

ПРИКЛАД

Застосування основних методів математичної статистики

У 50 жінок, хворих на залізодефіцитну анемію, визначали вміст гемоглобіну в крові. Отримали наступні дані (г/л):

60, 61, 72, 54, 62, 84, 54, 55, 71, 58, 63, 61, 64, 65, 68, 72, 81, 82, 73, 61, 60, 54, 58, 91, 69, 85, 42, 56, 72, 54, 63, 78, 55, 68, 71, 82, 90, 64, 75, 84, 62, 93, 49, 64, 75, 89, 67, 72, 75, 58.

Після лікування залізовмісним препаратом N визначення вмісту гемоглобіну в крові цих хворих дало наступні результати (г/л):

100; 105; 107; 109; 109; 109; 109; 111; 111; 113; 113; 117; 117; 117; 119; 119; 121; 121; 121; 121; 121; 121; 121; 123; 123; 123; 123; 123; 123; 123; 125; 125; 125; 125; 125; 125; 127; 127; 127; 127; 129; 129; 131; 90; 133; 133; 137; 87; 137; 143; 141; 143

Для контролю визначили вміст гемоглобіну в крові 50 здорових жінок. Отримали наступні дані (г/л):

100; 102; 104; 106; 106; 106; 108; 108; 110; 111; 114; 114; 114; 116; 116; 118; 118; 118; 118; 118; 118; 120; 120; 120; 120; 120; 120; 120; 122; 122; 122; 122; 122; 122; 124; 124; 124; 124; 126; 126; 128; 128; 129; 130; 134; 134; 140; 138; 142.

Крім того визначали кількість еритроцитів у крові здорових жінок ($\times 10^{12}/л$):

1,53; 2,83; 2,83; 3,64; 2,55; 2,34; 2,42; 3,68; 3,74; 3,36; 3,48; 3,25; 2,91; 2,59; 2,95; 3,52; 2,89; 2,65; 3,80; 3,46; 3,22; 4,25; 3,73; 4,04; 4,22; 2,87; 3,95; 3,00; 3,02; 3,12; 4,22; 3,22; 3,92; 3,01; 4,00; 3,35; 4,06; 4,46; 4,13; 4,16; 3,44; 3,97; 4,48; 3,85; 4,52; 3,73; 3,97; 3,92; 3,33; 4,30

Визначимо ефективність застосування препарату N для лікування залізодефіцитної анемії у жінок, а також наявність кореляційного зв'язку між вмістом гемоглобіну та кількістю еритроцитів у крові здорових жінок.

1. Дослідимо дані вмісту гемоглобіну в крові у жінок, хворих на залізодефіцитну анемію.

1.1. Побудуємо інтервальний варіаційний ряд. Впорядкуємо дані.

42, 49, 54, 54, 54, 55, 55, 56, 58, 58, 58, 60, 60, 61, 61, 61, 62, 62, 63, 63, 64, 64, 64, 65, 67, 68, 68, 69, 71, 71, 72, 72, 72, 72, 73, 75, 75, 75, 78, 81, 82, 82, 84, 84, 85, 89, 90, 91, 93.

$x_{\min}=42$, $x_{\max}=93$, $n=50$.

Будемо розбивати наш варіаційний ряд на 7 класів ($K=7$). Тоді величина класового інтервалу

$$\lambda = \frac{x_{\max} - x_{\min}}{K} = \frac{93 - 42}{7} = 7,28 ,$$

отже, приймаємо $\lambda=8$.

Будуємо ряд

$[x_i; x_{i+1})$	[38;46)	[46;54)	[54;62)	[62;70)	[70;78)	[78;86)	[86;94)
n_i	1	1	15	12	10	7	4

Перетворимо цей ряд у безінтервальний, замінюючи інтервали на їх серединні значення.

x_i	42	50	58	66	74	82	90
n_i	1	1	15	12	10	7	4

1.2. Побудуємо гістограму, полігон частот, кумулятивну криву.

