

	Клієнти банку здійснюють операції зняття грошей з рахунку, поповнення рахунку, платежі, оформлення кредитів. Час операцій, відповідно, 20 ± 10 , 30 ± 10 , 40 ± 10 , 40 ± 10 хвилин. Визначити відсоток операцій поповнення рахунку.	
--	---	--

Питання для самоконтролю

1. Що таке випадкова величина?
2. Який інструментарій дозволяє генерувати «випадкові числа» в MS Excel?
3. Як здійснюється Моделювання дискретних випадкових величин з відомим розподілом ймовірностей?
4. Для чого використовується поняття інтегральної функції розподілу?
5. Для чого використовується поняття диференціальної функції щільності розподілу?
6. Що таке математичне очікування?
7. Охарактеризувати рівномірний розподіл випадкових чисел?
8. Яка мінімальна вибірка рекомендується для експонентного розподілу?
9. Який розподіл залежить від великого числа незалежних випадкових факторів, кожний з яких окремо впливає на результат незначно?
10. Що таке потік подій Пуассона?
11. Який інструмент застосовується при проведенні імітаційних експериментів?
11. Яким образом можуть бути сформовані графіки диференціальних і інтегральних функцій розподілу?

Лабораторна робота № 7

Регресійний аналіз.

Мета роботи: *освоїти методи регресійного аналізу при побудові різного роду математичних моделей*

Вступ

Регресія, у теорії ймовірностей і математичній статистиці, залежність середнього значення якої-небудь величини від деякої іншої величини або від декількох величин. На відміну від чисто функціональної залежності $y=f(x)$, коли кожному значенню незалежної змінної x відповідає одне певне значення величини y , при регресійному зв'язку тому самому значенню x можуть відповідати залежно від випадку різні значення величини y . Якщо при кожному значенні $x=x_i$ спостерігається n_i значень $y_{i1} \dots y_{in_i}$ величини y , то залежність середніх арифметичних $\bar{y}_i = (y_{i1} + \dots + y_{in_i}) / n_i$ від $x=x_i$ і є регресією в статистичному розумінні цього терміна.

Рівняння лінійної парної регресії.

Рівняння лінійної парної регресії виглядає в такий спосіб:

$$Y = a_0 + a_1 X$$

За допомогою цього рівняння змінна Y виражається через константу a_0 і кут нахилу прямої (або кутовий коефіцієнт) a_1 , помножений на значення змінної X . Константу a_0 також називають вільним членом, а кутовий коефіцієнт - коефіцієнтом регресії. Параметри рівняння можуть бути визначені за допомогою методу найменших квадратів (МНК).

Метод найменших квадратів (у довідкових системах англійських програм - Least Squares Method, LS) є одним з основних методів визначення параметрів регресійних рівнянь, що дає найкращі лінійні незміщені оцінки.

Саме він використовується в MS Excel. Лінійні – належить до характеру взаємозв'язку змінних. Незміщені виходить, що очікувані значення коефіцієнтів регресії повинні бути

дійсними коефіцієнтами. Тобто крапки, побудовані за вихідними даними (x_i, y_i) , повинні лежати якнайближче до крапок лінії регресії. Сутність даного методу полягає в знаходженні параметрів моделі, при яких сума квадратів відхилень емпіричних (фактичних) значень результуючого ознаки від теоретичних, отриманих по обраному рівнянню регресії, тобто

$$S = \sum_{i=1}^n (y_i^{\delta} - y_i)^2 = \sum_{i=1}^n (y_i^p - a_0 - a_1 x)^2 \rightarrow \min,$$

де y_i^p – значення, обчислене по рівнянню регресії; $(y_i^p - y_i)$ – відхилення ε (помилка, остача) (рис. 1); n – кількість пар вихідних даних.

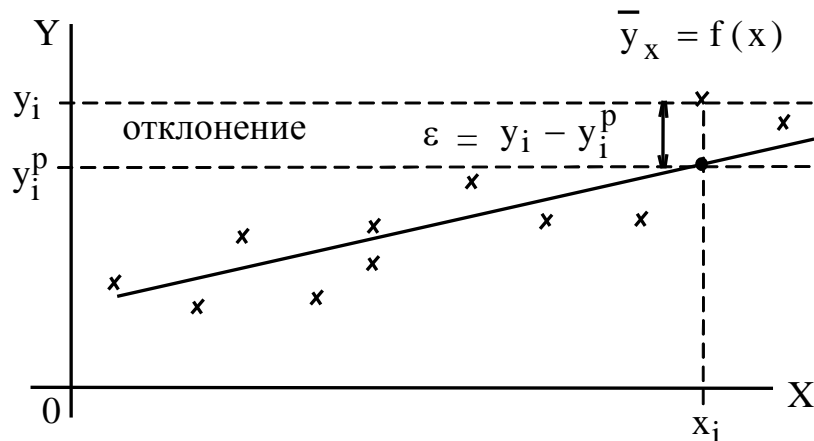


Рис. 1. Поняття відхилення ε для випадку лінійної регресії

У регресійному аналізі передбачається, що математичне очікування випадкової величини ε дорівнює нулю і її дисперсія однакова для всіх спостережуваних значень Y . Звідси випливає, що розсіювання даних біля лінії регресії повинне бути однаково при всіх значеннях параметра X . У випадку, показаному на рис. 2 дані розподіляються уздовж лінії регресії нерівномірно, тому метод найменших квадратів у цьому випадку не застосуємо.

