

Лабораторна робота. Побудова й аналіз моделі з фіктивними незалежними змінними

Мета – закріплення теоретичного й практичного матеріалу за темою “Моделювання економічних процесів зі специфічними змінними”, придбання навичок побудови й аналізу економетричних моделей з фіктивними незалежними змінними у середовищі *Statistica*.

Завдання – необхідно перевірити наявність структурного зрушення в даних. Завдання вирішується в модулі *Multiple Regression* ППП *Statistica*:

1. Побудувати “збірну” регресію, регресії для кожної з досліджуваних сукупностей, перевірити гіпотезу про відсутність структурного зрушення в вибікових даних на основі тесту Чоу.
2. Об’єднати всі спостереження та оцінити параметри лінійного рівняння: $y_t = a_0^1 + a_0^2 \cdot d_i + a_1^1 \cdot x_i + a_1^2 \cdot d_i x_i + \varepsilon_i$. Оцінити значущість диференційованих коефіцієнтів перетину та нахилу.
3. Побудувати рівняння для кожної з наведених сукупностей. Дати економічну інтерпретацію результатів моделювання.
4. Побудувати прогнози для об’єктів з відповідних досліджуваних сукупностей.

Методичні рекомендації

Завдання вирішується в модулі *Multiple Regression* (*Множинна регресія*). Вихідні дані для побудови моделі наведені на рис. 1.1.

| | 1 Xi | 2 Yi | 3 di |
|----|---------|---------|---------|
| 1 | 14772,9 | 370,64 | 1 |
| 2 | 11854 | 297,67 | 1 |
| 3 | 10735,2 | 269,7 | 1 |
| 4 | 8028,5 | 202,03 | 1 |
| 5 | 5446,4 | 137,48 | 1 |
| 6 | 1052,1 | 27,62 | 1 |
| 7 | 5614,7 | 141,69 | 1 |
| 8 | 3747,9 | 95,02 | 1 |
| 9 | 6054,4 | 152,68 | 1 |
| 10 | 4653,2 | 117,65 | 1 |
| 11 | 114,7 | 24,2 | 0 |
| 12 | 308,1 | 62,92 | 0 |
| 13 | 549,6 | 111,17 | 0 |
| 14 | 375,3 | 76,29 | 0 |
| 15 | 59,1 | 13,04 | 0 |
| 16 | 106,7 | 22,57 | 0 |
| 17 | 195,9 | 40,45 | 0 |
| 18 | 268,5 | 54,98 | 0 |
| 19 | 164,4 | 34,11 | 0 |
| 20 | 295,4 | 60,34 | 0 |

Рис. 1.1. Вихідні дані

В якості вихідних даних розглядається Y_i – дохід i -го комерційного банку; X_i – величина залучених коштів i -го комерційного банку; d_i – фіктивна змінна, яка приймає значення ноль, якщо банк відноситься до групи середніх банків і одиниця, якщо банк відноситься до великих банків.

Розглянемо порядок розрахункових процедур. Щоб приступити до обчислювальних процедур, необхідно ввійти в позицію *Statistics / Multiple Regression* (рис. 1.2).

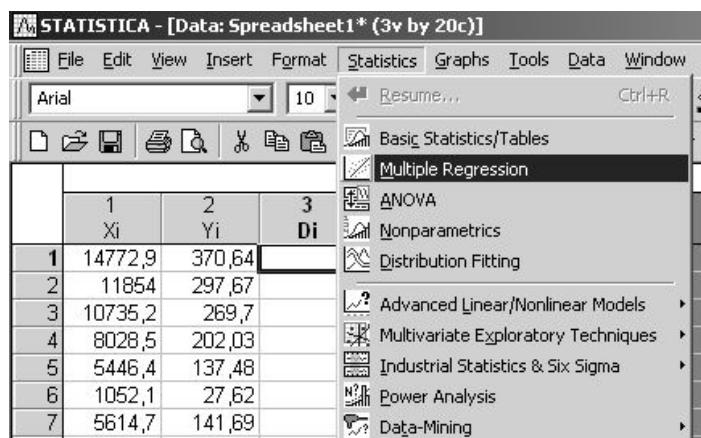


Рис. 1.2. Вибір модуля *Multiple Regression*

Після підтвердження вибору модуля перед вами з'явиться стартова панель даного модуля, де необхідно задати змінні для аналізу (рис. 1.3).

