# Розділ 3. Перевірка гіпотез про розташування та розсіювання

## 3.1. Перевірка статистичних гіпотез (СГ).

Загальні положення.

СГ – це гіпотези, які відносяться до вигляду або окремих параметрів розподілу випадкових величин.

Нехай $f(X,\Theta)$ – закон розподілу випадкової величини $X$ з деяким параметром $\Theta$.

Тоді:

– $H_0$ (нульова гіпотеза) $\Theta=\Theta_0$;

– $H_1$ (альтернативна або конкуруюча гіпотеза) $\Theta=\Theta_1$;

$H_0$ відкидається у тому випадку, коли ймовірність того, що вона правильна, виявляється нижчим за деякий рівень, що називається рівнем значущості.

При аналізі гіпотез можливі помилки двох видів:

– $H_0$ відкидається, коли вона правильна;

– $H_0$ приймається, коли правильна $H_1$.

Знижуючи рівень значущості ми зменшуємо ймовірність помилки першого роду, але при цьому зростає ймовірність помилки другого роду. Тому вводиться поняття потужності критерію, яка є ймовірністю відхилення $H_0$. Оскільки ця ймовірність змінюється при зміні параметрів сукупності (наприклад, розмір вибірки), то зазвичай розглядують криву потужності. На рисунку 3.1 наведена крива потужності для перевірки біноміальної гіпотези $p$=0.5.

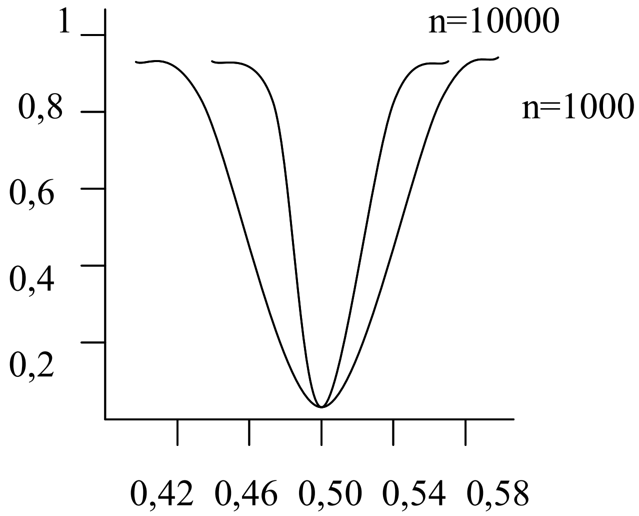


Рис. 3.1. Крива потужності для перевірки біноміальної гіпотези $p$=0.5.

На рис. 3.1 видно, що значення параметра, який перевіряється, відхиляється від 0,5 в ту або іншу сторону, тому ймовірність зробити помилку другого роду зменшується.

Перевірка гіпотез зазвичай проходить наступні етапи:

1. Визначення використовуваної статистичної моделі. Тут висувають деякий набір передумов щодо закону розподілу ВВ і його параметрів. Наприклад, ЗР нормальний, величини незалежні тощо.
2. Формулюють гіпотези $H_0$ і $H_1$.

3. Вибирають критерій (критеріальну статистику), який підходить до висунутої статистичної моделі.

4. Вибирають рівень значущості α в залежності від необхідної надійності висновків.

5. Визначають критичну область для перевірки $H_0$. Якщо значення критерію потрапляє в цю область, то $H_0$ відхиляється. За умови, що $H_0$ правильна, вірогідність попадання в критичну область дорівнює α. Вид цієї області (однобічна або двостороння) залежить від прийнятої $H_0$.

6. Розраховують значення вибраного статистичного критерію для наявних даних.

7. Розраховане значення критерію порівнюють з критичним (його також інколи називають табличним), а потім вирішують приймати або відкидати $H_0$.

Односторонні і двосторонні критерії

У випадку, якщо $H_1$ сформульована у вигляді $\Theta \neq \Theta_0$, використовуються двосторонній критерій (рис.3.2а.).

Якщо формулюється гіпотеза $H_1$ у вигляді $\Theta > \Theta_0$ (або $\Theta < \Theta_0$), то в цьому випадку використовується односторонній критерій (рис 3.2б).

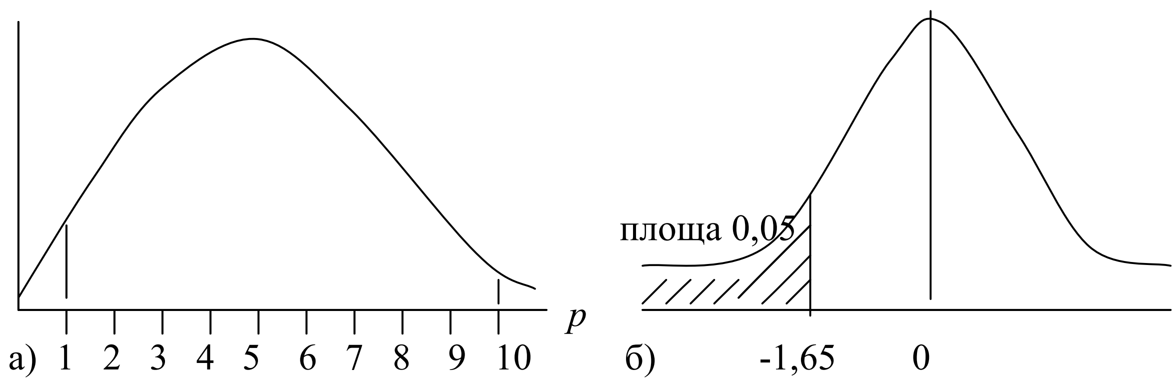


Рис.3.2. Приклади критичної області (а – двосторонній критерій; б – односторонній критерій).

Стійкість критерію

Будь-які методи перевіряють, висуваючи спочатку комплекс деяких передумов про ЗР ВВ. Невиконання передумов робить висновки з відповідних перевірок невідповідними істині. Тобто вірогідність неправильних висновків зростає. Ступінь зменшення надійності висновків при різних критеріях відрізняється. Стійкими (робастними) називають такі критерії, для яких малі відхилення від прийнятих передумов (статистичної моделі) незначно впливають на надійність висновків, що зроблені за ними.

В зв'язку з цим при розв’язанні реальної задачі необхідно підібрати критерій, що відповідає умовам саме цієї задачі. Зважаючи на велику кількість різних критеріїв (особливо непараметричних), це може викликати деякі труднощі у фахівців, для яких статистичні методи є всього лише рідко використовуваним інструментом. Тому пропонується певна послідовність дій, виконання якої можна зробити правильний вибір. Тут розглядуються найчастіше використовувані. Інші можна знайти в Ю.Н. Тюріна, Дж. Полларда, Л. Закса, М. Холлендера.

Послідовність операцій при виборі критерію

1. Сформулювати задачу. Можливі класи задач наведені раніше. Тут розглядуються задачі, що пов'язані з перевіркою яких-небудь параметрів ЗР. Можливі види прикладних задач наведено в таблиці 3.1.
2. Визначити клас використовуваних критеріїв. Необхідно зробити вибір між параметричними і непараметричними критеріями перевірки гіпотез.
3. Визначити додаткові умови для вибору критерію. Багато критеріїв вимагають виконання додаткових умов, без яких їх використання буде некоректним.
4. Вибір конкретного критерію. Для багатьох ситуацій існує декілька приблизно рівнозначних критеріїв для перевірки гіпотези.

Розглянемо розв'язання задачі вибору критерію більш детально.

Постановка задачі

Розглянемо кожну можливу прикладну постановку задачі:

– порівняння показників контрольної і експериментальної вибірок. Є дві незалежних вибірки, середні значення деяких параметрів нам потрібно порівняти. Наприклад, дві групи пацієнтів, лікування яких проводиться різними методами;

– порівняння показників вибірки до і після експерименту. Тут мають місце пов'язані вибірки. Наприклад, значення деякого показника в одній і тій же групі пацієнтів до і після лікування;

– чи можна вважати, що середнє значення показника дорівнює деякому номінальному значенню? Для деяких показників (наприклад, артеріальний тиск, частота пульсу тощо) може існувати деяке значення, що вважається за норму. Перевірити, чи можна вважати, що середнє значення показника в групі, що вивчається, дорівнює нормі. Після перевірки цієї гіпотези для середнього обов'язково треба побудувати довірчий інтервал і простежити, щоб для вибірки виконувалися необхідні умови;

– порівняння розсіювання показника в двох вибірках. У деяких біологічних експериментах важливе не середнє значення показника, а його розсіювання. Наприклад, необхідно вибрати препарат (метод лікування), для яких розсіювання контрольованої ознаки після вживання буде мінімальним;

– чи можна вважати, що в декількох вибірках має місце одне і теж значення показника? Задача є аналогічною попередній, але порівнюється не два види взаємодії, а три і більше;

– чи можна вважати, що в декількох вибірках має місце одне і те ж значення показника? Наприклад, для лікування різних груп хворих застосовуються декілька препаратів-аналогів. Чи можна вважати, що результати лікування статистично не відрізняються?

