МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БУДІВНИЦТВА І АРХІТЕКТУРИ

*С.В. Цюцюра, О.А. Поплавський*

# СПЕЦКУРС ЗА НАУКОВОЮ СПЕЦІАЛЬНІСТЮ

Методичні вказівки до виконання практичних робіт

Київ, 2020

**ЗМІСТ**

1 Стислі відомості про систему Statistica V6.0.....……..... 6
2 Практична робота 1. Знаходження середнього значення змінної, середнього квадратичного відхилення змінної й області прогнозів для даної змінної. Побудування рівняння прямої регресії $y = b_0 + b_1 x$. Прогноз за моделлю..... 12
3 Практична робота 2. Вибір моделі однофакторної регресії. ……. 20
4 Практична робота 3. Перевірка однофакторної лінійної регресії на адекватність.....……..... 29
5 Практична робота 4. Прогноз на підставі лінійної регресії. Точність прогнозу.....……..... 37
6 Практична робота 5. Перевірка факторів на мультиколінеарність. Вибір моделі багатофакторної регресії..... 48
7 Практична робота 6. Аналіз часових рядів.....…..... 57
Література.....……..... 83

# 1 СТИСЛІ ВІДОМОСТІ ПРО СИСТЕМУ STATISTICA V6.0

## 1.1 Структура пакету STATISTICA V6.0

Програма STATISTICA містить декілька незалежно працюючих модулів, що відкриваються за допомогою пункту меню **Statistics** (рис. 1). У кожному модулі зібрані логічно пов'язані між собою статистичні процедури. Завантажити можна відразу кілька модулів. Кнопки цих модулів знаходяться у нижній частині екрана. Переходити між ними можна стандартним чином, клацнувши лівою клавішею миші по відповідній кнопці.

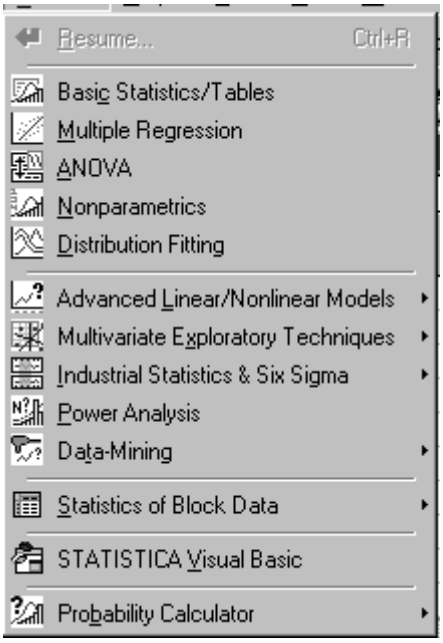


*Рисунок 1*

## 1.2 Створення нової таблиці даних

Для цього треба вибрати пункт меню **File – New – Ok.** Відкриється порожня електронна таблиця розміром 10x10 (рис. 2). У стовпчиках розташовані змінні (Vars), в рядках – випадки ( Cases).

### 1.3 Вилучення і додання нових змінних і випадків

Виконуються командами Delete (вилучити) і Add (додати). Після виділення рядка або стовпця натиснути кнопку Vars, якщо вилучаються або додаються змінні, і вказати, скільки елементів вилучається чи додається, Ok. Для випадків – аналогічно, але з кнопкою Cases.

| | 1<br>Var1 | 2<br>Var2 | 3<br>Var3 | 4<br>Var4 | 5<br>Var5 | 6<br>Var6 | 7<br>Var7 | 8<br>Var8 | 9<br>Var9 | 10<br>Var10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | | | |
| 2 | | | | | | | | | | |
| 3 | | | | | | | | | | |
| 4 | | | | | | | | | | |
| 5 | | | | | | | | | | |
| 6 | | | | | | | | | | |
| 7 | | | | | | | | | | |
| 8 | | | | | | | | | | |
| 9 | | | | | | | | | | |
| 10 | | | | | | | | | | |

*Рисунок 2*

### 1.4 Коректування таблиці

Коректування таблиці зводиться до коректування назв змінних, вмісту стовпців цілком і окремих кліток. Для цього робиться щиглик по імені стовпця, щиглик – по кнопці Vars (змінні), за пунктом меню **Current Specs** (поточні специфікації). Після цього можна задавати нове ім'я стовпцю в рядку Name (ім'я) і нові значення випадків за допомогою формули у віконці Long name (повне ім'я) (рис. 3). Також можна змінити подання числа (автоматично – 8 позицій (Column width), число позицій після коми (Decimals) – 3).

Коректування числових значень в окремих осередках виконується, як в EXCEL.

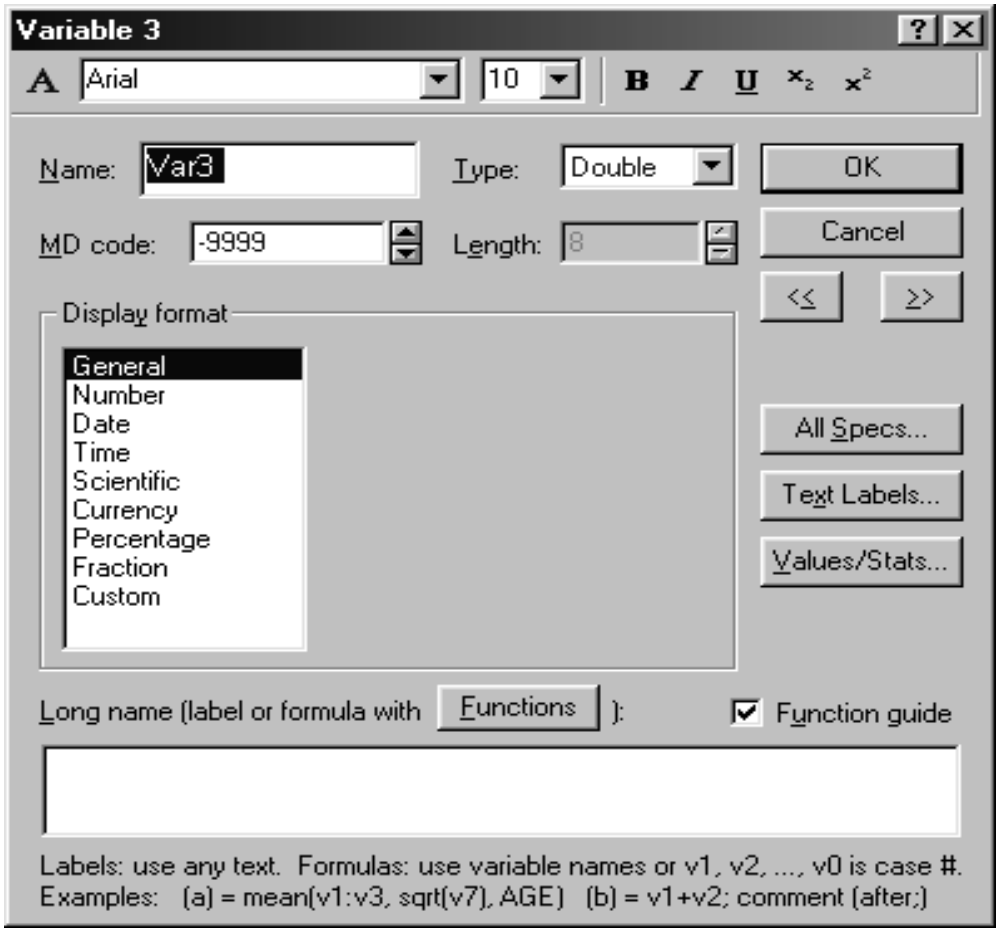


*Рисунок 3*

**1.5 Обчислення статистичних характеристик для значень змінних** (наприклад, максимальне, мінімальне, середнє значення, дисперсія і т.ін.).

Активізувати таблицю даних, потім активізувати пункти меню **Statistics – Basic Statistics/Tables – Descriptive Statistics** – вкладка **Advanced** (аналіз – описові статистики – усі статистики) (рис. 4), виділити змінні, для яких шукають характеристики (кнопка Variables), Ok, вибрати зі списку потрібні статистичні характеристики (наприклад, Min – мінімальне значення, Max – максимальне значення, Valid N – обсяг вибірки, Mean – середнє значення, Standard Deviation – середнє квадратичне відхилення, Variance – дисперсія), кнопка Summary. На екрані з'явиться таблиця з потрібними характеристиками.

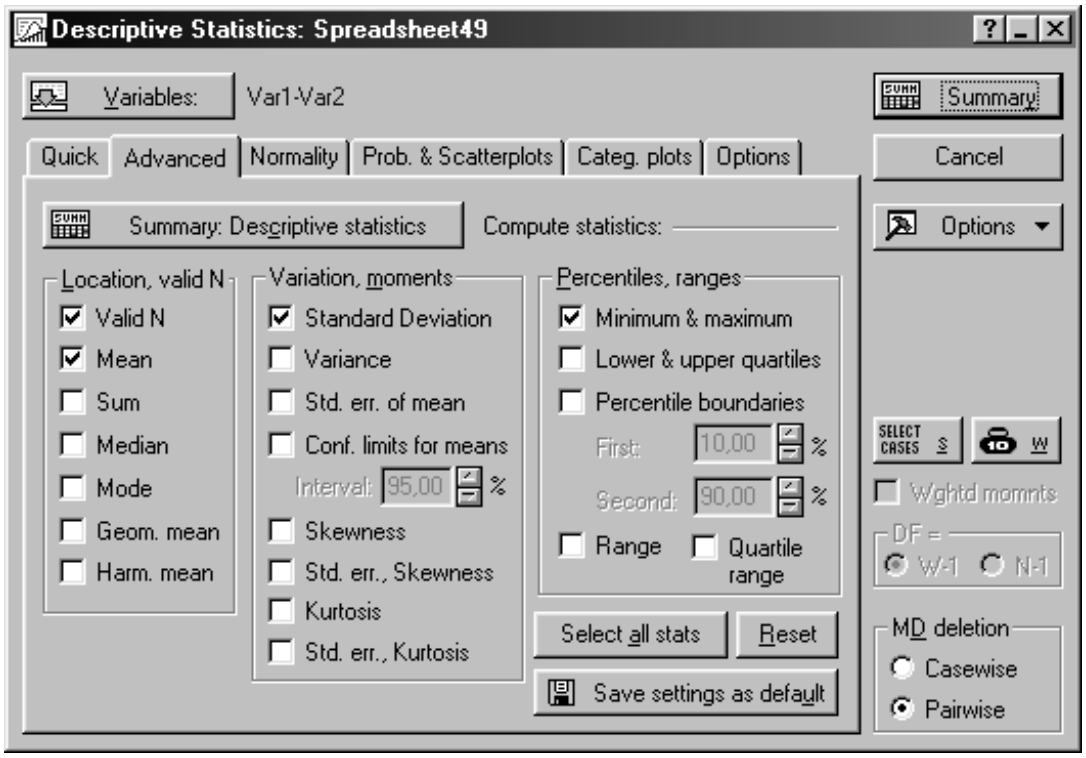


*Рисунок 4*

### 1.6 Одержання графіка і рівняння лінійної регресії

Активізувати таблицю. Вибрати пункт меню **Craphs – 2D Craphs – Scatterplots** – вкладка **Advanced**, (графіки, статистичні двомірні графіки, точковий графік), вибрати змінні Variables (для аргументу – *x* і функції – *y*), Ok, вибрати опції Regular, Linear, (регулярний, лінійний), Ok (рис. 5).

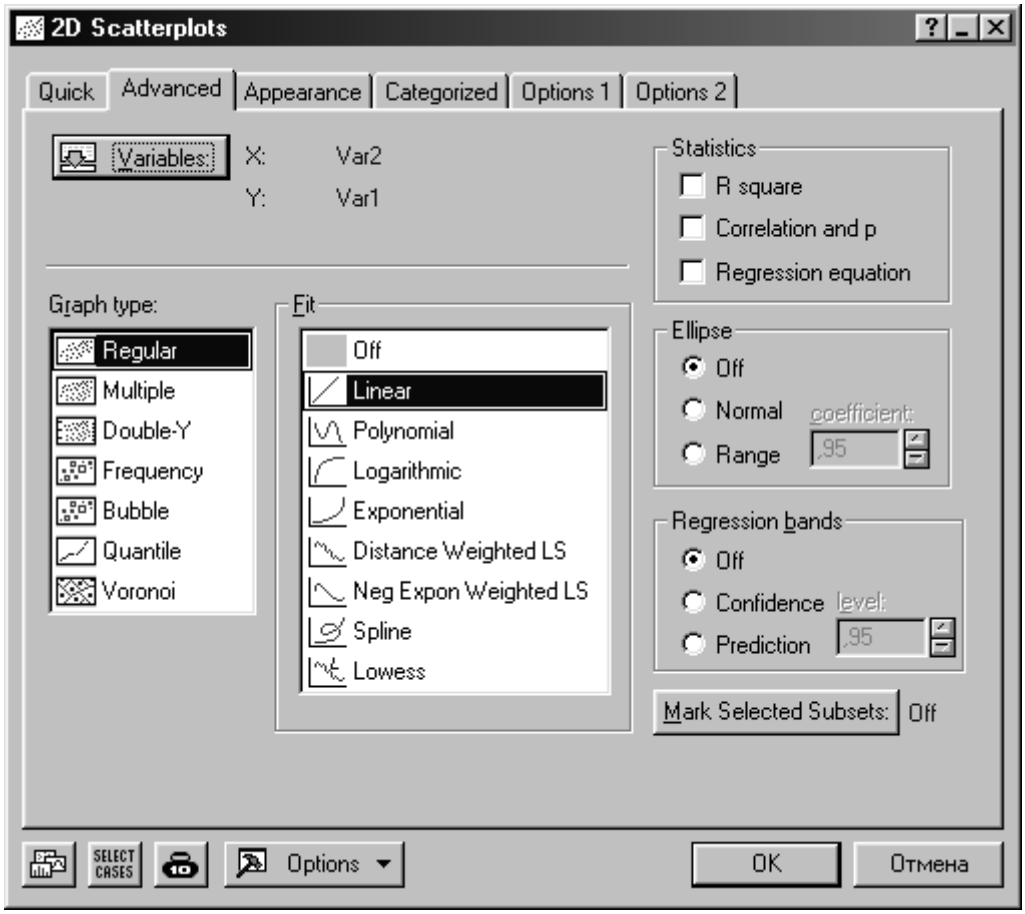


*Рисунок 5*

З'явиться графік лінійної регресії, над яким записане рівняння регресії (рис. 6).

### 1.7 Типи файлів у системі Statistica

Типи файлів:

*.sta – початкові дані;

*.stw – результати обробки даних, Workbook;

*.str, або *.rtf– звіт.

Довжина імені файла, як і будь-якого іншого ідентифікатора, не більше 8 символів.

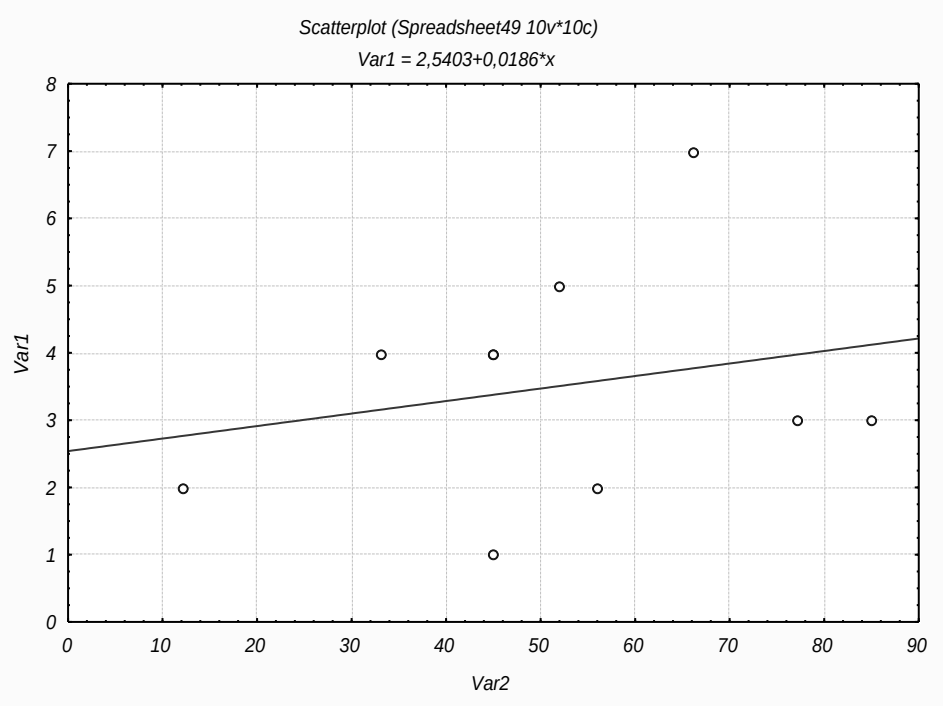


*Рисунок 6*

### 1.8 Створення автозвіту

Автозвіт  бажано створювати при кожному сеансі роботи з пакетом для того, щоб усі результати роботи (таблиці і графіки) запам’ятовувалися в автозвіті.

Далі визначений шлях для створення стислого автозвіту: File, Output Manager. З'явиться багатосторінкове меню.

На сторінці Output Manager треба відмітити опції: Single Workbook, Place results in Workbook automatically, Also send to Report Window, Single Report (рис. 7). На сторінці Workbook, крім вже заданих опцій, у полі Add to Workbook performs відмітити опцію Copy. На сторінці Report, крім вже заданих опцій, у полі Add to Report performs відмітити опцію Copy. Після цього натиснути Ok. Тепер усі розрахункові таблиці і графіки автоматично заносяться в звіт.

Файл звіту треба зберегти з розширенням *.rtf. Тепер його можна редагувати в WORD.

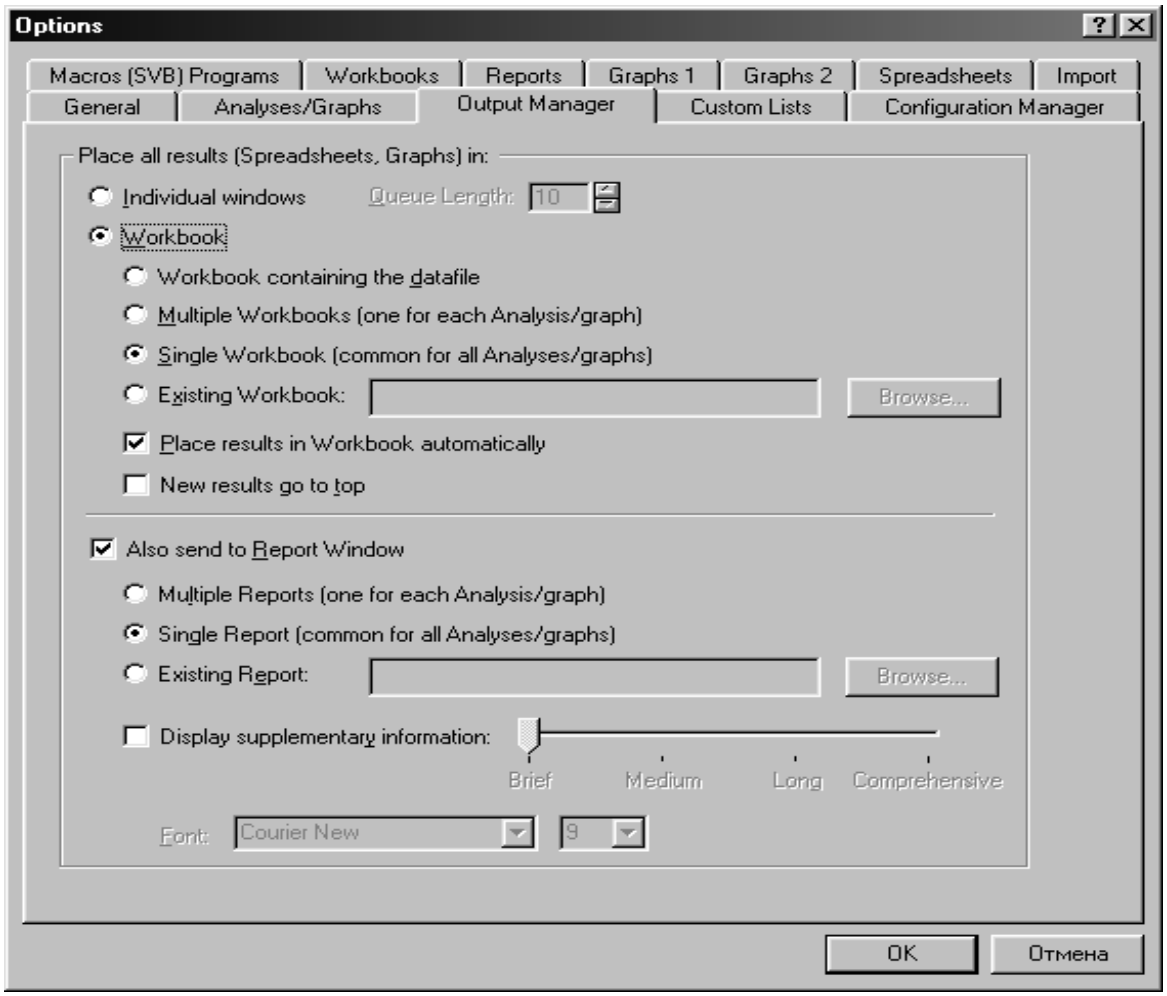


*Рисунок 7*

## ПРАКТИЧНА РОБОТА №1

**Тема**: Знаходження середнього значення змінної, середнього квадратичного значення змінної й області прогнозів для даної змінної. Знаходження рівняння прямої регресії $y = b_0 + b_1 x$. Прогноз за моделлю.

### 2.1 Стислі теоретичні відомості

*Вибірка* – сукупність випадково відібраних даних ($x_i$,$y_i$) (табл. 1), де n – обсяг вибірки; $x$ – фактор; $y$ – відклик.

*Таблиця 1*

| | | | |
|---|---|---|---|
| $x_1$ | $x_2$ | . . . | $x_n$ |
| $y_1$ | $y_2$ | . . . | $y_n$ |

*Кореляційне поле (діаграма розсіювання)* – графічне зображення точок вибірки (рис. 8).

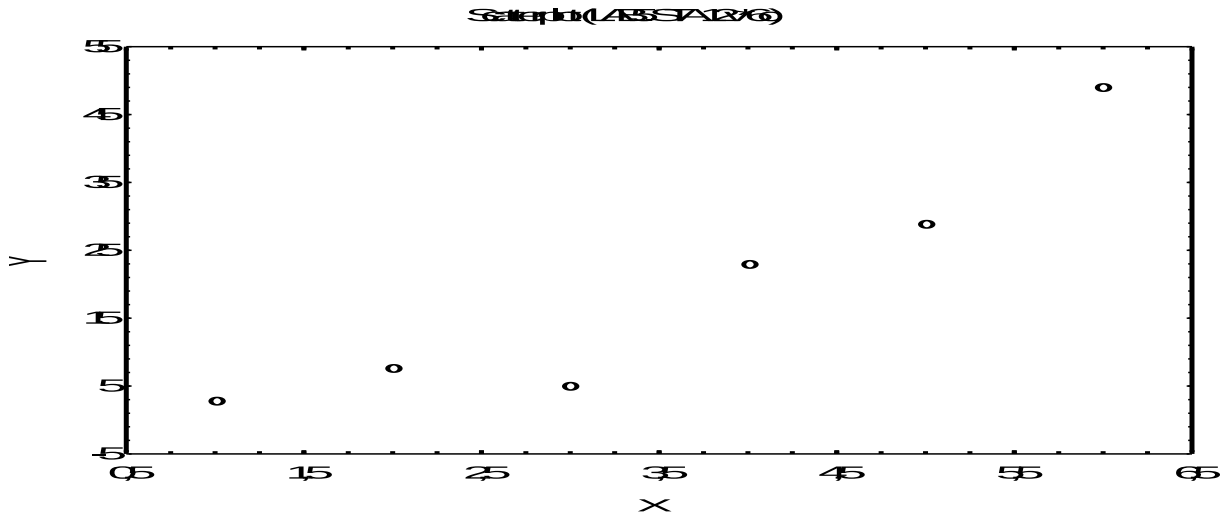


*Рисунок 8*

*Генеральна сукупність* – сукупність об'єктів, з яких беруть вибірку.

*Математична модель* – це наближений опис якого-небудь явища за допомогою математичної символіки. У найпростішому випадку однофакторної регресії математична модель – це формула виду y = F(x). Якщо модель лінійна, то $y = b_0 + b_1 x$ – рівняння лінійної регресії.

*Середні значення* фактора *x* і відклику *y* обчислюються за формулами

$$x_{cp} = \frac{1}{n}\sum_{i=1}^{n} x_i \;\; ; \quad y_{cp} = \frac{1}{n}\sum_{i=1}^{n} y_i \; .$$

Точка ($x_{ср}$,$y_{ср}$) називається *центром розсіювання.* Графік лінійної регресії завжди проходить через центр розсіювання.

*Середнє квадратичне відхилення* фактора обчислюється за формулою $\sigma_x = \sqrt{\frac{1}{n}\Sigma\left(x_i - x_{cp}\right)^2}$ і характеризує, наскільки в середньому значення

фактора $x_i$ відхиляються від $x_{ср}$. З двох вибірок з однієї генеральної сукупності більш якісною є та, де $\sigma_x$ більше.

*Область прогнозів* розташована між мінімальним і максимальним значеннями фактору $x$. Прогноз відклику $y$ робиться за рівнянням моделі.

### 2.2 Мета практичної роботи

Мають бути придбані наступні вміння:

1) створення й коректування таблиці даних;
2) створення автозвіту й робота з ним;
3) знаходження графіка й рівняння лінійної регресії й прогнозів за ним.

Мають бути засвоєні наступні поняття: модель, середнє значення, середньоквадратичне відхилення, кореляційне поле, область прогнозу, прогноз.

Робота розрахована на 4 години.

### 2.3 Завдання до практичної роботи

1) Створити таблицю даних.
2) Створити автозвіт.
3) Внести в автозвіт створену таблицю даних.
4) Знайти середнє значення, середньоквадратичне відхилення й область прогнозів для фактора $x$.
5) Знайти графік і рівняння прямої регресії.
6) Знайти прогноз $y$ у точці $x_{ср}$ і в будь-якій довільній точці з області прогнозів.

### 2.4 Зміст звіту

Звіт про практичну роботу повинен містити:

1) Тема роботи, завдання.
2) Роздрук таблиць і графіків.
3) Пояснення отриманих таблиць і графіків з погляду економетрії.

## 2.5 Приклад виконання практичної роботи в пакеті Statistica6

*Економічні дані*

Реальний обсяг випуску продукції (Y, млн т) і рівні факторів, її формувальних – капітальних витрат (X1, млн грн) і питомої ваги простоїв устаткування (X2, %) по металургійних підприємствах країни за минулий рік задані в таблиці 2.

