

ТЕМА 9. КОРЕЛЯЦІЙНО-РЕГРЕСІЙНИЙ АНАЛІЗ

Мета заняття:

Ознайомитися з можливостями та особливостями використання методу кореляційно-регресійного аналізу; навчити розраховувати коефіцієнти кореляції та регресії та їх аналізувати.

Обґрунтування мети. Кореляційно-регресійний аналіз зв'язку між різними ознаками чи явищами в досліджуваній сукупності дозволяє встановити її напрямок і силу, а також визначити вірогідність цієї залежності в генеральній сукупності. Практичне опанування цих методів допомагає в обґрунтуванні висновків, прийнятті рішень щодо розробки та застосування методів діагностики, тактики лікування, напрямів профілактики і т.д.

Основні поняття теми: кореляційно-регресійний аналіз, функціональний та кореляційний зв'язок, коефіцієнт кореляції, коефіцієнт регресії.

Навчально-цільові задачі:

Студенти повинні:

знати:

- види зв'язку між різними ознаками чи явищами;
- особливості застосування методу Пірсона та Спірмена;
- сутність регресійного аналізу;

вміти:

- визначати коефіцієнт кореляції лінійним та ранговим методами;
- оцінювати одержані результати, робити висновки.

Питання для передаудиторної підготовки

1. Вивчення зв'язку між кількісними перемінними.
2. Поняття про функціональний та кореляційний зв'язок.
3. Сила та напрям зв'язку.
4. Види коефіцієнтів кореляції.
5. Коефіцієнт лінійної кореляції Пірсона, його оцінка, характеристика.
6. Непараметричні методи оцінки зв'язку – ранговий коефіцієнт кореляції Спірмена.
7. Парні та множинні коефіцієнти кореляції.
8. Регресійний аналіз, коефіцієнт регресії, рівняння регресії.
9. Використання регресійного аналізу для прогнозування.

Чому так відрізняються показники частоти госпіталізації населення, тривалості лікування окремих пацієнтів, летальності в різних відділеннях тощо? З чим це пов'язано? Які чинники впливають на вищезазначені показники? В клінічній практиці **аналіз можливих зв'язків** між різними ознаками (явищами), між чинниками та явищами проводиться за допомогою кореляційно-регресійного аналізу. При вивченні зв'язків різних величин (ознак, явищ) в сукупності виявляється тенденція, характер, напрямок і сила їх, а також визначається вірогідність цієї залежності в генеральній сукупності, що є серйозною підставою в формуванні висновків і прийнятті рішень (напрямки профілактики, діагностика, тактика лікування і т.д.).

Додаткова література

1. Мармоза А.Т. Практикум по математической статистике. – К. – 1990. - С. 140-164.
2. Платонов А.Е. Статистический анализ в медицине и биологии: задачи, терминология, логика, компьютерные методы. - М.: Издательство РАМН, 2000. – С. 26-32.
3. С. Гланц. Медико-биологическая статистика. Пер. с англ. — М., Практика, 1998. — С. 221-275

Типовий приклад 1. Визначення коефіцієнта рангової кореляції (Спірмена)

При вивченні стану надання стаціонарної допомоги в районі Б. необхідно дослідити зв'язок між частотою оперативних втручань (т.зв. хірургічною активністю) та післяопераційною летальністю пацієнтів, що були пізно доставлені (після 24 годин від початку захворювання).

Нозологічна форма	Частота операцій (x)	Частота летальності (y)	Порядкові номера (ранги) ознак		Різниця рангів d (x-y)	Квадрат різниці рангів d ²
			x	y		
Шлунково-кишкова кровотеча	0,5	6,6	1	6	-5	25
Гострий панкреатит	0,8	12,7	2	8	-6	36
Проривна виразка шлунка та 12-палої кишки	1,5	12,6	3	7	-4	16
Гостра непрохідність кишок	1,7	4,6	4	5	-1	1
Травми внутрішніх органів грудної та черевної порожнини і тазу	1,9	4,2	5	4	1	1
Защемлена грижа	2,8	2,6	6	3	3	9
Гострий холецистит	4,6	0,56	7	2	5	25
Гострий апендицит	16,4	0,1	8	1	7	49
						$\sum d^2=162$

Коефіцієнт рангової кореляції

$$\rho = 1 - \frac{6 \cdot \sum d^2}{n(n^2 - 1)}$$

$$\rho = 1 - \frac{6 \cdot 162}{8(64 - 1)} = -0,92$$

$$t = \frac{\rho}{m} = \frac{0,92}{0,15} = 6,1$$

де ρ - коефіцієнт рангової кореляції;
d- різниця рангів;
n- число порівнюваних пар.