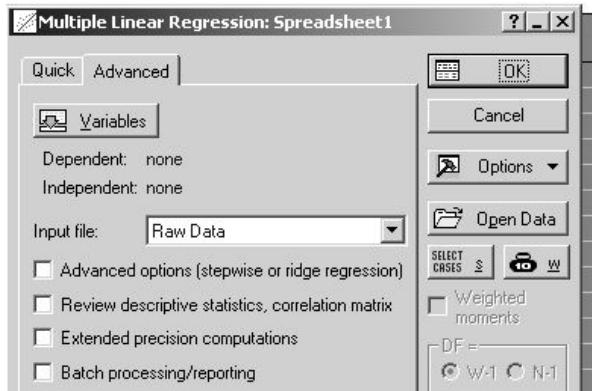


Рис. 1.3. Стартова панель модуля *Multiple Regression*

Ініціюйте кнопку *Variables* (Змінні) і у вікні, що з'явилося, вкажіть *Dependent* (залежну) і *Independent* (незалежну) змінні для побудови регресійної моделі. Вибір змінних наведений на рис. 1.4. Після вказівки змінних підтвердіть свій вибір натисканням кнопки *OK* (рис. 1.5).

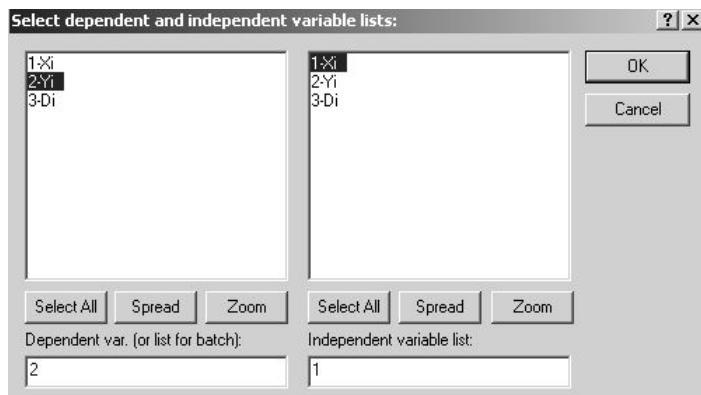


Рис. 1.4. Вибір змінних для аналізу

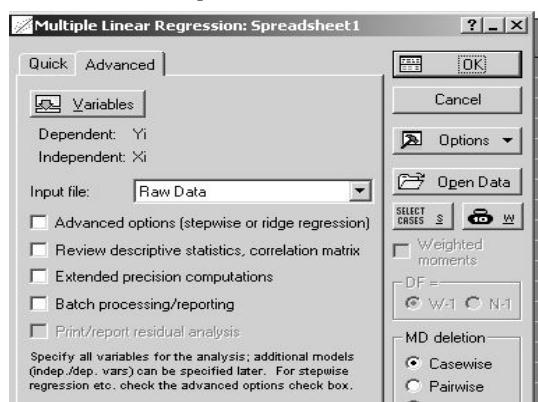


Рис. 1.5. Підтвердження вибору змінних

Побудуємо “збірну” регресію і визначимо всі її характеристики. Результати побудови лінійної регресійної моделі будуть подані в

діалоговому вікні (рис. 1.6). У верхній частині вікна подана основна інформація моделі, у нижній частині знаходяться функціональні кнопки, що дозволяють всебічно розглянути результати аналізу.

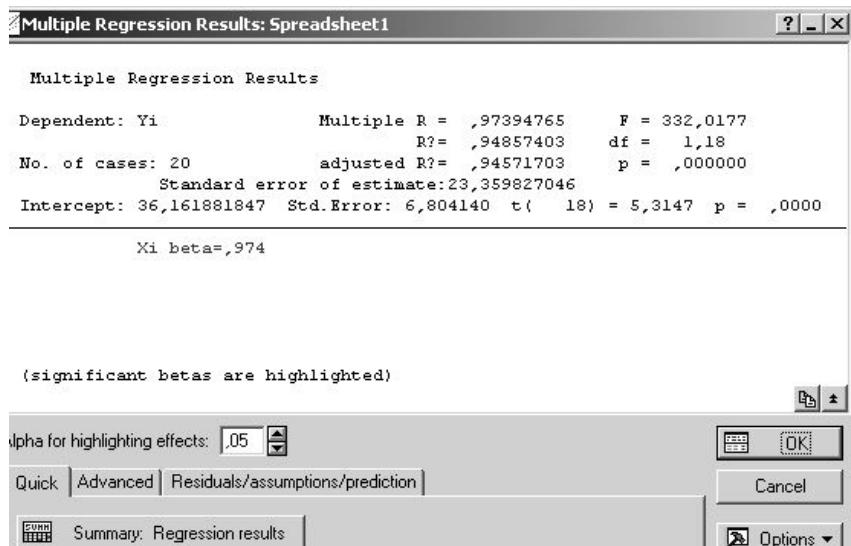


Рис. 1.6. Вікно результатів регресійного аналізу

Ініціювавши кнопку *Summary: Regression results* (на вкладці *Quick*) (*Результати регресійного аналізу*), визначимо найважливіші характеристики моделі й ступінь її адекватності (рис. 1.7).

