НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ЦИВІЛЬНОГО ЗАХИСТУ УКРАЇНИ

Кафедра психології діяльності в особливих умовах

В. Ф. Боснюк

МАТЕМАТИЧНІ МЕТОДИ В ПСИХОЛОГІЇ

Навчальний посібник

Черкаси
2025

УДК 519.22(075.8)
Б 85

Рекомендовано до друку вченою радою
Національного університету цивільного захисту України,
(протокол від 26.09.2025. № 1)

Рецензенти:
В. О. Олефір, доктор психологічних наук, доцент, завідувач кафедри загальної психології Харківського національного університету ім. В.Н. Каразіна
В. В. Кердивар, доктор філософії (PhD) з психології, начальник навчально-наукової лабораторії екстремальної та кризової психології навчально-наукового інституту оперативно-рятувальних сил Національного університету цивільного захисту України

Боснюк В. Ф.
Б 85 Математичні методи в психології: навчальний посібник. Черкаси : НУЦЗ України, 2025. 151 с.

Навчальний посібник призначений для здобувачів вищої освіти психологічних спеціальностей і спрямований на формування знань та практичних навичок використання математико-статистичних методів у психологічних дослідженнях.

У виданні послідовно розглядаються питання описової статистики, параметричних і непараметричних критеріїв перевірки гіпотез, методів аналізу взаємозв'язків та відмінностей, а також підходів до інтерпретації результатів. Особливу увагу приділено не лише визначенню статистичної значущості, але й оцінці розміру ефекту, що дозволяє визначати практичну важливість отриманих результатів.

Інтерактивною особливістю посібника є наявність QR-кодів після прикладів застосування статистичних критеріїв. Вони містять відео-огляди роботи у програмному середовищі SPSS Statistics, де покроково продемонстровано процес аналізу даних та інтерпретації результатів. Це робить навчальний матеріал доступним, наочним і практично орієнтованим.

Посібник поєднує теоретичний виклад із прикладами з різних галузей психології, завданнями для самостійної підготовки. Він буде корисним здобувачам вищої освіти, викладачам, практичним психологам та всім, хто використовує кількісні методи у гуманітарних дослідженнях.

УДК 519.22(075.8)

ЗМІСТ

ПЕРЕДМОВА

У сучасних умовах розвитку науки та освіти особливе місце займає психологія як галузь знань, що поєднує гуманітарні та природничо-наукові підходи до вивчення людини. Складність психологічних явищ, багатогранність поведінкових проявів і внутрішніх психічних процесів зумовлюють потребу у використанні точних та об'єктивних методів дослідження. Традиційні описові моделі поступово поступаються місцем системному аналізу, який базується на математичних методах. Це зумовлено не лише вимогами до наукової обґрунтованості результатів, але й потребою інтеграції психології з іншими науками.

Математичні методи сьогодні виступають невід'ємним інструментом психологічної науки. Вони забезпечують можливість адекватного опису психічних явищ, перевірки гіпотез, прогнозування поведінки та прийняття науково обґрунтованих рішень у практиці. Психолог, який володіє базовим арсеналом математичних методів, здатний якісніше будувати дослідницькі програми, аналізувати результати експериментів і впроваджувати ефективні психодіагностичні технології. Знання математичного апарату стає також важливим чинником професійної конкурентоспроможності майбутніх фахівців.

Метою цього навчального посібника є формування у здобувачів вищої освіти системного розуміння сутності математичних методів, що застосовуються в психології, а також розвиток умінь та навичок їх практичного використання у дослідницькій і професійній діяльності. Автор посібника прагнув зробити виклад матеріалу максимально доступним і водночас науково обґрунтованим, поєднавши теоретичні положення з прикладами та завданнями, що відображають реалії психологічних досліджень.

Посібник структуровано відповідно до логіки наукового дослідження: від загальних положень математичної статистики та вимірювання психологічних ознак – до складніших методів аналізу взаємозв'язків, виявлення групових відмінностей. Така послідовність дозволяє читачеві поступово засвоювати матеріал, формуючи цілісне уявлення про можливості й обмеження математичного підходу.

Особливий акцент зроблено на тому, що автор не обмежується лише визначенням статистичної значущості результатів розрахунків. Адже для психологічних досліджень не менш важливим є показник розміру ефекту, який дозволяє оцінити практичну значущість отриманих результатів. Такий підхід формує у здобувачів освіти більш критичне й науково обґрунтоване ставлення до аналізу даних, адже він виходить за межі традиційної схеми «значуще – незначуще» й орієнтує на комплексне осмислення результатів. Такий підхід сприяє формуванню критичного мислення й відповідального ставлення до обробки даних.

Важливою особливістю посібника є його інтерактивність. Після навчальних прикладів, де описані алгоритми застосування певних статистичних критеріїв, розміщено QR-коди, у яких закодовані авторські відео-огляди. У цих відео покроково продемонстровано процедури аналізу даних та інтерпретації результатів у програмному середовищі SPSS Statistics 23.0. Такий підхід дозволяє поєднати теоретичний виклад з практичними діями, забезпечуючи глибше розуміння матеріалу та формування навичок, необхідних для самостійного проведення досліджень.

Посібник орієнтований насамперед на здобувачів вищої освіти психологічних спеціальностей, але буде корисним і для аспірантів (ад'юнктів), викладачів, практичних психологів, а також дослідників у суміжних галузях. Його зміст відповідає сучасним освітнім стандартам та враховує потреби підготовки фахівців, здатних ефективно застосовувати наукові методи у практичній роботі.

Важливо наголосити, що цей навчальний посібник не претендує на вичерпне охоплення усіх аспектів математичної статистики. Його завдання – надати базові знання та сформувати у здобувачів навички, необхідні для подальшого поглибленого вивчення спеціалізованих методів. Автори прагнули знайти баланс між математичною строгістю та зрозумілістю викладу, щоб уникнути надмірної формалізації та зберегти доступність матеріалу для гуманітарної аудиторії.

Підготовка посібника стала можливою завдяки багаторічному досвіду викладання дисциплін математико-статистичного циклу у психологічній освіті, а також завдяки аналізу практичних потреб здобувачів Було враховано труднощі, з якими найчастіше стикаються здобувачі під час опанування математичного апарату, та намагалися подати матеріал у формі, що сприяє поступовому та усвідомленому засвоєнню.

