

Лабораторна робота. Побудова моделей з дискретними залежними змінними

Мета – закріплення теоретичного й практичного матеріалу за темою моделі з дискретними залежними змінними "Побудова Logit- та Probit- моделей", придбання навичок побудови й аналізу економетричних моделей з бінарними результативними змінними у модулі *Nonlinear Estimation*.

Завдання – необхідно побудувати моделі з бінарними результативними змінними (logit- та probit- моделі) за представленими вихідними даними у модулі *Nonlinear Estimation* ППП *Statistica*:

1. Побудувати логістичну модель регресії, визначити всі її характеристики, провести оцінку її адекватності, навести вид моделі.
2. Побудувати графік моделі, визначити помилки моделі та провести їх аналіз за гістограмою та розсівом.
3. Побудувати модель пробіт- регресії, визначити всі її характеристики, провести оцінку її адекватності, навести вид моделі.
4. Побудувати графік моделі, визначити помилки моделі та провести їх аналіз за гістограмою та розсівом.
5. Провести порівняльний аналіз моделей, навести матрицю класифікацій, визначити прогнозні значення ймовірностей, розрахувати маржинальні ефекти за побудованими моделями.
6. Визначити найбільш адекватну модель, представити економічну інтерпретацію результатів моделювання, зробити висновки.

Методичні рекомендації

Для побудови й всебічного аналізу моделей з бінарними результативними змінними у ППП *Statistica* передбачений модуль *Nonlinear Estimation* (*Нелінійне оцінювання*). Розглянемо порядок роботи в даному модулі при побудові logit- та probit- моделей. Вихідні дані, що відображують залежність між досвідом роботи науково-педагогічних кадрів за спеціальністю (X) та їх участю в міжнародних грантах (Y) представлені на рис. 1.1. Бінарна змінна (Y) приймає значення 1, якщо науковець є учасником гранту, значення 0 в протилежному випадку.

	1 X	2 Y
1	7	0
2	15	0
3	16	0
4	15	1
5	8	1
6	4	0
7	18	0
8	2	0
9	22	1
10	6	0
11	30	1
12	1	0
13	30	1
14	5	0
15	20	1
16	13	0
17	9	0
18	32	1
19	4	0
20	13	1
21	9	0
22	4	0
23	28	1
24	22	1
25	18	1

Рис. 1.1. Вихідні дані

Перевірку існування взаємозв'язку між досліджуваними змінними на основі побудови моделей з дискретними змінними будемо проводити в модулі *Advanced Linear/Nonlinear Models* (*Додаткові лінійні /Нелінійні моделі*). Вибір модуля *Nonlinear Estimation* (*Нелінійне оцінювання*) наведено на рис. 1.2.

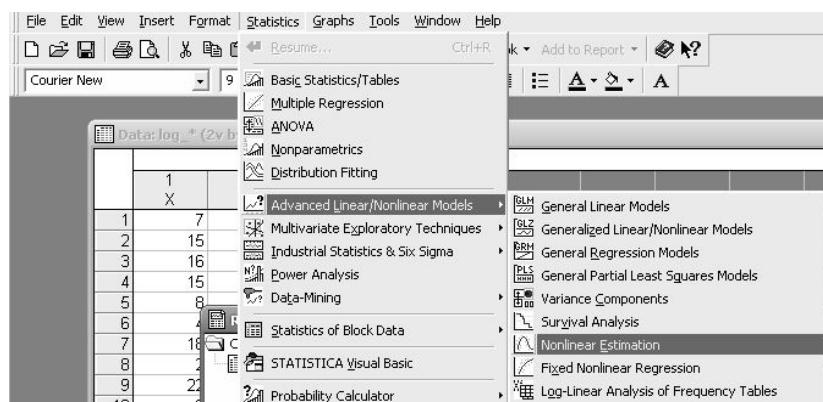


Рис. 1.2. Вибір модуля оцінювання

Вид стартової панелі модуля наведений на рис. 1.3. У вікні модуля подано наступні види нелінійної оцінки:

