ТЕМА 3
ЗАСТОСУВАННЯ СТАТИСТИЧНИХ КРИТЕРІЇВ ПРИ ОБРОБЦІ ДАНИХ НА КОМП'ЮТЕРІ

1. Загальні відомості
2. Двовибіркові тести
3. Непараметричні тести

1.Загальні відомості

Під *статистичною гіпотезою* розуміють усяке висловлення про генеральну сукупність (випадковій величині), що перевіряється по вибірці (за результатами спостережень). Процедуру зіставлення висловленої гіпотези з вибірковими даними називають *перевіркою статистичної гіпотези.* По прикладному змісті можна виділити наступні основні види висловлюваних у ході статистичної обробки даних гіпотез:

- про тип закону розподілу досліджуваної випадкової величини;
- про однорідність двох чи декількох оброблюваних вибірок або деяких характеристик аналізованих сукупностей;
- про числові значення досліджуваної генеральної сукупності;
- про тип залежності між компонентами досліджуваної багатомірної ознаки;
- про незалежність і стаціонарність оброблюваного ряду спостережень.

Статистична гіпотезу, яку перевіряють, прийнято *називати основною (чи нульовою*) гіпотезою (*позначається* H_0), а суперечну їй гіпотезу - *альтернативною (чи конкуруючою*) гіпотезою (*позначається* H_1).

Оскільки при перевірці статистичних гіпотез приходиться мати справу зі статистичним матеріалом, то, відкидаючи чи приймаючи нульову гіпотезу, завжди ризикуємо зробити помилку.

Помилку, що полягає в тім, що нульова гіпотеза відкидається, тоді як *вона в дійсності вірна, називають **помилкою першого роду**.*

*Помилку, що складається в тім, що нульова гіпотеза не відкидається, тоді як вона в дійсності невірна, називають **помилкою другого роду**.*

Перевірка статистичних гіпотез здійснюється за допомогою різних статистичних критеріїв. Як критерій використовується деяка випадкова величина, значення якої можуть бути обчислені на основі наявних даних. У безлічі можливих значень критерію вибирається підмножина, називана *критичною областю.* Якщо обчислене значення критерію належить критичної області, то нульова гіпотеза відкидається. Критична область вибирається таким чином, щоб ймовірність зробити помилку першого роду не перевершувала деякого заздалегідь визначеного позитивного числа □. Це число □ називають *рівнем значущості* і говорять: "нульова гіпотеза відкидається на рівні значущості □". У якості □ звичайно беруть одне з чисел: 0,05; 0,01; 0,001.

Імовірність зробити помилку другого роду позначається □. Величина 1-□ називається *потужністю критерію;* вона дорівнює імовірності відкинути невірну гіпотезу.

Порівняння середніх значень різних вибірок відноситься до найбільш уживаних методів статистичного аналізу. В таких випадках завжди вирішується питання, чи можливо пояснити відмінності середніх значень статистичними коливаннями чи ні. В останньому випадку відмінності вважаються значимими.

Найчастіше безліч можливих значень критерію належить деякому інтервалу. Інтервалом є і критична область. Граничні точки критичної області називаються *критичними точками.* Критичні точки вибираються таким чином, щоб при обраному рівні значимості □ потужність критерію (1-□) була *найбільшою.*

Можливі три види розташування критичної області (у залежності від виду нульової та альтернативної гіпотез, виду і розподілу статистичного критерію □):

1) правобічна критична область, що складається з інтервалу $\left(\chi^{кр}_{пр\alpha}, +\infty\right)$,

де точка $\chi^{кр}_{пр,\alpha}$ визначається з умови:

$$P\left(\chi^2>\chi^{кр}_{пр,\alpha}\right)=\alpha$$

і називається *правобічною критичною точкою, що* відповідає рівню значимості □;

2) лівостороння критична область, що складається з інтервалу $\left(-\infty;\chi^{кр}_{лев,\alpha}\right)$, де точка $\chi^{кр}_{лев,\alpha}$ визначається з умови:

$$P\left(\chi^2<\chi^{кр}_{лев,\alpha}\right)=\alpha$$

і називається *лівосторонньою критичною точкою, що* відповідає рівню значимості □;

3) двостороння критична область, що складається з наступних двох інтервалів: $\left(-\infty;\chi^{кр}_{лев,\alpha/2}\right)$ і $\left(\chi^{кр}_{пр,\alpha/2};+\infty\right)$, де точки $\chi^{кр}_{лев,\alpha/2}$ і $\chi^{кр}_{пр,\alpha/2}$ визначаються з умов:

$$P\left(\chi^2<\chi^{кр}_{лев,\alpha/2}\right)=\alpha/2 \text{ і } P\left(\chi^2>\chi^{кр}_{пр,\alpha/2}\right)=\alpha/2$$

