DOI: 10.32347/2076-815X.2023.83.38-50
УДК 528.482.5
канд. техн. наук, доцент **Гладілін В.М.**,
vgladilin.55@gmail.com, ORCID: 0000-0002-0492-3510,
канд. екон. наук, доцент **Сіроштан Т.М.**,
tanya3031@i.ua, ORCID: 0000-0001-6791-7081,
канд. геогр. наук, доцент **Гамалій І.П.**,
gurgev@gmail.com, ORCID: 0000-0002-3469-4798,
Свідерська Т.О., tsv245@gmail.com, ORCID: 0000-0001-7623-6958,
Білоцерківський національний аграрний університет,
Шудра Н.С., shudranatasha1984@gmail.com, ORCID: 0000-0001-5416-7680,
Чуланов П.О., chulanov.po@knuba.edu.ua, ORCID: 0000-0002-6735-3770,
Київський національний університет будівництва і архітектури

ВИРАХУВАННЯ КОЕФІЦІЄНТА КОРЕЛЯЦІЇ ПРИ ГЕОДЕЗИЧНИХ ВИМІРЮВАННЯХ

Пропонується формула для визначення коефіцієнтів кореляції, за якою при малій кількості вимірювань значення коефіцієнтів будуть більшими в порівнянні з величинами обчисленими за відомими формулами, тобто ці значення будуть менше зміщеними по відношенню до істинних величин коефіцієнтів кореляції.

В геодезії взагалі обмежуються невеликою кількістю вимірювань, в фотограмметрії при взаємному орієнтуванні знімків стереопари в автоматичному режимі (на станції Дельта) при ідентифікації точок на лівому і правому знімках необхідно точно обчислювати значення коефіцієнтів кореляції, що дає можливість побудувати кращу фотограмметричну модель зображення місцевості.

Ключові слова: коефіцієнт кореляції; коефіцієнти лінійної і нелінійної регресії в геодезичних вимірюваннях.

Вступ. Відомо [6,7], що при невеликій кількості вимірювань у випадку залежних рядів вимірювань коефіцієнт кореляції, обчислений за відомою формулою

$$r_0 = \left[\sum_{i=1}^{n}\delta(x_i)\delta(y_i)\right]/[(n-1)\sigma(x)\sigma(y)], \qquad (1)$$

виявляться заниженими. В цій формулі відхилення $\delta(x_i)$ і $\delta(y_i)$ знаходяться з виразів

$$\delta(x_i) = x_i - x_c;\ \ \delta(y_i) = y_i - y_c\ , \qquad (2)$$

де: x_i , y_i – вимірювання в рядах X та Y; $x_{ср}$, $y_{ср}$ – середні арифметичні значення за цими рядами вимірювань.

Емпіричні стандарти або середні квадратичні помилки результатів вимірювань за рядами *X* та *Y* визначаються за відомими формулами:

$$\sigma(x) = \sqrt{\sum_{i=1}^{n} \delta(x_i)/(n-1)};$$
$$\sigma(y) = \sqrt{\sum_{i=1}^{n} \delta(y_i)/(n-1)}, \qquad (3)$$

де n – кількість вимірювань в ряду.

Є формула, за якою можна знайти більш точні значення коефіцієнта кореляції:

$$r' = r_0 \ \{1 + (1 - r_0^2)/[2 * (n-3)]\}, \qquad (4)$$

але і ця формула, як виявляється при кількості вимірювань $n > 5$ дає декілька занижене значення коефіцієнта кореляції.

Спочатку зупинимось на рівняннях прямих регресії для двох залежних рядів вимірювань *X* і *Y*, такими будуть:

$$y = b_1 + k_1 * x; \quad x = b_2 + k_2 * y. \qquad (5)$$

В цих рівняннях величини b_1 і b_2 представимо виразами

$$b_1 = y_c - k_1 * x_c,$$
$$b_2 = x_c - k_2 * y_c. \qquad (6)$$

Коефіцієнти регресії або кутові коефіцієнти k_1 і k_2 знайдемо за формулами

$$k_1 = \sum_{i=1}^{n} \delta(x_i) * \delta(y_i) / \sum_{i=1}^{n} \delta^2(x_i);$$
$$k_2 = \sum_{i=1}^{n} \delta(x_i) * \delta(y_i) / \sum_{i=1}^{n} \delta^2(y_i). \qquad (7)$$

Формулу (1) для визначення коефіцієнта кореляції можливо представити з урахуванням (7) у такому вигляді:

$$|r_0| = \sqrt{k_1 * k_2}. \qquad (8)$$

Якщо рівняння регресії (5) співпадають, тобто кут між цими прямими $\varphi=0$, то коефіцієнт кореляції буде дорівнювати 1, $r_0 = 1$. Із збільшенням кута кореляційний зв'язок зменшується і коли $\varphi = \pi/2$, цей зв'язок дорівнює нулю, тобто $r_0 = 0$. Таким чином коефіцієнт кореляції буде залежати від кута між прямими рівняннями регресії (5), таку залежність пропонується представити у вигляді

$$r = 1 - 2 * \varphi/\pi, \qquad (9)$$

де φ – чисельне значення величини кута в радіанах.