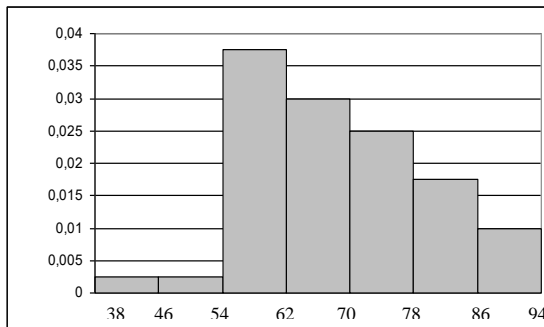
Для побудови гістограми знайдемо частоти, що будуть відповідати площам прямокутників:

$[x_i; x_{i+1})$	[38;46)	[46;54)	[54;62)	[62;70)	[70;78)	[78;86)	[86;94)
n_i	1	1	15	12	10	7	4
$\frac{n_i}{n}$	0,02	0,02	0,3	0,24	0,2	0,14	0,08

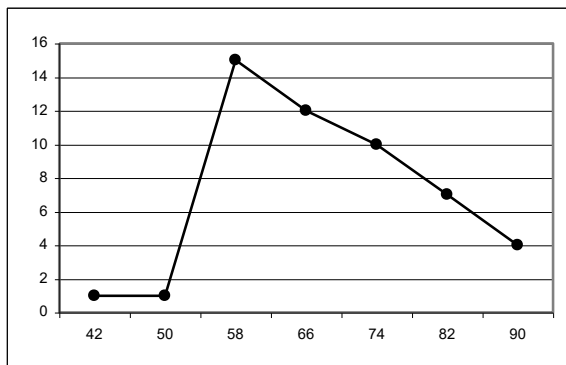
Оскільки довжина інтервалів $\lambda=8$, висоти прямокутників дорівнюватимуть $\frac{1}{8} \cdot \frac{n_i}{n}$, тобто

$[x_i; x_{i+1})$	[38;46)	[46;54)	[54;62)	[62;70)	[70;78)	[78;86)	[86;94)
n_i	1	1	15	12	10	7	4
$\frac{n_i}{n}$	0,02	0,02	0,3	0,24	0,2	0,14	0,08
$\frac{1}{8} \cdot \frac{n_i}{n}$	0,0025	0,0025	0,0375	0,03	0,025	0,0175	0,01

За даними таблиці будуємо гістограму



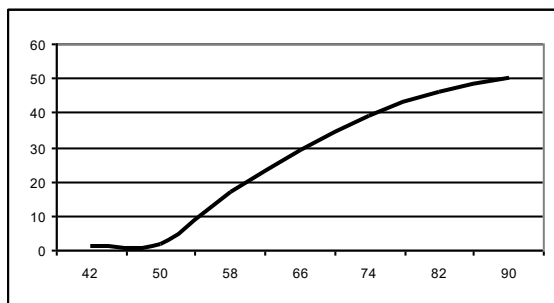
Будуємо полігон частот, відкладаючи на вісі Ох середини інтервалів, а на вісі Оу - частоти:



Для побудови кумулятивної кривої знайдемо накопичені частоти

Інтервал	[38;46)	[46;54)	[54;62)	[62;70)	[70;78)	[78;86)	[86;94]
Частоти	1	1	15	12	10	7	4
Кумуляти частот	1	2	17	29	39	46	50

Будуємо кумулятивну криву, відкладаючи на вісі Ох середини інтервалів, а на вісі Оу – накопичені частоти



1.3. Обчислимо основні статистичні показники вибірки:

Знайдемо середнє значення вибірки $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{3396}{50} = 67,92$

Похибка середнього арифметичного $s_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{11,71}{\sqrt{50}} \approx 1,66$

Знайдемо дисперсію $D = \overline{x^2} - (\bar{x})^2 = 4750,2 - 67,92^2 \approx 137,07$

Похибка дисперсії $s_{\sigma^2} = \frac{\sigma^2}{\sqrt{2n}} = \frac{137,07}{\sqrt{100}} \approx 13,7$