і називаються *двосторонніми критичними точками.*

При порівнянні середніх значень вибірок виділяють чотири різні тестові ситуації:

- порівняння двох незалежних вибірок,
- порівняння двох залежних вибірок,
- порівняння більше двох незалежних вибірок,
- порівняння більше двох залежних вибірок,

В цих випадках відповідно застосовуються такі статистичні тести:

- *t*-тест для незалежних вибірок (тест Ст'юдента),
- *t*-тест для залежних вибірок,
- однофакторний дисперсійний аналіз,
- однофакторний дисперсійний аналіз з повторними вимірюваннями.

5.2. Двовибіркові тести

5.2.1. Загальні положення

При порівнянні середніх значень вибірок, передбачається, що обидві вибірки відповідають нормальному закону розподілу. В іншому випадку розраховуються медіани та для аналізу використовуються непараметричні тести.

5.2.2. Двовибірковий z-тест для середніх

Розглядається критерій перевірки гіпотези про рівність середніх (математичних сподівань) двох нормальних розподілів з відомими дисперсіями, що знаходить важливе практичне застосування.

Дійсно, іноді виявляється, що середній результат $\bar{x}$ однієї серії спостережень відрізняється від середнього результату іншої серії. Виникає питання: чи можна це розходження пояснити випадковою помилкою експериментів чи ця відмінність не випадкова? Інакше кажучи, чи можна вважати, що результати експериментів являють собою вибірки з двох генеральних сукупностей з однаковими середніми чи середні цих сукупностей не рівні?

Постановка даної задачі формулюється в такий спосіб.

Вважаються нормально розподіленими величини $Y=N(a_y,\sigma_y)$, числові значення дисперсій σ_x^2 σ_y^2 яким відомі, числові значення a_y і

$Y=N(a_y,\sigma_y)$, числові значення дисперсій σ_x^2 і σ_y^2 яким відомі; числові значення середніх a_x і a_y невідомі.

Нехай $x_1, x_2, \ldots, x_n$ - результати незалежних, проведених в однакових умовах спостережень величини X, а $y_1, y_2, \ldots, y_n$ - результати незалежних, проведених в однакових умовах спостережень величини Y.

Потрібно перевірити гіпотезу про рівність математичних очікувань випадкових величин X та Y, тобто гіпотезу:

$$H_0: a_X = a_Y.$$

Якщо гіпотеза $H_0: a_X = a_Y$ приймається, то говорять, що розходження вибіркових середніх $\bar{x}$ і $\bar{y}$ *статистично незначуще.*

У математичній статистиці доводиться, що якщо дана гіпотеза виконується, то величина

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$$

має нормальний закон розподілу з нульовим математичним сподіванням і одиничною дисперсією, тобто

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} = N(0,1).$$

Величину z використовують як критерій при перевірці гіпотези

$$H_0 : a_X = a_Y.$$

5.2.3. Двохвибірковий t-тест з однаковими і різними дисперсіями

Розглянемо процедуру перевірки гіпотез про рівність середніх (математичних сподівань) двох нормальних розподілів з невідомими дисперсіями.

Щодо параметрів σ_x^2 і σ_y^2 можна висунути такі два припущення:

1) обидві дисперсії невідомі, але передбачається, що вони рівні між собою ($\sigma_x^2 = \sigma_y^2$);

2) обидві дисперсії невідомі, їхня рівність не передбачається ($\sigma_x^2 \neq \sigma_y^2$).

У випадку коли обидві дисперсії невідомі, але передбачаються рівними між собою, маємо справа з двома оцінками $s_x^2 = s_y^2$ однієї і тієї ж величини дисперсії $\sigma_x^2 = \sigma_y^2$. У зв'язку з цим розумно перейти до об'єднаної оцінки σ^2:

$$s^2 = \frac{s_X^2(n-1) + s_Y^2(m-1)}{(n-1)+(m-1)}.$$

У математичній статистиці доводиться, що якщо гіпотеза $H_0 : a_X = a_Y$ виконується, тоді величина

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{(1/n)+(1/m)}}$$

має розподіл Ст'юдента з $k = n + m - 2$ ступенями свободи, тобто

$$\frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n} + \frac{1}{m}}} = t(k = n + m - 2).$$

Величину t і використовують як критерій при перевірці гіпотези $H_0 : a_X = a_Y$.

