

Лабораторна робота № 3

ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ ЩОДО ВЛАСТИВОСТЕЙ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ

Мета роботи:

Ознайомитися з непараметричним критерієм Колмогорова-Смирнова, призначеним для перевірки гіпотези однорідності двох груп даних за допомогою програми МАТЛАБ.

Основні теоретичні відомості

При обробці даних нерідко виникає необхідність перевірки гіпотези однорідності декількох груп даних, тобто рівності функцій розподілу вибірок. Іноді буває необхідно перевірити збіг центрів розподілів (медіан) або однакове розсіювання декількох вибірок. Можна також перевірити певні властивості однієї вибірки, наприклад відповідність її заданій функції розподілу, гіпотези про симетрію, незалежність і однаковий розподіл (однорідності) членів вибірки. Всі ці гіпотези можна перевірити за допомогою непараметричних критеріїв.

Більшість непараметричних методів заснована на варіаційному ряду, який отримуємо, якщо члени початкової вибірки розташувати в порядку зростання: $x'_1 \leq x'_2 \leq \dots \leq x'_n$.

Для перевірки однорідності двох вибірок часто використовується критерій Колмогорова-Смирнова, який заснований на порівнянні емпіричних функцій розподілів.

Нехай $F_{1n}(x)$ – емпірична функція розподілу випадкової величини X , побудована по вибірці $x_1 \leq x_2 \leq \dots \leq x_n$ і $F_{2n}(x)$ – емпірична функція розподілу випадкової величини Y , побудована по вибірці $y_1 \leq y_2 \leq \dots \leq y_n$:

$$F_{1n}(x) = \frac{1}{n} \sum_{i=1}^n I_{x_i} \leq x, \quad \text{ää} \quad I_{x_i} = \begin{cases} 1, & x_i \leq x \\ 0, & x_i > x \end{cases} \quad (3.1)$$

$$F_{2n}(x) = \frac{1}{n} \sum_{i=1}^n I_{y_i} \leq x, \quad \text{ää} \quad I_{y_i} = \begin{cases} 1, & y_i \leq x \\ 0, & y_i > x \end{cases} \quad (3.2)$$

і обчислюють відстані між ними:

$$\begin{aligned} D_n^+ &= \max_{1 \leq i \leq n} \{F_{1n}(x) - F_{2n}(x)\}; \\ D_n^- &= \max_{1 \leq i \leq n} \{F_{2n}(x) - F_{1n}(x)\}; \\ D_n &= \max_{1 \leq i \leq n} (D_n^+, D_n^-). \end{aligned} \quad (3.3)$$

Для цих статистик відомі асимптотичні розподіли. Обчислені значення

$$z_n = D_n(\sqrt{n} + 0,275 - 0,04/\sqrt{n}), \quad (3.4)$$

$$z_n^{+(-)} = D_n^{+(-)}(\sqrt{n} + 0,12 + 0,11/\sqrt{n}) \quad (3.5)$$

порівнюють з критичними значеннями статистик Стефенса $\tilde{D}_n(q)$ і $\tilde{D}_n^{+(-)}(q)$ для декількох рівнів значущості q , приведених в табл. 3.1.

Таблиця 3.1

	Рівень значущості q				
	0,01	0,025	0,05	0,10	0,15
$\tilde{D}_n^{+(-)}(q)$	1,518	1,358	1,224	1,073	0,973
$\tilde{D}_n(q)$	1,628	1,480	1,358	1,224	1,138

При $z_n > \tilde{D}_n(q)$ або $z_n^{+(-)} > \tilde{D}_n^{+(-)}(q)$ гіпотеза про однорідність двох вибірок на рівні значущості q відхиляється.

При використанні критерію Колмогорова-Смирнова корисно також накреслити графіки емпіричних функцій F_{1n} і F_{2n} , які дозволяють виявити особливості вибірок.

До переваг критерію Колмогорова-Смирнова можна віднести те, що в разі відміни справжнього розподілу вибірки від гіпотетичного за допомогою цього критерію така відмінність буде виявлена (при досить великому числі спостережень n).

Приклад. Перевірити гіпотезу однорідності двох вибірок $X=[x_1, x_2, \dots, x_{10}]$ і $Y=[y_1, y_2, \dots, y_{10}]$, приведених в таблиці 3.2.

Таблиця 3.2

X	4,14	5,68	5,20	3,86	4,31	4,93	5,91	3,75	4,40	4,51
Y	5,98	6,76	6,42	5,65	5,98	5,10	0,50	3,74	4,72	5,34

Розташуємо члени початкові вибірки в порядку зростання: $x'_1 \leq x'_2 \leq \dots \leq x'_{10}$ і $y'_1 \leq y'_2 \leq \dots \leq y'_{10}$ і приведемо їх в табл. 3.3.

Таблиця 3.3

X	3,75	3,86	4,14	4,31	4,40	4,51	4,93	5,20	5,68	5,91
Y	0,50	3,74	4,72	5,10	5,34	5,65	5,98	5,98	6,42	6,76

Відстані між емпіричними функціями розподілу випадкових величин $F_{1n}(x)$ і $F_{2n}(x)$ приведені в табл. 2.4.

З табл.3.4 значення $D_n = 0,4$ і $D_n^- = 0,2$. Отримаємо у відповідності з (3.4) і (3.5) значення $z_n = 1,323$ і $z_n^- = 0,6617$. При рівні значущості $q=0,05$ $z_n < \tilde{D}_n(q)$ і $z_n^- < \tilde{D}_n^{+(-)}(q)$, що підтверджує однорідність двох груп даних.

Таблиця 3.4

x	$F_{1n}(x)$	$F_{2n}(x)$	$F_{1n}(x) - F_{2n}(x)$
0,5	0	0,1	-0,1
3,74	0	0,2	-0,2
3,75	0,1	0,2	-0,1
3,86	0,2	0,2	0
4,14	0,3	0,2	0,1
4,31	0,4	0,2	0,2
4,4	0,5	0,2	0,3
4,51	0,6	0,2	0,4
4,72	0,6	0,3	0,3
4,93	0,7	0,3	0,4
5,1	0,7	0,4	0,3
5,2	0,8	0,4	0,4
5,34	0,8	0,5	0,3
5,65	0,8	0,6	0,2
5,68	0,9	0,6	0,3
5,91	1	0,6	0,4
5,98	1	0,7	0,3
5,98	1	0,8	0,2
6,42	1	0,9	0,1
6,76	1	1	0

Порядок виконання лабораторної роботи

1. Провести перевірку гіпотези про однорідність двох груп даних, представлених двома варіантами завдання в табл. 3.2.
2. Зробіть висновок по отриманих результатах.

Контрольні запитання

1. У чому полягає різниця між параметричними і непараметричними методами обробки даних?
2. Як проводиться перевірка гіпотези про нормальний закон розподілу помилок вимірювань по критерію W?
3. Як проводиться перевірка однорідності двох вибірок по критерію Пірсону χ^2 ?
4. Як проводиться перевірка однорідності двох вибірок по критерію Колмогорова-Смирнова?
5. Як проводиться перевірка однорідності двох вибірок по критеріях знаків і серій?
6. Як проводиться перевірка однорідності двох вибірок по рангових критеріях Манна-Уїтні і Вілкоксона?
7. Як проводиться перевірка однорідності двох вибірок по ранговому критерію Сижела-Тьюки?