Середнє квадратичне відхилення $\sigma = \sqrt{D} = \sqrt{137,07} \approx 11,71$

Похибка середнього квадратичного відхилення

$$s_{\sigma} = \frac{\sigma}{\sqrt{2n}} = \frac{11,71}{\sqrt{100}} = 1,171$$

Знайдемо медіану. Оскільки вибірка складається з парної кількості елементів (50), медіаною буде середнє значення 25-го та 26-го

елементів впорядкованої за зростанням вибірки. $Me = \frac{65 + 67}{2} = 66$

Похибка медіани $s_{Me} = 1,2533 \frac{\sigma}{\sqrt{n}} = 1,2533 \frac{11,71}{\sqrt{50}} \approx 2,076$

Поділивши вибірку навпіл, з першої половини

42, 49, 54, 54, 54, 54, 55, 55, 56, 58, 58, 58, 60, 60, 61, 61, 61, 62, 62, 63, 63, 64, 64, 64, 65

знаходимо нижній кuartиль (медіану цієї половини). Оскільки дана вибірка має непарну кількість елементів (25), нижнім кuartилем буде значення 13-го елемента, тобто 60.

Верхній кuartиль – це значення 13-го елемента вибірки

67, 68, 68, 69, 71, 71, 72, 72, 72, 72, 73, 75, 75, 75, 78, 81, 82, 82, 84, 84, 85, 89, 90, 91, 93,

або, що те саме, значення 38-го елемента вихідної вибірки. Отже, верхній кuartиль дорівнює 75.

Різниця між медіаною і нижнім кuartилем (6) менша за різницю між верхнім кuartилем і медіаною (9), що вказує на асиметричність ряду.

Знайдемо коефіцієнт асиметрії $As = \frac{\overline{(x - \bar{x})^3}}{\sigma^3} = \frac{502,26}{11,71^3} \approx 0,31$

$As > 0$, отже ряд асиметричний, але, оскільки $As < 1$, асиметричність незначна.

Коефіцієнт варіації $C = \frac{\sigma}{M} \cdot 100\% = \frac{11,71}{67,92} \cdot 100\% \approx 17\%$

Коефіцієнт ексцесу $\varepsilon = \frac{\overline{(x - \bar{x})^4}}{\sigma^4} - 3 = \frac{46297,66}{11,71^4} - 3 \approx -0,54$

Оскільки $\varepsilon < 0$, крива розподілу має плоску вершину.

Ліміти вибірки: $x_{\min} = 42$, $x_{\max} = 93$

Розмах вибірки $R = x_{\max} - x_{\min} = 93 - 42 = 51$

1.4. Перевіримо розподіл вибірки на нормальність.

I. Перевіряємо за правилом “трьох сигм”.

$$\sigma=11,71; 0,625\sigma=7,32; 3\sigma=35,13$$

Перевіряємо, скільки спостережень відрізняються від середнього менше, ніж на 3σ

$$\bar{x} + 3\sigma = 103,05, \quad \bar{x} - 3\sigma = 32,79$$

В проміжок (32,79;103,05) попадають всі спостереження.

Перевіряємо, скільки спостережень відрізняються від середнього менше, ніж на σ

$$\bar{x} + \sigma = 79,63, \quad \bar{x} - \sigma = 56,21$$

У проміжок (56,21;79,63) попадають 32 з 50 (64%) спостережень.

Перевіряємо, скільки спостережень відрізняються від середнього менше, ніж на $0,625\sigma$

$$\bar{x} + 0,625\sigma = 75,24, \quad \bar{x} - 0,625\sigma = 60,60$$

У проміжок (60,60;75,24) попадають 25 з 50 спостережень, тобто не більше, ніж на $0,625\sigma$ від середнього відрізняються 50% спостережень.

Таким чином, за правилом “трьох сигм” дану вибірку не можна віднести до нормально розподілених.

II. За основними числовими характеристиками закон розподілу даної вибірки також не є нормальним. Дійсно,

$$Me = 66 \neq \bar{x}$$

$$As = 0,31 > 0$$

$$\varepsilon = -0,54 < 0,$$

а як відомо, медіана і середнє при нормальному розподілі співпадають, а коефіцієнт асиметрії та ексцес прямують до нуля.

Отже, не можна стверджувати, що закон розподілу нашої вибірки нормальний

1.5. Перевіряємо вибірку на однорідність.