Таблиця 3.1. Вибір методу для розв'язання задачі про порівняння параметрів розподілу вибірок.

| Формулювання задачі в прикладній постановці | Формулювання задачі в статистичній постановці | Додаткові умови | | Вживаний метод |
|---|---|---|---|---|
| Порівняння показників контрольної і експериментальної вибірок | Перевірка гіпотези про рівність середніх (центрів розподілу) у двох незалежних вибірках | Нормальний закон розподілу | Дисперсії вибірок однакові | t-критерий (Стьюдента) при рівних дисперсіях |
| | | | Дисперсії вибірок не однакові | t-критерий (Стьюдента) при нерівних дисперсіях |
| | | | Без припущення про дисперсії (але при однаковому розмірі вибірок) | t-критерий (Стьюдента) без припущень про дисперсії |
| | | Закон роз-поділу відрізняєтьсявід нормального, або дані вимірюються в нечисловій шкалі | Дисперсії вибірок однакові | Манна-Уїтні (U-критерий Уїлкоксона-Манна-Уїтні) |
| | | | Без передбачень про дисперсії | Двовибірковий Уїлкоксона, медіанний |
| Порівняння показників вибірки до і після експерименту | Перевірка гіпотези про рівність середніх в двох залежних вибірках | Нормальний закон розподілу | | t-критерий (Стьюдента) для зв'язаних вибірок |
| | | Закон розподілу відрізняється від нормального або дані вимірюються в нечисловій шкалі | | Знаковий, одно-вибірковий кри-терий Уїлкоксона |
| Чи можна вважати, що середнє значення показника дорівнює деякому номінальному значенню? | Перевірка гіпотези про рівність середнього константі | Нормальний закон розподілу | | t-критерий (Стьюдента) |
| | | Закон розподілу відрізняється від нормального або дані вимірюються в нечисловій шкалі | | Гупта, знаковий |
| Порівняння розсіювання показника в двох вибірках | Перевірка гіпотези про рівність дисперсій (про приналежність дисперсій до однієї генеральної сукупності) | Нормальний закон розподілу | | F-критерий Фішера |
| | | Закон розподілу відрізняється від нормального або дані вимірюються в нечисловій шкалі | | Зігеля-Тьюки, Мозеса |
| Чи можна вважати, що в декількох вибірках має місце одне і те ж значення показника? | Перевірка гіпотези про рівність дисперсій (про приналежність дисперсій до однієї гене-ральної сукупності) | Нормальний закон розподілу | | Шеффе, Діксона дисперсійний аналіз, LSD |
| | | Закон розподілу відрізняється від нормального або дані вимірюються в нечисловій шкалі | | Краскела-Уолліса, медіанний, ранго вих сум Фрідмана |
| Чи можна вважати, що в декількох вибірках має місце одне і те ж значення розсіювання показника? | Перевірка гіпотези про рівність середніх (про належність середніх до однієї гене-ральної сукупності) | Нормальний закон розподілу | | G-критерий при рівному розмірі виб., Бартлета |
| | | Закон розподілу відрізняється від нормального або дані вимірюються в нечисловій шкалі | | Фрідмана |

## Визначення класу використовуваних критеріїв

Необхідно зробити вибір між параметричними і непараметричними критеріями. Вибір залежить від закону розподілу вибірки і шкали виміру, в якій представлені початкові дані (класифікації, порядку, інтервалів, відношень).

Якщо дані представляються в шкалі найменувань і шкалі порядку (нечислові шкали), то застосовуються рангові (непараметричні) критерії. Нагадаємо, коли значення якісне – маємо справу з нечисловою природою шкали, хоча в таблиці можуть бути записані числа (бали, номер препарату і тому подібне).

Рангові (непараметричні) критерії застосовуються також і для числових даних, у тому випадку, коли закон розподілу вибірки відрізняється від нормального.

Сувора перевірка ЗР вимагає великої кількості даних, тому на практиці зазвичай застосовують спрощені критерії. Розглянемо один з них.

Можна вважати, що ВВ розподілена за НЗ, якщо виконуються умови, що є умовою НЗ розподілу:

1. майже 99.7% відхилень від середнього значення менше 3σ ($\varepsilon_1<3\sigma$);
2. дві третини (68.3%) відхилень менше σ;
3. половина відхилень менша, ніж 0,625σ.

Якщо ці умови виконуються, то можна вважати, що гіпотеза про НЗР не протирічить наявним даним. Всі ці перевірки засновані на властивостях НР.

## Визначення додаткових умов для вибору критерію

Найбільш поширеними додатковими умовами для вибору критерію є:

– однакові або неоднакові розміри вибірок?

– однакові або неоднакові дисперсії порівнюваних вибірок?

– чи однакові закони розподілу порівнюваних вибірок?

Остання умова є вимогою майже будь-якого критерію, але ніколи реально не піддається перевірці. Вона має бути забезпечене правильним формуванням вибірок.

Перша умова перевіряється простим порівнянням, а для перевірки другої – використовуються відповідні критерії, які вибираються аналогічно.

## Вибір конкретного критерію

Якщо є декілька варіантів, критерій вибирається виходячи з наявності програмних засобів або можливості перевірки передумов для його використання.

### Вимоги до вибірок

При проведенні досліджень (особливо клінічних) необхідно забезпечити наступні властивості вибірки:

– однорідність. При вибірці вплив сукупності факторів, які вивчаються, на ознаки, що нас цікавлять, не повинен суперечити один одному. Наприклад, вплив кави на організм людини у вибірці досліджуваних одночасно, не повинен бути у людей, яких кава збуджує, і тих, яких від неї хилить в сон. У ряді випадків причини неоднорідності можуть бути невідомі і тому перед аналізом даних бажана перевірка методом кластерного аналізу.

При вибірці не повинно бути факторів, що значущо впливають на досліджуваний параметр, окрім тих, які ми вивчаємо.

Репрезентативність (структурна відповідність).

Вибірка, що вивчається, має бути репрезентативна генеральній сукупності. Це означає, що коли формується вибірка з сукупності, вона повинна відповідати наступним вимогам:

1. вид статистичного розподілу вибірки повинен відповідати розподілу генеральної сукупності;
2. величина вибірки має бути достатня для віддзеркалення структури генеральної сукупності.

У тих випадках, коли порівнюються деякі параметри двох вибірок необхідно забезпечити рівність розподілу частот впливаючих факторів (стать, вік пацієнтів, серйозність захворювання тощо) в порівнюваних вибірках.

Співпадання умов спостережень.

Умови спостереження для окремих елементів вибірки або для двох порівнюваних вибірок повинні збігатися. Якнайкращим способом забезпечення цієї властивості є подвійний сліпий метод, при якому ні пацієнт, ні лікар, ні середній медичний персонал не знає, які ліки видаються конкретному хворому. Це дозволяє позбавитися від ефекту навіюваності (вплив якого можливий на 30-50% пацієнтів) і ефекту упередженості.

## 3.2 Параметричні критерії перевірки гіпотез про середні та дисперсії

### 3.2.1 Критерій Фішера

Призначення. Перевірка гіпотези про належність двох дисперсій до однієї генеральної сукупності і отже їх рівності.

Нульова гіпотеза. $S_1^2=S_2^2$.

Альтернативна гіпотеза. Існують такі варіанти $H_1$, в залежності від яких розрізняють критичні області:

1. $S_1^2>S_2^2$ найбільш часто використовуваний варіант $H_1$. Критична область – верхній хвіст $F$-розподілу.
2. $S_1^2<S_2^2$ критична область – нижній хвіст $F$-розподілу. Зважаючи на часту відсутність нижнього хвоста в таблицях критичну область зазвичай зводять до варіанту 1, міняючи місцями дисперсії.
3. Двостороння $S_1^2\neq S_2^2$. Комбінація перших двох.

Передумови. Дані незалежні і розподілені за НЗ.

Короткі теоретичні відомості. Гіпотеза про рівність дисперсій двох нормальних генеральних совокупностей приймається, якщо відношення більшої дисперсії до меншої менше критичного (табличного) значення розподілу Фішера.

$$F_{розр} = \frac{S_1^{\ 2}}{S_2^{\ 2}}, \ F_{розр} < F_{\alpha, v1, v2},$$

де $\alpha$ – рівень значущості; і $v_2$ – степені свободи для дисперсії в чисельнику і знаменнику відповідно. $v_1 = N_1 - 1$, $v_2 = N_2 - 1$, де $N_1$, $N_2$ – розміри вибірок.

Примітка. При описуваному способі перевірки значення $F_{розр}$ повинно бути більше 1. Критерій чутливий до порушення припущення про нормальність.

Для двосторонньої альтернативності $S_1^2 \neq S_2^2$ нульова гіпотеза приймається при виконанні умови $F_{1-\frac{\alpha}{2}, v1, v2} < F_{розр} < F_{\frac{\alpha}{2}, v1, v2}$.

### 3.2.2 Аналіз однорідності дисперсій за Кохреном

Призначення. Перевірка гіпотези про належність декількох дисперсій до однієї генеральної сукупності.

Нульова гіпотеза. $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_i^2 = \ldots \sigma_k^2 = \sigma^2$.

Передумови. Дані незалежні і розподілені за НЗ. Кількість спостережень однакова в кожній вибірці.

Короткі теоретичні відомості. Критеріальне значення розраховується за наступною формулою:

$$G_{розр} = \frac{S_{\max}^{\ 2}}{\sum_{i=1}^{k} S_i^{\ 2}},$$

де $S_{\max}^2$ – найбільша із дисперсій в вибірках; $S_i^2$ – емпіричні дисперсії, що розраховані в кожній вибірці.

