

**МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ВНУТРІШНІХ СПРАВ**

Факультет № 4

Кафедра інформаційних технологій

**НАВЧАЛЬНО-МЕТОДИЧНІ ВКАЗІВКИ
ДО САМОСТІЙНОЇ РОБОТИ**

з дисципліни «МАТЕМАТИЧНІ МЕТОДИ В ПСИХОЛОГІЇ»

Галузь знань	05 Соціальні та поведінкові науки
Спеціальність	053 "Психологія"
Ступень вищої освіти	Бакалавр
Форма навчання	денна, заочна

**м. Харків
2018 рік**

Передмова

СХВАЛЕНО

Науково-методичною радою ХНУВС

_____ Протокол № _____

(дата, місяць, рік)

ЗАТВЕРДЖЕНО

Вченою радою факультету № 4

ХНУВС

_____ Протокол № _____

(дата, місяць, рік)

_____ (підпис)

_____ (П.І.Б.)

ПОГОДЖЕНО

Секцією науково-методичної ради
ХНУВС з технічних дисциплін

_____ Протокол № _____

(дата, місяць, рік)

_____ (підпис)

_____ (П.І.Б.)

ЗАТВЕРДЖЕНО

На засіданні кафедри інформаційних
технологій

_____ Протокол № _____

(дата, місяць, рік)

_____ (підпис)

_____ (П.І.Б.)

Рецензенти:

Гнусов Ю.В., кандидат технічних наук, доцент, начальник кафедри кібербезпеки факультету № 4 Харківського національного університету внутрішніх справ

Яськов Г.М., кандидат технічних наук, доцент, науковий співробітник відділу математичного моделювання та оптимального проектування Інституту проблем машинобудування ім. А.М. Підгорного НАН України

Розробники: Мелашенко О.П., Шеховцов С.Б. - Харків, Харківський національний університет внутрішніх справ, 2018

© Мелашенко О.П., Шеховцов С.Б., 2018

© Харківський національний університет внутрішніх справ

1. Загальні методичні вказівки

«Зрілість науки звичайно вимірюється тим, якою мірою вона використовує математику. Сама ж математика не є наукою в емпіричному значенні, але являє собою формальну логічну, символічну систему, свого роду гру знаків і правил» - так починає С.С.Стивенс свою капітальну працю «Експериментальна психологія», що зробила великий вплив на становлення психології не тільки за кордоном, але й у нашій країні. Як же психологи використовують математику?

З історії психології добре відомо, що наприклад, психофізика почала свій розвиток із установа математичних закономірностей (формула Вебера-Фехнера). В даний час математичні процедури обов'язково входять у такі розділи психології як психометрика, психодіагностика, диференціальна психологія. Сучасна психогенетика, наприклад, широко використовує такий розділ вищої математики, як структурне моделювання.

З іншого боку, багато фундаментальних психологічних теорій, наприклад: теорія діяльності А.Н.Леонтьєва, теорія розвиваючого навчання В.В.Давидова, психоаналіз Фрейда й ін. добре відомі теорії, були створені без всякої опори на математику.

Тому виникає відчуття, що знання математичних методів психологу не потрібно.

Помітимо, що вище наведені теорії минулого сформульовані в минулому. Тепер же для захисту дисертації по психології часто потрібно статистичне підтвердження висунутих гіпотез. Крім того, зараз психологічні концепції піддаються сумніву на підставі того, що вони не були підтверджені статистично.

Зрозуміло, не всяке психологічне дослідження використовує математичні методи. Психологічні явища не мають математичних моделей (як у фізиці, наприклад). Але використання математичних методів дозволяє науково відібрати і систематизувати дані, а також дозволяє виявити закономірності, на перший погляд не завжди очевидні.

Головна відзнака галузей психологічних знань, що використовують математичні методи, полягає в тому, що їхній предмет дослідження не тільки може бути описаний, але й обмірюваний. Можливість виміру того чи іншого психологічного феномена, властивості, характеристики, риси і т.і. відкриває доступ для застосування методів кількісного аналізу, а виходить, і відповідних обчислювальних процедур.

Перша завдання математичних методів у психології – вказати прийоми збору й групування статистичних даних, отриманих у результаті спостережень або у результаті спеціально поставлених експериментів.

Друга задача математичних методів – розробити методи аналізу статистичних даних у залежності від цілей дослідження.

2. Методичні вказівки до самостійної роботи

Тема 1. Основні поняття в математичній обробці психологічних даних

1.1. Ознаки і перемінні

Ознаки і перемінні – це вимірювані психологічні явища.

Такими явищами можуть бути

- час рішення задачі;
- кількість допущених помилок;
- рівень тривожності;
- інтенсивність агресивних реакцій;
- кут повороту корпусу в бесіді;
- показник соціометричного статусу;
- і інші перемінні.

Психологічні перемінні є випадковими величинами, оскільки заздалегідь не відомо, яке саме значення вони приймуть.

Математична обробка – це дії зі значеннями ознаки, отриманими у психологічному дослідженні. Такі індивідуальні результати називаються також «варіантами», «датами», «індивідуальними показниками» і ін.

Значення ознаки визначаються за допомогою спеціальних шкал виміру.

1.2. Шкали виміру

Шкали виміру можна класифікувати на 4 типи:

- 1) номінативна чи шкала найменувань;
- 2) порядкова шкала;
- 3) інтервальна чи шкала рівних інтервалів;
- 4) шкала рівних відносин.

Розглянемо кожну з них.

Номінативна шкала – це шкала, що класифікує чи об'єкти явища за назвою. Назва не вимірюється кількісно, вона лише дозволяє відрізнити один об'єкт від іншого.

Найпростіший випадок номінативної шкали – дихотомічна шкала, що складається усього лише з двох осередків, наприклад:

- «іноземець»–«співвітчизник»;
- «відповів на запитання» - «не відповів на запитання»;
- «проголосував «за»» - «проголосував «проти»» і т.п.

Ознака, що вимірюється по дихотомічній шкалі найменувань, називається альтернативною. Вона може приймати всього два значення. При цьому дослідник найчастіше зацікавлений в одному з них, і тоді він говорить, що ознака «проявилася», якщо вона прийняла значення, яке його цікавить, і що ознака «не проявилася», якщо вона прийняла протилежне значення. Наприклад: “Ознака ліворукості проявилася у 8 випробуваних з 20”. У принципі номінативна шкала може складатися з осередків “ознака проявилася – ознака не проявилася”.

Більш складний варіант номінативної шкали – класифікація з трьох і більш осередків, наприклад «холерики – сангвініки - меланхоліки – флегматики».

Коли всі об'єкти, чи реакції усіх випробуваних розподілено по осередках класифікації, ми одержуємо можливість від найменувань перейти до чисел, підрахувавши кількість спостережень у кожному з осередків.

Номінативна шкала дозволяє нам підраховувати частоти різних значень ознаки і потім працювати з цими частотами за допомогою математичних методів.

Одиниця виміру, який ми при цьому оперуємо – кількість спостережень (випробуваних, реакцій, виборів і т.п.), чи частота. Точніше, одиниця виміру – це одне спостереження. Такі дані можуть бути оброблені за допомогою методу χ^2 , біноміального критерію m і кутового перетворення Фішера ϕ (ці методи ми розглянемо пізніше).

Порядкова шкала – це шкала, що класифікує об'єкти за принципом «більше – менше». Якщо номінативній шкалі було байдуже, у якому порядку ми розташуємо класифікаційні осередки, то в порядковій шкалі вони утворять послідовність від осередку «найменше значення» до осередку «найбільше значення» (чи навпаки). Осередки тепер можливо називати класами, оскільки стосовно класів використовують визначення «низький», «середній» і «високий» клас, чи 1-й, 2-й, 3-й клас, і т.д.

У порядковій шкалі повинне бути не менш трьох класів, наприклад «позитивна реакція – нейтральна реакція – негативна реакція» чи «підходить для заняття вакантної посади – підходить із застереженнями – не підходить» і т.п.

У порядковій шкалі ми не знаємо відстані між класами, а знаємо лише, що вони утворять послідовність. Наприклад, класи «підходить для заняття вакантної посади» і «підходить із застереженнями» можуть бути реально ближче друг до друга, чим клас «підходить із застереженнями» до класу «не підходить».

Від класів легко перейти до чисел, якщо ми умовимося вважати, що нижчий клас одержує ранг 1, середній клас – ранг 2, а вищий клас – ранг 3, чи навпаки. Чим більше класів у шкалі, тим більше в нас можливостей для математичної обробки отриманих даних і перевірки статистичних гіпотез.

Наприклад, ми можемо оцінити розходження між двома вибірками випробуваних по перевазі в них більш високих чи більш низьких рангів.

Усі психологічні методи, що використовують ранжирування, побудовані на застосуванні шкали порядку. Якщо випробуваному пропонується упорядкувати 18 цінностей по ступені їхньої значимості для нього, випробуваний робить так назване примусове ранжирування, при якому кількість рангів відповідає кількості об'єктів.

Незалежно від того, чи приписуємо ми кожній якості чи випробуваному один з 3-4 рангів або робимо процедуру примусового ранжирування, ми одержуємо в обох випадках ряди значень, обмірювані по порядковій шкалі. Правда, якщо в нас всього 3 можливих класи і, отже, 3 ранги, і при цьому, скажемо, треба здійснити ранжирування 20 випробуваних, то деякі з них неминуче одержать однакові ранги. Усе різноманіття життя не може уміститися в 3 градації, тому в один клас можуть потрапити люди, що досить серйозно розрізняються між собою. З іншого боку, примусове ранжирування, тобто утворення послідовності з багатьох випробуваних, може дуже перебільшити розходження між людьми. Крім того, дані, отримані в різних групах, можуть виявитися непорівнянними, тому що групи можуть споконвічно розрізнятися за рівнем досліджуваної якості, і випробуваний, що одержав в одній групі вищий ранг, в іншій одержав би усього лише середній.

Вихід з положення може бути знайдений, якщо задавати досить дробову систему, наприклад, з 10 класів. Переважна більшість психологічних методик, що використовують експертну оцінку, побудовано на вимірі тим самим «аршином» з 10, 20, чи навіть 100 градацій різних випробуваних у різних вибірках.

Отже, одиниця виміру в порядковій шкалі – відстань у 1 клас або у 1 ранг, при цьому відстань між класами і рангами може бути різною. До даних, що отримані по порядковій шкалі,

застосовують багато математичних методів (наприклад, метод χ^2 , метод кутового перетворення Фішера ϕ , метод рангової кореляції).

Інтервальна шкала – це шкала, що класифікує об'єкти за принципом «більше на визначену кількість одиниць – менше на визначену кількість одиниць». Кожне з можливих значень ознаки відстоїть від іншого на рівній відстані.

Можна припустити, що якщо ми вимірюємо час рішення задачі в секундах, то це вже явно шкала інтервалів. Однак насправді це не так, оскільки психологічно розходження в 20 секунд між випробуваними А і Б може бути аж ніяк не дорівнює розходженню в 20 секунд між випробуваними В і Г, якщо випробуваний А вирішив задачу за 2 секунди, Б – за 22 секунди, В – за 222 секунди, а Г – за 242 секунди.

Аналогічним образом, кожна секунда після того, як минуло півтори хвилини у досвіді з виміром м'язового вольового зусилля на динамометрі, по «ціні», може дорівнювати 10 чи навіть більш секундам у перші півхвилини досвіду.

Спроби вимірювати психологічні явища у фізичних одиницях зрозумілі, адже це виміри в одиницях об'єктивно існуючого часу і простору. Однак жоден досвідчений дослідник при цьому не зваблює себе думкою, що він робить виміри за психологічною інтервальною шкалою. Ці виміри належать порядковій шкалі.

Ми можемо у вище приведеному прикладі з визначеною часткою впевненості стверджувати лише, що випробуваний А вирішив задачу швидше випробуваного Б, Б швидше В, В швидше Г.

Аналогічним образом, значення, отримані випробуваними в балах по будь-якій не стандартизованій методиці, виявляються обмірюваними лише в порядковій шкалі. Насправді рівноінтервальними можна вважати лише шкали в одиницях стандартного відхилення, і лише за умови, що розподіл значень у вибірці підпорядковується нормальному закону.

Методи статистичної обробки, використовувані в психології, здебільшого не вимагають перевірки збігу отриманого емпіричного розподілу з нормальним. Вони побудовані на підрахунку частот і ранжируванні. Перевірка необхідна тільки у випадку застосування дисперсійного аналізу.

В усіх інших випадках немає необхідності перевіряти ступінь збігу отриманого емпіричного розподілу з нормальним, і тим більше прагнути перетворити порядкову шкалу в рівноінтервальну. У яких би одиницях ні були обмірювані перемінні - у секундах, міліметрах, градусах, кількості виборів і т.п. - усі ці дані можуть бути оброблені за допомогою непараметричних критеріїв.

Шкала рівних відносин - це шкала, що класифікує об'єкти чи суб'єкти пропорційно ступеня виразності вимірюваної властивості. У шкалах відносин класи позначаються числами, що пропорційні один одному: 2 так відноситься до 4, як 4 до 8. Це припускає наявність абсолютної нульової крапки відліку; У фізиці абсолютна нульова крапка відліку зустрічається при вимірі довжин чи відрізків фізичних об'єктів і при вимірі температури по шкалі Кельвіна з абсолютним нулем температур. Вважається, що в психології прикладами шкал рівних відносин є шкали порогів абсолютної чутливості (Стівенс С., 1960; Гайда В. К., Захаров В. П., 1982). Можливості людської психіки настільки великі, що важко уявити собі абсолютний нуль у який-небудь вимірюваний психологічний перемінний. Абсолютна дурість і абсолютна чесність поняття скоріше життєвої психології.

Те ж відноситься і до встановлення рівних відносин: тільки метафора повсякденної мови допускає, щоб Іванов був у 2 рази (3, 100, 1000) розумніше Петрова чи навпаки.

Абсолютний нуль, що правда, може мати місце при підрахунку кількості об'єктів чи суб'єктів. Наприклад, при виборі однієї з 3 альтернатив випробувані не вибрали альтернативу А жодного раз, альтернативу Б - 14 разів і альтернативу В - 28 разів. У цьому випадку ми можемо затверджувати, що альтернативу В вибирають у два рази частіше, ніж альтернативу Б. Однак при цьому обмірювана не психологічна властивість людини, а співвідношення виборів у 42 чоловік.

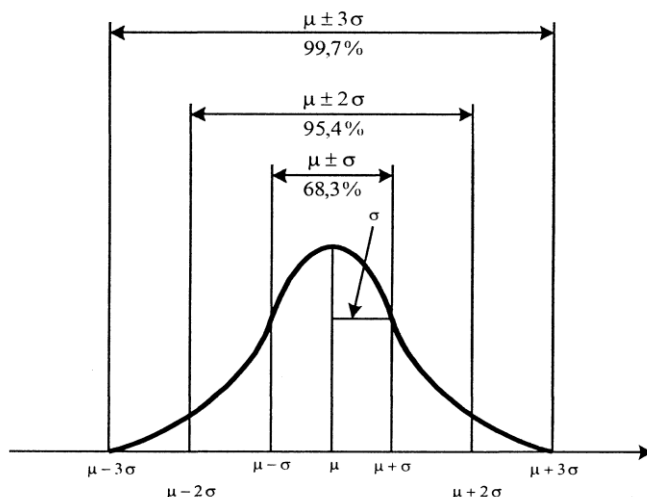
Стосовно показників частот можливо застосовувати всі арифметичні операції: додавання, віднімання, ділення і множення. Одиниця виміру в цій шкалі відносин - 1 спостереження, 1 вибір, 1 реакція і т.п. Ми повернулися до того, з чого почали: до універсальної шкали виміру в частотах того чи іншого значення ознаки і до одиниці виміру, що являє собою 1 спостереження. Розділив випробуваних по осередках номінативної шкали, ми можемо застосувати потім вищу шкалу виміру - шкалу відносин між частотами.

1.3. Розподіл ознаки. Параметри розподілу

Розподілом ознаки називається закономірність різних його значень.

У психологічних дослідженнях найчастіше посилаються на **нормальний розподіл**.

Нормальний розподіл характеризується тим, що крайні значення ознаки в ньому зустрічаються досить рідко, а значення, близькі до середньої величини - досить часто. Нормальним такий розподіл називається тому, що воно дуже часто зустрічалось в природно-наукових дослідженнях і здавався "нормою" усякого масового випадкового прояву ознак. Цей розподіл діє за законом, відкритому трьома вченими в різний час: Муавром у 1733 р. в Англії, Гауссом у 1809 р. у Німеччині і Лапласом у 1812 р. у Франції. Графік нормального розподілу являє собою звичну ока психолога-дослідника криву.



Параметри розподілу - це його числові характеристики, що вказують, де "у середньому" розташовуються значення ознаки, наскільки ці значення мінливі і чи спостерігається переважна поява визначених значень ознаки. Найбільше практично важливими параметрами є математичне чекання, дисперсія, показники асиметрії й ексцесу.

У реальних психологічних дослідженнях ми оперуємо не параметрами, а їхніми наближеними значеннями, так називаними оцінками параметрів. Це порозумівається обмеженістю обстежених вибірок. Чим більше вибірка, тим ближче може бути оцінка параметра до його щирого значення. Надалі, говорячи про параметри, ми будемо мати на увазі їхньої оцінки.