- *User – specified regression, least squares* (визначена користувачем регресія з похибками за методом найменших квадратів);

- *User – specified regression, custom loss function* (визначена користувачем регресія з заданою функцією похибок);
- *Quick Logit regression* (логіт- регресія);
- *Quick Probit regression* (пробіт- регресія);
- *Exponential growth regression* (регресія експоненційного зростання);
- *Piecewise linear regression* (кусково - лінійна регресія)

Для вирішення поставленої задачі необхідно вибрати опцію *Quick Logit regression* (рис. 1.3).

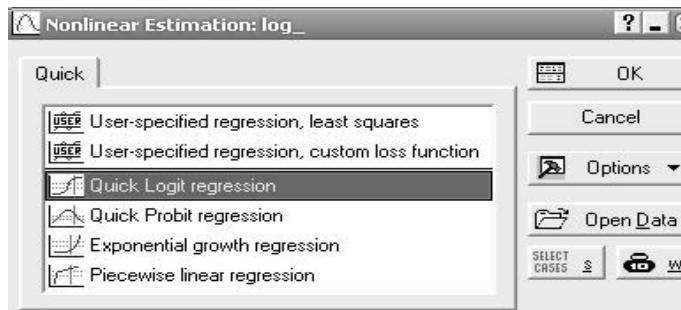


Рис. 1.3. Стартова панель модуля *Nonlinear Estimation*

Підтвердіть вибір напряму аналізу і у вікні, що з'явилося, вкажіть *Dependent* (залежну) та *Independent* (незалежну) змінні для побудови моделі, а також коди залежної змінної. Після вказівки змінних підтвердіть свій вибір натисканням кнопки *OK* (рис. 1.4).

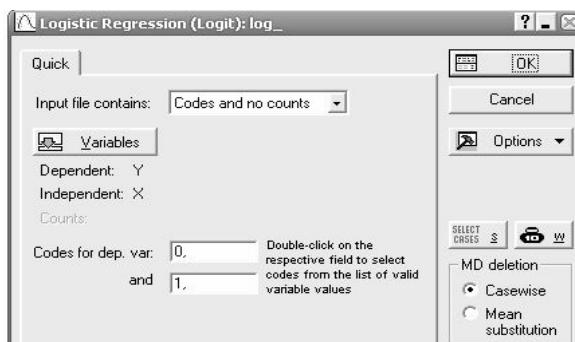


Рис. 1.4. Підтвердження вибору змінних

У наступному вікні необхідно задати метод оцінювання параметрів моделі, при необхідності на вкладці *Advanced* можна також визначити додаткові параметри ітеративної процедури та початкові значення оцінюваних параметрів і запустити процедуру оцінювання (рис. 1.5).

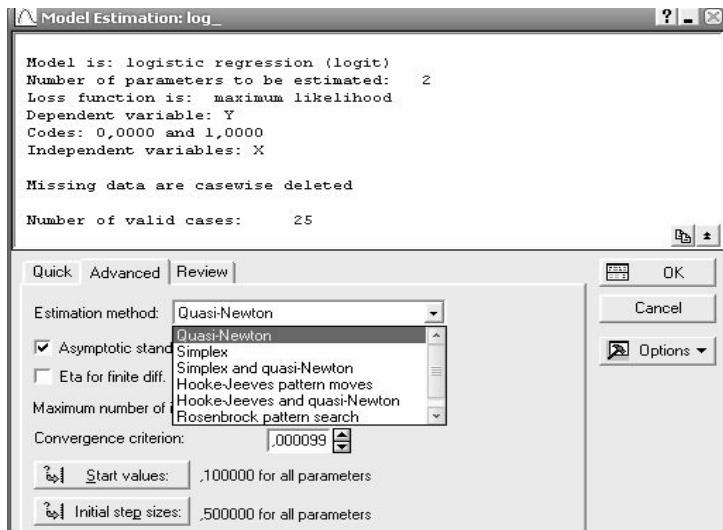


Рис. 1.5. Вибір методу оцінювання параметрів моделі

В модулі можливо вибір наступних методів оцінки параметрів:

- *Quasi-Newton* – Квазі-Ньютонівський метод;
- *Simplex* – Симплекс-метод;
- *Simplex and quasi-Newton* – Симплекс і Квазі-Ньютонівський метод;
- *Hooke-Jeeves pattern moves* – метод Хука-Дживіса;
- *Hooke-Jeeves and quasi-Newton* – метод Хука-Дживіса і Квазі-Ньютона;
- *Rosenbrock pattern search* – метод Розенброка.

Після проведення розрахунків на екрані з'явиться діалогове вікно результатів оцінювання за логіт-регресією (рис. 1.6). В верхній частині вікна подано значення основних результатів: значення функції втрат (*Final value*), логарифм функції правдоподібності ($-2 * \log(\text{likelihood})$), статистика χ^2 – критерій відношення правдоподібності (*Chi-square*), число ступенів свободи (*df*) та рівень значущості (*p*). Порівняння значення χ^2 з табличним при відповідних ступенях свободи і рівні значущості α є перевіркою значущості моделі. Якщо $\chi^2 > \chi^2_{\text{табл}}$, то модель є адекватною і може використовуватися для прогнозування, $\chi^2_{\text{табл}} (\alpha = 0,05; k = 1) = 3,85$. $\chi^2 > \chi^2_{\text{табл}} (\alpha, k)$, тобто в цілому рівняння є значимим, і, відповідно, досвід роботи науково-педагогічних кадрів за спеціальністю впливає на ймовірність участі в міжнародних проектах.

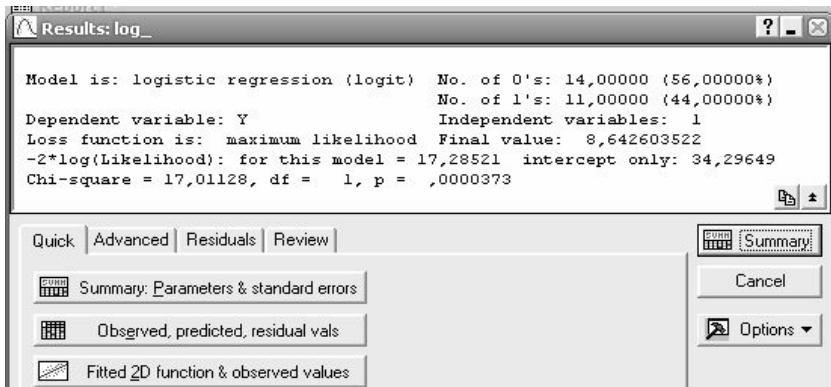


Рис. 1.6. Вікно результатів аналізу за logit- моделлю

В нижній частині вікна знаходяться опції для всебічного аналізу отриманої моделі. Графік логістичної функції одержимо, ініціювавши клавішу *Fitted 2D function & observed values*, який наведено на рис. 1.7.

