Практичне заняття

Тема. Кореляційний аналіз

Цілі та задачі: сформувати поняття функціональної, статистичної, кореляційної залежності; коефіцієнта кореляції; розглянути властивості коефіцієнта кореляції.

Знання та вміння: студенти повинні знати формули для обчислення числових характеристик за допомогою коефіцієнта кореляції; вміти застосовувати їх для встановлення виду залежності між випадковими величинами, вміти будувати графік кореляційної залежності..

Форми та методи контролю: розв'язання задач біля дошки, самостійна робота, поточний контроль по карточкам.

Час:2 год.

Домашнє завдання: [1] № 442; 451; 462; 467.

Теоретичні відомості.

Кореляційний аналіз

Функціональним називають зв'язок між ознаками, при якому кожному значенню однієї змінної відповідає чітко окреслене значення іншої змінної.

Кореляційним (статистичним) зв'язком називається такий зв'язок, при якому чисельному значенню однієї змінної відповідає кілька значень іншої.

Кореляційною залежністю y від x називається така залежність, при якій зміни випадкової величини x спричинюють зміни середнього значення змінної y ($\bar{y}_x$), тобто $\bar{y}_x = f(x)$.

Вибірковим коефіцієнтом кореляції називається число:

$$r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{n\sigma_x \sigma_y},$$

де $\bar{x}, \bar{y}$ - вибіркові середні для x і y, тобто $\bar{x} = \frac{\sum x_i}{n}$, $\bar{y} = \frac{\sum y_i}{n}$.

σ_x, σ_y - вибіркові середньоквадратичні відхилення для x і y

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$$

$$\sigma_y = \sqrt{\sigma_y^2} = \sqrt{\frac{\sum y_i^2}{n} - \bar{y}^2}.$$

Властивості коефіцієнта кореляції:

a) абсолютна величина коефіцієнта кореляції не перевищує 1

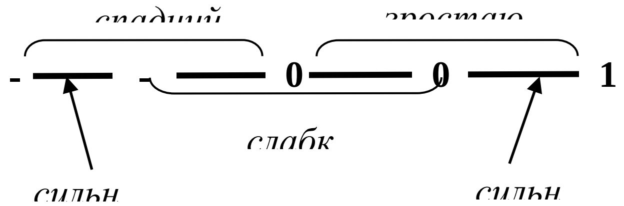
$$|r| \le 1$$

b) якщо $r = \pm 1$, то x і y зв'язані точкою лінійного зв'язку: $y = ax + b$;

c) якщо $r = 0$, то між x і y немає лінійного зв'язку, але криволінійна залежність можлива;

d) чим ближче r до ± 1, тим сильніше лінійний зв'язок між x і y і, чим ближче r до 0, тим він слабший;

e) якщо $r > 0$, зв'язок між x і y зростаючий, $r < 0$, зв'язок – спадний.



<u>Рівняння лінійної регресії</u> $y = ax + b$.

Параметри лінійної регресії дорівнюють:

$$a = r\frac{\sigma_y}{\sigma_x}, \quad b = \bar{y} - a\bar{x}.$$

<u>Перевірка гіпотези про значимість коефіцієнта кореляції .</u>

Гіпотеза H_o : лінійного кореляційного зв'язку для даної генеральної сукупності немає.

а) Визначаємо значення критерію, що спостерігається

$$t_p = r\sqrt{\frac{(n-2)}{(1-r^2)}} = r\sqrt{\frac{n-2}{1-r^2}}.$$

б) по таблиці Сть'юдента визначають $t_m = (\alpha,\ k = n-2)$.

в) при $|t_p| \geq t_m$ - нульову гіпотезу відкидають, при $|t_p| < t_m$ - H_o приймають.

Практична частина

Приклад. Задана залежність врожайності ***y*** (ц/га) від якості ґрунту ***x*** (у балах).

x_i	-1	0	1	4
y_i	-1	2	0	2

Знайти:

- коефіцієнт кореляції;
- рівняння лінійної регресії.
- перевірити коефіцієнт кореляції на значимість.

<u>Рішення**:**</u> Для зручності обчислень складемо розрахункову таблицю:

	x_i	y_i	$x_i\,y_i$	x_i^2	y_i^2
	-1	*-1*	*1*	*1*	*1*
	0	*2*	*0*	*0*	*4*
	1	*0*	*0*	*1*	*0*
	4	*2*	*8*	*16*	*4*
$\sum$	***4***	***3***	***9***	***18***	***9***

Обчислимо середнє значення для x и y:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{4}{4} = 1; \qquad \bar{y} = \frac{\sum y_i}{n} = \frac{3}{4}; \qquad (n=4)$$

Обчислимо середнє квадратичне відхилення для x і y:

$$\sigma_x = \sqrt{\frac{18 - 4 \cdot 1^2}{4}} = 2,16; \qquad \sigma_y = \sqrt{\frac{9 - 4 \cdot (0,75)^2}{4}} = 1,5 \cdot$$

Обчислюємо коефіцієнт кореляції:

$$r = \frac{9 - 4 \cdot 1 \cdot 0,75}{4 \cdot 2,16 \cdot 1,5} = 0,62 \cdot$$

На підставі властивостей коефіцієнта кореляції робимо висновок.

Оскільки $r=0,62 > 0$, то між x і y сильний, зростаючий лінійний кореляційний зв'язок.