Результат порівнюється з табличним значенням. Якщо $G_{розр} < G_{табл,k,n-1}$, то гіпотеза про однорідність приймається. Тут $n$ – кількість дослідів за якими обчислюють оцінку дисперсії $S_i^2$.

Примітка. Відхилення нульової гіпотези може означати відмінність закону розподілу даних від нормального, а неоднорідність дисперсій.

### 3.2.3. Критерій Бартлета

Призначення. Перевірка гіпотези про однорідність декількох дисперсій. Застосовується в випадку, коли вибірки, за якими визначаються дисперсії, мають неоднаковий розмір (на відмінність від критерію Кохрена, що вимагає вибірок однакового розміру).

Нульова гіпотеза. Дисперсії, що вивчаються, належать до однієї генеральної сукупності:

$$\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_i^2 \ldots = \sigma_k^2 = \sigma^2.$$

Альтернативна гіпотеза. $\sigma_i^2 \neq \sigma^2$ для конкретного $i$. Тобто, хоч би одна з дисперсій належить іншій генеральній сукупності.

Передумови. Розподіл даних має бути нормальним, а дані – незалежні.

Короткі теоретичні відомості.

Розрахункове значення критерію $\chi^2$ розраховують за формулою:

$$\chi^2_{розр} = \frac{2}{C}\left(2.3026\left(v \ln S^2 - \sum_{i=1}^{k} v_i \ln S_i^{\ 2}\right)\right),$$

$$\text{де} \qquad C = \frac{\sum_{i=1}^{k} \frac{1}{v_i} - \frac{1}{v}}{3(k-1)} + 1; \; S^2 = \frac{\sum_{i=1}^{k} v_i S_i^2}{v}$$

При цьому $v = \sum_{i=1}^{k} v_i$ – загальна кількість степенів свободи; $k$ – кількість груп; $v_i=n_i-1$ – кількість степенів свободи в кожній групі ($n_i$ – об'єм $i$-ої групи); $S_i^2$ – оцінка дисперсії в кожній групі; $S^2$ – зважена оцінка дисперсії.

Якщо розраховане значення $\chi^2_{розр}$ більше або дорівнює критичному (табличному) значенню, взятому з рівнем значущості $\alpha$ і кількістю степенів свободи $v$, то нульова гіпотеза приймається.

Примітка. Даний критерій дуже чутливий до відхилення даних від НЗР. Якщо виникли сумніви в тому, що досліджувані дані розподілені відповідно до НЗР, даний критерій краще не використовувати.

### 3.2.4. Перевірка гіпотез про рівність середніх при неоднакових дисперсіях вибірок

Критерій Стьюдента використовується для перевірки різних гіпотез про середні при нормальному законі розподілу.

Призначення. Перевірка рівності середніх двох генеральних совокупностей, з яких взято дві вибірки.

Передумови. Обидві вибірки взято з сукупності, що має нормальний закон розподілу, дані незалежні, дисперсії вибірок відрізняються, вибірки незалежні.

Короткі теоретичні відомості. Критеріальне значення розраховується за формулою:

$$t = \frac{\left(\overline{X}_1 - \overline{X}_2\right)}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}.$$

При цьому кількість степенів свободи для $t$-критерію дорівнює:

$$v = \frac{\left(\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}\right)^2}{\frac{\left(\frac{S_1^2}{N_1}\right)^2}{N_1+1} + \frac{\left(\frac{S_2^2}{N_2}\right)^2}{N_2+1}} - 2$$

где $N_1$ и $N_2$ – розміри першої та другої вибірок; $S_1^2$ и $S_2^2$ – емпіричні дисперсії; $\overline{X}_1$ и $\overline{X}_2$ – оцінки середніх значень в вибірках. При обчисленнях вважають, що $\overline{X}_1$ більше $\overline{X}_2$.

Нульова гіпотеза.

– $H_0$: $\overline{X}_1 = \overline{X}_2$ проти $H_1$: $\overline{X}_1 \neq \overline{X}_2$. В цьому випадку гіпотеза про рівність середніх відкидається, якщо по абсолютній величині критеріальне значення більше верхньої 100α/2 % точки $t$-розподілу, взятого з $v$ степенями свободи, тобто при $|t|>t_{v,\alpha/2}$;

– $H_0$: $\bar{X}_1 \le \bar{X}_2$ проти $H_1$: $\bar{X}_1 > \bar{X}_2$. Нульова гіпотеза відкидається, якщо критеріальне значення більше верхньої 100α% точки $t$-розподілу, взятого з $v$ степенями свободи, тобто при $|t|>t_{v,\alpha}$.

Примітка. Критерій стійкий при малих відхиленнях від НР.

### 3.2.5. Перевірка гіпотези про рівність середніх при однакових дисперсіях

Призначення. Перевірка рівності середніх двох генеральних совокупностей, з яких вибирають дві вибірки.

Передумови. Обидві вибірки вибрані з сукупностей, що мають НР. Дані незалежні. Дисперсії вибірок однакові.

Короткі теоретичні відомості. Критеріальне значення обчислюється за наступною формулою:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{(\frac{1}{N_1}+\frac{1}{N_2})\left(S_1^2(N_1-1)+S_2^2(N_2-1)\right)}{N_1+N_2-2}}},$$

где $N_1$ и $N_2$ – розміри першої та другої вибірок; $S_1^2$ и $S_2^2$ – емпіричні дисперсії; $\bar{X}_1$ и $\bar{X}_2$ – оцінки середніх значень.

Кількість степенів свободи для перевірки $t$-критерію дорівнює:

$$v=N_1+N_2-2.$$

Нульова гіпотеза.

– $H_0$: $\bar{X}_1 = \bar{X}_2$ проти $H_1$: $\bar{X}_1 \ne \bar{X}_2$. В цьому випадку гіпотеза про рівність середніх відкидається, якщо по абсолютній величині критеріальне значення більше верхньої 100α/2% точки $t$-розподілу, взятого з $v$ степенями свободи, тобто при $|t|>t_{v,\alpha/2}$.

– $H_0$: $\bar{X}_1 \le \bar{X}_2$ проти $H_1$: $\bar{X}_1 > \bar{X}_2$. Нульова гіпотеза відкидається, якщо критеріальне значення більше верхньої 100α% точки $t$-розподілу, взятого з $v$ степенями свободи, тобто при $|t|>t_{v,\alpha}$.

Примітка. Критерій стійкий при малих відхиленнях від НР. При виконанні умови рівності дисперсії критерій стійкий тільки для випадків вибірок однакового об'єму.

### 3.2.6. Перевірка гіпотези про рівність середніх без припущення про дисперсії

Призначення. Перевірка рівності середніх двох генеральних совокупностей, з яких вибрано дві вибірки.

Передумови. Обидві вибірки вибрано з сукупностей, що мають нормальний закон розподілу. Дані незалежні. Розміри вибірок однакові, вибірки незалежні.

Короткі теоретичні відомості. Критеріальне значення обчислюється за формулою:

$$t = \frac{\left(\overline{X}_1 - \overline{X}_2\right)\sqrt{N}}{\sqrt{\left(\sum_{i=1}^{N}\left(X_{1i} - X_{2i}\right) - \left(\overline{X}_1 - \overline{X}_2\right)\right)^2 / (N-1)}},$$

де $N$ – розмір першої та другої вибірок; $\overline{X}_1$ і $\overline{X}_2$ – оцінки середніх у вибірках; $X_{1i}$ і $X_{2i}$ – поточні значення змінних. Кількість степенів свободи для $t$-критерію дорівнює $v=N-1$.

Нульова гіпотеза.

$H_0$: $\overline{X}_1 = \overline{X}_2$ проти $H_1$: $\overline{X}_1 \neq \overline{X}_2$. В цьому випадку гіпотеза про рівність середніх відкидається, якщо по абсолютній величині критеріальне значення більше верхньої 100α/2% точки $t$-розподілу, взятого з $v$ степенями свободи, тобто при $|t|>t_{v,\alpha/2}$.

Примітка. Критерій стійкий при малих відхиленнях від НР.

### 3.2.7. Перевірка гіпотез про рівність середніх заданому значенню *A*

Призначення. Перевірка рівності середнього певному значенню.

Передумови. Вибірки вибрано з сукупності, що має нормальний закон розподілу, дані незалежні.

Короткі теоретичні відомості. Критеріальне значення обчислюється за формулою:

$$t = \frac{\left(\overline{X} - A\right)\sqrt{N}}{S^2},$$

де $N$ – розмір вибірки; $S^2$ – емпірична дисперсія вибірки; $A$ – передбачувана величина середнього значення; $\overline{X}$ – середнє значення.

Кількість степенів свободи для $t$-критерия дорівнює $v=N-1$.

Нульова гіпотеза.

– $H_0$: $\overline{X} = A$ проти $H_1$: $\overline{X} \neq A$. В цьому випадку гіпотеза про рівність середніх відкидається, якщо по абсолютній величині критеріальне значення більше верхньої 100α/2% точки $t$-розподілу, взятого з $v$ степенями свободи, тобто при $|t|>t_{v,\alpha/2}$.

– $H_0$: $\overline{X} \leq A$ проти $H_1$: $\overline{X} > A$. Нульова гіпотеза відкидається, якщо критеріальне значення більше верхньої 100α% точки $t$-розподілу, взятого з $v$ степенями свободи, тобто при $|t|>t_{v,\alpha}$.