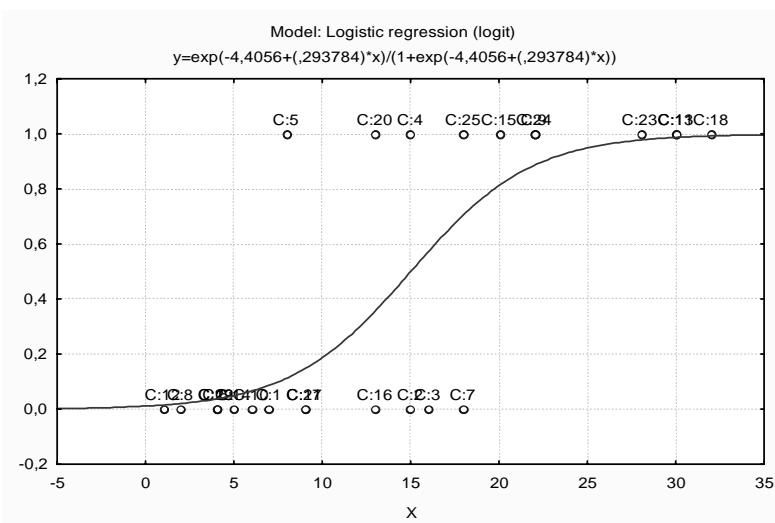


Рис. 1.7. Графік логістичної функції

Оцінки параметрів моделі (*Estimate*), їх стандартні помилки (*Standard Error*) та значущість за критерієм Стьюдента (*t*) можна одержати, ініціювавши опцію *Summary: Parameters & standard errors* (Результат: параметри моделі), результат реалізації якої наведено на рис. 1.8.

N=25	Model: Logistic regre Dep. var: Y Loss: Max Final loss: 8,6426035	
	Const.B0	X
Estimate	-4,40560	0,294
Standard Error	1,72744	0,114
t(23)	-2,55037	2,584

Рис. 1.8. Коефіцієнти логістичної регресії

Таким чином, модель залежності між досвідом роботи науково-педагогічних кадрів та їх участю в міжнародних грантах має наступний вид:

$$F(u) = \Lambda(u) = \frac{e^u}{1+e^u} = \frac{e^{(-4,406+0,294)}}{1+e^{(-4,406+0,294)}}.$$

Додатній коефіцієнт при незалежній змінній визначає, що збільшення досвіду роботи збільшує ймовірність участі у грантах.

Для всебічного дослідження помилок моделі та результатів класифікації необхідно ініціювати вкладку *Residuals* у нижній частині інформаційного вікна (рис. 1.9).

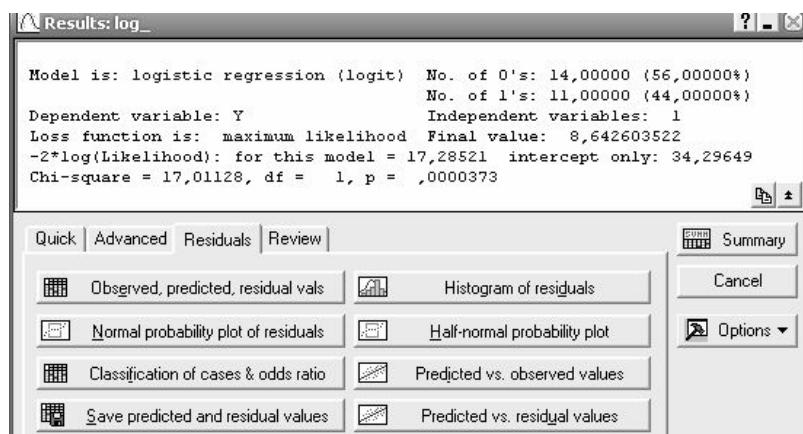


Рис. 1.9. Опції дослідження помилок моделі

Кнопка аналізу помилок *Observed, predicted residual vals* відображає спостережувані значення залежної змінної (*Observed*), теоретичні значення залежної змінної (прогнозовані значення ймовірностей за логіт-моделлю) (*Predicted*) і помилки моделі (*Residuals*) (рис. 1.10).

Model is: (log_)			
Dep. Var. : Y			
	Observed	Predicted	Residuals
1	0,000000	0,087136	-0,087136
2	0,000000	0,500288	-0,500288
3	0,000000	0,573204	-0,573204
4	1,000000	0,500288	0,499712
5	1,000000	0,113515	0,886485
6	0,000000	0,038035	-0,038035
7	0,000000	0,707341	-0,707341
8	0,000000	0,021499	-0,021499
9	1,000000	0,886717	0,113283
10	0,000000	0,066428	-0,066428
11	1,000000	0,987966	0,012034
12	0,000000	0,016114	-0,016114
13	1,000000	0,987966	0,012034
14	0,000000	0,050370	-0,050370
15	1,000000	0,813068	0,186932
16	0,000000	0,357458	-0,357458
17	0,000000	0,146596	-0,146596
18	1,000000	0,993277	0,006723
19	0,000000	0,038035	-0,038035
20	1,000000	0,357458	0,642542
21	0,000000	0,146596	-0,146596
22	0,000000	0,038035	-0,038035
23	1,000000	0,978550	0,021450
24	1,000000	0,886717	0,113283
25	1,000000	0,707341	0,292659