*Таблиця 2*

| № | X1 | X2 | Y |
|---|---|---|---|
| 1 | 1,033 | 1,45 | 1,83 |
| 2 | 0,012 | 4,295 | 0,58 |
| 3 | 0,045 | 3,553 | 1,34 |
| 4 | 0,243 | 1,568 | 1,34 |
| 5 | 0,266 | 1,52 | 1,64 |
| 6 | 0,302 | 0,512 | 1,65 |
| 7 | 0,451 | 0,457 | 1,91 |
| 8 | 1,041 | 1,822 | 1,96 |
| 9 | 1,423 | 0,442 | 2,08 |
| 10 | 1,914 | 0,498 | 2,18 |

Знайти залежність між показником $y$ (обсяг випуску продукції) і фактором $x_1$ (капітальні витрати) виду $y = b_0 + b_1 x$.

*Виконання завдання*

1) Створюємо таблицю даних. Таблиця буде мати дві змінні ($x$ і $y$) і десять випадків. Назва таблиці – lab1.sta.

2) Створюємо автозвіт, як розглянуто вище.

3) Внесемо в автозвіт створену таблицю даних (кнопка Add to Report на панелі інструментів) (рис. 9).

| | 1<br>X | 2<br>Y |
|---|---|---|
| 1 | 1,033 | 1,83 |
| 2 | 0,012 | 0,58 |
| 3 | 0,045 | 1,34 |
| 4 | 0,243 | 1,34 |
| 5 | 0,266 | 1,64 |
| 6 | 0,302 | 1,65 |
| 7 | 0,451 | 1,91 |
| 8 | 1,041 | 1,96 |
| 9 | 1,423 | 2,08 |
| 10 | 1,914 | 2,18 |

*Рисунок 9*

4) Знайдемо середнє значення $x_{ср}$, середньоквадратичне відхилення й область прогнозів для фактора *x*.

Активувати таблицю з даними. Далі: **Statistics – Basic Statistics/Tables – Descriptive statistics** (описові статистики) – **OK** (рис. 10).

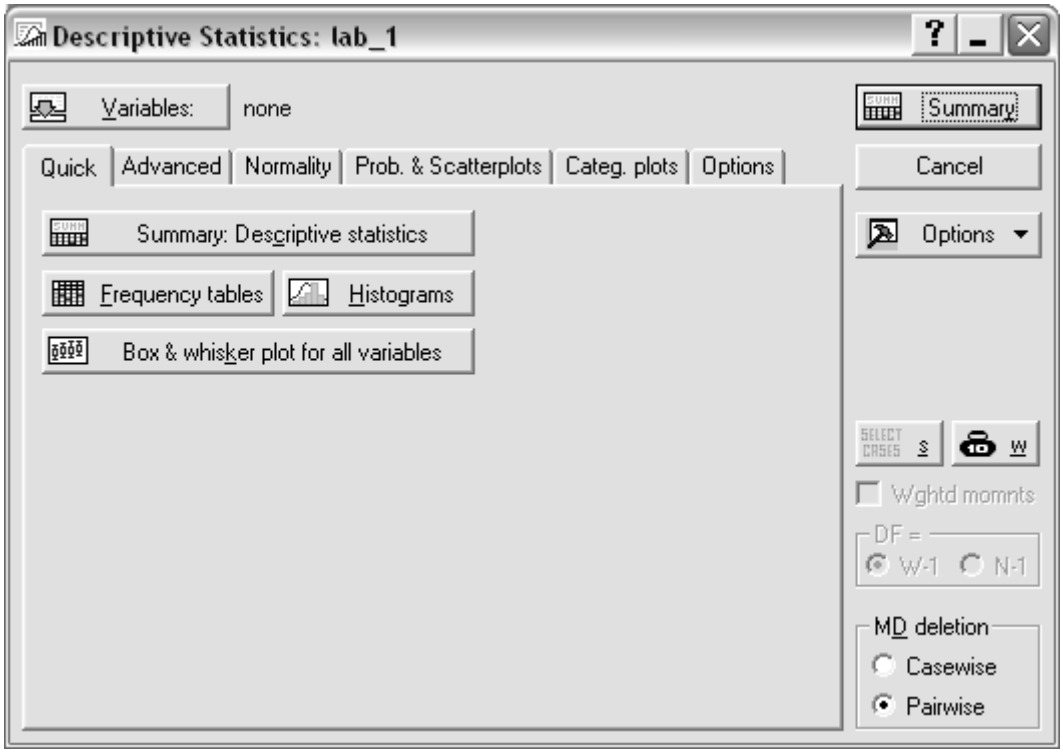


*Рисунок 10*

Далі: Variables – виділити потрібну змінну (у цьому випадку $x$) – OK.

На вкладці Advanced виділяємо (рис. 11): Mean (середнє значення), Standard Deviation (середнє квадратичне відхилення), Minimum і Maximum (мінімальне і максимальне значення) – Summary.

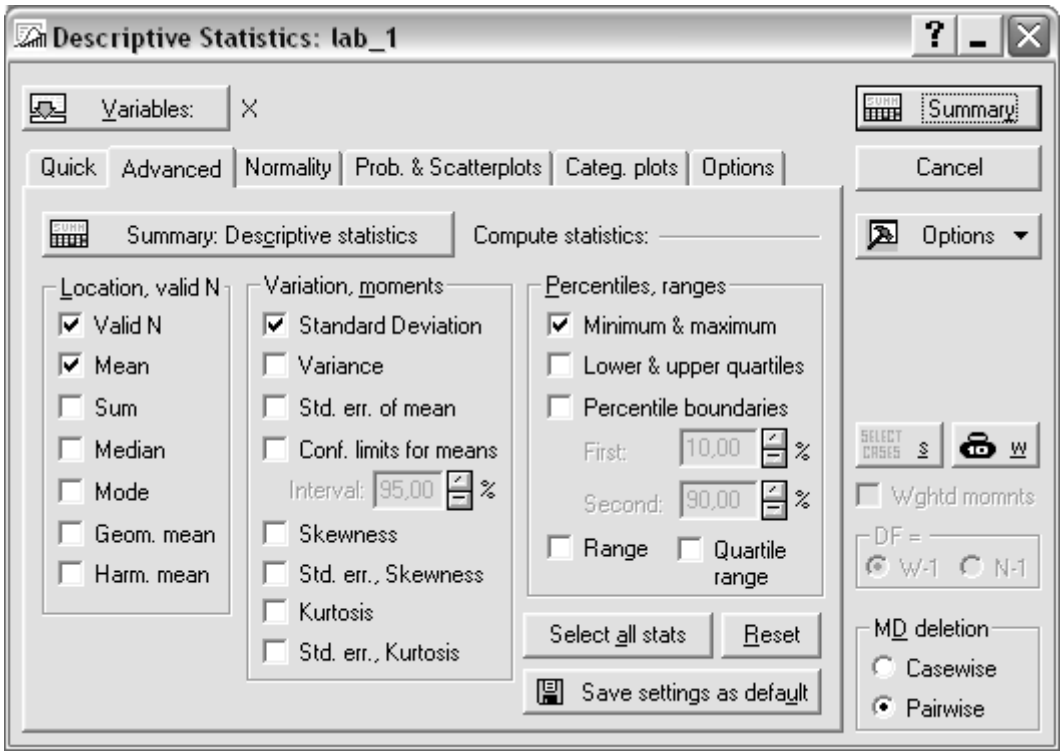


*Рисунок 11*

Потрібна таблиця автоматично вставляється в автозвіт (рис. 12).

| Variable | Descriptive Statistics (Исх_данные) | | | | |
|---|---|---|---|---|---|
| | Valid N | Mean | Minimum | Maximum | Std.Dev. |
| x | 10 | 0,67300 | 0,01200 | 1,91400 | 0,64432 |

*Рисунок 12*

5) Знайдемо графік і рівняння прямої регресії $y = b_0 + b_1 x$.

Активувати таблицю з даними: **Graphs – Scatterplots** (графіки – точкові графіки) – **Variables** (виділити аргумент $x$ і функцію $y$) – **OK**. Далі на вкладці Advanced вибрати опції **Regular, Linear fit, Regression bands – Off – OK**, як це показано на рис. 13.

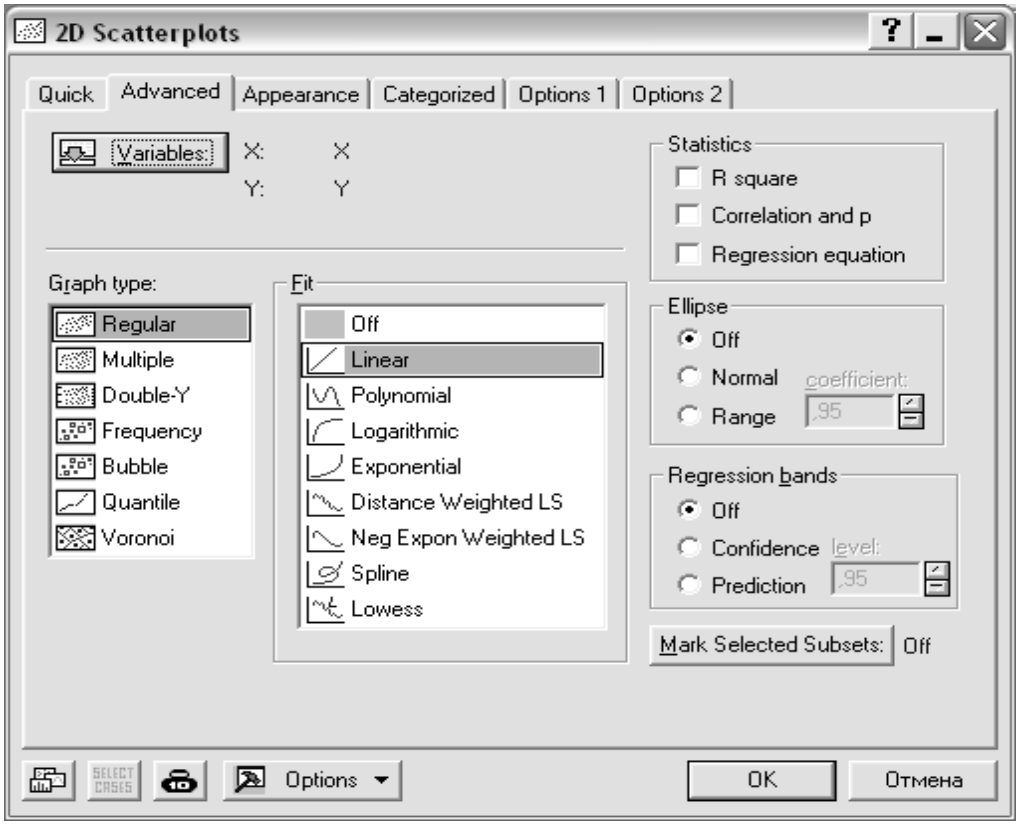


*Рисунок 13*

Над графіком рівняння прямої регресії y=1,265+0,573x (рис. 14).

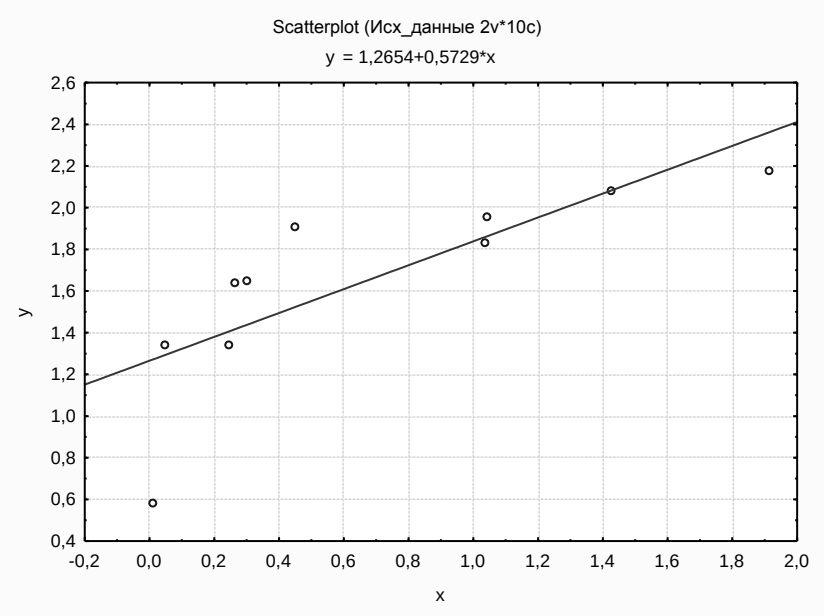


*Рисунок 14*

6) Знайдемо прогноз $y$ у точці $x_{ср}$ і в будь-якій довільній точці з області прогнозів.

Щоб зробити прогноз у точці $x_{ср}$=0,673 і, наприклад, у точці 1,5, потрібно виконати наступні дії:

а) додати в таблиці два рядки, у яких у стовпці X додати числа 0,673 і 1,5;

б) додати в таблиці 3-й стовпець Y_REGR. Подвійним щигликом по імені стовпця Y_REGR увійти у вікно редагування стовпця. У вікні Long Name (унизу екрана) вписати формулу =1,265+0,573*x (рівняння регресії), і нажати ОК. У стовпці Y_REGR з'являться значення *y*, розраховані за рівнянням прямої регресії y=1,265+0,573x для всіх *x*, перерахованих в 1-му стовпці, у тому числі й для значень 0,673 і 1,5 (рис. 15);

в) нова таблиця поміститься у звіт.

| | 1<br>x | 2<br>y | 3<br>Y_REGR |
|---|---|---|---|
| 1 | 1,033 | 1,83 | 1,856909 |
| 2 | 0,012 | 0,58 | 1,271876 |
| 3 | 0,045 | 1,34 | 1,290785 |
| 4 | 0,243 | 1,34 | 1,404239 |
| 5 | 0,266 | 1,64 | 1,417418 |
| 6 | 0,302 | 1,65 | 1,438046 |
| 7 | 0,451 | 1,91 | 1,523423 |
| 8 | 1,041 | 1,96 | 1,861493 |
| 9 | 1,423 | 2,08 | 2,080379 |
| 10 | 1,914 | 2,18 | 2,361722 |
| 11 | 0,673 | | 1,650629 |
| 12 | 1,5 | | 2,1245 |

*Рисунок 15*

### 2.6 Висновки

Середнє значення $x_{ср}$ (Mean), середнє квадратичне відхилення (Std.Dev.) і область прогнозів (Minimum- Maximum) для фактора *x* наведені на рисунку 16.

| | Descriptive Statistics (Исх_данные) | | | | |
|---|---|---|---|---|---|
| Variable | Valid N | Mean | Minimum | Maximum | Std.Dev. |
| x | 10 | 0,673000 | 0,012000 | 1,914000 | 0,644324 |

*Рисунок 16*

Область прогнозів 0,012000<= X <= 1,914000 задає діапазон, з якого припустимо вибирати значення фактора *x*, капітальні витрати (мільйонів гривень) для прогнозу обсягу випуску продукції *y* (мільйонів тонн).

Середнє значення $x_{ср}$=0,673 задає центр області прогнозів.

Середнє квадратичне відхилення 0,6443 характеризує середнє значення розсіювання значень капітальних витрат по металургійних підприємствах країни відносно $x_{ср}$.

Рівняння прямої регресії: Y=1,265+1,573X. За цим рівнянням розраховується прогноз обсягу випуску продукції *y* в залежності від капітальних витрат *x*. При капітальних витратах X=0,673 млн грн. обсяг випуску продукції буде дорівнювати Y= 1,651 млн т. При капітальних витратах X=1,500 млн грн. обсяг випуску продукції Y= 2,125 млн т.

## ПРАКТИЧНА РОБОТА №2

**Тема**: Вибір моделі однофакторної регресії.

### 3.1 Стислі теоретичні відомості

Рівняння лінійної регресії $y=b_0+b_1x$ знаходять за методом найменших квадратів. Відхилення і-ї точки кореляційного полю від лінії регресії дорівнює ($y_{лін} - y_i$). *Метод найменших квадратів* полягає в тому, щоб мінімізувати суму квадратів відхилень (залишків):

$$s = \Sigma( y_{лін} - y_i )^2 .= \Sigma(( b_0+b_1x_i-y_i)^2 \ (\min).$$

Мінімум досягається за умови рівності нулю часткових похідних:

$$\begin{cases} \partial s / \partial b_0 = 0, \\ \partial s / \partial b_1 = 0. \end{cases}$$

За цією системою рівнянь знаходять коефіцієнти регресії $b_0$ і $b_1$.

Для нелінійної моделі суму квадратів відхилень знаходять аналогічно:

$$s = \Sigma((\ y_{нелін} - y_i)^2.$$

З двох моделей оптимальною є та, у якої сума квадратів відхилень менше.

### 3.2 Мета практичної роботи

Мають бути придбані наступні вміння:

1) знаходження графіка й рівняння нелінійної регресії й прогнозів за ним;

2) вибір оптимальної моделі.

Мають бути засвоєні наступні поняття: оптимальна модель, залишки, метод найменших квадратів.

Робота розрахована на 4 години.

### 3.3 Завдання до практичної роботи

Використовуючи дані з практичної роботи 1, за мінімумом суми квадратів залишків вибрати оптимальну модель із двох моделей: лінійної й експоненційної $y = Ae^{bx}$. Для цього необхідно:

1) Знайти графік і рівняння прямої регресії.
2) Знайти прогноз *y* у кожній точці *x* за лінійною моделлю.
3) Знайти графік і рівняння експоненціальної регресії.
4) Знайти прогноз *y* у кожній точці *x* за нелінійною моделлю.
5) Знайти розбіжність лінійного й експоненціального прогнозів у відсотках.
6) Знайти квадрати відхилень для нелінійної моделі.
7) Знайти суму квадратів залишків для лінійної моделі.
8) Знайти суму квадратів залишків для лінійної моделі.
9) Вибрати оптимальну модель

### 3.4 Зміст звіту

Звіт про практичну роботу повинен містити:

1) тему роботи, завдання;
2) роздрук таблиць і графіків;
3) статистичні й економічні висновки за результатами роботи.

### 3.5 Приклад виконання практичної роботи у пакеті Statistica6

*Економічні дані*

Економічні дані взяти з практичної роботи 1.

Вихідна таблиця даних (рис. 17) вставляється у звіт, як у практичній роботі 1.

| | 1<br>X | 2<br>Y |
|---|---|---|
| 1 | 1,033 | 1,83 |
| 2 | 0,012 | 0,58 |
| 3 | 0,045 | 1,34 |
| 4 | 0,243 | 1,34 |
| 5 | 0,266 | 1,64 |
| 6 | 0,302 | 1,65 |
| 7 | 0,451 | 1,91 |
| 8 | 1,041 | 1,96 |
| 9 | 1,423 | 2,08 |
| 10 | 1,914 | 2,18 |

*Рисунок 17*

*Виконання завдання*

1) Графік і рівняння прямої регресії знаходимо так само, як у лабораторній роботі 1 (рис. 18).

Рівняння лінійної регресії Y=1,265+0,573*X.

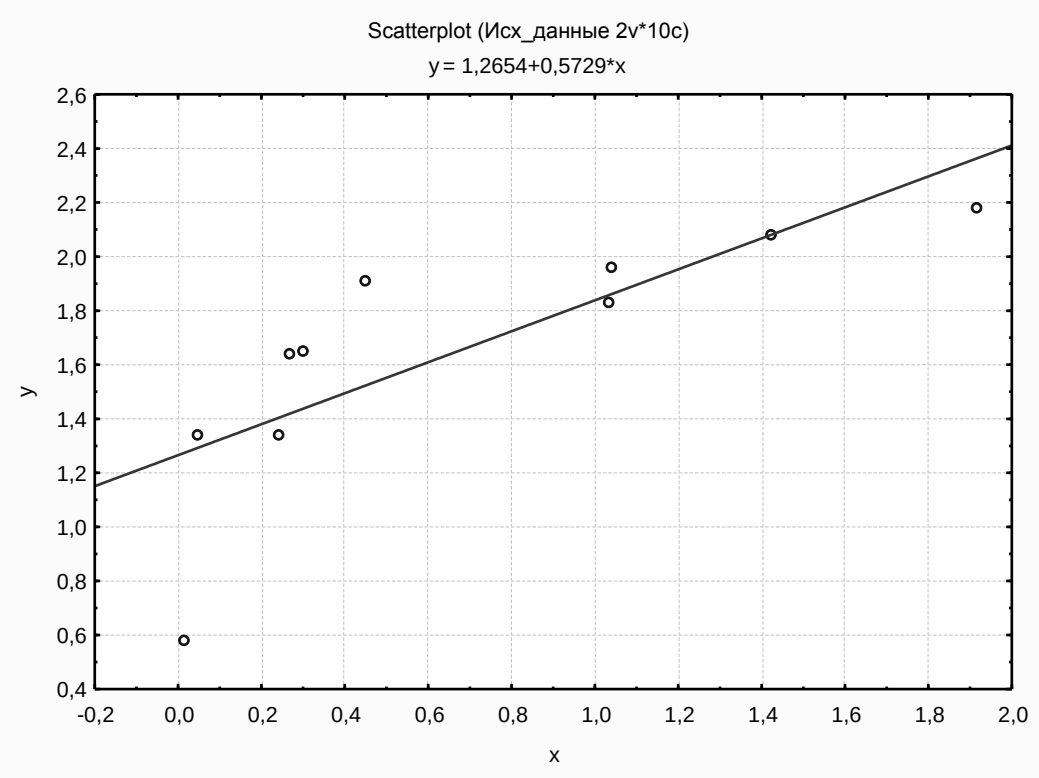


*Рисунок 18*

2) Прогноз *y* у кожній точці *x* за лінійною моделлю знаходимо так само як у практичній роботі 1. Позначаємо прогноз LIN_PR (рис. 19).

| | 1<br>X | 2<br>Y | 3<br>LIN_PR |
|---|---|---|---|
| 1 | 1,033 | 1,83 | 1,857 |
| 2 | 0,012 | 0,58 | 1,272 |
| 3 | 0,045 | 1,34 | 1,291 |
| 4 | 0,243 | 1,34 | 1,404 |
| 5 | 0,266 | 1,64 | 1,417 |
| 6 | 0,302 | 1,65 | 1,438 |
| 7 | 0,451 | 1,91 | 1,523 |
| 8 | 1,041 | 1,96 | 1,861 |
| 9 | 1,423 | 2,08 | 2,080 |
| 10 | 1,914 | 2,18 | 2,362 |

*Рисунок 19*

3) Графік і рівняння експонентної регресії отримують подібно графікові й рівнянню лінійної регресії. Експоненційну модель Statistica6 будує автоматично.

Активувати таблицю. Вибрати пункт меню **Craphs – Scatterplots –** (Графіки, точковий графік) – вибрати змінні **Variables** (для аргументу – *x* і функції – *y*) – **Ok,** вибрати вкладку **Advanced**, опції Regular, Exponential (регулярний, експоненційний) (рис. 20).

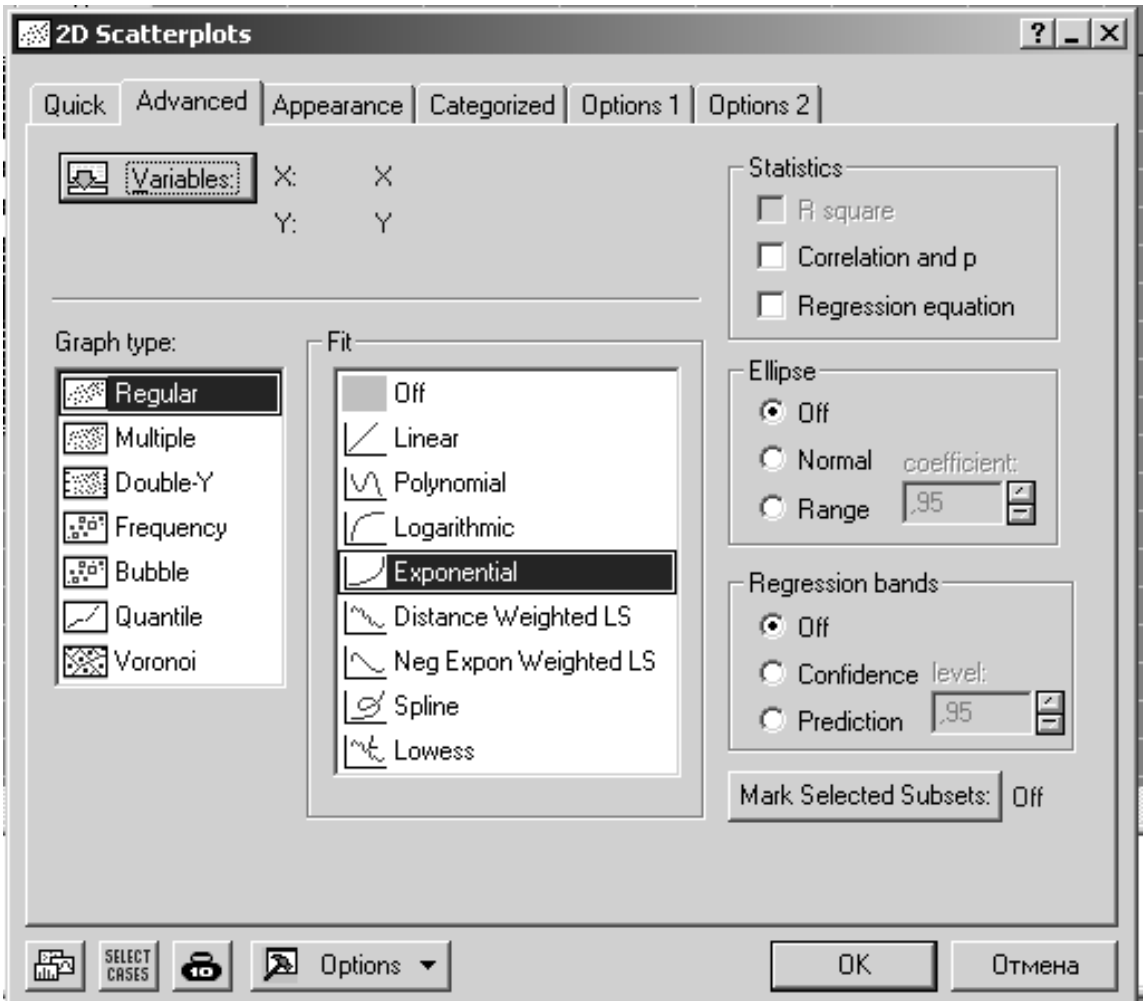


*Рисунок 20*

З'явиться графік експоненційної регресії, над яким записане рівняння регресії (рис. 21).