Середнє арифметичне (оцінка математичного чекання) обчислюється по формулі:

$$\bar{\tilde{o}} = \frac{\sum \tilde{o}_i}{n}$$

де x_i , - кожне значення ознаки, що спостерігається;

i - індекс, що вказує на порядковий номер даного значення ознаки;

n - кількість спостережень;
 знак підсумовування.
 Оцінка дисперсії визначається по формулі:

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

x_i - кожне значення ознаки, що спостерігається;
 \bar{x} - середнє арифметичне значення ознаки;
 n - кількість спостережень.

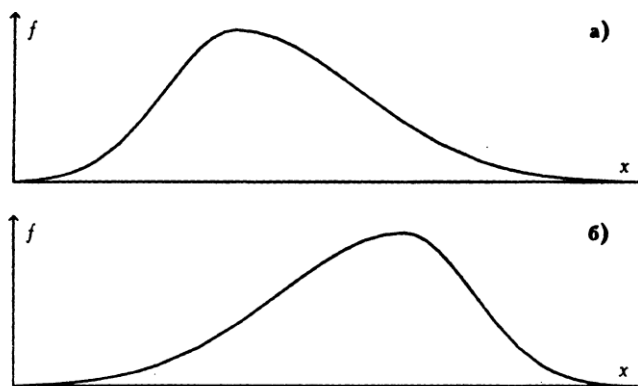
Величина, що представляє собою квадратний корінь з незміщеної оцінки дисперсії (S), називається стандартним чи відхиленням середнім квадратичним відхиленням. Для більшості дослідників звично позначати цю величину грецькою буквою σ (сигма), а не S . Насправді, σ - це стандартне відхилення в генеральній сукупності, а S - незміщена оцінка цього параметра в дослідженій вибірці. Але, оскільки S - краща оцінка σ , цю оцінку стали часто позначати вже не як S , а як σ :

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

У тих випадках, коли які-небудь причини сприяють більш частій появі значень, що чи вище, навпаки, нижче середнього, утворюються асиметричні розподіли. При лівосторонньої, чи додатної асиметрії в розподілі частіше зустрічаються більш низькі значення ознаки, а при правобічної, чи від'ємної - більш високі.

Показник асиметрії (A) обчислюється по формулі: $A_s = \frac{\sum (x_i - \bar{x})^3}{n \cdot \sigma^3}$

Для симетричних розподілів $A=0$.



Асиметрія розподілів

а) ліва, додатна; б) права, від'ємна

У тих випадках, коли які-небудь причини сприяють переважній появі середніх чи близьких до середніх значень, утвориться розподіл з додатним ексцесом. Якщо ж у розподілі переважають крайні значення, причому одночасно і більш низькі, і більш високі, то такий розподіл

характеризується від'ємним ексцесом, і в центрі розподілу може утворитися западина, що перетворює його в криву з двома вершинами. (див. Рис. 1.6).

Показник ексцесу (E) визначається по формулі:
$$E_k = \frac{\sum (x_i - \bar{x})^4}{n \cdot \sigma^4} - 3$$

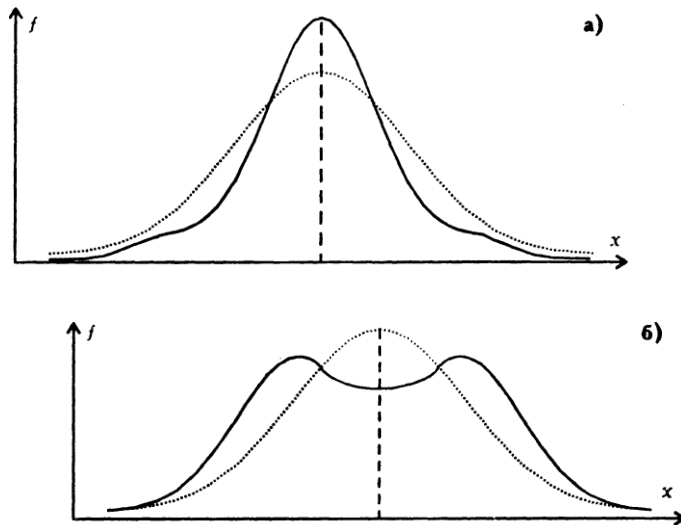


Рис. 1.6. Ексцес:

а) позитивний; б) негативний

У розподілах з нормальною опуклістю $E=0$. Параметри розподілу виявляється можливим визначити тільки стосовно даних, представленим принаймні в інтервальній шкалі. Фізичні шкали довжин, часу, кутів є інтервальними шкалами, і тому до них застосовні способи розрахунку оцінок параметрів, принаймні, з формальної точки зору. Параметри розподілу не враховують щирої психологічної нерівномірності секунд, міліметрів і інших фізичних одиниць виміру.

На практиці психолог-дослідник може розраховувати параметри будь-якого розподілу, якщо одиниці, що він використовував при вимірі, визнаються розумними в науковому співтоваристві.

1.4. Статистичні гіпотези

Формулювання гіпотез систематизує припущення дослідника і представляє їх у чіткому і лаконічному виді. Завдяки гіпотезам дослідник не втрачає дороговказної нитки в процесі розрахунків і йому легко зрозуміти після їхнього закінчення, що, власне, він знайшов.

Статистичні гіпотези підрозділяються на нульові й альтернативні, спрямовані і ненаправлені.

Нульова гіпотеза - це відсутності розходжень. Вона позначається як H_0 і називається нульовою тому, що містить число 0: $X_1 - X_2 = 0$, де X_1, X_2 - значення ознак, що зіставляються. Нульова гіпотеза - це те, що ми хочемо спростувати, якщо перед нами стоїть задача довести значимість розходжень.

Альтернативна гіпотеза - це гіпотеза про значимість розходжень. Вона позначається як H_1 . Альтернативна гіпотеза - це те, що ми хочемо довести, тому іноді її називають експериментальною гіпотезою.

Бувають задачі, коли ми хочемо довести саме незначимість розходжень, тобто підтвердити нульову гіпотезу. Наприклад, якщо нам потрібно переконатися, що різні випробувані одержують хоча і різні, але урівноважені по труднощам завдання, чи що експериментальна і контрольна вибірки не розрізняються між собою по якихось значимих характеристиках. Однак частіше нам

усе-таки потрібно довести **значимість розходжень**, тому що вони більш інформативні для нас у пошуку нового. Нульова й альтернативна гіпотези можуть бути спрямованими і ненаправленими.

Спрямовані гіпотези

H_0 : X_1 не перевищує X_2

H_1 : X_1 перевищує X_2

Ненаправлені гіпотези

H_0 : X_1 не відрізняється від X_2

H_1 : X_1 відрізняється від X_2

Якщо ви помітили, що в одній із груп індивідуальні значення випробуваних по якій-небудь ознаці, наприклад по соціальній сміливості, вище, а в іншій нижче, те для перевірки значимості цих розходжень нам необхідно сформулювати спрямовані гіпотези.

Якщо ми хочемо довести, що в групі А під впливом якихось експериментальних впливів відбулися більш виражені зміни, чим у групі Б, то нам теж необхідно сформулювати спрямовані гіпотези.

Якщо ж ми хочемо довести, що розрізняються форми розподілу ознаки в групі А и Б, то формулюються ненаправлені гіпотези.

При описі кожного критерію в керівництві дані формулювання гіпотез, що він допомагає нам перевірити.

Перевірка гіпотез здійснюється за допомогою критеріїв статистичної оцінки розходжень.

1.5. Статистичні критерії

Статистичний критерій - це вирішальне правило, що забезпечує надійне поводження, тобто прийняття щирої і відхилення помилкової гіпотези з високою імовірністю.

Статистичні критерії позначають також метод розрахунку визначеного числа і саме це число.

Коли ми говоримо, що вірогідність розходжень визначалася за критерієм χ^2 маємо на увазі, що використовували метод χ^2 розрахунку визначеного числа.

Коли ми говоримо, далі, що $\chi^2 = 12,676$, те маємо на увазі визначене число, розраховане по методу χ^2 . Це число позначається як емпіричне значення критерію.

По співвідношенню емпіричного і критичного значень критерію ми можемо судити про тім, чи підтверджується чи спростовується нульова гіпотеза. Наприклад, якщо χ^2 емпіричне більше χ^2 критичного H_0 відкидається.

У більшості випадків для того, щоб ми визнали розходження значимими, необхідно, щоб емпіричне значення критерію перевищувало критичне, хоча є критерії (наприклад, критерій Манна-Уитні чи критерій знаків), у яких ми повинні дотримувати протилежного правила.

Ці правила обмовляються в описі кожного з представлених у керівництві критеріїв.

У деяких випадках розрахункова формула критерію містить у собі кількість спостережень у досліджуваній вибірці, що позначається як n . У цьому випадку емпіричне значення критерію одночасно є тестом для перевірки статистичних гіпотез. По спеціальній таблиці ми визначаємо, якому рівню статистичної значимості розходжень відповідає дана емпірична величина. Прикладом такого критерію є критерій ϕ^* , що обчислюється на основі кутового перетворення Фішера.

У більшості випадків, однак, те саме емпіричне значення критерію може виявитися значимим чи незначущою в залежності від кількості спостережень у досліджуваній вибірці (n) чи від так названої кількості ступенів волі, що позначається як V чи як df .

Число ступенів волі V дорівнює числу класів варіаційного ряду мінус число умов, при яких він був сформований. До числа таких умов відносяться обсяг вибірки (n), середні і дисперсії.

Якщо ми розподілили спостереження по класах якої-небудь номінативної шкали і підраховували кількість спостережень у кожному осередку класифікації, то ми одержуємо так називаний частотний варіаційний ряд. Єдина умова, що дотримується при його формуванні - обсяг вибірки n . Допустимо, у нас 3 класи: "Вміє працювати на комп'ютері - вміє виконувати лише визначені операції - не вміє працювати на комп'ютері". Вибірка складається з 50 чоловік. Якщо в перший клас віднесені 20 випробуваних, у другий - теж 20, то в третьому класі повинні виявитися всі інші 10 випробуваних. Ми обмежені однією умовою - обсягом вибірки. Тому навіть якщо ми втратили дані про тім, скількох чоловік не вміють працювати на комп'ютері, ми можемо визначити це, знаючи, що в першому і другому класах - по 20 випробуваних. Ми не вільні у визначенні кількості випробуваних у третьому, розряді, "воля" простирається тільки на перші двох осередків класифікації:

$$V = c - l = 3 - 1 = 2$$

Аналогічним образом, якби в нас була класифікація з 10 розрядів, то ми були б вільні тільки в 9 з них, якби в нас було 100 класів - те в 99 з них і т.д.

Знаючи n чи число ступенів волі, ми по спеціальних таблицях можемо визначити критичні значення критерію і зіставити з ними отримане емпіричне значення. Звичайно це записується так: "при $n=22$ критичні значення критерію складають ..." чи "при $V=2$ критичні значення критерію складають ..." і т.п.

Критерії поділяються на параметричні і непараметричні.

Параметричні критерії – це критерії, що включають у формулу розрахунку параметри розподілу, тобто середні і дисперсії (t-критерій Стюдента, критерій F і ін.)

Непараметричні критерії - це критерії, що не включають у формулу розрахунку параметри розподілу і засновані на діях з частотами або рангами (критерій Q Розенбаума, критерій T Вілкоксона й ін.)

Можливості і обмеження параметричних критеріїв

1. Дозволяють прямо оцінити розходження в середніх, отриманих у двох вибірках (t - критерій Стюдента).
2. Дозволяють прямо оцінити розходження в дисперсіях (критерій Фішера).
3. Дозволяють виявити тенденції зміни ознаки при переході від умови до умови (дисперсійний однофакторний аналіз), але лише за умови нормального розподілу ознаки.
4. Дозволяють оцінити взаємодію двох і більш факторів у їхньому впливі на зміни ознаки (двохфакторний дисперсійний аналіз).
5. Експериментальні дані повинні відповідати двом, а іноді трьом, умовам:
 - а) значення ознаки обмірювані за інтервальною шкалою;
 - б) розподіл ознаки є нормальним;
 - в) у дисперсійному аналізі повинна дотримуватися вимога рівності дисперсій в осередках комплексу.
6. Математичні розрахунки досить складні.
7. Якщо умови, перераховані в п.5, виконуються, параметричні критерії виявляються трохи більш могутніми, чим непараметричні.

Можливості і обмеження непараметричних критеріїв

1. Дозволяють оцінити лише середні тенденції, наприклад, відповісти на запитання, чи частіше у вибірці А зустрічаються більш високі, а у вибірці Б - більш низькі значення ознаки (критерії Q , U , Φ^* і ін.).
2. Дозволяють оцінити лише розходження в діапазонах варіативності ознаки (критерій Фішера з кутовим перетворенням).
3. Дозволяють виявити тенденції зміни ознаки при переході від умови до умови при будь-якому розподілі ознаки (критерії тенденцій L і S).
4. Відсутня можливість оцінити взаємодію двох і більш факторів у їхньому впливі на зміни ознаки (двохфакторний дисперсійний аналіз).
5. Експериментальні дані можуть не відповідати жодній з умов:
 - а) значення ознаки можуть бути представлені в будь-якій шкалі, починаючи від шкали найменувань;
 - б) розподіл ознаки може бути будь-яким, збіг його з яким-небудь теоретичним законом розподілу є необов'язковим і не має потреби в перевірці;
 - в) вимога рівності дисперсій відсутня.
6. Математичні розрахунки по більшій частині прості і займають мало часу (за винятком критеріїв χ^2 і λ).
7. Якщо умови, перераховані в п.5, не виконуються, непараметричні критерії виявляються більш могутніми, чим параметричні.

З перерахованого вище видно, що параметричні критерії можуть виявитися трохи більш могутніми, ніж непараметричні, але тільки в тому випадку, якщо ознака обмірювана по інтервальній шкалі і нормально розподілена. З інтервальною шкалою є визначені проблеми. Лише з деякою натяжкою ми можемо вважати дані, представлені не в стандартизованих оцінках, як інтервальні. Крім того, перевірка розподілу "на нормальність" вимагає досить складних розрахунків, результат яких заздалегідь невідомий. Може виявитися, що розподіл ознаки відрізняється від нормального, і так чи інакше все рівно доведеться звернутися до непараметричних критеріїв.

Непараметричні критерії позбавлені всіх цих обмежень і не вимагають таких тривалих і складних розрахунків. У порівнянні з параметричними критеріями вони обмежені лише в одному - з їхньою допомогою неможливо оцінити взаємодію двох чи більш умов або факторів, що впливають на зміну ознаки. Цю задачу може вирішити тільки дисперсійний двофакторний аналіз.

1.6. Рівні статистичної значимості

Рівень значимості - це імовірність того, що ми рахували розходження істотними, а вони насправді випадкові.

Коли ми вказуємо, що розходження достовірні на 5%-ом рівні значимості, чи при $p \leq 0,05$, то ми маємо на увазі, що імовірність того, що розходження недостовірні, складає 0,05.

Коли ми вказуємо, що розходження достовірні на 1%-ом рівні значимості, чи при $p \leq 0,01$, то ми маємо на увазі, що імовірність того, що розходження недостовірні, складає 0,01.

Якщо перевести все це на більш формалізовану мову, то рівень значимості - це імовірність відхилення нульової гіпотези, у той час як вона вірна.

Помилка, що складається в тім, що ми відхилили нульову гіпотезу, у той час як вона вірна, називається помилкою 1 роду.

Імовірність такої помилки звичайно позначається як α . По суті, можна вказувати в дужках не $p \leq 0,05$ чи $p \leq 0,01$, а $\alpha \leq 0,05$ чи $\alpha \leq 0,01$. У деяких посібниках так і робиться.

Якщо імовірність помилки - це α , то імовірність правильного рішення: $1-\alpha$. Чим менше α , тим більше імовірність правильного рішення.

Історично склалося так, що в психології прийнято вважати:

- нижчим рівнем статистичної значимості 5% рівень ($p \leq 0,05$);
- достатнім - 1% рівень ($p \leq 0,01$);
- і вищим 0,1% рівень ($p \leq 0,001$).

Тому в таблицях критичних значень звичайно приводяться значення критеріїв, що відповідають рівням статистичної значимості $p \leq 0,05$ і $p \leq 0,01$, іноді - $p \leq 0,001$. Для деяких критеріїв у таблицях указаний точний рівень значимості їх різних емпіричних значень. Наприклад, для $\phi^* = 1,56$ $p = 0,06$. Доти, однак, поки рівень статистичної значимості не досягне $p = 0,05$, ми ще не маємо права відхилити нульову гіпотезу. Будемо дотримуватися наступного правила відхилення гіпотези про відсутність розходжень (H_0) і прийняття гіпотези про статистичну вірогідність розходжень (H_1).

Правило відхилення H_0 і прийняття H_1 :

Якщо емпіричне значення критерію дорівнює критичному значенню, що відповідає $p \leq 0,05$ чи перевищує його, то H_0 відхиляється, але ми ще не можемо безперечно прийняти H_1 . Якщо емпіричне значення критерію дорівнює критичному значенню, що відповідає $p < 0,01$ чи перевищує його, то H_0 відхиляється і приймається H_1 .

