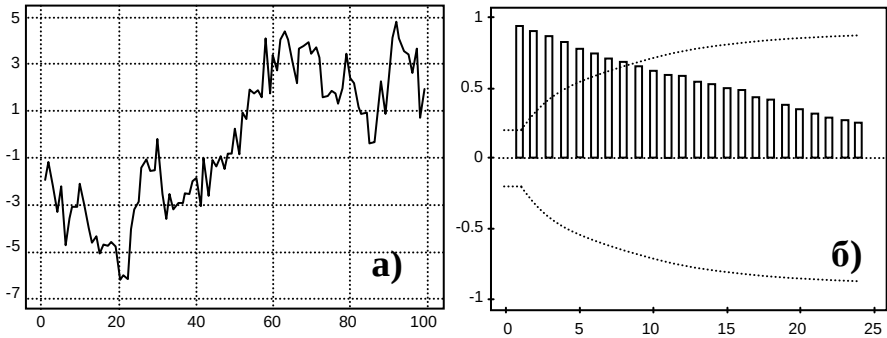
Приведемо два типові графіки поведінки вибірових автокореляційних функцій цих процесів. Автокореляційні функції цих процесів з ростом лага або просто експоненційно загасають, або являють собою експоненційно загасаючі синусоїдальні хвилі.

На рис. 2.11 приведені графіки ста значень реалізації процесу авторегресії другого порядку AR(2):

$$X(t) = \phi_1 X(t - 1) + \phi_2 X(t - 2) + \varepsilon_t,$$

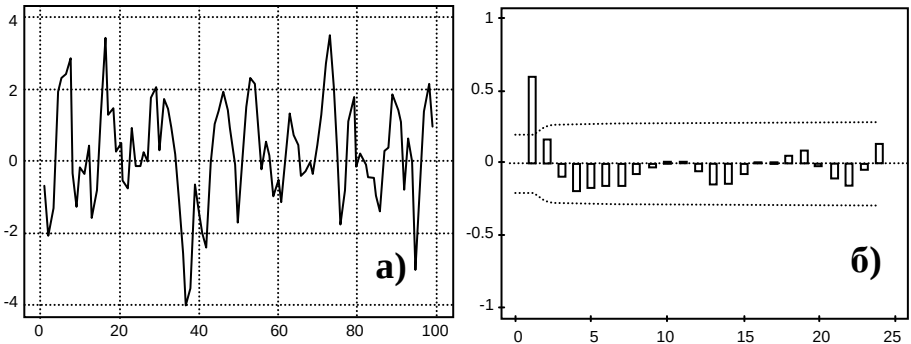
при $\phi_1 = 0.7$ і $\phi_2 = 0.25$. Тут автокореляційна функція процесу і відповідно корелограма експоненційно загасають з ростом лага.

На рис. 2.12 приведені графіки ста значень реалізації AR(2) процесу при $\phi_1 = 0.7$ і $\phi_2 = -0.25$. Автокореляційна функція цього процесу і відповідно корелограма поведуться з ростом лага як експоненційно загасаюча синусоїда.



а) вихідний ряд; б) корелограма

Рисунок 2.11 - Корелограма AR(2) процесу при $\phi_1 = 0.7$, $\phi_2 = 0.25$



а) вихідний ряд; б) корелограма

Рисунок 2.12 - Корелограма AR(2) процесу при $\phi_1 = 0.7$, $\phi_2 = -0.25$

2.3.3 Завдання на виконання роботи

Виконати наступні завдання:

- побудувати часові ряди за даними, що отримані при генерації рядів в лабораторній роботі №1;
- підібрати лінії тренда для кожного ряду;
- порівняти коефіцієнти в підібраних моделях тренда і коефіцієнт Пірсона з SPSS, TableCurve і MathCAD;
- виділити тренди з часових рядів і провести аналіз залишків (випадкової складової часового ряду), побудувавши графіки автокореляційної функції;
- порівняти отримані графіки корелограмми з SPSS і MathCAD.

