МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»
«Кафедра програмної інженерії та інформаційних технологій управління»

Звіт з лабораторних робіт
з дисципліни «Математична статистика»

Виконав:
ст. гр. КН-221в
Шулюпов Є.Р.

Перевірив:
проф. каф. ПІІТУ
Козуля Т.В.

Харків 2022

Задача №1

Постановка завдання

В результаті вимірювання(X) температури розділу фракції бензин-авіакеросин на установці первинної переробки нафти отримано значення температур(Y), що наведенні в таблиці 1.

Таблица 1

N	Значение	N	Значение	N	Значение	N	Значение
1	133,5	14	141,5	27	144,0	40	137,5
2	142,0	15	139,0	28	142,5	41	141,5
3	145,5	16	140,5	29	139,0	42	141,0
4	144,5	17	139,0	30	137,0	43	142,5
5	134,5	18	143,5	31	136,0	44	143,5
6	138,5	19	139,5	32	137,0	45	141,0
7	144,0	20	140,5	33	138,5	46	147,0
8	141,0	21	140,0	34	139,0	47	139,5
9	141,5	22	138,5	35	139,5	48	136,5
10	139,5	23	135,0	36	140,5	49	142,0
11	140,0	24	139,5	37	139,5	50	140,0
12	145,0	25	139,0	38	140,0		
13	141,5	26	138,0	39	140,5		

Обчислити основні числові характеристики для обсягу вибірки (n=50) використовуючи статистичний додаток Excel.

Розв’язання

Аби обчислити основні числові характеристики скористуємося аналізом данних статистичного додатку Excel(рис. 1):

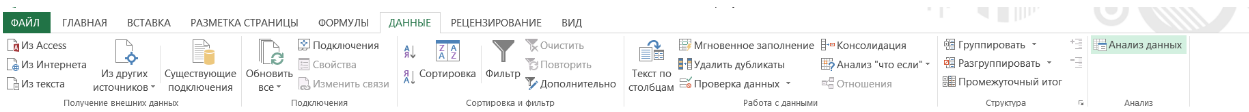

Рисунок 1.

Далі нам потрібно обрати пункт “Описова сатистика”, щоб отрмати необхідні данні.(рис. 2)

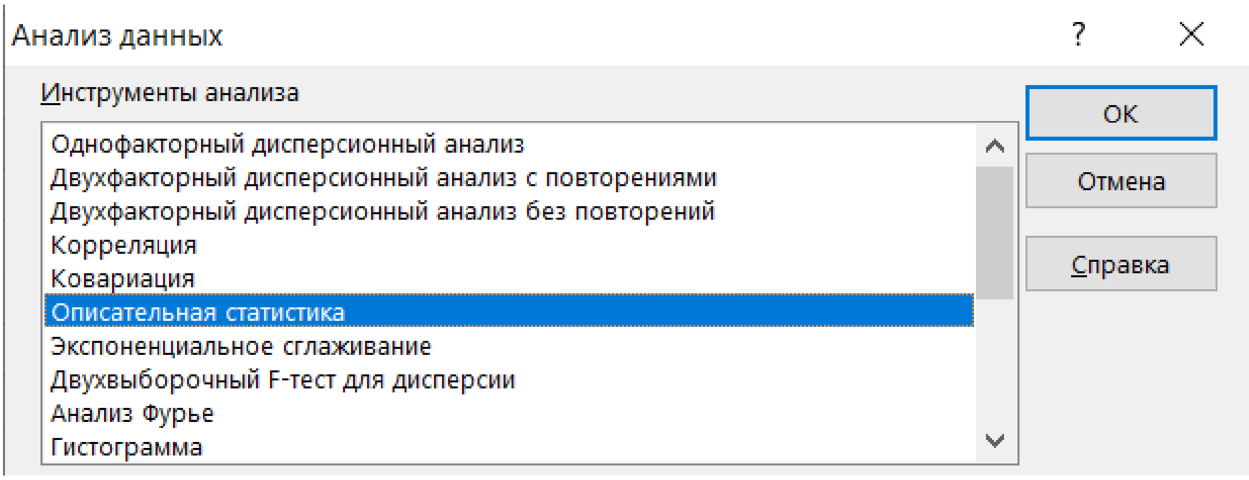

Рисунок 2.