Обчислюємо коефіцієнти лінійної регресії $y = ax + b$:

$$a = \frac{0,62 \cdot 1,5}{2,16} = 0,43; \qquad b = 0,75 - 0,43 = 0,32$$

Рівняння лінійної регресії має вид: $y = 0,43x + 0,32$

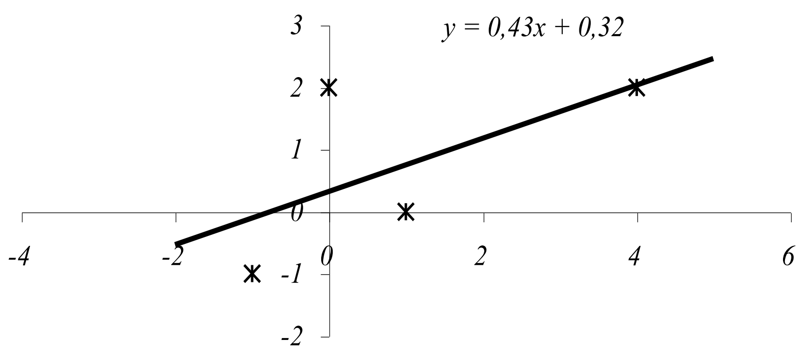
Побудуємо на координатній площині задані пари точок і отриману пряму.

Перевіримо коефіцієнта кореляції на значимість (критерій Сть'юдента).

$H_o : \rho = 0$ - для даної генеральної сукупності лінійного кореляційного зв'язку немає.

а) Обчислюємо значення критерію, що спостерігається

$$t_p = 0,62 \cdot \sqrt{\frac{(4-2)}{(1-0,62^2)}} \approx 1,12,$$



б) по таблиці Сть'юдента визначаємо

$$t_m(\alpha = 0,05;\ k = n - 2 = 4 - 2 = 2) = 4,30 \cdot$$

Оскільки $|t_p| < t_m$ (1,12 < 4,30) нульову гіпотезу відкидаємо, тобто коефіцієнт кореляції для всієї генеральної сукупності не дорівнює нулю.

Приклади для самостійного розв'язання

Наведені данні. Визначити:

- коефицієнт корреляції, зробити висновок;
- параметри лінійної регрессії $y = ax + b$;
- надійність коефицієнта корреляции;
- значимість коефицієнта корреляції;
- побудувати графік залежності $y = f(x)$

2.1	X	2	4	5	6	7	**2.2**	***X***	1	2	3	4	5
	Y	-3	5	9	4	2		***Y***	1	-5	-4	-3	-1
2.3	***X***	0	1	2	3,5	4	**2.4**	***X***	0	1	2	3	4
	Y	4	2,5	3	-2	3		***Y***	1	3	2	4	6
2.5	***X***	-2	-1	0	1	2	**2.6**	***X***	-1	0	3	4	5
	Y	1	1,5	4	3	1		***Y***	2	1	4	-2	3
2.7	X	-1	0	1	2	3	**2.8**	***X***	-2	0	1	2	3
	Y	-3	0	2	4	2		***Y***	1	-5	2	-4	2
2.9	***X***	1	2	3	4	5	**2.10**	***X***	-1	0	1	2	4
	Y	2	3	4	2	5		***Y***	0	3	9	4	2
2.11	***X***	-2	0	1	2	3	**2.12**	***X***	-1	0	3	4	5
	Y	-1	-5	5	3	2		***Y***	-1	5	2	4	2
2.13	X	-3	-2	-1	0	2	**2.14**	***X***	-1	0	1	3	5
	Y	-3	3	2	4	2		***Y***	3	2	1	4	2
2.15	***X***	-1	0	1	2	4	**2.16**	***X***	-1	0	1	2	6
	Y	2	1	9	4	2		***Y***	2	8	3	4	3
2.17	***X***	-2	-1	0	1	2	**2.18**	***X***	1	2	3	4	5
	Y	-2	5	2	4	2		***Y***	-1	-5	2	4	2
2.19	X	-1	0	1	2	5	**2.20**	***X***	-3	-2	-1	0	2
	Y	1	2,5	2	3	1		***Y***	1	3	4	3	1
2.21	***X***	1	2	3	4	5	**2.22**	***X***	-2	-1	0	2	3
	Y	1	-3	4	3	1		***Y***	-5	2	4	-3	1
2.23	***X***	-2	0	1	2	3	**2.24**	***X***	-3	-2	-1	0	2
	Y	1	2,5	4	3	1		***Y***	1	3	4	3	1
2.25	X	-3	-2	-1	0	1	**2.26**	***X***	1	2	3	4	5
	Y	2	5	3	3	1		***Y***	5	3	5	3	1
2.27	***X***	-3	-2	-1	0	2	**2.28**	***X***	-2	-1	0	1	2
	Y	6	4	2	3	1		***Y***	8	3	0	2	-1
2.29	***X***	-3	-1	0	1	2	**2.30**	***X***	-4	-2	0	1	3
	Y	5	3	0	1	1		***Y***	1	4	2	0	-3

Питання і завдання для самоконтролю.

1. Сформулюйте поняття функціональної, статистичної, кореляційної залежності.
2. Дайте означення коефіцієнта кореляції. Назвіть властивості коефіцієнта кореляції.
3. Записати формули для обчислення числових характеристик статистичного ряду розподілу за допомогою коефіцієнта кореляції.

Список використаної літератури.

1. Гурман В.Е. “Руководство к решению задач по теории вероятностей и математической статистике ”, учеб.пособие для ВТУЗов.Изд. 2-е доп.,М.: Высшая школа, 1975. -333 с.
2. Кремер Н.Ш. “Теория вероятностей и математическая статистика ”, Учебник для вузов.- М.: ЮНИТИ-ДАНА, 2001. – 543 с.