

# **ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ**

**Змістовий модуль 3**

# Основні поняття

Існує велика кількість різноманітних методів перевірки статистичних гіпотез. При виборі методу для вирішення певного конкретного завдання необхідно виходити з відповідей на такі питання:

- якою є мета перевірки гіпотези;
- у яких шкалах виміряні аналізовані дані;
- чи є аналізовані вибірки незалежними або спряженими;
- скільки вибірок необхідно порівняти.

Розглянуті в цьому розділі методи застосовують при порівнянні двох вибірок. При більшій кількості вибірок використовують методи дисперсійного аналізу.

# Нульова та альтернативна гіпотези

Гіпотезу, що перевіряють, називають нульовою гіпотезою ( $H_0$ ). Прикладами нульових гіпотез можуть бути такі твердження: “Середні значення двох вибірок суттєво не відрізняються одне від одного”; “Дисперсія першої вибірки суттєво перевищує дисперсію другої”; “Розподіл вибірки відповідає нормальному закону з певними параметрами” тощо. Гіпотезу, що суперечить нульовій, називають конкуруючою, або альтернативною гіпотезою ( $H_1$ ). Для вказаних вище нульових гіпотез конкуруючими можуть бути такі твердження:

“Середні значення двох вибірок суттєво розрізняються одне від одного”;

“Дисперсія першої вибірки не перевищує істотно дисперсію другої”;

“Розподіл вибірки не відповідає нормальному закону із вказаними параметрами”.

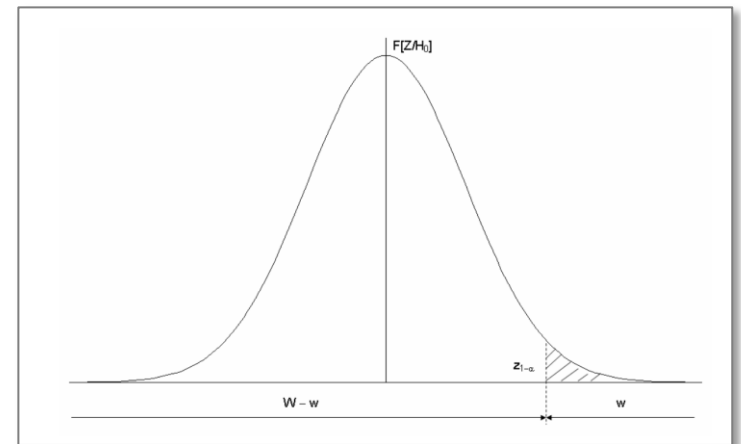
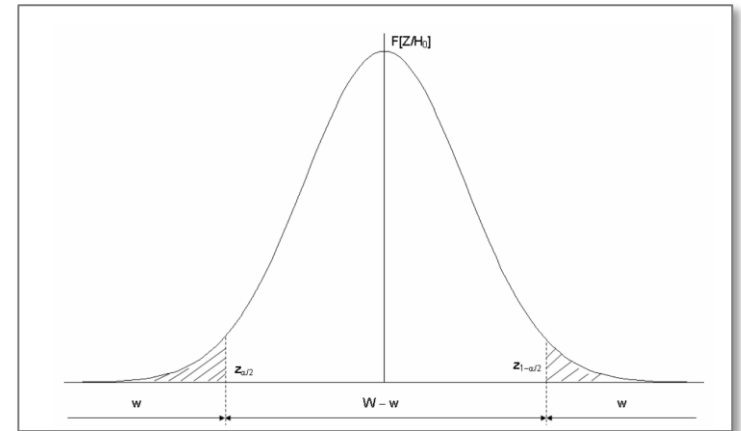
Для однієї нульової гіпотези у загальному випадку можна сформулювати багато різних альтернативних гіпотез.

Розрізняють прості та складні гіпотези. **Простою** називають гіпотезу, що містить тільки одне твердження. **Складні** гіпотези складаються з декількох простих (при цьому кількість простих гіпотез може бути нескінченно великою).

