МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ВНУТРІШНІХ СПРАВ
КРЕМЕНЧУЦЬКИЙ ЛЬОТНИЙ КОЛЕДЖ

Циклова комісія природничих дисциплін

ТЕКСТ ЛЕКЦІЇ

з навчальної дисципліни «Теорія ймовірностей і математична статистика» обов'язкових компонент освітньо-професійної програми першого (бакалаврського) рівня вищої освіти
Облік і аудит

за темою №9– Найпростіші випадки криволінійної кореляції. Поняття багатофакторної регресії

Харків 2022

ЗАТВЕРДЖЕНО
Науково-методичною радою
Харківського національного
університету внутрішніх справ
Протокол від 30.08.2022 №8

СХВАЛЕНО
Методичною радою Кременчуцького
льотного коледжу Харківського
національного університету
внутрішніх справ
Протокол від 22.08.2022 №1

ПОГОДЖЕНО
Секцією Науково-методичної ради
ХНУВС з технічних дисциплін
Протокол від 29.08.2022 №8

Розглянуто на засіданні циклової комісії природничих дисциплін, протокол від 10.08.2022 № 1

Розробник: доцент циклової комісії природничих дисциплін, к.ф. –м.н. Семенов В.О.

Рецензенти:
1. Доцент кафедри автомобілі та трактори Кременчуцького національного університету імені Михайла Остроградського, кандидат технічних наук Черниш А.А.
2. Ст.викладач циклової комісії економіки та управління Кременчуцького льотного коледжу Харківського національного університету внутрішніх справ, кандидат економічних наук Цимбалістова О.А.

План лекції

1. Нелінійна кореляційна залежність. Гіперболічна кореляція. Кореляційне відношення.
2. Параболічна кореляція.

Рекомендована література:

Основна

1. Барковський В. В. Теорія ймовірностей та математична статистика ; 5-те видання : навчальний посібник / В. В. Барковський, Н. В. Барковська, О. К. Лопатін. – К. : Центр учбової літератури, 2010. – 424 с.
2. Волощенко А. Б. Теорія ймовірностей і математична статистика : навчаль-но-методичний посібник для самостійного вивчення дисципліни / А. Б. Волощенко, І. А. Джалладова. – К. : КНЕУ, 2003. – 256 с.
3. Жильцов О.Б. Теорія ймовірностей та математична статистика у прикладах і задачах: навч. посіб. для студ. вищ. навч. закл. / О.Б. Жильцов ; за ред. Г.О. Михаліна. — К. : Київ. ун-т ім. Б. Грінченка, 2015. — 336 с.
4. Конспект лекцій.

Інформаційні ресурси в інтернеті

5. Жильцов О.Б. Теорія ймовірностей та математична статистика у прикладах і задачах: навч. посіб. для студ. вищ. навч. закл. / О.Б. Жильцов ; за ред. Г.О. Михаліна. — К. : Київ. ун-т ім. Б. Грінченка, 2015. — 336 с. [https://yandex.ua/search/?lr=143&oprnd=9346860898&text=жильцов%20теорія%20ймовірностей

Текст лекції

1. ***Нелінійна кореляційна залежність. Гіперболічна кореляція.***

Якщо відображені на площині XOY групи точок $(x_i, \overline{y}_{x_i})$ розміщуються, нагадуючи деякі криві, то доцільно вважати, що між досліджуваними величинами існує нелінійна залежність. Виникає завдання підібрати таку криву, яка б на основі методу найменших квадратів мала найменші відхилення від точок, здобутих при спостереженні, знайти її рівняння і визначити тісноту зв'язку. Розглянемо деякі найпростіші види нелінійної кореляційної залежності. Нехай зі зростанням однієї випадкової величини умовні середні другої спадають, але не на ту саму величину, як це буває в разі лінійної залежності, а розмір зміни ніби згасає. У такому разі можна вважати, що залежність *гіперболічна*:

$$\overline{y}_x = \frac{a}{x} + b.$$

Параметри a і b за методом найменших квадратів визначаються із системи рівнянь:

$$\begin{cases} a\sum_{i=1}^{m}\frac{1}{x_i}n_{x_i}+b\sum_{i=1}^{m}n_{x_i}=\sum_{i=1}^{m}\overline{y}_{x_i}n_{x_i}, \\ a\sum_{i=1}^{m}\frac{1}{x_i^2}n_{x_i}+b\sum_{i=1}^{m}\frac{1}{x_i}n_{x_i}=\sum_{i=1}^{m}\frac{1}{x_i}\overline{y}_{x_i}n_{x_i}. \end{cases}$$

У разі нелінійної кореляційної залежності тіснота зв'язку між величинами характеризується кореляційним відношенням. ***Кореляційним відношенням*** називається відношення середніх квадратичних відношень умовних середніх до загального середнього квадратичного відхилення: $\eta_{y/x}=\frac{\delta_y}{s_y}$, де

$$\delta_y=\sqrt{\frac{\sum_{i=1}^{m}\left(\overline{y}_{x_i}-\overline{y}\right)^2 n_{x_i}}{n}}.$$

Кореляційне відношення набуває значення на відрізку [0;1].Якщо кореляційне відношення дорівнює нулю, то кореляційний зв'язок відсутній, якщо η =1, то випадкові величини зв'язані функціональною залежністю. Зі зростан-ням значення η тіснота кореляційного зв'язку збільшується.