Виключення: критерій знаків G, критерій Т Вілкоксона і критерій U Манна-Уитні. Для них установлюються зворотні співвідношення.

Для полегшення процесу ухвалення рішення можна всякий раз вичерчувати «вісь значимості».

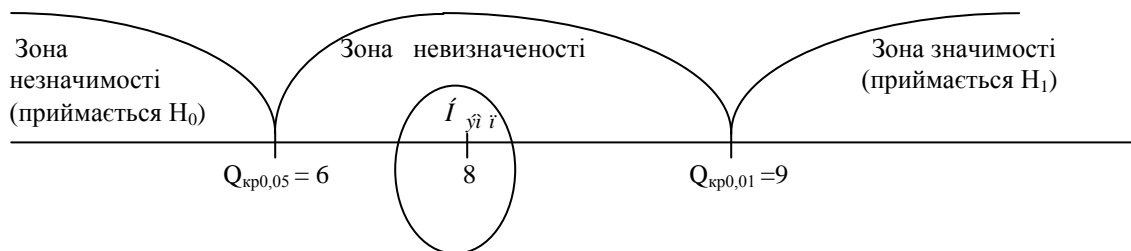


Рис. 1.7. Приклад "осі значимості" для критерію Q Розенбаума

Критичні значення критерію позначені як $Q_{кр0,05}$ і $Q_{кр0,01}$, емпіричне значення критерію як $Q_{емп}$. (укладено в еліпс).

Вправо від критичного значення $Q_{кр0,01}$ простирається "зона значимості" - сюди потрапляють емпіричні значення, що перевищують $Q_{кр0,01}$ і, отже, безумовно значимі.

Уліво від критичного значення $Q_{кр0,05}$ простирається "зона незначимості", - сюди потрапляють емпіричні значення Q , що нижче $Q_{кр0,05}$ і отже, безумовно незначимі.

Ми бачимо, що $Q_{кр0,05} = 6$; $Q_{кр0,01} = 9$; $Q_{емп} = 8$

Емпіричне значення критерію попадає в область між $Q_{кр0,05}$ і $Q_{кр0,01}$. Ця зона "невизначеності": ми вже можемо відхилити гіпотезу про невірогідність розходжень (H_0), але ще не можемо прийняти гіпотези про їхню вірогідність (H_1).

Практично, однак, дослідник може вважати достовірними вже ті розходження, що не попадають у зону незначимості, заявивши, що вони достовірні при $p \leq 0,05$.

Рівень статистичної значимості чи критичні значення критеріїв визначаються по-різному при перевірці направлених і ненаправлених статистичних гіпотез.

При направленій статистичній гіпотезі використовується однобічний критерій, при ненаправленій гіпотезі - двосторонній критерій. Двосторонній критерій більш строгий, оскільки він перевіряє розходження в обидва боки, і тому то емпіричне значення критерію, що раніше відповідало рівню значимості $p \leq 0,05$, тепер відповідає лише рівню $p \leq 0,10$.

Досліднику доводиться всякий раз самостійно вирішувати, чи використовує він однобічний чи двосторонній критерій. Але в багатьох посібниках таблиці критичних значень критеріїв підібрані таким чином, що направленим гіпотезам відповідає однобічний, а ненаправленим - двосторонній критерій, і приведені значення задовольняють тим вимогам, що пред'являються до кожного з них. Досліднику необхідно лише стежити за тим, щоб його гіпотези збігалися за змістом і за формою з гіпотезами, пропонованими в описі кожного з критеріїв.

1.7. Потужність критеріїв

Потужність критерію - це його здатність виявляти розходження, якщо вони є. Іншими словами, це його здатність відхилити нульову гіпотезу про відсутність розходжень, якщо вона невірна.

Помилка, що складається в тому, що ми прийняли нульову гіпотезу, у той час як вона невірна, називається помилкою II роду.

Імовірність такої помилки позначається як β . Потужність критерію - це його здатність не припуститися помилки II роду, тому:

Потужність $= 1 - \beta$

Потужність критерію визначається емпіричним шляхом. Ознакові задачі можуть бути вирішені за допомогою різних критеріїв, при цьому виявляється, що деякі критерії дозволяють виявити розходження там, де інші виявляються нездатними це зробити, чи виявляють більш високий рівень значимості розходжень. Виникає питання: а навіщо ж тоді використовувати менш могутні критерії? Справа в тім, що підставою для вибору критерію може бути не тільки потужність, але й інші його характеристики, а саме:

- а) простота;
- б) більш широкий діапазон використання (наприклад, стосовно даних, визначеним по номінативній шкалі, чи стосовно великих кількостей спостережень n);
- в) застосовність відносно нерівних по обсязі вибірок;
- г) велика інформативність результатів.

1.8. Класифікація задач і методів їхнього рішення

Безліч задач психологічного дослідження припускає ті чи інші зіставлення. Ми зіставляємо групи випробуваних по якій-небудь ознаці, щоб виявити розходження між ними за цією ознакою. Ми зіставляємо те, що було "до" з тим, що стало "після" наших експериментальних чи будь-яких інших впливів, щоб визначити ефективність цих впливів. Ми зіставляємо емпіричний розподіл значень ознаки з яким-небудь теоретичним законом чи розподілу два емпіричних розподіли між собою, для того, щоб довести не випадковість вибору чи альтернатив розходження у формі розподілів.

Ми, далі, можемо зіставляти дві ознаки, обмірювані на одній і тій же вибірці випробуваних, для того, щоб установити ступінь узгодженості їхніх змін, їхню спряженість, кореляцію між ними.

Нарешті, ми можемо зіставляти індивідуальні значення, отримані при різних комбінаціях яких-небудь істотних умов, для того щоб виявити характер взаємодії цих умов у їхньому впливі на індивідуальні значення ознаки.

Саме ці задачі дозволяє вирішити той набір методів, що приведено в таблиці. Усі ці методи можуть бути використані при так називаній "ручній" обробці даних.

Класифікація задач і методів їхнього рішення

N	Задачі	Умови	Методи
1.	Виявлення розходжень у рівні досліджуваної ознаки	а) 2 вибірки випробуваних	Q - критерій Розенбаума; U - критерій Манна-Уйтні; ϕ^* - критерій (кутове перетворення Фішера)
		б) 3 і більш вибірок випробуваних	S - критерій тенденцій Джонкіра; H - критерій Крускала-Уолліса.
2.	Оцінка зрушення значень досліджуваної ознаки	а) 2 виміри на одній і тієї ж вибірки випробуваних	T - критерій Вілкоксона; G - критерій знаків; ϕ^* - критерій (кутове перетворення Фішера)
		б) 3 і більш замірів на одній і тієї ж вибірці випробуваних	χ_r^2 - критерій Фрідмана; L - критерій тенденцій Пейджа
3.	Виявлення розходжень у розподілі ознаки	а) при зіставленні емпіричного розподілу з теоретичним;	χ^2 - критерій Пірсона; λ - критерій Колмогорова-Смірнова; m - біноміальний критерій.
		б) при зіставленні двох емпіричних розподілів	χ^2 - критерій Пірсона; λ - критерій Колмогорова-Смірнова; ϕ^* - критерій (кутове перетворення Фішера)
4.	Виявлення ступеня погодженості змін	а) двох ознак	r_s - коефіцієнт рангової кореляції Спірмена.
		б) двох ієрархій чи профілів	r_s - коефіцієнт рангової кореляції Спірмена.
5.	Аналіз змін ознаки під впливом контролюючих умов	а) під впливом одного фактора	S - критерій тенденцій Джонкіра; L - критерій тенденцій Пейджа; - однофакторний дисперсійний аналіз Фішера.
		б) під впливом двох факторів одночасно	Двофакторний дисперсійний аналіз Фішера.

ТЕМА №2. Виявлення розходжень у рівні досліджуваної ознаки

2.1. Обґрунтування задачі зіставлення і порівняння

Дуже часто перед дослідником у психології стоїть задача виявлення розходжень між двома, трьома і більш вибірками випробуваних. Це може бути, наприклад, задача визначення психологічних особливостей хронічно хворих дітей у порівнянні зі здоровими, юних правопорушників у порівнянні з законослухняними однолітками чи розходжень між працівниками державних підприємств і приватних фірм, між людьми різної національності чи різної культури і, нарешті, між людьми різного віку в методі "поперечних зрізів".

Іноді по виявленні у дослідженні статистично достовірним розходженням формується «груповий профіль» чи «усереднений портрет» людини тієї чи іншої професії, статусу ін. ..

В останні роки всі частіше встає задача виявлення психологічного портрета фахівця нових професій: "успішного менеджера", "успішного політика", "успішного торгового представника", "успішного комерційного директора" і ін. Такого роду дослідження не завжди мають на увазі участь двох чи більш вибірок. Іноді обстежить одна, але досить представницька вибірка кількістю не менш 60 чоловік, а потім усередині, цієї вибірки виділяються групи більш і менш успішних фахівців, і їхні дані по дослідженим перемінним порівнюються між собою. У найпростішому випадку критерієм для поділу вибірки на "успішних" і "неуспішних" буде середня величина по показнику успішності. Однак такий розподіл є досить грубим: особи, що одержали близькі оцінки по успішності, можуть виявитися в протилежних групах, а особи, що помітно розрізняються по оцінках успішності, - в одній групі. Це може викривити результати порівняння груп чи зробити розходження між групами менш помітними.

Щоб уникнути цього, можна спробувати виділити групи «успішних» і «неуспішних» фахівців більш строго, включаючи в першу з них тільки тих, чий значення *перевищують* середню величину не менш чим на $1/4$ стандартного відхилення, а в другу групу - тільки тих, чий значення не менш чим на $1/4$ стандартного відхилення *нижче* середньої величини. При цьому усі, хто виявляється в зоні середніх величин, $\bar{x} \pm \frac{1}{4} \sigma$, випадають з подальших зіставлень. Якщо розподіл близький до нормального, то випаде приблизно 19,8% випробуваних. Якщо розподіл відрізняється від нормального, то таких випробуваних може бути і більше. Щоб уникнути утрат, можна зіставляти не двох, а три групи випробуваних: з високою, середньою і низькою професійною успішністю.

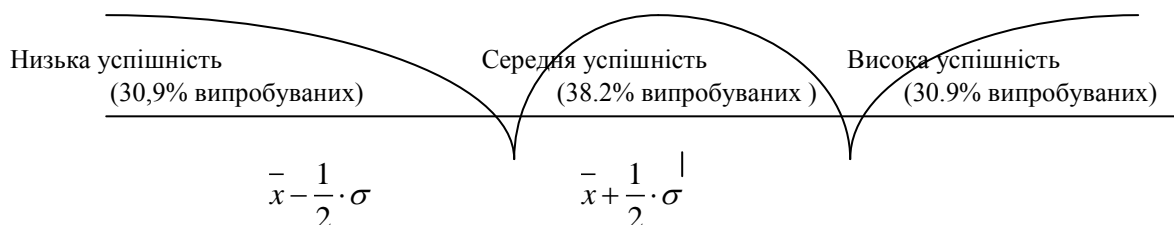


Рис 2.1. Схематичне зображення процесу поділу вибірки на групи з низькою, середньою і високою професійною успішністю

На Рис. 2.1 представлено схему поділу вибірки на групи з низькою, середньою і високою професійною успішністю за критерієм відхилення значень від середньої величини на $1/2$

стандартного відхилення. При такому строгому критерії в "середню" групу попадають (при нормальному розподілі) близько 38,2% усіх випробуваних, а в крайніх групах виявляється по 30,9% випробуваних.

Чим менше випробуваних виявляється в групах, тим менше в нас можливостей для виявлення достовірних розходжень, тому що критичні значення більшості критеріїв при малих n суворіше, ніж при великих n .

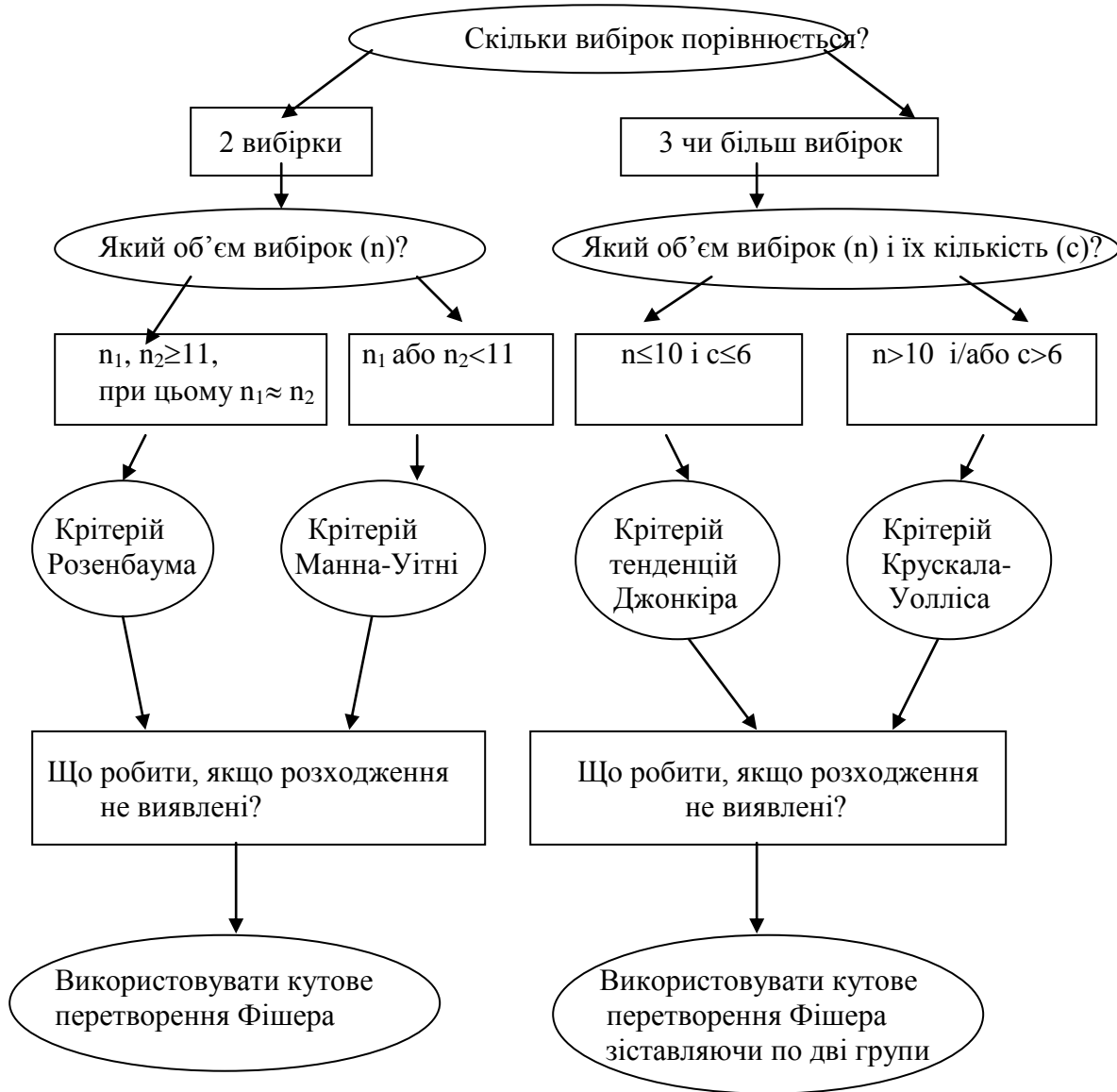
Таким чином, при нестрогому поділі випробуваних на групи ми втрачаємо в точності, а при строгому - у кількості випробуваних. При рішенні задач виявлення розходжень у рівневих показниках варто пам'ятати, що "усереднений профіль успішного фахівця" повинний розглядатися скоріше як дослідницький результат, що дозволяє сформулювати гіпотези для подальших досліджень, а не як підстава для професійного добору.

Тому є дві причини. По-перше, у жодного з успішних фахівців може не спостерігатися "усереднений профіль" - він, по суті, є абстрактним узагальненням; по-друге, у професійній діяльності наявність власного індивідуального стилю важніше відповідності середньому профілю. Р.Б. Кеттелл, з огляду на це, пропонував при дослідженні професійної успішності включати в розгляд індивідуальні профілі видатних представників тієї чи іншої професії.

Зіставлення рівневих показників у різних вибірках може бути необхідною частиною комплексних діагностичних, навчальних, психокоррекційних і інших програм. Воно допомагає нам звернути увагу на ті особливості обстежених вибірок, що повинні бути враховані і використані при адаптації програм до даної групи в процесі їхнього конкретного втілення.

Рішення про вибір того чи іншого критерію приймається на основі того, скільки вибірок зіставляється і який їхній обсяг (див. Алгоритм).

Алгоритм вибіру критерію оцінки вірогідності розходжень між незалежними вибірками по рівню ознаки



2.2. Q - критерій Розенбаума

Призначення критерію

Критерій використовується для оцінки розходжень між двома вибірками за рівнем будь-якої ознаки, кількісно обмірюваній. У кожній з вибірок повинне бути не менш 11 випробуваних.

Опис критерію

Це дуже простий непараметричний критерій, що дозволяє швидко оцінити розходження між двома вибірками за якою-небудь ознакою. Однак якщо критерій Q не виявляє достовірних розходжень, це ще не означає, що їх дійсно немає.