Для знаходження кута φ скористуємось формулою кута між двома прямими, для цього друге рівняння (5) знайдемо відносно y, тоді ця система набуде вигляду

$$y = b_1 + k_1 * x$$
$$y = -b_2/k_2 + (1/k_2) * x. \qquad (10)$$

Тоді кут між двома прямими визначиться як

$$\tan\varphi = (1/k_2 - k_1)/(1 + k_1/k_2) = (1 - k_1 * k_2)/(k_1 + k_2). \qquad (11)$$

Враховуючи (11) і (9) формула коефіцієнта кореляції буде мати вигляд

$$r = 1 - 2/\pi * \tan^{-1}[(1 - k_1 * k_2)/(k_1 + k_2)]. \qquad (12)$$

За одержаною формулою обчислюється коефіцієнт кореляції з додатнім кореляційним зв'язком. У виразі (7) коефіцієнти k_1 і k_2 можуть бути тільки з однаковими знаками, так як в їхніх виразах один і той же чисельник, а знаменники завжди додатні величини. Добуток коефіцієнтів k_1 і k_2 у відповідності з (8), і тим, що $|r_0| \leq 1$ буде

$$0 \leq k_1 * k_2 \leq 1. \quad (13)$$

Тому чисельник в (11) буде величиною додатною, тобто

$$1 - k_1 * k_2 \geq 0. \quad (14)$$

Відповідно знак дробу в (11) буде залежати від знаку знаменника, який в свою чергу залежить від знаку коефіцієнтів k_1 і k_2.

У випадку від'ємного кореляційного зв'язку, коли збільшення одної величини відповідає зменшенню другої, знаки коефіцієнтів k_1 і k_2 будуть від'ємними, тоді кут φ буде від'ємним, а формула коефіцієнту кореляції набуде вигляду

$$r = -1 - 2/\pi * \tan^{-1}[(1 - k_1 * k_2)/(k_1 + k_2)]. \quad (15)$$

Розглянемо приклади обчислення коефіцієнтів кореляції, спочатку з додатнім зв'язком. На пункті тріангуляції виміряні направлення дванадцятьма прийомами по способу кругових прийомів, відхилення від середніх значень направлень представлені в табл. 1.

Таблиця 1.

Відхилення виміряних напрямків від середніх значень

Прийоми	$\delta(x_i)$	$\delta(y_i)$	Прийоми	$\delta(x_i)$	$\delta(y_i)$
1	-2,18	+0,57	7	-0,48	-1,13
2	-2,38	-2,03	8	-0,98	-1,03
3	+0,62	+1,17	9	+2,32	+0,77
4	-0,38	+0,03	10	+0,42	-0,23
5	+0,82	+2,07	11	+2,32	+1,97
6	+0,02	-1,63	12	-0,12	-0,53

За даними табл. 1

$$\sum \delta(x_i) * \delta(y_i) = 13{,}807; \ \sum \delta^2(x_i) = 23{,}757;$$
$$\sum \delta^2(y_i) = 19{,}907; \ k_1 = 0{,}5812; \ k_2 = 0{,}6936.$$

Коефіцієнт кореляції обчислений за формулою (1) або (8) $r_0 = 0{,}635$, за формулою (4) $r' = 0{,}656$, за формулою (12) $r = 0{,}721$.

З цього прикладу видно, що коефіцієнт кореляції, обчислений за формулою (4) більше на величину 0.021, а за формулою (12) більший на величину 0.086 в порівнянні з його значенням обчисленим за формулами (1) або (8).

Розглянемо від'ємний кореляційний зв'язок. Для двох рядів вимірювань при n = 5, одержані такі значення $\sum \delta(x_i) * \delta(y_i) = -0{,}41;\ \sum \delta^2(x_i) = 0{,}72;$ $\sum \delta^2(y_i) = 0{,}84;\ \ k_1 = -0{,}5694;\ k_2 = -0{,}4881$, коефіцієнт кореляції в даному випадку обчислимо за формулою

$$r_0 = \sum_{i=1}^{n} \delta(x_i) * \delta(y_i) / \sqrt{\sum_{i=1}^{n} \delta^2(x_i) * \sum_{i=1}^{n} \delta^2(y_i)} = -0{,}527. \quad (16)$$

За формулою (4) коефіцієнт $r' = -0{,}622$, а за формулою (15) r = -0,619, тоді при кількості вимірювань n = 5 коефіцієнти кореляції, які обчислені за формулами (4) і (12) практично однакові, але необхідно зауважити, що при дослідженні кореляційної залежності кількість вимірів завжди більше п'яти, тому пропонується обчислювати коефіцієнт кореляції за формулою (12), значення якого буде правдоподібним в порівнянні з величинами, які знайдені за формулами (1), (8), (16) і за формулою (4) при кількості вимірювань більше п'яти.