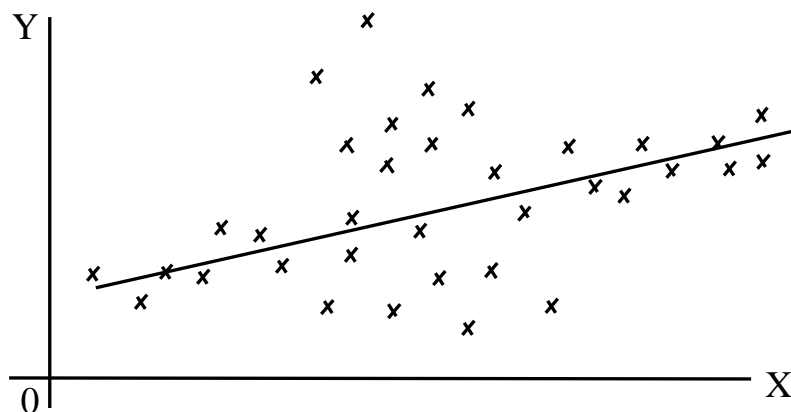


Рис.2. Нерівномірний розподіл вихідних точок уздовж лінії регресії

Провівши необхідні перетворення, одержимо систему двох рівнянь із двома невідомими a_0 і a_1 , які знайдемо, розв'язавши систему.

$$a_1 = \frac{n(\sum y_i x_i) - \sum y_i \sum x_i}{n(\sum x_i^2) - (\sum x_i)^2}; \quad (1)$$

$$a_0 = \frac{1}{n}(\sum y_i - a_1 \sum x_i) \quad (2)$$

Напрямок зв'язку між змінними визначається на підставі знаків (негативний або позитивний) коефіцієнта регресії (коефіцієнта a_1).

Якщо знак при коефіцієнті регресії - позитивний, зв'язок залежної змінної з незалежною буде позитивною. У нашому випадку знак коефіцієнта регресії позитивний, отже, зв'язок також є позитивним.

Якщо знак при коефіцієнті регресії - негативний, зв'язок залежної змінної з незалежною є негативною (зворотною).

Для аналізу загальної якості рівняння регресії використовують звичайно *множинний коефіцієнт детермінації* R^2 , називаний також квадратом коефіцієнта множинної кореляції R . R^2 (міра визначеності) завжди перебуває в межах інтервалу $[0;1]$.

Якщо значення R^2 близько до одиниці, це означає, що побудована модель пояснює майже всю мінливість відповідних змінних. І навпаки, значення R -Квдрата, близьке до нуля, означає погану якість побудованої моделі.

Коефіцієнт детермінації R^2 показує, на скільки відсотків ($R^2 \cdot 100\%$) знайдена функція регресії описує зв'язок між вихідними значеннями факторів X і Y

$$R^2 = \frac{\sum_{i=1}^n (y_i^p - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

де $(y_i^p - \bar{y})^2$ – пояснена варіація; $(y_i - \bar{y})^2$ – загальна варіація (рис.3).

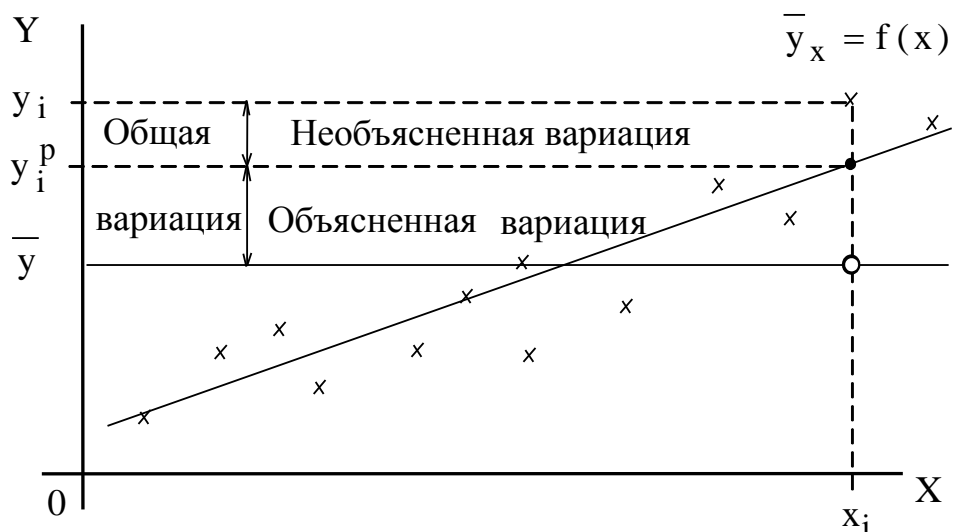


Рис. 3. Графічна інтерпретація коефіцієнта детермінації для випадку лінійної регресії

Відповідно, величина $(1 - R^2) \cdot 100\%$ показує, скільки відсотків варіації параметра Y обумовлені факторами, не включеними в регресійну модель. При високому ($R^2 \geq 75\%$) значенні коефіцієнта детермінації можна робити прогноз $y^* = f(x^*)$ для конкретного значення x^* .