У цьому випадку варто застосувати критерій φ^* Фішера. Якщо ж Q-критерій виявляє достовірні розходження між вибірками з рівнем значимості $p \leq 0,5$, можна обмежитися тільки їм і уникнути труднощів застосування інших критеріїв.

Критерій застосовується в тих випадках, коли дані представлені принаймні у порядковій шкалі. Ознака повинна варіювати в якомусь діапазоні значень, інакше зіставлення за допомогою Q-критерію просто неможливі. Наприклад, якщо в нас тільки 3 значення ознаки, 1, 2 і 3, - нам дуже важко буде встановити розходження. Метод Розенбаума вимагає, отже, досить тонко обмірюваних ознак.

Застосування критерію починаємо з того, що упорядковуємо значення ознаки в обох вибірках по наростанню (чи убуттю) ознаки. Найкраще, якщо дані кожного випробуваного представлені на окремій картці. Тоді нічого не варто упорядкувати два ряди значень по цікавлячому нас ознаці, розкладаючи картки на столі. Так ми відразу побачимо, чи збігаються діапазони значень, і якщо ні, то наскільки один ряд значень "вище" (S_1), а другий - "нижче" (S_2). Для того, щоб не заплутатися, у цьому й у багатьох інших критеріях рекомендується першим рядом (вбіркою, групою) вважати той ряд, де значення вище, а другим рядом - той, де значення нижче.

Гіпотези

H_0 : Рівень ознаки у вибірці 1 не перевищує рівня ознаки у вибірці 2.

H_1 : Рівень ознаки у вибірці 1 перевищує рівень ознаки у вибірці 2.

Графічне представлення критерію Q

На Рис. 2.2. представлені три варіанти співвідношення рядів значень у двох вибірках. У варіанті (а) усі значення першого ряду вище всіх значень другого ряду. Розходження, безумовно, достовірні, при дотриманні умови, що $n_1, n_2 \geq 11$.

У варіанті (б), навпроти, обидва ряди знаходяться на тому самому рівні: розходження недостовірні.

У варіанті (в) ряди частково перехрещуються, але все-таки перший ряд виявляється набагато вище другого, чи Досить великі зони S_1 і S_2 , у сумі складові Q, можна визначити по Таблиці I Додатка 1, де приведені критичні значення Q для різних n. Чим величина Q більше, тим більше достовірні розходження ми зможемо констатувати.

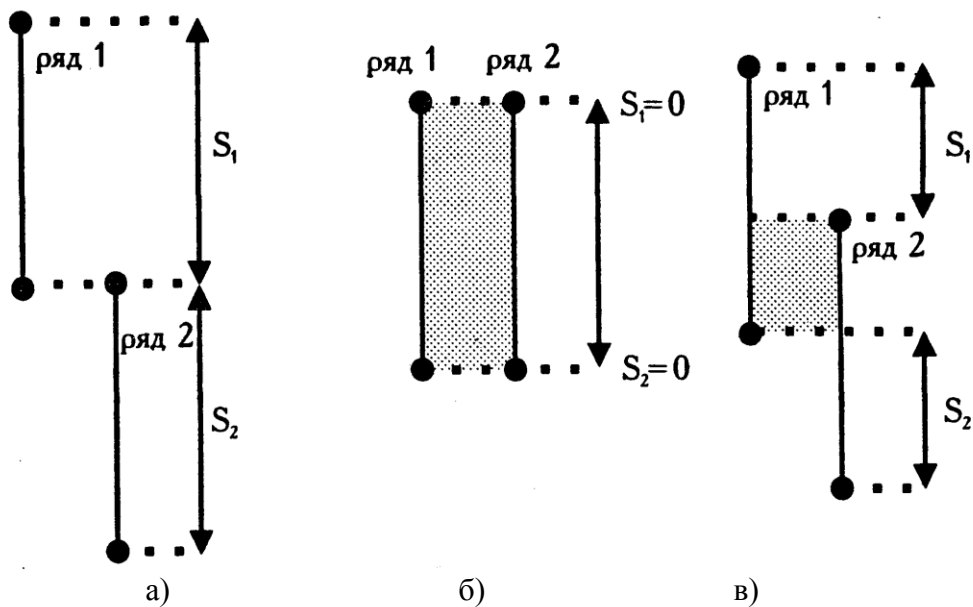


Рис.2.2. Три варіанти співвідношення рядів значень у двох вибірках.

Обмеження критерію Q

1. У кожній з вибірок, що зіставляються, повинне бути не менш 11 спостережень. При цьому обсяги вибірок повинні приблизно збігатися.

Існують наступні правила:

а) якщо в обох вибірках менше 50 спостережень, то абсолютна величина різниці між n_1 і n_2 не повинна бути більше 10 спостережень;

б) якщо в кожній з вибірок більше 51 спостереження, але менше 100, то абсолютна величина різниці між n_1 і n_2 не повинна бути більше 20 спостережень;

в) якщо в кожній з вибірок більше 100 спостережень, то допускається, щоб одна з вибірок була більше іншої не більш ніж у 1,5-2 рази.

2. Діапазони розкиду значень у двох вибірках повинні не збігатися між собою, у протилежному випадку застосування критерію безглуздо. Тим часом, можливі випадки, коли діапазони розкиду значень збігаються, але, унаслідок різнобічної асиметрії двох розподілів, розходження в середніх величинах ознак істотний (Рис. 2.3., 2.4).

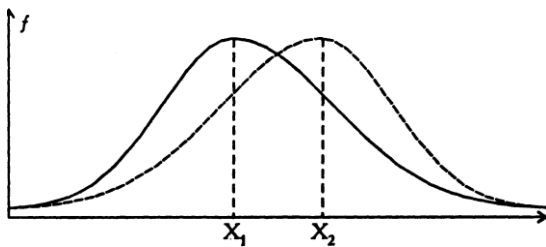


Рис. 2.3. Варіант співвідношення розподілів ознаки в двох вибірках, при якому критерій Q безпомічний

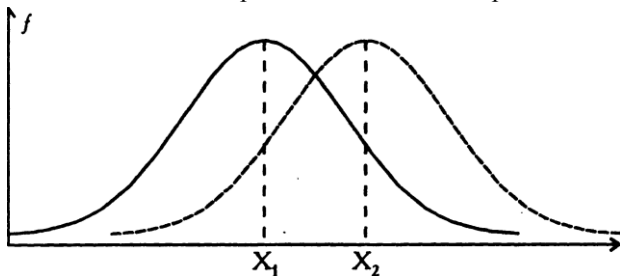


Рис. 2.4. Варіант співвідношення розподілів ознаки в двох вибірках, при якому критерій Q може бути могутнім

Підрахунок критерію Q Розенбаума

1. Перевірити, чи виконуються обмеження: $n_1, n_2 \geq 11$, $n_1 \approx n_2$
2. Упорядкувати значення окремо в кожній вибірці по ступені зростання ознаки. Вважати вибіркою 1 ту вибірку, значення в якій приблизно вище, а вибіркою 2 - ту, де значення приблизно нижче.
3. Визначити найвище (максимальне) значення у вибірці 2.
4. Підрахувати кількість значень у вибірці 1, що вище максимального значення у вибірці 2. Позначити отриману величину як S_1 .
5. Визначити найнижче (мінімальне) значення у вибірці 1.
6. Підрахувати кількість значень у вибірці 2, що нижче мінімальні значення вибірки 1. Позначити отриману величину як S_2 .
7. Підрахувати емпіричне значення Q по формулі: $Q = S_1 + S_2$.
8. По Табл. 2.1 визначити критичні значення Q для даних n_1 і n_2 . Якщо $Q_{\text{емп.}}$ дорівнює $Q_{0,05}$ чи перевищує його, H_0 відкидається.
9. При $n_1, n_2 > 26$ зіставити отримане емпіричне значення з $Q_{\text{кр}} = 8$ ($p \leq 0.05$) і $Q_{\text{кр}} = 10$ ($p \leq 0.01$). Якщо $Q_{\text{емп.}}$ чи перевищує принаймні дорівнює $Q_{\text{кр}} = 8$, H_0 відкидається.

Таблиця 2.1

Критичні значення критерію Q Розенбаума для рівнів статистичної значимості $p \leq 0,5$ і $p \leq 0,1$

n	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
$p=0,05$																
11	6															
12	6	6														
13	6	6	6													
14	7	7	6	6												
15	7	7	6	6	6											
16	8	7	7	7	6	6										
17	7	7	7	7	7	7	7									
18	7	7	7	7	7	7	7	7								
19	7	7	7	7	7	7	7	7	7							
20	7	7	7	7	7	7	7	7	7	7						
21	8	7	7	7	7	7	7	7	7	7	7					
22	8	7	7	7	7	7	7	7	7	7	7	7				
23	8	8	7	7	7	7	7	7	7	7	7	7	7			
24	8	8	8	8	8	8	8	8	8	8	7	7	7	7		
25	8	8	8	8	8	8	8	8	8	7	7	7	7	7	7	
26	8	8	8	8	8	8	8	8	8	8	7	7	7	7	7	7
$p=0,01$																
11	9															
12	9	9														
13	9	9	9													
14	9	9	9	9												
15	9	9	9	9	9											
16	9	9	9	9	9	9										
17	10	9	9	9	9	9	9									
18	10	10	9	9	9	9	9	9								
19	10	10	10	9	9	9	9	9	9							
20	10	10	10	10	9	9	9	9	9	9						
21	11	10	10	10	9	9	9	9	9	9	9					
22	11	11	10	10	10	9	9	9	9	9	9	9				
23	11	11	10	10	10	10	9	9	9	9	9	9	9			
24	12	11	11	10	10	10	10	9	9	9	9	9	9	9		
25	12	11	11	10	10	10	10	10	9	9	9	9	9	9	9	
26	12	12	11	11	10	10	10	10	10	9	9	9	9	9	9	9

Приклад

У учасників психологічного експерименту, що моделює діяльність повітряного диспетчера, був обмірюван рівень вербального і невербального інтелекту за допомогою методики Д. Векслера. Було обстежено 26 юнака у віці від 18 до 24 років (середній вік 20,5 років). 14 з них були студентами фізичного факультету, а 12 - студентами психологічного факультету Ленінградського університету (Сидоренко Е.В., 1978). Показники вербального інтелекту представлені в Табл. 2.2. Чи можна затверджувати, що одна з груп перевершує іншу за рівнем вербального інтелекту?

Таблиця 2.2

Індивідуальні значення вербального інтелекту у вибірках студентів фізичного ($n_1=14$) і психологічного ($n_2=12$) факультетів

Студенти-фізики			Студенти - психологи		
	П.І.	Показник вербального інтелекту		П.І.	Показник вербального інтелекту
1.	И.А	132	1.	Н.Т.	126
2.	К.А.	134	2.	О.В.	127
3.	К.Е.	124	3.	Е.В.	132
4.	П.А.	132	4.	Ф.О.	120
5.	С.А.	135	5.	И.Н.	119
6.	Ст.А.	132	6.	И.Ч.	126
7.	Т.А.	131	7.	И.В.	120
8.	Ф.А.	132	8.	К.О.	123
9.	Ч.И.	121	9.	Р.Р.	120
10.	Ц.А	127	10.	Р.И.	116
11.	См.А.	136	11.	О.К.	123
12.	К.Ан.	129	12.	Н.К.	115
13.	Б.Л.	136			
14.	Ф.В.	136			

Упорядкуємо значення в обох вибірках, а потім сформулюємо гіпотези:

H_0 : Студенти-фізики не перевершують студентів-психологів за рівнем вербального інтелекту.

H_1 : Студенти-фізики перевершують студентів-психологів за рівнем вербального інтелекту.

Таблиця 2.3

Упорядковані ряди індивідуальних значень вербального інтелекту у двох студентських вибірках

1 ряд- студенти-фізики		2 ряд- студенти-психологи	
1. См. А.	136 --		
2. Б.Л.	136		
3. Ф.В.	136 S1		
4. С.А.	135		
5. К.А.	134		

6. И.А.	132	1. Е.В.	132
7. П.А.	132		
8. Ст.А.	132		
9. Ф.А.	132		
10. Т.А.	131		
11. К.Ан.	129		
12. Ц.А.	127	2. О.В.	127
		3. Н.Т.	126
		4. И.Ч.	126
13 ДО.Е.	124		
		5. К.О.	123
		6. О.К.	123
14. Ч.И.	121		
		7. Ф.О.	120
		8. И.В.	120
		9. Р.Р.	120
		S2 10. И.Н.	119
		11. Р.И.	116
		12. Н.К.	115

Як видно з Табл. 2.3, ми правильно позначили ряди: перший, той, що "вище" - ряд фізиків, а другий, той, що "нижче" - ряд психологів.

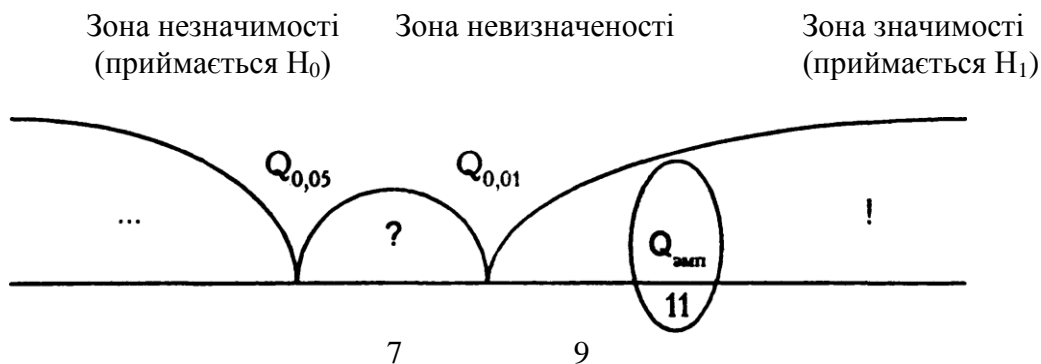
По Табл. 2.3 визначаємо кількість значень першого ряду, що більше максимального значення другого ряду: $S1=5$.

Тепер визначаємо кількість значень другого ряду, що менше мінімального значення першого ряду: $S2=6$.

Обчислюємо $Q_{\text{емп}}$ за формулою: $Q_{\text{емп}} = S1 + S2 = 5 + 6 = 11$

За табл.2.1. визначаємо критичні значення Q для $n_1=14$, $n_2=12$: $Q_{\text{кр}0,05}=7$ та $Q_{\text{кр}0,01}=9$.

Побудуємо "вісь значимості".



$$Q_{\text{емп}} > Q_{\text{кр}} (p \leq 0,1)$$

Відповідь: H_0 відхиляється.

Приймається H_1 . Студента-фізики перевершують студентів-психологів за рівнем вербального інтелекту ($p \leq 0,1$).

Відзначимо, що в тих випадках, коли емпіричне значення критерію виявляється на границі зони незначимості, ми маємо право стверджувати лише, що розходження достовірні при $p \leq 0,5$, якщо ж воно знаходиться між двома критичними значеннями, то ми можемо стверджувати, що $p < 0,5$.

Якщо емпіричне значення критерію виявляється на границі, ми можемо стверджувати, що $p \leq 0,1$, якщо воно попадає в зону значимості, ми можемо стверджувати, що $p \leq 0,1$.

Оскільки рівень значимості виявлених розходжень досить високий ($p \leq 0,1$), ми могли б на цьому зупинитися. Однак якщо дослідник сам психолог, а не фізик, навряд чи він на цьому зупиниться.

Він може спробувати зіставити вибірки за рівнем невербального інтелекту, оскільки саме невербальний інтелект визначає рівень інтелекту в цілому і ступінь його організованості.

2.3. U - критерій Манна-Уїтні

Призначення критерію

Критерій призначений для оцінки розходжень між двома вибірками за рівнем будь-якої ознаки, кількісно обмірюваної. Він дозволяє виявляти розходження між малими вибірками, коли $n_1, n_2 \geq 3$ чи $n_1=2, n_2 \geq 5$, і є більш могутнім, чим критерій Розенбаума.

Опис критерію

Існує кілька способів використання критерію і кілька варіантів таблиць критичних значень, що відповідають цим способам.

Цей метод визначає, чи досить мала зона перехресних значень між двома рядами. Ми пам'ятаємо, що 1-м рядом (вибіркою, групою) ми називаємо той ряд значень, у якому значення, по попередній оцінці, вище, а 2-м рядом - той, де вони приблизно нижче.

Чим менше область перехресних значень, тим більш імовірно, що розходження достовірні. Іноді ці розходження називають розходженнями в розташуванні двох вибірок.

Емпіричне значення критерію U відображає, наскільки велика зона збігу між рядами. Тому чим менше $U_{\text{емп}}$, тим більш імовірно, що розходження достовірні.

Гіпотези:

H_0 : Рівень ознаки в групі 2 не нижче рівня ознаки в групі 1.

H_1 : Рівень ознаки в групі 2 нижче рівня ознаки в групі 1.

Графічне представлення критерію U

На Рис. 2.5. представлені три можливих варіанта співвідношення двох рядів значень.