Проаналізуємо отримані результати моделі:

1) аналіз адекватності:

$R = 0,9739$ – коефіцієнт множинної кореляції (у випадку простої лінійної регресії дорівнює модулю коефіцієнта парної кореляції);

$R^2 = 0,9486$ – коефіцієнт детермінації моделі;

$Adjusted R^2 = 0,9457$ – скорегований коефіцієнт детермінації, який корегується на число спостережень і число параметрів моделі;

$F (1, 18) = 332,02$ – критерій Фішера статистичної значимості моделі з числом ступенів свободи та рівнем значимості p ;

$Std. Error of estimate = 23,36$ – середнє квадратичне відхилення помилок моделі; дана статистика є мірою розсіву досліджуваних значень відносно регресійної прямої.

| N=20 | Regression Summary for Dependent Variable:Yi (Spreadsheet1) | | | | | |
|-----------|---|------------------|----------|---------------|----------|----------|
| | Beta | Std.Err. of Beta | B | Std.Err. of B | t(18) | p-level |
| Intercept | | | 36,16188 | 6,804140 | 5,31469 | 0,000047 |
| Xi | 0,973948 | 0,053451 | 0,02136 | 0,001172 | 18,22135 | 0,000000 |

Рис. 1.7. Результати регресійного аналізу

2) аналіз параметрів та їх статистичної значимості:

$Beta (a_1) = (0,97394)$ – стандартизовані значення коефіцієнтів регресії (ваги);

$Std.Error of Beta (a_1) = 0,05345$ – середнє квадратичне відхилення стандартизованих параметрів моделі;

$B (a_0, a_1) = (36,1618; 0,0213)$ – параметри регресійної моделі, отже модель має вигляд:

$$\hat{Y} = 36,1618 + 0,0213 \cdot x_i ;$$

$Std.Error of B = (6,8041; 0,00117)$ – середнє квадратичне відхилення параметрів моделі;

$t(13) = (5,3147; 18,2213)$ – значимість параметрів моделі за критерієм Стьюдента;

$p-level = (0,00004; 0,000)$ – рівень значимості критерію Стьюдента.

Для перевірки гіпотези про значимість регресійної моделі використовується дисперсійний аналіз. Для цього необхідно ініціювати кнопку *Advanced / ANOVA* у нижній частині інформаційного вікна (рис. 1.8).

Результати дисперсійного аналізу для досліджуваної моделі наведені на рис. 1.9. У даній таблиці наведено суму квадратів відхилень за регресією (*Sums of Squares Regress*), суму квадратів відхилень похибок (*Sums of Squares Residual*), дисперсію похибок (*Mean Squares Residual*) та критерій Фішера.

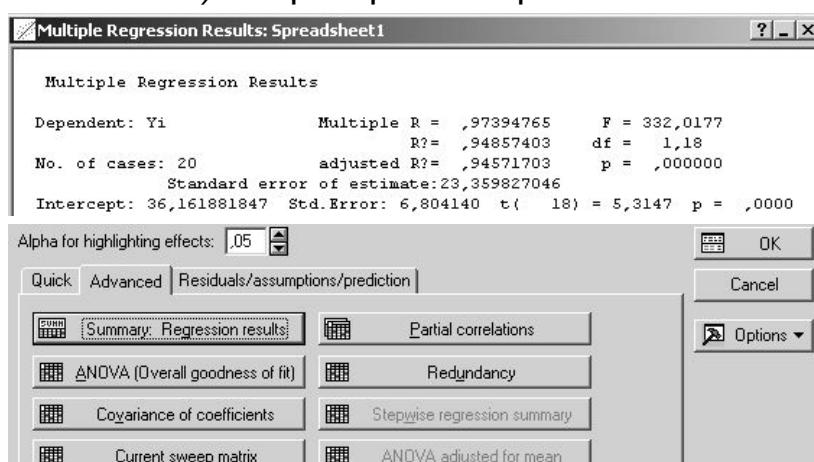


Рис. 1.8. Вибір вікна дисперсійного аналізу

| Effect | Analysis of Variance; DV: Yi (Spreadsheet1) | | | | |
|----------|---|----|--------------|----------|----------|
| | Sums of Squares | df | Mean Squares | F | p-level |
| Regress. | 181175,9 | 1 | 181175,9 | 332,0177 | 0,000000 |
| Residual | 9822,3 | 18 | 545,7 | | |
| Total | 190998,2 | | | | |