Нелінійна регресія

Розглянемо найбільш прості випадки *нелінійної* регресії: гіперболу, експоненту й параболу. При знаходженні коефіцієнтів гіперболи й експоненти використовують прийом приведення нелінійної регресійної залежності до лінійного виду. Це дозволяє використовувати для обчислення коефіцієнтів функцій регресії вище наведені формули.

Гіпербола. Для приведення рівняння виду $y = a_0 + \frac{a_1}{x}$ до лінійного виду вводять нову змінну $z = \frac{1}{x}$, тоді рівняння гіперболи ухвалює лінійний вид $y = a_0 + a_1 z$. Після цього використовують формули (1) і (2) для знаходження лінійної функції, але замість значень x_i використовуються значення $z_i = \frac{1}{x_i}$:

$$a_1 = \frac{n(\sum y_i z_i) - \sum y_i \sum z_i}{n(\sum z_i^2) - (\sum z_i)^2}; \quad a_0 = \frac{1}{n}(\sum y_i - a_1 \sum z_i). \quad (3)$$

Експонента. Для приведення до лінійного виду рівняння експоненти $y = a_0 e^{a_1 x}$ проведемо логарифмування

$$\begin{aligned} \ln y &= \ln(a_0 e^{a_1 x}); \\ \ln y &= \ln a_0 + \ln(e^{a_1 x}); \\ \ln y &= \ln a_0 + a_1 x. \end{aligned}$$

Введемо змінні $b_0 = \ln a_0$ й $b_1 = a_1$, тоді $\ln y = b_0 + b_1 x$, звідки випливає, що можна застосовувати формули (1) і (2), у яких замість значень y_i треба використовувати $\ln y_i$:

$$b_1 = \frac{n(\sum [\ln y_i] x_i) - \sum \ln y_i \sum x_i}{n(\sum x_i^2) - (\sum x_i)^2}; \quad b_0 = \frac{1}{n}(\sum \ln y_i - b_1 \sum x_i) \quad (4)$$

При цьому ми отримуємо чисельні значення коефіцієнтів b_0 і b_1 , від яких треба перейти до a_0 й a_1 , використовуваних у моделі експоненти. Виходячи з введених позначень і визначення логарифма, отримуємо

$$a_0 = e^{b_0}, \quad a_1 = b_1.$$

Парабола. Для знаходження коефіцієнтів рівняння параболи $y = a_0 + a_1 x + a_2 x^2$ необхідно розв'язати лінійну систему із трьох рівнянь:

$$\begin{cases} n \cdot a_0 + (\sum x_i) a_1 + (\sum x_i^2) a_2 = \sum y_i, \\ (\sum x_i) a_0 + (\sum x_i^2) a_1 + (\sum x_i^3) a_2 = \sum (y_i x_i), \\ (\sum x_i^2) a_0 + (\sum x_i^3) a_1 + (\sum x_i^4) a_2 = \sum (y_i x_i^2). \end{cases}$$

Чинність регресійного зв'язку для гіперболи й параболи визначається безпосередньо по тій же формулі що й для лінійної моделі. При обчисленні коефіцієнта детермінації для експоненти всі значення параметра Y (вихідні, регресійні, середні) необхідно замінити на їхні логарифми, наприклад, y_i^p – на $\ln(y_i^p)$ і т.д.

Якщо функція регресії визначена, інтерпретована й обґрунтована, і оцінка точності регресійного аналізу відповідає вимогам, можна вважати, що побудована модель і прогнозні значення мають достатню надійність.

Прогнозні значення, отримані таким способом, є середніми значеннями, які можна чекати.

Методичні рекомендації

Для проведення регресійного аналізу й прогнозування необхідно:

- 1) **побудувати графік** вихідних даних і спробувати зряче, приблизно визначити характер залежності;
- 2) **вибрати вид функції** регресії, яка може описувати зв'язок вихідних даних;
- 3) **визначити чисельні коефіцієнти** функції регресії методом найменших квадратів;
- 4) **оцінити чинність** знайденої регресійної залежності на основі коефіцієнта детермінації R^2 ;

5) **зробити прогноз** (при $R^2 \geq 75\%$) або зробити висновок про неможливість прогнозування за допомогою знайденої регресійної залежності. При цьому не рекомендується використовувати модель регресії для тих значень незалежного параметра X, які не належать інтервалу, заданому у вихідних даних.

Довідкова інформація з технології роботи з режимом "Регресія" надбудови Пакет аналізу MS Excel

Режим роботи "Регресія" слугує для розрахунків параметрів рівняння лінійної регресії й перевірки його адекватності досліджуваному процесу.

Для розв'язання задачі регресійного аналізу в MS Excel вибираємо в меню **Сервіс** команду **Аналіз даних** і інструмент аналізу "**Регресія**".