Дослідимо запропоновану формулу обчислення коефіцієнта кореляції. Спочатку покажемо, що кількість вимірювань n в рядах при якому коефіцієнти кореляції, які обчислені за формулами (4) і (12), приблизно однакове і коливається біля п'яти. Для цього формулу (12) представимо в іншому вигляді. Із формули (8) видно, що коефіцієнт кореляції дорівнює середньому геометричному значенню коефіцієнтів k_1 і k_2. Якщо $(k_1 + k_2)$:2 буде середнім арифметичним, $k_1 = k_2 = r_0$, що витікає із (8), то

$$k_1 + k_2 = 2 * r_0. \quad (17)$$

Враховуючи (17) вираз (12) набуде вигляду

$$r = 1 - 2/\pi * \tan^{-1}[(1 - r_0^2)/(2 * r_0)]. \quad (18)$$

Для наведеного раніш прикладу при додатному кореляційному зв'язку r_0=0,635, підставимо це значення в (4), одержимо r = 0.720. За формулою (12) r=0,721, видно, що ці одержані значення практично рівні, а різниця між коефіцієнтами $k_1 - k_2$ = 0,69 – 0,58 = 0,11 має доволі значну величину.

Для знаходження кількості вимірювань n при якому коефіцієнти кореляції, які обчислені за формулами (4) і (18) будуть рівні, для цього прирівняємо праві частини рівнянь (4) і (18), одержимо:

$$r_0 * \left[1 + \frac{1 - r_0^2}{2*(n-3)}\right] = 1 - \frac{2}{\pi} * \tan^{-1}\frac{1 - r_0^2}{2*r_0} \quad (19)$$

Із виразу (19) знайдемо:

$$n = \frac{r_0(1 - r_0^2)}{2*\left(1 - r_0 - 1 - \frac{2}{\pi}*\tan^{-1}\frac{1 - r_0^2}{2*r_0}\right)} + 3 \quad (20)$$

Задавшись різними значеннями коефіцієнта кореляції r_0 визначимо значення n за виразом (20), вони наведені у табл. 2.

Таблиця 2.

Кількість вимірювань

r_0	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,99
n	4,8	4,8	4,9	4,9	5,0	5,1	5,2	5,3	5,4	5,6	5,7

Із табл. 2 видно, що рівність коефіцієнтів кореляції, які знайдені за формулами (4) і (12) або (18), відповідає при кількості вимірювань близько п'яти.

Для порівняння значень коефіцієнтів, які одержані за різними формулами, спочатку необхідно одержати формули за якими можливо знаходити коефіцієнти k_1 і k_2 в залежності від заданих величин ρ – теоретичного значення коефіцієнта кореляції.

При рівноточних вимірюваннях в двох рядах x і y максимально допустиме відношення емпіричних дисперсій визначається нерівністю [7]

$$\frac{\sigma^2(x)}{\sigma^2(y)} \leq Fq,\ \left(\sigma(x) \geq \sigma(y)\right), \tag{21}$$

де Fq – дисперсійне відношення, яке вибирається з таблиці F – розподілу в залежності від рівня значимості q і числа степенів вільності $n-1$. Враховуючи формули (3) і (7) і (21) відношення дисперсій представиться як

$$\frac{\sigma^2(x)}{\sigma^2(y)} = \frac{\sum_{i=1}^{n} \delta^2(x_i)}{\sum_{i=1}^{n} \delta^2(y_i)} = \frac{k_2}{k_1} = Fq. \tag{22}$$

Розглянемо випадок, коли відношення коефіцієнтів k_2 і k_1 буде максимальним, тоді значення коефіцієнтів кореляції, які визначені за формулами (18) і (12) будуть мати найбільшу різницю. Складемо систему рівнянь із .(22) і (8), із рішення якої знайдемо коефіцієнти k_1 і k_2.

$$k_2/k_1 = Fq,\ k_1 * k_2 = r_0^2\,, \tag{23}$$

звідки

$$k_1 = \frac{r_0}{\sqrt{Fq}}\ ;\ k_2 = r_0 * \sqrt{Fq}\ . \tag{24}$$

Для рівня значущості $q = 0.10$ і числа ступенів вільності $n-1 = 6-1 = 5$, знайдемо по таблиці дисперсійне відношення $Fq = 3.5$, при таких вихідних даних різниця в значеннях коефіцієнта кореляції, які знайдені за формулами (18) і (12) буде найбільшою, тобто розглядається самий найгірший випадок.