У відкритому вікні нам необхідно обрати вхідний інтервал значень Y та визначити вихідний інтервал, також доцільно відмітити підсумкову статистику, аби отримати саме числові характеристики випадкових велечин. (рис. 3)

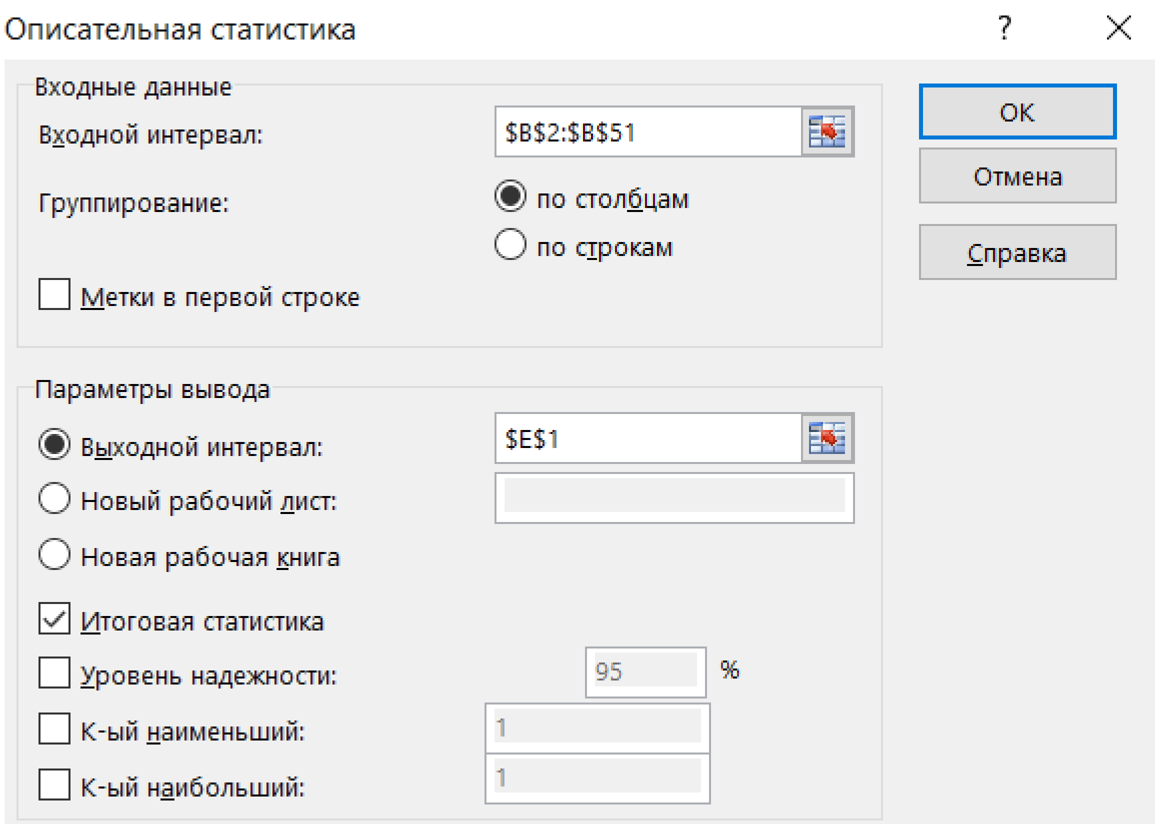

Рисунок 3.

Отриманні дані слід дослідити та визначити: “яку інформацію нам можуть надати ці данні?”(рис. 4)

Среднее	140,19
Стандартная ошибка	0,393801462
Медиана	140
Мода	139,5
Стандартное отклоне	2,784596845
Дисперсия выборки	7,753979592
Эксцесс	0,324636258
Асимметричность	0,019901466
Интервал	13,5
Минимум	133,5
Максимум	147
Сумма	7009,5
Счет	50

Рисунок 4.