Оскільки $n > 25$, перевіряємо за правилом трьох стандартних відхилень.

$$3\sigma=35,13$$

Для крайнього лівого числа 42 маємо

$$d_1 = |x_1 - \bar{x}| = |42 - 67,92| = 25,92 < 35,13$$

Для крайнього правого числа 93:

$$d_1 = |x_n - \bar{x}| = |93 - 67,92| = 25,08 < 35,13$$

Отже, і крайнє ліве і крайнє праве числа з вибірки не є промахами, тобто вибірка однорідна з рівнем значущості $\alpha < 0,01$.

2. Перевіримо, чи є нормальним закон розподілу вмісту гемоглобіну в крові здорових жінок

I. Перевіряємо за правилом “трьох сигм”.

$$\bar{x} = 120 ; \sigma = 9,5; 0,625\sigma = 5,9; 3\sigma = 28,4$$

Перевіряємо, скільки спостережень відрізняються від середнього менше, ніж на 3σ

$$\bar{x} + 3\sigma = 148,4 , \bar{x} - 3\sigma = 91,6$$

У проміжок (91,6;148,4) попадають всі спостереження.

Перевіряємо, скільки спостережень відрізняються від середнього менше, ніж на σ

$$\bar{x} + \sigma = 129,5 , \bar{x} - \sigma = 110,5$$

У проміжок (110,5;129,5) попадають 34 з 50 (68%) спостережень.

Перевіряємо, скільки спостережень відрізняються від середнього менше, ніж на $0,625\sigma$

$$\bar{x} + 0,625\sigma = 125,9 , \bar{x} - 0,625\sigma = 114,1$$

В проміжок (114,1;125,9) попадають 25 із 50 спостережень, тобто не більше ніж на $0,625\sigma$ від середнього відрізняються 50% спостережень.

Таким чином, за правилом “трьох сигм” дану вибірку можна віднести до нормально розподілених.

II. Перевіримо основні числові характеристики даної вибірки

$$Me = 120 = \bar{x} , Mo = 120 = \bar{x} , As = 0,094 \rightarrow 0 , \varepsilon = 0,099 \rightarrow 0 ,$$

Отже, і основні числові характеристики вибірки відповідають нормальному закону розподілу.

III. Скористаємося критерієм Пірсона на рівні значущості $\alpha=0,05$.

X	[100;106)	[106;112)	[112;118)	[118;124)	[124;130)	[130;136)	[136;142]
n_x	3	7	5	19	8	5	3

1) Формулюємо гіпотези:

H_0 : вміст гемоглобіну в крові підпорядковується нормальному закону розподілу $N(\mu; \sigma)$;

H_1 : вміст гемоглобіну в крові не підпорядковується нормальному закону розподілу.

Рівень значущості $\alpha = 0,05$.

2) обчислюємо теоретичні частоти.

Визначаємо з табл. 6 додатку Г значення функції нормального розподілу на межах класових інтервалів:

x	$-\infty$	106	112	118	124	130	136	∞
$\left(\frac{x - \bar{x}}{\sigma}\right)$	$-\infty$	-1,47	-0,84	-0,21	0,42	1,05	1,68	∞
$\Phi\left(\frac{x - \bar{x}}{\sigma}\right)$	-	0,5	-0,4292	-0,2995	-0,0832	0,1628	0,3531	0,4535

Обчислюємо ймовірності та теоретичні частоти попадання ознаки в класи за формулою

$$p_{xi}' = \Phi\left(\frac{x_i - \bar{x}}{\sigma}\right) - \Phi\left(\frac{x_{i-1} - \bar{x}}{\sigma}\right).$$

Теоретична частота попадання ознаки в i -ий клас $= np_i$. Результати розрахунків зведені в таблиці