Автор сподівається, що це видання не лише забезпечить здобувачів знаннями й навичками, а й сприятиме формуванню у них наукового стилю мислення, здатності критично оцінювати результати та прагнути до об'єктивності у професійній діяльності. Адже саме поєднання гуманітарної чутливості й наукової точності формує сучасного психолога нового покоління.

Завдання майбутнього психолога полягає не лише в інтерпретації поведінкових проявів, але й у вмінні ґрунтувати свої висновки на достовірних даних. Математичні методи є тим інструментом, який допомагає перетворити психологію з описової дисципліни на точну науку, що поєднує якісний аналіз з кількісною перевіркою. Володіння такими методами відкриває для здобувачів освіти нові горизонти професійного розвитку та сприяє підвищенню якості психологічної практики в Україні.

РОЗДІЛ 1
ОСОБЛИВОСТІ ВИМІРЮВАННЯ ОЗНАК ТА КІЛЬКІСНИЙ ОПИС ДАНИХ В ПСИХОЛОГІЇ

1.1. Генеральна сукупність та вибірка дослідження

Дослідник-теоретик, розмірковуючи над закономірностями психіки та поведінки, зазвичай має на увазі не конкретний об'єкт дослідження, а певну множину об'єктів. Наприклад, вивчаючи пам'ять людини, йдеться не про окрему особу, а про людину загалом – усіх людей, які живуть на Землі, жили раніше або ще будуть жити. Очевидно, що в цьому випадку мова йде про досить велику множину об'єктів, межі якої часто не є чітко визначеними. У математичній статистиці таку множину прийнято називати *генеральною сукупністю* або *популяцією* (позначається *N*).

Зазвичай дослідник прагне конкретизувати та звузити уявлення про генеральну сукупність, роблячи її більш компактною. Наприклад, замість пам'яті людини загалом можна розглядати пам'ять певної вікової групи – дітей певного віку або людей із конкретним захворюванням, наприклад амнезією. Однак навіть у такому випадку генеральна сукупність залишається досить розпливчастою категорією, що включає велику кількість об'єктів, поведінка яких і є предметом теоретичного аналізу. Мета дослідника – виявити закономірності поведінки цієї множини об'єктів і спрогнозувати її характер.

Проводячи дослідження, дослідник-експериментатор прагне перевірити гіпотези, сформульовані дослідником-теоретиком. Водночас він не має можливості охопити всі об'єкти, що належать до генеральної сукупності, через її великий обсяг та, як правило, нечіткі межі. Наприклад, ідея дослідити пам'ять усіх людей із різними формами амнезії у світі є не лише складною, а й практично нездійсненною. Навіть за умови реалізації такого масштабного дослідження, у подальшому з'являтимуться нові особи з подібними ознаками, що вимагатиме постійного повторного тестування теоретичних припущень.

Що ж говорити про ситуацію, коли теоретик міркує про людину загалом! На щастя, для цього немає потреби охоплювати всю сукупність. Як зазвичай кажуть: щоб дізнатися смак супу, не обов'язково з'їдати весь котел – достатньо лише однієї ложки, якщо суп попередньо добре розмішати. Точно так само дослідник працює не з усією генеральною сукупністю, а лише з її частиною, що називається вибірковою сукупністю або вибіркою.

Вибірка (n) – це обмежена за чисельністю група об'єктів (у психології – досліджуваних, респондентів), що спеціально відібрана з генеральної сукупності для вивчення її властивостей. Вибірка служить основою для узагальнення результатів дослідження на всю популяцію, оскільки вона є репрезентативним підмножиною генеральної сукупності.

Вибірка повинна в максимальному ступені відображати характеристики генеральної сукупності. Це забезпечує її репрезентативність,

тобто здатність відтворювати основні властивості популяції. Репрезентативність вибірки є критично важливою для того, щоб результати дослідження можна було узагальнити на всю генеральну сукупність.

Репрезентативність вибірки – це здатність вибірки адекватно відображати досліджувані явища, зокрема їхню мінливість, у межах генеральної сукупності. Цей критерій є основним для визначення меж генералізації висновків дослідження, оскільки гарантує, що отримані результати можуть бути поширені на всю популяцію.

Наведемо два класичних методи, що забезпечують репрезентативність вибірки:

1. *Метод випадкового відбору* (random sampling). У цьому методі кожен елемент генеральної сукупності має рівні шанси бути включеним до вибірки. Це дозволяє отримати вибірку, яка є статистично репрезентативною для генеральної сукупності, оскільки випадковість відбору мінімізує упередженість.

2. *Метод стратифікованого відбору* (stratified sampling). У цьому методі генеральна сукупність поділяється на підгрупи (страти), які є однорідними за певною ознакою. Потім вибірка випадковим чином формується з кожної підгрупи, що дозволяє забезпечити репрезентативність вибірки щодо всіх важливих характеристик генеральної сукупності.

Крім того *вибірки можуть бути залежними або незалежними*.

Якщо для кожного випадку в одній вибірці X можна встановити відповідний випадок в іншій вибірці Y (тобто, коли існує гомоморфна пара, де кожному елементу з вибірки X відповідає точно один елемент з вибірки Y, і навпаки), то такі вибірки називаються залежними (або зв’язаними, парними). Приклади залежних вибірок: група досліджуваних до проведення психологічного експерименту і після нього, пари близнюків, чоловіки та їхні дружини, тощо.

Вибірки називаються незалежними (незв’язними), якщо процедура дослідження та отримані результати вимірювання певної властивості у респондентів однієї вибірки не впливають на особливості проведення цього ж дослідження та результати вимірювання цієї ж властивості у респондентів іншої вибірки. У випадку незалежних вибірок кожен респондент у одній вибірці не має жодного зв’язку з респондентами в іншій вибірці.

Відповідно, залежні вибірки завжди мають однаковий обсяг досліджуваних, оскільки кожен випадок з однієї вибірки пов’язаний з відповідним випадком з іншої. У той час як обсяг незалежних вибірок може відрізнятися, оскільки між випадками в різних вибірках відсутній будь-який зв'язок чи взаємодія.

1.2. Проблема вимірювання в психології

У своїй роботі психолог часто стикається з проблемою вимірювання індивідуально-психологічних особливостей. Для цього в психодіагностиці

розробляються спеціальні вимірювальні процедури, мета яких полягає в тому, щоб забезпечити кількісне вираження психологічних параметрів у процесі дослідження. Кількісні дані, отримані в результаті ретельно спланованих досліджень з використанням відповідних вимірювальних інструментів, у подальшому піддаються математико-статистичній обробці.