Отже: ми отримали середнє аремфетичне значеня, стандартну помилку, медіану, моду, стандартне відхилення, дисперсію вибірки, єксцесс, асимметричність, інтервал, мінімум та максимум, а також загальну суму і об'єм вибірки. Випадкові велечини є дискретними, тому підпорядковуються відповідному закону розподілення.

Задача №2

Постановка завдання

З нормальної генеральної сукупності з відомою дисперсією $\sigma^2 = 1{,}44$ вилучено вибірку обсягу n=49 і за нею знайдено середнє значення $\bar{x} = 3{,}8;$ рівень значимості $\alpha = 0{,}05$, нульова гіпотеза $H_0 : \mu = 3$ при конкуруючій гіпотезі $H_1 : \mu > 3$.

1. **Знайти довірчий інтервал для математичного очікування.**
2. **Перевірити попадання μ_0 на інтервал.**
3. **Перевірити нульову гіпотезу при конкуруючій гіпотезі.**

Розв'язання

Предметом пошуку є не доказ однієї з гіпотез, а відкидання найменш вірогідної, аби звузити коло пошуку.

Нам вже відомо середнє значення вибірки, а для α = 0,05 відповідне табличне значення, згідно функції Лапласа, z_α =1,96.

$$z_\alpha = \frac{\delta\sqrt{n}}{\sigma}$$

α	0,01	0,05	0,1
z_α	2,576	1,96	1,64

Отже, ми отримаємо довірчий інтервал для параметра μ відомої дисперсії, скориставшись наступною формулою:

$$\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$$

Проведемо прості математичні операції:

$\sqrt{n} = 7;\ \sigma = 1{,}2;$

Перейдемо до підстановки числових характеристик до Excel.

=ДОВЕРИТ(L2; P3; P4)

J	K	L	M	N	O	P
α	0,01	0,05	0,1			
Zα	2,576	1,96	1,64		Станд. відхилення	1,2
					n	49
Довірчий інтервал	0,33599					
x	3,8					

Відповідно до отриманого довірчого інтервалу засобами формули «ДОВЕРИТ», ми можемо зробити висновок, що нульва гіпотеза не є адекватною для вибірки (μ належить [3,8 – 0,33599; 3,8 + 0,33599]) , що базується на основі генеральної сукупності, а конкуруюча гіпотеза, навпаки – є найбільш вірогідною.

Задача №3

Постановка завдання

За двома вибірками n1 і n2, вилученим з нормальних генеральних сукупностей, знайдено $\bar{x}, \bar{y}, s_1^2, s_2^2$. Перевірити нульову гіпотезу $H_0 : \mu_1 = \mu_2$ при конкуруючій гіпотезі $H_1 : \mu_1 \neq \mu_2$, якщо $n_1 = 30; n_2 = 20; \bar{x} = 10; \bar{y} = 12,5; s_1^2 = 12; s_2^2 = 10; \alpha = 0,05.$

1. **Отримати табличні значення t-критерію розподілу Стюдента, відповідно до отриманих ступеннів свободи.**
2. **Визначити t-критерій згідно розподілу Стюдента.**
3. **Перевірити нульову гіпотезу при конкуруючій гіпотезі.**

Розв'язання

Предметом пошуку є не доказ однієї з гіпотез, а відкидання найменш вірогідної, аби звузити коло пошуку.

Для побудови довірчого інтервалу використовується величина, що має розподіл Стьюдента з $\nu = n_1 + n_2 - 2$ ступенями свободи, використовуючи наступні формули:

$$s^2 = \frac{(n_1 - 1)\cdot s_1^2 + (n_2 - 1)\cdot s_2^2}{n_1 + n_2 - 2}; \qquad t = \frac{\bar{x} - \bar{y}}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

Критичне значення t знаходимо за таблицями двостороннього критерію Стьюдента для α = 0.05 та числа ступенів свободи $\nu = n_1 + n_2 - 2 = 48.$

$t_{\text{табл}}$ = 2, 0086

	n	s^2	x	y	α
1	30	12	10	12,5	0,05
2	20	10			

ss	11,20833333
t	2,586783681

Відповідно до отриманого значення критерія Стюдента, ми можемо зробити висновок, що так, як $t > t_{\text{табл}}$, то нульва гіпотеза нерелевантна.