Коли дисперсії невідомі і їхня рівність не передбачається ($\sigma_x^2 \neq \sigma_y^2$), використовується аналог *z-статистики* з заміною невідомих дисперсій їхніми оцінками:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}.$$

У цій ситуації вказати точний розподіл уведеної статистики важко. Відомо, однак, що цей розподіл близький до розподілу Ст'юдента з числом ступенів свободи, рівним:

$$k = \frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2}{\frac{\left(\frac{s_X^2}{n}\right)^2}{n-1} + \frac{\left(\frac{s_Y^2}{m}\right)^2}{m-1}}.$$

Останній статистичний критерій (при $\sigma_x^2 \neq \sigma_y^2$) називають також *критерієм Фішера-Беренса.* Даний критерій не часто застосовують на практиці, тому що, навіть коли невідомі дисперсії σ_x^2 та σ_y^2 істотно різні, припущення, що вони насправді рівні, дає результати, досить близькі до одержуваного за цим критерієм, але при набагато меншому обсязі обчислень.

5.2.4 Парний двохвибірковий t-тест для середніх

Процедури порівняння двох вибірок часто застосовуються для виявлення результату деякого впливу або, навпроти, для підтвердження його відсутності. Чим більш однорідними виявляться обрані для експерименту об'єкти (для контролю і впливу), чим меншими є їхні випадкові розходження, тим точніше можна буде дати відповідь на поставлене питання. Ясно, що розходження між об'єктами, обраними для дії і для контролю (чи для двох різних впливів, якщо інтерес представляє їхнє зіставлення), буде найменшим, якщо в обох якостях виступає той самий об'єкт. Якщо це можливо, то далі звичайним порядком складається група експериментальних об'єктів і потім для кожного об'єкта виміряються два значення досліджуваної характеристики наприклад, до впливу і після чи при двох різних впливах). Так виникають *пари спостережень* чи *парні дані*.

Нехай x_i та y_i — результати вимірів для об'єкта номер i, $i = 1, ..., n$, де n — чисельність експериментальної групи (число об'єктів). Тоді сукупність пар випадкових величин $(x_1, y_1), ..., (x_n, y_n)$ утворить парні дані.

Передбачається, що всі спостереження є реалізаціями випадкових величин і методика експерименту забезпечує їхню незалежність для різних об'єктів. Але спостереження, що входять в одну пару, не можна вважати незалежними, оскільки вони відносяться до тому самому об'єкта. Ці два спостереження відбивають властивості загального для них індивідуального об'єкта і тому можуть залежати друг від друга.

Для пар спостережень (x_i, y_i) введемо величину $z_i = y_i - x_i$, яку будемо вважати незалежною і нормально розподіленою. Тим самим задача про парні дані зводиться до задачі *про одну нормальну вибірку при невідомій дисперсії.*

При невідомому числовому значенні дисперсії σ_Z^2 в основу перевірки гіпотези $H_0: a_Z = a_0$, де a_0 - заздалегідь задане число, покладений критерій:

$$t = \frac{\bar{z} - a_0}{\frac{s}{\sqrt{n}}},$$

що при виконанні гіпотези $H_0: a_Z = a_0$ має t-розподіл з числом ступенів свободи $k = n - 1$, тобто

$$\frac{\bar{z} - a_0}{\frac{s}{\sqrt{n}}} = t(k = n - 1).$$

5.2.5. Двовибірковий F-тест для дисперсій

Двовибірковий F-тест для дисперсій використовується для перевірки гіпотези про рівність дисперсій двох нормальних розподілів.

Відзначимо, що ця задача має, і самостійне значення. Дисперсія характеризує спрямованість способів впливу, точність роботи і т.д.. Переконавшись в рівності двох дисперсій, ми тим самим переконуємося, наприклад, у тім, що обидва способи впливу мають однакову спрямованість.

У математичній статистиці доводиться, що якщо гіпотеза $H_0: \sigma_X^2 = \sigma_Y^2$ виконується, то величина

$$F = \frac{s_X^2}{s_Y^2}$$

має *F-розподіл* з $k = n - 1$ і $l = m - 1$ числом ступенів свободи, тобто

$$\frac{s_X^2}{s_Y^2} = F(k = n - 1;\ l = m - 1).$$

Величину *F*, яку називають дисперсійним відношенням Фішера (чи просто статистикою Фішера), використовують як критерій при перевірці гіпотези про рівність дисперсій $H_0: \sigma_X^2 = \sigma_Y^2$.

