

Курс: 2

Навчальна дисципліна: «Математичні методи в психології»

Викладач: проф.Олефір В.О.

Практична робота 5.

Тема: Одновимірна лінійна регресія

Мета роботи: визначення коефіцієнтів лінійної регресії.

Коротка теоретична довідка. Статистичні зв'язки між змінними досліджуються не тільки методами кореляційного, але й регресійного аналізу, які доповнюють один одного. Основне завдання кореляційного аналізу - визначення зв'язку між випадковими змінними та оцінки її інтенсивності і напрямку. Основне завдання регресійного аналізу – встановлення форми і вивчення залежності змінних.

Регресія дозволяє за величиною однієї ознаки (змінна X) знаходити середні (очікувані) значення іншої ознаки (змінна Y), пов'язаної з X кореляційно. Оскільки в дослідженнях конкретний вид взаємозв'язків невідомий, одна з головних задач регресійного аналізу полягає в підборі відповідного виразу $Y = f(X)$, графік якого проходить через емпіричні точки (або досить близько до них) і таким чином пов'язує змінні X і Y .

Вираз $Y = f(X)$ називається рівняння регресії функція $f(x)$ – функція регресії, а їхні графіки - лінії регресії. Регресійний аналіз виявляє кількісну залежність ознаки-фактору (залежною змінною) від одного або декількох ознак-факторів (незалежної змінної). Ця залежність може бути одновимірною або багатовимірною, як лінійною, так і нелінійною.

Одновимірна лінійна регресія передбачає тільки дві змінні, наприклад, незалежну X і залежну Y , а також рівняння лінійного типу $\hat{Y} = a_0 + a_1 \cdot X$. Лінійна регресія дає можливість виявляти, наскільки змінюється середня величина однієї ознаки при зміні іншого. Побудова лінійної регресії полягає в розрахунках коефіцієнтів лінійної регресії a_0 і a_1 :

Формули розрахунку коефіцієнтів лінійної регресії a_0 і a_1 :

$$a_1 = \frac{\sum(x_i - \bar{X}) \cdot (y_i - \bar{Y})}{\sum(x_i - \bar{X})^2} ;$$

$$a_0 = \bar{Y} - a_1 \cdot \bar{X} .$$

де $\bar{Y}$ і $\bar{X}$ – середні значення змінних Y і X . Вибір значень коефіцієнтів a_0 і a_1 виконується за методом «найменших квадратів» так, щоб сума

$$\sum (y_i - \hat{Y})^2 = \sum (y_i - a_0 - a_1 \cdot \bar{X})^2 \text{ була мінімальною.}$$

Якщо незалежною ознакою виступає Y а залежною – X , то рівняння лінійної регресії матиме інший вигляд типу $\hat{X} = b_0 + b_1 \cdot Y$

Коефіцієнти лінійної регресії b_0 і b_1 відрізнятися від коефіцієнтів a_0 і a_1 .

Приклад 5.1. Оцінити залежність успішності навчання (Y) від витраченого часу (X). Емпіричні дані представлені в таблиці 5.1.

○ Таблиця 5.1

Емпіричні дані			Розрахунки					Регресія	
i	x_i	y_i	$(x_i - \bar{X})$	$(y_i - \bar{Y})$	$(x_i - \bar{X})(y_i - \bar{Y})$	$(x_i - \bar{X})^2$	$(y_i - \bar{Y})^2$	Y	X
1	2,1	5	-0,29	0,91	-0,26	0,08	0,83	3,69	2,77
2	3,4	5	1,01	0,91	0,92	1,02	0,83	5,47	2,77
3	1,5	3	-0,89	-1,09	0,97	0,79	1,19	2,87	1,93
4	2,9	6	0,51	1,91	0,97	0,26	3,64	4,79	3,19
5	2,5	3	0,11	-1,09	-0,12	0,01	1,19	4,24	1,93
6	1,4	3	-0,99	-1,09	1,08	0,98	1,19	2,73	1,93
7	2,5	5	0,11	0,91	0,10	0,01	0,83	4,24	2,77
8	2,2	3	-0,19	-1,09	0,21	0,04	1,19	3,83	1,93
9	3	5	0,61	0,91	0,55	0,37	0,83	4,93	2,77
10	3,3	5	0,91	0,91	0,83	0,83	0,83	5,34	2,77
11	1,5	2	-0,89	-2,09	1,86	0,79	4,37	2,87	1,51
Суми					7,11	5,19	19,91		
Середні									
		r_{xy}	0,76						
		a_1	1,37			b_1	0,42		
		a_0	0,82			b_0	0,67		

Послідовність рішення:

Виконати розрахунки коефіцієнтів регресії a_0 і a_1 :

- розрахувати середні значення масивів $X = 2,39$ і $Y = 4,09$;
- розрахувати різниці, твори і квадрати різниць за допомогою відповідних формул, показано в табл. 5.1;
- розрахувати суми добутоків і квадратів різниць;

- розрахувати коефіцієнти лінійної регресії a_0 і a_1 за допомогою вище наведених формул: $a_1 = 7,11 / 5,19 = 1,37$ і $a_0 = 4,09 - 1,37 \times 2,39 = 0,82$;

Висновки:

Рівняння регресії $Y = 0,82 + 1,31 \times X$, а також $X = 0,67 + 0,42 \times Y$ дають можливість аналітичного прогнозування значень залежної змінної за допомогою незалежної змінної. Отримані регресійні рівняння мають різні коефіцієнти регресії і виконують різні прогнозуючі функції: перше прогнозує Y за значеннями X , друге – навпаки, X за значеннями Y (звичайно, якщо таке прогнозування має сенс).