Рис. 1.10. Аналіз помилок моделі

Для всебічного аналізу помилок моделі будується графік розсіювання на нормальному ймовірнісному папері (*Normal probability plot of residuals*) та гістограма розподілу помилок за нормальним законом, (*Histogram of residuals*), які наведено на рис. 1.11.

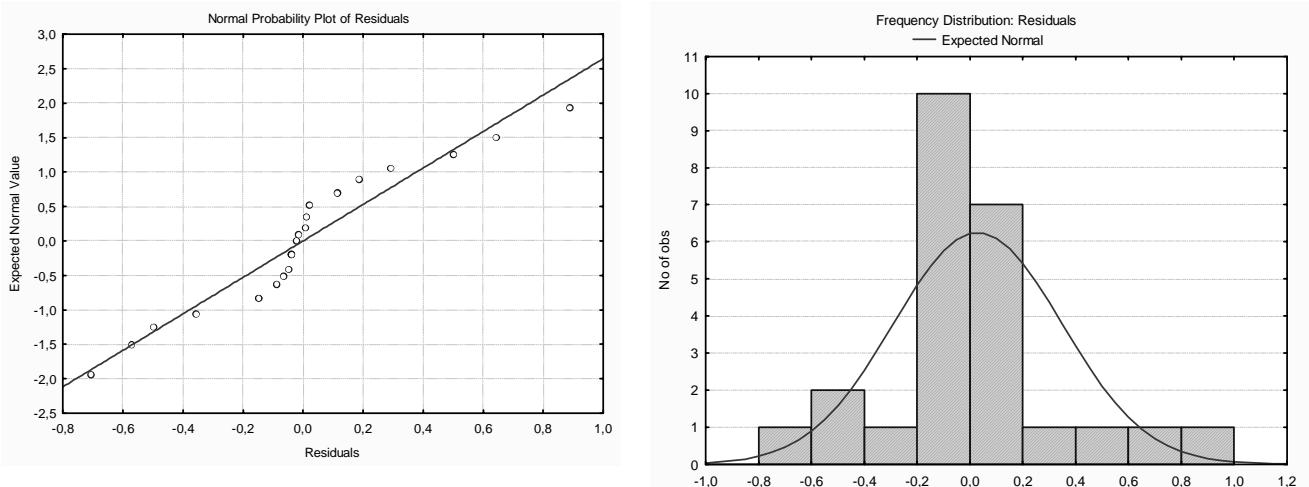


Рис. 1.11. Графік розсіювання та гістограма розподілу помилок

Ініціювавши опцію *Classification of cases & odds ratio* отримаємо

матрицю класифікації (рис. 1.12), в якій визначено кількість вірно та невірно класифікованих спостережень за побудованою моделлю, відсоток коректної класифікації (*Percent Correct*) та параметр відношення незгоди (*Odds ratio*) який визначає якість моделі.

		Classification of Cases (log_)		
		Odds ratio: 16,500		
Observed	Pred.	Pred.	Percent	
	0,000000	1,000000	Correct	
	0,000000	11	3	78,57143
	1,000000	2	9	81,81818

Рис. 1.12. Матриця класифікації за logit- моделлю

Для побудови пробіт- регресії для вирішення поставленої задачі необхідно вибрати опцію *Quick Probit regression* (рис. 1.13).