Оскільки величина *F* є ненегативною, критична область даної величини буде належати інтервалу (0; +∞).

5.3. Непараметричні тести

5.3.1.Загальні відомості

Непараметричні тести застосовуються в тих випадках, коли вибірки з інтервальними змінними не відповідають нормальному закону розподілу, або коли змінні відносяться до порядкової шкали. Одним з факторів, що обмежують застосування критеріїв, заснованих на припущенні нормальності, є об'єм вибірки. Доти поки вибірка досить велика (наприклад, 100 чи більше спостережень), можна вважати, що вибірковий розподіл нормальний, навіть якщо ви не упевнені, що розподіл змінної в популяції є нормальним. Проте, якщо вибірка *мала,* ці критерії варто використовувати тільки при наявності впевненості, що перемінна дійсно має *нормальний* розподіл. Однак немає способу перевірити це припущення на *малій* вибірці.

5.3.2. З’ясування відмінностей у рівнях досліджуваної ознаки

5.3.2.1. Обґрунтування задачі співставлення та порівняння

При дослідженнях залежностей часто виникає задача порівняння за отриманим результатом деяких різних способів впливу, які направлені на досягнення однієї мети. В таких випадках потрібно на основі тих або інших даних вирішити чи є справедливим певне судження.

Однією з характерних задач такого класу є задача порівняння двох вибіркових сукупностей. Наприклад, який з двох раціонів харчування забезпечує максимальну працездатність? Яка з двох методик професійної підготовки забезпечує вищу якість освіти? Які з двох ліків забезпечують швидше одужання?

Для дослідження такого типу необхідно мати однорідні об'єкти поділені на дві групи. Взаємний вплив та взаємодія об'єктів повинні бути виключені. Для кожного об'єкту реєструється його чисельна характеристика. Дві групи чисел, які при цьому отримують, можливо розглядати як дві незалежні вибірки.

5.3.2.2. Критерій Манна-Уітні.

Область застосування критерію Манна-Уітні — аналіз двох незалежних вибірок, розміри яких можуть розрізнятися.

Призначення критерію — перевірка гіпотези про статистичну однорідність двох вибірок. Іноді цю гіпотезу називають гіпотезою про відсутність ефекту обробки (маючи на увазі, що одна з вибірок містить характеристики об'єктів, які зазнали деякий вплив, інша - характеристики контрольних об'єктів).

Дані. Розглядаються дві вибірки $x_1,...,x_m$ (вибірка *x*) і $y_1,...,y_n$ (вибірка у) обсягів m і n. Позначимо закон розподілу першої вибірки через F, а другої — через G.

Припущення. Вибірки повинні бути незалежними з неперервними законами розподілу.

Гіпотеза. Твердження про однорідність вибірок $x_1,...,x_m$ і $y_1,...,y_n$ у введених вище позначеннях можна записати у вигляді $H_0: F = G$.

Альтернативи. Як альтернативи до $H_0: F = G$ можуть виступати всі можливості $F \neq G$. Однак критерій Манна-Уітні здатен виявляти аж ніяк не всі можливі відхилення від $H_0: F = G$. Цей критерій призначений, у першу чергу, для перевірки проти альтернативи $F \geq G$ (правобічна альтернатива) або альтернативи $F \leq G$ (лівостороння альтернатива).

Метод. Критерій Манна-Уітні заснований на попарному порівнянні результатів з першої і другої вибірок. При цьому всяка подія $x_i < y_j$, позначає "успіх", а всяка подія $x_i > y_j$ — "невдачу". Зміст такої термінології може бути зв'язаний з тим, що передбачається, що друга група краще першої. Змінюючи i від 1 до m і j від 1 до n, одержуємо mn парних порівнянь елементів вибірок x і y. Позначимо число успіхів у цих парних порівняннях через U. Ясно, що U може приймати будь-яке ціле значення від 0 до mn.

Визначення. *Уведена вище випадкова величина U називається статистикою Манна-Уітні.*

Обчисливши значення $U_{\text{набл}}$ можливо приступити до перевірки гіпотези $H_0: F = G$:

1. Задається рівень значимості α чи вибирається метод, пов'язаний з визначенням найменшого рівня значимості статистики U.