Завдання: за емпіричними даними оцінити залежність параметра (X) від параметра (Y).

i	1	2	3	4	5	6	7	8	9	10	11	12
X	2	8	12	3	1	6	7	10	4	9	11	5
Y	6	5	10	7	3	4	9	8	1	11	12	2

Результати оформити згідно вищенаведеному алгоритму та надіслати на електронну пошту vaolefir@gmail.com

Термін: 20.03.2020

Практична робота 6.

Тема Критерій t-Ст'юдента для незалежних вибірок

Мета роботи: визначення t-Ст'юдента для незалежних вибірок.

Коротка теоретична довідка. Метод дозволяє перевірити гіпотезу про те, що середні значення двох генеральних сукупностей, з яких вилучені порівнювані незалежні вибірки, відрізняються один від одного. Допущення незалежності передбачає, що представники двох вибірок не є парами ознак, що взаємно корелюють.

Статистична гіпотеза, що перевіряється $H_0: M_1 = M_2$.

При її відхиленні приймається альтернативна гіпотеза про те, що M_1 більше (менше) M_2 .

Вихідні припущення для статистичної перевірки:

- о одна вибірка витягується з однієї генеральної сукупності, а інша вибірка, незалежна від першої, витягується з іншої генеральної сукупності;
- о розподіл досліджуваної ознаки i в тій, i в іншій вибірці приблизно відповідає нормальному;
- о дисперсії ознаки у двох вибірках приблизно однакові (гомогенні).

Структура вихідних даних: досліджувану ознаку виміряно у випробовуваних, кожен з яких належить до однієї з двох порівнюваних незалежних вибірок.

Обмеження: розподіли ознаки i в тій, i в іншій вибірці суттєво не відрізняються від нормального; в разі різної чисельності порівнюваних вибірок їх дисперсії статистично достовірно не розрізняються (перевіряється за критерієм F-Фішера – при обчисленнях «вручну», за критерієм Лівена – при обчисленнях на комп'ютері).

Формули для розрахунку емпіричного значення критерію t-Ст'юдента:

$$t_3 = \frac{|M_1 - M_2|}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

або:

$$t_3 = \frac{|M_1 - M_2|}{\sqrt{\frac{(N_1 - 1)\sigma_1^2 + (N_2 - 1)\sigma_2^2}{N_1 + N_2 - 2} \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}} \quad (2)$$

де M_1 та M_2 – середні значення вибірок за досліджуваною ознакою; σ_1^2 та σ_2^2 – дисперсії вибірок; N_1 та N_2 – обсяги вибірок.

Приклад

Припустимо, вивчалоя відмінність в інтелекті студентів 1-го і 5-го курсів. Для цього випадковим чином були відібрані 30 студентів 1 курсу і 28 студентів 5 курсу, у яких інтелект визначався по одній і тій же методиці (наприклад, Амтхауера). Були отримані наступні результати:

Група 1: 1-й курс	Група 2: 5-й курс
$N_1 = 30$	$N_2 = 28$
$M_1 = 103$	$M_2 = 109$
$\sigma_1 = 10$	$\sigma_2 = 12$

Гіпотеза про відмінності інтелекту перевірялася па рівні $\alpha = 0,05$.

Крок 1. Обчислюємо емпіричне значення критерію t-Стюдента за формулою 1: $t_e = 2,06$ (за формулою 2: $t_e = 2,17$); $df = 56$.

Крок 2. Визначаємо по таблиці критичних значень критерію t-Стюдента p -рівень значимості. Для $df = 56$ емпіричне значення знаходиться між критичними для $p = 0,05$ і $p = 0,01$. Отже, $p < 0,05$.

Крок 3. Приймаємо статистичне рішення і формулюємо висновок. Статистична гіпотеза про рівність середніх значень відхиляється. **Висновок:** інтелект студентів 5 курсу статистично достовірно вище, ніж у студентів 1 курсу ($p < 0,05$).

Завдання 6.1

Умова задачі:

30 студентів (15 юнаків і 15 дівчат) були обстежені на рівень нейротизму по тесту Айзенка. Отримані наступні результати:

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ю	10	12	5	9	6	7	11	8	7	4	12	8	7	9	8
Д	5	9	9	13	8	8	10	7	13	11	11	13	10	16	11

Завдання:

Визначити достовірність відмінностей між юнаками та дівчатами за рівнем їх нейротизму.

Результати оформити згідно вищенаведеному алгоритму та надіслати на електронну пошту vaolefir@gmail.com

Контрольна дата: 27.03.2020