Рис. 1.9. Таблиця дисперсійного аналізу

Для побудови регресійного рівняння для групи великих банків необхідно ініціювати кнопку *Select cases* на стартовій панелі модуля та визначити спостереження, за якими буде здійснено аналіз (рис. 2.10).

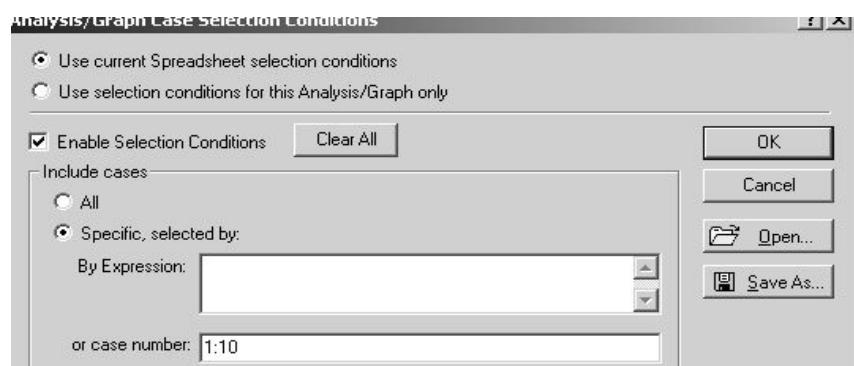


Рис. 1.10. Визначення вихідної сукупності

Побудова регресійного рівняння для групи великих банків здійснюється аналогічно до наведеної вище процедури побудови “збірної” регресії. Результати дисперсійного аналізу для групи великих банків наведені на рис. 1.11. Аналогічно знаходимо суму квадратів залишків регресійного рівняння залежності доходу від залучених коштів для групи середніх банків (рис. 1.12).

| Effect | Analysis of Variance; DV: Yi (Spreadsheet1) | | | | |
|----------|---|----|--------------|--------------|----------|
| | Sums of Squares | df | Mean Squares | F | p-level |
| Regress. | 97057,19033152 | 1 | 97057,19 | 2,726022E+10 | 0,000000 |
| Residual | 0,00002848 | 8 | 0,00 | | |
| Total | 97057,19036000 | | | | |

Рис. 1.11. Таблиця дисперсійного аналізу для групи великих банків

| Effect | Analysis of Variance; DV: Yi (Spreadsheet) | | | | |
|----------|--|----|--------------|----------|---------|
| | Sums of Squares | df | Mean Squares | F | p-leve |
| Regress. | 7859,350 | 1 | 7859,350 | 11320955 | 0,00000 |
| Residual | 0,006 | 8 | 0,001 | | |
| Total | 7859,356 | | | | |

Рис. 1.12. Таблиця дисперсійного аналізу для групи середніх банків

Для порівняння регресій за двома сукупностями застосовуємо тест Чоу:

$$F = \frac{982,3 - (0,0000284 + 0,006)}{0,0000284 + 0,006} \cdot \frac{10 + 10 - 2 \cdot 2}{2} = 13034654.$$

Визначимо теоретичне значення критерію за тестом Фішера. Для цього необхідно обрати модуль *Basic Statistics/Tables* (рис 2.13).

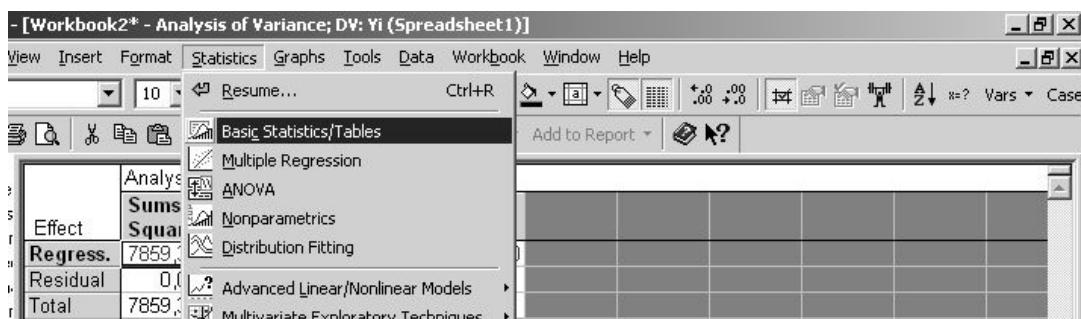


Рис. 1.13. Вибір модуля *Basic Statistics/Tables*

Після підтвердження вибору модуля з'явиться діалогове вікно, що дозволяє задати напрямок аналізу *Probability Calculator* (рис. 1.14).