2. Для правобічних альтернатив знаходиться по таблицях таке критичне значення $U_{\text{п}}(\alpha, m, n)$, що

$$P\{U \geq U_{\text{п}}(\alpha, m, n)\} \leq \alpha.$$

При цьому критична область для гіпотези $H_0: F = G$ проти правобічних альтернатив буде мати вигляд:

$$\{U \geq U_{\text{п}}(\alpha, m, n)\}.$$

При перевірці $H_0: F = G$ проти лівосторонніх альтернатив треба знайти критичне значення $U_{\text{л}}(\alpha, m, n)$, таке, що

$$P\{U \leq U_{\text{л}}(\alpha, m, n)\} \leq \alpha.$$

Тут критична область прийме вигляд:

$$\{U \leq U_{\text{л}}(\alpha, m, n)\}.$$

У таблицях звичайно приводяться критичні значення, що відповідають числам α з ряду 0,05; 0,025; 0,01; 0,005; 0,001. Через дискретний характер розподілу ймовірностей між можливими значеннями випадкової величини U, наведені вище рівняння не завжди мають точне рішення, і в таблицях вони приводяться приблизно. Для обчислення по таблицях значень $U_л(\alpha,m,n)$ можна скористатися співвідношенням

$$U_л(\alpha,m,n)+U_п(\alpha,m,n)=mn,$$

що випливає із симетрії розподілу статистики U щодо свого центру $mn/2$.

3. Відкинемо гіпотезу $H_0: F=G$ проти правобічних (лівосторонніх) альтернатив при влученні $U_{набл}$ у відповідну критичну область.

4. При перевірці $H_0: F=G$ проти двосторонніх альтернатив як критичну множину можна взяти об'єднання

$$\{U \le U_л(\alpha,m,n)\} \cup \{U \ge U_п(\alpha,m,n)\},$$

тобто відкинути $H_0: F=G$, якщо відбувається одне з двох раніше згаданих критичних подій. Через уже відзначену симетрію цьому критерію можна надати вигляд

$$\left|U-\frac{mn}{2}\right| \ge U_п(\alpha,m,n)-\frac{mn}{2}.$$

При такому виборі критичної множини рівень значимості подвоюється. Тепер він дорівнює 2α (з тими ж застереженнями щодо дискретності розподілу U, що були зроблені вище).

Розподіл ймовірностей U ***при гіпотезі*** $H_0: F=G$. Хоча статистика Манна-Уітні є сумою однаково розподілених випадкових величин, що приймають значення 0 і 1, вона не має біноміального розподілу. Тому розподіл статистики U приходиться розраховувати, використовуючи спеціальні таблиці або асимптотичне наближення.

Розрахунок розподілу статистики U значно спрощується тим, що при виконанні гіпотези $H_0: F = G$ цей розподіл не залежить від закону розподілу вибірок (якщо ці розподіли безупинні), а тільки від обсягів вибірок — m і n. У довідниках приводяться таблиці, по яких можна знайти імовірність $P(U \geq k)$ для різних k при невеликих значеннях m і n. При справедливості гіпотези $H_0: F = G$ (тобто при збігові законів розподілів F і G виконується $P(x_i > y_j) = P(x_i < y_j) = 0{,}5$. Тому при $H_0: F = G$ кількості успіхів і невдач повинні бути приблизно рівні, тобто U не повинно значно відхилятися від $mn/2$.

Розподіл статистики U при порушенні гіпотези. На відміну від поводження U при гіпотезі $H_0: F = G$, у цьому випадку розподіл U залежить від F і G, тому можливо описати його властивості лише для окремих типів альтернатив. Найпростіше указати властивості U для однобічних альтернатив: правобічних (якщо $F \geq G$), чи лівосторонніх (якщо $F \leq G$). Легко бачити, що для правобічних альтернатив виконується $P(x_i < y_j) > 0{,}5$, тому значення U, тобто загальне число успіхів $x_i < y_j$, швидше за все, повинне перевершувати $mn/2$ і тем значніше, чим більше $P(x_i < y_j)$. Для лівосторонніх альтернатив ($F \leq G$) співвідношення зворотне: $P(x_i < y_j < 0{,}5$, тому загальне число успіхів, як правило, повинне бути менше $mn/2$, і тем менше, чим менше $P(x_i < y_j)$.

Отже, для однобічних альтернатив статистика Манна-Уітні має ясні властивості, тому на її основі можна побудувати критерій для перевірки гіпотези $H_0: F = G$ проти таких альтернатив.

