Технічні науки

УДК 004.021

Тронько Олександра Олексіївна
студентка
Харківського національного університету радіоелектроніки
Тронько Александра Алексеевна
студентка
Харьковского национального университета радиоэлектроники
Tronko Oleksandra
Student of the
Kharkiv National University of Radio Electronics

Вінницька Єва Артемівна
студентка
Харківського національного університету радіоелектроніки
Винницкая Ева Артемовна
студентка
Харьковского национального университета радиоэлектроники
Vinnytska Yeva
Student of the
Kharkiv National University of Radio Electronics

ДОСЛІДЖЕННЯ МОЖЛИВОСТЕЙ ВИКОРИСТАННЯ КОРЕЛЯЦІЙНО-РЕГРЕСІЙНОГО АНАЛІЗУ ДЛЯ ПРОГНОЗУВАННЯ ДАНИХ
ИССЛЕДОВАНИЯ ВОЗМОЖНОСТЕЙ КОРРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА ДЛЯ ПРОГНОЗИРОВАНИЯ ДАННЫХ

THE RESEARCH OF THE POSSIBILITIES OF CORRELATION-REGRESSION ALALYSIS FOR DATA PROGNOSIS

Анотація. *Досліджено застосування кореляційно-регресійного аналізу при прогнозуванні даних.*

Ключові слова: *кореляційно-регресійний аналіз, математична статистика, прогнозування, дані.*

Аннотация. *Исследовано применение корреляционно-регрессионного анализа при прогнозировании данных.*

Ключевые слова: *корреляционно-регрессионный анализ, математическая статистика, прогнозирование, данные.*

Summary. *The use of correlation-regression analysis for data prognosis has been researched.*

Key words: *correlation and regression analysis, mathematical statistics, prognosis, data.*

Постановка проблеми. Збільшення обсягу інформації викликає потребу в переробці та аналізі інформації. Обробка даних, що представляють собою частину членів деякої досить великої сукупності, відібраних з неї для отримання відомостей про всю сукупність – одна з задач, що часто виникають у практичній діяльності. Обробкою цих даних займається розділ математичної статистики. Це дозволяє вирішувати різноманітні практичні завдання, в тому числі прогнозувати ймовірність появи різних подій.

Мета статті – дослідити можливості кореляційно-регресійного аналізу для прогнозувань явищ на основі заданих даних певного підприємства.

Виклад основного матеріалу. Застосування інструментів математичного аналізу дозволяє дати кількісну оцінку досліджуваних явищ, незалежно від їх природи. Інтерпретація отриманих результатів залежить від того, наскільки правильно були проведені розрахунки, а також обгрунтовані застосовувані методи. Одним з найбільш поширених видів аналізу статистичної інформації є кореляційний аналіз, сутність якого полягає в пошуку зв'язку між двома або більше змінними. Наявність зв'язку характеризується тим, наскільки сильно вона виражена, і обумовлюється критерієм, який отримав назву коефіцієнта кореляції (r). Грададія сили зв'язку описана шкалою Чертока, що розділяє силу (тісноту) зв'язку на функціональну, сильну, середню, слабку, нульову [1, с. 6]. Крім ступеню тісноти, зв'язки між явищами та його ознаками класифікуються за напрямком (прямий і зворотній) і аналітичним вираженням.

Кореляція взаємопов'язана з регресією, оскільки перша оцінює силу (тісноту) статистичної зв'язку, друга досліджує її форму. Регресійний аналіз полягає у визначенні аналітичного вираження зв'язку в вигляді рівняння регресії, що описує кореляційну залежність між результативною ознакою і однією або декількома факторними.

Дослідимо економічну ситуацію щодо вивчення зв'язку між кількістю співробітників на підприємстві та виготовленої продукції за робочу зміну. На рис. 1 зазначені значення змінної X (кількість співробітників на підприємстві) та значення змінної Y (кількість виготовленої продукції).