– $H_0$: $\overline{X} \geq A$ проти $H_1$: $\overline{X} < A$. Нульова гіпотеза відкидається, якщо критеріальне значення менше нижньої 100α% точки $t$-розподілу, взятого з $v$ степенями свободи.

Примітка. Критерій стійкий при малих відхиленнях від НР.

### 3.2.8. Перевірка гіпотези про рівність середніх при зв'язаних вибірках

Призначення. Перевірка рівності середніх двох генеральних совокупностей, з яких вибрано дві вибірки, за умови, що вибірки пов'язані.

Передумови. Обидві вибірки вибрано з сукупностей, що мають нормальний розподіл. Дані незалежні. Вибірки пов'язані.

Короткі теоретичні відомості. Критеріальне значення обчислюється за формулою:

$$t = \frac{\dfrac{\sum_{i=1}^{N}(x_i - y_i)}{N}}{\sqrt{\dfrac{\sum_{i=1}^{N}(x_i - y_i)^2 - \dfrac{\left(\sum_{i=1}^{N}(x_i - y_i)\right)^2}{N}}{N(N-1)}}},$$

де $x_i$ і $y_i$ – значення пов'язаних рядів спостережень; $N$ – розмір вибірки (кожної, оскільки вони однакові). Кількість степенів свободи для $t$-критерію

$v=N-1$.

Нульова гіпотеза.

– $H_0$: $\overline{X} = \overline{Y}$ проти $H_1$: $\overline{X} \neq \overline{Y}$. В цьому випадку гіпотеза про рівність середніх відкидається, якщо по абсолютній величині критеріальне значення більше верхньої 100α/2% точки $t$-розподілу, взятого з $v$ степенями свободи, тобто при $|t|>t_{v,\alpha/2}$.

– $H_0$: $\overline{X} \leq \overline{Y}$ проти $H_1$: $\overline{X} > \overline{Y}$. Нульова гіпотеза відкидається, якщо критеріальне значення більше верхньої 100α% точки $t$-розподілу, взятого з $v$ степенями свободи, тобто при $|t|>t_{v,\alpha}$.

– $H_0$: $\overline{X} \geq \overline{Y}$ проти $H_1$: $\overline{X} < \overline{Y}$. Нульова гіпотеза відкидається, якщо критеріальне значення менше нижньої 100α% точки $t$-розподілу, взятого з $v$ степенями свободи.

Примітка. Позитивною стороною цього критерію є те, що навіть при значних відхиленнях змінних $x_i$ і $y_i$ від НЗ їх різниця буде досить точно розподілена по НЗ.

### Про можливі парадокси при перевірці гіпотез про середні

При перевірці гіпотез цілком можливі наступні ситуації:

1. Підтверджені гіпотези $\overline{X}_1 = \overline{X}_2$ і $\overline{X}_3 = \overline{X}_2$ але відкинута $\overline{X}_1 = \overline{X}_3$.
2. Підтверджені гіпотези $\overline{X}_1 = A$ і $\overline{X}_2 = A$ але відкинута $\overline{X}_1 \neq \overline{X}_2$.

У зв'язку з цим, для перевірки рівності декількох середніх необхідно використовувати спеціальні критерії.

### Деякі зауваження до перевірок середніх

1. Треба пам'ятати, що у тому випадку, коли вибірка формується невипадковим способом, стандартні відхилення зменшуються, а різниця середніх збільшується.

2. Можливе порівняння параметрів на підставі зіставлення їх довірчих інтервалів. При цьому, якщо інтервали не перекриваються, то між параметрами є статистично значуща різниця. Якщо ж довірчі інтервали перекриваються частково, то з цього не слідує, що вони відрізняються незначуще.

## Визначення розміру вибірки

Для визначення мінімального об'єму вибірки, необхідного для оцінки значення досліджуваного параметра, можна скористатися наступною формулою:

$$n = \frac{N}{1+\alpha^2 N},$$

де $n$ – шуканий розмір вибірки; $N$ – розмір генеральної сукупності; α – рівень значущості при перевірці гіпотези про значення параметра.

## Перевірка рівності декількох середніх за допомогою дисперсійного аналізу

Призначення. Перевірка гіпотези про приналежність декількох середніх значень до однієї генеральної сукупності.

Нульова гіпотеза. $\overline{X}_1 = \overline{X}_2 = ... = \overline{X}_k$ .

Передумови. Розподіл даних має бути нормальним, дані незалежні.

Короткі теоретичні відомості. Використовується методологія дисперсійного аналізу.

Критеріальне значення розраховується за формулою:

$$F = \frac{\dfrac{SS_a}{k-1}}{\dfrac{SS_b}{n-k}},$$

де $n = \sum_{i=1}^{k} n_i$ – загальна кількість експериментів; $n_i$ – об'єм $i$-ої вибірки; $SS_a = \sum_{i=1}^{k} n_i \left( \overline{X}_i - \overline{\overline{X}} \right)^2$ – сума квадратів міжвибіркового розсіювання, де $\overline{X_i}$ – середнє значення по $i$-ій вибірці, а $\overline{\overline{X}}$ – загальне середнє значення по обєднаній вибірці; $SS_b = \sum_{i=1}^{k} \sum_{j=1}^{n} n_i \left( X_{ij} - \overline{\overline{X}}_i \right)^2$ – внутрішньовибіркова сума квадратів, де $X_{ij}$ – значення змінної в $j$-му експерименті $i$-ої вибірки.

Якщо розраховане критеріальне значення більше критичного розподілу Фішера, взятого з деяким рівнем значущості $\alpha$ і степенями свободи *(k-1)* і *(n-k)*, то нульова гіпотеза відкидається. Тобто середні значення (принаймні, деякі з них) належать різним генеральним совокупностям і є різними.

## 3.2.10. Метод множинних порівнянь Шефе (ММПШ)

Досить часто виникає ситуація, коли необхідно порівняти між собою не два значення середніх, а декілька. Порівняння за допомогою дисперсійного аналізу дозволяє з'ясувати, чи можемо ми вважати їх однаковими. У випадку, коли вони не однакові, представляє інтерес з'ясування питання, які середні значення однакові між собою, а які відрізняються, а також які групи середніх однакові між собою, а які – ні. Це дозволяє зробити ММПШ (інша назва – висновки про лінійні контрасти по Шефе) і LSD-критерій.

Для розв’язання цієї задачі наявні критерії неприпустимі, оскільки необхідно перевірити виконання декількох умов одночасно. Існують інші подібні методи, наприклад, метод Тьюкі і метод Стьюдента – Ньюмена – Кейлса. Але вони вимагають спеціальних таблиць.

Призначення. Перевірка гіпотези про приналежність декількох середніх значень до однієї генеральної сукупності або виділення груп середніх значень, що належить до однієї сукупності.

Нульова гіпотеза.

$H_0$: $\sum_{i=1}^{k} c_i \overline{\overline{X}}_i = 0$, де $c_i$ – деякі константи, причому на них накладається умова $\sum_{i=1}^{k} c_i = 0$.

Передумови. Розподіл даних має бути нормальним. Дані незалежні.

Короткі теоретичні відомості.

У даному контексті контрастом називають лінійну комбінацію середніх значень середніх вибірок. Змістовне походження цих умов розглянемо на прикладі. Нехай є 5 вибірок, середні значення яких $\overline{X}_i$. Вважаємо, що ці вибірки належать до двох різних генеральних сукупностей, середні значення в яких $\overline{X_A}$ і $\overline{X_B}$ відповідно. Тоді нульова гіпотеза може бути сформульована у вигляді: $H_0$: $\overline{\overline{X}}_A - \overline{\overline{X}}_B = 0$. При цьому залежно від складу вибірок, з яких можуть складатись групи $A$ і $B$, можливі наступні конкретні варіанти цієї гіпотези:

$$\frac{1}{2}\left(\overline{\overline{X}}_1 + \overline{\overline{X}}_2\right) - \frac{1}{3}\left(\overline{\overline{X}}_3 + \overline{\overline{X}}_4 + \overline{\overline{X}}_5\right) = 0,$$

$$\overline{\overline{X}}_1 - \frac{1}{4}\left(\overline{\overline{X}}_2 + \overline{\overline{X}}_3 + \overline{\overline{X}}_4 + \overline{\overline{X}}_5\right) = 0,$$

а інші варіанти можна скласти самому.

Зрозуміло, що коефіцієнти $C_i$ для першого випадку мають значення 1/2, 1/2 -1/3, -1/3, -1/3, а для другого – 1 -1/4, -1/4, -1/4, -1/4. Вони залежатимуть від того, які групи передбачається перевірити. Так, в першому випадку до однієї групи входять перша і друга вибірка, а в другу – третя, четверта і п'ята. Зрозуміло, найчастіше перевіряються не групи, а окремо взяті вибірки. Наприклад, якщо бажано порівняти між собою першу і четвертую вибірки, то коефіцієнти $C$ матимуть наступні значення 1, 0, 0 -1, 0. Таким чином, змінюючи $C$, можна перевірити будь-які комбінації пар вибірок. Критеріальне значення розраховується за формулою:

$$S = \frac{\left(\sum_{i=1}^{k} C_i \overline{\overline{X}}_i\right)^2}{(k-1)S_{BH}^2 \sum_{i=1}^{k} \frac{C_i^2}{n_i}} .$$

При цьому внутрішньогрупова дисперсія розраховується за формулою:

$$S^2{}_{BH} = \frac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{\bar{x}}_i\right)^2,$$

де $\bar{\bar{x}}_i = \sum_{j=1}^{n_i}\frac{x_{ij}}{n_i}$; $k$ – кількість вибірок; $n_i$ – кількість спостережень в кожній вибірці; $n = \sum_{i=1}^{n} n_i$ – загальна кількість спостережень. Якщо розраховане значення $S$ буде більше критичного значення розподілу $F_{k-1,\ n-k,\ \alpha}$, то гіпотези про рівність середніх відповідних вибірок або груп вибірок відкидається.