Вимірювання можна визначити як процес приписування чисел об'єктам або подіям відповідно до певних правил. Ці правила встановлюють відповідність між властивостями об'єктів, що досліджуються, і числовими значеннями. В загальному сенсі вимірювання є процедурою, під час якої об'єкт дослідження порівнюється з еталоном і отримує чисельне вираження в певній шкалі або масштабі.

У кожному конкретному випадку вимірювання є операцією, завдяки якій емпіричні дані набувають форми зв’язного числового повідомлення. Важливо розуміти, що приписування чисел об'єктам за певними принципами та правилами визначає тип шкали вимірювання. Саме закодована в числовій формі інформація дозволяє застосовувати математичні методи для виявлення прихованих закономірностей. Крім того, числове представлення об'єктів або подій дає змогу працювати зі складними поняттями в спрощеній формі. Це є основною причиною використання вимірювань у науці загалом, і в психології зокрема.

Будь-який різновид вимірювання передбачає наявність одиниць вимірювання. Одиниця вимірювання – це умовний еталон, що слугує «вимірювальною паличкою», як зазначав С. Стівенс, для здійснення вимірювальних процедур. У природничих науках і техніці використовуються стандартні одиниці вимірювання, такі як градус, метр, ампер тощо, які мають загальновизнане значення та забезпечують узгодженість вимірювань у різних контекстах.

Психологічні змінні, за винятком окремих випадків, не мають власних вимірювальних одиниць. Тому для визначення значень психологічних ознак зазвичай використовуються спеціальні вимірювальні шкали, які дозволяють трансформувати якісні психологічні характеристики в кількісні значення для подальшого аналізу.

Існує кілька типів шкал вимірювання. Операції вимірювання, або способи вимірювання об'єктів, визначають тип шкали. Шкала, в свою чергу, характеризується видом перетворень, які можна застосовувати до результатів вимірювання. Якщо ці правила не дотримуються, структура шкали буде порушена, і дані вимірювання стануть складними для осмисленої інтерпретації.

Тип шкали визначає набір статистичних методів, які можна використовувати для обробки даних вимірювання. Вибір шкали безпосередньо впливає на можливості застосування конкретних статистичних процедур, оскільки кожен тип шкали дозволяє виконувати різні види операцій з даними.

Шкала (від лат. *scala*) – це інструмент для вимірювання властивостей об'єкта, який представляє собою числову систему, в якій відносини між різними властивостями об'єктів виражаються через властивості числового ряду.

У психології для вивчення різних характеристик соціально-психологічних явищ використовуються різні шкали вимірювання. Виділяють чотири основні типи числових систем, які визначають відповідно чотири рівні вимірювання:

1) шкала номінальна (номінативна, найменувань, класифікаційна);

2) шкала порядку (рангова, ординарна);

3) шкала інтервалів (рівних інтервалів);

4) шкала відношень (пропорційна).

Низка фахівців виділяють також абсолютну шкалу і шкалу різниць.

Виділяють три основні атрибути вимірювальних шкал, наявність або відсутність яких визначає належність шкали до тієї чи іншої категорії:

1. Впорядкованість даних – означає, що один пункт шкали більший, менший або дорівнює іншому пункту. Це дозволяє встановити певний порядок між об'єктами або подіями.

2. Інтервальність пунктів шкали – передбачає, що інтервал між будь-якою парою чисел однаковий або порівняний з інтервалом між іншими парами чисел. Це означає, що різниця між значеннями має постійну величину.

3. Нульова точка – вказує на наявність абсолютного нуля, що позначає відсутність вимірюваної властивості. Цей атрибут дозволяє проводити операції множення і ділення, оскільки нуль вказує на повну відсутність властивості.

Крім того, виділяють такі групи шкал:

- неметричні шкали, у яких відсутні одиниці вимірювання та не можна визначити відстань між точками шкали. До цієї групи належать номінальна та порядкова шкали.

- метричні шкали, у яких існують чітко визначені одиниці вимірювання, що дозволяють вимірювати відстань між точками шкали. До цієї групи відносяться шкала інтервалів та шкала відношень.

1.3. Типи та характеристики вимірювальних шкал

Шкала номінальна. Шкала створюється шляхом привласнення «імен» об’єктам. У результаті аналізу об’єкти класифікуються за групами (класами), які можуть бути позначені номерами, назвами тощо. Позначення класу не вимірюється кількісно; воно лише дозволяє відрізнити один об’єкт від іншого за вимірюваною властивістю. Це і складає сутність номінальної шкали: її основне завдання – поділ на категорії.

Це найпростіша вимірювальна шкала, хоча фактично вона не асоціюється з вимірюванням і не пов'язана з поняттям «величина». Вона

використовується лише для того, щоб відрізнити один об’єкт від іншого. У номінальній шкалі відсутні всі головні атрибути вимірювальних шкал, а саме: упорядкованість, інтервальні характеристики та нульова точка. Тому вона не дозволяє проводити кількісні порівняння між об'єктами – лише класифікацію за певними категоріями.

Вимірювання полягає в тому, що після вивчення об’єкта визначають його належність до того чи іншого стану (якості) і записують це за допомогою символу (набору символів, назви, слова, числа), який позначає даний стан. Це дозволяє відобразити властивості об'єкта у числовій або категоріальній формі, що забезпечує подальшу обробку та аналіз отриманих даних.

Незважаючи на тенденцію «завищувати» потужність шкали, психологи часто застосовують номінальну шкалу в своїх дослідженнях. Багато вимірювальних процедур у діагностиці особистості зводяться до типологізації, тобто віднесення конкретної особистості до того чи іншого типу. Прикладом такої типології є класичні темпераменти: холерик, сангвінік, меланхолік і флегматик.

Найпростіша форма номінальної шкали – це дихотомічна шкала. За цією шкалою ознаки можна кодувати лише двома символами або цифрами, наприклад, 0 і 1, 2 і 6, або буквами А і В, а також будь-якими двома відмінними символами. Ознака, що вимірюється за дихотомічною шкалою, називається альтернативною.

Важливо розуміти, що позначення класів – це лише символи. Навіть якщо для цього використовуються номери, з ними не можна проводити математичні операції, як з числами. Наприклад, якщо одному спортсмену присвоєно номер 1, а іншому – номер 2, це не означає, що другий спортсмен «в два рази кращий». Можна лише зробити висновок, що це різні учасники змагань.

Порядок розташування класифікаційних назв не має значення. Той факт, що номер одного класу більший або менший за інший, не дає жодної інформації про властивості об'єктів, окрім того, що вони різні.

