Змістовий модуль 2. Поняття параметричної та непараметричної статистики

Лекція 2.2. Непараметрична статистика.

План:

1. Визначення непараметричних критерії перевірки гіпотез.
2. Операції з частотами та рангами
3. Критерій Колмогорова-Смірнова
4. Критерій Вілкоксона-Манна-Уітні.

Критерії значимості і перевірка гіпотез

Дана тема розглядає методи, що застосовуються тоді, коли належить перевірити якісь теоретичні припущення.

Дамо деякі визначення.

Статистична гіпотеза - називається твердження про розподіл генеральної сукупності, відповідне деяким уявленням про досліджуване явище. Зокрема, це може бути твердження про значення параметрів μ і σ нормально розподіленої генеральної сукупності.

Нульова гіпотеза (H_0) - гіпотеза, заснована на твердженні, що між двома генеральними сукупностями немає очікуваної відмінності:

$$\mu_1 = \mu_2$$

Наприклад, за μ_1 - взяті результати стрибків у довжину юних легкоатлетів, що тренуються за традиційною методикою; μ_2 - результати стрибків у довжину іншої групи юних легкоатлетів, які використовують новий комплекс спеціальних вправ.

Таким чином, нульова гіпотеза робить припущення, що генеральні середні арифметичні (результати всіх юних стрибунів даного класу, які могли б тренуватися за традиційною і новою програмою) не відрізняються $\mu_1=\mu_2$.

Альтернативна гіпотеза (H_1) - гіпотеза з твердженням, зворотним нульовій гіпотезі, тобто твердження про те, що в дійсності між генеральними сукупностями є відмінність: $\mu_1 \neq \mu_2$ (результати стрибків у довжину з розбігу юних легкоатлетів, які займаються за традиційною методикою і новою програмою не рівні).

Статистичні гіпотези, зокрема, нульова та альтернативна, перевіряються за допомогою якогось методу - *критерію*.

Існують критерії, засновані на нормальному розподілі даних (параметричні), до них відносяться: F - критерій Фішера; t-критерій Стьюдента; u-критерій.

Існують критерії, які порівнюють середні значення генеральних сукупностей μ_1 і μ_2, розподіл яких відхилився від нормального або параметри тих сукупностей, які вимірюються в шкалах порядку даних або найменувань (наприклад, довільна нумерація гравців футбольної команди або місця, зайняті спортсменами на змаганнях і т.д.). До них відносяться: критерій Вілкоксона або Уайта.

Крім цих критеріїв, існують критерії, за допомогою яких перевіряється припущення про нормальний розподіл генеральної сукупності. Вони називаються критеріями згоди. До них відносяться: Асиметрія (As); Ексцес (E_x); критерій - χ^2 (хі - квадрат); критерій Шапіро-Уілкі.

Перш ніж розглядати перераховані вище критерії, ми повинні визначитися з рівнями значущості, які застосовуються при перевірці гіпотез.

Рівні значущості

В ході дослідницької роботи дуже важливим моментом буває встановлення наявності або відсутності відмінностей в отриманих числових характеристиках при вивченні якихось результатів, показаних піддослідними контрольної та експериментальної груп.

Наприклад, перед дослідником ставиться завдання - розробити експериментальну методику навчання стрибків у довжину з розбігу для учнів загальноосвітньої школи. Після того як нова методика навчання розроблена і застосована в експериментальній групі школярів, їх середній результат виріс на 10 см (x = 10 см), а в контрольній групі цей показник збільшився всього на 4 см (y = 4 см). Перед дослідником постає питання: чи можна стверджувати, що нововведення ефективніше вплинуло на процес формування досліджуваного рухової дії в порівнянні з традиційною методикою або це випадковість?

Відповідаючи на це питання, дослідник перед проведенням експерименту формулює гіпотези:

а) *Нульова гіпотеза* (H_o) - передбачається, що новий комплекс вправ (методика навчання) недостатньо добре розроблений і незначно вплине на результат стрибків у довжину з розбігу, а відмінності в середніх значеннях контрольної і експериментальної груп (якщо вони виявляться), будуть обумовлені тільки дією випадковостей.

б) *Альтернативна гіпотеза* (H_1) - нововведення успішно вирішать завдання навчання в експериментальній групі, а отримані дані будуть перевершувати результати контрольної групи. Далі потрібно довести чи дійсно статистично достовірно, чи навпаки, недостовірно відмінність знайдених середніх приростів результатів стрибків у довжину з розбігу ($\bar{x}$-$\bar{y}$=6 см) контрольної та експериментальної груп.

Для цього обчислюють значення деякої величини, званої критерієм, яка найчастіше має стандартний розподіл (u-розподіл, t-розподіл і т.п.). Знайдена величина порівнюється з критичним (граничним) значенням критерію, узятим з відповідних таблиць, і за результатами порівняння визначається статистична вірогідність наявності або відсутності відмінностей між двома порівнюваними параметрами. Як вже зазначалося, зазвичай достатній рівень значущості α=0,05, більш серйозні висновки рекомендується давати, використовуючи рівень значущості α=0,01 або α=0,001.