Порядок виконання порівняння

Для того, щоб уникнути суперечностей в порівняннях, необхідно дотримуватися наступного порідку проведення порівнянь.

Спочатку середні значення впорядковуються за величиною. Потім проводиться порівняння найбільшого середнього з найменшим. Потім того ж більшого з наступним за величиною найменшим і так далі до тих пір, поки чергова перевірка приведе до прийняття нульової гіпотези про рівність середніх. Після цього найбільше середнє замінюється наступним за величиною (найбільшим, виключаючи самий великий). Всі перевірки знову проводяться розпочинаючи з найменшого середнього значення.

Утворення груп однорідних середніх

Часто виникає потреба в перевірці всіх пар з метою з'ясувати, чи не утворюють вони деяку кількість однорідних груп. Для цього використовується критерій LSD (least significant difference). При цьому необхідно виконати наступні дії.

1. Упорядкувати середні значення вибірок за убуванням.
2. Для кожної середньої пари, розпочинаючи з першої, виконати перевірки значущості різниці середніх.

Для перевірки розраховується значення LSD. Для випадку однакової кількості спостережень в кожній вибірці використовується формула:

$$LSD = t_{n-k,\alpha}\sqrt{\frac{2}{n_i}S^2{}_{BH}} = \sqrt{\frac{2}{n_i}S^2{}_{BH}F_{1,n-k,\alpha}}$$

Це значення буде використано для перевірки всіх пар.

В випадку, якщо об'єми вибірок розрізняються, використовується формула:

$$LSD_{a,b} = t_{n-k,\alpha}\sqrt{\frac{n_a + n_b}{n_a n_b}S^2{}_{BH}} = \sqrt{\frac{n_a + n_b}{n_a n_b}S^2{}_{BH}F_{1,n-k,\alpha}}$$

При цьому значення критерію розраховується для кожної сусідньої пари.

Тут $t_{n-k,\alpha}$ – табличне значення критерію Стьюдента; $S^2{}_{BH}$ - внутрішньогрупова дисперсія; $F_{1,n-k,\alpha}$ – табличне значення критерію Фишера; $n_i$ – кількість спостережень в кожній вибірці (якщо вибірки одинакового розміру); $n_a$ і $n_b$ – кількість спостережень в вибірках, що провіряються на рівність середніх; $n$ – загальна кількість спостережень; $k$ – кількість вибірок.

Якщо різниця середніх значень сусідньої пари менше значення LSD, то ці середні значення вважаються однаковими, а відповідні вибірки об'єднують в однорідну групу.

Примітка. При спільному використанні множинних порівнянь Шеффе і критерію LSD можлива поява кажущихся суперечностей. Наприклад, середні значення вибірок, що входять в різні однорідні групи відрізняються між собою незначимо. Суперечності тут насправді немає. У однорідній групі зібрані вибірки, для сукупності середніх яким приймається гіпотеза про їх рівність. Те, що деякі з них при парному порівнянні можуть дорівнювати середнім з інших вибірок, ніяк цьому не протирічить.

## 3.3. Непараметричні критерії (НК)

НК застосовується в тих випадках, якщо ЗР відрізняється від нормального, або дані вимірюють з використанням нечислових шкал виміру.

У всіх випадках використання наведених НК приймають наступні допущення:

– все ВВ взаємно незалежні;

– вибірки, що підлягають аналізу, розподілені по безперервному розподілу одного і того ж виду.

Часто перевіряють гіпотезу про рівність не середніх, а медіан. Це пов'язано з тим, що є для ЗР, що відрізняється від нормального. Тут медіана є більш стійкою і коректною оцінкою розташування центру розподілу.

Слід пам'ятати, що НК мають для випадку НЗ потужність меншу, ніж відповідні параметричні критерії. У зв'язку з цим не слід використовувати НК при нормальному законі розподілу випадкових величин в досліджуваних вибірках.

### 3.3.1. Поняття про ранги і їх побудова

Порядок значень випадкових величин називають рангами.

Рангом називають номер, який отримає це спостереження у впорядкованій сукупності всіх даних після впорядкування їх згідно певному правилу (наприклад, від меншого значення до більшого).

В деякий джерелах поняття рангу прирівнюється до поняття порядкового номера елементу варіаційного ряду.

Варіаційним рядом – називають значення випадкової вибірки ($x_1, x_2, \ , x_n$), що мають функцію розподілу $F(x)$, і розташовані в порядку їх зростання $x_1 \leq x_2 \leq x_n$.

Ранжирування – це процедура переходу від сукупності спостережень до послідовності їх рангів. Результат ранжирування називається ранжировкою.

## Вибір непараметричного критерію

Рекомендується застосовувати критерії в порядку їх перерахування. Кожен наступний критерій застосовується, якщо попередній не виявив відмінності порівняння вибірок.

В таблиці 3.2 наведено можливі непараметричні критерії та умови їх вибору. В ній прийнято такі умовні позначення:

z – серійний критерій (Вальда-Вольфовича);
ТМФ – точковий метод Фішера;
Ф – критерій Розенбаума;
U – критерій Вілкоксона-Манна-Уїтні;
КЗ – критерій знаків
Т – парний критерій Вілкоксона
Т – критерій Стьюдента.

Таблиці 3.2. Непараметричні критерії та умови їх вибору.

| № | У чому полягає нульова гіпотеза | Зв'язаність вибірок | Число членів кожної вибірки | Критерії |
|---|---|---|---|---|
| 1<br>2<br>3<br>4<br>5<br>6<br>7 | Немає відмінностей в центральних тенденціях розподілу | Зв'язані<br><br><br>Незалежні | 6-25<br>26-300<br>2-5<br>2-10<br>11-20<br>21-60<br>>60 | КЗ, Т, ММФ, U(t)<br>КЗ (t)<br>ММФ, U (t)<br>ТМФ, U (t)<br>Q, ТМФ, U (t)<br>Q, U (t)<br>Q (t) |
| 8<br>9 | Немає відмінностей в розподілах | незалежні | 2-20<br>>20 | z<br>U |
| 10 | Немає відмінностей в частоті появи деяких величин | | 2-20 | ТМФ |

Розглянемо процес ранжирування на такому прикладі. Нехай маємо вибірку що складається з 5 чисел: 8; 25; 42; 3; 1. Цим значенням будуть надано відповідні ранги: 3, 4, 5, 2, 1.

По суті, вони є позиціями елементів наведеної вибірки, якщо її відсортувати за збільшенням.

Інколи дані у вибірках співпадають. У цих випадках використовують так звані середні ранги. Сукупність елементів вибірки, що мають однакове значення, називають зв'язкою, а кількість однакових значень в зв'язці – її розміром. Середнім рангом є середнє арифметичне рангів елементів зв'язки, які б вони мали, якби однакові елементи зв'язки виявилися різними. Наприклад, маємо вибірку чисел: 15; 17; 12; 15; 7; 8; 5; 1; 8.

Цим значенням відповідатимуть ранги: 7,5; 9; 7,5; 6; 3; 4,5; 2; 1; 4,5.

### 3.3.2. Критичні значення статистики W-критерію Уїлкоксона

Критичні значення статистики W-критерію Уїлкоксона широко застосовують при використанні методів перевірки статичних гіпотез, що основані на рангах.

Тут пропонується використовувати для розрахунку критичних значень статистики W формулу, що заснована на її апроксимації з використанням нормального розподілу.

Для конкретного набору значень кількості рівнів рангів легко побудувати функцію розподілу Уїлкоксона. При цьому вважається, що ймовірність будь якого сполучення пар рангів двох експертів є рівноймовірною.

Нижню критичну точку статистики W досить точно обчислюють за апроксимуючою формулою:

$$w(Q,m,n) = \left[ \frac{m(m+n+1)-1}{2} - \Psi(1-Q)\sqrt{\frac{mn(n+n+1)}{12}} \right],$$

де [ ] – операція округлення до найближчого цілого; $Q$ – рівень значущості для одностороннього критерію; $m$, $n$ – розміри досліджуваних вибірок, причому $m \le n$, $\Psi(1-Q)$ – значення оберненої функції нормального розподілу з параметрами (0, 1). Наведена вище апроксимаційна формула дає задовільні результати при $n \ge m \ge 5$.

Верхня критична крапка пов'язана з нижньою співвідношенням:

$$W(Q,m,n)=m(m+n+1)-w(Q,m,n).$$

Таким чином, знаючи нижнє критичне значення статистики $w$, завжди можна обчислити верхнє.