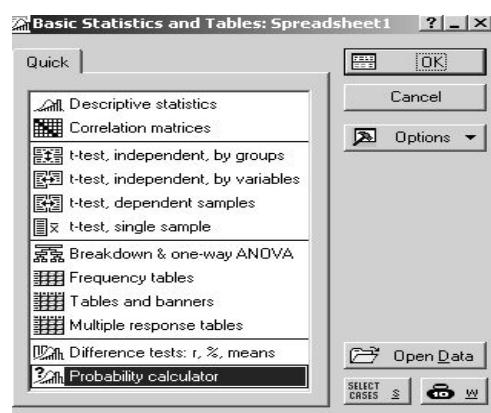


Рис. 1.14. Вибір напряму аналізу

Після вибору напряму аналізу з'явиться стартова панель модуля (рис. 1.15), де необхідно задати наступні вихідні параметри: *Distribution* (розділ), *df* (ступінь свободи), *p* (рівень значимості). Ініціюйте кнопку *Compute* і у вікні з'явиться критичне значення *F*-розділу Фішера.

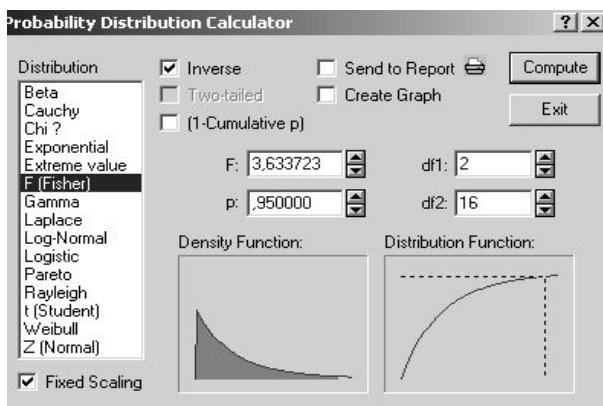


Рис. 1.15. Критичне значення *F*-розділу Фішера

Оскільки, обчислене значення *F*-критерію перевищує критичне, то гіпотеза про відсутність структурного зрушення в спостережуваних вибіркових даних відхиляється.

Об'єднаємо всі спостереження в єдину сукупність і оцінимо регресію з диференційованими коефіцієнтами, відповідно сформувавши фіктивні змінні нахилу (d_i) та перетину ($d_i x_i$) (рис. 1.16).

Побудуємо лінійну багатофакторну економетричну модель виду: $y_t = a_0^1 + a_0^2 \cdot d_i + a_1^1 \cdot x_i + a_1^2 \cdot d_i x_i + \varepsilon_i$ і визначимо характеристики її адекватності та значимості (рис. 1.17).

| | 1 Xi | 2 Yi | 3 Di | 4 Di*Xi |
|----|---------|---------|---------|------------|
| 1 | 14772,9 | 370,64 | 1 | 14772,9 |
| 2 | 11854 | 297,67 | 1 | 11854 |
| 3 | 10735,2 | 269,7 | 1 | 10735,2 |
| 4 | 8028,5 | 202,03 | 1 | 8028,5 |
| 5 | 5446,4 | 137,48 | 1 | 5446,4 |
| 6 | 1052,1 | 27,62 | 1 | 1052,1 |
| 7 | 5614,7 | 141,69 | 1 | 5614,7 |
| 8 | 3747,9 | 95,02 | 1 | 3747,9 |
| 9 | 6054,4 | 152,68 | 1 | 6054,4 |
| 10 | 4653,2 | 117,65 | 1 | 4653,2 |
| 11 | 114,7 | 24,2 | 0 | 0 |
| 12 | 308,1 | 62,92 | 0 | 0 |
| 13 | 549,6 | 111,17 | 0 | 0 |
| 14 | 375,3 | 76,29 | 0 | 0 |
| 15 | 59,1 | 13,04 | 0 | 0 |
| 16 | 106,7 | 22,57 | 0 | 0 |
| 17 | 195,9 | 40,45 | 0 | 0 |
| 18 | 268,5 | 54,98 | 0 | 0 |
| 19 | 164,4 | 34,11 | 0 | 0 |
| 20 | 295,4 | 60,34 | 0 | 0 |