Задавшись різними значеннями теоретичного коефіцієнта кореляції ρ, за формулами (24) обчислимо величини k_1 і k_2, за формулою (8) коефіцієнт r_0, за формулою (4) r' за формулами (18) і (12) r, всі ці значення наведені в табл. 3.

Таблиця 3.

За значеннями теоретичного коефіцієнта кореляції ρ, за формулами (24) обчислимо величини k_1 і k_2, за формулою (8) коефіцієнт r_0, за формулою (4) r' за формулами (18) і (12) r

ρ	Коефіцієнти		Значення коефіцієнтів кореляції			
	k_1	k_2	за (8) r_0	за (4)	за (18) r	за (12) r
с	0	0	0	0	0	0
0,1	0,05345	0,18708	0,100	0,116	0,127	0,152
0,2	0,10690	0,37417	0,200	0,232	0,251	0,296
0,3	0,16036	0,56125	0,300	0,346	0,371	0,427
0,4	0,21381	0,74833	0,400	0,456	0,484	0,543
0,5	0,26726	0,95541	0,500	0,562	0,590	0,645
0,6	0,32071	1,12250	0,600	0,664	0,688	0,734
0,7	0,37417	1,30958	0,700	0,760	0,778	0,813
0,8	0,42762	1,49666	0,800	0,848	0,859	0,882
0,9	0,48107	1,68375	0,900	0,928	0,933	0,944
1,0	0,53452	1,87083	1,000	1,000	1,000	1,000

Коефіцієнти кореляції, які обчислені за формулою (18) знаходились при $k_1 = k_2$, все ж таки значення r, які знайдені за (18) більші ніж за формулою (4). Найбільших значень коефіцієнти кореляції досягають при знаходженні за формулою (12).

Наведемо приклад, виміряні значення зміщень точок взяті з роботи [6,7], які наведені в табл. 4, за якими розраховані коефіцієнти кореляції, коефіцієнти регресійної моделі (лінійної моделі), а також проведено аналіз зв'язків між незалежними величинами $S(h)$, $S(x)$, $S\,(y)$.

Таблиця 4.

Виміряні значення зміщень точок

№ точок	$S(h)$, мм	$S(x)$, мм	$S(y)$, мм	№ точок	$S(h)$, мм	$S(x)$, мм	$S(y)$, мм	№ точок	$S(h)$, мм	$S(x)$, мм	$S(y)$, мм
1	-5,0	-12,0	5,0	**8**	-4,0	-10,0	-3,0	**15**	-11,0	-8,0	12,0
2	-5,5	-10,0	7,0	**9**	-2,0	-8,0	7,0	**16**	-13,0	-8,0	9,0
3	-6,0	-4,0	15,0	**10**	-2,5	-5,0	-1,0	**17**	-15,0	-7,0	-1,0
4	-7,0	-3,0	10,0	**11**	-4,0	-1,0	0,0	**18**	-13,5	-4,0	0,0
5	-5,0	-7,0	9,0	**12**	-5,0	-5,0	-4,0	**19**	-10,0	1,0	4,0
6	-6,5	5,0	8,0	**13**	-7,0	0,0	5,0	**20**	-9,0	2,0	7,0
7	-3,0	6,0	-1,0	**14**	-9,0	2,0	9,0	**21**	-7,0	5,0	8,0

За виміряними значеннями, наведеними у табл. 4, визначені лінійні рівняння коефіцієнтів регресії за формулами

$$\begin{cases} y_i = a_{yx} + b_{yx} \cdot x_i \\ x_i = a_{xy} + b_{xy} \cdot y_i \end{cases}, \tag{25}$$

тоді рівняння (25) набудуть вигляду

$$\begin{cases} y_i = 5{,}116393 + 0{,}034426 \cdot x_i \\ x_i = -3{,}5608085 + 0{,}035971 \cdot y_i \end{cases}. \tag{26}$$

Друге рівняння (26) виразимо відносно y відповідно до рівнянням (10), одержимо:

$$\begin{cases} y_i = 5{,}116392 + 0{,}034426 \cdot x_i \\ y_i = 98{,}99047627 + \frac{x_i}{0{,}0355971} \end{cases} \tag{27}$$

ці рівняння графічно представлені на рис. 1.