У варіанті (а) другий ряд нижче першого, і ряди майже не перехреснюються. Область накладення занадто мала, щоб скрадати розходження між рядами. Є шанс, що розходження між ними достовірні. Точно визначити це ми зможемо за допомогою критерію U.

У варіанті (б) другий ряд теж нижче першого, але й область перехресних значень у двох рядів досить велика. Вона може ще не досягати критичної величини, коли розходження прийдеться визнати несуттєвими. Але чи не так це, можна визначити тільки шляхом точного підрахунку критерію U.

У варіанті (в) другий ряд нижче першого, але зона накладення настільки велика, що розходження між рядами скрадаються.

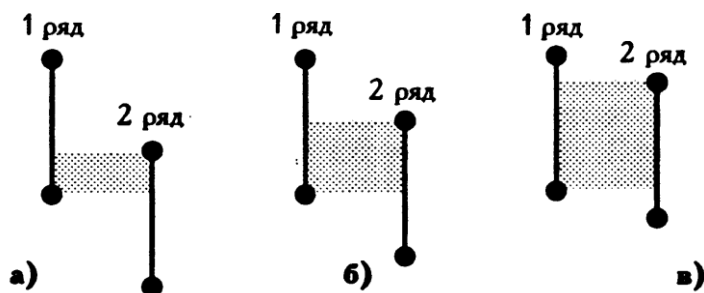


Рис. 2.5. Можливі варіанти співвідношень рядів значень у двох вибірках; штрихуванням позначені зони накладення

Обмеження критерію U - критерію Манна-Уїтні

1. У кожній вибірці повинне бути не менш 3 спостережень: $n_1, n_2 \geq 3$; допускається, щоб в одній вибірці було 2 спостереження, але тоді в другий їх повинно бути не менш 5.
2. У кожній вибірці повинне бути не більш 60 спостережень; $n_1, n_2 \leq 60$.
Однак при $n_1, n_2 > 20$ ранжирування стає досить трудомістким.
На наш погляд, у випадку, якщо $n_1, n_2 > 20$, краще використовувати інший критерій, а саме кутове перетворення Фідера в комбінації з критерієм λ , що дозволяє виявити критичну крапку, у якій накопичуються максимальні розходження між двома вибірками, що зіставляються.

Правила ранжирування

1. Якщо значення (число) у вибірці зустрічається один раз, то його ранг дорівнює порядковому номеру цього значення в упорядкованій вибірці.
2. У випадку, якщо кілька значень рівні, їм нараховується ранг, що представляє собою середнє значення з тих рангів, які вони одержали б, якби не були рівні.
3. Загальна сума рангів повинна збігатися з розрахунковою, котра визначається за формулою:

$$\sum(R_i) = \frac{N \cdot (N+1)}{2}$$

де N - загальна кількість спостережень (значень). Розбіжність реальної й розрахункової сум рангів буде свідчити про помилку, яка з'явилася при нарахуванні рангів або їхньому підсумовуванні. Перш ніж продовжити роботу, необхідно знайти помилку і усунути її.

Приклад.

Дано вибірку: 44, 47, 47, 42, 44, 44, 41, 50, 46. Потрібно проранжувати значення даної вибірки.

Рішення. Складемо таблицю, куди занесемо впорядковані за зростанням значення вибірки.

№	Значення вибірки	Ранг
1	41	1
2	42	2
3	44	4
4	44	4
5	44	4

6	46	6
7	47	7,5
8	47	7,5
9	50	9

Значення 41 має ранг 1. Значення 42 має ранг 2. Значення 46 має ранг 6. Ранги цих значень рівні їхньому порядковому номеру, тому що ці значення зустрічаються у вибірці один раз.

Щоб обчислити ранг значення 44 потрібно обчислити середнє арифметичне порядкових номерів значень 44:

$$\frac{3+4+5}{3} = 4$$

Щоб обчислити ранг значення 47 потрібно обчислити середнє арифметичне порядкових номерів значень 47:

$$\frac{7+8}{2} = 7,5$$

Переконаємося в правильності розрахунків, зрівнявши суму рангів, які ми одержали, з розрахунковою сумою рангів.

Сума рангів: $1+2+4+4+4+6+7,5+7,5+9=45$

Розрахункова сума рангів: $\sum(R_i) = \frac{N \cdot (N+1)}{2} = \frac{9 \cdot (9+1)}{2} = \frac{9 \cdot 10}{2} = 45$

Суми збігаються, виходить, розрахунок рангів виконаний вірно.

Підрахунок критерію U Манна-Уїтні.

1. Перенести всі дані випробуваних на індивідуальні картки.
2. Позначити картки випробуваних з вибірки 1 одним кольором, скажемо червоним, а всі картки з вибірки 2 - іншим, наприклад синім.
3. Розкласти всі картки в єдиний ряд за ступенем зростання ознаки, не зважаючи на те, до якої вибірки вони відносяться, як якби ми працювали з однією великою вибіркою.
4. Проранжирувати значення на картках, приписуючи меншому значенню менший ранг. Усього рангів вийде стільки, скільки в нас (n_1+n_2).
5. Знову розкласти картки на дві групи, орієнтуючись на кольорові позначення: червоні картки в один ряд, сині - в інший.
6. Підрахувати суму рангів окремо на червоних картках (вибірка 1) і на синіх картках (вибірка 2). Перевірити, чи збігається загальна сума рангів з розрахунковою.
7. Визначити більшу з двох рангових сум.
8. Визначити значення $U_{\hat{a} \hat{i}}$ за формулою:

$$U_{\hat{a} \hat{i}} = (n_1 \cdot n_2) + \frac{n_x \cdot (n_x + 1)}{2} - T_x$$

де n_1 - кількість випробуваних у вибірці 1;

n_2 - кількість випробуваних у вибірці 2;

T_x - більша з двох рангових сум;

n_x - кількість випробуваних у групі з більшою сумою рангів.

9. Визначити критичні значення $U_{\hat{e} \hat{\delta}}$ за таблицею критичних значень критерію U Манна-Уїтні.

Якщо $U_{\hat{a} \hat{i}} > U_{\hat{e} \hat{\delta} 0,05}$ гіпотеза H_0 приймається.

Якщо $U_{\hat{a} \hat{i}} \leq U_{\hat{e} \hat{\delta} 0,05}$ гіпотеза H_0 відхиляється.

Якщо $U_{\alpha} < U_{\alpha 0,01}$ гіпотеза H_0 відхиляється й приймається гіпотеза H_1 .

Чим менше значення U_{α} , тим вірогідність розходжень вище.

Приклад

Звернемося до результатів обстеження студентів фізичного й психологічного факультетів університету за допомогою методики Д. Векслера для виміру вербального і невербального інтелекту. Дані наведені в таблиці 3.

Чи можна стверджувати, що одна з вибірок перевершує іншу за рівнем невербального інтелекту?

Таблиця 3. Індивідуальні значення невербального інтелекту у вибірках студентів фізичного ($n_1=14$) і психологічного ($n_2=12$) факультетів

Студенти-Фізики		Студенти-Психологи	
П.І.	Показник невербального інтелекту	П.І.	Показник невербального інтелекту
1. И.А.	111	1. Н.Т.	113
2. К.А.	104	2. О.В.	107
3. К.Е.	107	3. Е.В.	123
4. П.А.	90	4. Ф.О.	122
5. С.А.	115	5. И.Н.	117
6. Ст.А.	107	6. И.Ч.	112
7. Т.А.	106	7. И.В.	105
8. Ф.А.	107	8. К.О.	108
9. Ч.И.	95	9. Р.Р.	111
10. Т.А.	116	10. Р.И.	114
11. В.А.	127	11. О.К.	102
12. Ан.	115	12. Н.К.	104
13. Б.Л.	102		
14. К.В.	99		

Рішення.

- Об'єднаємо дві вибірки в одну. Всі дані розташуємо в порядку зростання. (див.табл.4). Проранжиріємо дані об'єднаної вибірки (див. Правила ранжирування).

Таблиця 4

№	значення	Ранг
1	90	1
2	95	2
3	99	3
4	102	4,5
5	102	4,5
6	104	6,5
7	104	6,5
8	105	8
9	106	9
10	107	11,5
11	107	11,5
12	107	11,5
13	107	11,5
14	108	14
15	111	15,5
16	111	15,5
17	112	17
18	113	18
19	114	19
20	115	20,5

Таблиця 5

Студенти- фізики			Студенти-Психологи		
№	значенн	ранг	№	знач	ранг
1	111	15,5	1	113	18
2	104	6,5	2	107	11,5
3	107	11,5	3	123	25
4	90	90	4	122	24
5	115	20,5	5	117	23
6	107	11,5	6	112	17
7	106	9	7	105	8
8	107	11,5	8	108	14
9	95	2	9	111	15,5
10	116	22	10	114	19
11	127	26	11	102	4,5
12	115	20,5	12	104	6,5
13	102	4,5			
14	99	3			
Сума рангів			Сума		

21	115	20,5
22	116	22
23	117	23
24	122	24
25	123	25
26	127	26

2. Складемо таблицю для двох вибірок (див.табл.5). Для кожного значення з попередньої таблиці (табл.4) випишемо відповідний ранг. По кожній групі окремо підрахуємо суму рангів.

Таблиця 5

Студенти- фізики			Студенти-Психологи		
№	значення	ранг	№	значення	ранг
1	111	15,5	1	113	18
2	104	6,5	2	107	11,5
3	107	11,5	3	123	25
4	90	90	4	122	24
5	115	20,5	5	117	23
6	107	11,5	6	112	17
7	106	9	7	105	8
8	107	11,5	8	108	14
9	95	2	9	111	15,5
10	116	22	10	114	19
11	127	26	11	102	4,5
12	115	20,5	12	104	6,5
13	102	4,5			
14	99	3			
Сума рангів		165	Сума рангів		186

3. Перевіримо, чи однакові розрахункова і реальна суми рангів.

Загальна сума рангів: $165+186=351$. Розрахункова сума рангів дорівнює:

$$\sum(R_i) = \frac{N \cdot (N+1)}{2} = \frac{26 \cdot (26+1)}{2} = 351$$

Рівність реальної й розрахункової сум дотримано. Ми бачимо, що за рівнем невербального інтелекту більше "високим" рядом виявляється вибірка студентів-психологів. Саме на цю вибірку доводиться більша рангова сума: 186.

4. Сформулюємо **гіпотези**:

H_0 : Група студентів-психологів не перевершує групу студентів-фізиків за рівнем невербального інтелекту.

H_1 : Група студентів-психологів перевершує групу студентів-фізиків за рівнем невербального інтелекту.

5. Визначаємо емпіричну величину $U_{\hat{\alpha} \hat{\gamma}}$:

$$U_{\hat{\alpha} \hat{\gamma}} = (n_1 \cdot n_2) + \frac{n_x \cdot (n_x + 1)}{2} - T_x = (14 \cdot 12) + \frac{12 \cdot (12 + 1)}{2} - 186 = 60$$

За табл. 1 визначаємо критичні значення для відповідних об'ємів вибірок, причому менший об'єм позначаємо n_1 ($n_1=12$) і відшукуємо його у верхньому рядку таблиці, більший об'єм позначаємо n_2 , ($n_2=14$), і відшукуємо його в лівому стовпці таблиці.

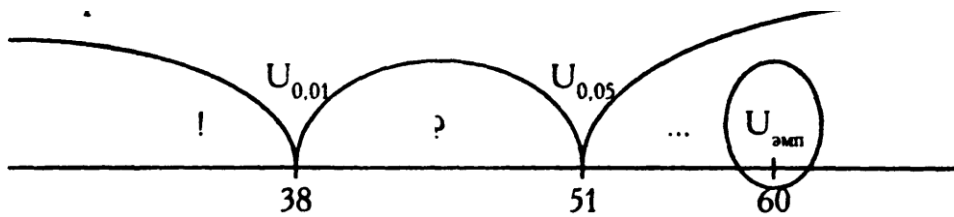
$$U_{\hat{\alpha} \hat{\gamma}} = \begin{cases} 51 & (p \leq 0,05) \\ 38 & (p \leq 0,01) \end{cases}$$

6. Побудуємо "вісь значимості".

Зона значимості
(H_1 приймається)

Зона невизначеності

Зона незначимості
(H_0 приймається)



Відзначимо на "осі значимості" $U_{y\bar{i}} = 60$. Тому що $U_{y\bar{i}} = 60$ знаходиться в зону незначимості приймається гіпотеза H_0 .

Відповідь: Гіпотеза H_0 приймається. Група студентів-психологів не перевершує групу студентів-фізиків за рівнем невербального інтелекту.

2.4. H - критерій Крускала-Уолліса

Призначення критерію

Критерій призначений для оцінки розходжень одночасно між *трьома, чотирма* і т.д. вибірками за *рівнем* будь-якої ознаки.

Він дозволяє встановити, що *рівень* ознаки *змінюється* при переході від групи до групи, але не вказує на напрямок цих змін.

Опис критерію

Критерій H іноді розглядається як непараметричний аналог методу дисперсійного однофакторного аналізу для незв'язних вибірок. Іноді його називають критерієм "суми рангів".

Даний критерій є продовженням критерію U на більше, ніж 2, кількості вибірок, що порівнюються. Всі індивідуальні значення ранжируються так, ніби це була одна велика вибірка. Далі всі індивідуальні значення повертаються у свої первісні вибірки, і ми підраховуємо суми отриманих ними рангів окремо по кожній вибірці. Якщо розходження між вибірками випадкові, суми рангів не будуть розрізнятися істотно, тому що високі і низькі ранги рівномірно розподіляться між вибірками. Але якщо в одній з вибірок будуть переважати низькі значення рангів, в іншій - високі, а в третьої - середні, то критерій H дозволить установити ці розходження.

Гіпотези

H_0 : Між вибірками 1, 2, 3 і т.д. існують лише випадкові розходження за рівнем досліджуваної ознаки.

H_1 : Між вибірками 1, 2, 3 і т.д. існують не випадкові розходження за рівнем досліджуваної ознаки.

Обмеження критерію H

1. При зіставленні 3-х вибірок допускається, щоб в одній з них $n=3$, а двох інших $n=2$. Але при таких об'ємах вибірок ми зможемо встановити розходження лише на нижчому рівні значимості ($p \leq 0,05$).

Для того, щоб виявилось можливим діагностувати розходження на більш високому рівні значимості ($p \leq 0,01$), необхідно, щоб у кожній вибірці було не менш 3 спостережень, чи щоб принаймні в одній з них було 4 спостереження, а в двох інші - по 2; при цьому неважливо, у якій саме вибірці скільки випробуваних, а важливе співвідношення 4:2:2.

2. Критичні значення критерію H і відповідні їм рівні значимості приведені в Табл. Кількість ступенів волі при цьому визначається по формулі:

$v = c - 1$, де c - кількість вибірок, що зіставляються.

3. При множинному зіставленні вибірок достовірні розходження між якою-небудь конкретною парою (чи парами) можуть виявитися стертими. Це обмеження можна перебороти, якщо провести всі можливі попарні зіставлення. Для таких попарних зіставлень використовується критерій для двох вибірок, наприклад U – Манна-Уїтні чи кутове перетворення Фішера ϕ^* .

Підрахунок критерію Н Крускала-Уолліса

1. Перенести всі показники випробуваних на індивідуальні картки.
2. Позначити картки випробуваних групи 1 визначеним кольором, наприклад червоним, картки випробуваних групи 2 - синім, картки випробуваних груп 3 і 4 - відповідно, зеленим і жовтим кольором і т.д. (Можна використовувати і будь-які інші позначення.)
3. Розкласти всі картки в єдиний ряд по ступені наростання ознаки, не зважаючи на те, до якої групи відносяться картки, як якби ми працювали з однією об'єднаною вибіркою.
4. Проранжирувати значення на картках, приписуючи меншому значенню менший ранг. Надписати на кожній картці її ранг. Загальна кількість рангів буде дорівнювати кількості випробуваних в об'єднаній вибірці.
5. Знову розкласти картки по групах, орієнтуючись на кольорові чи інші прийняті позначення.
6. Підрахувати суми рангів окремо по кожній групі. Перевірити збіг загальної суми рангів з розрахункової.

$$\sum(R_i) = \frac{N \cdot (N+1)}{2}$$

7. Підрахувати значення критерію Н по формулі:

$$H_{\text{yî}} = \left[\frac{12}{N \cdot (N+1)} \cdot \sum \frac{T_i^2}{n_i} \right] - 3 \cdot (N+1)$$

де N - загальна кількість випробуваних в об'єднаній вибірці;

n - кількість випробуваних у кожній групі;

T - суми рангів по кожній групі.

8. Визначити критичні значення χ^2 по Табл. 1. Якщо $H_{\text{емп}}$ дорівнює чи перевищує критичне значення χ^2 , H_0 відхиляється.

Таблиця 1.

Критичні значення критерію χ^2

p			p		
v	0,05	0,01	v	0,05	0,01
1	3.841	6.635	9	16.919	21.666
2	5.991	9.210	10	18.307	23.209
3	7.815	11.345	11	19.675	24.725
4	9.488	13.277	12	21.026	26.217
5	11.070	15.086	13	22.362	27.688
6	12.592	16.812	14	23.685	29.141
7	14.067	18.475	15	24.996	30.578
8	15.507	20.090	16	26.296	32.000

Приклад

В експерименті по дослідженню інтелектуальної наполегливості 22 випробуваним пред'являлися спочатку розв'язні чотирибуквені, п'ятибуквені і шестибуквені анаграми, а потім нерозв'язні анаграми, час роботи над якими не обмежувався. Експеримент проводився індивідуально з кожним випробуваним. Використовувалося 4 комплекти анаграм. Показники тривалості спроб у рішенні нерозв'язних анаграм представлені в таблиці. Усі випробувані були юнаками-студентами технічного вузу у віці від 20 до 22 років.