Задача №4

Постановка завдання

За двома незалежними вибірками обсягів n = 5 і m = 6, витягнутим із нормальних генеральних сукупностей, знайдені вибіркові середні $\bar{x} = 15,9; \bar{y} = 14,1;$ та вибіркові дисперсії $s_1^2 = 14,76; s_2^2 = 4,92.$ При рівній значимості $\alpha = 0,05.$

1. Обрахувати значення верхньої і нижньої границі згідно розподілу Пірсона.
2. Визначити t-критерій згідно розподілу Стюдента.
3. Знайти довірчий інтервал для дисперсії.
4. Знайти довірчий інтервал для математичного очікування для обох вибірок.
5. Перевірити нульову гіпотезу при конкуруючій гіпотезі у всіх трьох випадках.

а) $H_0 : \sigma_1^2 = \sigma_2^2,\ H_1 : \sigma_1^2 > \sigma_2^2.$ **б)** $H_0 : \mu_1 = \mu_2,\ H_1 : \mu_1 \neq \mu_2.$ **в)** $H_0 : \mu_1 = \mu_2,\ H_1 : \mu_1 > \mu_2.$

Розв’язання

Предметом пошуку є не доказ однієї з гіпотез, а відкидання найменш вірогідної, аби звузити коло пошуку.

а) При побудові довірчого інтервалу для дисперсії скористаємося тим, що величина $\frac{(n-1)s^2}{\sigma^2}$ належить розподілу Пірсона (χ^2) с (n-1, m-1) ступенями свободи.

Довірчий інтервал знаходитимемо з наступного співвідношення:

$$\underline{x}^2 \leq \frac{(n-1)s^2}{\sigma^2} \leq \overline{x}^2$$

$$P\left\{\chi_{n-1}^2 \geq \overline{x}^2\right\} = \frac{\alpha}{2};\ P\left\{\chi_{n-1}^2 \geq \underline{x}^2\right\} = 1 - \frac{\alpha}{2}$$

;

Значення $\underline{x}^2$ та $\overline{x}^2$ знаходимо з таблиць розподілу:

Число степеней свободы, n	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
4	13,2767	11,1433	9,49773	0,710721	0,484419	0,297110
5	15,0863	12,8325	11,0705	1,145476	0,831211	0,554300

$P\left\{\chi_4^2 \geq \overline{x}^2\right\} = 0,025;\ P\left\{\chi_4^2 \geq \underline{x}^2\right\} = 0,975;$

Отже, довірчий інтервал матиме такий вигляд:

$$\frac{(n-1)s^2}{\underline{x}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\overline{x}^2}$$

$$1)(n=5)\ \frac{4\cdot 14,76}{11,14} \leq \sigma^2 \leq \frac{4\cdot 14,76}{0,484} = 5,17 \leq \sigma^2 \leq 121,98;$$

$$2)(m=6)\ \frac{5\cdot 4,92}{12,83} \leq \sigma^2 \leq \frac{5\cdot 4,92}{0,831} = 1,91 \leq \sigma^2 \leq 29,6.$$

Довірчий інтервал має високий рівень розкиду, тому скористаємося критичним значенням статистики Стьюдента для $a = 0,05$ та чисел ступенів свободи та $v_1 = 4,\ v_2 = 5;\ t_{\hat{e}\delta} = 5,19.$

m_1 \ m_2	1	2	3	4
1	161,45	199,50	215,71	224,58
2	18,51	19,00	19,16	19,25
3	10,13	9,55	9,28	9,12
4	7,71	6,94	6,59	6,39
5	6,61	5,79	5,41	5,19

Обчислене значення статистики:

$$t = \frac{s_1^{\ 2}}{s_2^{\ 2}} = \frac{14,76}{4,92} = 3.$$

Так як $t < t_{\hat{e}\delta}$, робимо висновок, що дані вибірки нульвої гіпотези не суперечать, тобто. з надійністю $\gamma = 0,95$ можна стверджувати, що Дисперсії відрізняються незначно.