З даними, зафіксованими в номінальній шкалі, можна здійснювати лише операцію перевірки їх тотожності або відмінності.

Приклади номінальної шкали в психології включають: стать, національність, освіта, тип темпераменту, тип особистості тощо.

Шкала порядку. Якщо можна встановити порядок розташування психологічних об’єктів відповідно до вираженості в них якоїсь властивості, то використовується порядкова шкала.

У порядковій шкалі присутня упорядкованість, але відсутні атрибути інтервальності та нульової точки. Це означає, що хоча об'єкти можуть бути впорядковані за рівнем вираженості якоїсь ознаки, ми не знаємо точні величини цих відмінностей.

Ця шкала дозволяє здійснити лінійну впорядкованість об’єктів на деякій осі ознаки (відношення «більше» або «менше»). Наприклад, ми

можемо сказати, що один об'єкт має більшу або меншу вираженість певної властивості порівняно з іншим. Тим самим є можливість вивчити «лінійну» вираженість властивості, тоді як шкала найменувань дає лише «категоріальну» властивість (властивість є – властивості немає).

Конструювання шкали порядку – це складніша процедура, ніж створення шкали найменувань. Вона дозволяє зафіксувати ранг (місце) кожного значення змінної порівняно з іншими значеннями.

Одиницею вимірювання в порядковій шкалі є відстань між рангами або балами. Водночас відстань між рангами може бути різною і, як правило, невідома.

Для використання порядкової шкали повинно бути не менше трьох градацій (класів). Чим більше число градацій, на які розбивається експериментальна сукупність, тим більші можливості статистичної обробки отриманих даних і перевірки статистичних гіпотез.

Порядкові змінні можна кодувати будь-якими цифрами, головне – зберігати порядок, тобто кожна наступна цифра повинна бути більшою або меншою за попередню.

Особливістю порядкових шкал є те, що відношення порядку нічого не говорить про дистанцію між класами. Тому порядкові експериментальні дані, навіть якщо вони зафіксовані цифрами, не можна трактувати як числа. Наприклад, якщо об’єкти мають значення 1, 2 і 3, ми можемо сказати, що значення 3 є більшим за 2, а 2 – більшим за 1, але відстань між ними може бути різною.

Не рекомендується виконувати арифметичні операції (додавання, віднімання, множення, ділення) з числовими значеннями порядкової шкали або обчислювати середнє значення, оскільки в цій шкалі немає єдиної «еталонної» одиниці вимірювання, яка була б однаковою для всіх ділянок шкали.

Шкали порядку широко використовуються в психології. Ще сам засновник теорії вимірювання С. Стівенс висловив точку зору, що результати психологічних вимірювань здебільшого відповідають шкалам порядку.

Шкала інтервалів. Дана шкала є першою метричною шкалою. Саме з неї починається справжнє вимірювання в строгому сенсі цього слова.

У цій шкалі присутні атрибути впорядкованості та інтервальності, однак відсутня нульова точка, що вказує на відсутність властивості. Нульова точка в шкалі інтервалів є умовною і не має фізичного сенсу – це просто точка відліку, що використовується для вимірювання відстані між значеннями.

За допомогою шкали інтервалів можна порівнювати два об’єкти та визначати величину відмінностей між ними. Тобто ми можемо дізнатися, наскільки більше або менше виражена певна властивість у одного об’єкта порівняно з іншим. Проте, оскільки нульова точка є довільною, немає сенсу говорити, у скільки разів більша чи менша ця властивість у порівнянні з іншою. Якщо ігнорувати це правило, можна отримати некоректні результати.

Наприклад, не можна сказати, що температура води збільшилась у два рази при нагріванні з 9 до 18 градусів за шкалою Цельсія, оскільки для людини, яка користується шкалою Фаренгейта, зміна температури буде іншою – від 37 до 42 градусів.

Інтервальна шкала дозволяє застосовувати практично всю параметричну статистику для аналізу даних. Для характеристики центральної тенденції, крім медіани та моди, можна розраховувати середнє арифметичне, а для оцінки варіації – дисперсію та стандартне відхилення. Також можна визначати коефіцієнти асиметрії та ексцесу, а також інші параметри розподілу.

У психологічних дослідженнях до інтервальних шкал належать так звані квазіінтервальні шкали – штучно створені шкали. Це такі стандартизовані показники, як IQ, стени, Т-бали, стенайни та інші стандартизовані оцінки, що використовуються в різних психодіагностичних методиках.

Шкала відношень. У цій шкалі об’єкти класифікуються пропорційно ступеню вираженості певної властивості. Значення, отримані за цією шкалою, є «повноправними» числами, з якими можна здійснювати будь-які арифметичні операції, такі як додавання, віднімання, множення, ділення, а також використовувати різноманітні статистичні методи.

Ця шкала має всі атрибути вимірювальних шкал: впорядкованість, інтервальність і, найголовніше, нульову точку. Особливістю шкали відношень є те, що її нульова точка є абсолютною (істинною) – вона вказує на повну відсутність вираженості вимірюваної властивості. Це дозволяє стверджувати, що одна величина є в кілька разів більшою або меншою за іншу. Наприклад, можна сказати, що одна особа важить вдвічі більше за іншу, оскільки існує справжній нуль (нульова вага).

Приклади шкал відношень:

- фізичні величини: дистанція, вага, зріст, час, температура за Кельвіном (оскільки Кельвін має абсолютний нуль);
- у психології: шкали порогів абсолютної чутливості, час реакції, кількість об’єктів або суб’єктів (наприклад, кількість правильних відповідей у тесті).

Ця шкала є найбільш універсальною і дозволяє застосовувати широкий спектр математичних і статистичних методів для аналізу даних. Вона дає змогу точно вимірювати не лише величину властивості, але й порівнювати її пропорції між різними об’єктами.

1.4. Первинна описова статистика

Первинна статистична обробка даних спрямована на впорядкування інформації про об’єкти дослідження та забезпечує можливість у стислому вигляді охарактеризувати загальні властивості вибіркової сукупності: ступінь її однорідності чи неоднорідності, рівень компактності чи мінливості, відповідність або невідповідність певному закону розподілу тощо.

Первинний статистичний аналіз дозволяє дати відповіді на два ключові запитання: (1) яке значення є найбільш репрезентативним для вибірки?; (2) якою є варіативність даних відносно цього значення? Відповідь на перше запитання забезпечується шляхом обчислення мір центральної тенденції, тоді як друге – шляхом визначення мір мінливості.