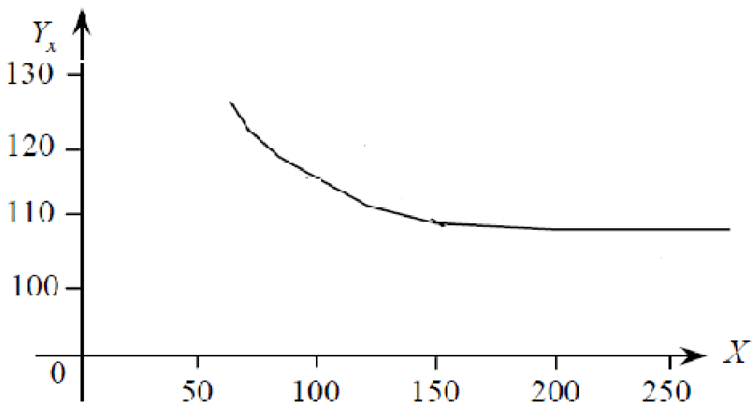
Приклад **1**. У результаті обстеження одержано статистичний розподіл 30 однотипних підприємств по добовому виробленню продукції X і собівартості одиниці цієї продукції Y. Установити форму залежності між X і Y, знайти рівняння лінії регресії і оцінити тісноту зв'язку.

x \ Y	100	110	120	130	n_{x_i}
50			1	3	4
100		3	3		6
150		6	2	1	9
200	1	4		1	6
250	4	1			5
n_{y_j}	5	14	6	5	30

Розв'язання. Знаходимо умовні середні значення $\overline{y}_{x_i}$. Результати обчислень заносимо в таблицю. У цій самій таблиці зроблено перехід до умовних змін-них. Переходячи до умовних змінних, ураховуємо, що C_1 =150, $h_1=50$, C_2 = 110, h_2=10.

$u \backslash v$	−1	0	1	2	n_{x_i}	$\bar{y}_{x_i}$
−2			1	3	4	127,5
−1		3	3		6	155
0		6	2	1	9	114,4
1	1	4		1	6	111,7
2	4	1			5	102

Зобразимо на координатній площині множину точок $(x_i, \bar{y}_{x_i})$.Згідно з рисун-ком ця група точок розміщена приблизно на деякій гіперболі, дещо відхиля-ючись від неї.



Рівняння гіперболи $\bar{y}_x = f(x)$ шукаємо у вигляді $\bar{y}_x = \dfrac{a}{x} + b$.Для визначення коефіцієнтів відповідної системи рівнянь складаємо таблицю:

x_i	n_{x_i}	$\dfrac{n_{x_i}}{x_i}$	$\dfrac{n_{x_i}}{x_i^2}$	$\bar{y}_{x_i}$	$\bar{y}_{x_i} n_{x_i}$	$\dfrac{\bar{y}_{x_i} n_{x_i}}{x_i}$
50	4	0,08	0,0016	127,5	510	10,2
100	6	0,06	0,0006	115	690	6,9
150	9	0,06	0,0004	114,4	1030	6,86
200	6	0,03	0,00015	111,7	670	3,35
250	5	0,02	0,00008	102	510	2,04
Сума	30	0,25	0,00283	—	3410	29,35

Невідомі параметри a і b знаходимо із системи рівнянь:

$$\begin{cases} 0,25a + 30b = 3410 \\ 0,00283a + 0,25b = 29,35. \end{cases}$$

Розв'язок системи $a \approx 1250,\ b \approx 103,3.$ Рівняння регресії має вигляд:

$$\overline{y}_x = \frac{1250}{x} + 103,3.$$

Тісноту зв'язку між випадковими величинами оцінимо за допомогою кореляційних відношень. Необхідні для розрахунків параметри знайдемо з допомогою умовних моментів розподілу:

$$\overline{v} = \frac{11}{30}; \quad \overline{v^2} = \frac{31}{30}; \quad \overline{y} = 10 \cdot \frac{11}{30} + 110;$$

$$s_y = 10\sqrt{\frac{31}{30} - \left(\frac{11}{30}\right)^2} \approx 9,5; \quad \delta_y = \sqrt{\frac{\sum_{i=1}^{5} (\overline{y}_i - 114)^2 n_{x_i}}{30}} \approx 7,2.$$

Отже, кореляційне відношення $\eta_{y/x} = \dfrac{7,2}{9,5} \approx 0,76.$ З огляду на значення коре-ляційного відношення можна стверджувати, що між добовим виробітком продукції і собівартістю одиниці продукції існує досить істотна кореляційна залежність.