Рис. 1.16. Вихідні дані з фіктивними змінними

| Regression Summary for Dependent Variable:Yi (Spreadsheet1) | | | | | | |
|---|----------|------------------|-----------|---------------|----------|----------|
| N=20 | Beta | Std.Err. of Beta | B | Std.Err. of B | t(16) | p-level |
| Intercept | | | 1,241252 | 0,011852 | 104,73 | 0,000000 |
| Xi | 9,12217 | 0,001922 | 0,200048 | 0,000042 | 4746,20 | 0,000000 |
| Di | 0,00041 | 0,000087 | 0,079343 | 0,017079 | 4,65 | 0,000269 |
| Di*Xi | -8,15170 | 0,001964 | -0,175048 | 0,000042 | -4150,44 | 0,000000 |

Рис. 1.17. Результати регресійного аналізу

Проаналізуємо отримані результати моделі:

1) аналіз адекватності:

$R = 0,9999$ – коефіцієнт множинної кореляції;

$R^2 = 0,9999$ – коефіцієнт детермінації моделі;

$Adjusted R^2 = 0,9999$ – скорегований коефіцієнт детермінації;

$F (3, 16)$ – критерій Фішера статистичної значимості моделі з числом ступенів свободи та рівнем значимості p ;

$Std.Error of estimate = 0,01868$ – середнє квадратичне відхилення помилок моделі;

2) аналіз параметрів моделі та їх статистичної значимості:

$B (a_0^1, a_1^1, a_0^2, a_1^2) = (1,24125; 0,200045; 0,07934; -0,17504)$ –

вектор параметрів моделі, отже модель має вигляд:

$$\hat{Y} = 1,24 + 0,2 \cdot x_i + 0,079 \cdot d_i - 0,18d_i x_i;$$

Регресійне рівняння для групи великих банків має вигляд:

$$\hat{Y} = (1,24 + 0,079) + (0,2 - 0,18) \cdot x_i = 1,321 + 0,025 \cdot x_i;$$

Відповідно, регресійне рівняння для групи середніх банків:

$$\hat{Y} = 1,24 + 0,2 \cdot x_i;$$

$Std.Error of B = (0,01185; 0,00004; 0,01707; 0,00004)$ – середнє квадратичне відхилення параметрів моделі;

$t(13) = (104,73; 4726,25; 4,65; -4150,45)$ – значимість параметрів за критерієм Стьюдента; $p-level$ – рівень значимості критерію Стьюдента.

Графічне відображення отриманих результатів моделювання для груп спостережень представлено на рис. 1.18.

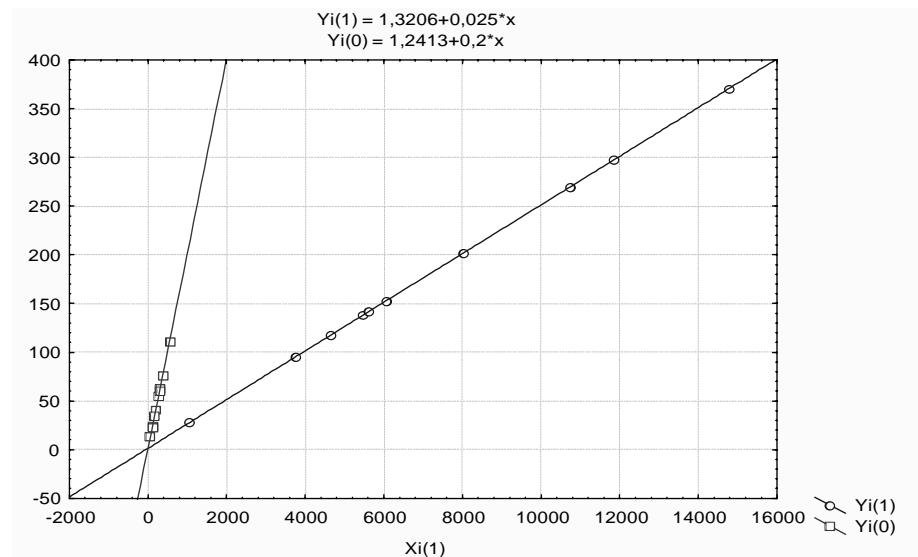


Рис. 1.18. Графік залежності для груп спостережень

Для аналізу помилок моделі у нижній частині інформаційного вікна результатів регресійного аналізу (див. рис. 2.6) необхідно вибрати напрям аналізу *Residuals/assumptions/prediction* та ініціювати опцію *Perform residual analysis* (*Всебічний аналіз залишків*). Ініціювавши дану опцію, одержимо меню для аналізу помилок моделі (рис. 1.19).