Чи можна стверджувати, що тривалість спроб рішення кожної з 4 нерозв'язних анаграм приблизно однакова?

Показники тривалості спроб рішення 4 нерозв'язних анаграм у секундах (N=22)

	Група 1: (n ₁ =4)	Група 2: (n ₂ =8)	Група 3: (n ₃ =6)	Група 4: (n ₄ =4)
	Анаграма ФОЛИТОН	Анаграма КАМУСТО	Анаграма СНЕРАКО	Анаграма ГРУТОСИЛ
1	145	145	128	60
2	194	210	283	2361
3	731	236	469	2416
4	1200	385	482	3600
5		720	1678	
6		848	2081	
7		905		
8		1080		

Сформулюємо гіпотези.

H₀: 4 групи випробуваних, що одержали різні нерозв'язні анаграми, не розрізняються по тривалості спроб їхнього рішення.

H₁: 4 групи випробуваних, що одержали різні нерозв'язні анаграми, розрізняються по тривалості спроб їхнього рішення.

Підрахунок рангових сум по групах випробуваних, що працювали над чотирма нерозв'язними анаграмами

Група 1: (n ₁ =4) анаграма ФОЛИТОН		Група 2: (n ₂ =8) анаграма КАМУСТО		Група 3: (n ₃ =6) анаграма СНЕРАКО		Група 4: (n ₄ =4) анаграма ГРУТОСИЛ	
Тривалість	Ранг	Тривалість	Ранг	Тривалість	Ранг	Тривалість	Ранг
						60	1
				128	2		
145	3.5	145	3.5				
194	5						
		210	6				
		236	7				
		385	9	283	8		
				469	10		
				482	11		
		720	12				
731	13						
		848	14				
		905	15				
		1080	16				
1200	17						
				1678	18		
				2081	19		
						2361	20

						2416	21
						3600	22
Суми	38,5		82,5		68		64
Середні	9.6		10.3		11.3		16.0

Загальна сума рангів = 38,5 + 82,5 + 68 + 64 = 253.

Розрахункова сума

$$\sum(R_i) = \frac{N \cdot (N+1)}{2} = \frac{22 \cdot (22+1)}{2} = 253$$

Рівність реальної і розрахункової сум дотримано.

Тепер визначаємо емпіричне значення Н:

$$H_{ei} = \left[\frac{12}{N \cdot (N+1)} \cdot \sum \frac{T_i^2}{n_i} \right] - 3 \cdot (N+1)$$

$$H_{ei} = \left[\frac{12}{22 \cdot (22+1)} \cdot \left(\frac{38,5^2}{4} + \frac{82,5^2}{8} + \frac{68^2}{6} + \frac{64^2}{4} \right) \right] - 3 \cdot (22+1) = 2,48$$

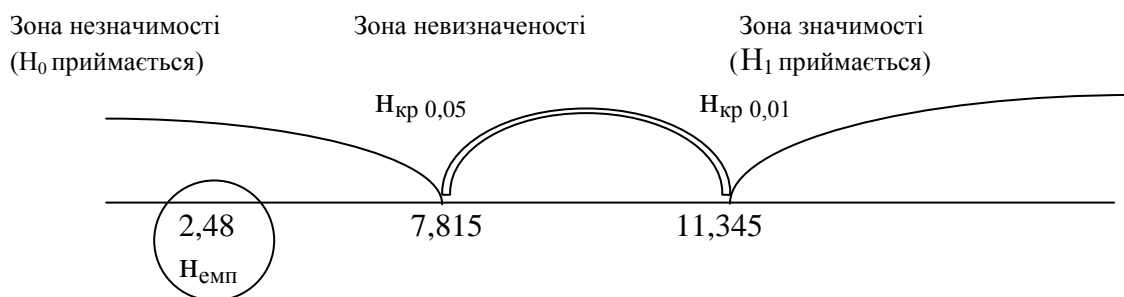
Оскільки таблиці критичних значень критерію Н передбачені тільки для кількості груп $c=3$, а в даному випадку в нас 4 групи, прийдеться зіставляти отримане емпіричне значення Н із критичними значеннями χ^2 . Для цього спочатку визначимо кількість ступенів волі v для $c=4$:
 $v = c - 1 = 4 - 1 = 3$

Тепер визначимо критичні значення по Табл. IX Додатка 1 для $v=3$:

$$\chi_{ed}^2 = \begin{cases} 7,815 & (p \leq 0,05) \\ 11,345 & (p \leq 0,01) \end{cases}$$

$$H_{ei} = 2,48$$

Побудуємо "вісь значимості".



Відповідь: H_0 приймається: 4 групи випробуваних, що одержали різні нерозв'язні анаграми, не розрізняються за тривалістю спроб їхнього рішення.

Тема 3. ВИЯВЛЕННЯ РОЗХОДЖЕНЬ У РОЗПОДІЛІ ОЗНАКИ

3.1. Обґрунтування задачі порівняння розподілів ознаки

Розподіли можуть розрізнятися по середнім, дисперсіям, асиметрії, ексцесу і по сполученнях цих параметрів. Розглянемо кілька прикладів.

На Рис. 4.1 представлено два розподіли ознаки. Розподіл 1 характеризується меншим діапазоном варіативності і меншою дисперсією, ніж розподіл 2. У розподілі 1 частіше зустрічаються значення ознаки, близькі до середнього, а в розподілі 2 частіше зустрічаються більш високі і більш низькі, чим середнє, значення ознаки.

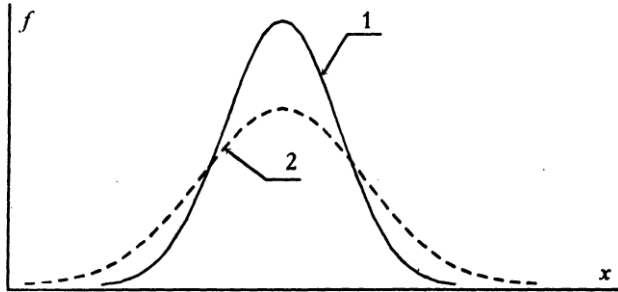


Рис. 4.1. Криві розподілу ознаки з меншим діапазоном варіативності ознаки (1) і більшим діапазоном розподілу ознаки (2); x - значення ознаки; f - відносна частота значень ознаки.

Саме таке співвідношення може спостерігатися в розподілі фенотипичних ознак у чоловіків (крива 2) і жінок (крива 1). Фенотипична дисперсія чоловічої статі повинна бути більшою, ніж жіноча. Чоловіки - це авангардна частина популяції, відповідальна за пошук нових форм пристосування, тому в них частіше зустрічаються рідкі крайні значення різних фенотипичних ознак. У той же час жіноча частина популяції відповідальна за збереження вже накопичених змін, тому в них частіше зустрічаються середні значення фенотипичних ознак.

Аналіз реально одержуваних у дослідженнях розподілів може дозволити нам підтвердити чи спростувати дані теоретичні припущення.

На Рис. 4.2 представлено два розподіли, що розрізняються за знаком асиметрії: розподіл 1 характеризується додатною (лівосторонньою) асиметрією, а розподіл 2 — від'ємною (правосторонньою) асиметрією.

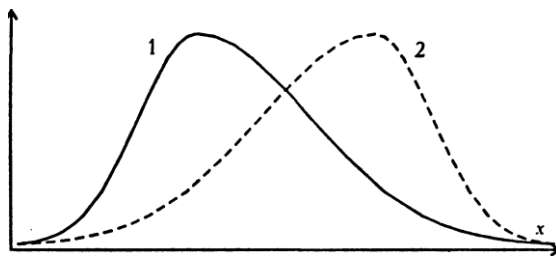


Рис. 4.2. Криві розподілу ознаки з додатною (лівосторонньою) асиметрією (1) і від'ємною (правосторонньою) асиметрією (2); x - значення ознаки; f відносна частота значень ознаки

Дані криві можуть відображати розподіл часу рішення простої задачі (крива 1) і важкої задачі (крива 2). Просту задачу більшість випробуваних вирішують швидко, тому велика частина значень групується ліворуч. У той же час сама простота задачі може привести до того, що деякі випробувані будуть думати над нею дуже, дуже довго, довше навіть, ніж над складною. Важку

задачу більшість випробуваних вирішують у тенденції довше, ніж просту, але в той же час майже завжди знаходяться люди, що вирішують її миттєво.

Якщо ми доведемо, що розподіли статистично вірогідно розрізняються, це може стати основою для побудови класифікацій задач і типологій випробуваних. Наприклад, ми можемо виявляти випробуваних зі стандартним співвідношенням ознак: просту задачу вони вирішують швидко, а важку - повільно, - і випробуваних з нестандартним співвідношенням: просту задачу вирішують повільно, а важку - швидко і т.п. Далі ми можемо порівняти виявлені групи випробуваних по показниках мотивації досягнення, тому що відомо, що особи з перевагою прагнення до успіху віддають перевагу задачі середніх труднощів, де імовірність успіху приблизно 0.5, а особи з перевагою прагнення уникати невдачі віддають перевагу або дуже легким, або, навпаки, дуже важким задачам. Таким чином, і тут зіставлення форм розподілу може дати початок науковому пошуку.

Часто нам буває корисно також зіставити отриманий емпіричний розподіл з теоретичним розподілом. Наприклад, для того, щоб довести, що воно підкоряється, чи навпаки, не підкоряється нормальному закону розподілу. Це краще робити за допомогою комп'ютерних програм обробки даних, особливо при великих обсягах вибірок.

У практичних цілях емпіричні розподіли повинні перевірятися на "нормальність" у тих випадках, коли ми маємо намір використовувати параметричні методи і критерії.

Традиційні для вітчизняної математичної статистики критерії визначення розбіжності чи згоди розподілів - це метод χ^2 К.Пірсона і критерій λ Колмогорова-Смірнова.

Обидва ці методи вимагають ретельного групування даних і досить складних обчислень. Крім того, можливості цих критеріїв повною мірою виявляються на великих вибірках ($n \geq 30$). Проте вони можуть виявитися настільки незамінними, що досліднику прийдеться зневажити економію часу і зусиль. Наприклад, вони незамінні в наступних двох випадках:

1) у задачах, що вимагають доказів не випадковості пріоритетів у виборі з декількох альтернатив;

2) у задачах, що вимагають виявлення крапки максимальної розбіжності між двома розподілами, яку потім використовують для перегрупування даних з метою застосування критерію ϕ^* (кутового перетворення Фішера).

Розглянемо метод χ^2 К.Пірсона - традиційний метод визначення розбіжності розподілів.

3.2 χ^2 - критерій Пірсона

Призначення критерію

Критерій χ^2 найбільше часто застосовується в двох цілях:

- 1) для зіставлення *емпіричного* розподілу ознаки з *теоретичним* - рівномірним, нормальним чи якимсь іншим;
- 2) для зіставлення *двох, трьох чи більш емпіричних* розподілів одного і того ж ознаки.

Опис критерію

Критерій χ^2 відповідає на запитання про те, чи з однаковою частотою зустрічаються різні значення ознаки в емпіричному і теоретичному розподілах чи у двох і більш емпіричних розподілах.

Перевага методу складається в тому, що він дозволяє зіставляти розподіл ознак, представлених у будь-якій шкалі, починаючи від шкали найменувань. У найпростішому випадку альтернативного розподілу "так - ні", "допустив помилку - не допустив помилку", "вирішив задачу - не вирішив задачу" і т.і. ми вже можемо застосувати критерій χ^2 .

Допустимо, деякий спостерігач фіксує кількість пішоходів, обравших праву чи ліву з двох симетричних доріжок на шляху з крапки А в крапку Б (див. Рис. 4.3).

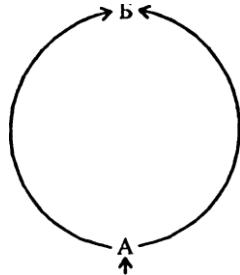


Рис. 4.3. Ілюстрація до приклада про теоретично рівноімовірний вибір із двох альтернатив - правої і лівої доріжок, що ведуть із крапки А в крапку Б

Допустимо, у результаті 70 спостережень установлено, що 51 чоловік вибрали праву доріжку, і лише 19 - ліву. За допомогою критерію χ^2 ми можемо визначити, чи відрізняється даний розподіл виборів від рівномірного розподілу, при якому обидві доріжки вибиралися б з однаковою частотою. Це варіант зіставлення отриманого *емпіричного* розподілу з *теоретичним*. Така задача може стояти, наприклад, у прикладних психологічних дослідженнях, зв'язаних із проектуванням в архітектурі, системах повідомлення та ін.

Але уявимо собі, що спостерігач вирішує зовсім іншу задачу: він зайнятий проблемами білатерального регулювання. Збіг отриманого розподілу з рівномірним його цікавить набагато в меншому ступені, чим збіг чи розбіжність його даних з даними інших дослідників. Йому відомо, що люди з перевагою правої ноги схильні робити коло проти годинної стрілки, а люди з перевагою лівої ноги - коло по ходу годинної стрілки, і що в дослідженні колег перевага лівої ноги була виявлена у 26 чоловік з 100 обстежених.

За допомогою методу χ^2 він може зіставити два емпіричних розподіли: співвідношення 51:19 у власній вибірці і співвідношення 74:26 у вибірці інших дослідників.

Це варіант *зіставлення двох емпіричних* розподілів по найпростішій альтернативній ознаці (звичайно, найпростішому з математичної точки зору, а аж ніяк не психологічній).

Аналогічним образом ми можемо зіставляти розподілу виборів із трьох і більш альтернатив.

Наприклад, якщо у вибірці з 50 чоловік 30 вибрали відповідь (а), 15 чоловік - відповідь (б) і 5 чоловік - відповідь (в), то ми можемо за допомогою методу χ^2 перевірити, чи відрізняється цей розподіл від рівномірного розподілу чи від розподілу відповідей в іншій вибірці, де відповідь (а) вибрали 10 чоловік, відповідь (б) - 25 чоловік, відповідь (в) - 15 чоловік.

У тих випадках, якщо ознака вимірюється кількісно, скажемо, у балах, чи секундах міліметрах, нам прийдеться об'єднати усі значення ознаки в кілька розрядів. Наприклад, якщо час рішення задачі варіює від 10 до 300 секунд, то ми можемо ввести 10 чи 5 розрядів, у залежності від обсягу вибірки. Наприклад, це будуть розряди: 0-50 секунд; 51-100 секунд; 101-150 секунд і т.д. Потім ми за допомогою методу χ^2 будемо зіставляти частоти зустрічальності різних розрядів ознаки, але в іншому принципова схема не міняється.

При зіставленні емпіричного розподілу з теоретичним ми визначаємо ступінь розбіжності між емпіричними і теоретичними частотами.

При зіставленні двох емпіричних розподілів ми визначаємо ступінь розбіжності між емпіричними частотами і теоретичними частотами, що спостерігалися б у випадку збігу двох цих емпіричних

розподілів. (Формули розрахунку теоретичних частот будуть спеціально дані для кожного варіанта зіставлень.)

Чим більше розбіжність між двома розподілами, що зіставляються, тим більше емпіричне значення χ^2 .

Гіпотези

Можливі кілька варіантів гіпотез, у залежності від задач, що ми перед собою ставимо.

Перший варіант:

H₀: Отриманий емпіричний розподіл ознаки не відрізняється від теоретичного (наприклад, рівномірного) розподілу.

H₁: Отриманий емпіричний розподіл ознаки відрізняється від теоретичного розподілу.

Другий варіант:

H₀: Емпіричний розподіл 1 не відрізняється від емпіричного розподілу 2.

H₁: Емпіричний розподіл 1 відрізняється від емпіричного розподілу 2.

Третій варіант:

H₀: Емпіричні розподіли 1, 2, 3, ... не розрізняються між собою.

H₁: Емпіричні розподіли 1, 2, 3, ... розрізняються між собою.

Критерій χ^2 дозволяє перевірити всі три варіанти гіпотез.

Обмеження критерію

1. Обсяг вибірки повинний бути досить великим: $n \geq 30$. При $n < 30$ критерій χ^2 дає дуже наближені значення. Точність критерію підвищується при великих n .
 2. Теоретична частота для кожного осередку таблиці не повинна бути менше 5: $f \geq 5$. Це означає, що якщо число розрядів задане заздалегідь і не може бути змінено, то ми не можемо застосовувати метод χ^2 , не накопичивши визначеного мінімального числа спостережень. Якщо, наприклад, ми хочемо перевірити наші припущення про те, що частота звертань у телефонну службу Довіри нерівномірно розподіляються по 7 дням тижня, то нам буде потрібно $5 \cdot 7 = 35$ звертань. Таким чином, якщо кількість розрядів (k) задано заздалегідь, як у даному випадку, мінімальне число спостережень (n_{\min}) визначається по формулі: $n_{\min} = k \cdot 5$.
 3. Обрані розряди повинні "вичерпувати" весь розподіл, тобто охоплювати весь діапазон варіативності ознак. При цьому групування на розряди повинне бути однаковим у всіх розподілах, що зіставляються.
 4. Необхідно вносити "виправлення на безперервність" при зіставленні розподілів ознак, що приймають усього 2 значення. При внесенні поправки значення χ^2 зменшується (див. Приклад з виправленням на безперервність).
 5. Розряди повинні бути неперехресними: якщо спостереження віднесені до одного розряду, то воно вже не може бути віднесено ні до якого іншого розряду.
- Сума спостережень по розрядах завжди повинна дорівнювати загальній кількості спостережень.