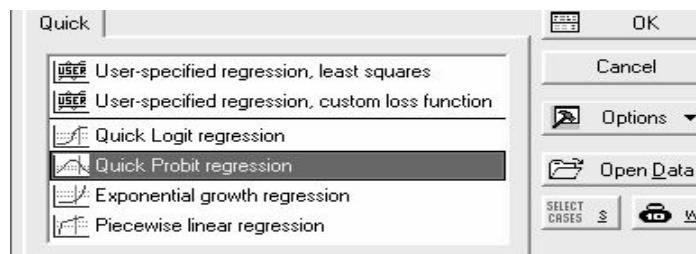


Рис. 1.13. Вибір напряму дослідження *Quick Probit regression*

Після проведення розрахунків на екрані з'явиться діалогове вікно результатів оцінювання за пробіт- регресією (рис. 1.14). Аналіз отриманих результатів оцінки адекватності моделі проводять аналогічно з логіт-моделлю. За критерієм χ^2 модель є адекватною, тобто в цілому рівняння є значимим.

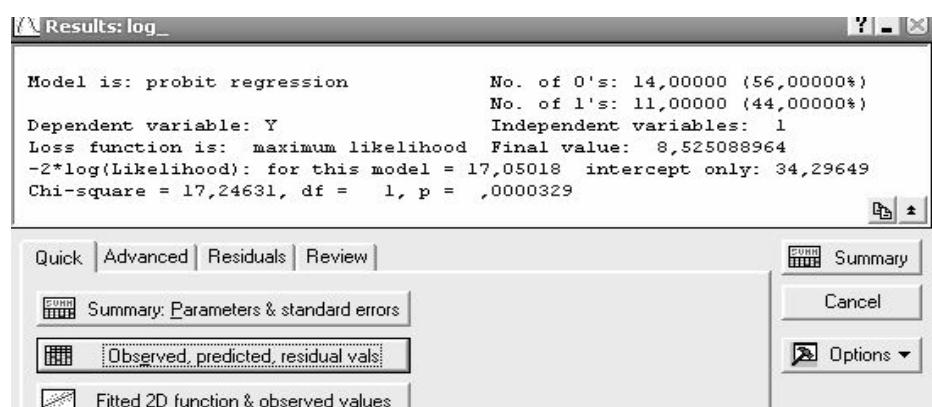


Рис. 1.14. Вікно результатів аналізу за probit- моделлю

Коефіцієнти моделі та їх характеристики наведено на рис. 1.15.

N=25	Model: Probit regression Dep. var: Y Loss: Max Final loss: 8,525088964	
	Const.B0	X
Estimate	-2,56613	0,172788
Std.Err.	0,90356	0,061572
t(23)	-2,84002	2,806291
p-level	0,00927	0,010024

Рис. 1.15. Коефіцієнти пробіт- регресії

Пробіт- модель залежності між досвідом роботи науково-педагогічних кадрів та їх участю в міжнародних грантах має вид:

$$F(u) = \Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{z^2}{2}} dz = \Phi(-2,56613 + 0,172788 \cdot x)$$

Прогнозовані значення ймовірностей за пробіт-моделлю (*Predicted*) і помилки моделі (*Residuals*) наведено на рис. 1.16.

	Model is: (log_)		
	Observed	Predicted	Residuals
1	0,000000	0,087450	-0,087450
2	0,000000	0,510250	-0,510250
3	0,000000	0,578667	-0,578667
4	1,000000	0,510250	0,489750
5	1,000000	0,118242	0,881758
6	0,000000	0,030398	-0,030398
7	0,000000	0,706800	-0,706800
8	0,000000	0,013191	-0,013191
9	1,000000	0,891624	0,108376
10	0,000000	0,063083	-0,063083
11	1,000000	0,995577	0,004423
12	0,000000	0,008348	-0,008348
13	1,000000	0,995577	0,004423
14	0,000000	0,044360	-0,044360
15	1,000000	0,813170	0,186830
16	0,000000	0,374529	-0,374529
17	0,000000	0,156000	-0,156000
18	1,000000	0,998477	0,001523
19	0,000000	0,030398	-0,030398
20	1,000000	0,374529	0,625471
21	0,000000	0,156000	-0,156000
22	0,000000	0,030398	-0,030398
23	1,000000	0,988459	0,011541
24	1,000000	0,891624	0,108376
25	1,000000	0,706800	0,293200

Рис. 1.16. Аналіз помилок моделі

Матриця класифікації за probit- моделлю представлена на рис. 1.17, параметр відношення незгоди (*Odds ratio*) перевищує одиницю, що підтверджує якість побудованої моделі.