Метод перевірки гіпотези. У зв'язку з таким поводженням статистики

U для перевірки гіпотези $H_0: F = G$ проти зазначених вище можливих альтернатив розумно запропонувати наступне правило: відкинути $H_0: F = G$, якщо спостережене $U_{\text{набл}}$ значно відхиляється від $mn/2$ — значення, очікуваного від U при гіпотезі $H_0: F = G$ (від математичного чекання U при гіпотезі $H_0: F = G$). Чим більше відхиляється від $mn/2$ спостережене значення $U_{\text{набл}}$, тим сильніше сумнів у тім, що $H_0: F = G$ вірна. Зрозуміло, U може значно відхилятися від $M(U \mid H_0)$ і за рахунок дії, коли $H_0: F = G$ виконується, але чим більше відхилення, тим воно при $H_0: F = G$ менш імовірно, і тем сутужніше пояснити це відхилення випадковістю. Швидше за все, якщо відхилення велике, воно викликано не випадком, а закономірною причиною — тим, що розподілу F і G не збігаються.

Силу таких доказів проти $H_0: F = G$ на користь, наприклад, правобічної альтернативи $F \geq G$ можна виразити кількісно, обчисливши $P(U \geq U_{\text{набл}} \mid H_0)$. Це імовірність того, що при незалежному повторенні експерименту ми одержимо таке ж чи ще більш сильне свідчення проти $H_0: F = G$ (на користь правобічної альтернативи), що як уже мається) $U_{\text{набл}}$. Якщо $U_{\text{набл}}$ велике, то вищезгадана імовірність мала, і навпаки. Якщо ця імовірність настільки мала, що подібна подія здається практично неможливим при $H_0: F = G$, гіпотезу $H_0: F = G$ варто відкинути (за наявним спостереженням $U_{\text{набл}}$), на користь правобічної альтернативи.

Рекомендація змінюється очевидним образом, якщо з $H_0: F = G$ конкурують лівосторонні альтернативи. Нарешті, у випадку двосторонніх альтернатив треба обчислити імовірність

$$P\left\{\left|U - \frac{mn}{2}\right| \geq \left|U_{\text{набл}} - \frac{mn}{2}\right|\right\}$$

і в залежності від того, наскільки вона мала, відкинути гіпотезу.

Описаний спосіб дій має визначені переваги перед стандартною процедурою перевірки статистичних гіпотез. **Головне те, що тут не приходиться заздалегідь вибирати рівень значимості, що завжди виглядає трохи довільно.** Описаний підхід автоматично доставляє нам той найменший рівень значимості, на якому (за наявними спостереженнями) можна відкинути гіпотезу $H_0 : F = G$ на користь відповідної альтернативи. У даному випадку є і ще одна додаткова перевага: через дискретність розподілу традиційні номінальні рівні значимості типу 0,05; 0,025; 0,001 і т.д. можуть бути досягнуті лише приблизно. В методі перевірки наближення зникає: виходить точне значення імовірності, якщо звернеться до досить докладних таблиць розподілів U.

Збіги. З умови безперервності розподілів F і G випливає відсутність повторень у вибірках. На практиці ж такі повторення зустрічаються часто. У багатьох випадках причиною цього є не порушення вихідних припущень, а обмежена точність при записі спостережень.

Допустимо, що деякі елементи вибірки x_i збіглися з деякими елементами з вибірки y_j, тобто $x_i = y_j$ для деяких i, j ($i = 1,\ldots,m$; $j = 1,\ldots,n$). У цьому випадку статистику U обчислюють так: до числа успіхів додають зменшене вдвічі число подій виду ($x_i = y_j$). Таким чином, кожен збіг x_i і y_j вважається за половину успіху.

При наявності співпадаючих спостережень одержувані при використанні описаних критеріїв висновки мають наближений характер, і ці наближення тим гірше (і висновки тим сумнівніше), чим більше серед спостережень співпадаючих, тобто чим сильніше відступ від вихідних математичних припущень. У тих випадках, коли результати можуть приймати лише обмежене число значень (що спричиняє велику кількість збігів), цей метод застосовувати не слід.

5.3.2.3. Критерій Краскела-Уоллеса.