За формулою (11) знайдемо кут між прямими (26) він буде

$$\varphi = 85{,}9682159 = 85^o58'05,'' 58$$

Коефіцієнти кореляції між цими величинами $r_{h,x}$ = 0,012274, $r_{h,y}$ = -0.,4465, $r_{x,y}$ = 0.,3519, тобто вони вказують на те, що лінійного зв'язку між величинами $S(h)$, $S(x)$, $S(y)$ не має, і на це ще вказує, те що кут φ між регресійними рівняннями досить близький до $\pi/2$

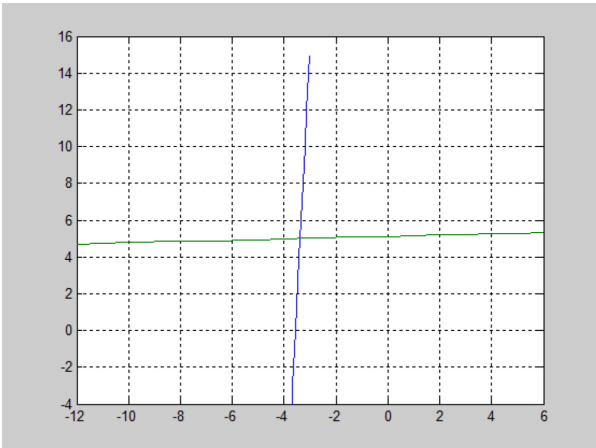

Рис. 1. Графічне представлення лінійних рівнянь регресії (26)

Лінії регресії перетинаються в точці центру ваги з координатами

$\bar{x}_{cp} = -3{,}381$ з СКП, $m_x = 5{,}23$мм; $\bar{y}_{cp} = 5{,}00$ з СКП, $m_y = 5{,}13$мм.

Коефіцієнт кореляції дорівнює середньому геометричному з обох коефіцієнтів регресії (25), підставивши значення з (26), одержимо

$$|r_{xy}| = \sqrt{b_{yx} \cdot b_{xy}} = \sqrt{0{,}034426 \cdot 0{,}035971} = 0{,}035190. \tag{28}$$

При знаходженні коефіцієнтів лінійної регресії середня квадратична помилка апроксимації рівняннями (26) $m_{anp} = 5{,}55$мм, кореляційне відношення 0,67; показник кореляційного зв'язку γ = 0,45 з СКП $m_\gamma = 0{,}161$;

критерій достовірності лінійного зв'язку *tg* = 2,816 досить низький, тому передбачається не лінійний зв'язок. Знайдемо рівняння регресії другої степені.

Не лінійні рівняння регресії другої степені мають вид

$$\begin{cases} y_i = 5.205025 + 0.0120003 \cdot x_i - 0{,}0042065 \cdot x_i^2 \\ x_i = -3{,}374994 + 0{,}374548 \cdot y_i - 0{,}0364964 \cdot y_i^2 \end{cases}, \quad (29)$$

вони визначені з середніми квадратичними помилками апроксимації $m_y = 5{,}74$ мм, $m_x = 5{,}71$ мм, графічно представлені на рис. 2

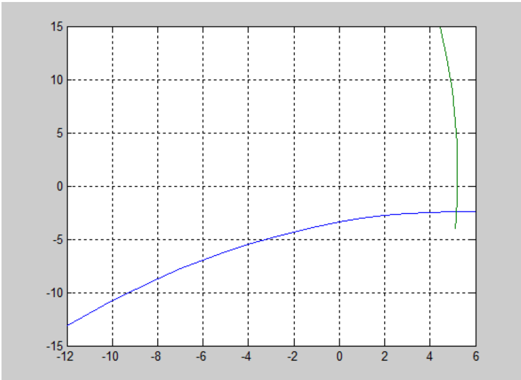

Рис. 2. Графічне представлення не лінійних рівнянь регресії (27)

Лінії регресії перетинаються в точці з координатами $\bar{x}_{cp} = -3{,}381$ з СКП, $m_x = 5{,}23$мм; $\bar{y}_{cp} = 5{,}00$ з СКП, $m_y = 5{,}13$мм.

При рівності СКП апроксимації рівнянь y_i і x_i (29) $m_x = m_y$, а також при теоретичному коефіцієнті кореляції $\rho = 0$, січення площиною $x0y$ розсіяння виміряних значень утворюють коло з радіусом який дорівнює СКП вимірювань, при $m_x \neq m_y$ розсіяння у вигляді еліпса, який звужується до регресійної прямої при теоретичному коефіцієнті кореляції $\rho \Longrightarrow 1$. Поле розсіяння вимірювань наведено на рис. 3.