Спостереження	Змінна *X*	Змінна *Y*
1	2	3
1	173	58
2	164	57
3	161	52
4	161	57
5	154	48
6	166	60
7	170	60
8	181	70
9	170	50
10	177	60
11	178	65

Рис. 1. Задані значення

Для подальшого розрахунку коефіцієнта лінійної кореляції розраховано необхідні величини. Отримані дані занесені в таблицю, що представлена на рис. 2.

Спостереження	Змінна *X*	Змінна *Y*	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	2	3	4	5	6	7	8
1	173	58	4,36	0,09	0,40	19,04	0,01
2	164	57	-4,64	-0,91	4,21	21,5	0,83
3	161	52	-7,64	-5,91	45,12	58,31	34,92
4	161	57	-7,64	-0,91	6,94	58,31	0,83
5	154	48	-14,64	-9,91	145,03	214,22	98,19
6	166	60	-2,64	2,09	-5,51	6,95	4,37
7	170	60	1,36	2,09	2,85	1,86	4,37
8	181	70	12,36	12,09	149,49	152,86	146,19
9	170	50	1,36	-7,91	-10,79	1,86	62,55
10	177	60	8,36	2,09	17,49	69,95	4,37
11	178	65	9,36	7,09	66,40	87,68	50,28
Сума	**1855,00**	**637,00**	**0,00**	**0,00**	**421,64**	**692,55**	**406,91**

Рис. 2. Розраховані величини

На основі розрахованих величин отримано коефіцієнт кореляції (r) рівний 0.79, що за градацією Чертока інтерпретує зв'язок як сильний.

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{L=1}^{n}(x_i-\bar{x})^2 \cdot \sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{421{,}64}{\sqrt{692{,}55 \cdot 406{,}91}} = 0{,}79.$$

За методом найменших квадратів визначено параметри рівняння лінійної регресії для подальшого прогнозування значень Y (кількості виробленої продукції) при заданні значення X (кількості співробітників).

Спостереження	X	Y	x_i^2	$x_i \cdot y_i$
1	2	3	4	5
1	173	58	29929	10034
2	164	57	26896	9348
3	161	52	25921	8372
4	161	57	25921	9177
5	154	48	23716	7392
6	166	60	27556	9960
7	170	60	28900	10200
8	181	70	32761	12670
9	170	50	28900	8500
10	177	60	31329	10620
11	178	65	31684	11570
Сума	**1855,00**	**637,00**	**313513,00**	**107843,00**

Рис. 3. Розрахунок необхідних величин для отримання значень параметрів методом найменших квадратів

Склавши систему рівнянь (рис. 4), знайдено необхідні параметри для визначення рівняння парної лінійної регресії.

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i, \\ a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i. \end{cases} \Rightarrow$$

$$\begin{cases} 313513a + 1855b = 107843, \\ 1855a + 11b = 637, \end{cases} \Rightarrow \begin{cases} a = 0,61, \\ b = -44,76, \end{cases} \Rightarrow \bar{y}_x = 0,61x - 44,76.$$

Рис. 4. Знаходження параметрів функції парної лінійної регресії

Для знаходження прогнозних значень кількості продукції, що виробить певна кількість робітників, необхідно у функцію парної лінійної регресії $\bar{y}_x = 0,61x - 44,76$. замість змінної x підставити у формулу передбачувану кількість робітників. Для кількості робітників X=150 прогнозне значення виробленої продукції становить приблизно $\bar{y}_x$= 0,61·

150 - 44,76 = 47 од. Для кількості робітників X=190 прогнозне значення виробленої продукції становить приблизно 71 од.

Отже, кореляційно-регресійний аналіз надає змогу знайти прогнозні значення на основі статистичних даних. Таким чином, проаналізувавши взаємозв'язок кількості співробітників підприємства та кількості виробленої продукції, можна зробити прогноз, що кількість виробленої продукції за участі 150 робітників складе 47 од., а за участі 190 робітників цей показник збільшиться до 71 од.

Література

1. Harchenko M.A. Uchebnoe posobie dlya vuzov. Korrelyatsionnyiy analiz. Voronezh: Voronezhskiy gosudarstvennyiy universitet, 2008. 30 s.