### Критичне значення статистики Манна-Уїтні

Критичне значення цієї статистики пов'язано з критичним значенням статистики $w$ співвідношенням

$$U(Q,m,n)=w(Q,m,n)-m(m+1)\ /2.$$

Наприклад, якщо $Q$=0.05; $m$=10; $n$=12, то

$$U(Q,m,n)=w(Q,m,n)-m(m+1)\ /2=w(0.05,10,12)-10(10+1)/2=89-55=34.$$

### 3.3.3. Критерій Краскела-Уолліса (H-критерій)

Призначення. Перевірка гіпотези про те, що порівнювані вибірки мають однаковий розподіл і отже середні значення цих вибірок однакові.

Нульова гіпотеза. Всі $k$ вибірок мають однаковий розподіл.

Альтернативна гіпотеза. Нульова гіпотеза неправильна.

Передумови. Всі випадкові величини взаємно незалежні.

Перевірка гіпотези виконується в такій послідовності.

1. Вибірки об'єднують в одну.
2. Для об'єднаної вибірки всі спостереження замінюють рангами.
3. Визначають критеріальне значення за формулою:

$$H = \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(N+1),$$

де N – загальна кількість спостережень $N = \sum_{i=1}^{k} n_i$, $n_i$ – кількість спостережень в кожній вибірці; $R_i$ – сума рангів в кожній вибірці; $R_i = \sum_{j=1}^{n_i} r_{ij}$, $r_{ij}$ – ранг $j$-го спостереження в $i$-ій вибірці, $k$ – кількість вибірок.

4. Виконують порівняння розрахункового значення з критичним (табличним), значенням. Якщо розрахункове значення більше критичного, то нульова гіпотеза відхиляється, і ми не можемо вважати, що середні значення однакові між собою.

При малій загальній кількості дослідів ($N$<15) використовуються спеціальні таблиці розподілу Краскела-Уолліса. Якщо загальна кількість дослідів більше вказаної величини, то як критичну використовують верхню 5% точку розподілу $\chi^2$ з $(k-1)$ кількістю степенів свободи.

Примітка. Якщо нульова гіпотеза відхилена, то для того, щоб з'ясувати, які пари вибірок мають однаковий розподіл, не можна використовувати будь-які два вибіркові критерії – необхідно застосовувати множинні порівняння. Цей критерій потужніший, ніж критерій Фішера, тому при числових даних і нормальному законі розподілу слід застосовувати саме критерій Фішера або Кохрена для декількох дисперсій.

Якщо вибірки мають різний розподіл, даний критерій неприйнятний. Замість нього треба використовувати критерій Фрідмана.

При наявності в стовпчику спостережень зв’язок значення критерію корегується за формулою:

$$H' = \frac{H}{1 - \frac{\sum_{i=1}^{L} T_i}{N^3 - N}},$$

де $L$ – кількість зв'язок, $T_i$ – кількість співпадаючих елементів в $i$-й зв'язці.

## 3.3.4. Критерій Зігеля-Тьюкі

Призначення. Критерій призначений для перевірки рівності дисперсій двох вибірок порядкових даних або при законі розподілу, що відрізняється від нормального.

Нульова гіпотеза. $S_1^2=S_2^2=S^2$.

Передумови. Всі випадкові величини взаємно незалежні. Аналізовані вибірки розподілені по безперервному розподілу одного і того ж виду.

Короткі теоретичні відомості.

Перевіряють гіпотезу про приналежність двох незалежних вибірок до однієї генеральної сукупності. Для цього формуємо об'єднану вибірку із спільним числом спостережень $n_1+n_2$ (при цьому $n_1<n_2$), яка упорядковується за величиною. Потім назначають ранги (при цьому ранги формуються незвичайним чином) так, щоб мінімальні і максимальні значення отримали низькі ранги, а по мірі наближення до середніх значень ранги збільшувалися. При цьому, якщо кількість спостережень непарна, то середнє значення (медіана) не отримує ніякого рангу. Ранг 1 надається найменшому значенню, ранги 2 і 3 двом найбільшим, 4 і 5 наступним найменшим, 6 і 7 – найбільшим і так далі Потім для кожної вибірки визначають суми наданих рангів $R_1$ і $R_2$. Для перевірки правильності проставляння рангів можна скористатися виразом $R_1+R_2=(n_1+n_2)(n_1+n_2+1)/2$ для непарної кількості спостережень або $R_1+R_2=(n_1+n_2)(n_1+n_2+1)/2-1$ для парного. Для малих об'ємів вибірки ($n_1\leq n_2\leq 20$) існують спеціальні таблиці, які дають критичні значення розподілу Уїлкоксона. В випадках, коли $n_1, n_2>9$ або $n_1>2$, $n_2>20$, можна застосовувати стандартну нормальну змінну:

$$\tilde{z}=\frac{2R_1-n_1(n_1+n_2+1)+1}{\sqrt{n_1(n_1+n_2+1)\left(\frac{n_2}{3}\right)}},$$

де $R_1$ – сума рангів меншої за об'ємом вибірки. Коли $2R_1>n_1(n_1+n_2+1)$ в чисельнику вираження для $z$ перед 1 знак потрібно замінити з (+) на (–). Якщо об'єми вибірок відрізняються, то $z$ необхідно скорегувати:

$$z'=z+(1/10n_1-1/10n_2)(z^3-3z).$$

У випадку, якщо більше 20% спостережень (різних вибірок!) пов'язано рівняннями або залежностями, знаменник вираження для $z$ набирає наступного вигляду:

$$\frac{\frac{n_1(n_1+n_2+1)n_2}{3}-4n_1n_2}{\sqrt{(n_1+n_2)(n_1+n_2-1)(S_1+S_2)}},$$

де $S_1$ – сума квадратів рангів залежних спостережень; $S_2$ – сума квадратів середніх рангів залежних спостережень (у кожній групі, залежних спостережень знаходиться середній ранг).

За обчисленим значенням $z$ з таблиці визначається ймовірність, яку помноживши на 2, отримують рівень значущості $\alpha$. Якщо ця ймовірність мала (менше 0,05), то з довірчою ймовірністю $1-\alpha$, то можна вважати, що дисперсії генеральних совокупностей розрізняються.

Примітка. Чим більше розрізняються вибірки, тим менш надійний цей критерій.

### 3.3.5. Критерій знаків

Призначення. Критерій застосовують у тому випадку, коли порівнювані вибірки однакового розміру розбиваються природним чином на пари (пов'язані вибірки). Інколи такі перевірки називають перевіркою наявності ефекту обробки. Наприклад, показники одних і тих же хворих до прийому препарату і після.

Нульова гіпотеза. Фактично тут застосовують гіпотезу про рівність нулю медіани різниці між парами двох вибірок.

Короткі теоретичні відомості. Розглядуються значення:

$$z_i = \begin{cases} 1, якщо Y_i - X_i > 0 \\ 0, якщо Y_i - X_i < 0 \end{cases},$$

де $Y_i$ і $X_i$ – значення елементів відповідних вибірок.

Оскільки передбачається, що загальний розподіл $X$ і $Y$ безперервний, то вважають, що ймовірність рівності значень $X$ і $Y$ дорівнює 0. Тому, в разі співпадання значень вони відкидаються, і кількість вимірів відповідно зменшується. Обчислюється число $m$, що дорівнює кількості $z_i$, значення яких дорівнюють 1.

Для невеликих вибірок можна обчислити рівень значущості (ймовірність відкинути гіпотезу про рівність нулю медіани різниці, якщо вона правильна) за формулою:

$$P = \frac{1}{2^n}\sum_{i=0}^{n-m}\frac{n!}{i!(n-i)!},$$

де $n$ – розмір вибірки.

Для великих вибірок можна скористатися наближеною нормальною апроксимацією:

$$P = 1 - \Phi\left(\frac{m - 1 - 0.5n + 0.5}{\sqrt{0.5n + 0.5}}\right).$$

Якщо отримане значення $P$ мале (менше 0.05), то гіпотезу про рівність нулю відкидають.

У тих випадках, коли розміри вибірок не однакові або природне розбиття на пари неможливе, приймають медіанний критерій для двох вибірок або двохвибірковий критерій Уїлкоксона.

### 3.3.6. Критерій Манна-Уїтні (U-критерий Уїлкоксона-Манна-Уїтні)

Призначення. Перевірка гіпотези про рівність середніх двох незалежних вибірок.

Нульова гіпотеза. Дві незалежні вибірки належать одній генеральній сукупності, їх функції розподілу однакові. Ця гіпотеза включає рівність середніх і медіан.

Можна перевірити і іншу гіпотезу: $\overline{X}_1 - \overline{X}_2 = A$, де $A$ – деяка константа. Щоб застосувати цю гіпотезу, необхідно зі всіх спостережень першої вибірки відняти $A$, а залишки вважати за нову першу вибірку. Після цього всі описані нижче дії виконуються аналогічно.

Передумови. Всі незалежні величини взаємно незалежні. Спостереження, що входять в одну вибірку, належать до однієї генеральної сукупності. Дисперсії вибірок однакові.

Короткі теоретичні відомості. Цей критерій – непараметричний аналог $t$-критерия для перевірки середніх значень. Він є найбільш суворим із непараметричних критеріїв. Його можна застосовувати тільки у разі, коли дисперсії вибірок однакові.

Формують єдину вибірку, для її значень визначають ранги (за збільшенням). Спостереженням, що повторюються, надають середній ранг визначають розрахункове значення критерію. Для малих вибірок:

$$U_1==n_1n_2+0.5n_1(n_1+1)-R_1,$$
$$U_2==n_1n_2+0.5n_2(n_2+1)-R_2,$$

де $R_1$ і $R_2$ – суми рангів, розрахованих для значень тих, що належать першій і другій вибіркам відповідно.