2.3.4 Хід роботи

2.3.4.1 Підготовка вихідних даних для обробки.

Відкрити в MS Excel файл з розширенням .txt, збережений у лабораторній роботі №1.

У вікні Мастер текстов, що відкрилося:

- крок 1 і 2 – натиснути кнопку Далее;
- крок 3 – натиснути кнопку Подробнее, потім вибрати в полі Разделитель целой и дробной части «точка», натиснути ОК;
- натиснути кнопку Готово.

Змінити формат комірок стовпчика A на числовий.

Додати новий стовпчик A, заповнити його значеннями від 1 до 100, змінити формат даних на числовий з 0 знаками після коми.

Зберегти файл під іншим ім'ям з вказівкою типу Текстовый файл (з роздільником табуляції).

2.3.4.2 Обробка числових даних часового ряду.

1. Запустити програму SPSS.
2. Відкрити раніше збережений текстовий файл. Для цього вибрати в меню File пункт Read Text Data.
3. У вікні, що відкрилося, виконати наступне:
 - кроки з 1 по 3 – натиснути кнопку Далее;
 - крок 4 (рис.2.13) – убрати «пташку» з пункту Comma;
 - крок 5 (рис.2.14) – виділити колонку V1:
 - поле Variable name – t;
 - поле Data format – Numeric;

виділити колонку V2:
 поле Variable name – x;
 поле Data format – Dot;
 крок 6 – натиснути кнопку Далее.

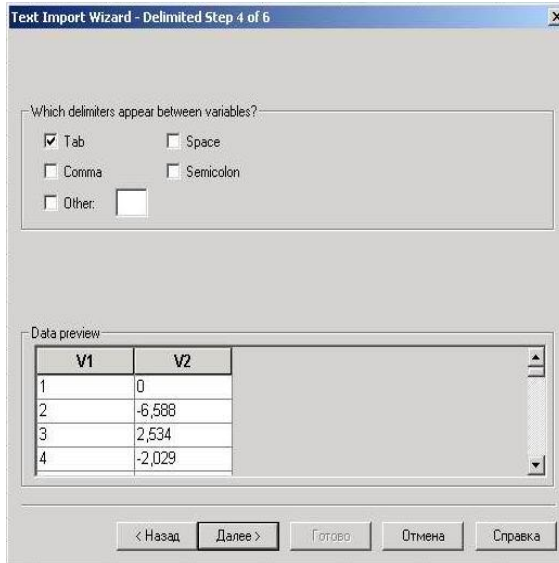


Рисунок 2.13 – Імпорт текстового файлу

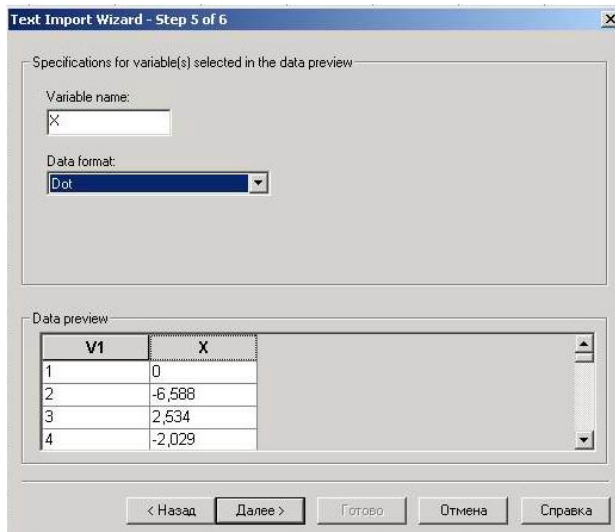


Рисунок 2.14 – Завдання назв змінних

4. Створення графіка часового ряду.