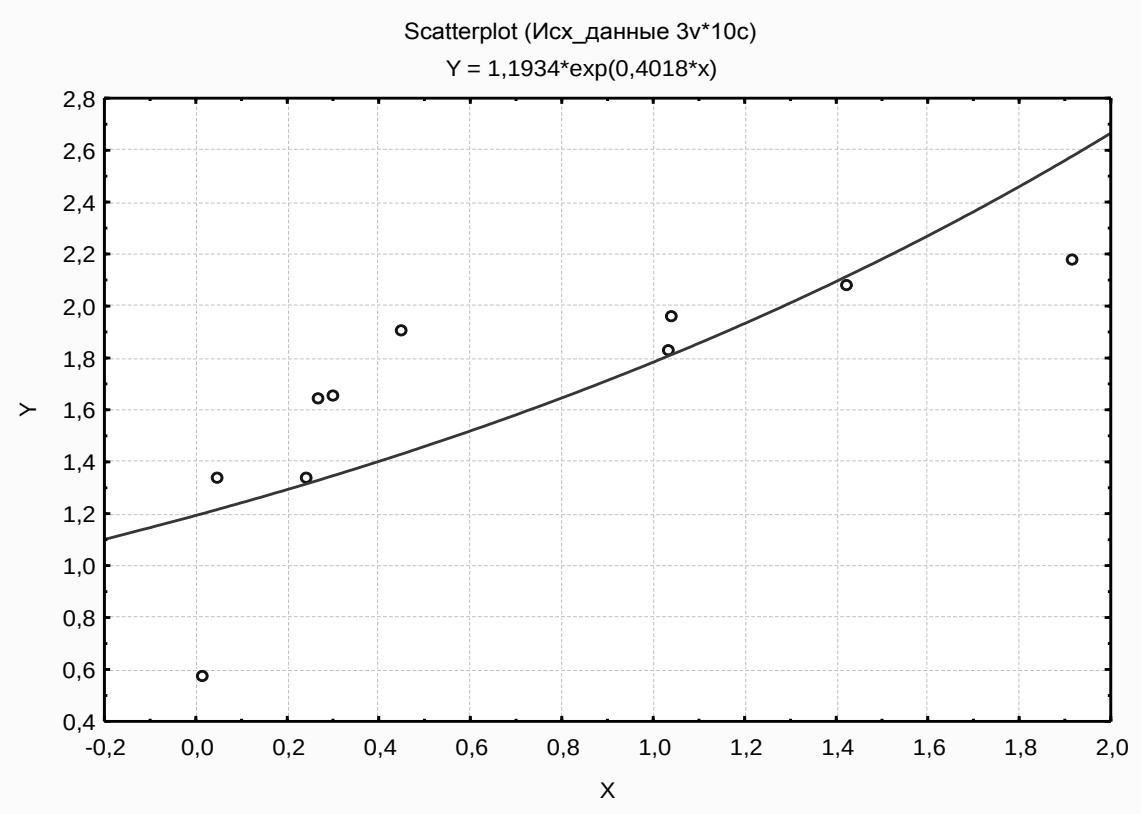


*Рисунок 21*

Рівняння експонентної регресії Y=1,193*exp(0,402*X).

4) Прогноз за експоненційною (нелінійною) моделлю будуємо так: подвійним щигликом по імені стовпця Y_NELIN увійти у вікно редагування стовпця. У вікні Long Name (унизу екрана) вписати формулу =1,193*EXP(0,402*X). Таблиця набуде наступного виду (рис. 22).

| | 1<br>X | 2<br>Y | 3<br>LIN_PR | 4<br>NELIN_PR |
|---|---|---|---|---|
| 1 | 1,033 | 1,83 | 1,857 | 1,807 |
| 2 | 0,012 | 0,58 | 1,272 | 1,199 |
| 3 | 0,045 | 1,34 | 1,291 | 1,215 |
| 4 | 0,243 | 1,34 | 1,404 | 1,315 |
| 5 | 0,266 | 1,64 | 1,417 | 1,328 |
| 6 | 0,302 | 1,65 | 1,438 | 1,347 |
| 7 | 0,451 | 1,91 | 1,523 | 1,430 |
| 8 | 1,041 | 1,96 | 1,861 | 1,813 |
| 9 | 1,423 | 2,08 | 2,080 | 2,114 |
| 10 | 1,914 | 2,18 | 2,362 | 2,575 |

*Рисунок 22*

5) Знаходимо розбіжність лінійного й експонентного прогнозу у відсотках. Для цього додаємо в таблицю змінну RASH, формула =(LIN_PR-NELIN_PR)/LIN_PR*100.

Формула вписується так: подвійним щигликом по імені стовпця RASH увійти у вікно редагування стовпця. У вікні Long Name (унизу екрана) вписати формулу.

Таблиця набуде наступного виду (рис. 23).

| | 1<br>X | 2<br>Y | 3<br>LIN_PR | 4<br>NELIN_PR | 5<br>RASH |
|---|---|---|---|---|---|
| 1 | 1,033 | 1,83 | 1,857 | 1,807 | 2,681 |
| 2 | 0,012 | 0,58 | 1,272 | 1,199 | 5,748 |
| 3 | 0,045 | 1,34 | 1,291 | 1,215 | 5,888 |
| 4 | 0,243 | 1,34 | 1,404 | 1,315 | 6,325 |
| 5 | 0,266 | 1,64 | 1,417 | 1,328 | 6,334 |
| 6 | 0,302 | 1,65 | 1,438 | 1,347 | 6,332 |
| 7 | 0,451 | 1,91 | 1,523 | 1,430 | 6,123 |
| 8 | 1,041 | 1,96 | 1,861 | 1,813 | 2,608 |
| 9 | 1,423 | 2,08 | 2,080 | 2,114 | -1,610 |
| 10 | 1,914 | 2,18 | 2,362 | 2,575 | -9,036 |

*Рисунок 23*

Негативне значення змінної RASH означає, що експоненційний прогноз перевищує лінійний прогноз на 1,610 і 9,036 % відповідно.

6) Обчислення квадратів відхилень для нелінійної регресії.

Додаємо змінну:

KV_OST, формула =(Y-NELIN_PR)^2.

Формули вписуються так: подвійним щигликом по імені стовпця KV_OST увійти у вікно редагування стовпця. У вікні Long Name (унизу екрана) вписати формулу: =(Y-NELIN_PR)^2-ОК. Таблиця набуде наступного виду (рис. 24).

| | 1<br>X | 2<br>Y | 3<br>LIN_PR | 4<br>NELIN_PR | 5<br>RASH | 6<br>KV_OST |
|---|---|---|---|---|---|---|
| 1 | 1,033 | 1,830 | 1,857 | 1,807 | 2,681 | 0,001 |
| 2 | 0,012 | 0,580 | 1,272 | 1,199 | 5,748 | 0,383 |
| 3 | 0,045 | 1,340 | 1,291 | 1,215 | 5,888 | 0,016 |
| 4 | 0,243 | 1,340 | 1,404 | 1,315 | 6,325 | 0,001 |
| 5 | 0,266 | 1,640 | 1,417 | 1,328 | 6,334 | 0,098 |
| 6 | 0,302 | 1,650 | 1,438 | 1,347 | 6,332 | 0,092 |
| 7 | 0,451 | 1,910 | 1,523 | 1,430 | 6,123 | 0,230 |
| 8 | 1,041 | 1,960 | 1,861 | 1,813 | 2,608 | 0,022 |
| 9 | 1,423 | 2,080 | 2,080 | 2,114 | -1,610 | 0,001 |
| 10 | 1,914 | 2,180 | 2,362 | 2,575 | -9,036 | 0,156 |

*Рисунок 24*

7) Обчислення суми квадратів залишків для нелінійної регресії.

Активувати таблицю .sta – **Statistics** – **Basic Statistics/Tables** – **Descriptive Statistics** – **Advanced** – залишити прапорець тільки в опції Sum – виділити змінну KV_OST – Summary (рис. 25).

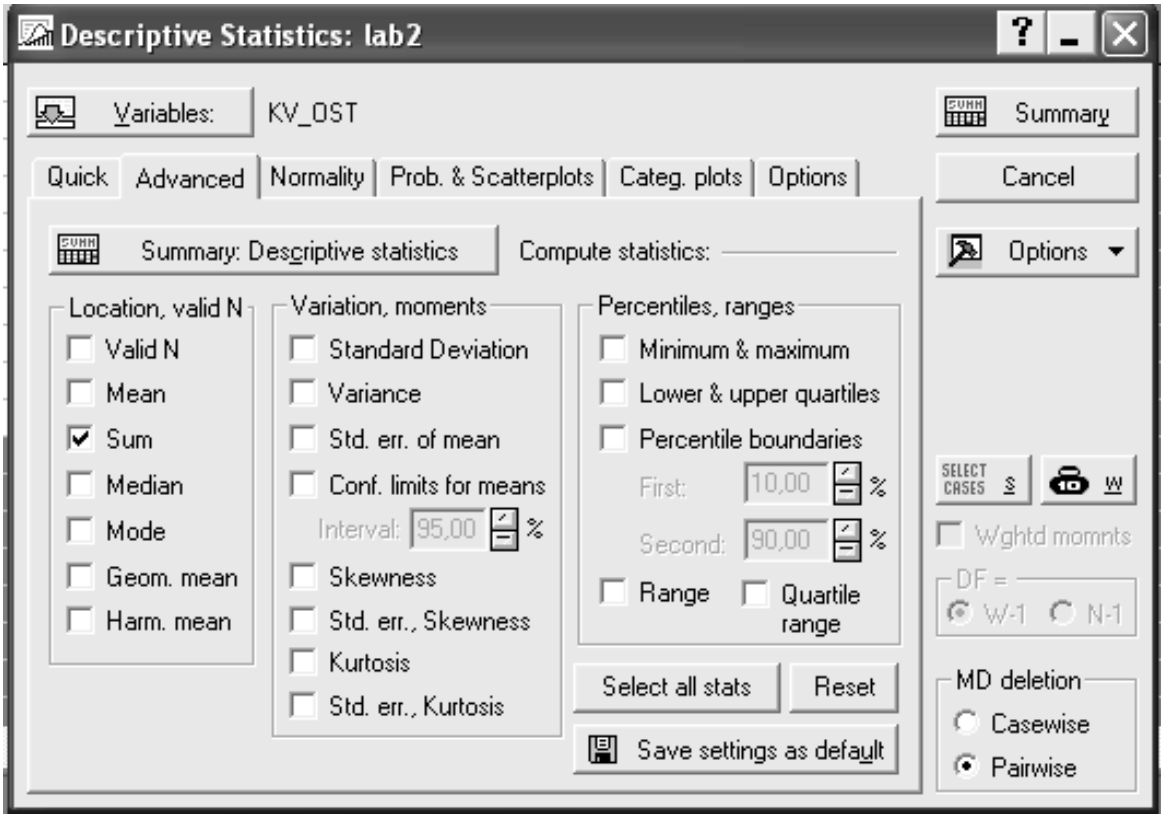


*Рисунок 25*

Одержуємо суму квадратів залишків для нелінійної моделі (рис. 26).

| Variable | Descriptive Statistics (lab2) |
|---|---|
| | **Sum** |
| **KV_OST** | 0,998226 |

*Рисунок 26*

8) Сума квадратів залишків лінійної моделі можна знайти як суму доданків виду (Y-LIN_PR)^2, де LIN_PR=1,265+0,573*X. Однак суму квадратів залишків лінійної моделі простіше знайти за допомогою модуля **Multiple Regression. Statistics – Multiple Regression – Variables**(Dependent Y - Independent X) – **OK** – **Advanced** – **ANOVA**(Overall goodness of fit) – **OK**. Сума квадратів залишків лінійної моделі(Residual) дорівнює 0,7726 (рис. 27).

| Effect | Analysis of Variance; DV: Y (lab2) | | | | |
|---|---|---|---|---|---|
| | **Sums of Squares** | df | Mean Squares | F | p-level |
| Regress. | 1,226490 | 1 | 1,226490 | 12,69988 | 0,007361 |
| **Residual** | 0,772600 | 8 | 0,096575 | | |
| Total | 1,999090 | | | | |

*Рисунок 27*

9) Остаточний вибір моделі – за сумою квадратів залишків. Тому що сума квадратів залишків для лінійної моделі 0,7726 менше, ніж сума квадратів залишків для експонентної моделі 0,9982, то вибираємо лінійну модель.

Примітка. Якщо значення *x* набагато більше, ніж *y*, то на графіку один з параметрів рівняння регресії може дорівнювати нулю. Тоді потрібно ввести нову змінну X_N, що виходить із *x* діленням на 1000 або 100, так, щоб X_N і *y* були одного порядку. Одержати залежність *y* від X_N. Щоб одержати залежність *y* від *x*, потрібно параметр b розділити на те ж число, що й при переході від *x* до X_N.

### 3.6 Висновки

Рівняння лінійної регресії $y$=1,265+0,573$x$.

Рівняння експонентної регресії $y = 1{,}193e^{0{,}402x}$.

Тому що сума квадратів залишків для лінійної моделі 0,7726 менше, ніж сума квадратів залишків для експонентної моделі 0,9982, то лінійна модель є оптимальною.

При неправильному виборі моделі (експонентна модель замість лінійної) максимальне завищення прогнозу дорівнює 9% у точці $x$=1,914, а максимальне заниження прогнозу дорівнює 6,334% у точці $x$=0,266.

## ПРАКТИЧНА РОБОТА №3

**Тема**: Перевірка однофакторної лінійної регресії на адекватність.

### 4.1 Стислі теоретичні відомості

*Статистична гіпотеза* – це припущення або про закон розподілу випадкової величини, або про значення числових характеристик (статистик) випадкової величини. *Нульовою* (основною) називають гіпотезу $H_0$, висунуту першою. *Конкуруючою* (альтернативною) називають гіпотезу, що суперечить основній гіпотезі. *Помилка першого роду* – відкинута правильна гіпотеза. *Помилка другого роду* – прийнята неправильна гіпотеза. Рівень значущості гіпотези $\alpha$ – імовірність відкинути правильну гіпотезу. Звичайно $\alpha$=0,05 чи $\alpha$=0,01. *Статистичний критерій* – випадкова величина, що служить для перевірки нульової гіпотези. Значення критерію, що спостерігається, обчислюється за вибіркою. *Область прийняття гіпотези* – сукупність значень критерію, при яких нульову гіпотезу приймають. *Критична область* – сукупність значень критерію, при яких нульову гіпотезу відкидають. Критичні точки (критичні значення критерію) відокремлюють область прийняття гіпотези

від критичної області. При дослідженні однофакторної регресії використовують два критерії:

- критерій Стьюдента з числом степенів вільності k=n-2: T(x,k), де n – обсяг вибірки (рис. 28);

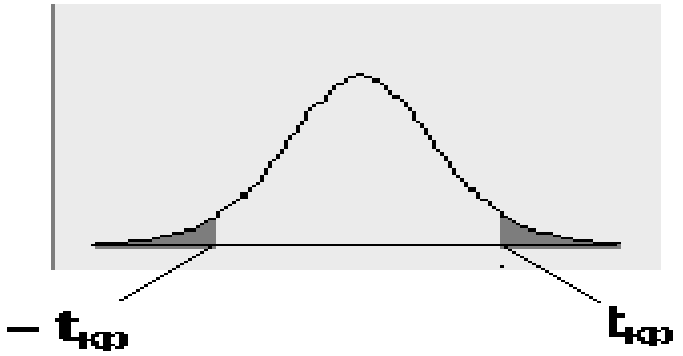


*Рисунок 28*

- критерій Фішера з двома числами степенів вільності – k1=1 і k2=n-2: F(x,k1,k2) (рис. 29).

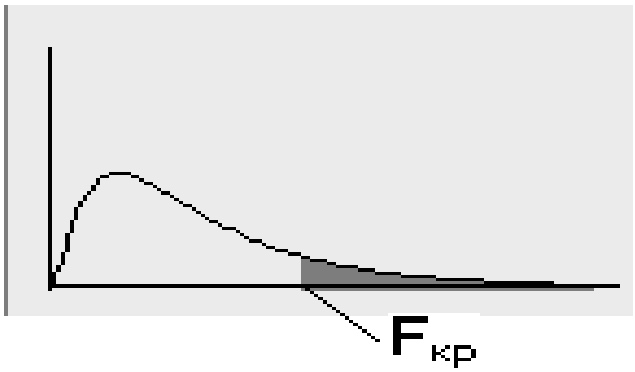


*Рисунок 29*

Критерій Стьюдента двосторонній – у нього дві симетричні критичні точки: $t_{кр}$ і $– t_{кр}$. Сумарна площа виділених ділянок дорівнює рівню значущості $\alpha$ нульової гіпотези $H_0$.

Критерій Фішера однобічний – у нього одна критична точка $F_{кр}$. Площа виділеної ділянки дорівнює значущості $\alpha$ нульової гіпотези $H_0$.

Кількість степенів вільності статистики дорівнює обсягу вибірки мінус кількість накладених зв'язків.

Статистична значущість коефіцієнта рівняння лінійної регресії з рівнем значущості $\alpha$ означає наступне: імовірність того, що даний

коефіцієнт, відмінний від нуля, дорівнює $\alpha$. Перевіряється значущість коефіцієнтів рівняння за допомогою критерію Стьюдента: знаходять за даними вибірки $t_{спост}$ і критичне значення критерію $t_{кр}$. Якщо $t_{спост} > t_{кр,}$ коефіцієнт рівняння статистично значущий. Якщо $t_{спост} < t_{кр,}$ коефіцієнт рівняння статистично не значущий.

Адекватність рівняння лінійної регресії з рівнем значущості $\alpha$ означає наступне: імовірність того, що відклик *y* залежить від фактору *x*, дорівнює $\alpha$. Перевіряється адекватність рівняння за допомогою критерію Фішера: знаходять за даними вибірки $F_{спост}$ і критичне значення критерію $F_{кр}$. Якщо $F_{спост} > F_{кр}$, рівняння адекватне. Якщо $F_{спост} < F_{кр}$ , рівняння не адекватне.

*Коефіцієнт кореляції* $r_{xy}$ характеризує щільність лінійного зв'язку між фактором *x* і відкликом *y*. Якщо $0,9<|r_{xy}|<1$ – зв'язок щільний; якщо $0,6<|r_{xy}|<0,9$ – зв'язок достатній; якщо $0,3<|r_{xy}|<0,6$ – зв'язок слабкий; якщо $0<|r_{xy}|<0,3$ – зв'язок відсутній. Знак коефіцієнта $r_{xy}$ характеризує характер лінійного зв'язку: при $r_{xy}>0$ зв'язок між *x* і *y* позитивний (зі зростом фактора *x* відклик *y* зростає ), при $r_{xy}<0$ зв'язок між *x* і *y* зворотний (зі зростом фактора *x* відклик *y* зменшується).

*Коефіцієнт детермінації* $R^2$ для лінійної регресії дорівнює квадрату коефіцієнта кореляції $r_{xy}$: $R^2 = r_{xy}^2$, $0 \le R^2 \le 1$. $R^2$ показує, яка частка дисперсії відклику *y* пояснюється рівнянням регресії.

### 4.2 Мета практичної роботи

Мають бути придбані наступні вміння: перевірка значущості коефіцієнтів лінійної моделі, перевірка адекватності лінійної моделі.

Мають бути засвоєні наступні поняття: статистична гіпотеза, рівень значущості гіпотези, статистичний критерій, значення критерію що спостерігаються, критичне значення критерію, кількість степенів вільності статистики, критерії Стьюдента і Фішера, статистична значущість коефіцієнтів моделі, адекватність моделі, коефіцієнти кореляції і детермінації, їхні властивості.

Робота розрахована на 2 години.

### 4.3 Завдання до практичної роботи

Використовуючи дані з практичної роботи 1, виконати наступні завдання:

1) Знайти коефіцієнт кореляції, коефіцієнт детермінації, значення критерію Фішера, що спостерігається, кількість степенів вільності критеріїв Фішера і Стьюдента.

2) Знайти коефіцієнти рівняння лінійної регресії $b_0$ і $b_1$.

3) Знайти значення критерію Стьюдента, що спостерігаються, для коефіцієнтів $b_0$ і $b_1$.

4) Знайти критичне значення критерію Стьюдента з рівнем значущості α=0.05.

5) Перевірити статистичну значущість коефіцієнтів $b_0$ і $b_1$.

6) Знайти критичне значення критерію Фішера з рівнем значущості α=0,05.

7) Перевірити лінійну модель на адекватність за допомогою критерію Фішера.

8) За значенням коефіцієнта кореляції зробити висновок про близькість зв'язку до лінійного.

### 4.4 Зміст звіту

Звіт про практичну роботу повинен містити:

1) Тему роботи, завдання.
2) Роздрук таблиць.
3) Пояснення отриманих значень коефіцієнтів з погляду економетрії.
4) Відповіді на питання завдання.
5) Статистичний і економетричний аналізи отриманих результатів.

### 4.5 Приклад виконання практичної роботи у пакеті Statistica 6

*Економічні дані*

Економічні дані взяти з лабораторної роботи 1.

Вихідна таблиця даних вставляється у звіт так само, як у практичній роботі 1 (рис. 30).

| | 1<br>X | 2<br>Y |
|---|---|---|
| 1 | 1,033 | 1,83 |
| 2 | 0,012 | 0,58 |
| 3 | 0,045 | 1,34 |
| 4 | 0,243 | 1,34 |
| 5 | 0,266 | 1,64 |
| 6 | 0,302 | 1,65 |
| 7 | 0,451 | 1,91 |
| 8 | 1,041 | 1,96 |
| 9 | 1,423 | 2,08 |
| 10 | 1,914 | 2,18 |

*Рисунок 30*

*Виконання завдання*

1) Знайдемо коефіцієнт кореляції, коефіцієнт детермінації, значення критерію Фішера, що спостерігається, кількість степенів вільності критеріїв Фішера і Стьюдента. **Statistics – Multiple Regression** (множинна регресія) – **Variables** – (залежна змінна dependent *y* – незалежна змінна independent *x*) – **Ok** – **Advanced** – **Summary**: Regression results (підсумки регресійного аналізу) (рис. 31).

| | Regression Summary for Dependent Variable: Y (Исх_данные)<br>R= ,78327797 R?= ,61352438 Adjusted R?= ,56521493<br>F(1,8)=12,700 p<,00736 Std.Error of estimate: ,31077 | | | | | |
|---|---|---|---|---|---|---|
| N=10 | Beta | Std.Err.<br>of Beta | B | Std.Err.<br>of B | t(8) | p-level |
| Intercept | | | 1,265414 | 0,146166 | 8,657390 | 0,000025 |
| X | 0,783278 | 0,219794 | 0,572936 | 0,160771 | 3,563689 | 0,007361 |

*Рисунок 31*

У таблиці, що з'явилася, знаходяться всі необхідні відомості: коефіцієнт кореляції (R)=0,783, коефіцієнт детермінації (R2)=0,613, значення критерію Фішера, що спостерігається $F_{спост}$ = (F(1, 8))=12,7, кількість степенів вільності критерію Фішера: k1=1 і k2=8, кількість степенів вільності критерію Стьюдента (t(8)): k=8.

2) Знайдемо коефіцієнти рівняння лінійної регресії $b_0$ і $b_1$.

У стовпці B таблиці Regression Summary знаходяться значення параметрів: $b_0 = 1{,}265$, $b_1 = 0{,}573$.

3) Для перевірки значущості коефіцієнтів $b_0$ і $b_1$ використовуються спостережувальні значення критерію Стьюдента для кожного коефіцієнта. Ці значення знаходяться у стовпці t(8) таблиці Regression Summary: $t_{набл}(b_0) = 8{,}657$, $t_{набл}(b_1) = 3{,}564$

4) Знайдемо критичне значення критерію Стьюдента з рівнем значущості $\alpha = 0{,}05$ (рис. 32): **Statistics – Probability Calculator – Distributions – t(Student)** (статистика – підрахунок імовірності – розподіл – t-розподіл).

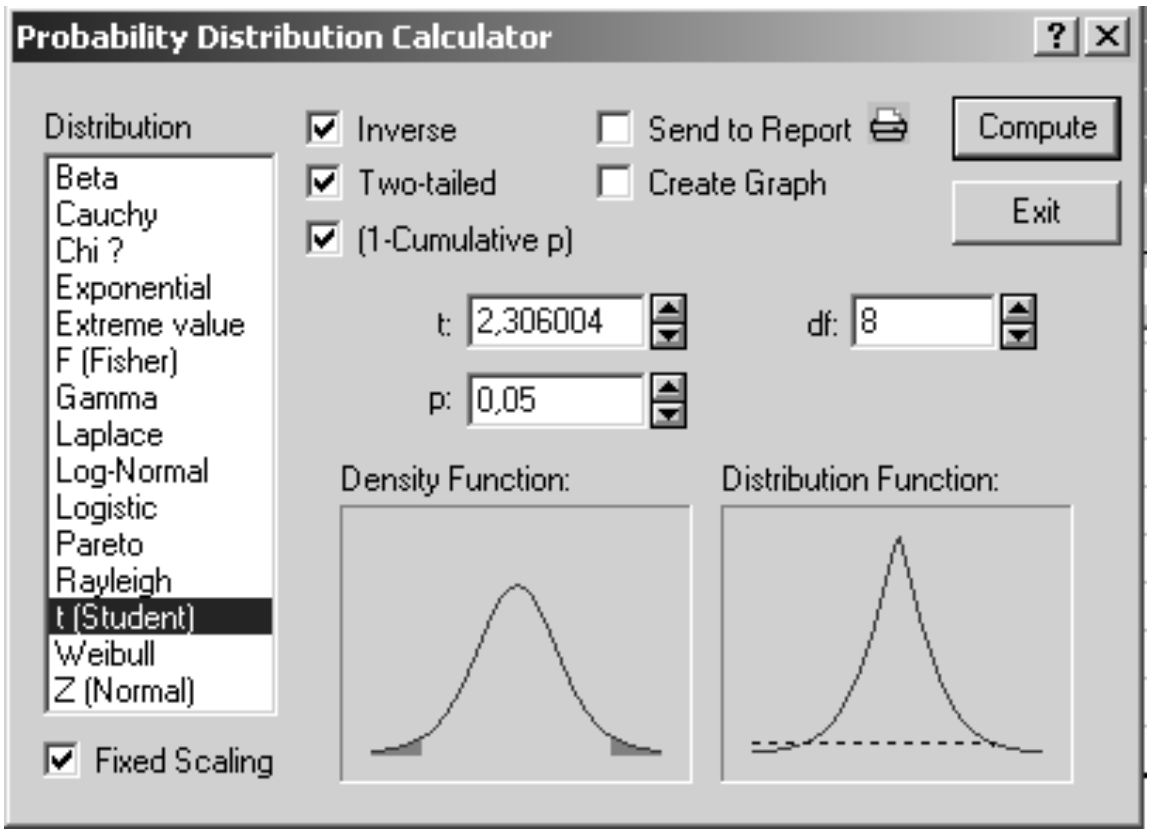


*Рисунок 32*

У віконце df внести число ступенів вільності 8 і у віконце p внести рівень значущості, який дорівнює 0,05, натиснути Compute (підрахунок).

5) Визначаємо статистичну значущість коефіцієнтів рівняння лінійної регресії $b_0$ і $b_1$. Тому що спостережувальні значення критерію Стьюдента для кожного коефіцієнта $t_{набл}(b_0) = 8{,}657$ і $t_{набл}(b_1) = 3{,}564$

більше критичного значення t(8) = 2,306, то коефіцієнти $b_0$ = 1,265 і $b_1$ = 0,573 статистично значущі.

6) Знайдемо критичне значення критерію Фишера з рівнем значущості α = 0,05 (рис. 33): **Statistics – Probability Calculator – Distributions – F (Fisher**)(статистика - підрахунок імовірності – розподілу – F(Fisher)).

У віконця df1, df2 внести числа степенів вільності 1 і 8, у віконце p внести рівень значущості гіпотези – 0,05, нажати Compute (підрахунок).

F(1,8)=5,317.

