УДК: 371.26

Павло ВОЛОВИК
м.Київ

Педагогічна технологія застосування регресійного і кореляційного аналізів в педагогічних дослідженнях

(продовження статті, початок в журналі «Неперервна професійна освіта: теорія і практика», Вип. I-II. – 2007. – С.13-24)

У статті аналізуються проблеми кореляції педагогічних явищ і статистичні оцінки зв'язків педагогічних показників. Сформульовані завдання і способи організації досліджень, які дають можливість встановлювати кореляційні зв'язки між змінними, що характеризують педагогічні явища і процеси. Дається класифікація зв'язків за їх формою (лінійні, криволінійні), розглядаються відмінності між ними, розкривається методика виявлення лінійних і криволінійних зв'язків та методи побудови кореляційних рівнянь.

У праці висвітлюються технології дослідження: тісноти взаємозв'язку між якісними ознаками педагогічних явищ за допомогою коефіцієнта асоціації та коефіцієнта взаємної спряженості; динамічної тісноти зв'язку між двома взаємозв'язаними варіаційними рядами за допомогою лінійного коефіцієнта кореляції, а також викладена методика визначення надійності вибіркових лінійних коефіцієнтів кореляції.

1. Кореляційний зв'язок.

Лінійна кореляція. Рівняння зв'язку, розглянуті раніше, стосувалися тих випадків, коли кожному значенню незалежної ознаки x відповідало єдине значення залежної ознаки y. Такі залежності, як відомо, називаються функціональними.

Одним з основних завдань наукових досліджень якраз і є встановлення зв'язку між величинами або факторами, зміна яких визначає суть досліджуваного процесу. Проте в багатьох випадках зв'язки в навколишньому світі настільки багатогранні й складні, що їх не можна описати за допомогою строгої функціональної залежності. Зокрема, в статистиці часто трапляються такі залежності, коли одному значенню незалежної ознаки x відповідає не одне значення ознаки y, а сукупність значень y або розподіл залежної y. Такі залежності називаються стохастичними залежностями або стохастичними зв'язками.

Проілюструємо сказане прикладом. Після пояснення нової теми з алгебри розв'язуємо з учнями вправи: з першою групою – 1 вправу, з другою – 2 вправи, з третьою – 3, з четвертою – 4, з п'ятою – 5, з шостою – 6 вправ. У кожній групі по 10 осіб, за підготовкою групи підібрані рівнозначні. Після розв'язування відповідно кількості вправ даємо учням контрольну роботу. Результати контрольної роботи занесені в табл. 1.

Таблиця 1

Кількість виконаних вправ x / **Оцінка за контрольну роботу y**	**1**	**2**	**3**	**4**	**5**	**6**	n_x
«1»	1						1
«2»	2	2	2				6
«3»	5	5	3	4	3	3	23
«4»	1	2	3	5	5	6	22
«5»	1	1	2	1	2	1	8
m_x	10	10	10	10	10	10	60

Аналіз таблиці показує, що між кількістю виконаних вправ і результатами контрольних робіт (засвоєння теми) існує певний зв'язок. Встановимо цей зв'язок. Для цього знайдемо середні бали за виконані учнями контрольні роботи після розв'язання відповідної кількості вправ:

$$\bar{y}_1 = \frac{1\cdot 1 + 2\cdot 2 + 3\cdot 5 + 4\cdot 1 + 5\cdot 1}{10} = 2{,}9;$$

$$\bar{y}_1 = \frac{2\cdot 2 + 3\cdot 5 + 4\cdot 2 + 5\cdot 1}{10} = 3{,}2;$$

$$\bar{y}_3 = \frac{2\cdot 2 + 3\cdot 3 + 4\cdot 3 + 5\cdot 2}{10} = 3{,}5;$$

$$\bar{y}_4 = \frac{3\cdot 4 + 4\cdot 5 + 5\cdot 1}{10} = 3{,}7;$$

$$\bar{y}_5 = \frac{3\cdot 3 + 4\cdot 5 + 5\cdot 2}{10} = 3{,}9;$$

$$\bar{y}_2 = \frac{3\cdot 3 + 4\cdot 6 + 5\cdot 2}{10} = 3{,}8.$$

Величини $\bar{y}_1$; $\bar{y}_2$; …; $\bar{y}_6$ називаються частковими середніми або умовними середніми. Якщо позначити умовні середні бали $\bar{y}_x$, які відповідають кількості виконаних вправ *x*, то можна їх звести в таку таблицю:

Таблиця 2

x	1	2	3	4	5	6
$\bar{y}_x$	2,9	3,2	3,5	3,7	3,9	3,8

З таблиці видно, що результати контрольних робіт кращі в тих учнів, які перед контрольними роботами виконали більше вправ, проте це зростання небезмежне. Учні, які виконали значну кількість однотипних вправ, у контрольних роботах, виконуючи правильно основні дії, допускають механічні помилки в допоміжних діях.