- а) Для цього вибрати в меню Graph→Scatter→Simple→Define. У вікні, що відкрилося, Simple Scatterplot (рис.2.15) додати змінну t на вісь X и змінну x на вісь Y .

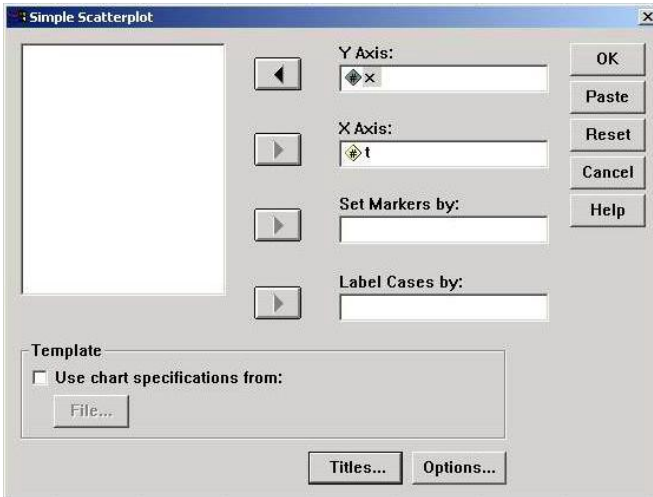


Рисунок 2.15 – Завдання змінних для графіка часового ряду

- б) На створеному графіку двічі клацнути мишею, потім у вікні, що відкрилося, натиснути кнопку з зображенням ламаної кривої на панелі інструментів. У вікні LineInterpolation вибрати Straight, потім натиснути Apply і Close. Закрити вікно Chart1.
5. Виділення складових часового ряду – тренда і залишку.
- а) Вибрати в меню Analyze(Regression(Curve Estimation.
- б) У вікні Curve Estimation, що відкрилося (рис.2.16) додати змінну x у поле Dependent і змінну t у поле Independent. У розділі Models вибрати відповідну модель.

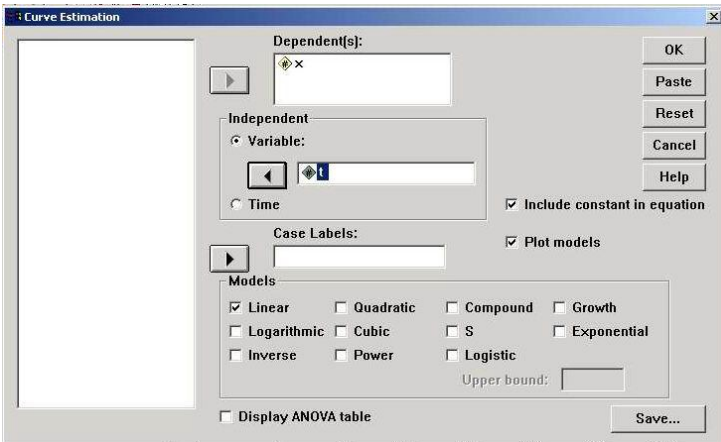


Рисунок 2.16 – Вибір моделі для тренда

- в) Натиснути кнопку Save, поставити «пташку» на Predicted Values i Residuals (рис.2.17). Натиснути Continue, потім OK.

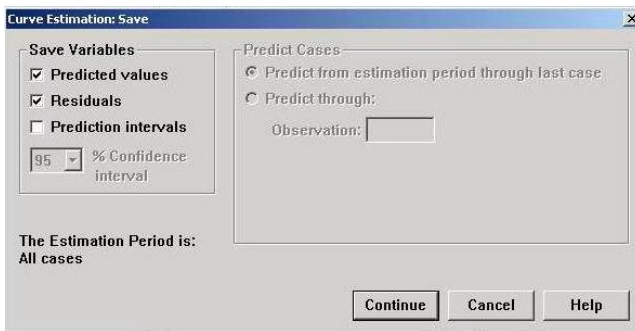


Рисунок 2.17 – Завдання параметрів для тренда

- г) Переглянути отриманий результат на графіку (рис.2.18).