Щоб уникнути однозначних, переконливих відповідей на поставлені серйозні питання, діють у такий спосіб: рівень значущості до експерименту не встановлюється точно, а за експериментальними даними обчислюється ймовірність *p* того, що критерій вийде за межі значення, розрахованого у вибірці. *p* - експериментальний рівень значущості. Точне значення *p* також не вказують, зазвичай це робиться в такий спосіб:

а) якщо обчислене значення критерію (наприклад, t-критерію Стьюдента) не перевищує критичного значення (табличне, t-критерій Стьюдента) на рівні значущості α=0,05, то відмінності вважаються статистично недостовірними, записується - (p> 0,05)

б) якщо обчислене значення критерію перевищує критичні значення при α=0,05; α=0,01 або α=0,001, то записується - (p<0,05), (p<0,01), (p<0,001).

Це означає, що спостерігаємі відмінності статистично достовірні на рівнях значущості - 0,05; 0,01 або 0,001.

Статистичний ряд розподілу — впорядкований розподіл одиниць досліджуваної сукупності на групи за групувальною (варіативною) ознакою. Вони характеризують склад (структуру) досліджуваного явища, дозволяють судити про однорідність сукупності, межі її зміни, закономірності розвитку досліджуваного об'єкта. Залежно від ознаки статистичні ряди розподілу діляться на:

- атрибутивні (якісні);
- варіаційні (кількісні):
  - ✓ дискретні;
  - ✓ інтервальні.

Непараметричні критерії

Якщо розподіл відхиляється від нормального, використовуються допоміжні критерії, що допомагають зробити оцінку розбіжностей з генеральними параметрами. В їх основі лежить порівняння не самих середніх значень вибірок, а порядкові числа в ранжированном ряду їх окремих вибіркових значень. Критерії, засновані на цьому принципі, називають порядковими або непараметричними.

Критерій Вілкоксона або Уайта

Одним з найбільш простих і поширених непараметричних критеріїв є W-критерій Вілкоксона (Уайта), який використовується при порівнянні пов'язаних і незв'язаних вибірок.

Для того, щоб скористатися цим критерієм, ми повинні дати визначення рангу.

Ранг - порядковий номер вибіркового значення в ранжированій вибірці. Величина рангу збігається зі значенням вибірки, якщо немає збігів. Якщо ж вони є, то величина рангу визначається як середнє арифметичне порядкових номерів співпадаючих значень.

Наприклад, нехай є вибірка n = 7, яка після ранжирування виглядає наступним чином:

№	1	2	3	4	5	6	7
X_i	3	5	7	7	9	10	12

Значення порядкових номерів 3,4 збіглися, тому величина їх рангів (R) буде дорівнювати:

$$R=(3+4)/2=3,5$$

Таким чином, ранжируваний ряд даної вибірки буде наступним:

X_i	3	5	7	7	9	10	12
R_i	1	2	3,5	3,5	5	6	7

Далі розглянемо практичне застосування W - критерію Вілкоксона, використовуючи дані прикладу, наведеного на сторінці 39. Потрібно визначити існує чи ні достовірне розходження в середніх значеннях незв'язаних вибірок (x_i; y_i) на рівні значущості $\alpha = 0,05$:

1. Об'єднуємо обидві вибірки в одну, $n = n_x + n_y = 20$. Ранжируємо об'єднану вибірку, маючи в своєму розпорядженні дані в порядку зростання (стовпець 1, табл.1). При цьому зазначаємо зірочкою дані, що відносяться до однієї з вибірок, наприклад другої;

2. Знаходимо ранги R_i об'єднаної вибірки. Відзначаємо зірочкою ранги, що відносяться до другої вибірки (стовпець 3);

3.Знаходимо суму рангів, окремо першої ΣR_x і другої ΣR_y вибірок;

4. Меншу суму рангів (в нашому прикладі $\Sigma R_y = 63,5$) приймаємо в якості значення W-критерію;

Таблиця 1 - Розрахунки W-критерію Вілкоксона для незалежних вибірок

№	X_i, $У_i$	R_i	№	X_i, $У_i$	R_i
1	*2*	*3*	*1*	*2*	*3*
1	0*	1*	11	7	10,5
2	1*	2*	12	8*	12,5*
3	2*	3*	13	8	12,5
4	3*	4,5*	14	9	14
5	3	4,5	15	10	15,5
6	4*	6,5*	16	10	15,5

7	4*	6,5*	17	11	17
8	5*	8*	18	12	18
9	6*	9*	19	13	19
10	7*	10,5*	20	17	20
Суммы рангов: $\sum R_x$=146,5; $\sum R_y$ = 63,5					

5. З таблиці * W - критерій (Інтернет) знаходимо критичне значення W_α - критерію Вілкоксона при рівні значущості α = 0,05 і при обсягах вибірки n_1=10; n_2=10:

$$W_{0,05} = 78$$

1. Висновок: оскільки W <$W_{0,05}$, то на рівні значущості α=0,05 ми відкидаємо нульову гіпотезу H_0, тобто за допомогою W-критерію Вілкоксона ми довели, що знайдені відмінності в показниках приросту результатів у стрибках в довжину з розбігу контрольної і експериментальної груп дійсно статистично достовірні на рівні значущості α = 0,05 (p <0,05).