У діалоговому вікні, що з'явилося, задаємо наступні параметри:

1. **Вхідний інтервал Y** - це діапазон даних за результативною ознакою. Він повинен складатися з одного стовпця.
2. **Вхідний інтервал X** - це діапазон комірок, що містять значення факторів (незалежних змінних). Число вхідних діапазонів (стовпців) повинне бути не більше 16.
3. Прапорць **Мітки**, установлюється в тому випадку, якщо в першому рядку діапазону коштує заголовков.
4. Прапорць **Рівень надійності** активізується, якщо в поле, що перебуває поруч із ним необхідно ввести рівень надійності, відмінний від установленого за замовчуванням. Використовується для перевірки значимості коефіцієнта детермінації R^2 і коефіцієнтів регресії.
5. **Константа нуль**. Даний прапорць необхідно встановити, якщо лінія регресії повинна пройти через початок координат ($a_0=0$).
6. **Вихідний інтервал/ Новий робочий лист/ Нова робоча книга** – указати адресу верхньої лівої комірки вихідного діапазону.
7. Прапорці в групі **Залишки** встановлюються, якщо необхідно включити у вихідний діапазон відповідні стовпці або графіки.
8. Прапорць **Графік нормальної ймовірності** необхідно зробити активним, якщо потрібно вивести на лист крапковий графік залежності спостережуваних значень Y від автоматично формованих інтервалів перцентілей.

Після натискання кнопки ОК у вихідному діапазоні одержуємо звіт.

Приклад виконання роботи

Задача: Деяка фірма займається поставками різних вантажів на короткі відстані усередині міста. Оцінити вартість таких послуг, що залежить від затрачуваного на поставку часу. У якості найбільш важливого фактору, що впливає на час поставки, обрана пройдена відстань. Були зібрані вихідні дані про десять поставок (таблиця 1)

Таблиця 1

Відстань, миль	3,5	2,4	4,9	4,2	3,0	1,3	1,0	3,0	1,5	4,1
Час, хв.	16	13	19	18	12	11	8	14	9	16

Визначте характер залежності між відстанню й витраченим часом, використовуючи майстер діаграм MS excel, проаналізуйте застосовність методу найменших квадратів, побудуйте рівняння регресії, використовуючи МНК, проаналізуйте силу регресійного зв'язку. Проведіть регресійний аналіз, використовуючи режим роботи "Регресія" в MS Excel і порівняйте з результатами, отриманими раніше. Зробіть прогноз часу поїздки на 2 милі. Порахувати й побудувати графічно міру помилки регресійної моделі використовуючи табличний процесор Excel.

Рішення

На графіку (рис. 4) будуємо вихідні дані по десяти поїздкам.