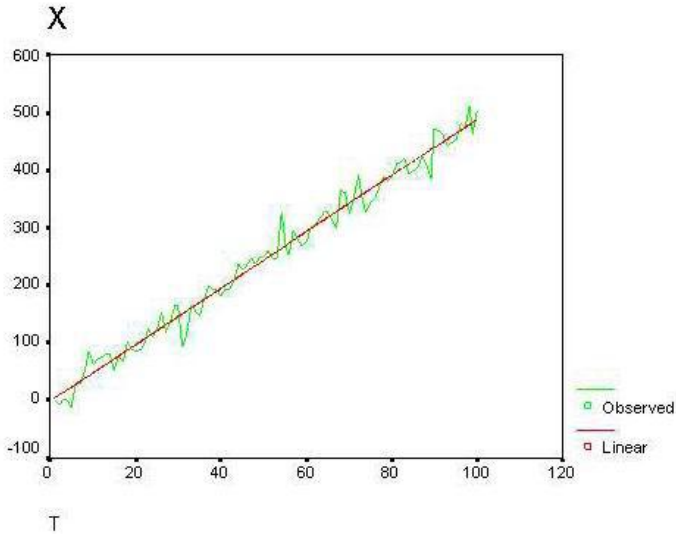


Рисунок 2.18 – Графіки часового ряду і його тренда

6. Створення графіка автокореляційної функції.

- Вибрати в меню Graphs→Time series→Autocorrelations.
- Додати змінну Error for X у поле Variables (рис.2.19). В розділі Display убрати «пташку» на пункті Partial autocorrelations.

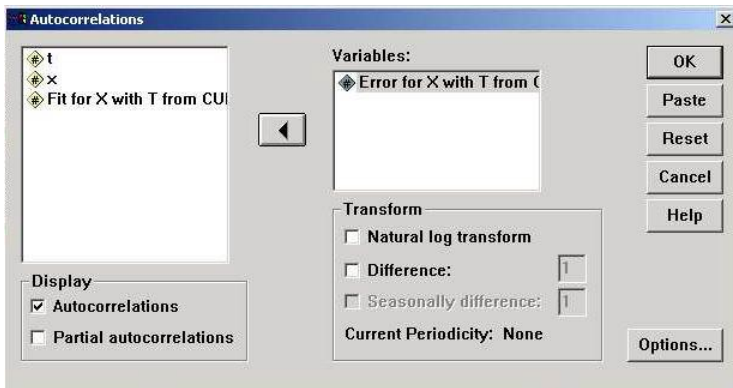


Рисунок 2.19 – Завдання параметрів для автокореляційної функції

- в) Натиснути Options. В полі Maximum Number of Lags установити значення 20. Натиснути Continue, потім OK.
7. Порівняти отриманий графік автокореляційної функції (рис.2.20) з графіком, отриманим у лабораторній роботі №1.

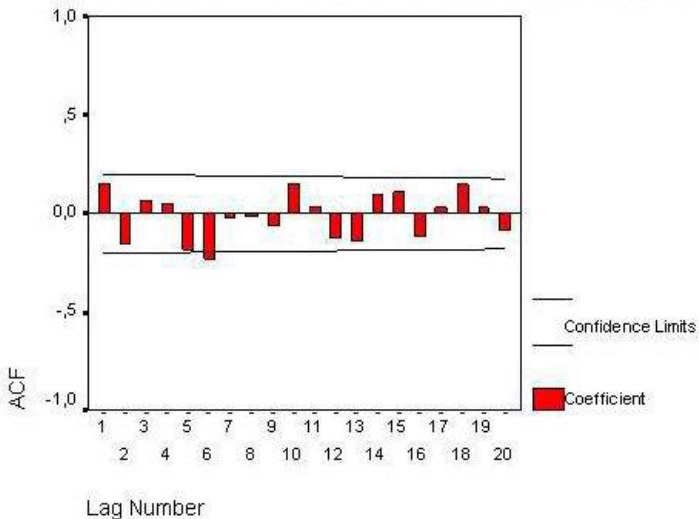


Рисунок 2.20 – Графік автокореляційної функції

2.3.5 Зміст звіту

Звіт включає в себе:

- титульний лист;
- мета роботи;
- графіки часового ряду і підібраного тренда для кожного ряду, що досліджується;
- числові параметри підібраного тренда для кожного часового ряду, що досліджується та результати їхнього порівняння з відповідними параметрами з лабораторних робіт №1 і №2;
- графіки автокореляційної функції для кожного ряду, що досліджується та результати їхнього порівняння з відповідними графіками з лабораторної роботи №1;
- висновки.

2.3.6 Контрольні питання

2.3.6.1 Який ряд називається стаціонарним?

2.3.6.2 Які цілі ставляться при вивченні ряду, отриманого з вихідного часового ряду після виключення детермінованої компоненти?

2.3.6.3 Які особливості корелограми часового ряду, що містить тренд?

2.3.6.4 Які особливості корелограми часового ряду, що містить сезонну компоненту?

2.3.6.5 Які особливості корелограми білого шуму?

2.3.6.6 Які особливості корелограми процесів ковзного середнього?

2.3.6.7 Які особливості корелограми процесів авторегресії?

2.4 Лабораторна робота №4. “ Документування результатів аналізу часових рядів за допомогою спеціалізованого пакета LaTeX”

2.4.1 Мета роботи

Метою роботи є розробка на рівні світових стандартів наукових документів, насичених математичними формулами за допомогою спеціалізованого редактора LaTeX.

2.4.2 Теоретичні відомості

LaTeX – система верстки, що орієнтована на виробництво наукових математичних документів високої типографської якості. LaTeX працює на більшості комп'ютерів, починаючи з IBM PC чи Mac, і закінчуючи великими системами UNIX чи VMS. Дуже багато наукових видань у світі приймають до публікації наукові праці, підготовлені саме в LaTeX. У багатьох університетських мережах система уже встановлена і готова до роботи.

TeX - це комп'ютерна програма, створена Дональдом Кнотом і призначена для верстки тексту і математичних формул. LaTeX – макропакет, що дозволяє авторам верстати і друкувати їхні роботи з високою типографською якістю, за допомогою заздалегідь визначених професійних макетів. Як механізм для верстки він використовує TeX. LaTeX бере на себе роль дизайнера книги, використовуючи TeX у якості верстальника. Але LaTeX – це лише програма і має потребу в чітких інструкціях. Автор повинний надати додаткову інформацію, що описує логічну структуру своєї роботи. Ця інформація записується в текст у виді команд LaTeX.

Це в корені відрізняється від WYSIWYG підходу, прийнятого в більшості сучасних текстових процесорів, наприклад, таких як MS Word. У цих програмах автори форматують документ інтерактивно в процесі набору тексту на комп'ютері. У процесі роботи вони можуть бачити на екрані, як буде виглядати їхня робота, коли вона буде надрукована. При використанні LaTeX звичайно не можна побачити готову картину під час набору тексту, однак її можна побачити на екрані після обробки файлу LaTeX. Потім при необхідності можна внести виправлення.

Основні переваги LaTeX перед звичайними текстовими проце-

сорами:

- готові професійно виконані макети, що роблять документи дійсно, що виглядають «як видані»;
- зручна підтримка верстки математичних формул;
- користувачу потрібно лише вивчити кілька зрозумілих команд, що задають логічну структуру документа, йому практично ніколи не потрібно возитися з макетом документа;
- легко виготовляються навіть складні структури, типу приміток, змістів, бібліографій та інше;
- для рішення багатьох типографських задач, що не підтримуються прямо базовим LaTeX, є вільно розповсюджувані додаткові пакети;
- LaTeX заохочує авторів писати добре структуровані документи, тому що саме так LaTeX і працює – шляхом специфікації структури;
- TeX, форматуючий засіб LaTeX надзвичайно мобільний і вільно доступний, тому система працює практично на всіх існуючих платформах.