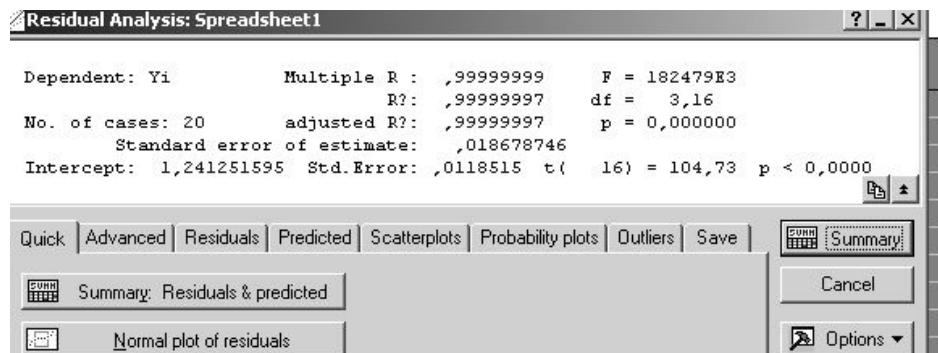


Рис. 1.19. Меню аналізу помилок моделі

Кнопка аналізу помилок *Quick / Summary: Residuals & Predicted* відображає спостережувані значення залежної змінної (*Observed value*), теоретичні значення залежної змінної (*Predicted value*) і помилки моделі (*Residual*) як різницю спостережуваних і теоретичних значень (рис. 1.20).

| Case No. | Predicted & Residual Values (S) | | |
|----------|---------------------------------|----------------|-----------------|
| | Dependent variable:Yi | Observed Value | Predicted Value |
| 1 | 370,6400 | 370,6414 | -0,0013 |
| 2 | 297,6700 | 297,6692 | 0,0008 |
| 3 | 269,7000 | 269,6993 | 0,0006 |
| 4 | 202,0300 | 202,0322 | -0,0021 |
| 5 | 137,4800 | 137,4799 | 0,0000 |
| 6 | 27,6200 | 27,6230 | -0,0029 |
| 7 | 141,6900 | 141,6874 | 0,0025 |
| 8 | 95,0200 | 95,0177 | 0,0023 |
| 9 | 152,6800 | 152,6799 | 0,0001 |
| 10 | 117,6500 | 117,6500 | -0,0000 |
| 11 | 24,2000 | 24,1868 | 0,0132 |
| 12 | 62,9200 | 62,8761 | 0,0439 |
| 13 | 111,1700 | 111,1877 | -0,0177 |
| 14 | 76,2900 | 76,3193 | -0,0293 |
| 15 | 13,0400 | 13,0641 | -0,0241 |
| 16 | 22,5700 | 22,5864 | -0,0163 |
| 17 | 40,4500 | 40,4307 | 0,0193 |
| 18 | 54,9800 | 54,9542 | 0,0258 |
| 19 | 34,1100 | 34,1292 | -0,0191 |
| 20 | 60,3400 | 60,3355 | 0,0045 |

Рис. 1.20. Аналіз помилок моделі

У меню аналізу помилок, ініціювавши кнопку *Advanced / Durbin-Watson statistic* (Статистика Дарбіна – Уотсона), одержуємо значення автокореляції помилок моделі за критерієм Дарбіна – Уотсона та значення нециклічного коефіцієнта автокореляції (рис. 2.21).

| Durbin-Watson d (Sp and serial correlation) | | |
|--|-----------------|----------|
| Durbin- Watson d | Serial Corr. | |
| Estimate | 1,630424 | 0,183472 |

Рис. 1.21. Автокореляція помилок моделі

Значення отриманих коефіцієнтів порівнюються з табличними значеннями і робиться висновок про наявність у моделі автокореляції.

Оскільки одна з основних гіпотез щодо випадкової величини говорить, що помилки повинні бути розподілені за нормальним законом, представимо гістограму розподілу помилок (*Residuals / Histogram plot of residuals*) і проаналізуємо її (рис. 1.22).