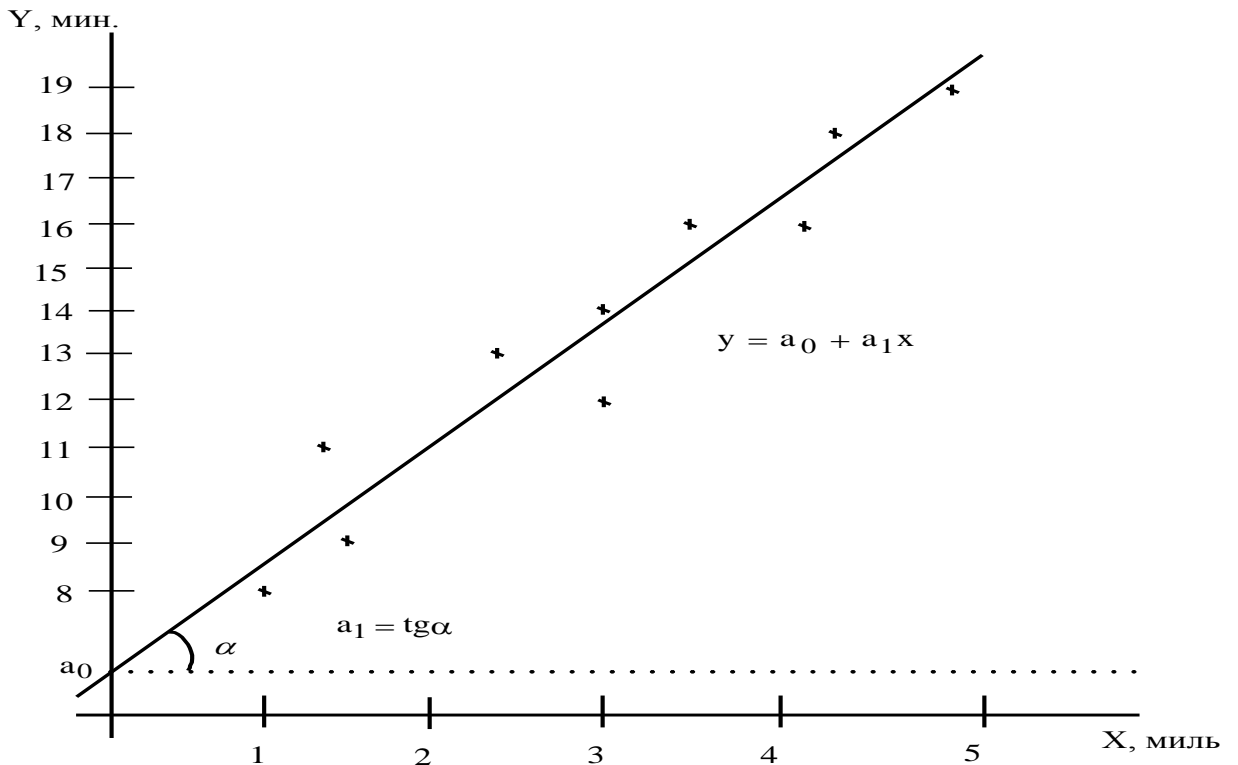


Рис.4. Графік вихідних даних і передбачувана лінія регресії

Крім відстані на час поставки впливають пробки на дорогах, час доби, дорожні роботи, погода, кваліфікація водія, вид транспорту. Побудовані крапки не перебувають точно на лінії, що обумовлене описаними вище факторами. Але ці крапки зібрані навколо прямої лінії, тому можна припустити лінійний зв'язок між параметрами. Усі вихідні точки рівномірно розподілені уздовж передбачуваної прямої лінії, що дозволяє застосувати метод найменших квадратів.

Обчислимо суми, необхідні для розрахунків коефіцієнтів рівняння лінійної регресії й коефіцієнта детермінації R^2 за допомогою допоміжної таблиці (таблиця 2).

Таблиця 2

X_i	Y_i	x_i^2	$x_i Y_i$	y_i^p	$(y_i^p - \bar{y})^2$	$(y_i - \bar{y})^2$
3,5	16	12,25	56,00	15,223	2,634129	5,76
2,4	13	5,76	31,2	12,297	1,697809	0,36
4,9	19	24,01	93,1	18,947	28,59041	29,16
4,2	18	17,64	75,60	17,085	12,14523	19,36
3,0	12	9,00	36,00	13,893	0,085849	2,56
1,3	11	1,69	14,30	9,371	17,88444	6,76
1,0	8	1,00	8,00	8,573	25,27073	31,36

3,0	14	9,00	42,00	13,893	0,085849	0,16
1,5	9	2,25	13,50	9,903	13,66781	21,16
4,1	16	16,81	65,60	16,819	10,36196	5,76
$\Sigma = 28,9$	$\Sigma = 136$	$\Sigma = 99,41$	$\Sigma = 435,30$	-	112,4242	122,4

$$\bar{y} = \frac{\sum y_i}{n} = \frac{16+13+19+18+12+11+8+14+9+16}{10} = 13,6.$$

Обчислимо коефіцієнти лінійної регресії по формулах (1) і (2):

$$a_1 = \frac{10 \cdot 435,30 - 136 \cdot 28,9}{10 \cdot 99,41 - 835,21} = 2,660 ;$$

$$a_0 = 0,1 \cdot (136 - 2,660 \cdot 28,9) = 5,913.$$

Таким чином, шукана регресійна залежність має вигляд:

$$y^p = 5,913 + 2,660x.$$

Нахил лінії регресії $a_1 = 2,66$ хвилин на милю – це кількість хвилин, що доводиться на одну милю відстані. Координата точки перетинання прямої з віссю Y $a_0 = 5,913$ хвилин – це час, який не залежить від пройденої відстані, а обумовлюється всіма іншими можливими факторами, явно не врахованими при аналізі.

Обчислимо коефіцієнт детермінації:

$$R^2 = \frac{112,424}{122,400} = 0,918 \text{ або } 91,8\%.$$

Проведемо регресійний аналіз із використанням режиму Регресія MS Excel. Значення параметрів, установлених в однойменному діалоговому вікні, представлені на рис.5.

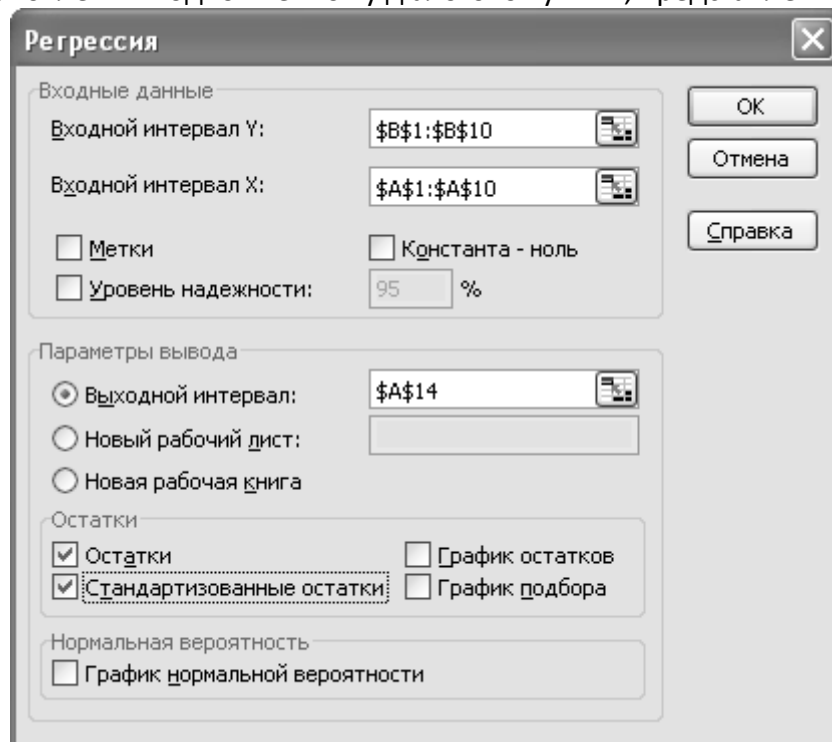


Рис. 5.