Правильність розрахунку $U_1$ і $U_2$ перевіряють співвідношенням

$$U_1+U_2= n_1n_2.$$

Критеріальне значення визначають таким чином $U=min(U_1,U_2)$.

Гіпотезу про рівність вибірок відкидають, якщо $U>(n_1,n_2,\alpha)$, де $U(n_1,n_2,\alpha)$ – критичне значення статистики Манна-Уїтні. Крім того, в пункті 3.3.2 детально наведено, яким чином можна точно обчислити критичне значення статистики $U$ через критичне значення статистики $W$.

Для вибірок, розмір яких досить великий ($n_1+n_2>60$), можна скористатися апроксимацією:

$$U\left(n_1, n_2, \alpha\right)=\frac{n_1 n_2}{2}\tilde{z}_a\sqrt{\frac{n_1 n_2\left(n_1+n_2+1\right)}{12}},$$

де $\tilde{z}$ відповідає значенню стандартного нормального розподілу. Для випадку $n_1 \geq 8$ і $n_2 \geq 8$ можна скористатися нормальною апроксимацією:

$$\tilde{z}=\frac{\left(U-\frac{n_1 n_2}{12}\right)}{\sqrt{\frac{n_1 n_2\left(n_1+n_2+1\right)}{12}}},$$

Якщо розраховане значення $\tilde{z}$ більше табличного значення нормального стандартного розподілу, взятого з відповідним рівнем значущості, то гіпотеза відкидається.

Примітка. Якщо нульова гіпотеза відкидається, то це значить, що вибірки мають різні закони розподілу. Але в загальному випадку це не означає, що середні значення однакові.

Критерій малочутливий до малих значень дисперсій вибірок.

$t$-критерий потужніше за критерій Манна-Уїтні, тому останній не треба використовувати в разі нормального розподілу вибірок. Цей критерій не можна застосовувати до пов'язаних вибірок. З передбачуваної безперервності розподілів аналізованих вибірок виходить відсутність повторень у вибірках. Проте на практиці співпадання зустрічаються досить часто, що пов'язане, наприклад, з округленням даних при вимірюванні. В цьому випадку зроблені за допомогою цього критерію висновки є приблизними (при цьому, чим більше у вибірках співпадаючих значень, тим більш сумнівні висновки). При великій кількості повторень даний метод застосовувати не варто, краще скористатися, наприклад, двохвибірковим критерієм Уїлкоксона.

### 3.3.7. Медіанний критерій для декількох вибірок

Призначення. Використовують для перевірки приналежності $k$ вибірок до однієї генеральної сукупності. З точки зору експериментатора – це перевірка одночасної рівності медіан декількох вибірок.

Нульова гіпотеза. Все сукупності, з яких отримані вибірки, мають однаковий закон розподілу (отже, медіани однакові).

Передумови. Всі випадкові величини взаємно незалежні. Спостереження, що входять в одну вибірку, належать до генеральної сукупності.

Короткі теоретичні відомості. Спочатку знаходять медіану об'єднаної вибірки і після цього формують таблицю, форма якої відповідає таблиці 3.10.

Таблиця 3.10. Вид таблиці для перевірки медіанного критерію.

| | Вибірка | | | | Всього |
|---|---|---|---|---|---|
| | 1 | 2 | … | $k$ | |
| Число спостережень, яке більше медіани | $L_1$ | $L_2$ | ... | $L_k$ | $\sum_{j=1}^{k} L_j$ |
| Число спостережень, яке менше медіани | $n_1$- $L_1$ | $n_2$- $L_2$ | ... | $n_k$- $L_k$ | $n-\sum_{j=1}^{n-k} L_j$ |
| Всього | $n_1$ | $n_2$ | ... | $n_k$ | $n$ |

Обчислюють очікувану кількість спостережень для кожної клітинки (при виконанні нульової гіпотези). Для цього елемент рядка „Всього" (відповідний даній клітинці) умножають на елемент стовпця „Всього" (цієї ж клітки) і ділять на загальну кількість спостережень $n$. Потім обчислюють розрахункове значення критерію за формулою:

$\chi^2$=∑(спостережуване значення – очікуване значення)$^2$/очікуване значення.

Сума знаходиться за всіма $2k$ клітинками таблиці.

Очікуване значення для першого рядка (у якому знаходиться значення $L_i$) визначається за формулою:

$$k_{1i} = \frac{n_i\left(\sum_{j=1}^{k} L_j\right)}{n}.$$

Для другого рядка, в якому знаходиться значення ($n_i$–$L_i$), очікуване значення визначають за формулою:

$$k_{2i} = \frac{(n_i - L_i)\left(n - \sum_{j=1}^{k} L_j\right)}{n}.$$

Якщо розраховане значення $\chi^2$ більше верхнього критичного значення розподілу $\chi^2$, взятого з вибраним рівнем значущості $\alpha$ і числом степенів свободи ($k$-1), то гіпотеза про рівність середніх відкидається.

### 3.3.8. Двовибірковий критерій Уїлкоксона

Призначення. Перевірка гіпотези про рівність середніх двох незалежних вибірок.

Нульова гіпотеза. Обидві вибірки мають однаковий розподіл, тобто взяті з однієї генеральної сукупності. Внаслідок чого є рівність двох середніх.

Передумови. Всі випадкові величини взаємно незалежні.

Короткі теоретичні відомості. Аналізовані вибірки можуть бути різних розмірів – $n_1$ і $n_2$, причому $n_1 \le n_2$. Якщо ж $n_1 > n_2$, то вибірки треба поміняти місцями. Для подальшого аналізу ці вибірки об'єднують в одну, розмір якої буде $n_1+n_2$. для об'єднаної вибірки будують ранги і розраховують значення суми рангів, відповідних обом вибіркам, – $R_1$ і $R_2$.

Якщо розміри вибірок однакові, то критеріальне значення $W$ дорівнюватиме одній з сум рангів – $R_1$ або $R_2$. Якщо ж вибірки не однакові, то як критеріальнене значення буде сума рангів, відповідних меншій вибірці ($W_{спост}=R_1$).

Далі необхідно прийняти або відкинути нульову гіпотезу, скориставшись одним з трьох способів.

Спосіб 1. Обчислене критеріальне значення $W_{спост}$ порівнюється з верхнім $(W(Q,n_1,n_2))$ і нижним $(w(Q,n_1,n_2))$ критичними значеннями статистики Уїлкоксона. Для малих значень вибірок існують спеціальні таблиці. Але їх можна досить точно розрахувати самостійно. Розрахунок критичного значення статистики $W$ детально розглянуто в п.3.3.2.

Гіпотеза про рівність середніх відкидається на рівні значущості $Q$, якщо розраховане критеріальне значення $W_{спост} \le w(\alpha,n_1,n_2)$ або $W_{спост} \ge W(\alpha,n_1,n_2)$ де $\alpha=Q/2$.

Спосіб 2. Даний спосіб краще застосовувати для великих вибірок, розмір кожної з яких перевищує 25. В цьому випадку від обчисленого значення $W_{спост}$ переходить до $W^*$, яке розраховується по формулі:

$$W^* = \frac{2W_{спост} - n_1(n_1+n_2+1)+1}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{3}}}.$$

Якщо обчислене критеріальне значення $|W_{спост}|$ більше значення стандартного нормального розподілу $z_\alpha$, знайденого для рівня значущості $\alpha=Q/2$, то нульова гіпотеза про рівність відкидається на рівні значущості $Q$.

Треба пам'ятати, що цей спосіб можна використовувати тільки в тому випадку, якщо розмір хоч би однієї з вибірок перевищує 25 (при нерівних вибірках). Якщо розміри вибірок однакові, то дана апроксимація добре працює до $n_1=n_2=8$.

Якщо у вибірках є однакові значення (зв'язки), то критеріальне значення $W^*$ розраховують за формулою:

$$W^* = \frac{2W_{спост} - n_1(n_1 + n_2 + 1) + 1}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12} - \frac{\sum_{i=1}^{k} t_i (t_i^2 - 1)}{(n_1 + n_2)(n_1 + n_2 + 1)}}},$$

де $k$ – число тих груп зв'язок, в які входять ранги як однієї так і другої вибірок; $t_i$ – розмір $i$-тої зв'язки.

Спосіб 3. Цей спосіб пов'язаний з обчисленням найменшого рівня значущості $p_{гр}$. Найменшим рівнем значущості називають такий рівень, вище якого можна прийняти нульову гіпотезу, якщо вона правильна. Для сформульованої нульової гіпотези про рівність середніх найменший рівень значущості обчислюють за формулою:

$$p_{гр}=2(1-\Phi(|W^*|)),$$

де $\Phi(u)$ – функція нормального стандартного розподілу (функція Лапласа). Якщо обчислений найменший рівень значущості дуже малий, то нульову гіпотезу треба відкинути.

## 3.3.9. Критерій Уїлкоксона для парних спостережень (одновибірковий критерій Уїлкоксона, критерій знакових рангів).

Призначення. Це аналог $t$-критерия для парних спостережень в разі нечислових даних або закону розподілу, що відрізняється від нормального. Застосовують для пов'язаних пар спостережень (перевірка відсутності ефекту обробки).