Якщо ми не можемо сказати що-небудь визначене про альтернативи до H_0, можна скористатися для її перевірки вільним від розподілу *критерієм Краскела-Уоллеса.* Для цього замінимо спостереження x_{ij} їхніми рангами r_{ij}, упорядковуючи всю сукупність $\| x_{ij} \|$ у порядку зростання (для визначеності). Потім для кожної обробки j (тобто для кожного стовпця вихідної таблиці) треба обчислити

$$R_j=\sum_{i=1}^{n_j} r_{ij} \text{ і } R_{\cdot j}=\frac{1}{n_j}\sum_{i=1}^{n_j} r_{ij},$$

де $R_{\cdot j}$ — це середній ранг, розрахований по стовпцю. Якщо між стовпцями немає систематичних розходжень, середні ранги $R_{\cdot j}$, $j=1,\ldots,k$ не повинні значно відрізнятися від середнього рангу, розрахованого по всій сукупності $\| r_{ij} \|$. Ясно, що останній дорівнює $(N+1)/2$. Тому величини

$$\left(R_{\cdot 1}-\frac{N+1}{2}\right)^2, \ldots, \left(R_{\cdot k}-\frac{N+1}{2}\right)^2$$

при H_0 в сукупності повинні бути невеликими. Складаючи загальну характеристику, розумно врахувати розходження в числі спостережень для різних обробок і взяти як міру відступу від чистої випадковості величину

$$H=\frac{12}{N(N+1)}\sum_{j=1}^{k} n_j\left(R_{\cdot j}-\frac{N+1}{2}\right)^2 .$$

Ця величина називається *статистикою Краскела-Уоллеса.* Множник $12[N(N+1)]^{-1}$ потрібний для стабілізації її розподілу при великому числі спостережень. Інша форма для обчислення:

$$H=\frac{12}{N(N+1)}\sum_{j=1}^{k}\frac{R_j^2}{n_j}-3(N+1).$$

Якщо є співпадаючі значення, треба при ранжуванні використовувати середні ранги. Якщо збігів багато, рекомендують використовувати модифіковану форму статистики H':

$$H'=\frac{H}{1-\left(\frac{\sum\limits_{j=1}^{g}T_j}{N^3-N}\right)},$$

де g - число груп співпадаючих спостережень; $T_j=t_j^3-t_j$, t_j - число співпадаючих спостережень у групі з номером *j*.

5.3.2.4. Критерій Джонкхієра

Нерідко заздалегідь відомо, що наявні групи результатів упорядковані по зростанню впливу фактора. У таких випадках можна використовувати критерій Джонкхієра, більш чуттєвий (більш могутній) проти альтернатив про упорядкований вплив фактора. Однак проти інших альтернатив властивості цього критерію можуть виявитися гірше властивостей критерію Краскела-Уоллеса.

Статистика Джонкхієра є узагальненням статистики Манна-Уітні на k способів обробки. Звернувшись до такого загального випадку, порівняємо k способів обробки. Для кожної пари натуральних чисел u і v, де $1\le u<v\le k$, складаємо по вибірках з номерами u,v статистику Манна-Уітні:

$$U_{uv}=\sum_{\substack{i=1..n_u\\ j=1..n_v}}\varphi(x_{iu},y_{jv}).$$

Статистику Джонкхієра J визначають як

$$J = \sum_{1 \le u < v \le k} U_{uv} .$$

Свідченням на користь альтернативи упорядкованості ефектів (проти гіпотези однорідності) служать великі значення статистики J, отримані в експерименті.

При невеликих обсягах вибірок і невеликим k розподіл статистики J табульовано. Для великих вибірок у відношенні J діє нормальна апроксимація:

$$J \approx N(M, D),$$

де $M = \frac{1}{4}\left(N^2 - \sum_{j=1}^{k} n_j^2\right)$, $D = \frac{1}{72}\left[N^2(2N+3) - \sum_{j=1}^{k} n_j^2(2n_j+3)\right]$.

Свідченням проти гіпотези однорідності служать великі (порівняно з процентними крапками стандартного нормального розподілу) значення статистики $(J-M)/\sqrt{D}$, отримані в експерименті.

5.3.3. Оцінювання достовірності змін у значеннях досліджуваної ознаки

5.3.3.1. Обґрунтування задачі дослідження змін

Порівняння сукупностей спостережень (вибірок) часто проводиться для виявлення результату якого-небудь впливу (виявлення ефекту обробки), або навпроти, для підтвердження його відсутності. Ніж більш однорідними виявляться обрані для експерименту об'єкти (для контролю і впливу), чим менше їхні випадкові розходження, тим точніше (і по меншому числу спостережень) можна буде дати відповідь на питання. Формування однорідної групи експериментальних об'єктів складає важливу і не завжди просту задачу.