Якщо побудувати на площині точки з координатами (1; 2,9), (2; 3,2), …, (6; 3,8), то побачимо, що вони лежать майже на одній прямій (рис.1). Тому ламану лінію можна спрямити за допомогою прямої. Рівняння цієї лінії можна взяти за залежність, яка добре відображатиме залежність між кількістю виконаних вправ перед контрольною робою і середнім балом, здобутим учнями за контрольну роботу. Для визначення коефіцієнтів лінійної залежності $\overline{y}_x = a + bx$ застосуємо метод найменших квадратів. Складемо спочатку таблицю, за допомогою якої побудуємо нормальну систему Гаусса, з якої і визначимо коефіцієнти a і b.

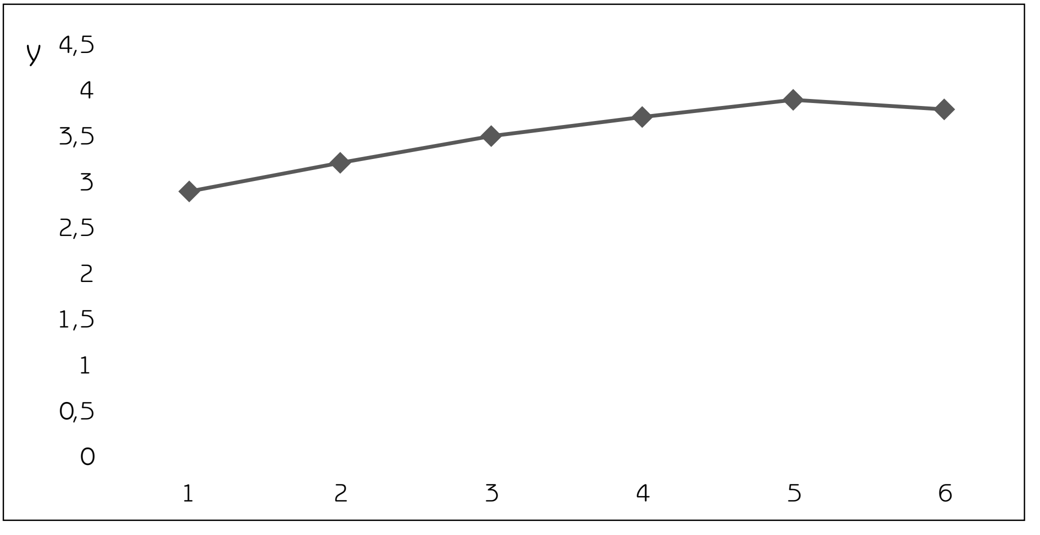

Рис.1

Таблиця 3

x	$\overline{y}_x$	x^2	$x\overline{y}_x$
1	2,9	1	2,9
2	3,2	4	6,4
3	3,5	9	10,5
4	3,7	16	14,8
5	3,9	25	19,5
6	3,8	36	22,8
$\sum x = 21$	$\sum \overline{y}_x = 21{,}0$	$\sum x^2 = 91$	$\sum x\overline{y}_x = 76{,}9$

Підставивши знайдені значення в систему нормальних рівнянь Гаусса:

$$\left.\begin{array}{l} na + b\sum x = \sum \overline{y}_x; \\ a\sum x + b\sum x^2 = \sum x\overline{y}_x \end{array}\right\},$$

дістанемо

$$\left.\begin{array}{l} 6a + 21b = 21; \\ 21a + 91b = 76{,}9 \end{array}\right\},$$