Алгоритм розрахунку критерію Пірсона

1. Занести в таблицю найменування розрядів і відповідні їм емпіричні частоти (перший стовпець).
2. Поруч із кожною емпіричною частотою записати теоретичну частоту (другий стовпець).

3. Підрахувати різниці між емпіричною й теоретичною частотою по кожному розряду(рядку) і записати їх у третій стовпець.
4. Визначити число ступенів волі по формулі: $\nu = k - 1$, де k – кількість розрядів ознаки.
Якщо $\nu = 1$, внести виправлення на «безперервність», замінивши $(f_{Ei} - f_T)$ на $(f_{Ei} - f_T - 0,5)$.
5. Піднести до квадрата отримані різниці й занести їх у четвертий стовпець.
6. Розділити отримані квадрати різниць на теоретичну частоту й записати результати в п'ятий стовпець.
7. Знайти суму значень п'ятого стовпця. Отриману суму позначити як $\chi^2_{\text{д.і.}}$.
8. Визначити за табл. критичні значення для даного числа ступенів волі.
Якщо $\chi^2_{\text{д.і.}}$ менше критичного значення, розбіжності між розподілами статистично недостовірні.
Якщо $\chi^2_{\text{д.і.}}$ дорівнює критичному значенню або перевищує його, розбіжності між розподілами статистично достовірні.

Приклад.

У таблиці дані частоти звернень до телефонної служби Довіри. Висунуто припущення про те, що частоти звернень до телефонної служби Довіри нерівномірно розподіляються по 7 дням тижня. Перевірити це припущення за допомогою критерію Пірсона.

День тижня	Понеділок	Вівторок	Середа	Четвер	П'ятниця	Субота	Неділя
Частота звернень	6	8	5	9	4	5	3

Рішення.

Сформулюємо гіпотези:

H_0 : Розподіл частот звернень до телефонної служби Довіри по 7 дням тижня не відрізняється від рівномірного розподілу.

H_1 : Розподіл частот звернень до телефонної служби Довіри по 7 дням тижня відрізняється від рівномірного розподілу.

Визначимо теоретичну частоту звернень до служби Довіри при рівномірному розподілі. Для цього складемо частоти всіх днів тижня й розділимо на 7 днів тижня.

$$f_T = \frac{n}{k}, \quad \text{де } n - \text{кількість спостережень; } k - \text{кількість розрядів ознаки.}$$

$$f_T = \frac{n}{k} = \frac{6+8+5+9+4+5+3}{7} = \frac{40}{7} = 5,71$$

Складемо таблицю

	$f_{\text{д.і.}}$	f_T	$(f_{\text{д.і.}} - f_T)$	$(f_{\text{д.і.}} - f_T)^2$	$(f_{\text{д.і.}} - f_T)^2 / f_T$

Понеділок	6	5,71	0,29	0,08	0,01
Вівторок	8	5,71	2,29	5,22	0,91
Середа	5	5,71	- 0,71	0,51	0,09
Четвер	9	5,71	3,29	10,8	1,89
П'ятниця	4	5,71	- 1,71	2,94	0,51
Субота	5	5,71	- 0,71	0,51	0,09
Неділя	6	5,71	- 2,71	7,37	1,29
СУМА	40				4,79

$$\chi^2_{\hat{a} \hat{i}} = \sum \frac{(f_{\hat{y} \hat{i}} - f_T)^2}{f_T} = 4,79$$

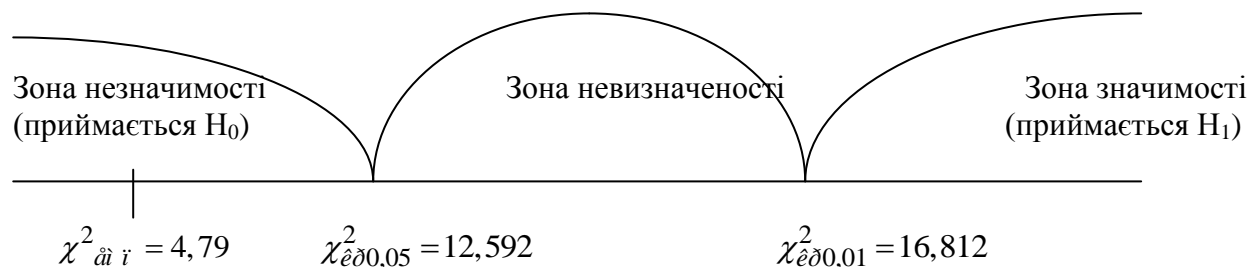
Для того, щоб установити критичні значення χ^2 , потрібно визначити число ступенів волі ν за формулою: $\nu = k - 1$, де k – кількість розрядів. У нашому випадку $\nu = 7 - 1 = 6$.

p			p		
ν	0,05	0,01	ν	0,05	0,01
1	3.841	6.635	9	16.919	21.666
2	5.991	9.210	10	18.307	23.209
3	7.815	11.345	11	19.675	24.725
4	9.488	13.277	12	21.026	26.217
5	11.070	15.086	13	22.362	27.688
6	12.592	16.812	14	23.685	29.141
7	14.067	18.475	15	24.996	30.578
8	15.507	20.090	16	26.296	32.000

$$\chi^2_{\hat{e} \hat{d} 0,05} = 12,592$$

$$\chi^2_{\hat{e} \hat{d} 0,01} = 16,812$$

Побудуємо «вісь значимості».



Тому що $\chi^2_{\hat{y} \hat{i}}$ потрапило в зону незначимості, приймаємо гіпотезу H_0 : Розподіл частот звернень до телефонної служби Довіри за 7 днями тижня не відрізняється від рівномірного розподілу.

Висновок: висловлене припущення про те, що частоти звернень до телефонної служби Довіри нерівномірно розподіляються за 7 днями тижня, було невірне.

Тема 5. Метод рангової кореляції

Кореляційний зв'язок - це погоджені зміни двох ознак чи більшої кількості ознак (множинний кореляційний зв'язок). Кореляційний зв'язок відображає той факт, що мінливість однієї ознаки знаходиться в деякій відповідності з мінливістю іншого.

Кореляційна залежність - це зміни, що вносять значення однієї ознаки в імовірність появи різних значень іншої ознаки.

У кореляційних зв'язках кожному значенню однієї ознаки може відповідати визначений розподіл значень іншої ознаки, але не визначене його значення.

Обидва терміни - кореляційний зв'язок і кореляційна залежність - часто використовуються як синоніми. Тим часом, погоджені зміни ознак і кореляційний зв'язок, що відображає це, між ними може свідчити не про залежність цих ознак між собою, а залежності обох цих ознак від якоїсь третьої ознаки чи сполучення ознак, не розглянутих у дослідженні.

Залежність має на увазі вплив, зв'язок - будь-які погоджені зміни, що можуть порозуміватися сотнями причин. Кореляційні зв'язки не можуть розглядатися як свідчення причинно-наслідкового зв'язку, вони свідчать лише про те, що змінам однієї ознаки, як правило, супроводжують визначені зміни іншого, але чи знаходиться причина змін в одній з ознак чи вона виявляється за межами досліджуваної пари ознак, нам невідомо.

Ступінь, сила чи тіснота кореляційного зв'язку визначається по величині коефіцієнта кореляції.

Сила зв'язку не залежить від її спрямованості і визначається за абсолютним значенням коефіцієнта кореляції. Максимальне можливе абсолютне значення коефіцієнта кореляції $r = 1,00$; мінімальне $r = 0$.

Використовується дві системи класифікації кореляційних зв'язків по їхній силі: загальна і часткова.

Загальна класифікація кореляційних зв'язків :

- | | |
|---------------------|--------------------------|
| 1) сильна, чи тісна | при $ r > 0,7$; |
| 2) середня | при $0,5 < r < 0,69$; |
| 3) помірна | при $0,3 < r < 0,49$; |
| 4) слабка | при $0,2 < r < 0,29$; |
| 5) дуже слабка | при $ r < 0,19$. |

Часткова класифікація кореляційних зв'язків:

- | | |
|-----------------------------------|---|
| 1) висока значима кореляція | при r , що відповідає рівню статистичної значимості $p \leq 0,01$; |
| 2) значима кореляція | при r , що відповідає рівню статистичної значимості $p \leq 0,05$; |
| 3) тенденція достовірного зв'язку | при r , що відповідає рівню статистичної значимості $p \leq 0,1$; |
| 4) незначуща кореляція | при r , не сягаючого рівня статистичної значимості. |

Дві ці класифікації не збігаються. Перша орієнтована тільки на величину коефіцієнта кореляції, а друга визначає, якого рівня значимості досягає дана величина коефіцієнта кореляції при даному обсязі вибірки. Чим більше обсяг вибірки, тим меншої величини коефіцієнта кореляції виявляється досить, щоб кореляція була визнана достовірною. У результаті при малому обсязі вибірки може виявитися, що сильна кореляція виявиться недостовірною. У той же час при великих обсягах вибірки навіть слабка кореляція може виявитися достовірною.

Звичайно прийнято орієнтуватися на другу класифікацію, оскільки вона враховує обсяг вибірки. Разом з тим, необхідно пам'ятати, що сильна, чи висока, кореляція - це кореляція з коефіцієнтом $|r| > 0,7$, а не просто кореляція високого рівня значимості.

Коефіцієнт рангової кореляції r_s Спірмена

Призначення рангового коефіцієнта кореляції

Метод рангової кореляції Спірмена дозволяє визначити тісноту (силу) і напрямок кореляційного зв'язку між двома ознаками чи двома профілями (ієрархіями) ознак.

Опис методу

Для підрахунку рангової кореляції необхідно мати два ряди значень, що можуть бути проранжировані.

Спочатку ранжуються індивідуальні значення по першій ознаці, отримані різними випробуваними, а потім індивідуальні значення по другій ознаці.

Якщо дві ознаки зв'язані позитивно, то випробувані, що мають низькі ранги по одному з них, будуть мати низькі ранги і по іншому, а випробувані, які мають високі ранги за одною ознакою, будуть мати по іншій ознаці також високі ранги. Для підрахунку r_s необхідно визначити різниці (d) між рангами, отриманими даним випробуваним по обох ознаках. Потім ці показники d певним чином перетворюються і віднімаються з 1. Чим менше різниці між рангами, тим більше буде r_s , тим ближче він буде до +1.

Якщо кореляція відсутня, то всі ранги будуть перемішані і між ними не буде ніякої відповідності. Формула складена так, що в цьому випадку r_s виявиться близьким до 0.

У випадку негативної кореляції низьким рангам випробуваних по одній ознаці будуть відповідати високі ранги по іншій ознаці, і навпаки.

Чим більше розбіжність між рангами випробуваних по двох перемінним, тим ближче r_s до (-1).

Значимість отриманого коефіцієнта кореляції визначається по кількості ранжированих значень N .

Ця кількість буде збігатися з обсягом вибірки n . Якщо абсолютна величина r_s досягає критичного значення чи перевищує його, кореляція достовірна.

Гіпотези

H_0 : Кореляція між перемінними А і В не відрізняється від нуля.

H_1 : Кореляція між перемінними А і В вірогідно відрізняється від нуля.

Коефіцієнт рангової кореляції Спірмена r_s при великій кількості однакових рангів по одній чи обох перемінним дає приблизні значення. В ідеалі обидва ряди повинні являти собою дві послідовності незбіжних значень. У випадку, якщо ця умова не дотримується, необхідно вносити виправлення на однакові ранги.

Алгоритм підрахунку коефіцієнта рангової кореляції r_s Спірмена

1. Визначити, які дві ознаки будуть брати участь у зіставленні як перемінні А і В.
2. Проранжировувати значення перемінної А, нараховуючи ранг 1 найменшому значенню, відповідно до правил ранжирування. Занести ранги в перший стовпець таблиці один по одному номерів випробуваних чи ознак.
3. Проранжировувати значення перемінної В, відповідно до тих же правил. Занести ранги в другий стовпець таблиці один по одному номерів випробуваних чи ознак.
4. Підрахувати різниці d між рангами А і В по кожному рядку таблиці і занести в третій стовпець таблиці.
5. Звести кожен різницю в квадрат: d^2 . Ці значення занести в четвертий стовпець таблиці.

6. Підрахувати суму квадратів Σd^2 .

7. При наявності однакових рангів розрахувати виправлення:

$$T_a = \frac{\sum (a^3 - a)}{12} \quad T_b = \frac{\sum (b^3 - b)}{12}$$

де a - обсяг кожної групи однакових рангів у ранговому ряді А;

b - обсяг кожної групи однакових рангів у ранговому ряді В.

8. Розрахувати коефіцієнт рангової кореляції r_s за формулою:

а) при відсутності однакових рангів

$$r_{s\hat{a}i} = 1 - 6 \cdot \frac{\sum d^2}{N \cdot (N^2 - 1)}$$

б) при наявності однакових рангів

$$r_{s\hat{a}i} = 1 - 6 \cdot \frac{\sum d^2 + T_a + T_b}{N \cdot (N^2 - 1)}$$

Σd^2 – сума квадратів різниць між рангами;

T_a і T_b - виправлення на однакові ранги;

N - кількість випробуваних чи ознак, що брали участь у ранжируванні.

9. Визначити по Табл. критичні значення r_s для даного N . Якщо r_s перевищує критичне значення чи принаймні дорівнює йому, кореляція вірогідно відрізняється від 0.

Таблиця. Критичні значення вибіркового коефіцієнта рангової кореляції

n	p		n	p		n	p	
	0,05	0,01		0,05	0,01		0,05	0,01
5	0,94	-	17	0,48	0,62	29	0,37	0,48
6	0,85	-	18	0,47	0,60	30	0,36	0,47
7	0,78	0,94	19	0,46	0,58	31	0,36	0,46
8	0,72	0,88	20	0,45	0,57	32	0,36	0,45
9	0,68	0,83	21	0,44	0,56	33	0,34	0,45
10	0,64	0,79	22	0,43	0,54	34	0,34	0,44
11	0,61	0,76	23	0,42	0,53	35	0,33	0,43
12	0,58	0,73	24	0,41	0,52	36	0,33	0,43
13	0,56	0,70	25	0,40	0,51	37	0,33	0,43
14	0,54	0,68	26	0,39	0,50	38	0,32	0,41
15	0,52	0,66	27	0,38	0,49	39	0,32	0,41
16	0,50	0,64	28	0,38	0,48	40	0,31	0,40

Приклад

У дослідженні випробувані повинні були вирішувати задачі на вибір оптимального типу злітно-посадочної смуги для заданого типу літака. Чи зв'язана кількість помилок, допущених випробуваними в тренувальній сесії, з показниками вербального інтелекту, обмірjованими за методикою Д. Векслера?

Показники кількості помилок у тренувальній сесії і показники рівня вербального інтелекту в студентів-фізиків.

	Випробувані	Кількість помилок	Показник вербального інтелекту
--	-------------	-------------------	--------------------------------

1.	Т.А.	29	131
2.	П.А.	54	132
3.	Ч.И.	13	121
4.	Ц.А.	8	127
5.	К.М.	14	136
6.	П.Н.	26	124
7.	Р.А.	9	134
8.	Д.Л.	20	136
9.	Р.О.	2	132
10.	Г.Е.	17	136

Сформулюємо гіпотези.

H_0 : Кореляція між показником кількості помилок у тренувальній сесії і рівнем вербального інтелекту не відрізняється від нуля.

H_1 : Кореляція між показником кількості помилок у тренувальній сесії і рівнем вербального інтелекту статистично значимо відрізняється від нуля.

Розрахунок d^2 для рангового коефіцієнта кореляції Спірмена r_s при зіставленні показників кількості помилок і вербального інтелекту в студентів-фізиків ($N=10$):

	Випробуваний	Змінна А: кількість помилок		Змінна В: вербальний інтелект		$d(\text{ранг А}-\text{ранг В})$	d^2
		значення	ранг	значення	ранг		
1.	Т.А.	29	9	131	4	5	25
2.	П.А.	54	10	132	5,5	4,5	20,25
3.	Ч.И.	13	4	121	1	3	9
4.	Ц.А.	8	2	127	3	-1	1
5.	К.М.	14	5	136	9	-4	16
6.	П.Н.	26	8	124	2	6	36
7.	Р.А.	9	3	134	7	-4	16
8.	Д.Л.	20	7	136	9	-2	4
9.	Р.О.	2	1	132	5,5	-4,5	20,25
10.	Г.Е.	17	6	136	9	-3	9
сума			55	сума	55	сума	156,5

Коефіцієнт рангової кореляції Спірмена підраховується за формулою:

$$r_{sei} = 1 - 6 \cdot \frac{\sum d^2}{N \cdot (N^2 - 1)} = 1 - 6 \cdot \frac{156,5}{10 \cdot (10^2 - 1)} = 1 - \frac{939}{990} = 0,052$$

де d - різниця між рангами по двох перемінним для кожного випробуваного;

N - кількість ранжируваних значень, у даному випадку кількість випробуваних.