Добуток відхилень від відповідних середніх значень є мірою степені взаємопов'язаності вимірювань

$$Q_{xy} = \sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - {}^{(\sum x \cdot \sum y)}\!/_{n}, \quad (30)$$

Середній добуток відхилень є оцінкою (S_{xy}) – коваріації σ_{xy}

$$S_{xy} = {}^{Q_{xy}}\!/_{(n-1)} = {}^{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}\!/_{(n-1)}. \quad (31)$$

Стандартне відхилення значень для *y* при умові, що *x* приймає визначене значення, дорівнює

$$S_{y.x} = \sqrt{\Sigma(y - a - b_{yx} \cdot x)^2 / (n-2)} \cdot \quad (32)$$

Стандартна помилка прогнозу передбачається при n – 2, за вимірюваннями оцінюються два параметри a та b_{yx} .

$$S_{b_{yx}} = {S_{y.x}} / {\sqrt{Q_x}} , \quad (33)$$

$$S_{a_{yx}} = S_{b_{yx}} \sqrt{\Sigma x^2 / n} \quad . \quad (34)$$

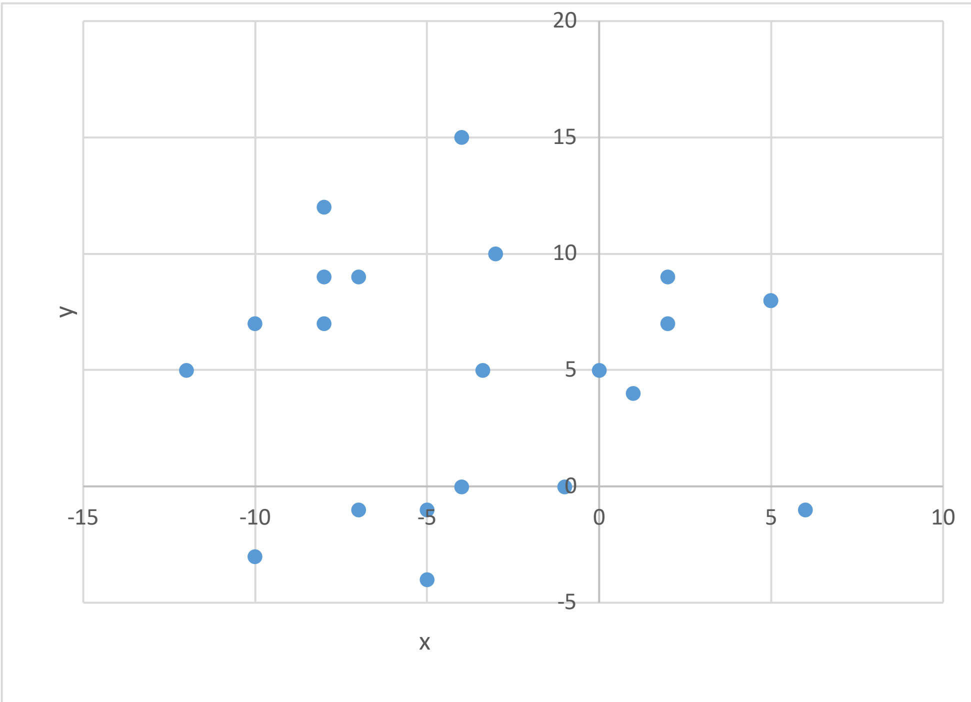

Рис. 3. Поле розсіяння виміряних величин

Квадрат стандартної помилки оцінювання є остаточною дисперсією $S_{y.x}^2$-розсіювання відносно прямої регресії – це дисперсія y після виключення впливу розсіяння x, тобто

$$S_{y.x}^2 = S_y^2(1 - r^2)\frac{n-1}{n-2}. \quad (35)$$

Кількість градацій які розрізнюються у вимірювальній величині

$$N = {L_y} / {L_{\Delta y}} , \quad (36)$$

кожний з розмірів L_y і $L_{\Delta y}$ характеризується відповідним значенням середнього квадратичного відхилення окремих точок від їх середнього арифметичного $\bar{y}$ у вигляді σ_y та окремих точок від лінії $y = kx$ у вигляді $\sigma_{\Delta y}$.

Приведена помилка γ, яка використовується в вимірювальній техніці є відношення половини ширини полоси невизначеності $L\_\Delta y\ /\ 2$ до довжини діапазону L_y, що наведено на рис. 4, тобто

$$\gamma = {}^{L_{\Delta y}}\!/_{2L_y} = {}^{1}\!/_{(2N)}\,, \tag{37}$$

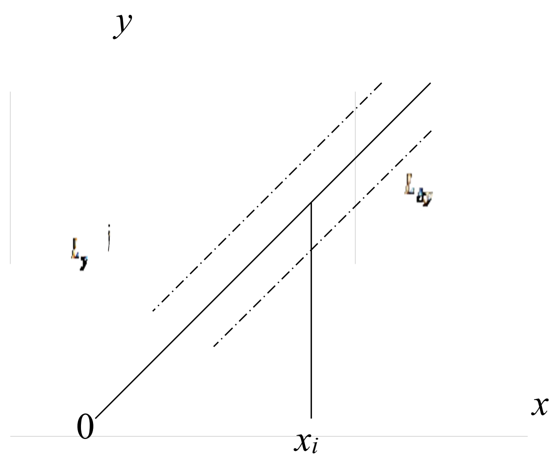

Рис. 4.