LaTeX має і деякі недоліки:

- хоча визначені макети мають безліч параметрів, що можна налаштовувати, створення цілком нового макету документа не дуже просто і займає багато часу;
- дуже складно писати неструктуровані і неорганізовані документи.

Вихідними даними для LaTeX є звичайний текстовий файл (з розширенням .tex) у ASCII, який можна створити в будь-якому текстовому редакторі. Він містить текст документа разом з командами, що вказують LaTeX як верстати текст.

Для зручності написання вихідного файлу, його компіляції і перегляду результату (файл із розширенням .div) необхідно запустити оболонку LaTeX, зайшовши в директорію TexShell і запустивши файл texshell.exe. Потім вибрати в меню File і створити новий чи відкрити вже існуючий файл відповідною командою. Звичайно усі файли, з якими працює LaTeX зберігаються в директорії Tex_files, там вже знаходяться готові файли, якими можна скористатися як прикладами. Після написання тексту документа його необхідно зберегти, а потім вибрати цей файл у якості того, що компілюють командою File>Main file. Зовнішній вигляд оболонки LaTeX у режимі редагування представле-

ний на рис.2.21.

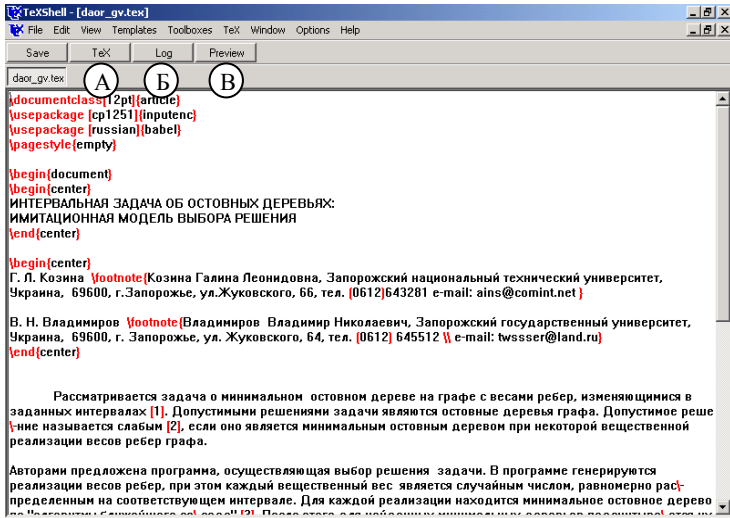


Рисунок 2.21 – Оболонка LaTeX в режимі редагування

Для компіляції файлу необхідно натиснути кнопку TeX (кнопка A на рис.2.21), після чого вихідний файл буде відкомпільований. Якщо в процесі компіляції не з'явилися повідомлення про помилки, то вона пройшла успішно. Для перегляду результату необхідно натиснути кнопку Preview (кнопка B на рис.2.21), після чого на екран буде виведений файл результату (рис.2.22).

Якщо в процесі компіляції були видані повідомлення про помилки, необхідно повернутися до вихідного .tex файлу і внести виправлення. Для перегляду повідомлень компілятора з діагностикою помилок можна скористатися спеціальним файлом з розширенням .log, зміст якого виводиться на екран при натисканні кнопки Log (кнопка B на рис.2.21).