5.3.3.2. Критерій знаків

Парні спостереження. Порівняння двох сукупностей спостережень (двох вибірок) часто проводиться для виявлення результату якого-небудь впливу (виявлення ефекту обробки), або навпроти, для підтвердження його відсутності. Розходження між об'єктами, обраними для впливу і для контролю (чи для двох різних впливів, якщо інтерес представляє їхнє зіставлення) буде найменшим, якщо в обох якостях виступає той самий об'єкт. Якщо це можливо, то далі звичайним порядком складається групу експериментальних об'єктів (як і раніше прагнучи до того, щоб вони були однорідні). Далі для кожного об'єкта виміряється два значення необхідної характеристики (наприклад, до впливу і після чи при двох різних впливах). Так виникають пари спостережень і парні дані.

Призначення. Критерій знаків використовується для перевірки гіпотези про однорідність спостережень усередині кожної пари (іноді говорять — для перевірки гіпотези про відсутність ефекту обробки).

Дані. Розглянемо сукупність випадкових пар (x_1, y_1), …, (x_n, y_n) обсягу n. Уведемо величини $z_i = x_i - y_i$, $i = 1,...,n$.

Припущення. 1. Усі z_i передбачаються взаємно незалежними. При цьому не потрібно незалежності між елементами x_i і y_i з однаковим номером. Це дуже важливо на практиці, коли спостереження робляться для одного об'єкта і тим самим можуть бути залежні. 2. Усі z_i мають рівні нулю медіани, тобто $P(z_i < 0) = P(z_i > 0) = \frac{1}{2}$. При цьому розподілу різних z_i можуть не збігатися.

Гіпотеза. Твердження про відсутність ефекту обробки для повторних парних спостережень (x_1, y_1), …, (x_n, y_n) можна записати у виді

$$H_0: P(z_i < 0) = P(z_i > 0) = \frac{1}{2} \text{ для усіх } i = 1,...,n.$$

Метод. 1. Перейдемо від повторних парних спостережень (x_1, y_1), …, (x_n, y_n) до величин $z_i = x_i - y_i$, $i = 1,...,n$.

2. До сукупності $z_i = x_i - y_i$, $i = 1,...,n$ застосовується критерій знаків для перевірки гіпотези про рівність нулю медіан розподілів величин $z_i = x_i - y_i$, $i = 1,...,n$.

Зв'язані дані. Якщо серед значень z_i є нульові, то їх варто відкинути і відповідно зменшити n до числа ненульових значень z_i .

Оцінка ефекту обробки. Нерідко для z_i, розглядають моделі $z_i = \theta + e_i$, $i = 1,...,n$,де θ - деяка константа яка характеризує положення одного розподілу щодо іншого (скажемо, до впливу і після), e_i — випадкові величини, що не спостерігаються. Цю константу часто іменують ефектом обробки. Прийняті допущення 1 і 2 переносяться на величини $e_1,...,e_n$. Гіпотеза однорідності формулюється у виді гіпотези про нульовий ефект обробки $H_0 : \theta = 0$

Уведені величини θ і представлення $z_i = \theta + e_i$, виявляються корисними, якщо в ході перевірки гіпотези з'ясовується, що $\theta \neq 0$ і що тому треба оцінити кількісно те розходження, що привносить обробка (вплив).

Одне з головних достоїнств критерію знаків — його простота. Іншою важливою особливістю цього критерію є скромні вимоги до первісного статистичного матеріалу. Ці вимоги описуються за допомогою моделі парних спостережень.

5.3.3.3. Критерій Вілкоксона

Призначення. Критерій Вілкоксона використовується для перевірки гіпотези про однорідність двох вибірок. Нерідко одна з вибірок представляє характеристики об'єктів, які піддані перед тим якомусь впливу (обробці). У цьому випадку гіпотезу однорідності можна назвати *гіпотезою про відсутність ефекту обробки.*

5.3.3.4. Критерій Фрідмана

Призначення. Непараметричний критерій Фрідмана для перевірки гіпотези H_0 проти альтернативи про наявність впливу фактора A використовується у випадку, якщо про розподіл випадкових величин e_{ij}, $i=1,\ldots,n$, $j=1,\ldots,k$ відомо тільки те, що воно безупинно, а самі величини e_{ij} незалежні в сукупності. Критерій заснований на ідеї переходу від значень величин e_{ij} до їхніх рангів. На відміну від однофакторного аналізу, ранжирування відбувається не по всій сукупності величин e_{ij}, а поблочно, тобто розглядаються дані при фіксованому індексі i і здійснюється ранжирування величин x_{ij} при $j=1,\ldots,k$. Тим самим усувається вплив фактора B, "що заважає", значення якого для i індексу постійно.