вона як і будь – яка інша оцінка помилки є негативною характеристикою тісноти кореляційного зв'язку, тобто відносно витягнутості поля експериментальних точок, для цієї цілі користуються позитивною оцінкою у вигляді коефіцієнта кореляції

$$\rho = 1/\sqrt{1 + \left(\sigma_{\Delta y}/\sigma_y\right)^2} \approx \sqrt{1 - \left({}^{\sigma_{\Delta y}}\!/_{\sigma_y}\right)^2}\,. \tag{38}$$

Враховуючи, що при вимірюваннях помилка складає одиниці або долі відсотка, тобто $\sigma_{\Delta y} \le \sigma_y$, практично

$$\rho = \sqrt{1 - \left({}^{\sigma_{\Delta y}}\!/_{\sigma_y}\right)^2} = \sqrt{1 - (2\gamma)^2}\,, \tag{39}$$

у цих виразах $\sigma_{\Delta y}$ середнє квадратичне відхилення експериментальних точок від прямої $y = \bar{y}$, σ_y – середнє квадратичне відхилення тих самих точок від горизонтальної прямої на рівні $\bar{y}$. Оцінка $\sigma_{\Delta y}$ – є абсолютна оцінка ширини полоси розсіяння, а оцінка σ_y – є оцінка ширини діапазону змінення значень y_i . Визначаючи приведену відносну помилку як

$$\gamma = {}^{\Delta y}\!/_{Y_K}\,, \tag{40}$$

під Δy приймають не всю ширину полоси невизначеності, а лише її половину, в той час як Y_K – це не відхилення значень y_i від $\bar{y}$, а весь діапазон змінення від y_{min} до y_{max}, тоді співвідношення (38) приймає вигляд (39), або

$$\gamma = \frac{1}{2}\sqrt{1 - \rho^2}, \quad \text{якщо } \rho < 0{,}9\,. \tag{41}$$

Під приведеною відносною помилкою розуміють не середнє квадратичне відхилення виміряних величин, а інтервальну (довірчу або ентропійну) оцінку, співвідношення між якими визначається законом розподілу. Розподіл середньо квадратичного відхилення σ_y визначається розміщенням експериментальних точок по осі y найбільш близько до рівномірного розподілу, тому для нього ентропійне значення дорівнюється $\Delta_{ep} \approx \sqrt{3}\,\sigma_y$. Розподіл який характеризується середнім квадратичним відхилення σ_Δ, що визначається помилками експериментальних значень вимірювань є колоколоподібний, тобто близький до нормального, його ентропійне значення $\Delta_{en} = 2{,}066\,\sigma_\Delta$, тоді ентропійне значення приведеної відносної помилки експериментальних даних буде

$$\gamma_e = \frac{\Delta_{en}}{2\Delta_{ep}} = \frac{2{,}066\sigma_\Delta}{2\sqrt{3}\cdot\sigma_y} = 0{,}5964 \cdot {}^{\sigma_\Delta}/_{\sigma_y}\,, \tag{42}$$

$${}^{\sigma_\Delta}/_{\sigma_y} = 1{,}67672\gamma_e\,, \tag{43}$$

з урахування цього з виразів (39) і (41) одержимо:

$$\rho = \sqrt{1 - 3\cdot\gamma_e^2}, \tag{44}$$

і

$$\gamma_e = \sqrt{(1-\rho^2)/3} = 0{,}57735\sqrt{1-\rho^2}, \text{ якщо } \rho > 0{,}9. \tag{45}$$

Ці співвідношення дають можливість простої оцінки коефіцієнта кореляції ρ

$$\rho^2 = \frac{(\sum_1^n x_i\cdot y_i)^2}{\sum_1^n x_i^2\cdot\sum_1^n y_i^2}\,. \tag{46}$$

Одержавши більш точне значення за формулою (46), яке обчислюється на ЕОМ можна за формулами (41) і (45) розрахувати більш точне значення ентропійної оцінки помилки виміряних значень, яка базується на середній квадратичній оцінці відхилень усіх експериментальних точок від кривої регресії.

Висновки. Таким чином, як видно із табл. 3 оцінка коефіцієнта кореляції починаючи з $n = 6$, яка виконана за формулою (4), буде зміщеною, тому точнішою оцінку можливо одержати за формулою (12), формулу (18) в розрахунок не приймаємо, тому що вона одержана для ідеального випадку коли $k_1 = k_2 = r_0$, така ситуація буває зрідка. Значення r знайдені за формулою (18) будуть меншими, ніж знайдені за формулою (12). За суттю коефіцієнти, які одержані за формулою (12) будуть максимальними при відповідному рівню значущості і числі степенів свободи, яка одержана на основі того, що рівень значущості $q > 0{,}25$ і при великій кількості вимірювань, $Fq \to 1$.