Міри центральної тенденції – це числові показники, які відображають типові властивості емпіричних даних. Вони дають змогу відповісти на запитання на кшталт: «Який середній рівень інтелекту здобувачів вищої освіти університету?» або «Яке характерне значення показника відповідальності певної групи осіб?». Кількість таких показників є порівняно невеликою, а основними серед них виступають мода, медіана та середнє арифметичне. Кожна з цих мір має власні особливості, що зумовлює її цінність у характеристиці об'єкта дослідження за певних умов.

Мода (Mo) – це значення ознаки, яке найчастіше трапляється у вибірці, тобто те, що має найбільшу частоту. Для її визначення необхідно проаналізувати варіаційний ряд і встановити значення, яке повторюється найчастіше (у разі згрупованого розподілу виділяється модальний інтервал). Важливо підкреслити, що модою є саме значення ознаки, а не величина частоти.

Розподіл може мати одну або кілька мод. Якщо мода одна, такий розподіл називається унімодальним; якщо дві – бімодальним; за наявності більшої кількості мод він вважається мультимодальним. У випадку, коли всі значення трапляються з однаковою частотою, вважається, що мода відсутня.

Обираючи моду як міру центральної тенденції, необхідно враховувати, що вона адекватно відображає характерну тенденцію лише за умови, якщо модальна частота суттєво (у кілька разів) перевищує найближчі частоти.

Приклад 1.1. У дослідженні двох різних вибірок отримані наступні результати:

Таблиця 1.1 – Опис розподілів даних у двох вибірках

Вибірка № 1

Тип темпераменту	Холерик	Сангвінік	Флегматик	Меланхолік	Σ
f_i	16	17	13	14	60
p_i	0,27	0,28	0,22	0,23	1,00

Вибірка № 2

Тип темпераменту	Холерик	Сангвінік	Флегматик	Меланхолік	Σ
f_i	10	22	5	3	40
p_i	0,25	0,55	0,12	0,08	1,00

У першій вибірці відсутні підстави стверджувати про наявність вираженої тенденції у представленості певного типу темпераменту, оскільки всі вони трапляються з приблизно однаковою частотою. Водночас, якщо формально визначати моду, то модальним виступає сангвінічний тип темпераменту, адже його частота є найвищою у даній вибірці.

У другій вибірці модою також є сангвінічний тип темпераменту (оскільки він має найбільшу частоту у варіаційному ряді). Однак у цьому випадку його частота майже вдвічі перевищує найближче за величиною значення. Така диспропорція дає підстави говорити про виражену тенденцію у представленості саме сангвінічного типу темпераменту в досліджуваній групі.

Медіана (*Me* або *Md*) – це значення ознаки, яке поділяє впорядковану вибірку на дві рівні частини: 50 % спостережень мають значення ознаки, менші за медіану, і 50 % – більші за неї. Іншими словами, медіана відповідає середині впорядкованої послідовності емпіричних даних.

Слід зауважити, що медіана не обов'язково розташовується посередині варіаційного розмаху. Її положення визначається розподілом спостережень у межах впорядкованого ряду.

Також медіана не завжди збігається з конкретним емпіричним значенням. У разі непарної кількості спостережень вона дорівнює значенню, що стоїть посередині ряду. Якщо ж кількість спостережень парна, медіана визначається як середнє арифметичне двох центральних значень.

Приклад 1.2. Розрахуємо медіану для розподілу емпіричних даних: 90, 66, 106, 84, 105, 83, 104, 82, 107, 97.

Спочатку необхідно впорядкувати дані: 66, 82, 83, 84, 90, 97, 104, 105, 106, 107.

У нашому прикладі парна кількість чисел, центральними значеннями є 90 та 97.

$$Md = \frac{90+97}{2} = 93,5.$$

Для симетричних розподілів медіана співпадає з модою, для асиметричних – не співпадає.

Медіану використовують як міру мінливості для даних виміряних у порядковій шкалі.

Середнє арифметичне значення, середнє значення – це значення ознаки, яке відображає середній рівень вираженості параметра у досліджуваних.

Середнє арифметичне значення визначається за такими формулами:

А) для вибіркової оцінки:

$$\bar{x} = \frac{\sum x_i}{n}, \text{ де}$$

x_i – значення ознаки x у конкретного досліджуваного;

n – обсяг вибірки досліджуваних.

Б) для генеральної сукупності:

$$\mu = \frac{\sum x}{N}$$

x – значення ознаки у всіх членів генеральної сукупності;

N – число членів генеральної сукупності.

Для аналізу результатів дослідження бажано використовувати всі три міри центральної тенденції – моду, медіану й середнє арифметичне значення.

Особливості мір центральної тенденції:

1. У малих за обсягом вибірках мода може бути не стабільна.

Приклад 1.3. Для вибірки 1, 1, 1, 3, 5, 7, 7, 8. *Mo* = 1. Припустимо в цій вибірці один досліджуваний був замінений іншим. Значення стали такими: 1, 1, 3, 5, 7, 7, 7, 8. *Mo* = 7.

2. На медіану не впливають найбільші та найменші значення (аутлаєри).

Приклад 1.4. Для вибірки 1, 1, 3, 5, 7. *Md* = 3. Для вибірки 1, 1, 3, 5, 20. *Md* = 3.

3. На величину середнього значення впливає кожен елемент вибірки, якщо який-небудь елемент вибірки зміниться на величину c, то середнє значення зміниться на величину c/n.

Приклад 1.5. Для вибірки 1, 1, 3, 5, 7. $\bar{x}$ = 3,4. Для вибірки 1, 1, 3, 5, 16. $\bar{x}$ = 5,2. Одне значення у вибірці збільшилось на 9 одиниць. Середнє арифметичне збільшилось на 9/5 = 1,8.

4. Деякі вибірки взагалі не можна охарактеризувати за допомогою мір центральної тенденції. Особливо це актуально для вибірок, що мають декілька мод.

Приклад 1.6. Викладач за допомогою тесту досягнень визначив рівень знань здобувачів вищої освіти за спеціальною бальною шкалою.

