Міністерство освіти та науки України
Національна металургійна академія України
Національний педагогічний університет
імені М.П. Драгоманова
Державний інститут підготовки
та перепідготовки кадрів промисловості
Інститут інформаційних технологій
і засобів навчання АПН України

Теорія та методика навчання фундаментальних дисциплін у вищій школі

Збірник наукових праць
Випуск VI

Кривий Ріг
Видавничий відділ НМетАУ
2010

ДЕЯКІ ЗАСТОСУВАННЯ СТАТИСТИЧНИХ МЕТОДІВ ДО ПЕДАГОГІЧНИХ ДОСЛІДЖЕНЬ

В.П. Ржепецький

м. Кривий Ріг, Криворізький державний педагогічний університет

1. *Встановлення кореляції між результатами сумісних вимірювань*

Припустимо, що для знаходження функції $q(x, y)$ ми вимірюємо дві величини x та y декілька разів і одержуємо n пар даних (x_1, y_1), …, (x_n, y_n). За результатами вимірювань обчислюємо $\bar{x}$, $\bar{y}$ та дисперсії σ_x^2 і σ_y^2:

$$\sigma_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}; \qquad \sigma_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}. \tag{1}$$

Для n пар значень x та y можна знайти n значень величини q:

$$q_i = (x_i, y_i), \text{ де } i = 1, 2, \ldots, n.$$

Будемо вважати, що всі похибки малі, тобто всі числа x_1, ..., x_n близькі до $\bar{x}$, а всі числа y_1, ..., y_n – до $\bar{y}$. В цьому випадку справедливе наближення:

$$q_i = q(x_i, y_i) \approx q(\bar{x}, \bar{y}) + \frac{\partial q}{\partial x}(x_i - \bar{x}) + \frac{\partial q}{\partial y}(y_i - \bar{y}). \tag{2}$$

В рамках цього наближення середнє набуває виду:

$$\bar{q} = \frac{1}{n}\sum_{i=1}^{n} q_i = \frac{1}{n}\sum_{i=1}^{n}\left[q(\bar{x}, \bar{y}) + \frac{\partial q}{\partial x}(x_i - \bar{x}) + \frac{\partial q}{\partial y}(y_i - \bar{y})\right],$$

з якого випливає, що середнє значення функції q дорівнює [2, 182]:

$$\bar{q} = q(\bar{x}, \bar{y}). \tag{3}$$

Дисперсія для n значень q_1, ..., q_n визначається виразом:

$$\sigma_q^2 = \frac{\sum (q_i - \bar{q})^2}{n-1}.$$

Підставивши в цей вираз (2) і (3), одержимо:

$$\sigma_q^2 = \frac{1}{n-1}\sum\left[\frac{\partial q}{\partial x}(x_i - \bar{x}) + \frac{\partial q}{\partial x}(y_i - \bar{y})\right]^2 =$$

$$= \left(\frac{\partial q}{\partial x}\right)^2 \frac{\sum (x_i - \bar{x})^2}{n-1} + \left(\frac{\partial q}{\partial y}\right)^2 \frac{\sum (y_i - \bar{y})^2}{n-1} + 2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}.$$

Суми в перших двох доданках – це дисперсії σ_x^2 і σ_y^2 (див. (1)); сума в третьому доданку – це так званий ***змішаний другий момент*** змінних x та y або ***коваріація***:

$$\sigma_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}. \tag{4}$$

З врахуванням цих позначень вираз для дисперсії σ_q^2 матиме вид:

$$\sigma_q^2 = \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^2 + 2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\sigma_{xy}. \tag{5}$$

Якщо значення x та y ***незалежні***, то після багатьох вимірювань σ_{xy} буде наближатись до нуля і ми приходимо до відомого виразу для незалежних і випадкових похибок:

$$\sigma_q^2 = \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^2. \tag{6}$$

Якщо вимірювання x та y ***не є незалежними,*** то змішаний другий момент σ_{xy} не повинен дорівнювати нулю. Якщо σ_{xy} не дорівнює нулю, то говорять, що похибки в x та y ***мають кореляцію,*** і похибка σ_q повинна обчислюватись за виразом (5), а не (6). Проте можна показати, що і при наявності кореляції в похибках σ_q дорівнюватиме:

$$\sigma_q \le \left|\frac{\partial q}{\partial x}\right| \cdot \sigma_x + \left|\frac{\partial q}{\partial y}\right| \cdot \sigma_y, \tag{7}$$

тобто (7) є верхньою межею похибки в будь-якому випадку.