X	[100;106)	[106;112)	[112;118)	[118;124)	[124;130)	[130;136)	[136;142]
n_x	3	7	5	19	8	5	3
X'	$(-\infty; 1,47)$	$[-1,47; 0,84)$	$[-0,84; -0,21)$	$[-0,21; 0,42)$	$[0,42; 1,05)$	$[1,05; 1,68)$	$[1,68; \infty)$
p_x'	0,071	0,130	0,216	0,246	0,190	0,100	0,047
np_x'	3,54	6,485	10,815	12,3	9,515	5,02	2,325
$(n_i - np_i')^2$	0,292	0,265	33,814	44,890	2,295	0,0004	0,456
$\frac{(n_i - np_i')^2}{np_i'}$	0,082	0,041	3,127	3,650	0,241	0,0001	0,196

Розраховуємо значення χ^2 – критерію

$$\chi^2 = \sum_i (n_i - n_i')^2 / n_i' = 7,34$$

Критичне значення визначаємо з таблиці χ^2 – критерію при $k = 7 - 1 - 2 = 4$; $p = 1 - 0,05 = 0,95$;

$$\chi_{кр}^2(0,95; 4) = 9,488$$

Оскільки $\chi^2 < \chi_{кр}^2$, то немає підстав заперечувати нормальний розподіл вмісту гемоглобіну в крові здорових жінок.

3. Для двох незалежних вибірок – вміст гемоглобіну в крові здорових жінок та вміст гемоглобіну в крові жінок, хворих на залізодефіцитну анемію, після лікування при рівні значущості 0,01 перевіряємо нульову гіпотезу $H_0: \bar{X} = \bar{Y}$, при конкуруючій гіпотезі $H_1: \bar{X} \neq \bar{Y}$.

Знайдемо значення критерію, що спостерігається:

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{D(X)/n + D(Y)/m}} = \frac{120 - 121,08}{\sqrt{92/50 + 132,5/50}} = -0,51$$

За умовою, конкуруюча гіпотеза має вигляд $\bar{X} \neq \bar{Y}$, тому критична область – двостороння. За таблицею функції Лапласа можна знайти критичну точку згідно рівності $\Phi_{Z_{кр}} = (1 - \alpha)/2$.

Знайдемо праву критичну точку:

$$\Phi(z_{кр}) = (1 - \alpha)/2 = (1 - 0,01)/2 = 0,495.$$

За таблицею функції Лапласа знаходимо $z_{кр} = 2,58$.

Оскільки $|Z| < z_{кр}$ – нульову гіпотезу приймаємо. Іншими словами, вибіркові середні відрізняються незначимо.

4. Для двох незалежних вибірок – вміст гемоглобіну в крові здорових жінок та вміст гемоглобіну в крові жінок, хворих на залізодефіцитну анемію, після лікування при рівні значущості 0,05 перевіримо нульову гіпотезу $H_0 : D(X) = D(Y)$ про рівність генеральних дисперсій при конкуруючій гіпотезі $H_1: D(X) > D(Y)$.

Нам відомо, що $n = 50, m=50$, виправлені вибіркові дисперсії: $s_x=92, s_y=97$.

Знайдемо відношення більшої виправленої дисперсії до меншої: $F_{набл} = 97/92 = 1,05$

Конкуруюча гіпотеза має вигляд $D(X) > D(Y)$, тому критична область – правостороння.

За таблицею при рівні значущості $\alpha = 0,05$ і числам ступенів вільності $k_1= 50 - 1 = 49$ та $k_2 = 50- 1 = 49$ знаходимо критичну точку $F_{кр}(0,05; 49, 49) \approx 1,6$

Оскільки $F_{набл} < F_{кр}$ немає підстав відкинути нульову гіпотезу про рівність генеральних дисперсій.

5. Визначимо ефективність застосування препарату N для лікування залізодефіцитної анемії у жінок.

5.1. Перевіримо, чи суттєво змінилося значення вмісту гемоглобіну в крові у жінок після лікування. Вибірки до лікування та після лікування залежні, їх розподіли відрізняються від нормального, тому скористаємося T-критерієм Уїлкоксона.

Знайдемо різниці між значеннями до лікування та після лікування для кожного хворого, а також ранги модулів цих різниць.