Згенеруються результати по регресійній статистиці, представлені в таблиці 3.

Таблица 3.

ВИВІД ПІДСУМКІВ	
-----------------	--

<i>Регресійна статистика</i>	
Множинний R	0,958275757
R-Квадрат	0,918292427
Нормований R-Квадрат	0,90807898
Стандартна помилка	1,11809028
Спостереження	10

Розглянемо представлену в таблиці 3 регресійну статистику.

Величина R-Квадрат, називана також заходом визначеності, характеризує якість отриманої регресійної прямої. Ця якість виражається ступенем відповідності між вихідними даними й регресійною моделлю (розрахунковими даними). Захід визначеності завжди перебуває в межах інтервалу [0;1]. У нашому прикладі захід визначеності рівний 0,91829, що говорить про дуже гарне припасування регресійної прямої до вихідних даних і збігається з коефіцієнтом детермінації R^2 , обчисленим по формулі.

Таким чином, лінійна модель пояснює 91,8% варіації часу доставки, що означає правильність вибору фактору (відстані). Не пояснюється $100\% - 91,8\% = 8,2\%$ варіації часу поїздки, які обумовлені іншими факторами, що впливають на час поставки, але не включеними в лінійну модель регресії.

Розрахований рівень значимості $\alpha_p = 1,26E-05 < 0,05$ (показник значимість F у таблиці Дисперсійний аналіз) підтверджує значимість R^2 .

Множинний R - коефіцієнт множинної кореляції R - виражає ступінь залежності незалежних змінних (X) і залежної змінної (Y) і дорівнює квадратному кореню з коефіцієнта детермінації, ця величина ухвалює значення в інтервалі від нуля до одиниці. У простому лінійному регресійному аналізі множинний R дорівнює коефіцієнту кореляції Пірсона. Дійсно, множинний R у нашому випадку дорівнює коефіцієнту кореляції Пірсона (0,95827), який обчислюється по формулі:

$$r_{xy} = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] \cdot [n \sum y^2 - (\sum y)^2]}}$$

Тепер розглянемо середню частину розрахунків, представлену в таблиці 4 (наведена в скороченому варіанті). Тут дані коефіцієнт регресії a_1 (2,65970168) і зсув по осі ординат, тобто константа a_0 (5,913462144).

Таблиця 4

	<i>Коефіцієнти</i>	<i>Стандартна помилка</i>	<i>t-статистика</i>	<i>P-значення</i>
Y-Перетинання	5,913462144	0,884389599	6,686489927	0,00015485
Змінна X 1	2,65970168	0,280497238	9,482095791	1,26072E-05

Виходячи з розрахунків, можемо записати рівняння регресії в такий спосіб:

$$y^{\hat{}} = 5,91346 + 2,65970x. (*)$$

Бачимо, що це рівняння, збігається з рівнянням, отриманим нами при розрахунках по МНК вручну з точністю до помилки округлення.

Напрямок зв'язку між змінними визначається на підставі знаків (негативний або позитивний) коефіцієнта регресії (коефіцієнта a_1). У нашому випадку знак коефіцієнта регресії позитивний, отже, зв'язок також є позитивним.

Далі перевіримо значимість коефіцієнтів регресії: a_0 і a_1 . Порівнюючи попарно значення стовпців *Коефіцієнти* й *Стандартна помилка* в таблиці 4, бачимо, що абсолютні значення коефіцієнтів більше чим їх стандартні помилки. До того ж ці коефіцієнти є значущими, про що

можна судити за значеннями показника Р-Значення в таблиці 4, які менше заданого рівня значимості $\alpha=0,05$.

Таблиця 5

ВИВІД ЗАЛИШКУ			
Спостереження	Передвіщене Y	Залишки	Стандартні залишки
1	15,22241803	0,777581975	0,737641894
2	12,29674618	0,703253823	0,667131568
3	18,94600038	0,053999622	0,051225961
4	17,0842092	0,915790799	0,868751695
5	13,89256718	-1,892567185	-1,795356486
6	9,371074328	1,628925672	1,545256778
7	8,573163824	-0,573163824	-0,543723571
8	13,89256718	0,107432815	0,101914586
9	9,903014664	-0,903014664	-0,8566318
10	16,81823903	-0,818239033	-0,776210624

У таблиці 5. представлені результати виводу залишків. За допомогою цієї частини звіту ми можемо бачити відхилення кожної точки від побудованої лінії регресії. Найбільше абсолютне значення залишку в нашому випадку - 1,89256, найменше - 0,05399. Для кращої інтерпретації цих даних скористаємося графіком вихідних даних і побудованою лінією регресії, представленими на рис. 6. Як бачимо, лінія регресії добре "підігнана" під значення вихідних даних.