Таблиця 1.2 – Варіаційний ряд за результатами тестування

Кількість балів	0	1	2	3	4	5	6	7	8
f_i	15	10	3	0	0	0	4	12	15

Результати можна представити у вигляді гістограми:

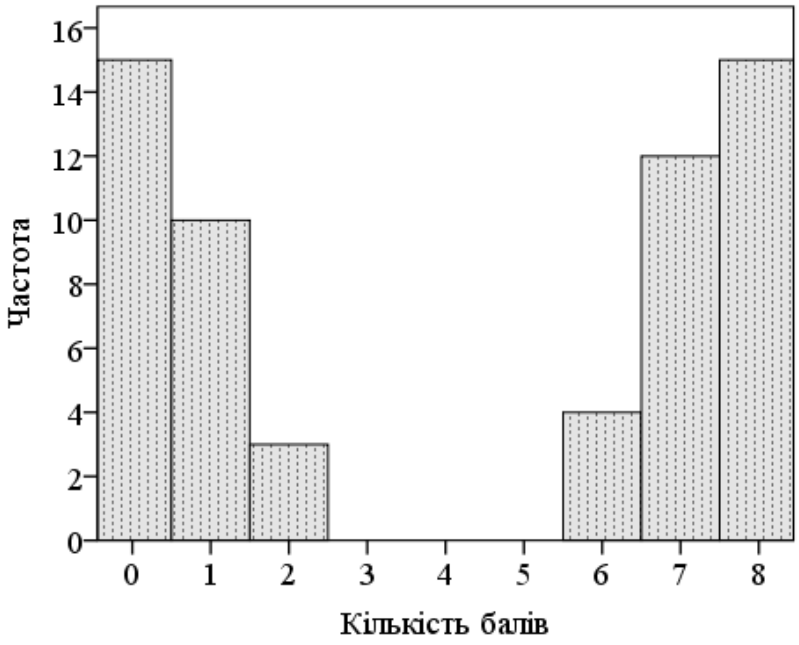


Рис. 1.1. Розподіл оцінок

У даному прикладі середнє арифметичне значення $\bar{x} = 3{,}85$, хоча жоден здобувач освіти не отримав трійки чи четвірки. Медіана цієї вибірки $Md = 4$, хоча є досить велика кількість значень 7 та 8. У даному прикладі ні медіана, ні середнє арифметичне значення не дають адекватного уявлення про досліджувану вибірку. Найбільш коректною характеристикою для даної вибірки є таке твердження: «Гістограма є бімодальною й має V-подібну форму, $Mo_1 = 0$, $Mo_2 = 8$».

5. Якщо вибірка унімодальна та симетрична (відповідає закону нормального розподілу даних), то мода, медіана і середнє значення співпадають.

6. Найпростіше з розглянутих трьох мір центральної тенденції визначається мода, її можна легко обчислити за гістограмою або полігоном частот (рис. 1.2).

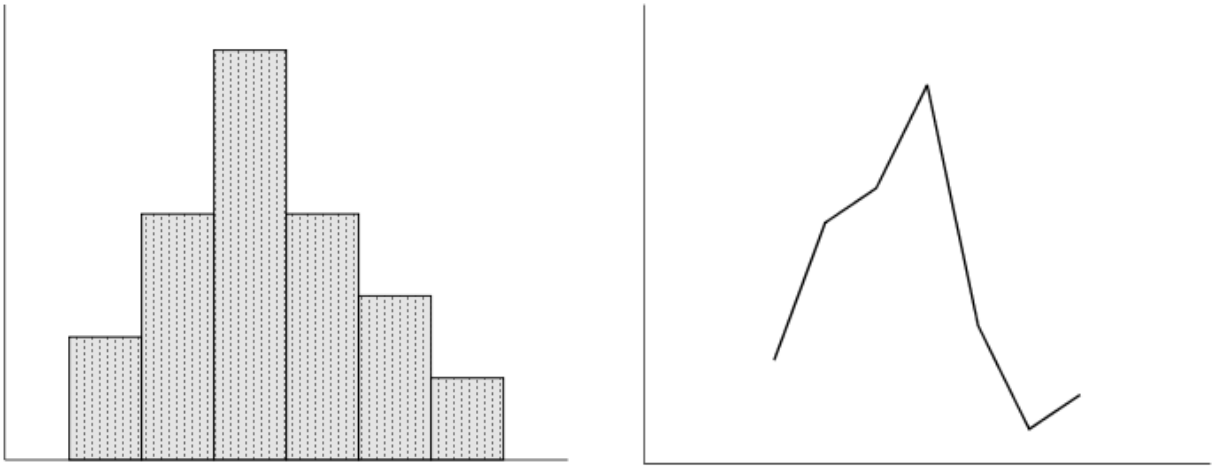


Рис. 1.2. Гістограма та полігон частот

За складністю розрахунку медіана займає проміжне положення між модою й середнім значенням.

Розглянемо приклад, як змінюються міри центральної тенденції, якщо замінити одне значення вибірки.

Приклад 1.7. Перша вибірка 1, 3, 3, 5, 6, 7, 8.

$$Mo = 3;\ Md = 5;\ \bar{x} = 4{,}71.$$

Друга вибірка 1, 3, 3, 5, 6, 7, 16

$$Mo = 3;\ Md = 5;\ \bar{x} = 5{,}86.$$

7. У вибірках великого обсягу мода й медіана є більш стійкими характеристиками, ніж середнє значення.

Загалом не можна однозначно сказати, яка із мір центральної тенденції краще характеризує вираженість ознаки у вибірці.

Середнє геометричне значення (G) відображає середній приріст значень ознаки, а *середнє гармонійне значення (H)* показує середню швидкість зміни ознаки. Вказані параметри має сенс розраховувати для динамічних ознак, які порівняно швидко змінюються в часі. Наприклад, характеристики психічних станів або ефективність роботи протягом робочого дня. Такого типу ознаки в практиці психологічних досліджень трапляються

не часто. Тому формули для розрахунку цих параметрів ми розглядати не будемо (вони достатньо складні). При необхідності слід звернутися до підручника Дж. Гласс та Дж. Стенлі.

Міри мінливості – це статистичні показники, які характеризують ступінь однорідності або різнорідності сукупності даних, а також відображають їхню компактність чи розсіювання. У психологічних дослідженнях найбільш поширеними показниками цього типу є: варіаційний розмах, дисперсія, стандартне відхилення, коефіцієнт варіації, процентилі, квартилі та довірчий інтервал.

Варіаційний розмах (R) – це інтервал між максимальним і мінімальним значеннями ознаки. Це найпростіша міра мінливості.

$R = x_{\max} - x_{\min}$, де

$x_{\max}$ – максимальне значення у вибірці;

$x_{\min}$ – мінімальне значення у вибірці.