До лікування	Після лікування	Різниця	Модулі різниць	Ранги модулів додатних різниць	Ранги модулів від'ємних різниць
60	100	40	40	7,5	
61	105	44	44	10	
72	107	35	35	3	
54	109	55	55	24,5	
62	109	47	47	13,5	
84	109	25	25	1	
54	111	57	57	29,5	
55	111	56	56	27	
71	113	42	42	9	
58	113	55	55	24,5	
63	117	54	54	21,5	
61	117	56	56	27	
64	117	53	53	18,5	
65	119	54	54	21,5	
68	119	51	51	17	

72	121	49	49	16	
81	121	40	40	7,5	
82	121	39	39	6	
73	121	48	48	15	
61	121	60	60	31	
60	121	61	61	32	
54	123	69	69	40	
58	123	65	65	36	
91	123	32	32	2	
69	123	54	54	21,5	
85	123	38	38	5	
42	123	81	81	46	
56	123	67	67	38	
72	125	53	53	18,5	
54	125	71	71	44,5	
63	125	62	62	33,5	
78	125	47	47	13,5	
55	125	70	70	42,5	
68	125	57	57	29,5	
71	127	56	56	27	
82	127	45	45	11,5	
90	127	37	37	4	
64	127	63	63	35	
75	129	54	54	21,5	
84	129	45	45	11,5	
62	131	69	69	40	
93	90	-3	3		2
49	133	84	84	48	
64	133	69	69	40	
75	137	62	62	33,5	
89	87	-2	2		1
67	137	70	70	42,5	
72	143	71	71	44,5	
75	141	66	66	37	
58	143	85	85	47	

Сума рангів додатних різниць $R_1=1176$, сума рангів від'ємних різниць $R_2=3$. Для рівня значущості 0,05 критичне значення $T_{0,05}=434,4$, для рівня значущості 0,01 критичне значення $T_{0,01}=370,2$.

Отже, з похибкою не більше ніж 1% ($\alpha < 0,01$) можна стверджувати, що після лікування препаратом зросла кількість гемоглобіну в крові у жінок, хворих на залізодефіцитну анемію.

5.2. Перевіримо, чи наблизилось значення вмісту гемоглобіну в крові у хворих жінок після лікування до такого у здорових жінок. Вибірка для хворих жінок розподілена ненормально, тому застосуємо непараметричний U-критерій Уїлкоксона. Об'єднаємо обидві вибірки в один ряд та знайдемо ранги елементів цього ряду.

	<i>Об'єднаний ряд</i>	<i>Ранги</i>		<i>Об'єднаний ряд</i>	<i>Ранги</i>
<i>Хворі</i>	100	3,5	<i>Здорові</i>	100	3,5
	105	7		102	5
	107	11		104	6
	109	15		106	9
	109	15		106	9
	109	15		106	9
	111	19,5		108	12,5
	111	19,5		108	12,5
	113	21,5		110	17,5
	113	21,5		110	17,5
	117	29		114	24
	117	29		114	24
	117	29		114	24
	119	37,5		116	26,5
	119	37,5		116	26,5
	121	48,5		118	33,5
	121	48,5		118	33,5
	121	48,5		118	33,5
	121	48,5		118	33,5
	121	48,5		118	33,5
	123	61		120	42
	123	61		120	42
	123	61		120	42
	123	61		120	42
	123	61		120	42
	123	61		120	42
	123	61		120	42
	123	61		120	42
	125	71,5		122	54,5
	125	71,5		122	54,5
	125	71,5		122	54,5
	125	71,5		122	54,5
	125	71,5		122	54,5
	125	71,5		122	54,5
127	78,5	124	66,5		
127	78,5	124	66,5		
127	78,5	124	66,5		
127	78,5	124	66,5		
129	83,5	126	75,5		
129	83,5	126	75,5		
131	87	128	81,5		
90	2	128	81,5		
133	88,5	130	85,5		
133	88,5	130	85,5		
137	93,5	134	91		
87	1	134	91		
137	93,5	134	91		
143	99,5	140	96		
141	97	138	95		
143	99,5	142	98		

Знайдемо суму рангів для елементів першого ряду $R_1=2688,5$ та суму рангів елементів другого ряду $R_2=2361,5$.