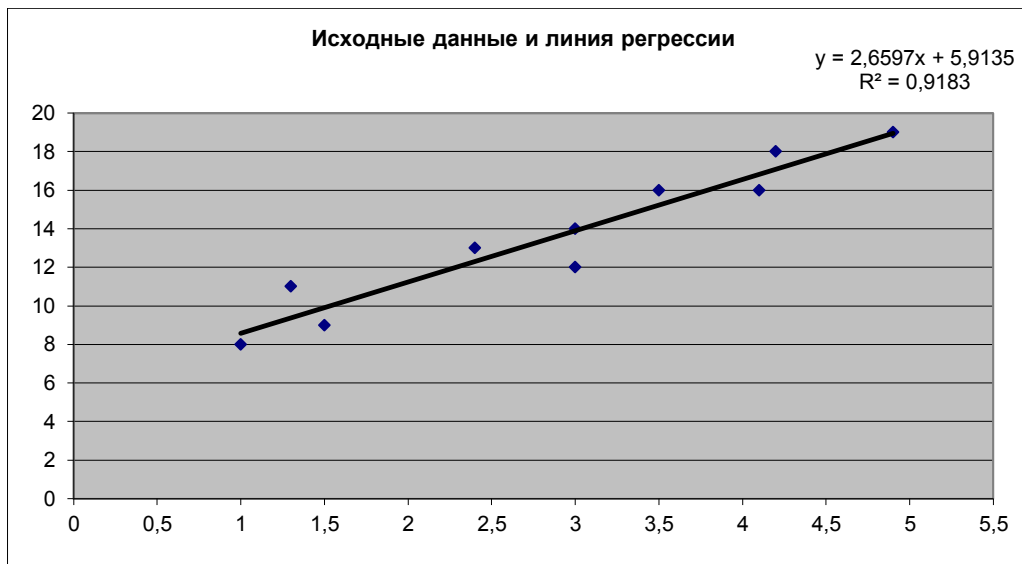


Рис.6.

Приблизним, але найпростішим і наочним способом перевірки задовільності регресійної моделі є графічна представлення відхилень:

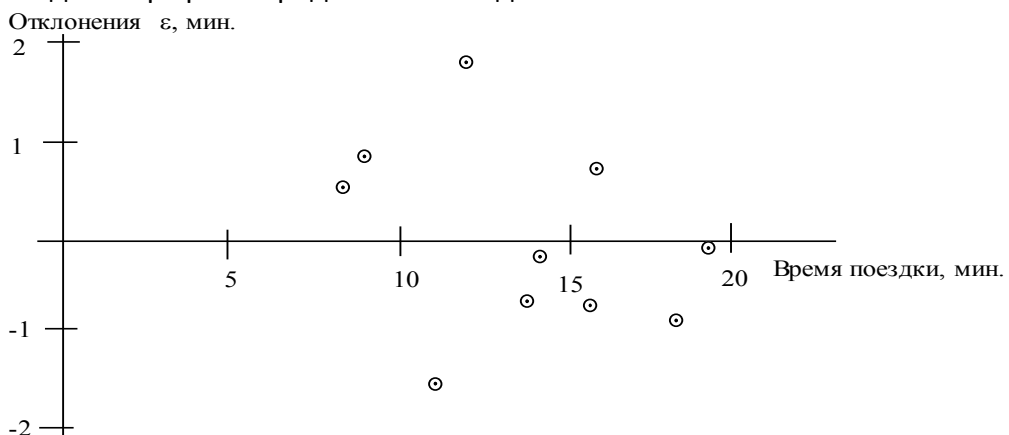


Рис. 7. Графік відхилень

Відкладемо відхилення $(y_i^p - y_i)$ по осі Y, для кожного значення y_i . Якщо регресійна модель близька до реальної залежності, то відхилення будуть носити випадковий характер і їх

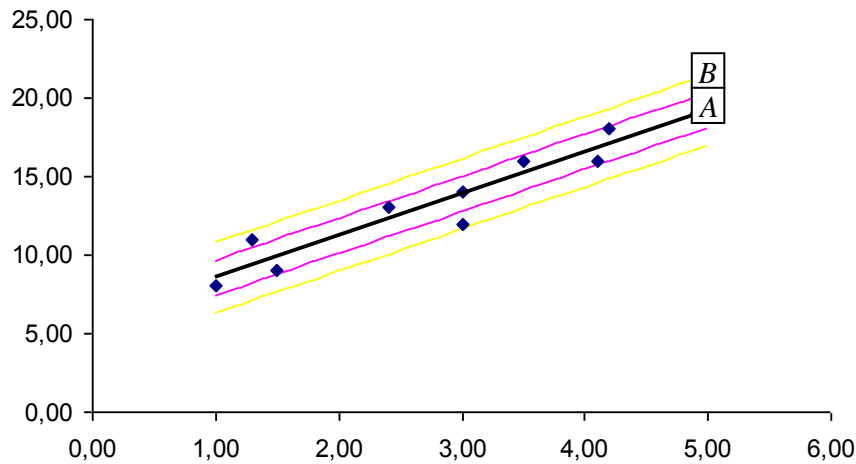


Рис.8. Регресійна модель і відхилення

сума буде близька до нуля. У розглянутому прикладі $\sum_{i=1}^n (y_i^p - y_i) = 0,004$.