Змішаний другий момент дозволяє відповісти на питання, наскільки добре набір результатів вимірювань підтверджує гіпотезу про наявність лінійної залежності між x та y.

Нехай ми маємо n пар значень (x_1, y_1), ..., (x_n, y_n) двох змінних, які, як ми вважаємо, повинні бути зв'язані лінійною залежністю $y=ax+b$. Підкреслимо, що x_i (як і y_i) в даному випадку − це не результати вимірювання однієї й тієї ж величини; це результати вимірювань n ***різних*** значень однієї змінної (наприклад, n різних висот, з яких кидали кульку, оцінки студентів за контрольну роботу і т.ін.). Ступінь підтвердження лінійної залежності між величинами x та y визначається ***коефіцієнтом лінійної кореляції*** (або просто ***коефіцієнтом кореляції***):

$$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}, \tag{8}$$

де змішаний другий момент σ_{xy} і стандартні відхилення σ_x і σ_y визначаються формулами (1) і (4). Підставимо ці вирази у (8):

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}. \qquad (9)$$

Коефіцієнт кореляції може мати значення між +1 і –1. Якщо r_{xy} близький до ±1, то точки лежать близько до деякої прямої; якщо r_{xy} близький до 0, то точки не корельовані і не групуються поблизу прямої.

Кількісним критерієм значущості r_{xy} є імовірність того, що n вимірювань двох змінних x та y, кореляція між якими відсутня, дадуть значення r, не менше, ніж будь-яке задане значення r_0: $P_n(|r| \geq |r_0|)$. Значення цих ймовірностей наведено в табл. 1. ***Значущою*** вважатиметься кореляція r_0, якщо одержати коефіцієнт $|r| \geq |r_0|$ для некорельованих змінних менша 5%. (При P_n < 1% кореляція ***високозначуща***).

Таблиця 1.

Подана у процентах імовірність $P_n(|r| \geq |r_0|)$ того, що результати n вимірювань двох некорельованих змінних дадуть коефіцієнт кореляції $|r| \geq |r_0|$, як функція n і r_0. (Прочерк вказує на імовірність, яка менша, ніж 0,05%)

N	r_0										
	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
3	100	94	87	81	74	67	59	51	41	29	0
4	100	90	80	70	60	50	40	30	20	10	0
5	100	87	75	62	50	39	28	19	10	3,7	0
6	100	85	70	56	43	31	21	12	5,6	1,4	0
7	100	83	67	51	37	25	15	8,0	3,1	0,6	0
8	100	81	63	47	33	21	12	5,3	1,7	0,2	0
9	100	80	61	43	29	17	8,8	3,6	1,0	0,1	0
10	100	78	58	40	25	14	6,7	2,4	0,5	-	0
11	100	77	56	37	22	12	5,1	1,6	0,3	-	0
12	100	76	53	34	20	9,8	3,9	1,1	0,2	-	0
13	100	75	51	32	18	8,2	3,0	0,8	0,1	-	0
14	100	73	49	30	16	6,9	2,3	0,5	0,1	-	0
15	100	72	47	28	14	5,8	1,8	0,4	-	-	0
16	100	71	46	26	12	4,9	1,4	0,3	-	-	0
17	100	70	44	24	11	4,1	1,1	0,2	-	-	0
18	100	69	43	23	10	3,5	0,8	0,1	-	-	0
19	100	68	41	21	9,0	2,9	0,7	0,1	-	-	0
20	100	67	40	20	8,1	2,5	0,5	0,1	-	-	0
25	100	63	34	15	4,8	1,1	0,2	-	-	-	0
30	100	60	29	11	2,9	0,5	-	-	-	-	0
35	100	57	25	8,0	1,7	0,2	-	-	-	-	0

N	r_0										
	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
40	100	54	22	6,0	1,1	0,1	-	-	-	-	0
45	100	51	19	4,5	0,6	-	-	-	-	-	0
50	100	73	49	30	16	8,0	3,4	1,3	0,4	0,1	
60	100	70	45	25	13	5,4	2,0	0,6	0,2	-	
70	100	68	41	22	9,7	3,7	1,2	0,3	0,1	-	
80	100	66	38	18	7,5	2,5	0,7	0,1	-	-	
90	100	64	35	16	5,9	1,7	0,4	0,1	-	-	
100	100	62	32	14	4,6	1,2	0,2	-	-	-	

Так, якщо для m=10 одержано r_0=0,8, то $P_n(|r| \geq 8) = 0,5\%$, тобто ймовірність того, що між значеннями існує лінійна залежність, висока.