звідки

$$a = \frac{\begin{vmatrix} 21{,}0 & 21 \\ 76{,}9 & 91 \end{vmatrix}}{\begin{vmatrix} 6 & 21 \\ 21 & 91 \end{vmatrix}} = \frac{21 \cdot 91 - 76{,}9 \cdot 21}{6 \cdot 91 - 21 \cdot 21} = 2{,}82;$$

$$b = \frac{\begin{vmatrix} 6 & 21 \\ 21 & 76{,}9 \end{vmatrix}}{105} = \frac{6 \cdot 76{,}9 - 21 \cdot 21}{105} = 0{,}194.$$

Отже, рівняння, яке виражає залежність між кількістю виконаних вправ і середнім балом, одержаним учням за виконання контрольної роботи, матиме вигляд:

$$\overline{y}_x = 2{,}82 + 0{,}194x \quad (1).$$

Рівняння (1) називається кореляційним рівнянням або рівнянням регресії *y* на *x*.

Якщо із збільшенням незалежної ознаки, або фактор-ознаки, збільшується також в середньому значення залежної або результативної ознаки $\overline{y}_x$, то зв’язок між x і $\overline{y}_x$ називається прямим; якщо ж значення $\overline{y}_x$ зменшується, то зв’язок називається оберненим. У нашому прикладі ми мали справу з прямим зв’язком.

Рівняння регресії *y* на *x* і *x* на *y* можна записати в дещо іншій формі. Щоб знайти величини *a* і *b* в рівнянні регресії $\overline{y}_x = a + bx$, складемо нормальну систему рівнянь Гаусса:

$$\left.\begin{array}{l} na + b\sum x = \sum y; \\ a\sum x + b\sum x^2 = \sum xy \end{array}\right\}.$$

Розв’язавши цю систему, дістанемо:

$$a = \frac{\begin{vmatrix} \sum y & \sum x \\ \sum xy & \sum x^2 \end{vmatrix}}{\begin{vmatrix} n & \sum x \\ \sum x & \sum x^2 \end{vmatrix}} = \frac{\sum x^2 \sum y - \sum x \sum xy}{n\sum x^2 - \left(\sum x\right)^2}.$$

Поділивши чисельник і знаменник на n^2, дістанемо:

$$a=\frac{\frac{\sum x^2}{n}\cdot\frac{\sum y}{n}-\frac{\sum x}{n}\cdot\frac{\sum xy}{n}}{\frac{\sum x^2}{n}-\left(\frac{\sum x}{n}\right)^2}=\frac{\overline{x^2}\,\overline{y}-\overline{x}\cdot\overline{xy}}{\overline{x^2}-\left(\overline{x}\right)^2}=\frac{\overline{x^2}\,\overline{y}-\overline{x}\cdot\overline{xy}}{\sigma_x^2};$$

$$a=\frac{\begin{vmatrix} n & \sum y \\ \sum x & \sum xy \end{vmatrix}}{\begin{vmatrix} n & \sum x \\ \sum x & \sum x^2 \end{vmatrix}}=\frac{n\sum xy-\sum x\cdot\sum y}{n\sum x^2-\left(\sum x\right)^2}=\frac{\frac{\sum xy}{n}-\frac{\sum x}{n}\cdot\frac{\sum y}{n}}{\frac{\sum x^2}{n}-\left(\frac{\sum x}{n}\right)^2}=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\overline{x^2}-\left(\overline{x}\right)^2}=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\sigma_x^2}.$$

Величину b= $\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\sigma_x^2}$, яка є кутовим коефіцієнтом прямої регресії, позначають ρ_{yx} і називають коефіцієнтом регресії *y* на *x*.

Якщо розглядати рівняння регресії $\overline{x}_y=\alpha+\beta y$, то для кутового коефіцієнта β знайдемо: $\beta=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\sigma_y^2}$.

Величину $\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\sigma_y^2}$ позначають через ρ_{xy} і називають коефіцієнтом регресії *x* на *y*.

Ввівши ці позначення, рівняння регресії можна записати так:

$$\overline{y}_x=a+\rho_{yx}\cdot x, \quad (2)$$

$$\overline{x}_y=\alpha+\rho_{xy}\cdot y. \quad (3)$$

Рівняння зв'язку є методом узагальнення статистичних зв'язків, які спостерігаються в дослідах, і методом їх вивчення.