Отримане емпіричне значення r_s близько до 0.

Визначимо критичні значення r_s при $N=10$ за Табл. 1: $r_{s \text{ } \hat{=} 0,05} = 0,64$, $r_{s \text{ } \hat{=} 0,01} = 0,79$

$$r_{s \text{ емп}} \geq r_{s \text{ кр}}$$

Відповідь: H_0 приймається. Кореляція між показником кількості помилок у тренувальній сесії і рівнем вербального інтелекту не відрізняється від нуля: кількість помилок, допущених випробуваними в тренувальній сесії, не зв'язана з показниками вербального інтелекту.

Тема 6. Дисперсійний аналіз

Поняття дисперсійного аналізу

Дисперсійний аналіз – це аналіз мінливості ознаки під впливом яких-небудь контрольованих перемінних факторів. У зарубіжній літературі дисперсійний аналіз часто позначається як ANOVA, що перекладається як аналіз варіативності (Analysis of Variance). Автором методу є Р.А.Фішер (1918,1938).

Задача дисперсійного аналізу полягає в тому, щоб із загальної варіативності ознаки вичленувати варіативність троякого роду:

- а) варіативність, обумовлену дією кожної з досліджуваних незалежних перемінних;
- б) варіативність, обумовлену взаємодією досліджуваних незалежних перемінних;
- в) випадкову варіативність, обумовлену всіма іншими невідомими перемінними.

Варіативність, обумовлена дією досліджуваних перемінних і їхньою взаємодією, співвідноситься з випадковою варіативністю.

Показником цього співвідношення є критерій F Фішера.

(Примітка. Критерій F Фішера і метод кутового перетворення Фішера – це зовсім різні методи, що мають різне призначення і різні способи обчислення.)

$$F_{\text{емп A}} = \frac{\text{Варіативність, обумовлена перемінної A}}{\text{Випадкова варіативність}}$$

$$F_{\text{емп B}} = \frac{\text{Варіативність, обумовлена перемінної B}}{\text{Випадкова варіативність}}$$

$$F_{\text{емп AB}} = \frac{\text{Варіативність, обумовлена взаємодією перемінних A и B}}{\text{Випадкова варіативність}}$$

У формулу розрахунку критерію F входять оцінки дисперсій, тобто параметрів розподілу ознаки, тому критерій F є параметричним критерієм.

Чим у більшій ступені варіативність ознаки обумовлена досліджуваними перемінними (факторами) чи їх взаємодією, тим вище емпіричні значення критерію F.

Підготовка даних до дисперсійного аналізу

1) Створення комплексів

Найкраще для кожного випробуваного створити окрему картку, куди були б занесені дані по всіх досліджених ознаках. Справа в тому, що в процесі аналізу у дослідника можуть змінитися гіпотези. Буде потрібно створювати, бути може, не один, а безліч дисперсійних комплексів, що розрізняються як по факторах, так і по результативних ознаках. Картки допоможуть досліднику швидко створювати нові дисперсійні комплекси. Завдяки карткам ми відразу побачимо, чи рівномірно розподіляються дані по градаціях у випадку, якщо за фактор ми вирішили прийняти один з досліджених психологічних ознак. За допомогою карток ми можемо допомогти собі виділити три, чотири чи більш градації цього фактора, наприклад, рівні мотивації, наполегливості, креативності та ін.

2) Зрівноважування комплексів

Комплекс, у якому кожен осередок представлений однаковою кількістю спостережень, називається рівномірним. Рівномірність комплексу дозволяє нам обійти вимогу рівності дисперсій у кожному з осередків комплексу.

У випадку, якщо в різних градаціях комплексу виявилася нерівна кількість спостережень, необхідно відсіяти деякі з них шляхом випадкового вибору необхідної кількості карток.

3) Перевірка нормальності розподілу результативності ознаки

Перед тим, як застосовувати дисперсійний аналіз, ми повинні переконатися, що розподіл ознаки є нормальним. Нормальність розподілу можна перевірити шляхом розрахунку показників асиметрії й ексцесу і зіставлення їх із критичними значеннями.

Показники асиметрії та ексцесу свідчать про достовірну відмінність емпіричних розподілів від нормального в тому випадку, якщо вони перевищують по абсолютній величині свою помилку репрезентативності в 3 і більш раз:

$$t_A = \frac{|A|}{m_A} \geq 3 \quad t_E = \frac{|E|}{m_E} \geq 3$$

Показники асиметрії й ексцесу з їхніми помилками репрезентативності визначаються по наступним формулам:

$$A = \frac{\sum (x_i - \bar{x})^3}{n \cdot \sigma^3} \quad m_A = \sqrt{\frac{6}{n}}$$

$$E = \frac{\sum (x_i - \bar{x})^4}{n \cdot \sigma^4} - 3 \quad m_E = 2 \cdot \sqrt{\frac{6}{n}}$$

де x_i – кожне значення ознаки, що спостерігається;

\bar{x} – середнє арифметичне;

n – кількість спостережень;

σ – середнє квадратичне відхилення $\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$

4) Перетворення емпіричних даних з метою спрощення розрахунків

Можливі наступні перетворення:

А) усі значення, що спостерігаються, можна розділити на те саме число k ;

Б) усі значення, що спостерігаються, можна помножити на те саме число k ;

В) від усіх значень, що спостерігаються, можна відняти те саме число A , наприклад найменше значення;

Г) можна зробити подвійне перетворення: з кожного значення відняти число A , а отриманий результат розділити на інше число k .

При всіх цих перетвореннях результативної ознаки показники співвідношення дисперсій виходять точними і не вимагають ніяких виправлень.

Середні величини змінюються, але їх можна відновити, множачи середню величину на число k чи поділяючи її на k , чи додаючи до середнього число A . Стандартне відхилення змінюється тільки при введенні множника або дільника; отриманий результат потім прийдеться або розділити на число k , або помножити на нього.

Однофакторний дисперсійний аналіз для незв'язаних вибірок

Призначення методу

Метод однофакторного дисперсійного аналізу застосовується в тих випадках, коли досліджуються зміни результативної ознаки під впливом умов, що змінюються, чи градацій будь-якого фактора. У даному варіанті методу впливу кожної з градацій фактора піддаються різні вибірки випробуваних. Градацій фактора повинне бути не менш трьох.

Непараметричним варіантом цього виду аналізу є критерій Н Крускала-Уолліса.

Опис методу

Роботу починаємо з того, що представляємо отримані дані у виді стовпців індивідуальних значень. Кожний зі стовпців відповідає той чи інший з досліджуваних умов.

Після цього нам потрібно підсумувати індивідуальні значення по стовпцях і суми звести в квадрат.

Суть методу полягає в тому, щоб зіставити суму цих зведених у квадрат сум із сумою квадратів усіх значень, отриманих у всьому експерименті.

Гіпотези

H_0 : Розходження між градаціями фактора (різними умовами) є не більш вираженими, чим випадкові розходження усередині кожної групи.

H_1 : Розходження між градаціями фактора (різними умовами) є більш вираженими, чим випадкові розходження усередині кожної групи.

Обмеження методу

1. Однофакторний дисперсійний аналіз вимагає не менш трьох градацій фактора і не менш двох випробуваних у кожній градації.
2. Повинне дотримуватися правило рівності дисперсій у кожному осередку дисперсійного комплексу.
3. Результативна ознака повинна бути нормально розподілена у досліджуваній вибірці.

Послідовність операцій в однофакторному дисперсійному аналізі

1. Підрахувати $SS_{\text{факт}}$

$$SS_{\text{факт}} = \frac{\sum T_c^2}{n} - \frac{(\sum x_s)^2}{N}$$

де

x_i - індивідуальні значення

T_c - суми індивідуальних значень по кожному з умов;

n - кількість випробуваних у кожній групі

N - загальна кількість індивідуальних значень

2. Підрахувати $SS_{\text{заг}}$ (сума квадратів загальна)

$$SS_{\text{заг}} = \sum x_i^2 - \frac{(\sum x_i)^2}{N}$$

- 3 Підрахувати випадкову залишкову величину $SS_{\text{вип}}$ (сума квадратів випадкова)

$$SS_{\text{вип}} = SS_{\text{заг}} - SS_{\text{факт}}$$

- 4 Визначити число ступенів волі.

$$df_{\text{факт}} = 3 - 1$$

$$df_{\text{заг}} = N - 1$$

$$df_{\text{вип}} = df_{\text{заг}} - df_{\text{факт}}$$

- 5 Розділити кожен суму квадратів на відповідне число ступенів волі

$$MS_{\text{факт}} = SS_{\text{факт}} / df_{\text{факт}}$$

$$MS_{\text{вип}} = SS_{\text{вип}} / df_{\text{вип}}$$

- 6 Підрахувати значення $F_{\text{емп}}$

$$F_{\text{емп}} = MS_{\text{факт}} / MS_{\text{вип}}$$

- 7 Визначити з таблиці критичні значення $F_{\text{кр}}$.

- 8 Зіставити емпіричне і критичне значення $F_{\text{кр}}$.

Якщо $F_{\text{емп}} > F_{\text{кр}}$. Гіпотеза H_0 відхиляється. Гіпотеза H_1 приймається.

Приклад.

Три різні групи з шести випробуваних одержали списки з десяти слів. Першій групі слова пред'являлися з низькою швидкістю, другій групі із середньою швидкістю й третій групі з великою швидкістю. Було припущено, що показники відтворення будуть залежати від швидкості пред'явлення слів. Результати представлені у таблиці.

№ випробуваного	Група 1: низька швидкість	Група 2: середня швидкість	Група 3: висока швидкість
1	8	7	4
2	7	8	5
3	9	5	3
4	5	4	6
5	6	6	2
6	8	7	4
Суми	43	37	24
Загальна сума	$(\sum \tilde{o}_j) = 43 + 37 + 24 = 104$		

Сформулюємо гіпотези:

H_0 : Розходження в обсязі відтворення слів між групами є не більш вираженими, чим випадкові розходження усередині кожної групи.

H_1 : Розходження в обсязі відтворення слів між групами є більш вираженими, чим випадкові розходження усередині кожної групи.

Розрахуємо основні величини для однофакторного дисперсійного аналізу.

1. T_c – суми індивідуальних значень по кожній з умов
 $T_1=43; \quad T_2=37; \quad T_3=24.$
2. $\sum(T_c^2)$ – сума квадратів сумарних значень по кожній з умов:
 $\sum(T_c^2) = 43^2 + 37^2 + 24^2 = 3794$
3. Визначимо c – кількість умов (градацій фактора): $c = 3$;
 n – кількість випробуваних у кожній групі: $n = 6$;
 N – загальна кількість випробуваних: $N = 18.$
4. $(\sum x_i)^2$ – квадрат загальної суми індивідуальних значень
 $(\sum x_i)^2 = 104^2$
5. $\frac{(\sum x_i)^2}{N}$ – константа, яку треба відняти з кожної суми квадратів

$$\frac{(\sum x_i)^2}{N} = \frac{104^2}{18} = 600,89$$
6. $\sum(x_i^2)$ – сума квадратів індивідуальних значень
 $\sum(x_i^2) = 8^2 + 7^2 + 9^2 + 5^2 + 6^2 + 8^2 + 7^2 + 8^2 + 5^2 + 4^2 + 6^2 + 7^2 + 4^2 + 5^2 + 3^2 + 6^2 + 2^2 + 4^2 = 664$

Зауваження. Відзначимо різницю між $\sum(x_i^2)$, у якій всі індивідуальні значення спочатку зводяться у квадрат, а потім підсумуються, і $(\sum x_i)^2$, де індивідуальні значення спочатку підсумуються, а потім уже ця сума зводиться у квадрат.

Тепер послідовно **виконаємо розрахунки:**

1. Підрахувати $SS_{\text{факт}}$

$$SS_{\text{факт}} = \frac{\sum T_c^2}{n} - \frac{(\sum x_i)^2}{N} = \frac{3794}{6} - \frac{104^2}{18} = 31,44$$

де

x_i – індивідуальні значення

T_c – суми індивідуальних значень по кожній з умов;

n – кількість випробуваних у кожній групі

N – загальна кількість індивідуальних значень

2. Підрахувати $SS_{\text{заг}}$

$$SS_{\text{заг}} = \sum x_i^2 - \frac{(\sum x_i)^2}{N} = 664 - 600,89 = 63,11$$

3. Підрахувати випадкову залишкову величину $SS_{\text{вип}}$

$$SS_{\text{вип}} = SS_{\text{заг}} - SS_{\text{факт}} = 63,11 - 31,44 = 31,67$$

4. Визначити число ступенів волі.

$$df_{\text{факт}} = 3 - 1 = 3 - 1 = 2$$

$$df_{\text{заг}} = N - 1 = 18 - 1 = 17$$

$$df_{\text{вип}} = df_{\text{заг}} - df_{\text{факт}} = 17 - 2 = 15$$

5. Розділити кожен суму квадратів на відповідне число ступенів волі

$$MS_{\text{факт}} = SS_{\text{факт}} / df_{\text{факт}} = 31,44 / 2 = 15,72$$

$$MS_{\text{вип}} = SS_{\text{вип}} / df_{\text{вип}} = 31,67/15 = 2,11$$

6. Підрахувати значення $F_{\text{емп}}$

$$F_{\text{емп}} = MS_{\text{факт}} / MS_{\text{вип}} = 15,72/2,11 = 7,45$$

7. Визначити за таблицею критичні значення $F_{\text{кр}}$

$$df_1 = df_{\text{факт}} = 2$$

$$df_2 = df_{\text{вип}} = 15$$

$$\text{З таблиці виписуємо: } F_{\text{кр } 0,05} = 3,68 \quad F_{\text{кр } 0,01} = 6,36$$

8. Зіставити емпіричне й критичне значення F

Тому що $F_{\text{емп}} > F_{\text{кр}}$ ($7,45 > 6,36$) гіпотеза H_0 відхиляється. Приймається гіпотеза H_1 : розходження в обсязі відтворення слів між групами є більш вираженими, чим випадкові розходження усередині кожної групи.

Висновок: швидкість пред'явлення слів впливає на об'єм їхнього відтворення.

Рекомендована література

Базова

1. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. - М.: Изд-во «Прогресс», 1976. – 496 с.
2. Ермолаев О.Ю. Математическая статистика для психологов. –М.: Московский психолого-социальный институт: Флинта, 2004. – 336 с
3. Лапач С.Н., Чубенко А.В., Бабич П.Н. Статистические методы в медико-биологических исследованиях с использованием Excel. – К.: МОРИОН, 2001.– 408 с.
4. Навч. посіб. / А. Б. Телейко, Р. К. Чорней. — К. : МАУП, 2007. — 424 с.— Бібліогр. : с. 411–412
5. Руденко В. М. Математичні методи в психології [Текст] : підручник / В. М. Руденко, Н. М. Руденко. - К. :Академвидав, 2009. - 384 с. : табл., іл. - (Альма-матер). - Бібліогр.: с. 378-383
6. Остапенко Р. И. Математические основы психологии. Учебно – методическое пособие. – Воронеж.: ВГПУ, 2010. – 76 с., ил.
7. Сидоренко Е.В. Методы математической обработки в психологии. СПб.: ООО «Речь», 2001. -350с.

Допоміжна

8. Артемьева Е.Ю. Вероятностные методы в психологии. / Е.Ю. Артемьева, Е.М. Мартыанов. – Изд-во Московского университета. – М., 1975. – 206 с.
9. Артемьева Е.Ю. Сборник задач по теории вероятностей и математической статистике для психологии. – Изд-во Московского университета. – М., 1965. – 90 с.
10. Афифи А., Эйзен С. Статистический анализ. – М.: Мир 1982. – 450 с.
11. Вайнберг Д., Шуменер Д. Статистика. – М.: Статистика, 1979. – 390 с.
12. Калинина В.Н., Панкин В.Ф. Математическая статистика. Учебное пособие для вузов. – М.: Высш. Школа, 1994. – 336 с.
13. Лупандин В. И., Зайцев А. В. Сборник задач по курсу «Математические методы в психологии». Учебно – методическое пособие. Екатеринбург: УрГУ, 2000 г.
14. Макаров Ф.А., Тюрин Ю.Н. Анализ данных на компьютере. – М.: Финансы и статистика, 1995. – 400 с.
15. Минзов А. С. Математические методы в психологии. Контрольные работы и методические указания по их выполнению / Под ред. В. В. Грачева. – М. : Издательство СГИ, 2000. – 40 с.
16. Митина О. В. Математические методи в психологи: Практикум / О. В. Митина. – М.: Аспект Пресс, 2008. – 238 с.
17. Погребницкая М.В. Математические методы в психологии: Учебно-методическое пособие. – Петропавловск, СКГУ, 2004.- 228 с.
18. Статистические методы в психологи: учебно – методический комплекс / сост. Ю. В. Насонова. – Витебск: УО «ВГУ им. П. М, Машерова», 2010. – 237 с.