		Classification of Cases (log_)		
		Odds ratio: 16,500		
Observed	Pred.	Pred.	Percent	
	0,000000	1,000000	Correct	
	0,000000	11	3	78,57143
1,000000		2	9	81,81818

Рис. 1.17. Матриця класифікацій за probit- моделлю

Графік розсіювання та гістограма розподілу помилок представлено на рис. 1.18.

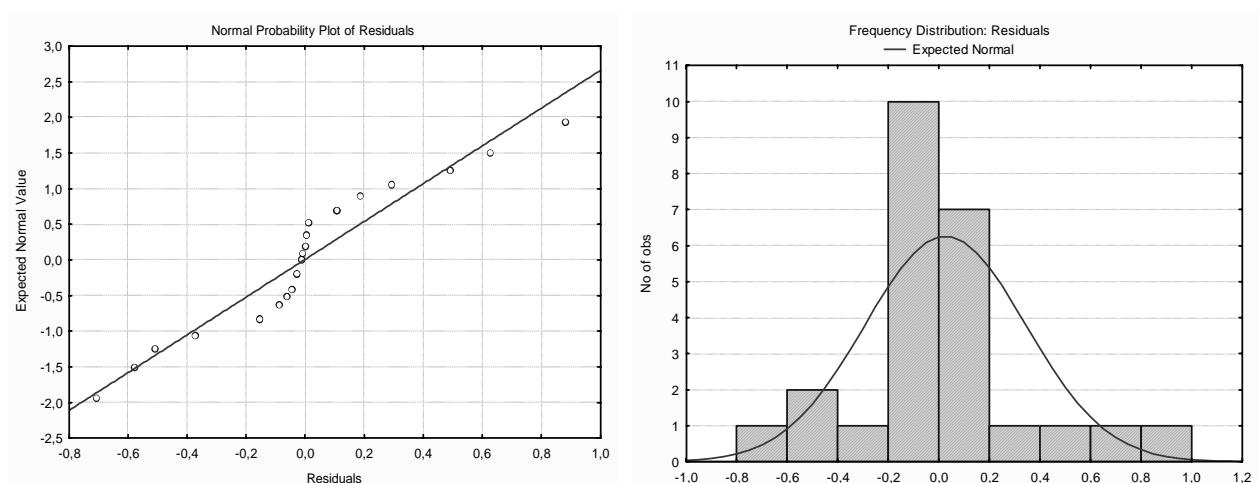


Рис. 1.18. Графік розсіювання та гістограма розподілу помилок

Так як досліджувані залежності є нелінійними, то при різних значеннях залежної змінної її зростання на одиницю дає різний приріст ймовірності. Розрахунок приросту ймовірності (маржинального ефекту) наведено в табл. 1.1.

Таблиця 1.1

Маржинальний ефект

Досвід роботи	Маржинальний ефект	
	logit- модель	probit- модель
1	0,004658	0,003932
5	0,014052	0,01619
15	0,073446	0,06891
20	0,044652	0,046405
30	0,003493	0,002242

Як видно з табл. 1.1, прогнози за моделями є досить близькими, особливо для науково-педагогічних кадрів з досвідом роботи від 5 до 20 місяців. Таким чином, найбільший приріст ймовірності участі в міжнародних грантах спостерігається для науково-педагогічних кадрів з досвідом роботи 15 – 20 місяців. При незначному досвіді, як і при великому, приріст ймовірності є досить незначним.