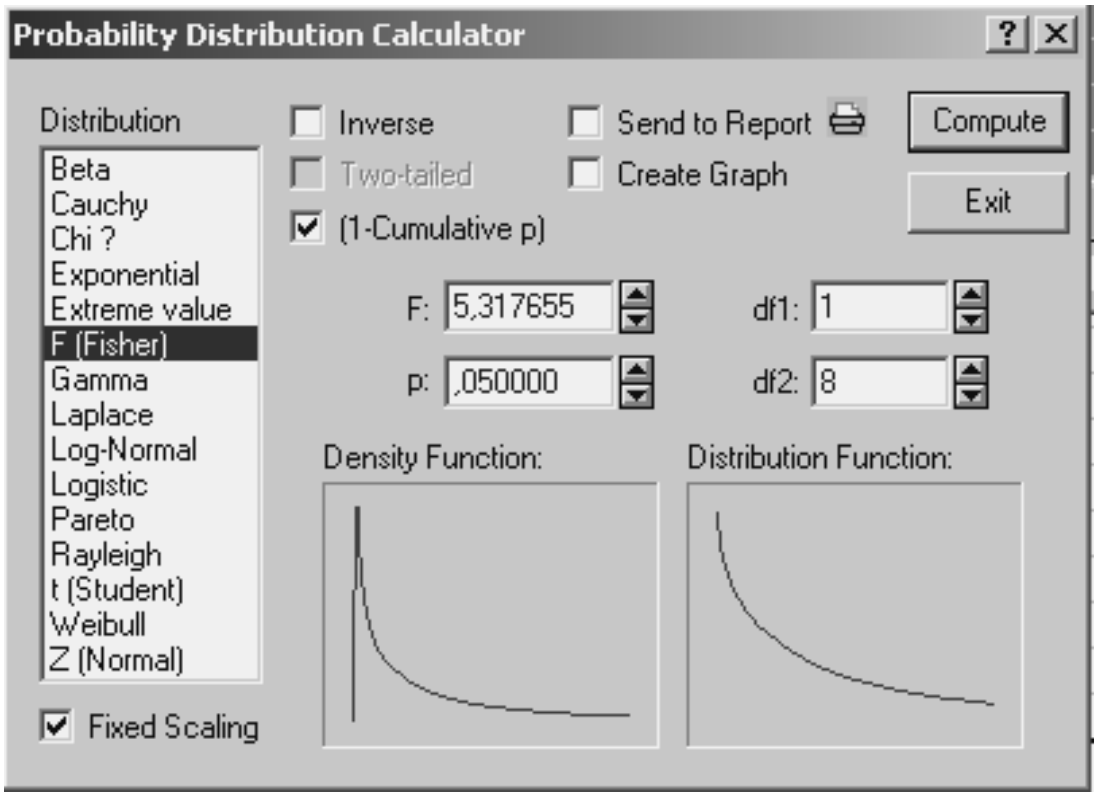


*Рисунок 33*

7) Перевірка лінійної моделі на адекватність за допомогою критерію Фішера. Тому що спостережувальне значення критерію Фішера $F_{набл}$ = (F(1, 8))=12,7 більше критичного значення F(1,8) =5,317, то лінійна модель адекватна.

8) Висновок про близькість зв'язку до лінійного. Коефіцієнт кореляції R = 0,783 належить до діапазону (0,6; 0,9). Лінійний зв'язок достатній.

Примітка. У звіт рисунки 15 і 16 можна не включати.

### 4.6 Висновки

1) Статистичні характеристики вибірки беремо з таблиці (рис. 31): коефіцієнт кореляції (R)=0,783, коефіцієнт детермінації (R2)=0,613, значення критерію Фішера, що спостерігаються $F_{спост}$=(F(1, 8))=12,7, число степенів вільності критерію Фішера k1=1 і k2=8 , число степенів вільності критерію Стьюдента (t(8)) k=8.

2) Коефіцієнти лінійної регресії $b_0$ і $b_1$ і значення критерію Стьюдента, що спостерігаються для кожного коефіцієнта, беремо у стовпцях B і t(8) таблиці (рис. 31): $b_0$=1,265; $b_1$=0,573; $t_{спост}(b_0)$=8,657; $t_{спост}(b_1)$=3,564. Критичне значення критерію Стьюдента з рівнем значущості $\alpha$=0,05 знаходимо за допомогою імовірнісного калькулятора (Probability Calculator): t(8)=2,306004. Тому що значення критерію Стьюдента, що спостерігаються, для кожного коефіцієнта більше критичного значення, коефіцієнти $b_0$ і $b_1$ статистично значущі.

3) Критичне значення критерію Фішера з рівнем значущості $\alpha$=0.05 знаходимо за допомогою імовірнісного калькулятора ($F_{кр}$= F(1,8)=5,317) і порівнюємо зі значенням критерію Фішера, що спостерігається: $F_{спост}$ = F(1, 8)=12,7. Тому що значення критерію Фішера, що спостерігається, більше критичного значення, лінійна модель адекватна.

4) Виконаний статистичний аналіз показує, що залежність обсягу випуску металургійної продукції Y (мільйонів тонн) від капітальних витрат X (мільйонів гривень) відбувається за лінійним законом Y=1,265+1,573X. При цьому всі коефіцієнти рівняння статистично значущі, а саме рівняння адекватне.

## ПРАКТИЧНА РОБОТА №4

**Тема**: Прогноз на підставі лінійної регресії. Точність прогнозу.

### 5.1 Стислі теоретичні відомості

Рівняння лінійної регресії $y = b_0+b_1x$ знаходиться за даними вибірки методом найменших квадратів. На рисунку 17 це похила пряма, зображена лінією. Точна лінійна залежність між *x* і *y*: $y = \beta_0+\beta_1x + \varepsilon$, де $\varepsilon$ – випадковий член, невідома Можна тільки стверджувати, що вона з імовірністю $\gamma$ розташована в довірчій області, обмеженій лініями гіперболи (рис. 34).

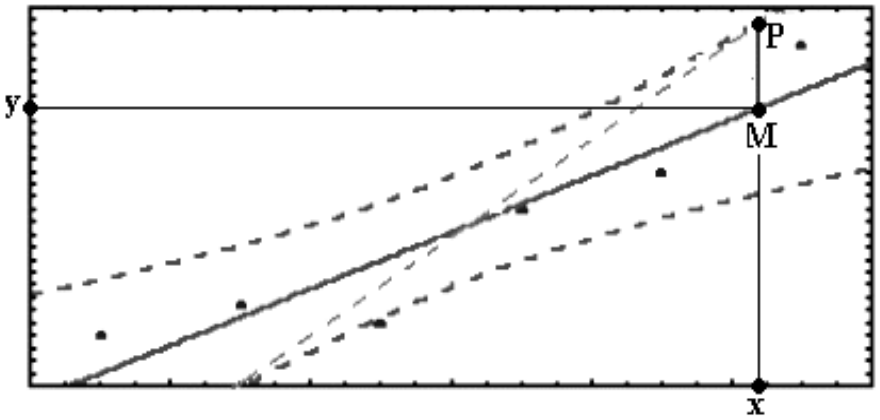


*Рисунок 34*

Імовірність $\gamma$ називається *рівнем довіри*. Звичайно $\gamma$=0,95 або $\gamma$=0,99. Точна лінія регресії $y = \beta_0+\beta_1x + \varepsilon$ зображена на рисунку 34 пунктирною прямою. Прогноз *y* в точці *x* роблять за рівнянням $y = b_0+b_1x$. Точне значення прогнозу може з імовірністю $\gamma$ відповідати будь-якій точці довірчого інтервалу PQ (рис. 35).

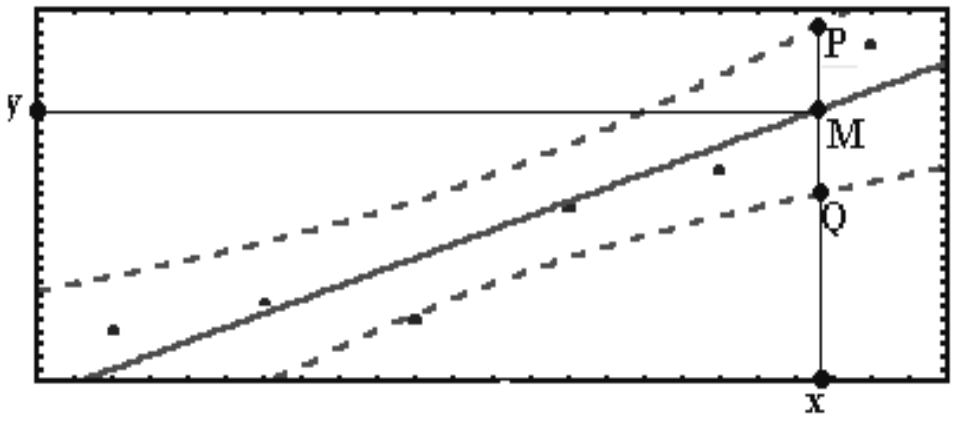


*Рисунок 35*

*Напівширина довірчого інтервалу* $\delta$=MP=MQ обчислюється за формулою

$$\delta = \sigma_\varepsilon t_\gamma \sqrt{1+\frac{1}{n}+\frac{(x-x_{cp})^2}{\sum_i (x-x_{cp})^2}},$$

де $\sigma_\varepsilon$ – середньоквадратична помилка залишків;

$t_\gamma$ – критична точка розподілу Стьюдента, що відповідає рівню довіри $\gamma$ (рис. 37, площа виділеної області дорівнює рівню довіри $\gamma$);

n – обсяг вибірки;

$x_{cp}$ – середнє значення фактору $x$.

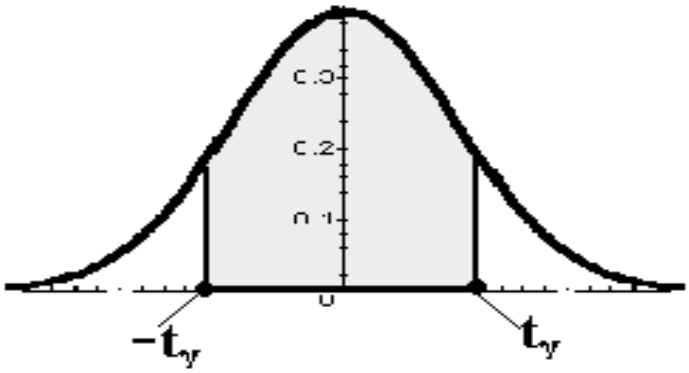


*Рисунок 36*

## 5.2 Мета практичної роботи

Мають бути придбані наступні вміння:

1) побудування довірчого інтервалу для прогнозу. Пояснення розходження в ширині довірчих областей в залежності від рівня довіри;

2) розрахунку максимальної відносної помилки прогнозу.

Мають бути засвоєні наступні поняття: довірча область, довірчий інтервал, рівень довіри, напівширина довірчого інтервалу, формула напівширини довірчого інтервалу, зміст параметрів, що входять до формули.

Робота розрахована на 4 години.

## 5.3 Завдання до практичної роботи

Використовуючи дані з практичної роботи 1, виконати такі завдання:

1) Знайти рівняння лінійної регресії.

2) Побудувати графіки лінії регресії з 80, 95 і 99%-ми довірчими областями.

3) Розрахувати напівширину довірчого інтервалу для всіх точок вибірки, а також для двох будь-яких точок з області прогнозів для всіх трьох значень коефіцієнта довіри (80, 95 і 99%).

4) Оцінити максимальну відносну помилку прогнозу (у відсотках) для всіх трьох значень коефіцієнта довіри (80, 95 і 99%).

## 5.4 Зміст звіту

Звіт про практичну роботу повинен містити:

1) Тему роботи, завдання.

2) Роздрук таблиць і графіків.

3) Пояснення отриманих значень коефіцієнтів з погляду економетрії.

4) Прогноз для двох точок з області прогнозу для всіх трьох коефіцієнтів довіри і висновок про зміну помилки прогнозу в залежності від коефіцієнта довіри.

## 5.5 Приклад виконання практичної роботи в пакеті Statistica6

*Економічні дані*

Економічні дані взяти з практичної роботи 1.

Вихідна таблиця даних (рис. 37) вставляється у звіт так само, як у лабораторній роботі 1.

| | 1<br>X | 2<br>Y |
|---|---|---|
| 1 | 1,033 | 1,83 |
| 2 | 0,012 | 0,58 |
| 3 | 0,045 | 1,34 |
| 4 | 0,243 | 1,34 |
| 5 | 0,266 | 1,64 |
| 6 | 0,302 | 1,65 |
| 7 | 0,451 | 1,91 |
| 8 | 1,041 | 1,96 |
| 9 | 1,423 | 2,08 |
| 10 | 1,914 | 2,18 |

*Рисунок 37*

*Виконання завдання*

1) Побудова графіків лінійної регресії з 80%, 95%і 99% довірчими областями. **Graphs – 2D Graphs – Scatterplots – Variables** (X, Y) – **Ok** - далі на вкладці **Advanced** (рис. 38) відмітити опції **Regular – Linear - Confidence level – 0,8 – Ok** (Графіки – 2D графіки – Графіки розсіювання – Змінні (X, Y) – Ok, Регулярний – Лінійний – Рівень довіри – 0,8 – Ok).

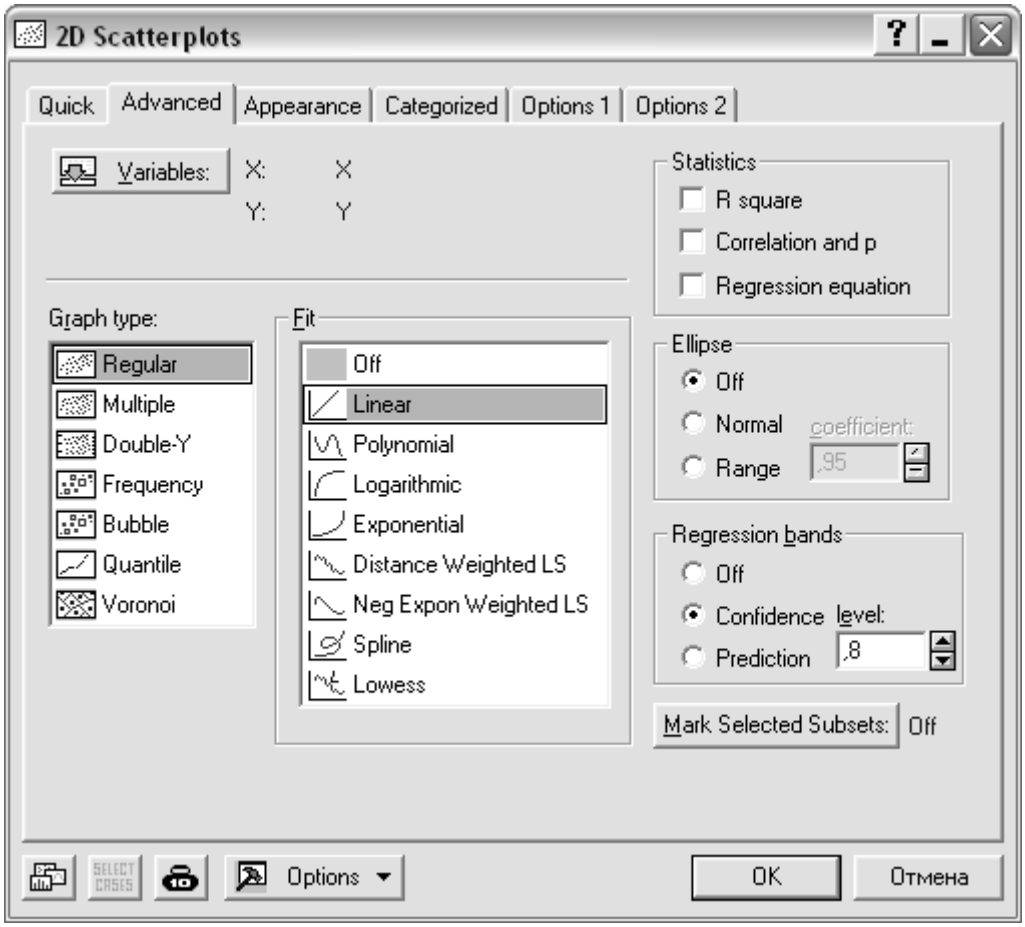


*Рисунок 38*

Аналогічно будуємо довірчі області з рівнями довіри 0,95 і 0,99.

Одержимо три графіки.

Рівень довіри 80% (рис. 39).

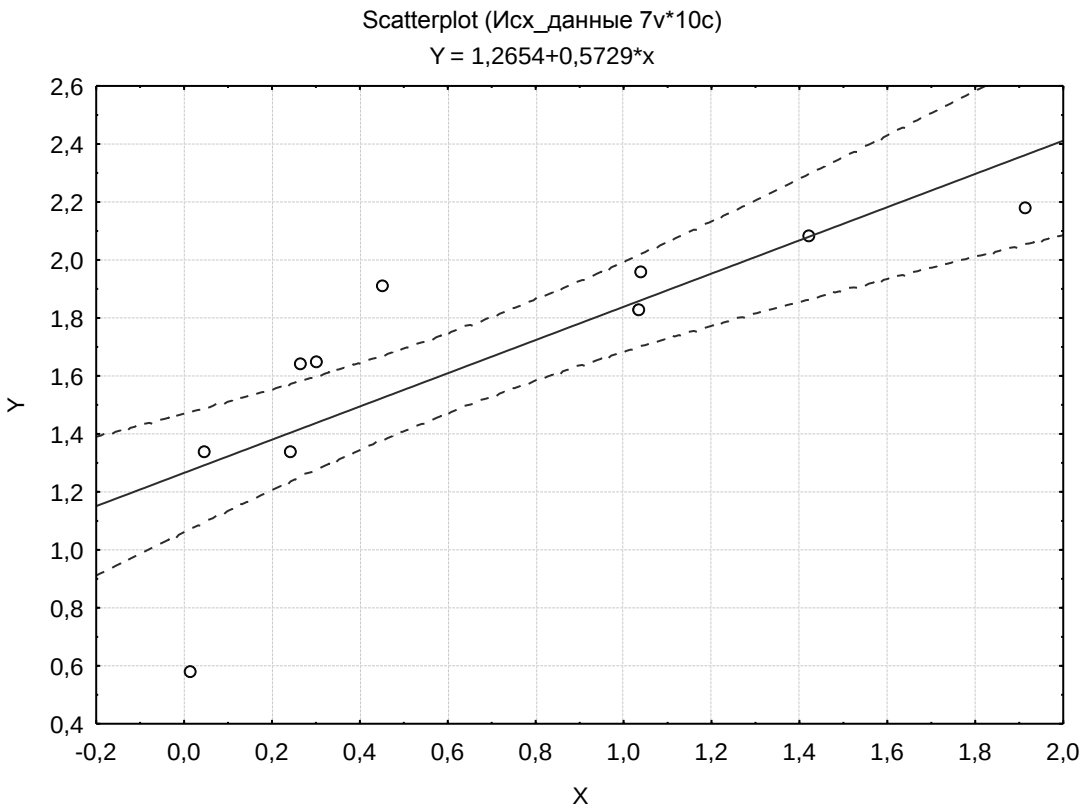


*Рисунок 39*

Рівень довіри 95% (рис. 40).

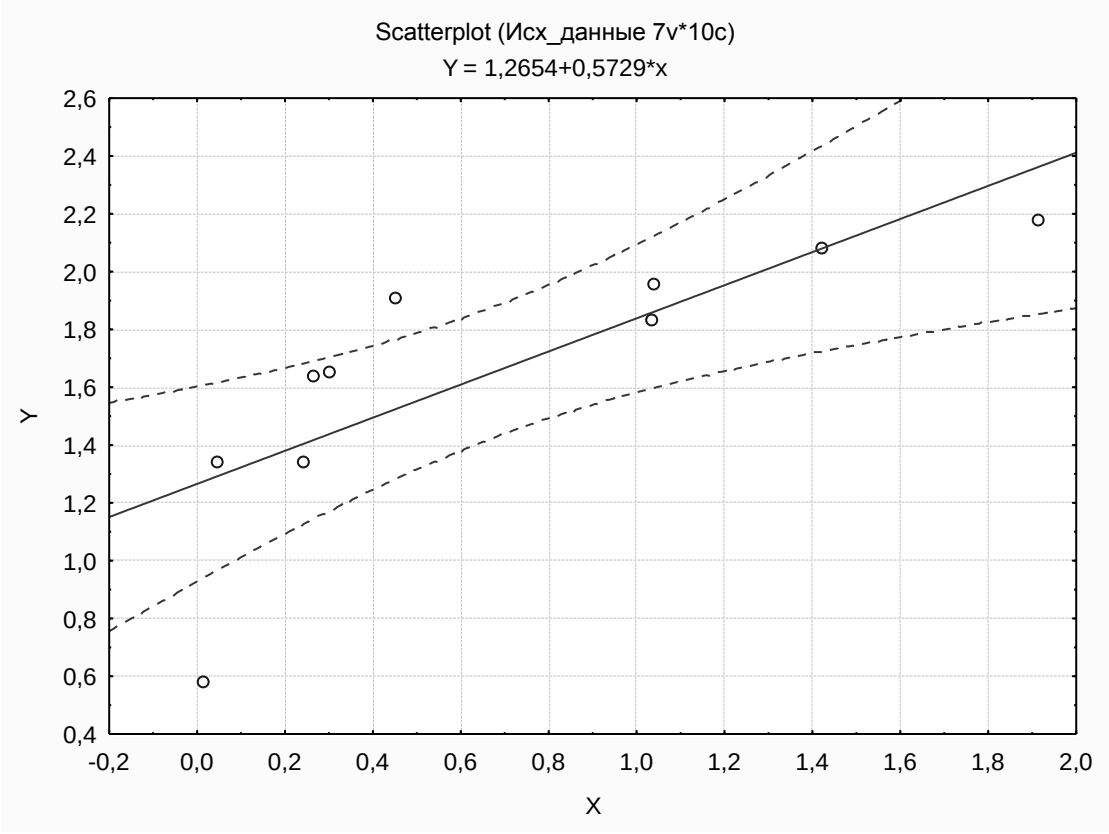


*Рисунок 40*

Рівень довіри 99% (рис. 41).

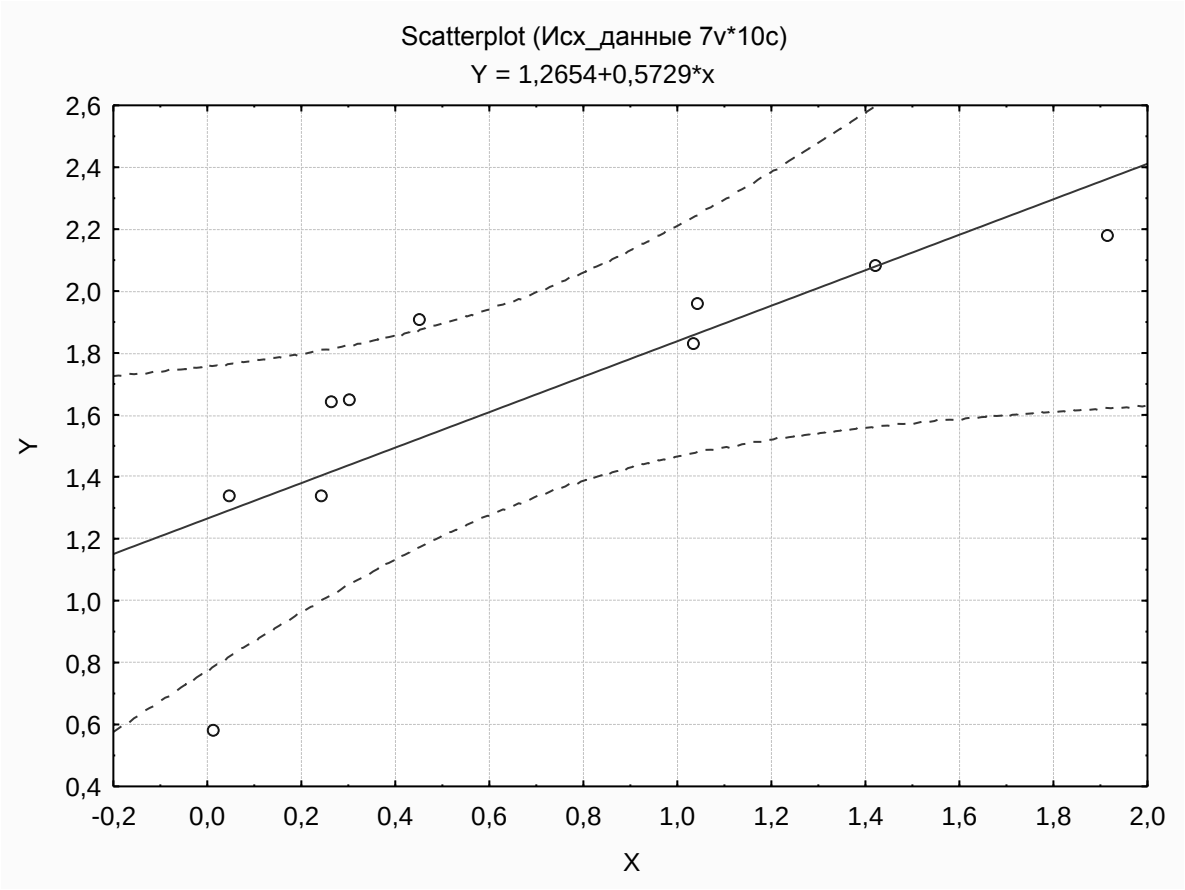


*Рисунок 41*

2) Визначаємо величини параметрів, необхідних для розрахунку напівширини довірчого інтервалу в точці $x$.

Напівширина $\delta$ довірчого інтервалу в точці $x$ розраховується за формулою

$$\delta = \sigma_\varepsilon t_\gamma \sqrt{1+\frac{1}{n}+\frac{(x-x_{cp})^2}{\sum_i (x-x_{cp})^2}}.$$

Визначаємо величину параметрів, що входять до формули для напівширини довірчого інтервалу:

а) $t_\gamma$ – критична точка розподілу Стьюдента, що відповідає рівню довіри $\gamma$ (рис. 42) **Statistics – Probability Calculator – Distributions – t** (Student) (статистика – імовірнісний калькулятор – розподілу – t-розподіл).