2. Параболічна кореляція.

Нехай зі зростанням однієї випадкової величини умовні середні другої зростають (спадають), досягають максимуму (мінімуму), а потім спадають (зростають). Тоді можна вважати, що між ними існує параболічна залежність виду:

$$\overline{y}_x = a_2 x^2 + a_1 x + a_0.$$

За методом найменших квадратів для визначення значень параметрів a_2 , a_1, a_0 потрібно скласти і розв'язати систему рівнянь:

$$\begin{cases} a_2 \sum_{i=1}^{m} x_i^2 n_{x_i} + a_1 \sum_{i=1}^{m} x_i n_{x_i} + a_0 \sum_{i=1}^{m} n_{x_i} = \sum_{i=1}^{m} \overline{y}_{x_i} n_{x_i}, \\ a_2 \sum_{i=1}^{m} x_i^3 n_{x_i} + a_1 \sum_{i=1}^{m} x_i^2 n_{x_i} + a_0 \sum_{i=1}^{m} x_i n_{x_i} = \sum_{i=1}^{m} x_i \overline{y}_{x_i} n_{x_i}, \\ a_2 \sum_{i=1}^{m} x_i^4 n_{x_i} + a_1 \sum_{i=1}^{m} x_i^3 n_{x_i} + a_0 \sum_{i=1}^{m} x_i^2 n_{x_i} = \sum_{i=1}^{m} x_i^2 \overline{y}_{x_i} n_{x_i}. \end{cases}$$

***Приклад* 2.** Дано кореляційну таблицю

X	Y					n_x
	2	3	5	7	8	
12	5	–	–	2	6	13
14	1	8	–	12	3	24
17	–	6	11	10	1	28
18	–	4	15	8	–	27
20	–	–	7	1	–	8
n_y	6	18	33	33	10	$n = 100$

Знайти вибіркове рівняння параболічної регресії Y на X. Оцінити коефіцієнт кореляції за допомогою вибіркового кореляційного відношення.

Розв'язання. Складемо розрахункову таблицю для обчислення коефіцієнтів вищенаведеної системи рівнянь.

X	n_{x_i}	$\overline{y}_{x_i}$	$n_{x_i}x_i$	$n_{x_i}x_i^2$	$n_{x_i}x_i^3$	$n_{x_i}x_i^4$	$n_{x_i}\overline{y}_{x_i}$	$n_{x_i}\overline{y}_{x_i}x_i$	$n_{x_i}\overline{y}_{x_i}x_i^2$
12	13	5,538	156	1872	22464	269568	72	864	10368
14	24	5,583	336	4704	65856	921984	134	1876	26264
17	28	5,393	476	8092	137564	2338588	151	2567	43639
18	27	5,296	486	8748	157464	2834352	143	2574	46332
20	8	5,25	160	3200	64000	1280000	42	840	16800
Σ	100		1614	26616	447348	7644492	542	8721	143403

Підставивши числа з останнього рядка таблиці у систему рівнянь , маємо систему рівнянь відносно невідомих коефіцієнтів a_0, a_1, a_2:

$$\begin{cases} 26616a_2 + 1614a_1 + 100a_0 = 542, \\ 447348a_2 + 26616a_1 + 1614a_0 = 8721, \\ 7644492a_2 + 447348a_1 + 26616a_0 = 143403. \end{cases}$$

Розв'язавши одержану систему рівнянь, маємо:

$$a_2 \approx -0,00424;\ a_1 \approx 0,08868;\quad a_0 \approx 5,1434.$$

Отже, рівняння параболічної регресії $\overline{y}_x = a_2x^2 + a_1x + a_0$ має вигляд:

$$y = –0,0434x^2 + 0,08868x + 5,1434.$$

Для того щоб знайти вибіркове кореляційне відношення $\eta_{y/x}$, визначимо вибіркове середнє $\overline{y}$, вибіркове середнє квадратичне відхилення s_y і міжгрупове середнє квадратичне відхилення δ_y:

$$\overline{y} = \frac{\sum_{j=1}^{5} y_j n_{y_j}}{n} = \frac{2\cdot 6 + 3\cdot 18 + 5\cdot 33 + 7\cdot 33 + 8\cdot 10}{100} \approx 5,42;$$

$$s_y = \sqrt{\frac{\sum_{j=1}^{5} y_j^2 n_{y_j}}{n} - \overline{y}^2} \approx 1,82;\quad \delta_y = \sqrt{\frac{\sum_{i=1}^{m}\left(\overline{y}_{x_i} - \overline{y}\right)^2 n_{x_i}}{n}} \approx 0,122;$$

Таким чином, шукане вибіркове кореляційне відношення

$$\eta_{y/x} = \frac{\delta_y}{s_y} = \frac{0,122}{1,82} \approx 0,067.$$

Це кореляційне відношення показує, що кореляційний зв'язок між випадко-вими величинами несуттєвий.