Зазвичай при перевірці нульової гіпотези використовують певні модельні розподіли, що приблизно відповідають розподілу досліджуваного параметра. Їх називають статистичними критеріями. На практиці як критерії найчастіше використовують нормальний розподіл,  $\chi^2$ -розподіл, розподіли Стюдента і Фішера. Значенням критерію, що спостерігається, називають його величину, яку розраховують за досліджуваними вибірками.

Для перевірки гіпотези весь вибірковий простір поділяють на дві області, що не перетинаються: критичну ( $w$ ) та область прийняття ( $W - w$ ). Критичною областю називають сукупність значень критерію, за яких нульову гіпотезу слід відхилити. Областю прийняття гіпотези (областю допустимих значень) називають сукупність значень критерію, за яких нульову гіпотезу приймають. Перевірка гіпотези передбачає розрахунок значення критерію і перевірку його потрапляння до області прийняття гіпотези.

Вирізняють двобічні й однобічні (лівобічні, правобічні) критичні області. Їх використання залежить від вибору конкуруючої гіпотези.



Якщо розподіл імовірності спостережень, що відповідає нульовій гіпотезі  $H_0$ , є відомим, то критичну область визначають так, щоб при виконанні  $H_0$  імовірність її відхилення була рівною заздалегідь заданій малій величині (рівню значущості)  $\alpha$ .

$$P(x \in w | H_0) = \alpha,$$

Замість рівня значущості можна використовувати також довірчий рівень  $p = 1 - \alpha$ .

Критерії, що базуються на використанні заздалегідь заданого рівня значущості, називають критеріями значущості. Рівень значущості визначає розмір критичної області: що більшим є рівень значущості, то ширшою буде критична область.

Розглядають два типи помилок, що можуть виникати при перевірці статистичних гіпотез:

- помилкою першого роду є відхилення правильної нульової гіпотези, рівень значущості  $\alpha$  є ймовірністю такої помилки;
- помилкою другого роду є прийняття помилкової нульової гіпотези.

У деяких застосуваннях помилки першого та другого роду називають, відповідно, ризиком виробника та ризиком споживача.

Зменшення ймовірності помилки першого роду водночас призводить до підвищення ймовірності помилки другого роду  $\beta$ . З огляду на це додатково вводять поняття потужності критерію  $1 - \beta$ , яка є ймовірністю відхилення помилкової нульової гіпотези, тобто ймовірністю потрапляння критерію до критичної області за умови, що правильною є конкуруюча гіпотеза:

$$P\{x \in w | H_1\} = 1 - \beta.$$

Потужність критерію можна підвищити, збільшуючи обсяг вибірки. При визначенні критичної області її зазвичай будують так, щоб максимізувати потужність обраного критерію. За наявності декількох критеріїв, що можуть використовуватися для перевірки досліджуваної гіпотези, рекомендується обирати більш потужні з них, якщо їх застосування є обґрунтованим.

Загальна методика отримання висновків при перевірці гіпотез передбачає, що на першому етапі необхідно задати рівень значущості. Найчастіше його беруть рівним 0,01; 0,05 або 0,1. Обираючи рівень значущості, слід пам'ятати, що його зменшення знижує ймовірність помилки першого роду, але збільшує ймовірність помилки другого роду. Тому, виходячи з конкретних умов, потрібно знайти певний компроміс між ймовірностями припустити помилки різного типу.

На другому етапі за даними вибірки розраховують значення критерію та порівнюють його з обчисленими для заданого рівня значущості межами критичної області. Якщо розраховане значення критерію потрапляє до них, то нульову гіпотезу відхиляють. В іншому випадку вважають, що немає підстав для відхилення нульової гіпотези і або приймають її на заданому рівні значущості, або здійснюють додаткову перевірку. Для визначення меж критичної області застосовують спеціальні таблиці або розраховують їх на основі відомих законів розподілу використовуваних критеріїв.