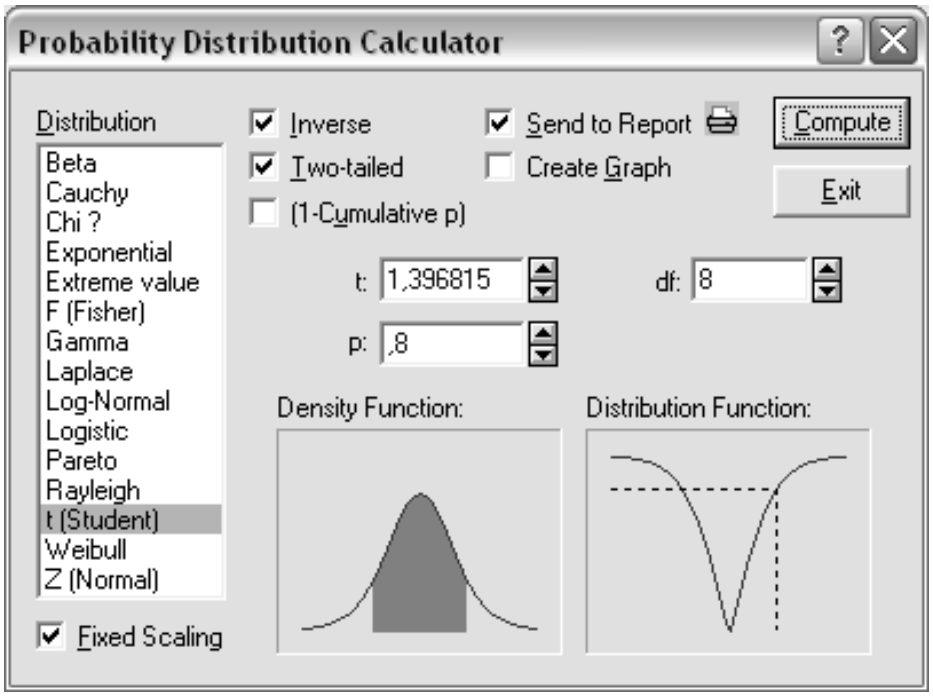


*Рисунок 42*

У віконце df внести число ступенів волі =n-2 і у віконце p внести рівень значущості, який дорівнює 0,05, натиснути Compute.

Для рівнів довіри γ=0,95 і γ=0,99 аналогічно:

t(8)=1,396815 p=0,8,

t(8)=2,306004 p=0,95,

t(8)=3,355385 p=0,99.

б) Значення $\sigma_\varepsilon$ (середньоквадратична погрішність залишків) беремо з таблиці Regression Summary (підсумки регресійного аналізу) (див. лабораторну роботу 3 – Std.Error of estimate = $\sigma_\varepsilon$ = 0,31);

в) n – обсяг вибірки, n=10;

г) Значення $x_{cp}$ (середнє значення фактора $x$) було знайдено в лабораторній роботі 1: $x_{cp}$=0,673.

Для кожного значення фактора $x$ знаходимо квадрат відхилення від середнього значення $(x\text{-}x_{cp})^2$. Для цього додаємо новий стовпець: подвійним щигликом по заголовку входимо у вікно властивостей, даємо заголовок, наприклад, KV, у вікні Long Name уводимо формулу = (x-0,673) ^2 (x – незалежний фактор, 0,673 – середнє значення).

Знаходимо суми значень $(x\text{-}x_{ср})^2$ (суми значень стовпця KV): **Statistics – Basic Statistics/Tables – Descriptive Statistics – Variables** (KV) – **Ok** – на вкладці **Advanced** – залишити прапорець тільки в опції Sum - Summary (рис. 43).

| Variable | Descriptive Statistics (Исх_данные) |
|---|---|
| | **Sum** |
| **KV** | 3,736384 |

*Рисунок 43*

Отже, $\Sigma((x\text{-}x_{ср})\ ^2)=3,73$.

3) Знаходимо прогноз у точках вибірки і двох додаткових точках. У вихідну таблицю додамо дві точки з області прогнозів, наприклад, 0,6 і 1,8. Для цього додаємо два випадки, використовуючи кнопку Cases. Також додаємо стовпці: Y_REGR – прогноз за моделлю (див. лабораторну роботу1), DELTA – напівширина довірчого інтервалу і POGR – максимальна помилка прогнозу. Для кожного рівня довіри ($\gamma$=0,80, $\gamma$=0,95, $\gamma$=0,99) розраховуємо окрему таблицю.

Для розрахунку DELTA при $\gamma$=0,80 у вікно Long Name вводимо формулу =1,397*0,31*(1+1/10+KV/3,73)^0,5.

4) Оцінюємо максимальну відносну помилку прогнозу (у відсотках) для всіх трьох значень коефіцієнта довіри (80, 95 і 99%).

Максимальна помилка прогнозу POGR у відсотках розраховується за формулою = DELTA/abs(Y_REGR)*100. Для $\gamma$=0,80 отримаємо таблицю (рис. 44).

| | 1<br>X | 2<br>Y | 3<br>Y_REGR | 4<br>KV | 5<br>DELTA | 6<br>POGR |
|---|---|---|---|---|---|---|
| 1 | 1,033 | 1,830 | 1,857 | 0,130 | 0,461 | 24,844 |
| 2 | 0,012 | 0,580 | 1,272 | 0,437 | 0,478 | 37,565 |
| 3 | 0,045 | 1,340 | 1,291 | 0,394 | 0,476 | 36,841 |
| 4 | 0,243 | 1,340 | 1,404 | 0,185 | 0,464 | 33,066 |
| 5 | 0,266 | 1,640 | 1,417 | 0,166 | 0,463 | 32,685 |
| 6 | 0,302 | 1,650 | 1,438 | 0,138 | 0,462 | 32,110 |
| 7 | 0,451 | 1,910 | 1,523 | 0,049 | 0,457 | 29,993 |
| 8 | 1,041 | 1,960 | 1,861 | 0,135 | 0,462 | 24,800 |
| 9 | 1,423 | 2,080 | 2,080 | 0,562 | 0,484 | 23,281 |
| 10 | 1,914 | 2,180 | 2,362 | 1,540 | 0,533 | 22,554 |
| 11 | 0,600 | | 1,6088 | 0,005 | 0,455 | 28,251 |
| 12 | 1,800 | | 2,2964 | 1,270 | 0,520 | 22,634 |

*Рисунок 44*

Потім перераховуємо стовпець DELTA для γ=0,95: =2,306*0,31*(1+1/10+KV/3,73)^0,5 (для перерахування максимальної помилки прогнозу) **Vars – Recalculate – All variables** (змінні - перерахування – всі змінні). Вносимо отриману таблицю в автозвіт.

Для γ=0,95 отримаємо таблицю (рис. 45).

| | 1<br>X | 2<br>Y | 3<br>Y_REGR | 4<br>KV | 5<br>DELTA | 6<br>POGR |
|---|---|---|---|---|---|---|
| 1 | 1,033 | 1,830 | 1,857 | 0,130 | 0,762 | 41,009 |
| 2 | 0,012 | 0,580 | 1,272 | 0,437 | 0,789 | 62,008 |
| 3 | 0,045 | 1,340 | 1,291 | 0,394 | 0,785 | 60,812 |
| 4 | 0,243 | 1,340 | 1,404 | 0,185 | 0,766 | 54,582 |
| 5 | 0,266 | 1,640 | 1,417 | 0,166 | 0,765 | 53,953 |
| 6 | 0,302 | 1,650 | 1,438 | 0,138 | 0,762 | 53,004 |
| 7 | 0,451 | 1,910 | 1,523 | 0,049 | 0,754 | 49,510 |
| 8 | 1,041 | 1,960 | 1,861 | 0,135 | 0,762 | 40,936 |
| 9 | 1,423 | 2,080 | 2,080 | 0,562 | 0,799 | 38,430 |
| 10 | 1,914 | 2,180 | 2,362 | 1,540 | 0,879 | 37,230 |
| 11 | 0,600 | | 1,6088 | 0,005 | 0,750 | 46,633 |
| 12 | 1,800 | | 2,2964 | 1,270 | 0,858 | 37,362 |

*Рисунок 45*

Аналогічно перераховуємо стовпець DELTA для γ=0,99: =3,36*0,31*(1+1/10+KV/3,73)^0,5.

Для γ=0,99 отримаємо таблицю (рис. 46).

| | 1<br>X | 2<br>Y | 3<br>Y_REGR | 4<br>KV | 5<br>DELTA | 6<br>POGR |
|---|---|---|---|---|---|---|
| 1 | 1,033 | 1,830 | 1,857 | 0,130 | 1,110 | 59,753 |
| 2 | 0,012 | 0,580 | 1,272 | 0,437 | 1,149 | 90,350 |
| 3 | 0,045 | 1,340 | 1,291 | 0,394 | 1,144 | 88,608 |
| 4 | 0,243 | 1,340 | 1,404 | 0,185 | 1,117 | 79,529 |
| 5 | 0,266 | 1,640 | 1,417 | 0,166 | 1,114 | 78,613 |
| 6 | 0,302 | 1,650 | 1,438 | 0,138 | 1,111 | 77,231 |
| 7 | 0,451 | 1,910 | 1,523 | 0,049 | 1,099 | 72,139 |
| 8 | 1,041 | 1,960 | 1,861 | 0,135 | 1,110 | 59,647 |
| 9 | 1,423 | 2,080 | 2,080 | 0,562 | 1,165 | 55,996 |
| 10 | 1,914 | 2,180 | 2,362 | 1,540 | 1,281 | 54,247 |
| 11 | 0,600 | | 1,6088 | 0,005 | 1,093 | 67,948 |
| 12 | 1,800 | | 2,2964 | 1,270 | 1,250 | 54,439 |

*Рисунок 46*

### 5.6 Висновки

Побудовано довірчі області для лінійної регресії y=1,265+0,573x для трьох рівнів довіри 80, 95 і 99%. Розраховано прогноз за лінійною регресією у всіх точках вибірки і в двох додаткових точках з області прогнозів x= 0,6 і x=1,8, а також розраховані відносні помилки прогнозів.

З порівняння відносних похибок прогнозів видно, що підвищення рівня довіри з 80 до 99%, знижує точність прогнозу. Оцінюємо приблизно, в скільки разів знижується точність прогнозу. Для цього можна порівняти відносні похибки для того самого *x*, наприклад, *x*=1,800: POGR80=27,166; POGR95=44,843; POGR99=65,242.

З наведених значень помилок видно, що підвищення рівня довіри з 80 до 99% знижує точність прогнозу в 2,4 рази.

## ПРАКТИЧНА РОБОТА №5

**Тема**: Перевірка факторів на мультиколінеарність. Вибір моделі багатофакторної регресії

### 6.1 Стислі теоретичні відомості

Нехай ми маємо n спостережень для трифакторної регресійної моделі $Y = F(X_1, X_2, X_3)$, де $X_1, X_2, X_3$ – фактори, $y$ – відклик. Спостереження зведені в таблиці 3.

*Таблиця 3*

| № п/п | $X_1$ | $X_2$ | $X_3$ | Y |
|---|---|---|---|---|
| 1 | $x_{11}$ | $x_{21}$ | $x_{31}$ | $y_1$ |
| 2 | $x_{12}$ | $x_{22}$ | $x_{32}$ | $y_2$ |
| … | … | … | … | … |
| … | … | … | … | … |
| n | $x_{1n}$ | $x_{2n}$ | $x_{3n}$ | $y_n$ |

Числа, що записані у стовпцях, можна витлумачувати як координати n-мірних векторів:

$$\overline{X}_i = \begin{pmatrix} x_{11} \\ x_{12} \\ \dots \\ \dots \\ x_{1n} \end{pmatrix}, \ \text{i=1,2,3;} \qquad \overline{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ \dots \\ y_n \end{pmatrix}.$$

*Колінеарність* двох векторів означає, що один з них можна виразити через інший множенням на постійний, відмінний від нуля множник, наприклад: $\overline{X}_1 = k\overline{X}_2, \ k \neq 0.$

*Мультиколінеарність* декількох векторів означає, що один з них можна виразити через лінійну комбінацію інших, наприклад: $\overline{X}_1 = k_2\overline{X}_2 + k_3\overline{X}_3$, причому не всі множники $k_i$ дорівнюють нулю.

Колінеарність двох векторів $\overline{X}_i$ і $\overline{X}_j$ перевіряється за модулем коефіцієнта їхньої кореляції $r_{ij}$ : якщо $r_{ij}$ =1, то вектори колінеарні. У випадку колінеарності двох векторів рівняння лінійної регресії за методом найменших квадратів знайти неможливо. На практиці звичайно зустрічається неточна колінеарність, коли $r_{ij}$ близьке до одиниці. У випадку неточної колінеарності двох векторів рівняння лінійної регресії за методом найменших квадратів знайти можна, але точність визначення його коефіцієнтів низька. Крім того, критерії Фішера і Стьюдента в цьому випадку працюють погано.

Мультиколінеарність за коефіцієнтами парної кореляції $r_{ij}$ перевірити не можна. Застосовують інші, більш складні, методи перевірки. Але наслідки мультиколінеарності такі ж, як і при колінеарності.

### 6.2 Мета практичної роботи

Мають бути придбані наступні вміння:

1) побудування лінійної і нелінійної моделей для багатофакторної регресії;

2) перевірка факторів на колінеарність і мультиколінеарність;

3) знаходження еластичності відносно кожного з факторів.

Мають бути засвоєні наступні поняття: колінеарність і мультиколінеарність, їхні наслідки; часткова еластичність, її економічний зміст.

Робота розрахована на 4 години.

### 6.3 Завдання до практичної роботи

1) Перевірити фактори $x_1$ і $x_2$ на мультиколінеарність.

2) Побудувати дві моделі: лінійну і степеневу модель Кобба-Дугласа. При рівні значущості $\alpha$=0,05 перевірити коефіцієнти рівнянь на значущість; перевірити моделі на адекватність.

3) За мінімумом суми квадратів залишків вибрати оптимальну модель.

4) Знайти еластичність моделі відносно змінних $x_1$ і $x_2$ і пояснити її економічний зміст.

### 6.4 Зміст звіту

Звіт про практичну роботу повинен містити:

1) Тему роботи, завдання.
2) Роздрук таблиць і графіків.
3) Пояснення отриманих значень коефіцієнтів з погляду економетрики.
4) Пояснення вибору моделі і змісту еластичності.

### 6.5 Приклад виконання практичної роботи у пакеті Statistica6

*Економічні дані*

Реальний обсяг випуску продукції (Y, млн т) і рівні факторів, її формувальних, – капітальних витрат (X1, млн грн) і питомої ваги простоїв устаткування (X2, %) по металургійних підприємствах країни за минулий рік задані в таблиці 4.

*Таблиця 4*

| № | X1 | X2 | Y |
|---|---|---|---|
| 1 | 1,033 | 1,45 | 1,83 |
| 2 | 0,012 | 4,295 | 0,58 |
| 3 | 0,045 | 3,553 | 1,34 |
| 4 | 0,243 | 1,568 | 1,34 |
| 5 | 0,266 | 1,52 | 1,64 |
| 6 | 0,302 | 0,512 | 1,65 |
| 7 | 0,451 | 0,457 | 1,91 |
| 8 | 1,041 | 1,822 | 1,96 |
| 9 | 1,423 | 0,442 | 2,08 |
| 10 | 1,914 | 0,498 | 2,18 |

Вихідна таблиця даних (рис. 47) вставляється у звіт так само, як у попередніх роботах.

| | 1<br>X1 | 2<br>X2 | 3<br>Y |
|---|---|---|---|
| 1 | 1,033 | 1,45 | 1,830 |
| 2 | 0,012 | 4,295 | 0,580 |
| 3 | 0,045 | 3,553 | 1,340 |
| 4 | 0,243 | 1,568 | 1,340 |
| 5 | 0,266 | 1,52 | 1,640 |
| 6 | 0,302 | 0,512 | 1,650 |
| 7 | 0,451 | 0,457 | 1,910 |
| 8 | 1,041 | 1,822 | 1,960 |
| 9 | 1,423 | 0,442 | 2,080 |
| 10 | 1,914 | 0,498 | 2,180 |

*Рисунок 47*

*Виконання завдання*

1) Перевіряємо фактори на мультиколінеарність.

Тому що факторів всього два, їх варто перевіряти на колінеарність за значенням парного коефіцієнта кореляції $r_{x_1x_2}$. Для цього створюємо кореляційну таблицю (рис. 48): активуємо таблицю даних – **Statistics – Basic Statistics/Tables – Correlation matrices – OK – Two lists(rect.matrix) – X1 – X2 – OK – Summary**: Correlation matrix.

| | Correlations (lab5)<br>Marked correlations are significant at p < ,05000<br>N=10 (Casewise deletion of missing data) |
|---|---|
| Variable | **X2** |
| **X1** | -0,58 |

*Рисунок 48*

Парний коефіцієнт кореляції $r_{x_1x_2}$ = -0,58. Визначимо за допомогою критерію Стьюдента, чи є це значення коефіцієнта статистично значущім.

Знаходимо спостережуване значення критерію Стьюдента $t_{набл} = r_{x_1x_2}((n-2)/(1-r_{x_1x_2}^2))^{0,5}$ = -2,014. Критичне значення $t_{кр}$ визначається за допомогою імовірнісного калькулятора при рівні значущості $\alpha$=0,05 і числі ступенів волі n-2=8 (див. лаб. роботу 3): t(8)=2,306004; p=0,05.

Отже, $t_{кр}$ = 2,3.

Тому що abs ($t_{спост}$)< $t_{кр}$, коефіцієнт кореляції статистично не значущий. Отже, фактори $x_1$ і $x_2$ неколінеарні.

2) Будуємо дві моделі: лінійну $y=b_0+b_1x_1+b_2x_2$ і степеневу модель Кобба-Дугласа $y=AX_1^{a1}X_2^{a2}$. Перевіримо коефіцієнти рівнянь на значущість. Перевіримо моделі на адекватність.

Спочатку знайдемо рівняння лінійної регресії.

Для оцінки коефіцієнтів лінійної регресії $b_0$, $b_1$ і $b_2$ треба виділити таблицю даних – **Statistics - Multiple Regression (Множинна регресія) – Variables –** (dependent Y - independent X1, X2) – **Ok – Ok – Summary: Regression results** (вибір змінних Y,X1,X2 – Ok – Ok – Підсумки регресійного аналізу) ((рис. 49).

У стовпці B таблиці, що з'явилася, взяти параметри $b_0$, $b_1$і $b_2$.

Regression Summary for Dependent Variable: Y (lab5)
R= ,91924017 R?= ,84500249 Adjusted R?= ,80071749
F(2,7)=19,081 p<,00147 Std.Error of estimate: ,21039

| N=10 | Beta | Std.Err. of Beta | B | Std.Err. of B | t(7) | p-level |
|---|---|---|---|---|---|---|
| **Intercept** | | | 1,768423 | 0,184378 | 9,59128 | 0,000028 |
| X1 | 0,440721 | 0,182668 | 0,322369 | 0,133614 | 2,41269 | 0,046589 |
| X2 | -0,590613 | 0,182668 | -0,207469 | 0,064167 | -3,23327 | 0,014387 |

*Рисунок 49*

Перевіримо параметри $b_0$, $b_1$і $b_2$ на значущість. Для цих параметрів значення критерію Стьюдента, що спостерігаються, задані в стовпці t(7): $t_{спост}(b_0)$=9,59128, $t_{спост}(b_1)$=2,41269, $t_{спост}(b_2)$= -3,23327. Критичне значення критерію Стьюдента при рівні значущості $\alpha$=0,05 і числі степенів вільності k=7 знаходимо за допомогою імовірнісного калькулятора (див. лаб. роботу 3): $t_{кр}(7)$=2,364624. Тому що для всіх трьох коефіцієнтів $b_0$, $b_1$ і

$b_2$ abs($t_{спост}$)>$t_{кр,}$ всі три коефіцієнти значущі. Лінійне рівняння регресії: $y=1,768+0,322x_1-0,207x_2$.

Для перевірки рівняння на адекватність у шапці таблиці зчитуємо значення критерію Фішера, що спостерігається, $F_{спост}$ = F(2,7)=19,081, число степенів вільності критерію Фішера: k1=2 і k2=7. $F_{кр}$ знаходимо при рівні значущості $\alpha$=0,05, використовуючи імовірнісний калькулятор: $F_{кр}$ = F(2;7)=4,737416 (див. лабораторну роботу 3). Тому що $F_{спост}>F_{кр}$, рівняння лінійної регресії адекватно.

Знайдемо рівняння нелінійної (степеневої) регресії.

Щоб знайти рівняння степеневої моделі потрібно:

а) зробити лінеаризацію вибірки за формулами: V=ln(Y), U1=ln(X1), U2=ln(X2);

б) знайти рівняння лінійної регресії для змінних: V, U1, U2;

в) повернутися до вихідних змінних: Y, X1, X2.

Лінеаризуємо вибірку. Додаємо нові змінні і обчислюємо їхні значення за наведеними формулами (рис. 50). У пакеті Statistica натуральний логарифм ln(x) записується: log(x).

| | 1<br>X1 | 2<br>X2 | 3<br>Y | 4<br>U1 | 5<br>U2 | 6<br>V |
|---|---|---|---|---|---|---|
| 1 | 1,033 | 1,45 | 1,830 | 0,032 | 0,372 | 0,604 |
| 2 | 0,012 | 4,295 | 0,580 | -4,423 | 1,457 | -0,545 |
| 3 | 0,045 | 3,553 | 1,340 | -3,101 | 1,268 | 0,293 |
| 4 | 0,243 | 1,568 | 1,340 | -1,415 | 0,450 | 0,293 |
| 5 | 0,266 | 1,52 | 1,640 | -1,324 | 0,419 | 0,495 |
| 6 | 0,302 | 0,512 | 1,650 | -1,197 | -0,669 | 0,501 |
| 7 | 0,451 | 0,457 | 1,910 | -0,796 | -0,783 | 0,647 |
| 8 | 1,041 | 1,822 | 1,960 | 0,040 | 0,600 | 0,673 |
| 9 | 1,423 | 0,442 | 2,080 | 0,353 | -0,816 | 0,732 |
| 10 | 1,914 | 0,498 | 2,180 | 0,649 | -0,697 | 0,779 |

*Рисунок 50*

Побудування рівняння лінійної регресії для змінних V, U1, U2 виконується так само, як для змінних Y, X1, X2: виділити таблицю даних – **Statistics – Multiple Regression** (Множинна регресія) – **Variables** – (dependent V – independent U1, U2) – **Ok – Ok – Summary: Regression**

**results** (вибір змінних Y,X1,X2 – Ok – Ok – Підсумки регресійного аналізу) (рис. 51).

У стовпці B таблиці 28, що з'явилася, взяти параметри $b_0$, $b_1$і $b_2$.

| N=10 | Regression Summary for Dependent Variable: V (lab5)<br>R= ,92416891 R?= ,85408818 Adjusted R?= ,81239908<br>F(2,7)=20,487 p<,00119 Std.Error of estimate: ,16738 | | | | | |
|---|---|---|---|---|---|---|
| | **Beta** | Std.Err. of Beta | B | Std.Err. of B | t(7) | p-level |
| **Intercept** | | | 0,683253 | 0,071070 | 9,613774 | 0,000028 |
| U1 | 0,846927 | 0,213443 | 0,204528 | 0,051545 | 3,967924 | 0,005407 |
| U2 | -0,101413 | 0,213443 | -0,045885 | 0,096574 | -0,475127 | 0,649163 |

*Рисунок 51*

Два коефіцієнти $b_0$ і $b_1$ значущі при рівні значущості $\alpha$=0,05, тому що для них abs($t_{спост}$) > $t_{кр}$: 9,613774 > 2,364624, 3,967924 > 2,364624, а коефіцієнт $b_2$ не значущий, тому що для нього abs ($t_{набл}$) < $t_{кр}$: 0,475127<2,364624. Проте, якщо рівняння лінійної регресії V=0,683+0,204U1-0,045U2 адекватне, член –0,045U2 у цьому рівнянні варто зберегти, тому що вилученням доданка – 0,045U2 може порушитися специфікація моделі.

Перевіряємо лінеаризовану модель на адекватність. Значення критерію Фішера, що спостерігається, $F_{спост}$=20,487, число степенів вільності критерію Фішера k1=2 і k2=7. $F_{kr}$ знаходимо при рівні значущості $\alpha$=0,05, використовуючи імовірнісний калькулятор: F(2;7)=4,737416.

Тому що $F_{спост} > F_{кр}$ , лінеаризована модель адекватна.

в) Повертаємося до вихідних змінних:

$$A=e^{b_0},\ b_1=a_1,\ b_2=a_2,$$

$$y=1{,}97X_1^{0{,}204}X_2^{-0{,}045}.$$

3) Виберемо оптимальну модель за мінімумом суми квадратів залишків.

Суму квадратів залишків лінійної моделі знаходимо, як у лабораторній роботі 2: Analysis of Variance (аналіз залишків) і в таблиці,

що з'явилася (рис. 52), у рядку Residual (залишок) вибираємо суму квадратів залишків 0, 309854.

| Effect | Analysis of Variance; DV: Y (lab5) | | | | |
|---|---|---|---|---|---|
| | **Sums of Squares** | df | Mean Squares | F | p-level |
| **Regress.** | 1 689236 | 2 | 0,844618 | 19,08´01 | 0,001466 |
| Residual | 0 309854 | 7 | 0,044265 | | |
| Total | 1 999090 | | | | |

*Рисунок 52*

Для розрахунку квадратів залишків для степеневої моделі доповнюємо таблицю двома стовпцями: Y_NELIN= 1,97*X1^0,204*X2^(-0,046) і KV_OST_ NELIN =( Y_NELIN-Y)^2 (рис. 53).

| | 1 X1 | 2 X2 | 3 Y | 4 U1 | 5 U2 | 6 V | 7 Y_NELIN | 8 KV_OST_NELIN |
|---|---|---|---|---|---|---|---|---|
| 1 | 1,033 | 1,45 | 1,830 | 0,032 | 0,372 | 0,673 | 1,960 | 0,017 |
| 2 | 0,012 | 4,295 | 0,580 | -4,423 | 1,457 | -0,285 | 0,752 | 0,030 |
| 3 | 0,045 | 3,553 | 1,340 | -3,101 | 1,268 | -0,007 | 0,993 | 0,120 |
| 4 | 0,243 | 1,568 | 1,340 | -1,415 | 0,450 | 0,374 | 1,454 | 0,013 |
| 5 | 0,266 | 1,52 | 1,640 | -1,324 | 0,419 | 0,394 | 1,483 | 0,025 |
| 6 | 0,302 | 0,512 | 1,650 | -1,197 | -0,669 | 0,469 | 1,598 | 0,003 |
| 7 | 0,451 | 0,457 | 1,910 | -0,796 | -0,783 | 0,556 | 1,743 | 0,028 |
| 8 | 1,041 | 1,822 | 1,960 | 0,040 | 0,600 | 0,664 | 1,943 | 0,000 |
| 9 | 1,423 | 0,442 | 2,080 | 0,353 | -0,816 | 0,792 | 2,207 | 0,016 |
| 10 | 1,914 | 0,498 | 2,180 | 0,649 | -0,697 | 0,847 | 2,332 | 0,023 |

*Рисунок 53*

Суму квадратів залишків степеневої моделі знаходимо, як у лабораторній роботі 2: **Statistics – Basic Statistics/Tables – Descriptive Statistics – Advanced** – залишити прапорець тільки в опції Sum – виділити змінну KV_OST_NELIN – **Summary** (рис. 54).

| Variable | Descriptive Statistics (lab5) |
|---|---|
| | **Sum** |
| **KV_OST_NELIN** | 0,274381 |

*Рисунок 54*

Для степеневої моделі сума квадратів залишків менше, ніж для лінійної: 0,27 < 0,31. Отже, вибираємо степеневу модель.

4) Визначаємо еластичність моделі.

Еластичність моделі Y відносно змінних розраховується за формулами: $E_{x_1} = \frac{x_1}{y} y'_{x_1}$ $E_{x_2} = \frac{x_2}{y} y'_{x_2}$.