Варіаційний розмах – це показник, що відображає інтервал варіювання ознаки у вибірковій сукупності, тобто встановлює межі, в яких розташовані всі (100 %) емпірично зафіксовані значення варіаційного ряду. Його обчислення є простим та не потребує значних затрат часу, однак цей показник вирізняється високою чутливістю до випадкових коливань, що особливо виявляється за невеликого обсягу вибірки.

Дисперсія – це статистичний показник, що відображає середній квадрат відхилень усіх значень ознаки від її середнього арифметичного. Вона обчислюється лише для даних, отриманих у метричних шкалах вимірювання. Особливістю дисперсії є те, що її розмірність відповідає квадрату одиниці вимірювання ознаки, унаслідок чого цей показник не завжди є зручним для практичного використання. Формально дисперсія визначається за такими формулами:

А) для вибіркової оцінки:

$$s^2 = \frac{\sum\left(x_i - \bar{x}\right)^2}{n-1}$$, де

x_i – значення ознаки x у конкретного досліджуваного;

$\bar{x}$ – середнє арифметичне значення;

n – обсяг вибірки досліджуваних.

Б) для генеральної сукупності:

$$\sigma^2 = \frac{\sum\left(x_i - \mu\right)^2}{N}$$, де

x_i – значення ознаки у всіх членів генеральної сукупності;

μ – середнє арифметичне генеральної сукупності;

N – число членів генеральної сукупності.

Стандартне (середньоквадратичне) відхилення – це статистичний показник, який відображає середнє відхилення кожного значення ознаки від її середнього арифметичного. На відміну від дисперсії, стандартне

відхилення має ту саму розмірність, що й сама ознака, завдяки чому воно є зручнішим для інтерпретації. Обчислюється цей показник як квадратний корінь із дисперсії.

А) для вибіркової оцінки:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum\left(x_i - \bar{x}\right)^2}{n-1}}$$, де

x_i – значення ознаки x у конкретного досліджуваного;

$\bar{x}$ – середнє арифметичне значення;

n – обсяг вибірки досліджуваних.

Б) для генеральної сукупності:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum\left(x_i - \mu\right)^2}{N}}$$, де

x_i – значення ознаки у всіх членів генеральної сукупності;

μ – середнє арифметичне генеральної сукупності;

N – число членів генеральної сукупності.

Приклад 1.8. Необхідно розрахувати вибіркову дисперсію та стандартне відхилення тривожності для заданого набору даних.

Таблиця 1.3 – Таблиця первинних даних та проміжних розрахунків

№	X	$\left(x_i - \bar{x}\right)$	$\left(x_i - \bar{x}\right)^2$
1	7	1,7	2,89
2	9	3,7	13,69
3	2	-3,3	10,89
4	4	-1,3	1,69
5	5	-0,3	0,09
6	6	0,7	0,49
7	7	1,7	2,89
8	6	0,7	0,49
9	2	-3,3	10,89
10	5	-0,3	0,09
	$\bar{x} = 5,3$	$\Sigma = 0$	$\Sigma = 44,1$

$$s^2 = \frac{\sum\left(x_i - \bar{x}\right)^2}{n-1} = \frac{44,1}{10-1} = 4,9.$$

$$s = \sqrt{\frac{\Sigma\left(x_i - \bar{x}\right)^2}{n-1}} = \sqrt{\frac{44,1}{10-1}} = 2,21.$$

Таким чином, дисперсія тривожності складає 4,9, а стандартне відхилення – 2,21.

Слід ураховувати, що стандартне відхилення зумовлене не лише ступенем розсіювання даних, а й вибором одиниць вимірювання. У зв’язку з

цим воно може бути використане для аналізу мінливості лише однорідних показників, тоді як безпосереднє порівняння абсолютних значень стандартного відхилення різних параметрів є некоректним. З метою забезпечення можливості співставлення рівня варіації ознак різної розмірності (виражених у відмінних одиницях виміру) та нейтралізації впливу масштабу середнього арифметичного на величину стандартного відхилення застосовують коефіцієнт варіації.

Коефіцієнт варіації – процентне відношення стандартного відхилення до середнього арифметичного значення ознаки. Він використовується для порівняння рівня мінливості розподілів ознак, що виражені у різних одиницях вимірювання. По суті, коефіцієнт варіації нормує стандартне відхилення, зводячи його до єдиного масштабу, що забезпечує коректність зіставлення.

$$Cv = \frac{s}{\bar{x}} \cdot 100\%$$, де

s – стандартне відхилення;

$\bar{x}$ – середнє арифметичне значення.

Коефіцієнт варіації дає змогу здійснювати порівняння мінливості ознак, поданих у різних шкалах вимірювання, а також слугує індикатором однорідності вибіркової сукупності. Для інтерпретації його величини зазвичай використовують орієнтовні критерії: варіація вважається низькою за значень до 10 %, середньою — у межах 10–20 %, та високою — понад 20 %.

Приклад 1.9. Необхідно оцінити варіативність отриманих в результаті дослідження даних: $\bar{x}$ = 12,1; s =3,3.

$$Cv = \frac{s}{\bar{x}} \cdot 100\% = \frac{3,3}{12,1} \cdot 100\% = 27,3\%.$$

Таким чином, коефіцієнт варіації більше 20 %, що свідчить про велику мінливість даних.

У психологічних дослідженнях широкого застосування набули квантилі розподілу. *Квантиль* визначається як точка на числовій шкалі вимірюваної ознаки, що поділяє впорядковану сукупність емпіричних даних на дві групи із заздалегідь відомим співвідношенням їх чисельності. Одним із найпоширеніших квантилів є медіана, яка розділяє вибірку на дві рівні частини. Крім медіани, у практиці статистичного аналізу часто застосовуються також квартилі та процентилі, поряд з іншими різновидами квантилів.

Процентилі (P_1 … P_{99}) – це система з 99 точок, які поділяють впорядковану за зростанням множину емпіричних даних на 100 рівних за чисельністю частин. Алгоритм визначення конкретного процентиля є аналогічним до процедури обчислення медіани. Наприклад, для знаходження 10-го процентиля (P_{10}) спочатку всі значення ознаки впорядковуються за зростанням, після чого визначається 10 % досліджуваних із найменшими

значеннями ознаки. Процентиль P_{10} відповідає такому значенню, яке відділяє ці 10 % спостережень від решти 90 %.