До такого ж висновку ми прийшли при використанні t-критерію Стьюдента. Використовуючи приклад, в якому представлені результати вимірювання ЖЕЛ у школярів до і після перебування в спортивному таборі, доведемо існування відмінності в вибірках x_i і y_i для пов'язаних пар за допомогою W-критерію Вілкоксона (α = 0,05).

1. Відкидаємо пари з однаковими значеннями x_i і y_i (стовпець 2 і 3; табл.11) і для подальших розрахунків скорочуємо обсяг вибірки на число відкинутих пар (в нашому прикладі відкидаємо пару N7, отже n=9).
2. З решти пар утворюємо різниці d_i= x_i - y_i. Ці різниці наведені в стовпці 4.
3. Знаходимо ранги $R|d_i|$ абсолютних значень різниць d_i. Наприклад, розглянемо обчислення значень рангів $Rd_2=Rd_4=Rd_6=Rd_8$=2,5 (см.стовпець 5, табл.11). Отримані в стовпці 4 різниці ранжируємо незалежно від отриманих позитивних і негативних знаків різниць (-100; -100; -100; 100; -200; - 200; -300; -300; -400). Далі знаходимо

$$Rd_2=Rd_4=Rd_6=Rd_8=(1+2+3+4)/4=2,5$$

4. Відзначаємо зірочкою ранги, що відносяться до позитивного значення різниці.

Таблиця 11 - Розрахунки W-критерію Вілкоксона для пов'язаних вибірок

№	x_i	y_i	$D_i = x_i - y_i$	Ранги $\|d_i\|$
1	2	3	4	5
1	3400	3800	-400	9

2	3600	3700	-100	2,5
3	3000	3300	-300	7,5
4	3500	3600	-100	2,5
5	2900	3100	-200	5,5
6	3100	3200	-100	2,5
7	3200	3200	0	-
8	3400	3300	100	2,5*
9	3200	3500	-300	7,5
10	3400	3600	-200	5,5
Суми рангів: R* = 2,5*; R = 42,5				

5. Знаходимо окремо суми рангів негативних R і позитивних R* різниць.

6. Меншу з сум рангів приймаємо в якості значення W-критерію, для нашого прикладу W=R*=2,5.

7. З таблиці W - критерій (Інтернет) знаходимо критичне значення W-критерію Вілкоксона при рівні значущості α=0,05 і обсязі вибірки n = 10:

$$W_{0,05}=9.$$

8. Висновок: оскільки $W<W_{0,05}$, то на рівні значущості α=0,05 ми відкидаємо нульову гіпотезу H_0. Таким чином, спостерігаєма відмінність в нашому прикладі визнана достовірною на рівні значущості α=0,05 (P<0,05).

Кореляція

Для точного вираження залежності між змінними величинами *X* і *У* в математиці приймається поняття функції. При запису *У=f(X)*, певному значенню У, званому аргументом, відповідає тільки одне значення змінної *X*. Ця залежність називається функцією. Залежність між змінними, яким відповідають середні величини, називається кореляційною, або просто кореляцією: $У_x=f(X_i)$.

Щоб виявити кореляцію між швидкістю розбігу на останніх 3 метрах від місця відштовхування (x_i) і результатом в стрибках в довжину з розбігу (y_i), необхідно зіставити величину показників отриманих даних. При малій кількості випадків коефіцієнт кореляції між досліджуваними параметрами можна визначити за формулою:

$$r = \frac{\sum x_i y_i - n\overline{x}\overline{y}}{\sqrt{(\sum x_i^2 - n\overline{x}^2)(\sum y_i^2 - n\overline{y}^2)}}, \qquad (26)$$

де r - коефіцієнт кореляції, x_i і y_i- досліджувані параметри, $\bar{x}$ і $\bar{y}$ - середні значення досліджуваних параметрів.

Спряженість між *X* і *У* може приймати значення від -1 до +1.

Якщо коефіцієнт кореляції поставляє величину 0,3 - слабкий зв'язок, від 0,31 до 0,5 - помірний, від 0,51 до 0,7 - значний, від 0,71 до 0,9 - сильний; від 0,91 до 0,99 - дуже сильний.

Отримані коефіцієнти кореляції зіставляють з граничними (табл.6 Додатків).

Запитання для самоперевірки:

1. Які види середніх використовуються в статистиці?
2. Що таке генеральна сукупність; вибірка?
3. За допомогою яких критеріїв перевіряють статистичні гіпотези?
4. Основні форми графічного представлення даних
5. Характеристики розсіювання
6. Закон нормального розподілу
7. Межі довірчого інтервалу
8. Визначення необхідного обсягу вибірки для отримання оцінок заданої точності
9. Критерії значимості і перевірка гіпотез
10. f-критерій Фішера
11. t - критерій Стьюдента
12. Що таке кореляція?