Лінія регресії, побудована за рівнянням регресії, розміщується так, що сума вертикальних відхилень від неї дорівнює нулю, тобто вона збалансовує (мінімізує) суму додатніх і від'ємних відхилень. Маючи на увазі ці властивості, її називають лінією найкращого наближення. Згідно з критерієм найменших квадратів вона наближає розсіяння краще, ніж будь-яка інша пряма.

Застосовуючи ту чи іншу функцію як рівняння зв'язку, розрізняють зв'язки за їх формою: лінійний зв'язок і криволінійний (параболічний, гіперболічний та ін.).

Рівняння зв'язку можуть бути для залежностей від однієї ознаки при різних формах зв'язку (лінійному, криволінійному) і для множинного зв'язку.

Вище ми розглянули лінійний зв'язок.

Розглянемо тепер криволінійну кореляцію. Часто кореляційний зв'язок між незалежними ознаками *x* і середнім значенням залежної ознаки $\overline{y}_x$ не можна задати за допомогою лінійного рівняння $\overline{y}_x = a + bx$. Якщо на площині побудувати точки з координатами $(x; \overline{y}_x)$, де $\overline{y}_x$ – умовні середні, і якщо вони розмістяться близько від вдало проведеної кривої, то рівняння цієї кривої $\overline{y}_x = f(x)$ і виражатиме кореляційну залежність між *x* і $\overline{y}_x$. Кореляційний зв'язок при цьому називається криволінійним.

Побудова рівнянь регресії криволінійного кореляційного зв'язку не відрізняється від побудови рівнянь регресії лінійного кореляційного зв'язку.

Спочатку за заданою кореляційною таблицею знаходимо умовні середні $\overline{y}_x$ або $(\overline{x}_y)$, а потім на основі простої таблиці, в якій кожному значенню *x(y)* відповідає значення $\overline{y}_x(\overline{x}_y)$, складаємо систему нормальних рівнянь, з якої визначаємо коефіцієнти рівняння регресії.

Слід зазначити, що рівняння регресії є по суті емпіричними формулами, але встановлюють зв'язок між фактором-ознакою і умовним середнім значенням результативної ознаки.

2. Емпіричні міри тісноти зв'язку.

Емпіричні міри тісноти зв'язку дають змогу оцінити ступінь зв'язку між явищами або факторами, які залежать один від одного.

До емпіричних мір тісноти зв'язку відносять: коефіцієнти асоціації, коефіцієнт взаємної спряженості та ін.

Розглянемо, в чому полягає їх суть та як їх обчислювати.

1. Коефіцієнт асоціації.

Як міра тісноти зв'язку коефіцієнт асоціації застосовується для вивчення зв'язку двох ознак, що складаються лише в двох груп. Щоб обчислити його, треба побудувати чотирикліточну таблицю кореляції, яка виражає зв'язок між двома явищами, кожне з яких, в свою чергу, повинно складатися лише з двох видів, якісно відмінних один від одного.

Наприклад, щоб перевірити ефективність певної методики викладання теми, беремо 2 класи з однаковою кількістю рівнозначних учнів. В одному класі (контрольному) викладаємо дану тему за старою методикою, а в другому (експериментальному) – за новою, ефективність якої хочемо перевірити. Після вивчення даної теми проводимо контрольні роботи. Відповіді на запитання і розв'язані задачі оцінюємо так: «засвоїв», «не засвоїв».

Результати експерименту можна занести в таблицю (табл.4).

Міра тісноти зв'язку – коефіцієнт асоціації – обчислюється за формулою

$$A = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}. \quad (4)$$

Таблиця 4

Успішність	Класи		Всього учнів
	Експериментальний	Контрольний	
Засвоїли	a	b	a+b
Не засвоїли	c	d	c+d
Всього:	a+c	b+d	

Для прикладу обчислимо коефіцієнт асоціації за даними експериментальної перевірки методики викладання розділу «Оптика». Зокрема, розглянемо, як було перевірено засвоєння учнями питання про утворення дійсних зображень. Візьмемо лише 2 класи (всіх експериментальних класів було 28).

Таблиця 4а

Успішність	Класи		Всього учнів
	Експериментальний	Контрольний	
Засвоїли	28	21	49
Не засвоїли	4	11	15
Всього:	32	32	

Коефіцієнт асоціації буде:

$$A = \frac{28 \cdot 11 - 4 \cdot 21}{\sqrt{49 \cdot 15 \cdot 32 \cdot 32}} = \frac{308 - 84}{896} = \frac{224}{896} = 0{,}25.$$

Величина коефіцієнта асоціації достатня для того, щоб зробити висновок, що експериментальна методика ефективніша за ту, яка застосувалася у контрольному класі.