Наведемо приклад знаходження кореляції між оцінками за контрольну роботу з механіки і пропусками занять. Розв'яжемо цю задачу на прикладі потоку ФІ-06 (2007 р.) Використаємо електронні таблиці Excel.

Всього студентів на курсі 38. В комірки A1 – A38 заносимо число пропусків занять кожним зі студентів, в комірки B1 – B38 – їхні оцінки за контрольну роботу з механіки (максимальна оцінка – 8 балів). В комірку C1 записуємо знак =. Наводимо курсор на знак f_x («*Вставка функции*») і натискаємо ліву кнопку миші. З'являється діалогове вікно «*Мастер функций*». Наводимо курсор на кнопку у вікні «*Категория*», натискаємо ліву кнопку миші: випадає список функцій, в якому вибираємо (наводимо курсор) «*Статистические*». У вікні „*Выберите функцию*" знаходимо «КОРРЕЛ», натискаємо ліву кнопку, потім – «ОК». На екрані з'явиться вікно «*Аргументы функции*». У масив 1 заносимо A1:A38, у масив 2 – B1:B38, потім натискаємо ОК. В комірці C1 з'являється число -0,46413 (див. табл. 2). Це – значення коефіцієнта кореляції r_{xy}. Знак «–» указує на обернену залежність (більше пропусків – нижче оцінка). За табл. 1 знаходимо ймовірність того, що 38 вимірювань двох некорельованих змінних дадуть коефіцієнт кореляції $r \geq 0,46$. Для 40 вимірювань імовірність менша від 1,1%, тобто кореляція високозначуща. Немає сумніву в тому, що погані оцінки на контрольній роботі пов'язані з пропусками занять. Нижче наведені дані, що обробляються: верхній рядок – пропуски занять, нижній – оцінки за контрольну роботу. Під даними – частина табл. 2:

4	1	4	1	2	1	2	2	0	1
	0	0	0	2	0	3	0		
0	1	0	8	0,5	4	2,5	1	3	3
	7	4	2,5	0	1	1	2		
0	0	0	0	1	0	2	0	1	0

	1	5	0	2	0	0	0		
7	4	7	1	0	6	4,5	4	2	7
	1	0	1,5	2	3	0,5	0		

0	0	0	0
1	3	6	6

Табл. 2

	A	B	C
1	4	0	-0,46413
2	1	1	
3	4	0	
4	1	8	
5	2	0,5	
6	1	4	
7	2	2,5	
8	2	1	
9	0	3	
10	1	3	
11	0	7	
12	0	4	
13	0	2,5	
…	…	…	
35	0	1	
36	0	3	
37	0	6	
38	0	6	

2. *Перевірка статистичних гіпотез*

Спостереження можуть охоплювати всі члени сукупності, що вивчається, або обмежуватись дослідженням лише частини даної сукупності. Статистична сукупність, що досліджується і з якої роблять вибір, називається ***генеральною сукупністю***. Відібрана з генеральної сукупності деяка частина одиниць, яку будуть безпосередньо досліджувати, називається ***вибірковою сукупністю*** або ***вибіркою***.

Щоб вибірка найповніше відображувала структуру генеральної сукупності, слід проводити процес відбору випадковим чином, щоб забезпечити всім варіантам однакову можливість потрапити до вибірки. При дотриманні всіх вимог всі висновки, зроблені при дослідженні вибірки, можна перенести на генеральну сукупність. При вибірковому методі дослідження проводять на малій частині сукупності – від 5% до 15%.

Педагогічний експеримент полягає найчастіше в порівнянні результатів, наприклад, навчання експериментальної та контрольної груп учнів. В більшості випадків педагогічний експеримент є квазіекспериментом, оскільки на результати його впливає велика кількість неврахованих факторів. Крім того, майже завжди проводиться вибіркове дослідження, коли вивчається лише 10–15% членів сукупності. При формуванні вибірки практично неможливо дотриматись випадковості відбору, тому до початку експерименту слід виконати перевірку рівня підготовки та інших ознак у відібраних груп учнів і в якості контрольної вибрати групу з кращими результатами перевірки. В деяких випадках певну інформацію про досягнення учнів можна одержати при порівнянні результатів зрізів знань в одному й тому ж класі через деякий проміжок часу.