б) Для побудови довірчого інтервалу використовується величина, що має розподіл Стьюдента з $\nu = n_1 + n_2 - 2$ ступенями свободи, використовуючи наступні формули:

$$s^2 = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}; \qquad t = \frac{\bar{x} - \bar{y}}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

Критичне значення $t_{кр}$ знаходимо за таблицями двостороннього критерію Стьюдента для α = 0.05 та числа ступенів свободи $\nu = n_1 + n_2 - 2 = 9.$
$t_{кр} = 2,26.$

Число степеней свободы, n	Уровень значимости α (двусторонняя критическая область)					
	0,10	0,05	0,02	0,01	0,002	0,001
9	1,8331	2,2622	2,8214	3,2498	4,2968	4,7809

Знайдемо значення s^2:

$$s^2 = \frac{4 \cdot 14,76 + 5 \cdot 5,92}{9} = 9,84;$$

Обчислимо значення статистики t:

$$t = \frac{15,9 - 14,1}{9,84\sqrt{\frac{1}{5} + \frac{1}{6}}} = 0,3.$$

Відповідно до отриманого значення критерія Стюдента, ми можемо зробити висновок, що так, як $t_{кр} > t$, то нульва гіпотеза не відкидається і заслуговує розгляду.

Так як $t < t_{кр}$, робимо висновок, що дані вибірки нульвої гіпотези не суперечать, тобто. з надійністю $\gamma = 0,95$ можна стверджувати, що значення відрізняються незначно.

в) Скористуємося отриманими даними в варіанту “б” та перевіримо гіпотези.

Так як $t < t_{кр}$, робимо висновок, що дані вибірки нульової гіпотези не суперечать, тобто. з надійністю $\gamma = 0,95$ можна стверджувати, що значення відрізняються незначно.

Задача №5

Постановка завдання

Вісім разів при різних значеннях ознаки ς було виміряно значення ознаки η. Отримано такі результати:

x_i	0,30	0,91	1,50	2,00	2,20	2,62	3,00	3,30
y_i	0,20	0,43	0,35	0,52	0,81	0,68	1,15	0,85

1. Довести кореляційну взаємодію між ознаками.
2. Довести, що ознака η є результуючою, а ς - факторною.
3. Оцінити коефіцієнт кореляції між ознаками.
4. Отримати лінійну регресійну моделі $y(x) = a + b \cdot x$;

Розв'язання

Перевірка на значимість обчислених вибіркових коефіцієнтів кореляції є перевіркою наступної гіпотези: чи суттєво відрізняється від нуля розрахований за рядом вимірювань об'єму n емпіричний коефіцієнт кореляції?

Введемо нульову гіпотезу і альтернативну їй:

$H_0 : r_{xy} = 0;$

$H_1 : r_{xy} \neq 0.$

Об'єм вибірки сягає 8($n = 8$). Нехай рівень значимості $a = 0,01$.
Тоді, $v = n - 2 = 6,$ а значення статистики матиме такий вигляд:

$$t = \frac{r_{xy} \cdot v}{\sqrt{1-\left(r_{xy}\right)^2}}.$$

Скористуємося статистичним додатком Excel для обрахування коєфіцієнта кореляції.
*Також характерна формула для знаходження коєфіцієнта кореляції буде мати наступний вигляд:

$$r_{xy} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\,\bar{y}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right)}}$$

За умови, що ми працюємо з двомірною, нормальною генеральною сукупністю.
Перейдмо до вкладки “Дані”, та оберемо функцію “Аналіз даних”:

У списку, що з’явився обираємо “кореляцію”
Вибираємо необхідні інтервали:

x	y
0,3	0,2
0,91	0,43
1,5	0,35
2	0,52
2,2	0,81
2,62	0,68
3	1,15
3,3	0,85

	x	*y*
x	1	
y	0,87916	1

Це означає, що ознака X(η) **–факторна** і впливає на ознаку Y(ς) **- результуючу** на 88%, тобто коефіцієнт кореляції $r_{xy} = 0,88,$ коефіцієнт наближенний більше до 1 ніж до 0, а отже кореляційний момент присутній, нульова гіпотеза відкижається.
Побудуємо регресійну модель:

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множестве	0,879155
R-квадрат	0,772914
Нормиров	0,735067
Стандартн	0,159097
Наблюден	8

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*ачимость F*
Регрессия	1	0,516915	0,516915	20,42175	0,004022
Остаток	6	0,151872	0,025312		
Итого	7	0,668788			

	Коэффициент	*артная ош*	*татисти*	*P-Значение*	*ижние 95%*	*ерхние 95%*	*ижние 95,0*	*рхние 95,0%*
Y-пересеч	0,102139	0,128402	0,795466	0,456649	-0,212048	0,416326	-0,212048	0,416326
Перемен	0,263606	0,058332	4,519043	0,004022	0,120872	0,40634	0,120872	0,40634

ВЫВОД ОСТАТКА

Наблюдение	*дсказанно*	*Остатки*
1	0,181221	0,018779
2	0,342021	0,087979
3	0,497548	-0,147548
4	0,629352	-0,109352
5	0,682073	0,127927
6	0,792788	-0,112788
7	0,892958	0,257042
8	0,97204	-0,12204

Згідно отриманим даним ми встановлюємо лінійну регресійну модель:
Відповідь: $y = 0{,}26x + 0{,}102$

Задача №6

Постановка завдання

Залежність між ς та η задана таблицею:

x_i	–2	–1	0	1	2	3
y_i	–2	–3	–3	–1	3	7

1. **Довести кореляційну взаємодію між ознаками.**
2. **Довести, що ознака η є результуючою, а ς - факторною.**
3. **Оцінити коефіцієнт кореляції між ознаками.**
4. **Користуючись методом найменших квадратів знайти параметри a, b і c квадратичної регресійної моделі $y(x) = a + bx + cx^2$,**

Розв'язання

Об'єм вибірки сягає 6($n = 6$).

Відповідно до методу найменших квадратів, суму квадратів різниць між y_i та $y(x_i)$ становить:

$$F(a,b,c) = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

Система матиме такий вигляд:

$$\frac{\partial F}{\partial a} = \frac{\partial F}{\partial b} = \frac{\partial F}{\partial c} = 0$$

Складемо Систему Лінійних Алгебраїчних Рівнянь:

$$\begin{cases} a\sum_{i=1}^{n} x_i^4 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i^2; \\ a\sum_{i=1}^{n} x_i^3 + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i x_i; \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i + cn = \sum_{i=1}^{n} y_i. \end{cases}$$

$$\sum_{i=1}^{n} x_i^4 = 113;$$

$$\sum_{i=1}^{n} x_i^3 = 27;$$

$$\sum_{i=1}^{n} x_i^2 = 17;$$

$$\sum_{i=1}^{n} x_i = 3$$

$$\sum_{i=1}^{n} y_i x_i^2 = 63;$$

$$\sum_{i=1}^{n} y_i x_i = 33;$$

$$\sum_{i=1}^{n} y_i = -1;$$

$$\begin{cases} 113a + 27b + 17c = 63; \\ a27 + b17 + 3c = 33; \\ a17 + 3b + 6c = -1; \end{cases}$$

Використаємо метод Крамера, записавши визначники, враховуючи значення матриці А і матриці В:

a	113	27	17		b	63
	27	17	3			33
	17	3	6			-1

d	3976		
d1	2404	a	0,604628
d2	5724	b	1,439638
d3	-10336	c	-2,5996

	63	27	17
1	33	17	3
	-1	3	6
	113	63	17
2	27	33	3
	17	-1	6
	113	27	63
3	27	17	33
	17	3	-1

Отже, регресійна модель має наступний вигляд:

$$\tilde{y} = 0,6x^2 + 1,4x + 2,6.$$