Додаємо в таблицю дві змінні: Ex1 і Ex2. Знаходимо вираз для еластичності і вписуємо розрахункові формули в поле Long Name для кожної змінної (рис. 55).

| 11<br>Ex1 | 12<br>Ex2 |
|---|---|
| 0,204 | -0,046 |
| 0,204 | -0,046 |
| 0,204 | -0,046 |
| 0,204 | -0,046 |
| 0,204 | -0,046 |
| 0,204 | -0,046 |
| 0,204 | -0,046 |
| 0,204 | -0,046 |
| 0,204 | -0,046 |
| 0,204 | -0,046 |

*Рисунок 55*

### 6.6 Висновки

Тому що факторів всього два, перевірку на колінеарність виконуємо за значенням парного коефіцієнта кореляції $r_{x_1x_2}$. Парний коефіцієнт кореляції $r_{x_1x_2}$ =-0,58. За допомогою критерію Стьюдента визначаємо статистичну значущість коефіцієнта: при рівні значущості $\alpha$=0,05 abs($t_{спост}$) = 2,01.

Тому що abs ($t_{спост}$) < $t_{кр}$, коефіцієнт кореляції статистично не значущий, отже, фактори $x_1$ і $x_2$ неколінеарні.

Побудовані дві моделі: лінійна y=1,768+0,322$X_1$-0,207$X_2$ і степенева модель Кобба-Дугласа y = 1,97$X_1^{0,204}X_2^{-0,046}$.

Коефіцієнти обох моделей перевірені на статистичну значущість. При рівні значущості $\alpha = 0,05$ всі коефіцієнти лінійної моделі статистично значущі. Для моделі Кобба-Дугласа $y = AX_1^{a1}X_2^{a2}$ коефіцієнти A = 1,97 і a1 = 0,204 статистично значущі, а коефіцієнт a2 = -0,046 статистично не значущий.

При рівні значущості $\alpha = 0,05$ обидва рівняння адекватні.

Суми квадратів залишків для лінійної й статечної моделей рівні, відповідно, 0,31 і 0,27. Тому що для степеневої моделі сума квадратів залишків менше, ніж для лінійної, 0,27 < 0,31, оптимальною є степенева модель.

Для степеневої моделі Кобба-Дугласа часткова еластичність відносно кожного з факторів дорівнює показнику степеня при відповідній змінній: $E_{x_1}$ = 0,204, $E_{x_2}$ =-0,046. Отже, при збільшенні капітальних витрат (X1, млн грн) на 1% випуск продукції (Y, млн т) по металургійних підприємствах країни збільшиться на 0,204%, а при збільшенні питомої ваги простоїв устаткування (X2, %) на 1% випуск продукції (Y, млн т) по металургійних підприємствах країни зменшиться на 0,046%. Незначний вплив ваги простоїв устаткування на випуск продукції випливає також з результатів статистичного аналізу коефіцієнтів рівняння Кобба-Дугласа: показник a2= -0,046 при змінній $x_2$ статистично не значущий.

## ПРАКТИЧНА РОБОТА №6

**Тема**: Аналіз часових рядів.

### 7.1 Стислі теоретичні відомості

*Часовий ряд* – ряд послідовних значень, що характеризують зміну показника за часом.

*Лаг* – економічний показник, що відбиває відставання в часі одного економічного показника в порівнянні з іншим, зв'язаним з ним. Якщо показник x відстає на s періодів, то він записується $x_{t-s}$. *Дистрибутивно-лагові моделі* економічних процесів, що протікають у часі, мають вид

$y_t = a_0+b_0x_t+b_1x_{t-1}+b_2x_{t-2}+\ldots+u_t$, де $u_t$–випадковий член.

Модель є *авторегресійною*, якщо вона містить відклик із запізнюванням. Авторегресійні моделі можуть мати вид:

$y_t = a_0+b_0x_t+b_1y_{t-1}+u_t,$

$y_t = a_0+ b_1y_{t-1}+u_t,$

$y_t = a_0+b_0y_t+b_1y_{t-1}+b_2y_{t-2}+\ldots+u_t.$

Дистрибутивно-лагові моделі можна перетворити в авторегресійні. Дистрибутивно-лагові й авторегресійні моделі описують часові ряди.

*Тренд* – тривала тенденція зміни економічних показників. Це основна складова прогнозованого часового ряду, на яку накладаються сезонні коливання. Якщо сезонні коливання сумуються з трендом, то модель називається дистрибутивною. Якщо сезонні коливання перемножуються з трендом, то модель називається *мультиплікативною*.

Метод ковзного середнього і метод експоненціального згладжування – методи, що застосовуються при аналізі часових рядів. *Метод ковзного середнього* полягає в тому, що кілька значень часового ряду, які йдуть один за одним, заміняються їхнім середнім значенням, наприклад, $(x_t+x_{t-1}+ x_{t-2})/3$. Потім усереднюються доданки, починаючи з $x_{t-1}$ і т.д. *Метод експоненціального згладжування* також полягає в тому, що усереднюються кілька значень часового ряду, які йдуть один за одним, але з вагами $(w_1x_t+w_2x_{t-1} +w_3x_{t-2})$, $w_1+ w_2+ w_3=1$. Ваги $w_1$, $w_2$, $w_3$ беруться у вигляді експоненціальних функцій.

У пакеті STATISTICA використовуються два методи припасування моделі часового ряду: ARIMA і Exponential Smoothing & Forecasting. Метод ARIMA використовує ковзні середні, метод Exponential Smoothing & Forecasting – експоненціальне згладжування. При методі ARIMA контроль за якістю моделі відбувається автоматично, при методі Exponential Smoothing & Forecasting контроль за якістю моделі дослідник повинний робити сам за графіками залишків. Метод Exponential Smoothing & Forecasting (експоненціальне згладжування і прогноз) не будує довірчих інтервалів для зробленого прогнозу. Отже, неможливо розрахувати ризик

при використанні прогнозу. Контроль якості прогнозу ведеться непрямими методами. Для того, щоб прогноз мав 90%-ий довірчий інтервал, необхідно, щоб залишки задовольняли трьом вимогам:

1) У графіку залишків не має бути розгойдування (тобто систематичного зростання амплітуди), як на рисунку 56.

2) Між залишками не має бути залежності (залишки не повинні корелювати один з одним). При кореляції залишків за групою негативних залишків випливає група позитивних залишків, потім знову група негативних залишків і т.д. (рис. 57).

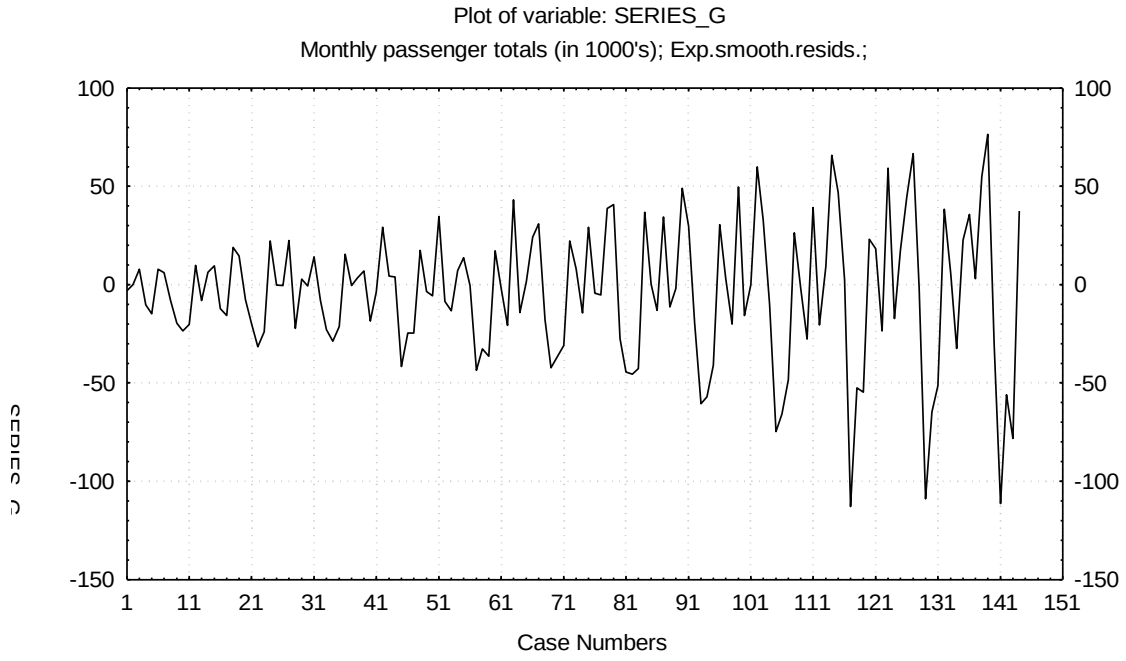


*Рисунок 56*

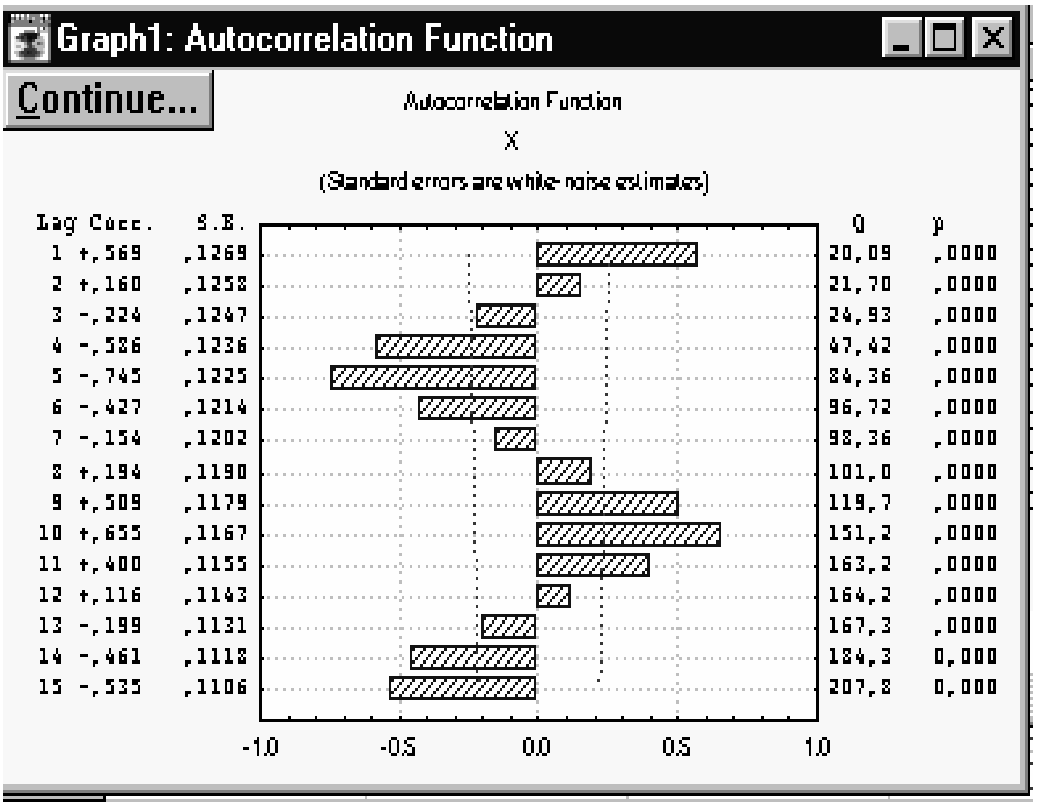


*Рисунок 57*

При правильно припасованій моделі графік автокореляції залишків повинний мати вид, як на рисунку 58: знаки залишків випадково чергуються, і графік автокореляції не виходить за межі вертикальних пунктирних ліній.

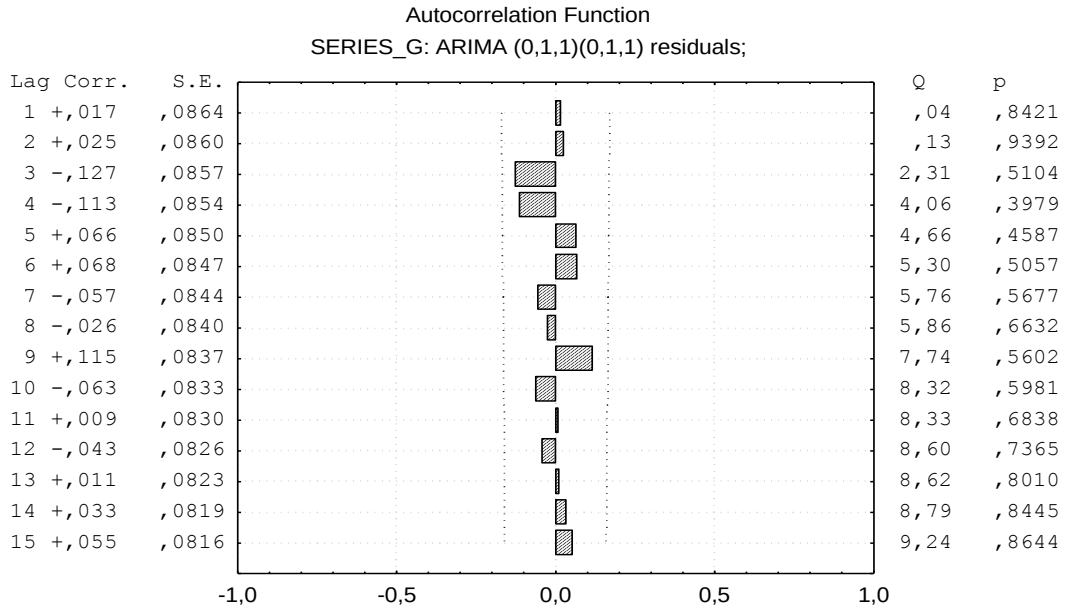


*Рисунок 58*

3) Залишки повинні мати нормальний розподіл: нормально розподілені залишки на імовірнісному папері лягають на пряму лінію, як показано на рисунку 59.

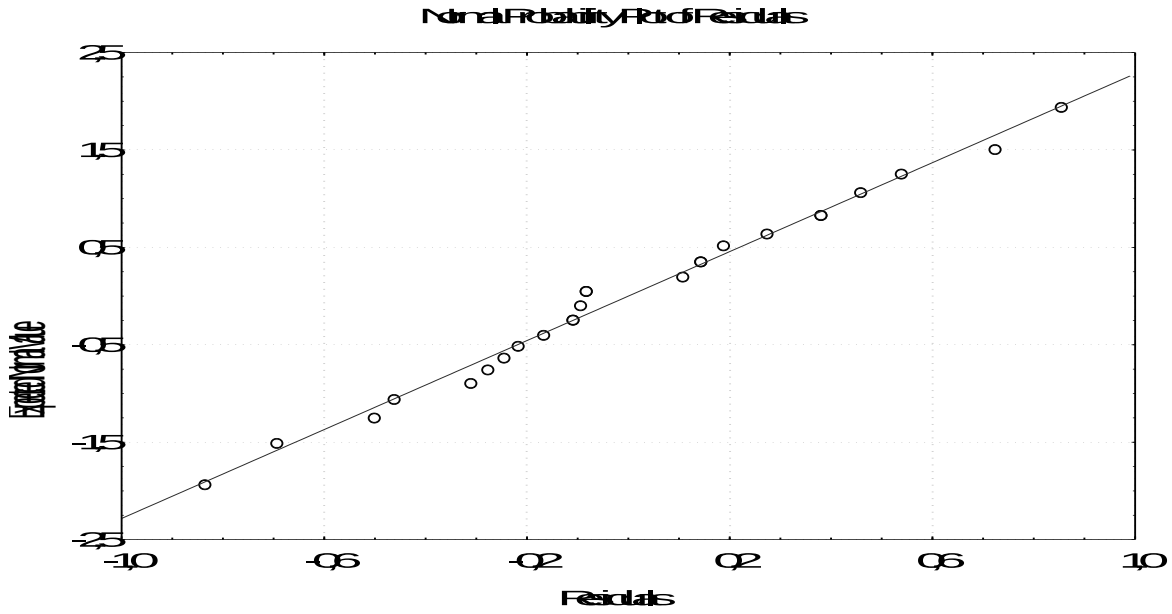


*Рисунок 59*

Прогноз за моделлю, припасованою методом ARIMA, можливий на будь-яку кількість пунктів уперед із заданим рівнем довіри. На графіку прогнозу показані довірчі інтервали.

Прогноз за моделлю, припасованою методом Exponential Smoothing & Forecasting, робиться завжди на 10 пунктів уперед. На графіку прогнозу довірчі інтервали не показані.

Більш якісний прогноз за методом ARIMA вимагає більш великих обсягів вибірки: мінімальний обсяг вибірки – 97. Для прогнозу за методом Exponential Smoothing & Forecasting досить мати вибірку, що містить 61 точку. Exponential Smoothing & Forecasting – метод першого, досить грубого наближення. Його застосовують або при першому дослідженні, або в тому випадку, коли інші методи застосувати не можна.

### 7.2 Мета практичної роботи

Мають бути придбані наступні вміння:

1) за даними файла даних визначити, яка часова модель може бути

використана для прогнозу;

2) з декількох припустимих моделей вибрати кращу на підставі аналізу графіків залишків, графіка автокореляції залишків і графіка залишків на нормальному папері;

3) за остаточно обраною моделлю зробити прогноз і оцінити довірчий інтервал для нього.

Мають бути засвоєні наступні поняття: часовий ряд, лаг, дистрибутивно-лагові моделі, авторегресійні моделі, тренд, дистрибутивні і мультиплікативні моделі, метод ковзного середнього, метод експоненціального згладжування, методи припасування моделі часового ряду в пакеті STATISTICA, їхня порівняльна характеристика, аналіз залишків, прогноз.

Робота розрахована на 6 годин.

### 7.3 Завдання до практичної роботи

1) За даними таблиці 5 визначити, яка часова модель може бути використана для прогнозу.

2) З попередньо обраних моделей вибрати кращу на підставі аналізу трьох графіків: графіку залишків, графіку автокореляції залишків і графіку залишків на нормальному папері.

3) На підставі обраної моделі зробити прогноз на півроку вперед. Узяти лаг часового ряду рівним 12 місяцям.

### 7.4 Зміст звіту

Звіт про практичну роботу повинен містити:

1) Тему роботи, завдання.
2) Роздрук таблиць і графіків.
3) Пояснення вибору моделі з погляду економетрики.
4) Прогноз.

## 7.5 Приклад виконання практичної роботи в пакеті Statistica6

*Економічні дані*

Відомості про щомісячні перевезення вугілля за період із січня 1995 р. по січень 2000 р.(млн т/міс.) задані в таблиці 5.

За даними треба визначити, яка часова модель може бути використана для прогнозу, зробити прогноз на півроку за обраною моделлю.

Тому що обсяг вибірки недостатній для використання методу ARIMA, застосовуємо метод Exponential Smoothing & Forecasting.

Таблицю початкових даних вставляємо в звіт (1 змінна, 61 випадок):

*Таблиця 5*

| № | x | № | x | № | x |
|---|---|---|---|---|---|
| *1* | *2* | *3* | *1* | *2* | *3* |
| 1 | 14 | 21 | 29 | 41 | 114 |
| 2 | 28 | 22 | 10 | 42 | 125 |
| 3 | 25 | 23 | 36 | 43 | 138 |
| 4 | 17 | 24 | 41 | 44 | 106 |
| 5 | 31 | 25 | 46 | 45 | 87 |
| 6 | 44 | 26 | 74 | 46 | 68 |
| 7 | 44 | 27 | 59 | 47 | 90 |
| 8 | 32 | 28 | 68 | 48 | 92 |
| 9 | 15 | 29 | 74 | 49 | 92 |
| 10 | 14 | 30 | 95 | 50 | 132 |
| 11 | 14 | 31 | 95 | 51 | 131 |
| 12 | 11 | 32 | 80 | 52 | 125 |
| 13 | 22 | 33 | 58 | 53 | 139 |
| 14 | 37 | 34 | 42 | 54 | 160 |
| 15 | 31 | 35 | 62 | 55 | 168 |
| 16 | 21 | 36 | 67 | 56 | 133 |
| 17 | 45 | 37 | 76 | 57 | 107 |
| 18 | 66 | 38 | 89 | 58 | 76 |
| 19 | 66 | 39 | 77 | 59 | 97 |

Продовження таблиці 5

| *1* | *2* | *3* | *1* | *2* | *3* |
|---|---|---|---|---|---|
| 20 | 54 | 40 | 79 | 60 | 100 |
| | | | | 61 | 84 |

Модель знаходимо методом експоненціального згладжування.

*Виконання завдання*

1) Вибір групи моделей. Щоб вибрати модель, потрібно побудувати графік вихідних даних: **Statistics**(Статистика) – **Advanced Linear/Nonlinear Models** (Додатково лінійні/нелінійні моделі) – **Time Series/Forecasting** (Часові ряди/Прогноз) – потім нажимати кнопку в правому верхньому куті діалогових вікон, що послідовно з'являються, – **Ok** (transformations, autocorrelations, ...plots) , **Ok** (Transform selected series). З'явиться графік змінної X+0.000 (рис. 60).

Вертаємося у вікно «Time series Analysis:...», вибираємо кнопку Exponential smoothing & forecasting, вкладку **Advanced** на стартову панель і порівнюємо вид цього графіка із пропонованими моделями. У діалоговому вікні, що з'явилося (рис. 62) пропонується 12 моделей тимчасових рядів.


*Рисунок 60*

З порівняння рисунків моделей із графіком змінної видно, що можуть підійти 4 моделі: дві адитивні з лінійним і експонентним трендом і дві мультиплікативні з лінійним і експонентним трендом.

2) Вибір кращої моделі із групи моделей. Перевірка обраних моделей за залишками.

Щоб перевірити за залишками якість обраної моделі, потрібно виконати наступні дії:

1) **Statistics**(Статистика) – **Advanced Linear/Nonlinear Models** (Додатково лінійні/нелінійні моделі) – **Time Series/Forecasting** (Прогноз/серія часу) – **Exponential smoothing & forecasting** – **Advanced**.

2) Вибрати потрібну модель. Для нашої задачі встановити лаг, який дорівнює 12 (опція праворуч угорі над рисунками моделей (див.рис. 62)).

3) Натиснути кнопку Grid search.

4) У вікні, що з'явилося, **Seasonal and Non-Seasonal Exponential Smoothing,** натиснути кнопку Perform grid search. STATISTICA6 розрахує три параметри – Alpha, Delta, Gamma, необхідні для побудови прогнозу. У таблиці, що з'явилася, вони займають верхній рядок (рис. 61).

Parameter grid search (Smallest abs. errors are highlighted) (lab_6)
Model: Linear trend, add.season (12); S0=9,542 T0=2,076
X

| Model Number | Alpha | Delta | Gamma | Mean Error | Mean Abs Error | Sums of Squares | Mean Squares | Mean % Error | Mean Abs % Error |
|---|---|---|---|---|---|---|---|---|---|
| **487** | 0,700000 | 0,100000 | 0,100000 | -0,723808 | 6,633651 | 4911,390 | 80,51459 | -0,000000 | -0,000000 |
| 568 | 0,800000 | 0,100000 | 0,100000 | -0,660097 | 6,784163 | 4952,425 | 81,18729 | -0,000000 | -0,000000 |
| 406 | 0,600000 | 0,100000 | 0,100000 | -0,807494 | 6,727871 | 4952,531 | 81,18904 | -0,000000 | -0,000000 |
| 649 | 0,900000 | 0,100000 | 0,100000 | -0,610889 | 7,035930 | 5054,889 | 82,86703 | -0,000000 | -0,000000 |
| 577 | 0,800000 | 0,200000 | 0,100000 | -0,664250 | 6,867598 | 5062,114 | 82,98547 | -0,000000 | -0,000000 |
| 496 | 0,700000 | 0,200000 | 0,100000 | -0,729406 | 6,735157 | 5064,980 | 83,03246 | -0,000000 | -0,000000 |
| 407 | 0,600000 | 0,100000 | 0,200000 | -0,735348 | 6,842459 | 5077,036 | 83,23011 | -0,000000 | -0,000000 |
| 488 | 0,700000 | 0,100000 | 0,200000 | -0,646265 | 6,878996 | 5099,700 | 83,60164 | -0,000000 | -0,000000 |
| 658 | 0,900000 | 0,200000 | 0,100000 | -0,613270 | 7,076913 | 5114,272 | 83,84052 | -0,000000 | -0,000000 |
| 325 | 0,500000 | 0,100000 | 0,100000 | -0,916076 | 6,998098 | 5117,252 | 83,88937 | -0,000000 | -0,000000 |

*Рисунок 61*

5) Значення Alpha=0,7, Delta=0,1, Gamma=0,1 потрібно ввести у відповідні віконця у вікні **Seasonal and Non-Seasonal Exponential Smoothing,** вкладка **Advanced** (рис. 62).

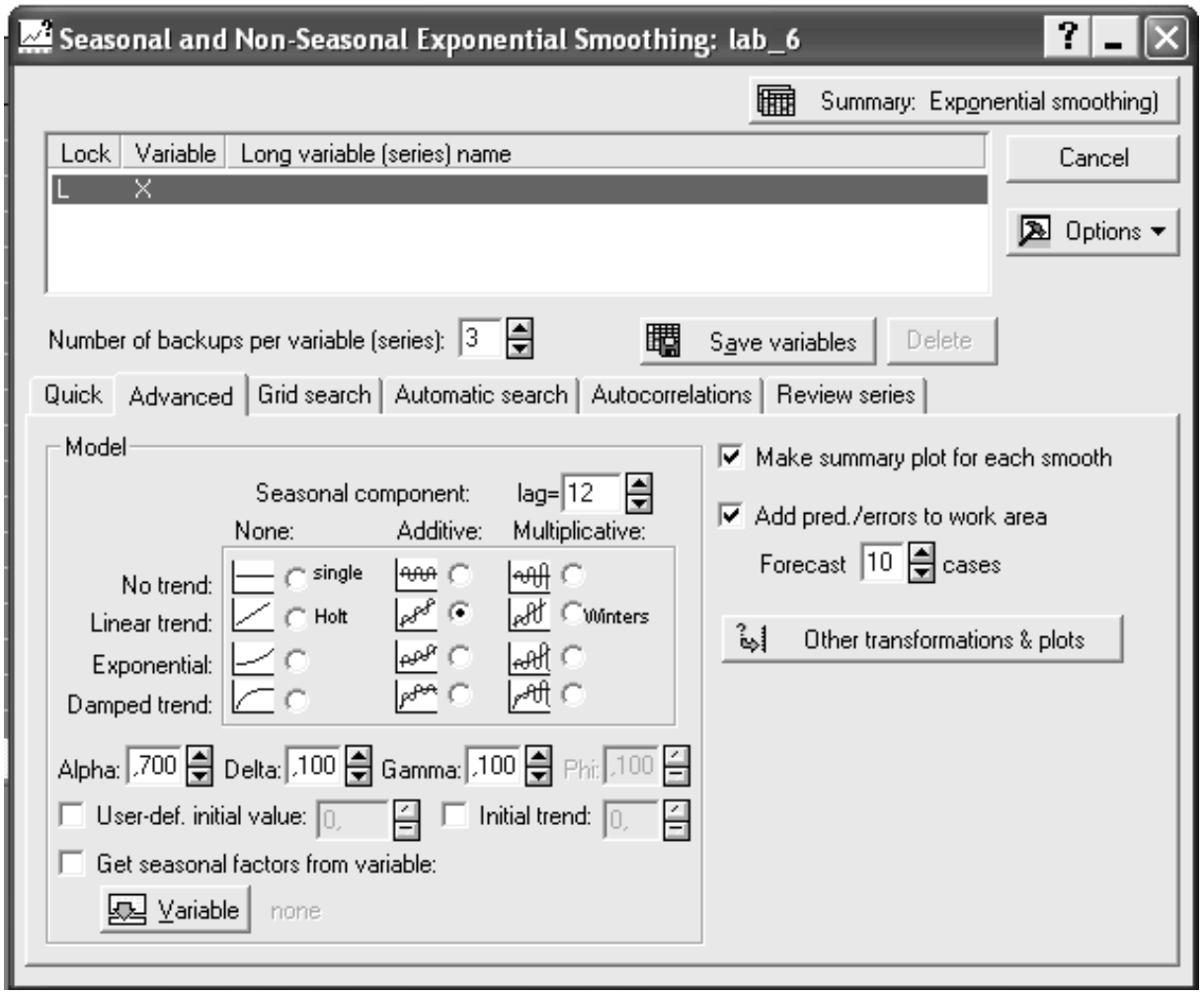


*Рисунок 62*

6) Натиснути праву верхню кнопку **Summary**: Exponential smoothing. STATISTICA6 роздрукує прогноз і залишки в таблиці Exp. Smoothing (експонентний прогноз), а також графік, на якому суцільною лінією нанесені вихідні дані, пунктирною – прогноз, а точковою – графік залишків.