Знайдемо величини $U_1 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 1413,5$ та

$U_2 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 1086,5$, де $n_1=n_2=50$ – об'єми вибірок.

Найменшу з цих величин $U_2=1086,5$ необхідно порівняти з критичним значенням. $U_{0,05} = 1010$, $U_{0,01} = 912$. Найменша з величин U_1 та U_2 більша за обидва критичні значення, тому приймаємо нульову гіпотезу. Тобто, лікування жінок було ефективним.

6. Встановимо наявність кореляційного зв'язку між вмістом гемоглобіну та кількістю еритроцитів у крові здорових жінок.

6.1. Знайдемо коефіцієнт кореляції.

Вибірки незалежні, розподілені нормально.

Знайдемо \bar{x} та \bar{y}

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 120; \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = 3,48.$$

Складемо таблицю для подальших обчислень:

x_i Вміст гемоглобіну в крові здорових жінок (г/л)	y_i Кількість еритроцитів в крові у здорових жінок ($\times 10^{12}$)	$x_i y_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
100	1,53	153	-20	400	-1,95	3,79
102	2,83	288,66	-18	324	-0,65	0,42
104	2,83	294,32	-16	256	-0,65	0,42
106	3,64	385,84	-14	196	0,16	0,03
106	2,55	270,3	-14	196	-0,93	0,86
106	2,34	248,04	-14	196	-1,14	1,29
108	2,42	261,36	-12	144	-1,06	1,12
108	3,68	397,44	-12	144	0,20	0,04
110	3,74	411,4	-10	100	0,26	0,07
110	3,36	369,6	-10	100	-0,12	0,01
114	3,48	396,72	-6	36	0,00	0,00
114	3,25	370,5	-6	36	-0,23	0,05
114	2,91	331,74	-6	36	-0,57	0,32
116	2,59	300,44	-4	16	-0,89	0,79

116	2,95	342,2	-4	16	-0,53	0,28
118	3,52	415,36	-2	4	0,04	0,00
118	2,89	341,02	-2	4	-0,59	0,35
118	2,65	312,7	-2	4	-0,83	0,68
118	3,8	448,4	-2	4	0,32	0,10
118	3,46	408,28	-2	4	-0,02	0,00
118	3,22	379,96	-2	4	-0,26	0,07
120	4,25	510	0	0	0,77	0,60
120	3,73	447,6	0	0	0,25	0,06
120	4,04	484,8	0	0	0,56	0,32
120	4,22	506,4	0	0	0,74	0,55
120	2,87	344,4	0	0	-0,61	0,37
120	3,95	474	0	0	0,47	0,22
120	3	360	0	0	-0,48	0,23
122	3,02	368,44	2	4	-0,46	0,21
122	3,12	380,64	2	4	-0,36	0,13
122	4,22	514,84	2	4	0,74	0,55
122	3,22	392,84	2	4	-0,26	0,07
122	3,92	478,24	2	4	0,44	0,20
122	3,01	367,22	2	4	-0,47	0,22
124	4	496	4	16	0,52	0,27
124	3,35	415,4	4	16	-0,13	0,02
124	4,06	503,44	4	16	0,58	0,34
124	4,46	553,04	4	16	0,98	0,97
126	4,13	520,38	6	36	0,65	0,43
126	4,16	524,16	6	36	0,68	0,47
128	3,44	440,32	8	64	-0,04	0,00
128	3,97	508,16	8	64	0,49	0,24
130	4,48	582,4	10	100	1,00	1,00
130	3,85	500,5	10	100	0,37	0,14
134	4,52	605,68	14	196	1,04	1,09
134	3,73	499,82	14	196	0,25	0,06
134	3,97	531,98	14	196	0,49	0,24
140	3,92	548,8	20	400	0,44	0,20
138	3,33	459,54	18	324	-0,15	0,02
142	4,3	610,6	22	484	0,82	0,68

За даними таблиці знаходимо:

$$\overline{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i = 421,14$$

Знаходимо виправлені дисперсії для x та y :

$$\sigma_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{4504}{49} \approx 91,92$$

$$\sigma_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{20,57}{49} \approx 0,42$$

Знайдемо коефіцієнт кореляції:

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x(X)\sigma_y(Y)} = \frac{421,14 - 120 \cdot 3,48}{\sqrt{91,92} \cdot \sqrt{0,42}} \approx 0,57$$

Отриманий результат свідчить про наявність середнього кореляційного зв'язку між вмістом гемоглобіну та кількістю еритроцитів у крові у здорових жінок.