Звичайно заходом помилки регресійної моделі служить середнє квадратичне відхилення

$$\sigma_{\varepsilon} = \left[\left(\sum_{i=1}^n \varepsilon_i^2 \right) / (n-2) \right]^{1/2} = \left\{ \left[\sum_{i=1}^n (y_i^p - y_i)^2 \right] / (n-2) \right\}^{1/2}$$

Для нормально розподілених процесів приблизно 67% крапок перебуває в межах одного відхилення σ_{ε} від лінії регресії й 95% - у межах $2\sigma_{\varepsilon}$ (на малюнку 8 трубки А и В відповідно).

Розв'яжемо задачу прогнозування. Оскільки коефіцієнт детермінації R^2 має досить високе значення й відстань 2 милі, для якого треба зробити прогноз, перебуває в межах діапазону вихідних даних (таблиця 1), те ми можемо використовувати отримане рівняння лінійної регресії для прогнозування

$$y^*(2 \text{мили}) = 5,913 + 2,660 \cdot 2 = 11,2 \text{ хвилин.}$$

При прогнозах на відстані, що не входять у діапазон вихідних даних, не можна гарантувати справедливості отриманої моделі. Це пояснюється тим, що зв'язок між часом і відстанню може змінюватися в міру збільшення відстані. На час далеких перевезень можуть впливати нові фактори такі, як використання швидкісних шосе, зупинки на відпочинок, обід і т.п.

Таким чином, у результаті використання регресійного аналізу в пакеті Microsoft Excel ми:

- побудували рівняння регресії;
- установили форму залежності й напрямок зв'язку між змінними - позитивна лінійна регресія, яка виражається в рівномірному росту функції;
- установили напрямок зв'язку між змінними;

- оцінили якість отриманої регресійної прямої;
- змогли побачити відхилення розрахункових даних від даних вихідного набору;
- передбачили майбутнє значення залежної змінної.

Варіанти завдань

Завдання №1

Побудуйте регресійну модель (лінійну) для вихідних даних, наведених у таблиці. Для полегшення розрахунків вихідні дані містять тільки чотири пари значень (x_i, y_i) .

Варіанти

№	Координати	Точки				x*
		1	2	3	4	
1	X	1	2	3	4	1.6
	Y	30	7	8	1	?
2	X	1	2	3	4	2.3
	Y	25	7	7	2	
3	X	9	5	2	3	2.9
	Y	25	7	7	2	?
4	X	1	2	3	4	2.6
	Y	15	10	7	0.5	?
5	X	10	3	6	4	8
	Y	25	7	7	2	?
6	X	9	5	2	3	2.5
	Y	15	8.5	7.5	5	?
7	X	2	3	7	8	7.5
	Y	11	8.5	6.5	5	?
8	X	10	3	6	4	9
	Y	15	7	8	6	?
9	X	2	3	4	5	4.5
	Y	13	9	8	7	?
10	X	1	2	3	4	1.5
	Y	7.5	7	5	3.5	?
11	X	1	2	3	4	3.6
	Y	13	9	8	7	?
12	X	3	4	6	10	8
	Y	7.5	7	6.5	3.5	?
13	X	3	4	5	6	7.8
	Y	9	7	5	3	?
14	X	7	5.6	13	14.7	15
	Y	7.5	7	5	3.5	?
15	X	9	5	2	3	5.7
	Y	13	9	8	7	?

Дослідіть модель за допомогою режиму Регресія в MS Excel і зробіть прогноз для x^* .

Завдання № 2

Для вихідних даних, представлених у таблиці, були побудовані наступні регресійні моделі:

- $y = 6,067 - 0,085x$;
- $y = -2,017 + 3,957x - 0,367x^2$;
- $y = 5,918e^{-0,043x}$.

Вихідні дані задачі №2

X	3	8	5	10	7	6	4	9	1	2
Y	6	5	9	1	8	9	8	4	2	4

За допомогою графіка відхилень виберіть задовільну модель і перевірте свій вибір за допомогою відповідного розрахунків.

Завдання №3

У таблиці представлені дані про ціни на комплектуючі для ПЕВМ. Комплектуючі проводяться різними компаніями-виробниками й розбиті на групи по своїх функціональних можливостях.

Вихідні дані задачі №3

Група	1	1	2	2	2	3	3	3	4	4
Ціна, \$	50	60	70	80	95	100	115	120	105	120
Група	4	5	5	5	6	6	6	7	7	7
Ціна, \$	130	110	150	190	120	130	220	145	265	270

Побудуйте графік вихідних даних і з його допомогою проаналізуйте застосовність методу найменших квадратів. Підтвердите свої висновки за допомогою розрахунків (для лінійної моделі). Прокоментуйте економічні причини отриманого результату.

Питання для самоконтролю

1. Що таке регресія?
2. Пояснити фізичний зміст МНК?
3. Який показник використовується для аналізу загальної якості рівняння регресії?
4. Що дозволяє судити про якість побудови регресійної моделі?
5. У чому відмінність лінійної й нелінійної регресії?
6. Який підхід використовується для спрощення розв'язання задач нелінійної регресії?
7. Який інструментарій MS "Excel" застосовується для розв'язання задач регресійного аналізу?
8. Що є мірою помилки регресійної моделі?