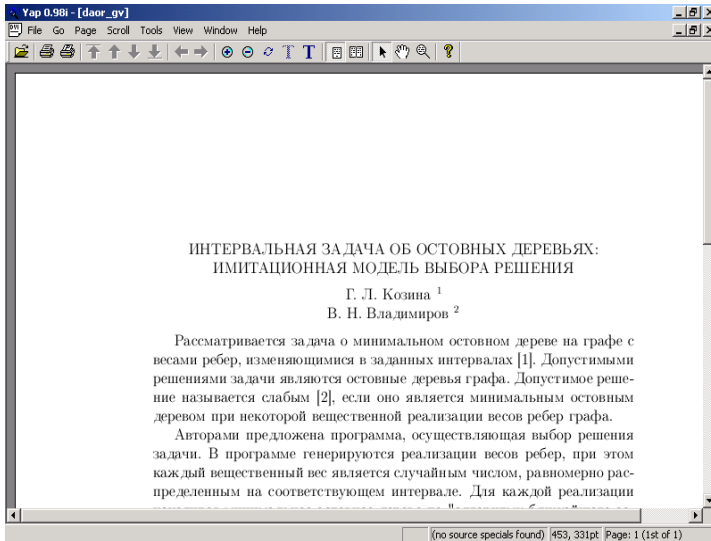


Рисунок 2.22 – Оболонка LaTeX в режимі перегляду результату

2.4.3 Завдання на виконання роботи

Створити звіт про проведення у лабораторних роботах №1-3 аналіз часових рядів. Звіт повинний містити:

- а) Титульний лист.
- б) Мету роботи.
- в) Результати, що отримані при генеруванні часових рядів у пакеті MathCAD:
 - 1) математичні формули, що описують тренд і випадкову складову часових рядів, що аналізуються;
 - 2) графіки трендів, випадкових складових і часових рядів;
 - 3) математичні формули, що описують числові характеристики часових рядів, що аналізуються та їхні значення;
 - 4) графіки автокореляційних функцій випадкових складових часових рядів, що аналізуються.
- г) Результати, отримані при аналізі часових рядів у пакеті Table Curve:
 - 1) математичні формули та параметри, що описують підібрані тренди для часових рядів, що аналізуються;

- 2) графіки трендів і часових рядів.
- д) Результати, отримані при аналізі часових рядів у пакеті SPSS:
 - 1) математичні формули, що описують підібрані тренди для часових рядів, що аналізуються;
 - 2) графіки трендів і часових рядів;
 - 3) графіки автокореляційних функцій випадкових складових часових рядів, що аналізуються.
- е) Висновки.

2.4.4 Контрольні питання

2.4.4.1 У чому відмінність редактора LaTeX від звичайних текстових редакторів?

2.4.4.2 Які основні достоїнства редактора LaTeX?

2.4.4.3 Які основні недоліки редактора LaTeX?

2.4.4.4 З яких функціональних частин складається редактор LaTeX?

2.4.4.5 Які формати файлів використовує LaTeX і яке їхнє призначення?

ЛИТЕРАТУРА

- 1 Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере / Под ред. В.Э. Фигурнова. - М.: ИНФА-М, 2003.- 544 с.: ил.
- 2 Рыжиков Ю.И. Решение научно-технических задач на персональном компьютере.- СПб: КОРОНА принт, 2000.- 270с.
- 3 Гайдышев И. Анализ и обработка данных: Специальный справочник.- СПб: Питер, 2001.- 752 с.
- 4 Бююль А. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей/ А.Бююль, П.Цефель; Под ред. В.Е.Момота.- СПб.: ДиаСофтИП, 2002.- 608с.
- 5 Кнут, Дональд, Э. Все про TEX: Перевод с англ. – М.: Издательский дом «Вильямс», 2003. – 560 с.: ил.
- 6 Тутубалин В.Н. Теория вероятностей и случайных процессов. - М.: Изд-во МГУ, 1992. - 400 с.
- 7 Вучков И., Бояджиева Л., Солаков Е. Прикладной линейный регрессионный анализ. - М.: Финансы и статистика, 1987. - 239с.
- 8 Кирьянов Д.В. Самоучитель MathCad 2001.- СПб: БХВ-Петербург, 2002.- 544с.