В будь-якому разі з точки зору математичної обробки результатів ми повинні дати відповідь на питання: чи суттєво відрізняються між собою одержані два масиви чисел? Чи вправі ми зробити висновок, що результати експериментальної групи суттєво відрізняються від результатів контрольної?

Розглянемо порядок дій, які виконуються в ході експерименту.

Підраховують середні арифметичні оцінок $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Знаходять суму квадратів відхилень $SS_x = \sum (x_i - \bar{x})^2$ і обчислюють дисперсію σ_x^2 та стандартне відхилення σ_x:

$$\sigma_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}; \qquad \sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}.$$

Стандартне відхилення використовують для підрахунку ***статистичної похибки середнього арифметичного***, яка виступає ***мірою точності***, з якою вибірковий показник репрезентує генеральний параметр:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma_x^2}{n}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n(n-1)}}.$$

Точність визначення вибіркової середньої визначається відношенням $\sigma_{\bar{x}}$ до $\bar{x}$. Це відношення позначають C_S:

$$C_S = \frac{\sigma_{\bar{x}}}{\bar{x}} \cdot 100\%$$

Точність визначення середніх показників вважається достатньою, якщо C_S не перевищує 15%.

При одержанні висновків про достовірність відмінностей між результатами формулюють дві гіпотези. ***Нульова гіпотеза*** – твердження, що новий метод не має яких-небудь переваг. ***Альтернативна гіпотеза*** – припущення про перевагу нового методу.

Альтернативну гіпотезу приймають, коли нульову спростовують. Це буває тоді, коли різниця, наприклад, в середніх арифметичних експериментальної і контрольної груп настільки значна, що ризик помилитись при відмові від нульової гіпотези і прийняти альтернативну не перевищує одного з трьох рівнів значущості статистичного висновку:

Перший рівень – 5% (або α=0,05) : ризик помилки в п'яти випадках зі ста теоретично можливих таких самих експериментів. Другий рівень – 1% (або α=0,01). Третій рівень – 0,1% (або α=0,001).

Останній рівень використовується рідко, оскільки він передбачає дуже високі вимоги до обґрунтування результатів експерименту.

Для перевірки прийнятої гіпотези використовують ***критерії достовірності***. Оскільки найчастіше в дослідженнях одержуємо сукупність даних, які можна описати нормальним законом розподілу, то скористаємося критерієм, що використовує значення середнього арифметичного і ширини довірчого інтервалу в залежності від обраного рівня значущості. Цей критерій називають ***t-критерієм Стьюдента***:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{S_{(\bar{x}_1 - \bar{x}_2)}}, \tag{10}$$

де $\bar{x}_1$, $\bar{x}_2$ – середні арифметичні в експериментальній та контрольній групах, а $S_{(\bar{x}_1 - \bar{x}_2)}$ – стандартна похибка різниці середніх арифметичних, яка для $n_1 \neq n_2$ визначається за формулою:

$$S_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2} = \sqrt{\frac{\sum(x_i - \bar{x}_1)^2 + \sum(x_i - \bar{x}_2)^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}. \tag{11}$$

Знайшовши фактичне значення t-критерія Стьюдента $t_ф$, порівнюємо його з теоретичним значенням t_{st} для обраного рівня значущості α і числа ступенів свободи $f=n_1+n_2-2$ (табл. 3). Якщо $t_ф \geq t_{st}$, то відмовляються від нульової гіпотези і обирають альтернативну.

Таблиця 3.

Критичні точки *t*-критерія Стьюдента.
Довірчі границі для *t* з *f* степенями свободи

f	Двосторонні границі			
	5%	2%	1%	0,1%
1	12,710	31,820	63,660	636.600
2	4,303	6,965	9,925	31,600
3	3,182	4,541	5,841	12,920
4	2,776	3,747	4,604	8,610