Нульова гіпотеза. Розподіл різниці пар симетричний щодо нуля.

Альтернативна гіпотеза. Нульова гіпотеза не правильна.

Передумови. Різниці незалежні, вони належать до безперервної сукупності (необов'язково одній і тій же), що симетрична відносно нуля.

Короткі теоретичні відомості.

Для використання критерію необхідно:

– розрахувати значення різниць пар двох вибірок;

– проранжирувати абсолютні значення різниць пар в зростаючому порядку;

– призначити їм відповідні значення рангів від 1 і вище. Нульові значення різниць не розглядуються (звідси – зменшення загальної кількості спостережень $N$). Загалом, побудову рангів виконують як завжди (див.3.1), тільки знаки значень різниць при цьому не враховують;

– за знак рангу вважати знак відповідної йому різниці;

– порахувати суму R додатних рангів;

– обчислити критеріальне значення:

$$T = \frac{\frac{R - N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}};$$

– перевірити критерій. Для малих вибірок ($N$<20), якщо значення $R$ більше табличного значення критерію знакових рангів Уїлкоксона $T$ (не треба плутати із статистикою рангових сум $W$!) з рівнем значущості α/2 і числом степенів свободи

$N$, нульова гіпотеза відкидається. Для великих вибірок значення $T$ порівнюється з верхньою α% точкою стандартного нормального розподілу.

Інші гіпотези.

$H_0$: Середня різниця між значеннями пар двох вибірок дорівнює $A$ (задана константа).

$H_1$: Середня різниця не дорівнює $A$.

В цьому випадку з кожного набутого значення різниці віднімається $A$. Подальша обробка виконується за вже відомою схемою.

Примітка. Даний критерій стійкий до помірних відхилень від прийнятих передумов. Критерій Уілкоксона не треба застосовувати до нормально розподілених сукупностей. Краще використовувати потужніший $t$-критерий. За наявності зв'язок у формулі розрахунку критеріального значення знаменник рекомендується замінити на вираз:

$$\frac{N(N+1)(2N+1)-\dfrac{\sum_{j=1}^{g} t_j(t_j-1)(t_j+1)}{2}}{24},$$

де $g$ – кількість зв'язок, $t_1, t_2, \ , t_g$ – їх розміри.

## 3.3.10. Медіанний одновибірковий критерій (медіана дорівнює константі)

Призначення. Використовується для перевірки припущення про те, що медіана вибірки має якесь певне значення. Наприклад, це може бути нормативне значення показника або ж відомо, яким воно має бути у здорових осіб.

Нульова гіпотеза.

$H_0$: медіана = $A$ проти $H_1$: медіана ≠ $A$. Гіпотеза про рівність медіан відкидається, якщо критеріальне значення більше верхньої *100α/2* % точки або менше нижньої *100α/2* % точки стандартного нормального розподілу.

$H_0$: медіана ≤ $A$ проти $H_1$: медіана > $A$. Нульова гіпотеза відкидається, якщо критеріальне значення більш ніж верхня *100α* % точка стандартного нормального розподілу.

$H_0$: медіана ≥ $A$ проти $H_1$: медіана < $A$. Нульова гіпотеза відкидається, якщо критеріальне значення менше нижньої *100α* % точки стандартного нормального розподілу.

Передумови. Спостереження незалежні і належать одній генеральній сукупності.

Опис методу:

– розраховують кількість вибіркових значень (позначимо його $K$), які більше $A$;

– розраховують критеріальне значення за формулою: $T=\dfrac{2K-N}{\sqrt{N}}$ ;

виконують перевірку гіпотези (див. вище).

Якщо кількість спостережень у вибірці менше 25, то як критеріальне значення використовується *K*, і порівнюють його з біноміальним розподілом таким чином.

Для випадку $H_0$: медіана = А проти $H_1$: медіана $\neq A$, гіпотеза про рівність медіан відкидається, якщо критеріальне значення вище, ніж верхня $100\alpha/2$ % точка або менше нижньої $100\alpha/2$ % точки біноміального розподілу з параметрами ($N$; 0.5).

Для випадку $H_0$: медіана $\leq$ А проти $H_1$: медіана $> A$, нульова гіпотеза відкидається, якщо критеріальне значення вище, ніж верхня $100\alpha$ % точка біноміального розподілу з параметрами ($N$; 0.5).

Для випадку $H_0$: медіана $\geq A$ проти $H_1$: медіана $< A$, нульова гіпотеза відкидається, якщо критеріальне значення менше нижньої $100\alpha$ % точки біноміального розподілу з параметрами ($N$; 0.5).

Примітка. При малій кількості експериментів потужність критерію мала. Цей критерій може застосовуватися і при нормальному законі розподілу.

### 3.3.11. Довірчий інтервал для медіани, що заснований на методі Уїлкоксона (Тьюкі)

Призначення. Використовується для парних спостережень, що дозволяє побудувати довірчий інтервал для різниці значень пар.

Опис методу. Визначення довірчого інтервалу з довірчою вірогідністю *p* виконують таким чином:

– знаходять ціле число *K* (при розрахунках необхідно округлити) за формулою:

$$K = n(n+1)/2 + 1 - t_{(1-p)/2,n} \ ,$$

де *n* – кількість пар спостережень; $t_{(1-p)/2,n}$ – процентна точка розподілу Стьюдента;

– розраховують значення $W_l=(z_i+z_j)/2$, де *z* – значення різниці між парами значень по абсолютній величині. $W_l$ – знаходять для всіх комбінацій пар (тобто 1-е з 1, 2, 3 ,…, *n*-тим; потім 2-е з 2, 3 ,…, *n*-тим і так далі);

– отримані значення $W_l$ упорядковують за збільшенням.

– визначають ліву і праву границі довірчого інтервалу. Ліва границя дорівнює елементу ранжируваного ряду $W_l$ за номером *K*, а права – значенню ряду за номером $M+1-K$. При цьому $M=n(n+1)/2$.

### 3.3.12. Довірчий інтервал для зрушення, що заснований на методі Уїлкоксона (Мозеса)

Призначення. Використовують для двох незалежних вибірок, що дозволяє побудувати довірчий інтервал для зрушення, точніше для величини, яка є різницею між значеннями параметрів зрушення в контрольній і експериментальній вибірках.

Опис методу. Визначення довірчого інтервалу з довірчою вірогідністю р виконують таким чином:

– знаходять ціле число $K$ (при розрахунку необхідно округлити) за формулою: $K=n(2m+n+1)/2+1-\omega_{(1-p)/2,m,n}$, де $m$ – кількість спостережень в першій вибірці, а $n$ – кількість спостережень в другій вибірці; $\omega_{(1-p)/2,m,n}$ – верхня процентна точка розподілу Уілкоксона.

– розраховують значення: $U_l=Y_j-X_i$, де $Y_j-X_i$ – значення різниці між значеннями першої і другої вибірок. $U_l$ обчислюють для всіх комбінацій пар (тобто 1-а з 1, 2, 3 ,…, $m$-тою; потім 2-а з 2, 3 ,…, $m$-тою і так далі);

– упорядковують значення $U_l$ за зростанням;

– визначають ліву і праву границі довірчого інтервалу. Ліва границя дорівнює елементу ранжируваного ряду $U_l$ за номером $K$, а права – значенню елементу ряду за номером ($nm+1-K$).

## 3.3.13. Множинні порівняння

Для множинних порівнянь є непараметричні критерії, які виконують ті ж або подібні задачі, що і критерій Шеффе. До них відносяться критерії Пейджа і Холлендера для альтернатив з впорядкуванням і множинні порівняння, що засновані на сумах рангів Фрідмана. Власне те ж задачі, що і метод множинних порівнянь Шеффе, якраз вирішує і останній з них. Його й розглянемо.

Нехай є $k$ вибірок, в кожної із них $n$ спостережень. Тобто є матриця початкових даних, в котрій $n$ рядків і $k$ стовпців.

Будують ранги для кожного рядка матриці. Після цього для кожного стовпця знаходять суму рангів:

$$R_j = \sum_{i=1}^{n} r_{i,j} ,$$

де $r_{i,j}$ – ранг $i$-го спостереження в $j$-тій вибірці. Після цього будують $k(k-1)$ різниць між сумами рангів, що взяті за абсолютною величиною $|R_u-R_v|$,   $u<v$.

Якщо виконується умова $|R_u-R_v|\geq r(\alpha,k,n)$, , то з рівнем значущості $\alpha$ можна вважати, що середні значення вибірок за номерами $u$ і $v$ не однакові. Тут $r(\alpha,k,n)$ задовольняють умовам:

$$P_0(|R_u-R_v|<r(\alpha,k,n))=1-\alpha;$$
$$u \in (1...k-1),\ v \in (u+1...k)$$
$$P_0(|R_u-R_v|<r(\alpha,k,n))=1-\alpha;$$
$$\forall u \in (1...k-1),\ \forall v \in (u+1...k).$$

Існують таблиці значень $r(\alpha,k,n)$. Кожному значенню $r(\alpha,k,n)$ відповідає своє значення $\alpha$. Це пов'язано з тим, що значення $r(\alpha,k,n)$ дискретні і розраховують за комбінаторними формулами для кожного поєднання $n$ і $k$. Значення $\alpha$ при цьому вибирають з фактично наявних в таблиці, які найбільш відповідають за значенням для перевірки гіпотези.