Оцінка вірогідності коефіцієнта кореляції:

$$m_2 = \sqrt{\frac{1 - \rho^2}{n - 2}}$$

$$m_2 = \sqrt{\frac{1 - 0,92^2}{8 - 2}} = 0,15$$

$$t = \frac{\rho}{m} > 3, \text{ вірогідно}$$

Висновок: в генеральній сукупності з ймовірністю більш, ніж 99%, між частотою оперативних втручань та післяопераційною летальністю пацієнтів, що були пізно доставлені, існує зворотній сильний вірогідний зв'язок. Тобто, чим більше виконується операцій, тим менша післяопераційна летальність (напевне, збільшується досвід персоналу). Проте необхідно зауважити, що частота операцій відображає частоту повторення певної патології (наприклад, апендицит найчастіше зустрічається, при цьому реєструється найменша летальність).

n	P<0,05	P<0,01
5	0,94	-
6	0,85	-
7	0,78	0,94
8	0,72	0,88
9	0,68	0,83

$$t_{\text{факт}} > t_{\text{табл}}$$

Спрощений метод оцінки коефіцієнта рангової кореляції передбачає порівняння його з критичним табличним значенням. Критичні значення коефіцієнта кореляції рангів (ρ) Спірмена (за В. Ю. Урбахом)

Рекомендації щодо застосування *методу рангової кореляції (метод Спірмена)*

- коли немає необхідності в точному встановленні сили зв'язку, а досить орієнтовних даних
- коли ознаки представлені не тільки кількісними, а й атрибутивними значеннями
- коли ряди розподілу ознак мають відкриті варіанти (наприклад, стаж роботи до 1 року та ін.)

Завдання. Відомо, що у населення існує значна потреба у стаціонарній допомозі. Вона може бути реалізована за наявності достатнього числа ліжок. Методом рангової та лінійної кореляції визначити парний коефіцієнт кореляції між забезпеченістю населення ліжками, частотою госпіталізації та середньою тривалістю перебування хворих у стаціонарі. Розрахувати коефіцієнт регресії.

Тут щось зайве...
Знайди помилки!
ПЕРЕВІР СЕБЕ!



Типовий приклад 2. Визначення коефіцієнта лінійної кореляції (Пірсона)

При вивченні стану надання стаціонарної допомоги в районі Б. необхідно дослідити зв'язок між ознакою x - відстанню від сільських населених пунктів району до ЦРЛ (км) та ознакою y - кількістю звернень по медичну допомогу до ЦРЛ протягом року на 1 жителя у 6-ти населених пунктах.

Населений пункт	x	y	d_x	d_y	d_x^2	d_y^2	$d_x \times d_y$
Д.	25	5	-4	-1	16	1	5
З.	20	6	-9	0	81	0	0
К.	10	11	-19	5	361	25	-95
В.	30	7	1	1	1	1	1
М.	40	4	11	-2	121	4	-22
Т.	50	4	21	-2	441	4	-42
	$\bar{X}_x = 29,1$	$\bar{X}_y = 6,1$			$\sum d_x^2 = 1021$	$\sum d_y^2 = 30$	$\sum d_x \times d_y = -154$

коефіцієнт кореляції:

$$r_{xy} = \frac{\sum (d_x \times d_y)}{\sqrt{\sum d_x^2 \times \sum d_y^2}} \quad r_{xy} = \frac{\sum -154}{\sqrt{\sum 30 \times 1021}} = \frac{-154}{175} = -0,88$$

Оцінка вірогідності коефіцієнта кореляції:

$$m_{r_{xy}} = \sqrt{\frac{1 - r_{xy}^2}{n - 1}} \quad m_{r_{xy}} = \sqrt{\frac{1 - (-0,88)^2}{6 - 1}} = 0,6 \quad t = \frac{r_{xy}}{m_r} \quad t = \frac{0,88}{0,6} = 1,4 \quad t = \frac{r}{m} > 3, \text{ вірогідно}$$

Висновок: між відстанню населеного пункту та кількістю відвідувань жителями ЦРЛ існує **зворотній сильний зв'язок, але невірогідний зв'язок** (через недостатню кількість спостережень).

Рекомендації до застосування методу Пірсона:

- коли потрібне точне встановлення сили зв'язку між ознаками
- коли ознаки мають лише кількісне вираження

Умови розрахунку коефіцієнта кореляції Пірсона (r):

- досліджувані змінні X і Y повинні бути розподілені нормально
- вимір зв'язку можливо тільки в якісно однорідних сукупностях (наприклад, вимір зв'язку між зростом і вагою в групах, однорідних за статтю та віком)
- для обчислення коефіцієнта кореляції використовуються незгруповані варіаційні ряди (ця вимога застосовується тільки при обчисленні коефіцієнта кореляції за методом квадратів)
- кількість значень по кожній з досліджуваних змінних X і Y має бути однаковим
- число спостережень має бути не менше 30

Типовий приклад 3. Визначення коефіцієнта регресії

Відомо, що із підвищенням температури тіла прискорюється частота пульсу. У хворих на пневмонію ($n=18$) вивчали взаємозв'язок між цими двома ознаками: коефіцієнт лінійної кореляції становив 0,9 ($t=4,6$), тобто зв'язок прямий, сильний, вірогідний.

Наявність вихідних даних дає можливість розрахувати: середню арифметичну, середнє квадратичне відхилення для кожної з ознак, і потім коефіцієнт регресії R_{xy}

$$R_{xy} = r_{xy} \cdot \frac{\delta_x}{\delta_y} \quad R_{xy} = 0,9 \cdot \frac{4,6}{0,43} = 9,2$$

Коефіцієнт регресії показує, на яку величину в середньому зміниться другий параметр при зміні першого на певну одиницю виміру.

Висновок: При підвищенні температури тіла на 1° частота пульсу прискорюється на 9,2 уд/хв.

Логічна структура теми 9. АНАЛІЗ ВЗАЄМОЗВ'ЯЗКУ МІЖ ДОСЛІДЖУВАНИМИ ПАРАМЕТРАМИ СТАТИСТИЧНИХ СУКУПНОСТЕЙ