7) Щоб одержати можливість надрукувати потрібні нам графіки для залишків, повернемося у вікно **Seasonal and Non-Seasonal Exponential Smoothing** (сезонне й несезонне експонентне згладжування). Вибираємо змінну *x* Exp.smooth.resids, після чого можна побудувати кожний з потрібних графіків:

а) щоб побудувати графік залишків, натиснути Review series, верхню кнопку Plot (графік) (рис. 63);

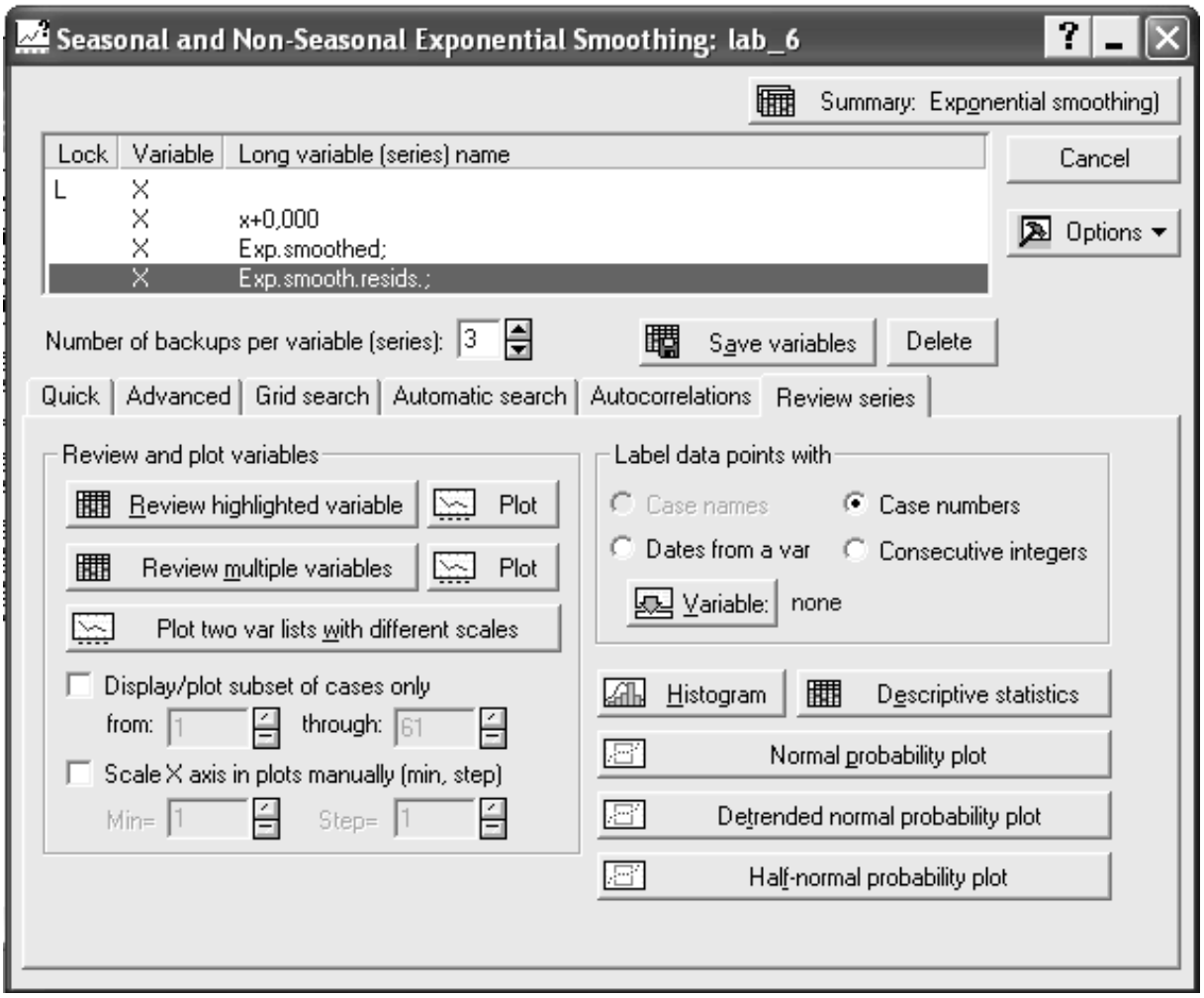


*Рисунок 63*

б) щоб побудувати графік автокореляції, нажати кнопки Autocorrelations (автокореляція), Autocorrelations (рис. 64);

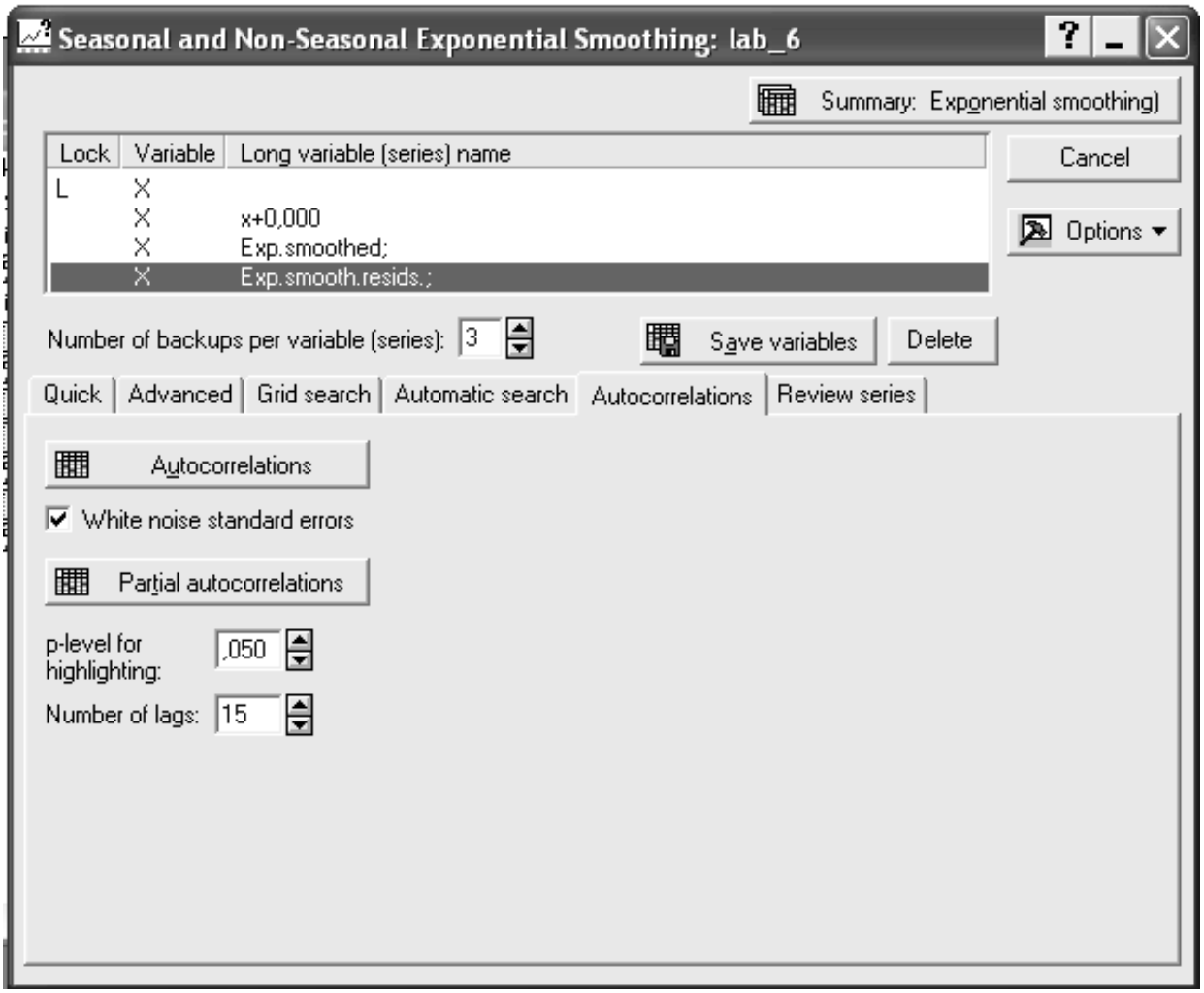


*Рисунок 64*

в) щоб побудувати графік на нормальному папері, нажати кнопки Review series, Normal probability plot (нормальний графік) (див. рис. 63).

У такий спосіб потрібно перевірити кожну із чотирьох моделей, проходячи заново шлях .1) – 7).

*Перевірка обраних моделей 1- 4 за графіками залишків*

Модель1: адитивна, тренд лінійний, зростаючий.

Графік залишків показаний на рисунку 65.

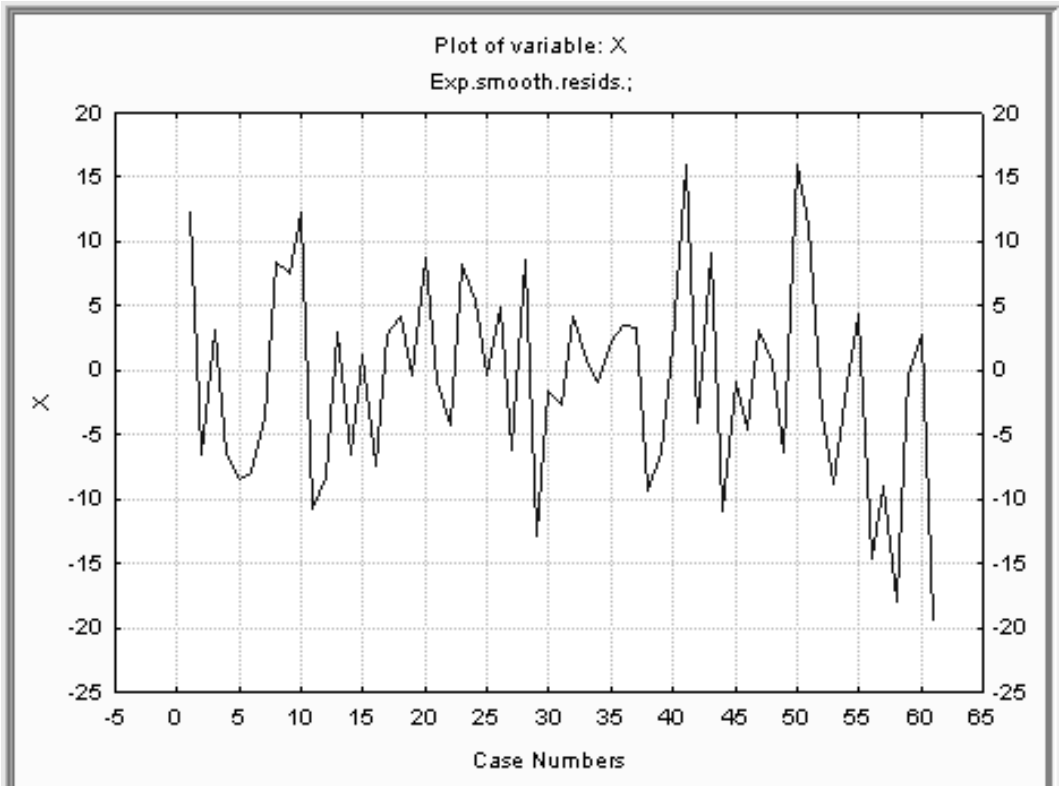


*Рисунок 65*

Графік автокореляційної функції показаний на рисунку 66.

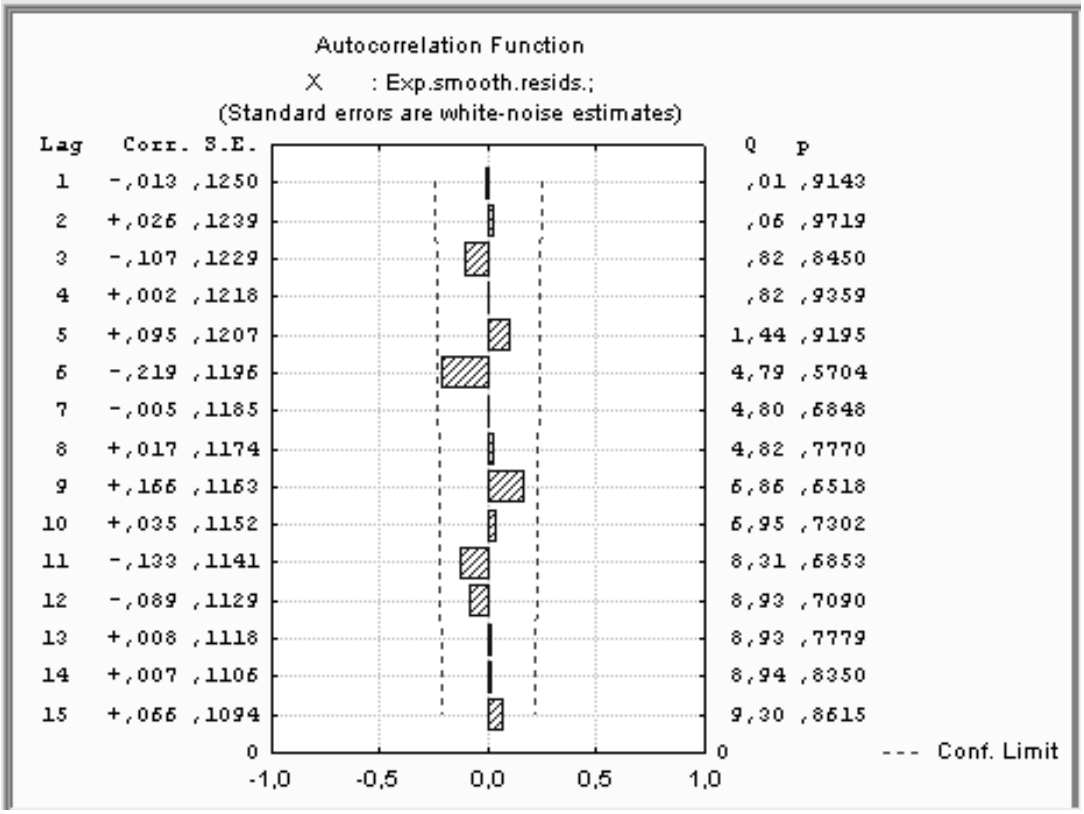


*Рисунок 66*

Графік на нормальному папері показаний на рисунку 67.

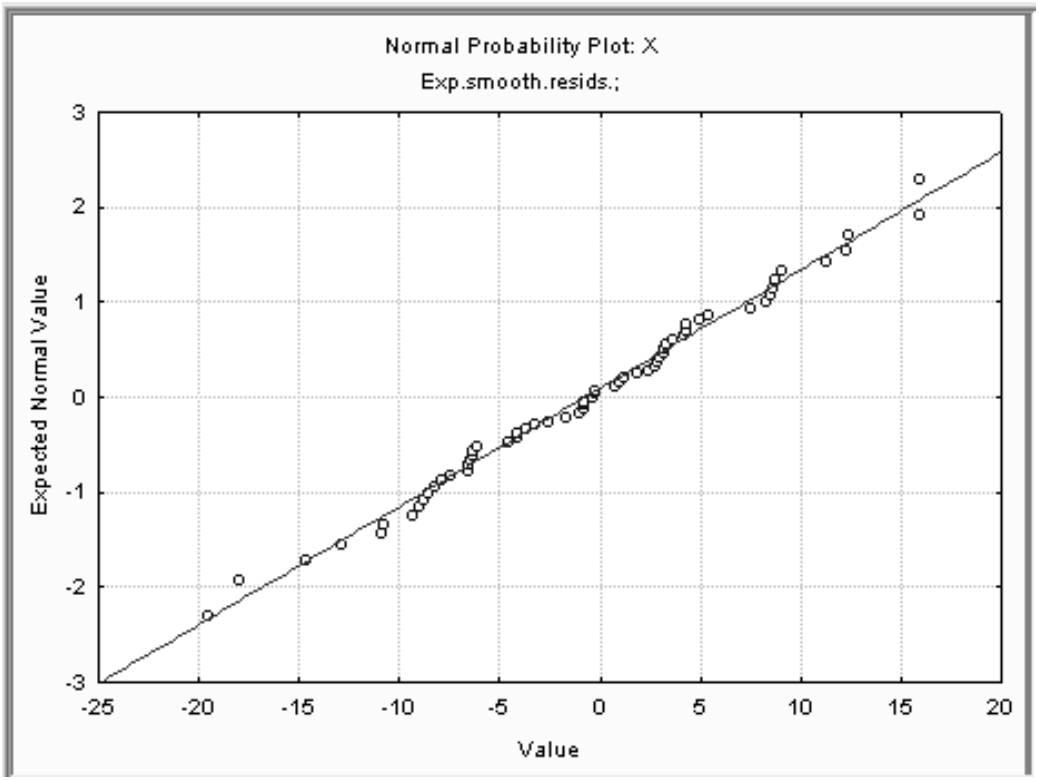


*Рисунок 67*

Модель2: аддитивна, тренд експонентний, зростаючий.

Графік залишків показаний на рисунку 68.

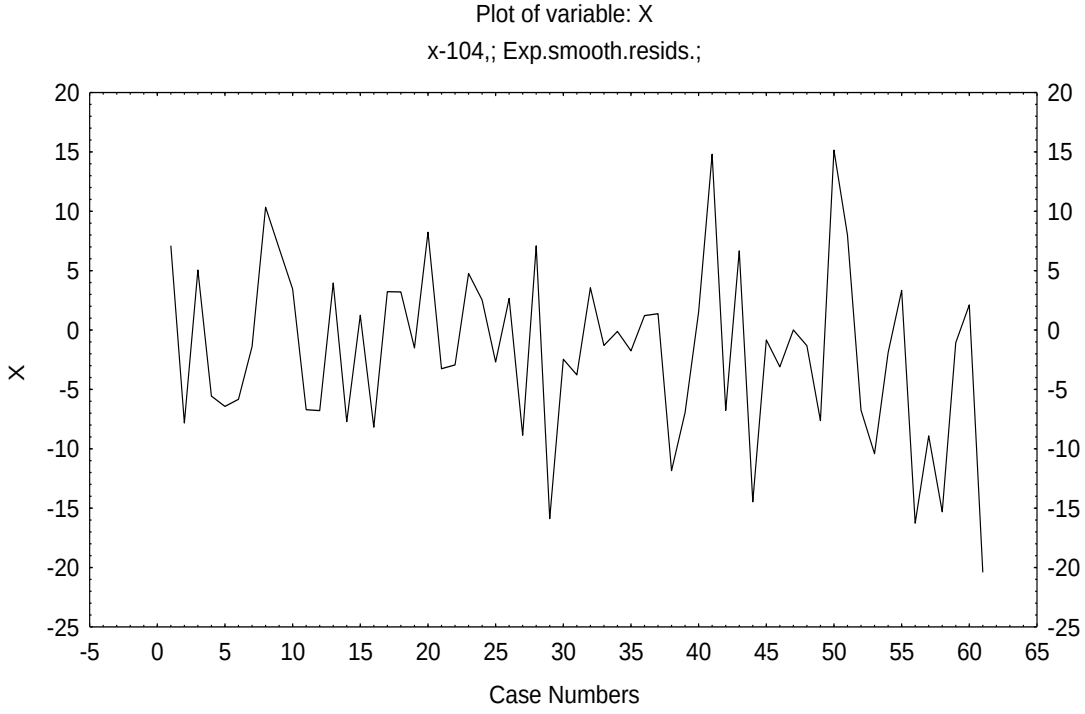


*Рисунок 68*

Графік автокореляційної функції показаний на рисунку 69.

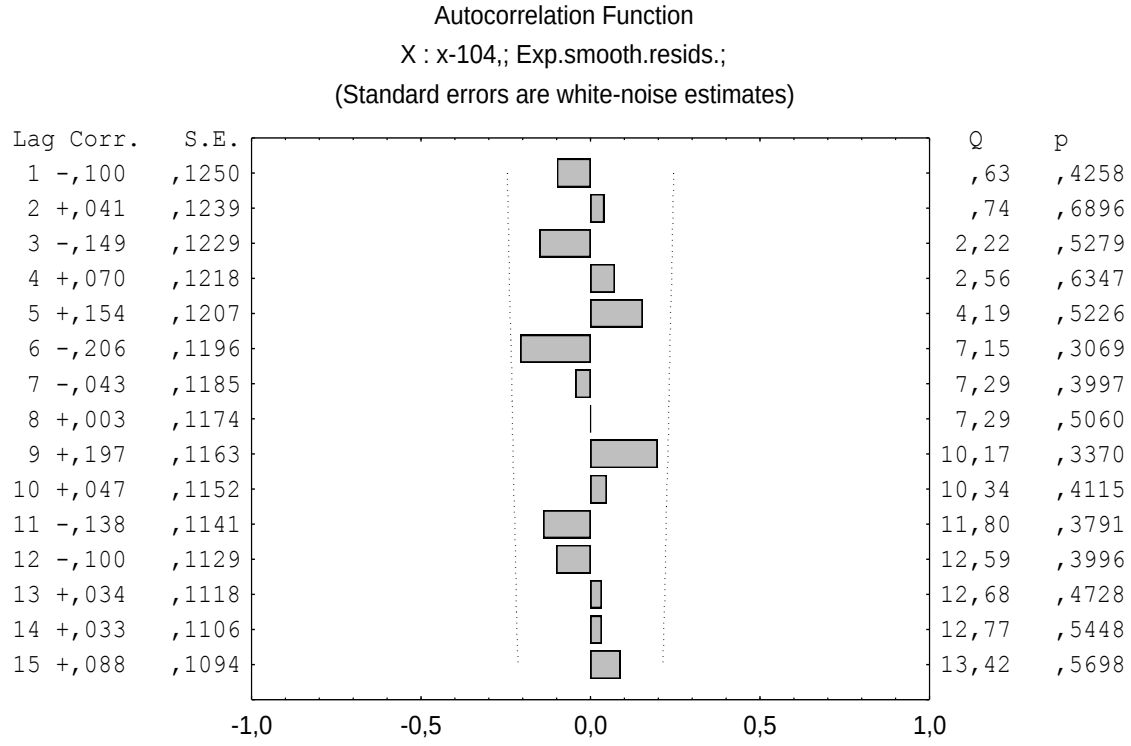


*Рисунок 69*

Графік на нормальному папері показаний на рисунку 70.

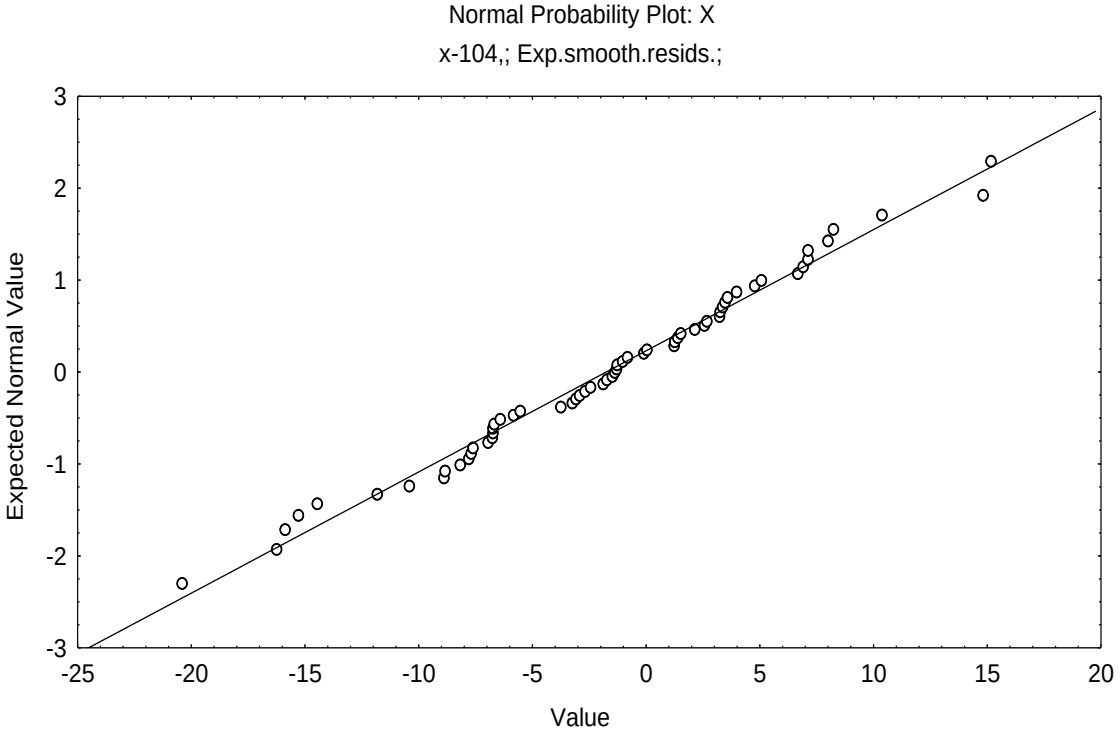


*Рисунок 70*

Модель3: мультиплікативна, тренд лінійний, зростаючий.

Графік залишків показаний на рисунку 71.