6.2. Знайдемо рівень значущості коефіцієнта кореляції.

Обчислимо експериментальне значення критерію t :

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,57\sqrt{50-2}}{\sqrt{1-0,57^2}} = \frac{3,95}{0,82} \approx 4,8$$

Враховуючи, що число ступенів вільності $k = n - 2 = 50 - 2 = 48$, шукаємо за таблицею значення $t_{кр}(0,01;48)=2,58$. Отже, коефіцієнт кореляції є значущим з високим рівнем значущості ($p < 0,01$).

6.3. Знайдемо рівняння регресії Y на X . Побудуємо пряму регресії y на x .

$$\rho_{yx} = \frac{\overline{xy} - \bar{x}\bar{y}}{(\sigma(X))^2} = \frac{421,14 - 120 \cdot 3,48}{91,92} = \frac{3,54}{91,92} \approx 0,04$$

Рівняння шуканої прямої має вигляд:

$$y - 3,48 = 0,04(x - 120),$$

або

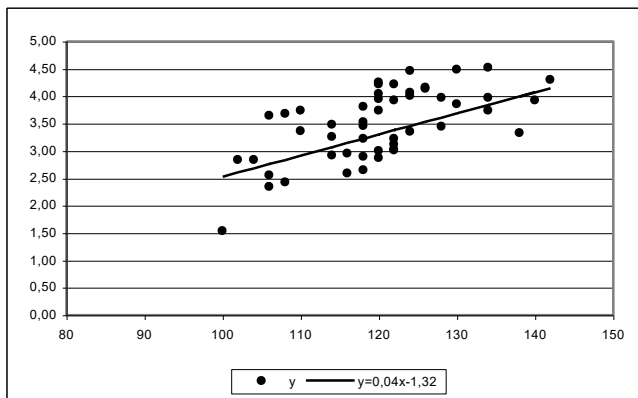
$$y = 0,04x - 1,32.$$

Обчислимо значення y за знайденими формулами, результати занесемо в таблицю:

x_i	y_i	$y = 0,04x - 1,32$
100	1,53	2,53
102	2,83	2,61
104	2,83	2,69
106	3,64	2,76
106	2,55	2,76
106	2,34	2,76
108	2,42	2,84
108	3,68	2,84
110	3,74	2,92
110	3,36	2,92
114	3,48	3,07
114	3,25	3,07
114	2,91	3,07
116	2,59	3,15
116	2,95	3,15
118	3,52	3,22
118	2,89	3,22
118	2,65	3,22
118	3,80	3,22
118	3,46	3,22
118	3,22	3,22
120	4,25	3,30
120	3,73	3,30
120	4,04	3,30
120	4,22	3,30

x_i	y_i	$y = 0,04x - 1,32$
120	2,87	3,30
120	3,95	3,30
120	3,00	3,30
122	3,02	3,38
122	3,12	3,38
122	4,22	3,38
122	3,22	3,38
122	3,92	3,38
122	3,01	3,38
124	4,00	3,46
124	3,35	3,46
124	4,06	3,46
124	4,46	3,46
126	4,13	3,53
126	4,16	3,53
128	3,44	3,61
128	3,97	3,61
130	4,48	3,69
130	3,85	3,69
134	4,52	3,84
134	3,73	3,84
134	3,97	3,84
140	3,92	4,07
138	3,33	3,99
142	4,30	4,15

За даними таблиці будуємо графік:



Висновок. Після прийому препарату N вміст гемоглобіну в крові у жінок, хворих на залізодефіцитну анемію, достовірно збільшився і досяг рівня гемоглобіну в крові здорових жінок. Отже, лікування було ефективним.

Між вмістом гемоглобіну та кількістю еритроцитів у крові здорових жінок існує значущий середній кореляційний зв'язок.