2. Коефіцієнта взаємної спряженості.

Тіснота зв’язку якісних ознак, кожна з яких складається більш як з двох груп, вимірюється коефіцієнтом взаємної спряженості.

Для розрахунку коефіцієнтів взаємної спряженості часто застосовують формулу К.Пірсона

$$C = \sqrt{\frac{\varphi^2}{1+\varphi^2}}, \quad (5)$$

де, φ^2 – показник взаємної спряженості.

Покажемо на конкретному прикладі, як обчислювати коефіцієнт взаємної спряженості.

Приклад. Обчислити коефіцієнт взаємної спряженості між кількістю виконаних вправ після проходження нової теми і оцінками за контрольні роботи з цієї теми (див. табл.5).

Таблиця 5

Кількість виконаних вправ *(x)* Оцінки за контрольні роботи *(y)*	1	2	3	4	5	6	n_x
«1»	1 (1) 0,1	–	–	–	–	–	1 __ – 0,1; 0,1
«2»	2 (4) 0,4	2 (4) 0,4	2 (4) 0,4	–	–	–	6 __ – 1,2; 0,2
«3»	5 (25) 2,5	5 (25) 2,5	3 (9) 0,9	4 (16) 1,6	3 (9) 0,9	3 (9) 0,9	23 __ – 9,3; 0,4
«4»	1 (1) 0,1	2 (4) 0,4	3 (9) 0,9	5 (25) 2,5	5 (25) 2,5	6 (36) 3,6	22 __ – 10; 0,46
«5»	1 (1) 0,1	1 (1) 0,1	2 (4) 0,4	1 (1) 0,1	2 (4) 0,4	1 (1) 0,1	8 __ – 1,2; 0,15
m_x	10	10	10	10	10	10	60; 1,31

Розв'язання. Складаємо таблицю, в якій у кожній клітці записуємо частоти, їх квадрати і квадрати частот, поділені на суму частот по колонці. В підсумкову колонку записуємо суми частот по рядку, суми результатів ділення квадратів частот на суму частот по колонці, а також результати ділення нижнього числа на верхнє. Сума цих чисел дорівнює показнику взаємної спряженості φ^2 плюс 1.

Отже, $\varphi^2 + 1$ в нашому випадку дорівнює 1,31, а $\varphi^2 = 1{,}31 - 1 = 0{,}31$.

Обчислюємо коефіцієнт взаємної спряженості за формулою Пірсона:

$$C_1 = \sqrt{\frac{0{,}31}{1{,}31}} = 0{,}49\ .$$

Величина коефіцієнта взаємної спряженості Пірсона показує, що існує залежність між кількістю вправ, розв'язаних після проходження нової теми, і результатами контрольних робіт з цієї теми.

Продовження статті в наступному номері журналу.

Стаття надійшла до редакції 15.0.5.08

Павел ВОЛОВИК

Педагогическая технология применения регрессионного и корреляционного анализов в педагогических исследованиях

Резюме

В статье анализируются проблемы корреляции педагогических явлений и статистическая оценка связей педагогических показателей. Сформулированы задачи и способы организации исследований, позволяющие установить корреляционные связи между переменными, характеризующими педагогические явления и процессы. Дается классификация связей по их форме (линейные, криволинейные), рассматриваются различия между ними и методика построения уравнений линейной и криволинейной корреляционных связей.

В работе излагается технология исследования: тесноты взаимосвязи между качественными признаками педагогических явлений при помощи коэффициента ассоциации и коэффициента взаимной сопряженности; динамической тесноты связи между двумя взаимосвязанными вариационными рядами при помощи линейного коэффициента корреляции, а также излагается методика определения надежности выборочных линейных коэффициентов корреляции.

Pavel VOLOVIK

Pedagogical technology of regression and correlation analyses in pedagogical research.

Summary

The paper deals with the analysis of pedagogical phenomena correlation and statistic estimation of pedagogical indexes, which allow to make correlation connections between variables characterizing pedagogical phenomena and processes. The connections classification due to the form (linear, non-linear), the differences between them and technique for construction of relations of linear and non-linear correlation connections are presented.