Можливості сучасної комп'ютерної техніки та наявного програмного забезпечення дають змогу отримувати висновки іншим шляхом. Якщо за наявними емпіричними даними розрахувати значення критерію, то на наступному етапі можна визначити, для якого рівня значущості це значення буде критичним. Ураховуючи, що рівень значущості є ймовірністю відхилення правильної нульової гіпотези, ми можемо за його значенням зробити висновок про ймовірність правильності або помилковості нульової гіпотези. Залежно від того, задовольняє нас отримана ймовірність помилки чи ні, нульову гіпотезу приймають або відхиляють.

При перевірці гіпотез доцільно застосовувати різні методи, призначені для вирішення одних й тих самих завдань та однакових типів даних.

Причинами розбіжності отримуваних при цьому результатів зазвичай є:

- помилки при введенні даних;
- непридатність окремих методик для типу даних, що розглядають;
- алгоритмічні помилки у програмах, що використовують для аналізу.

Залежно від наявності або відсутності можливості визначення напряму розбіжності порівнюваних вибірок, розрізняють **однобічні** та **двобічні** критерії. Перші застосовують, якщо наявні дані дають змогу вказати такий напрям, наприклад зробити висновок, що значення порівнюваної ознаки для одної вибірки є вищим, ніж в іншій. Двобічні критерії дають можливість зробити висновок лише про різницю вибірок за порівнюваною ознакою. Відповідно до цього говорять про однобічні й двобічні гіпотези.

Для двобічних критеріїв рівень значущості є вдвічі більшим, ніж для відповідних однобічних. При використанні однобічних критеріїв рекомендується спочатку розраховувати двобічні. **Якщо за двобічним критерієм різниці між вибірками немає, то наступне порівняння за однобічним є необґрунтованим.**



Дані реальних експериментів можуть бути подані **незалежними** або **спряженими** вибірками. Для незалежних вибірок критерії допомагають виявити статистичну значущість різниці, що спостерігається. Прикладами незалежних вибірок є:

- мешканці двох різних населених пунктів (при демографічних дослідженнях);
- дві партії однотипної продукції, виготовлені різними працівниками на різному обладнанні (при розробці технології виробництва);
- випускники різних шкіл (при аналізі результатів зовнішнього незалежного оцінювання).

Критерії, що застосовують до вибірок з попарно спряженими даними, називають **парними**. Прикладами спряжених вибірок є:

- дані опитування громадської думки до й після певної суспільно значущої події;
- дві партії однотипної продукції, виготовлені одними й тими самими працівниками на одному й тому самому обладнанні до й після внесення певних змін до технології;
- одна й та сама партія виробів до і після певної технологічної обробки.

# Параметричні тести

Критерії й тести, що застосовують для порівняння вибірок, поділяють на дві групи:

- параметричні
- непараметричні.

Особливістю параметричних критеріїв є припущення, що розподіл ознаки в генеральній сукупності підпорядковується певному відомому закону. Ця відповідність має бути доведена до застосування будь-якого з параметричних тестів. Переважна більшість параметричних тестів розроблена для нормально розподілених даних. Але для деяких типів гіпотез існують параметричні тести, призначені для вибірок, що підпорядковуються іншим законам розподілу.

Як правило, параметричні критерії є потужнішими за непараметричні. Застосування непараметричних критеріїв у випадках, коли можна використовувати параметричні, призводить до збільшення ймовірності прийняття помилкової нульової гіпотези, тобто помилки другого роду.

# Непараметричні тести

У багатьох випадках емпіричні дані не задовольняють нормальний розподіл. Тому для їх аналізу некоректно застосовувати параметричні тести. Серед непараметричних тестів важливе місце займають так звані робастні методи, що виявляють слабку чутливість до відхилень від стандартних умов і можуть використовуватися в широкому діапазоні реальних умов.

При перевірці нульової гіпотези про однорідність вибірок числових даних рекомендується використовувати омега-квадрат критерій або (за відсутності необхідних таблиць та програмного забезпечення) критерій Смірнова.