5	2,571	3,365	4,032	6,869
6	2,447	3,143	3,707	5,959
7	2,365	2,998	3,499	5,408
8	2,306	2,896	3,355	5,041
9	2,262	2,821	3,250	4,781
10	2,228	2,704	3,169	4,587
11	2,201	2,718	3,106	4,437
12	2,179	2,681	3,055	4,318
13	2,160	2,650	3,012	4,221
14	2,145	2,624	2,977	4,140
15	2,131	2,602	2,947	4,073
16	2,120	2,583	2,921	4.015
17	2,110	2,567	2,898	3,965
18	2,101	2,552	2,878	3,922
19	2,093	2,539	2,861	3,883
20	2,086	2,528	2,845	3,850
21	2,080	2,518	2,831	3,819
22	2,074	2,508	2,819	3,792
23	2,069	2,500	2,807	3,767
24	2,064	2,492	2,797	3,745
25	2,060	2,485	2,787	3,725
26	2,056	2,479	2,779	3,707
27	2,052	2,473	2,771	3,690
28	2,048	2,467	2,763	3,674
29	2,045	2,462	2,756	3,659
30	2,042	2,457	2,750	3,646
40	2,021	2,423	2,704	3,551
50	2,009	2,403	2,678	3,495
60	2,000	2,390	2,660	3,460
80	1,990	2,374	2,639	3,415
100	1,984	2,365	2,626	3,389
200	1,972	2,345	2,601	3,339
500	1,965	2,334	2,586	3,310
∞	1,960	2,326	2,576	3,291
f	2,5%	1%	0,5%	0,05%
	Односторонні границі			

Наведемо приклад застосування t-критерія Стьюдента. В комірки A1-A11 заносимо результати тестування експериментальної групи, в B1-B9 – контрольної (табл. 4). В комірках A12 і B12 обчислюємо середні арифметичні СРЗНАЧ(A1:A11) та СРЗНАЧ(B1:B9), в комірках A13 і

B13 – стандартне відхилення СТАНДОТКЛОН(A1:A11) та СТАНДОТ-КЛОН(B1:B9), в комірках A14 і B14 − суми квадратів відхилень КВАД-РОТКЛ(A1:A11) та КВАДРОТКЛ(B1:B9). В комірці C1 набираємо формулу =(A12-B12)/КОРЕНЬ(0,01122*(A14+B14)).

Табл. 4

	A	B	C
1	12	13	3,982072
2	14	9	
3	13	11	
4	16	10	
5	11	7	
6	9	6	
7	13	8	
8	15	10	
9	15	11	
10	18		
11	14		
12	13,63636	9,444444	
13	2,460599	2,185813	
14	60,54545	38,22222	

Число 0,01122 − це результат обчислення $\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\cdot\frac{1}{n_1+n_2-2}$ (в прикладі n_1=11, n_2=9). В комірці C1 з'являється число 3,982072 (табл. 4). Табличне значення критерію Стьюдента для f=18 дорівнює 2,878 при можливості зробити помилку в одному зі ста випадків. Оскільки $t_ф>t_{st}$, то ми відхиляємо нульову гіпотезу і приймаємо альтернативну, що учні експериментальної групи показують більш високий рівень знань.

Стандартне відхилення (комірки A13 і B13) можна використати для знаходження точності визначення вибіркової середньої C_S.

Смисл t-критерію Стьюдента стає особливо зрозумілим, якщо переписати вираз (10) для випадку, коли $n_1=n_2=n$. Підставимо в (10) вираз (11), використавши цю умову.

$$t=\frac{\left|\bar{x}_1-\bar{x}_2\right|}{\sqrt{\frac{\sum\left(x_i-\bar{x}_1\right)^2+\sum\left(x_i-\bar{x}_2\right)^2}{2\cdot(n-1)}\cdot\frac{2}{n}}}=\frac{\left|\bar{x}_1-\bar{x}_2\right|\cdot\sqrt{n}}{\sqrt{\frac{\sum\left(x_i-\bar{x}_1\right)^2+\sum\left(x_i-\bar{x}_2\right)^2}{(n-1)}}}.$$

В знаменнику під коренем стоїть сума дисперсій:

$$t = \frac{\left|\bar{x}_1 - \bar{x}_2\right| \cdot \sqrt{n}}{\sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2}} .$$

Отже, чим більша вибірка, чим більша різниця середніх арифметичних і чим менші дисперсії, тим надійніший висновок про значущу відмінність між результатами експериментальної і контрольної груп.

Література

1. Зайдель А. Н. Погрешности измерений физических величин / Зайдель А. Н. – Л. : Наука, 1985.
2. Тейлор Дж. Введение в теорию ошибок / Тейлор Дж. – М. : Мир, 1985. – 272 с.
3. Рего К. Г. Метрологическая обработка результатов технических измерений : справ. пособие / Рего К. Г. – К. : Техника, 1987. – 128 с.