Одержавши більш точне значення за формулою (46), яке обчислюється на ЕОМ можна за формулами (41) і (45) розрахувати більш точне значення ентропійної оцінки помилки виміряних значень, яка базується на середній квадратичній оцінці відхилень усіх експериментальних точок від кривої регресії.

Обчислення більш точних значень коефіцієнтів кореляції (нормованих кореляційних моментів) є важливим в цифровій фотограмметрії при взаємному орієнтуванні і геодезичній прив'язці стереопари знімків, а також при дослідженні систематичних помилок в рядах геодезичних вимірювань. Занижені значення коефіцієнтів кореляції будуть впливати на рівняння кореляційної функції, і в кінцевому випадку будуть давати викривлене представлення про величини систематичних помилок.

Список джерел

1. Видуев Н.Г., Кондра Г.С. Вероятностно – статистический анализ погрешностей измерений. М.: Недра, 1969. – 320 с.

2. Закс Л. Статистическое оценивание. М.: Статистика, 1976. – 598 с.

3. Gladilin V., Belenok V., Kryachok S., Siroshtan T., Hamalii I. New Formula for Finding the Correlation Coefficient in Geodetic Measurements for a Small Number of Observations. Geodetski list, 2022, 2, p. 153-168

4. Гладилин В.Н. Точность геодезических измерений при выверке промышленного оборудования. – К.: Техніка, 1996. – 224 с.

5. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. – М.: Наука, 1974. – 254 с.

6. Gladilin V., Belenok V., Hebrin – Baidy L., Chookarina N. Structural method for determining deformations by geodetic measurement. Geodesy and Cartography. Vol. 45, No 2, 2019. – P92 -95. https://doi.org/10.3846/gac2019.6692

7. Gladilin V., Belenok V. & Shudra N. Determining the form of error distribution of geodetic messuring. Geodesy and Cartography. Vol. 48, No 2, 2022. – P56 -61. https://doi.org/10.3846/gac.2022.14403

5. Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. – Л.: Энергоатомиздат, 1991. – 304 с.

Ph.D., associate professor **Gladilin Valeriy**,
Ph.D., associate professor **Siroshtan Tatiana**,
Ph.D., associate professor **Hamalii Iryna**, Assistant **Sviderska Tetyana**,
Belotserkovsky National Agrarian University,
Senior Lecturer **Shudra Nataliia**, Senior Lecturer **Chulanov Petro**,
Kyiv National University of Construction and Architecture

CALCULATION OF THE CORRELATION COEFFICIENT IN GEODETIC MEASUREMENTS

A formula for determining the correlation coefficients is proposed, according to which with a small number of measurements the values of the coefficients will be

larger compared to the values calculated by known formulas, i.e., these values will be less offset relative to the true values of correlation coefficients.

In geodesy in general are limited to a small number of measurements, in photogrammetry at mutual orientation of pictures of a stereo pair in an automatic mode at identification of points on the left and right pictures it is necessary to calculate precisely values of correlation coefficients.

Keywords: correlation coefficient; linear and nonlinear regression coefficients in geodetic measurements.

REFERENCES

1. Viduev N.G., Kondra G.S. Veroyatnostno – statisticheskij analiz pogreshnostej izmerenij. M.: Nedra, 1969. – 320 s. {in Russian}.

2. Zaks L. Statisticheskoe ocenivanie. M.: Statistika, 1976. – 598 s. {in Russian}.

3. Gladilin V., Belenok V., Kryachok S., Siroshtan T., Hamalii I. New Formula for Finding the Correlation Coefficient in Geodetic Measurements for a Small Number of Observations. Geodetski list, 2022, 2, p. 153-168. {in English}.

4. Gladilin V.N. Tochnost geodezicheskih izmerenij pri vyverke promyshlennogo oborudovaniya. – K.: Tehnika, 1996. – 224 s. { in Russian }.

5. Tihonov A.N., Arsenin V.Ya. Metody resheniya nekorrektnyh zadach. – M.: Nauka, 1974. – 254 s. {in Russian}.

6. Gladilin V., Belenok V., Hebrin – Baidy L., Chookarina N. Structural method for determining deformations by geodetic measurement. Geodesy and Cartography. Vol. 45, No 2, 2019. – P92 -95. . https://doi.org/10.3846/gac2019.6692 {in English}.

7. Gladilin V., Belenok V. & Shudra N. Determining the form of error distribution of geodetic messuring. Журнал VILNIUS TECH "GEODESY & CARTOGRAPHY". 2022. 56-61. https://doi.org/10.3846/gac.2022.14403. {in English}.

8. Novickij P.V., Zograf I.A. Ocenka pogreshnostej rezultatov izmerenij. – L.: Energoatomizdat, 1991. – 304 s. {in Russian}.