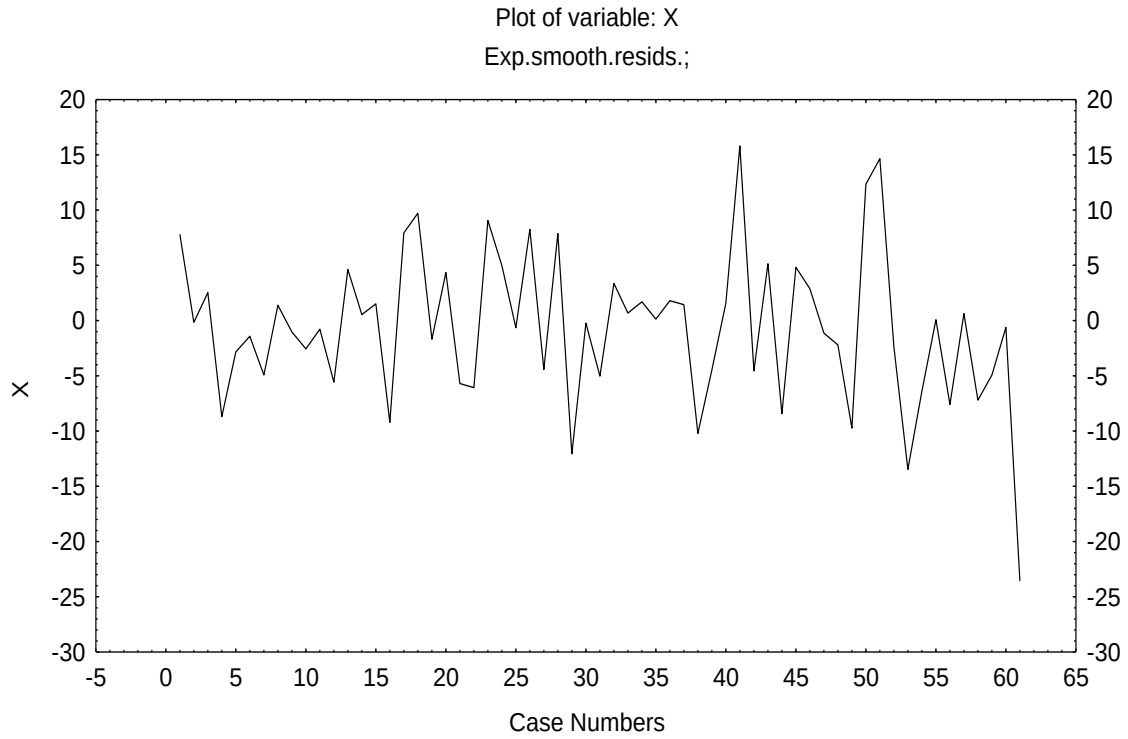


*Рисунок 71*

Графік автокореляційної функції показаний на рисунку 72.

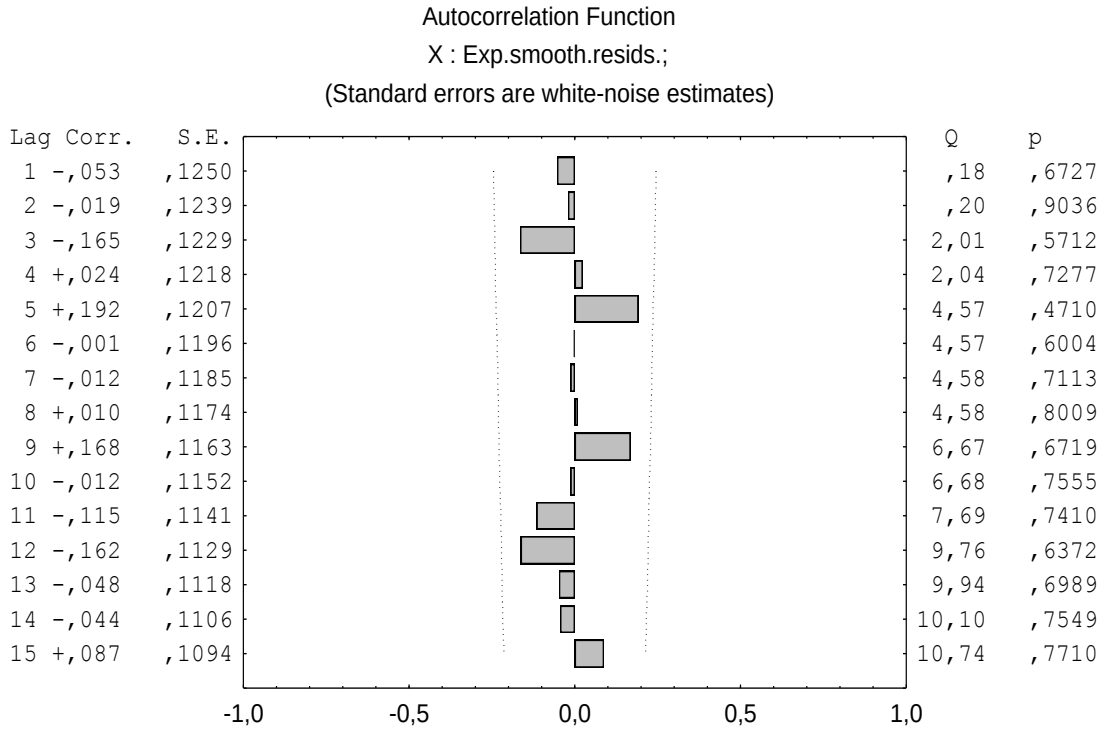


*Рисунок 72*

Графік на нормальному папері показаний на рисунку 73.

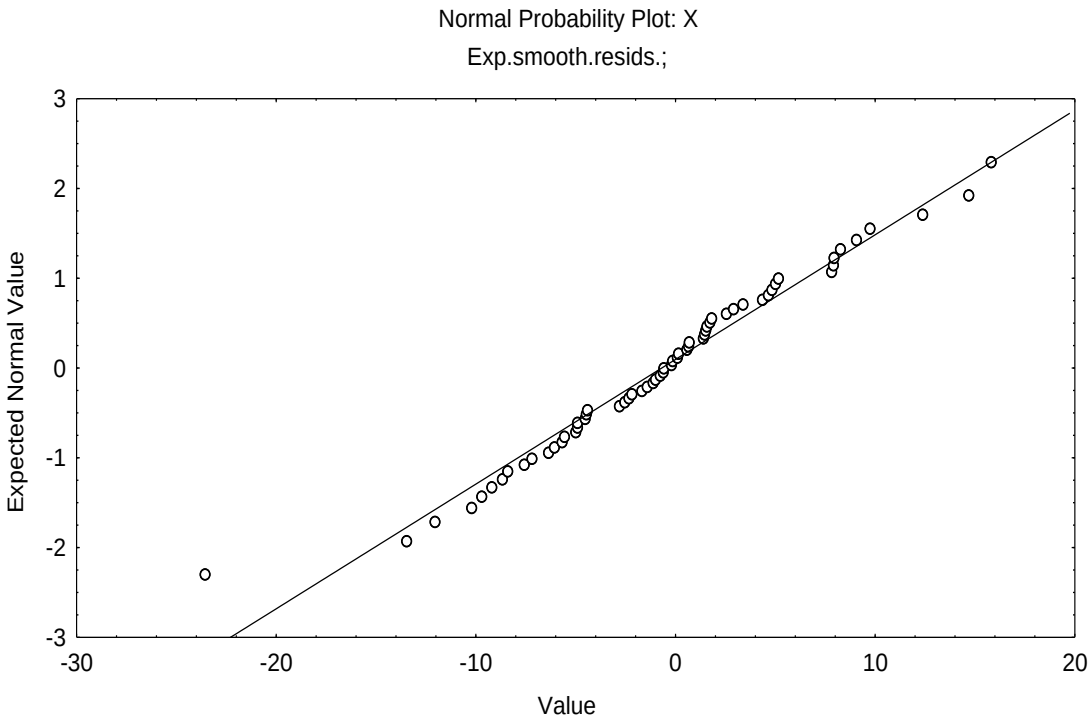


*Рисунок 73*

Модель 4: мультиплікативна, тренд експонентний, зростаючий. Графік залишків показаний на рисунку 74.

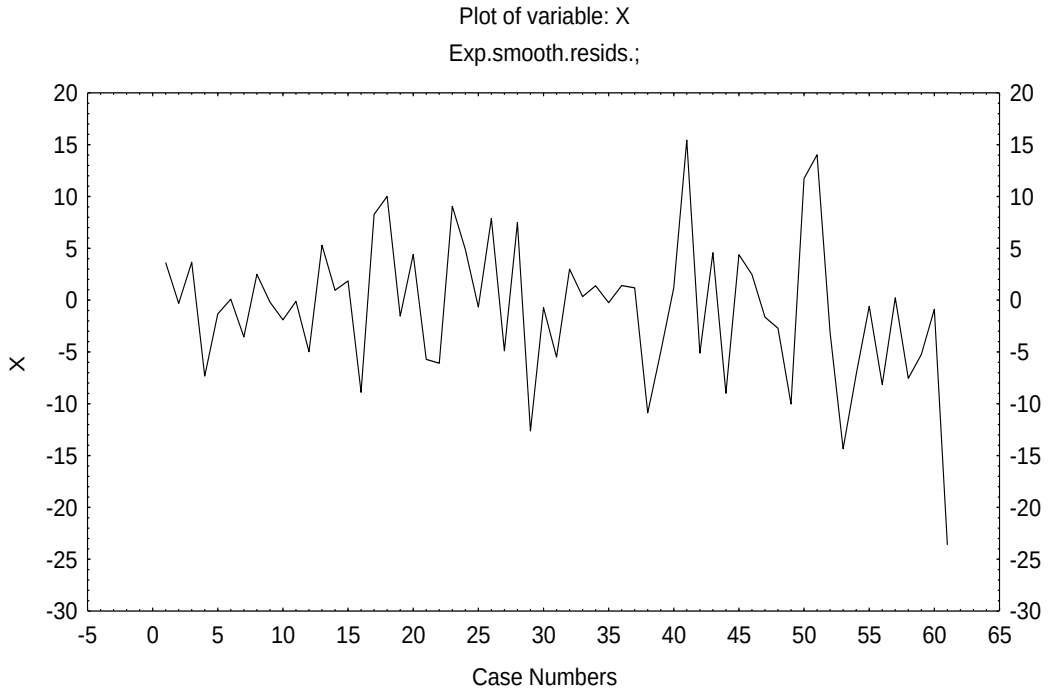


*Рисунок 74*

Графік автокореляційної функції показаний на рисунку 75.

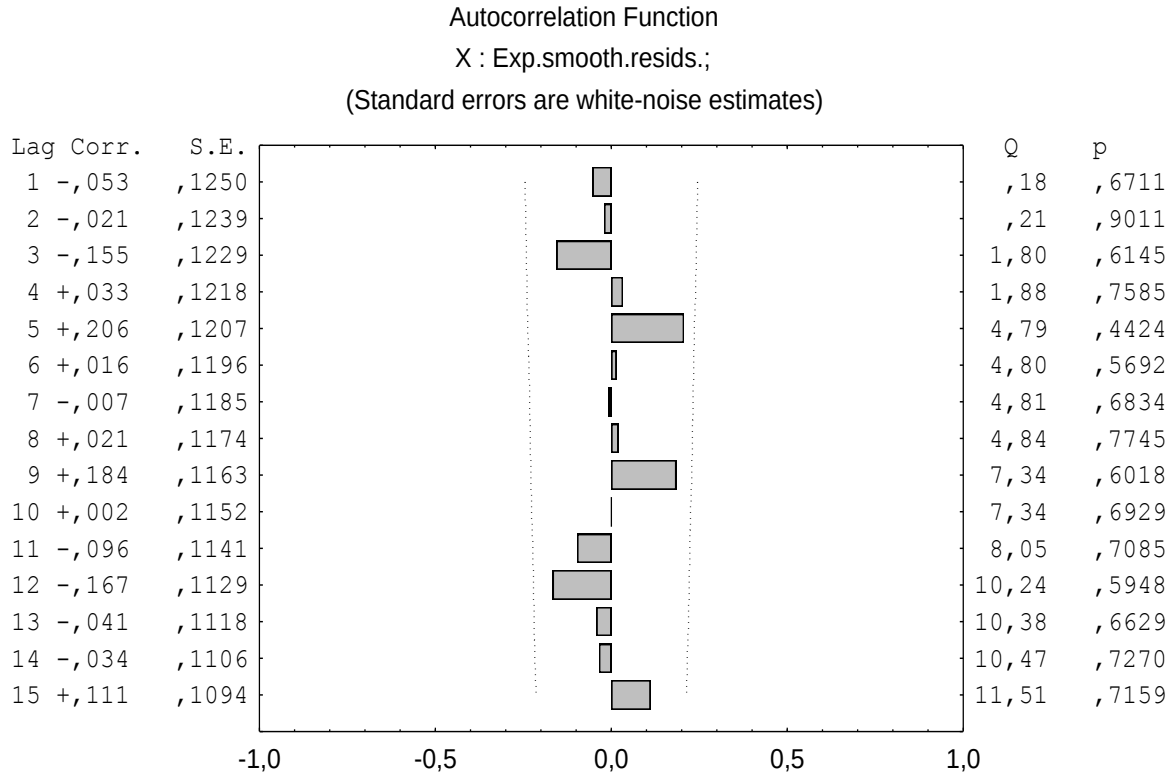


*Рисунок 75*

Графік на нормальному папері показаний на рисунку 76.

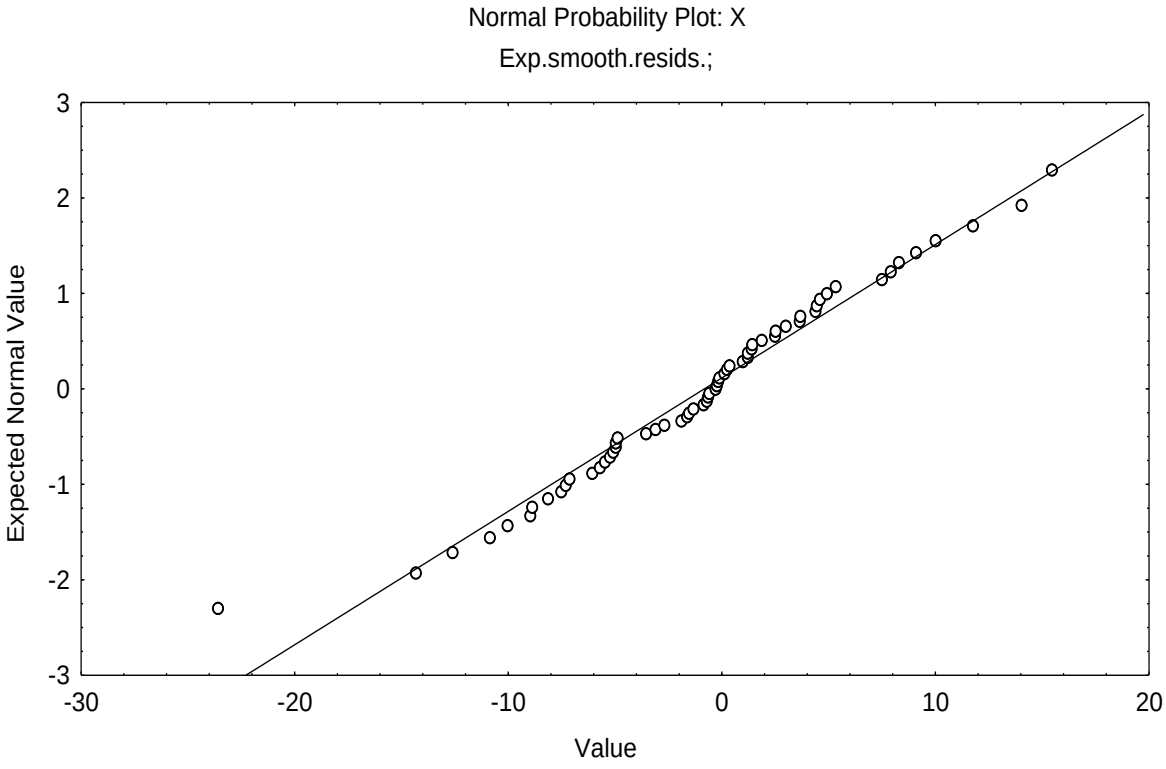


*Рисунок 76*

У моделях 3 і 4 амплітуда залишків незначно зростає, тому для подальшого аналізу залишаємо тільки моделі 1 і 2.

*Перевірка моделей 1 і 2 за автокореляційними функціями*

Моделі 1 і 2 мають практично однакові автокореляційні функції, які не виходять за пунктирну смугу, що не дає можливості зробити вибір між цими двома моделями.

*Перевірка моделей 1 і 2 на нормальність розподілу залишків*

В обох моделях точки досить добре лягають на пряму, але в 2-й моделі відхилення від нормального розподілу на деяких ділянках більше, ніж в 1-й.

Вибираємо 1-у модель – адитивну з лінійним трендом.

3) Прогноз за обраною моделлю. Після вибору моделі беремо з автозвіту таблицю із прогнозом. У першому стовпці перебувають значення *x*, у другому – прогноз за експонентним згладжуванням, у третьому – залишки й у четвертому – сезонний компонент (табл. 6).

*Таблиця 6*

| | **X** | **Smoothed** | **Resids** | **Seasonal** |
|---|---|---|---|---|
| | *2* | *3* | *4* | *5* |
| | 14,0000 | 1,0014 | 12,9986 | -11,6288 |
| | 28,0000 | 34,4843 | -6,4843 | 9,7878 |
| | 25,0000 | 21,8544 | 3,1456 | -0,8163 |
| | 17,0000 | 23,6231 | -6,6231 | -3,9830 |
| | 31,0000 | 39,4235 | -8,4235 | 14,1837 |

| | | | | |
|---|---|---|---|---|
| | 44,0000 | 51,9781 | -7,9781 | 30,9545 |
| | 44,0000 | 47,9402 | -3,9402 | 31,3795 |
| | 32,0000 | 23,6613 | 8,3387 | 9,0128 |
| | 15,0000 | 7,6614 | 7,3386 | -14,2538 |
| 0 | 0,0000 | -4,7166 | 4,7166 | -33,7122 |

Продовження таблиці 6

| | *2* | *3* | *4* | *5* |
|---|---|---|---|---|
| 1 | 14,0000 | 19,1085 | -5,1085 | -15,4622 |
| 2 | 11,0000 | 17,4485 | -6,4485 | -15,4622 |
| 3 | 22,0000 | 18,6224 | 3,3776 | |
| 4 | 37,0000 | 43,5198 | -6,5198 | |
| 5 | 31,0000 | 29,8852 | 1,1148 | |
| 6 | 21,0000 | 28,5284 | -7,5284 | |
| 7 | 45,0000 | 42,1668 | 2,8332 | |
| 8 | 66,0000 | 61,9282 | 4,0718 | |
| 9 | 66,0000 | 66,6035 | -0,6035 | |
| | 54 | 45, | 8, | |

| | | | | |
|---|---|---|---|---|
| 0 | ,0000 | 4195 | 5805 | |
| 1 | 29 ,0000 | 29, 9665 | - 0,9665 | |
| 2 | 10 ,0000 | 11, 5227 | - 1,5227 | |
| 3 | 36 ,0000 | 30, 0751 | 5, 9249 | |
| 4 | 41 ,0000 | 36, 2602 | 4, 7398 | |
| 5 | 46 ,0000 | 46, 5058 | - 0,5058 | |
| 6 | 74 ,0000 | 69, 0612 | 4, 9388 | |
| 7 | 59 ,0000 | 65, 1521 | - 6,1521 | |
| 8 | 68 ,0000 | 59, 4159 | 8, 5841 | |
| 9 | 74 ,0000 | 86, 7385 | - 12,7385 | |
| 0 | 95 ,0000 | 96, 6414 | - 1,6414 | |
| 1 | 95 ,0000 | 97, 7819 | - 2,7819 | |
| 2 | 80 ,0000 | 75, 8007 | 4, 1993 | |
| 3 | 58 ,0000 | 57, 1399 | 0, 8601 | |
| 4 | 42 ,0000 | 40, 2313 | 1, 7687 | |
| 5 | 62 ,0000 | 61, 8149 | 0, 1851 | |

| | | | | |
|---|---|---|---|---|
| 6 | 67,0000 | 64,0485 | 2,9515 | |

Продовження таблиці 6

| | *2* | *3* | *4* | *5* |
|---|---|---|---|---|
| 7 | 76,0000 | 72,8617 | 3,1383 | |
| 8 | 89,0000 | 98,3632 | -9,3632 | |
| 9 | 77,0000 | 83,3406 | -6,3406 | |
| 0 | 79,0000 | 77,1321 | 1,8679 | |
| 1 | 114,0000 | 97,8610 | 16,1390 | |
| 2 | 125,0000 | 129,0799 | -4,0799 | |
| 3 | 138,0000 | 129,0773 | 8,9227 | |
| 4 | 106,0000 | 116,9161 | -10,9161 | |
| 5 | 87,0000 | 87,9337 | -0,9337 | |
| 6 | 68,0000 | 70,0305 | -2,0305 | |
| 7 | 90,0000 | 88,8750 | 1,1250 | |
| 8 | 92,0000 | 91,8832 | 0,1168 | |
| 9 | 92,0000 | 98,5529 | -6,5529 | |
| | 13 | 116 | 15 | |

| | | | | |
|---|---|---|---|---|
| 0 | 2,0000 | ,0523 | ,9477 | |
| 1 | 13 1,0000 | 119 ,7666 | 11 ,2334 | |
| 2 | 12 5,0000 | 128 ,2649 | -3,2649 | |
| 3 | 13 9,0000 | 147 ,6285 | -8,6285 | |
| 4 | 16 0,0000 | 160 ,9693 | -0,9693 | |
| 5 | 16 8,0000 | 163 ,8177 | 4, 1823 | |
| 6 | 13 3,0000 | 147 ,6948 | -14,6948 | |
| 7 | 10 7,0000 | 116 ,0540 | -9,0540 | |
| 8 | 76 ,0000 | 91, 5524 | -15,5524 | |
| 9 | 97 ,0000 | 99, 1984 | -2,1984 | |
| 0 | 10 0,0000 | 97, 7895 | 2, 2105 | |
| 1 | 84 ,0000 | 103 ,8108 | -19,8108 | |
| 2 | | 109 ,8628 | | |

Продовження таблиці 6

| | *2* | *3* | *4* | *5* |
|---|---|---|---|---|
| 3 | | 98, 3140 | | |
| 4 | | 93, 7693 | | |

| | | | | |
|---|---|---|---|---|
| 5 | | 110,7414 | | |
| 6 | | 126,4398 | | |
| 7 | | 126,2770 | | |
| 8 | | 102,5232 | | |
| 9 | | 78,2296 | | |
| 0 | | 57,3963 | | |
| 1 | | 74,9437 | | |

Відповідно до умови задачі випадки 1…61 відповідають періоду із січня 1995 р. по січень 2000 р. Отже, випадки 62…67 відповідають періоду з лютого 2000 р. по липень 2000 р.

Прогноз на лютий – липень 2000 р., рівень довіри 90% (установлено з порівняння результатів аналізу методами ARIMA і Exponential Smoothing & Forecasting), даний у таблиці 7.

*Таблиця 7*

| | |
|---|---|
| Лютий | 109,8628 |
| Березень | 98,3140 |
| Квітень | 93,7693 |
| Травень | 110,7414 |
| Червень | 126,4398 |
| Липень | 126,2770 |

Відомо, що для тимчасових рядів довірчий інтервал практично симетричний щодо прогнозу. Його напівширина для рівня довіри 90% становить 20...30% від прогнозу. Беручи напівширину довірчого інтервалу

рівною 25% прогнозу, одержуємо, наприклад, на липень такий прогноз: перевезення вугілля будуть укладені в межах 126,27*(1 - 0,25) … 126,27*(1 + 0,25), що дає 94,70 … 157,83 млн т.

Графік прогнозу показаний на рисунку 77.

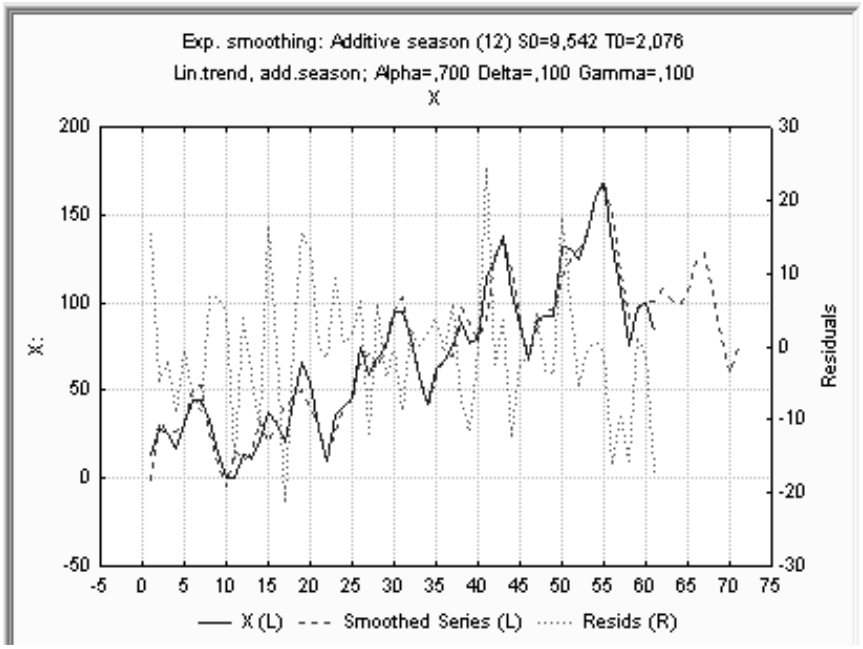


*Рисунок 77*

На цьому графіку суцільною лінією показаний графік тимчасового ряду, пунктирною лінією – графік прогнозу, точковою лінією – графік залишків. Масштаб для графіків тимчасового ряду й прогнозу заданий на лівій вертикальній осі, масштаб для графіка залишків – на правій вертикальній осі. З порівняння графіків тимчасового ряду й прогнозу видно їхній гарний збіг на всьому протязі графіка тимчасового ряду.

### 7.6 Висновки

На підставі аналізу залишків тимчасового ряду для припасування моделі обрана лінійна дистрибутивна модель. Прогноз на підставі цієї моделі наведений у таблиці 8.

*Таблиця 8*

| **Місяць** | **Прогноз** | **90% довірчий інтервал** |
|---|---|---|
| *1* | *2* | *3* |

| | | |
|---|---|---|
| Лютий | 109,8628 | 82,3971 – 137,328 |

*Продовження таблиці 8*

| *1* | *2* | *3* |
|---|---|---|
| Березень | 98,3140 | 73,74 – 122,89 |
| Квітень | 93,7693 | 70,33 –117,21 |
| Травень | 110,7414 | 83,06 – 138,43 |
| Червень | 126,4398 | 94,83 – 158,05 |
| Липень | 126,2770 | 94,71 –157,85 |

# ЛІТЕРАТУРА

1 Лук'яненко І. Економетрика/ І.Лук'яненко, Л.Краснікова. – Київ: Знання, 1998. – 493с.

2 Лук'яненко І. Економетрика: Практикум/ І.Лук'яненко, Л.Краснікова. – Київ: Знання, 1998. – 217с.

3 Боровиков В.П. STATISTICA/ В.П.Боровиков, И.П.Боровиков – М.: Информационно-издательский дом “Филинъ”, 1997. – 592с.

4 Доугерти К. Введение в эконометрику. – М.: Инфра-М, 2001. – 402с.

5 Эконометрика. Начальный курс: Ученик/ Я.Р.Магнус, П.К.Катышев, А.А.Пересецкий. – 4-е изд. – М.: Дело, 2000. – 400 с.