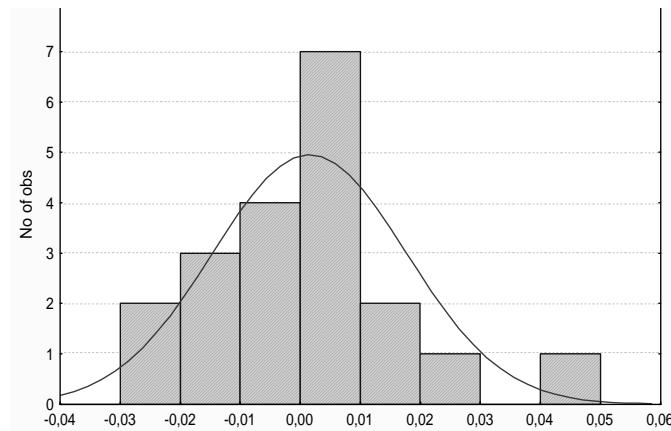


Рис. 1.22. Гістограма розподілу помилок

Оскільки модель є адекватною, її параметри значимі, то за моделлю можна скласти прогноз. Щоб розрахувати прогнозні значення залежної змінної, у нижній частині вікна результатів регресійного аналізу є опція *Predict dependent variable* (Прогнозування залежної змінної). Ініціювавши дану опцію, необхідно вказати значення незалежної змінної, для якої необхідно спрогнозувати залежну величину (рис. 1.23). Знайдемо прогнозні значення доходу комерційних банків, які відповідно відносяться до групи середніх та великих банків.

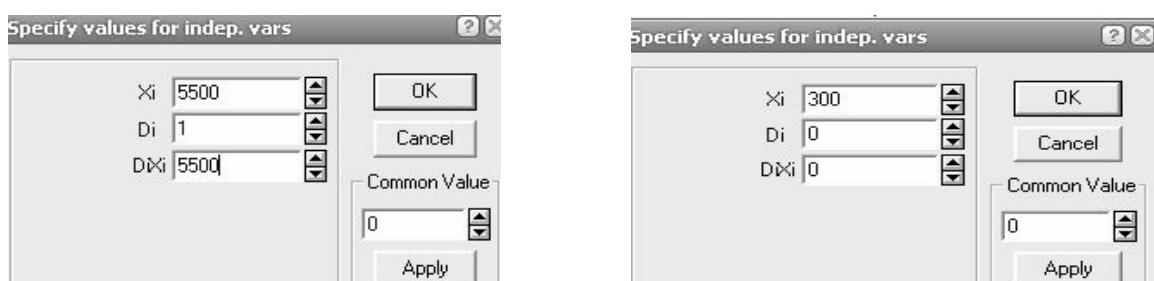


Рис. 1.23. Значення незалежних змінних

Результати прогнозу подаються у вигляді таблиці, у якій зазначені коефіцієнти моделі й порядок розрахунків (рис. 1.24).

Прогнозне значення доходу комерційного банку групи великих банків, якщо величина залучених коштів 5500 тис. грн складе 138,820 тис. грн, відповідно довірчі інтервали для прогнозного значення:

$$138,806 \leq 138,820 \leq 138,834.$$

| Variable | Predicting Values for (Чоу) variable: Yi | | |
|-----------|---|----------|---------------------|
| | B-Weight | Value | B-Weight * Value |
| Xi | 0,200048 | 5500,000 | 1100,265 |
| Di | 0,079343 | 1,000 | 0,079 |
| DiXi | -0,175048 | 5500,000 | -962,766 |
| Intercept | | | 1,241 |
| Predicted | | | 138,820 |
| -95,0%CL | | | 138,806 |
| +95,0%CL | | | 138,834 |

| Variable | Predicting Values for (Чоу) variable: Yi | | |
|-----------|---|----------|---------------------|
| | B-Weight | Value | B-Weight * Value |
| Xi | 0,200048 | 300,0000 | 60,01446 |
| Di | 0,079343 | 0,0000 | 0,000000 |
| DiXi | -0,175048 | 0,0000 | 0,000000 |
| Intercept | | | 1,24125 |
| Predicted | | | 61,25571 |
| -95,0%CL | | | 61,24222 |
| +95,0%CL | | | 61,26920 |

Рис. 1.24. Результати прогнозу

Прогнозне значення доходу комерційного банку групи середніх банків, якщо величина залучених коштів 300 тис. грн складе 61,256 тис. грн, відповідно довірчі інтервали для прогнозного значення:

$$61,242 \leq 61,256 \leq 61,269.$$

Якість отриманих моделей підтверджують досить точні інтервали зміни прогнозних значень рівня доходу.