Квартилі (Q_1, Q_2, Q_3) – це три точки, які поділяють впорядковану за зростанням множину даних на чотири рівні за чисельністю частини. Перший квартиль (Q_1) відповідає 25-му процентилю (P_{25}), другий квартиль (Q_2) збігається з 50-м процентилем (P_{50}), або медіаною (Md), а третій квартиль (Q_3) відповідає 75-му процентилю (P_{75}).

Децилі (D_1, D_2, …, D_9) – це дев'ять точок, які поділяють впорядковану за зростанням множину даних на десять рівних за чисельністю частин.

Процентилі, квартилі та децилі застосовуються для оцінювання частоти появи окремих значень (або їхніх інтервалів) вимірюваної ознаки, а також для виокремлення підгруп чи індивідів, що є найбільш типовими або, навпаки, нетиповими для даної сукупності спостережень.

Альтернативним підходом до оцінювання мінливості даних є інтервальне оцінювання. У цьому разі замість одного числового значення надаються нижня та верхня межі інтервалу, тобто визначається діапазон значень на числовій осі.

На практиці для характеристики варіативності порядкових змінних зазвичай застосовується *міжквартильний інтервал* – показник мінливості, що охоплює 50 % усіх значень варіаційного ряду та відображає статистичну норму для досліджуваної вибірки. Цей інтервал визначається як діапазон між першим (Q_1) та третім (Q_3) квартилями.

Довірчий інтервал (ДІ) є формою інтервального оцінювання, за якої інтервал визначається з урахуванням ймовірності того, що істинне значення параметра потрапить у його межі з певним рівнем надійності. Для середнього значення, як правило, розраховують 95 % довірчий інтервал, який охоплює 95 % середніх значень вибірок, сформованих із гіпотетичної нескінченної генеральної сукупності. Інакше кажучи, з 95 % рівнем надійності можна стверджувати, що істинне середнє значення в популяції міститься в межах відповідного довірчого інтервалу.

Довірчий інтервал є інтервальною оцінкою середнього значення, яка доповнює його точкову оцінку. Зі збільшенням обсягу вибірки довірчий інтервал звужується, що свідчить про зростання точності оцінювання параметра.

Якщо розподіл ознаки в популяції відповідає нормальному закону, то для обчислення нижньої та верхньої меж довірчого інтервалу можна використати відповідну формулу:

$\bar{x} - t \cdot \frac{s}{\sqrt{n}}$ та $\bar{x} + t \cdot \frac{s}{\sqrt{n}}$, де

$\bar{x}$ – середнє арифметичне значення;

t – значення розподілу критерію t-Стьюдента;

s – стандартне відхилення;

n – обсяг вибірки досліджуваних.

Значення критерію *t*-Стьюдента визначається за таблицею критичних значень *t*-розподілу (див. Додаток 3) з урахуванням бажаного рівня довіри та кількості ступенів свободи, що обчислюється як $df = n - 1$, де n – обсяг вибірки досліджуваних.

Приклад 1.10. Необхідно розрахувати 95 % довірчий інтервал екстраверсії нормально розподілених даних з параметрами: $\bar{x} = 17$; $s = 5{,}1$; $n = 30$.

$df = n - 1 = 30 - 1 = 29$.

$t = 2{,}05$ (так як ми розраховуємо 95 % ДІ, то беремо значення t при $p = 0{,}05$).

$$\bar{x} - t \cdot \frac{s}{\sqrt{n}} = 17 - 2{,}05 \cdot \frac{5{,}1}{\sqrt{30}} = 15{,}1.$$

$$\bar{x} + t \cdot \frac{s}{\sqrt{n}} = 17 + 2{,}05 \cdot \frac{5{,}1}{\sqrt{30}} = 18{,}9.$$

Таким чином, із 95 % ймовірністю можна стверджувати, що істинне середнє значення рівня екстраверсії в генеральній сукупності знаходиться в межах інтервалу від 15,1 до 18,9 бала.

Для будь-якого типу розподілу, зокрема й нормального, довірчий інтервал може бути розрахований із використанням методів чисельного ресамплінгу, або, як їх ще називають, методів повторного вибіркового моделювання. До таких підходів належать чотири основні техніки, які відрізняються алгоритмом, але є концептуально спорідненими: рандомізація (переставний тест, *permutation*), метод «складного ножа» (*jackknife*), крос-перевірка (*cross-validation*) та бутстреп (*bootstrap*). Ці методи дозволяють емпірично змоделювати розподіл вибіркових статистик і виступають сучасною альтернативою класичним параметричним підходам. Найпоширенішим та найбільш популярним з-поміж них є бутстреп-метод

Основна ідея бутстреп-методу полягає в застосуванні статистичних випробувань Монте-Карло для багаторазового формування повторних вибірок з емпіричного розподілу. Зокрема, беруть сукупність з n членів вихідної вибірки $x_1, x_2, ..., x_{n-1}, x_n$, звідки на кожному кроці з n послідовних ітерацій за допомогою генератора випадкових чисел «витягується» довільний елемент x_k, який знову «повертається» у вихідну вибірку (алгоритм «випадкового вибору з поверненням» – random sampling with replacement). Внаслідок цього деякі елементи можуть повторюватися в одній псевдовибірці кілька разів, тоді як інші можуть бути відсутні. Такий підхід дозволяє сформувати довільно велику кількість бутстреп-вибірок (зазвичай 5000–10000), кожна з яких є випадковою комбінацією елементів початкової вибірки.

На основі сформованих бутстреп-вибірок конструюється розподіл вибіркових середніх значень, з якого відсікаються по 2,5 % площі з обох кінців (відповідно до 2,5 % та 97,5 % процентилів). У результаті формується 95 % довірчий інтервал для середнього значення, обчислений за допомогою

бутстреп-процедури методом процентилів. Окрім цього, існують інші методи бутстреп-оцінювання, одним із найефективніших є метод BCa (Bias-Corrected and accelerated) – прискорений бутстреп із корекцією на зміщення.

1.5. Представлення результатів дослідження за допомогою параметрів розподілу

Результати, отримані за допомогою різних методів вимірювання, допускають застосування різних математичних процедур. Обробка цих результатів визначається:

1) шкалою, в якій були виміряні психологічні ознаки;
2) цілями дослідження, тобто тим, що саме прагне відобразити дослідник.

Отже, перед початком обробки даних та вибором відповідних методів математико-статистичного аналізу необхідно чітко визначити тип шкали, у якій здійснювалися вимірювання.

Для спрощення прийняття рішення щодо вибору параметрів розподілу доцільно скористатися таблицею 1.4. Параметри розподілу – це числові характеристики, які відображають основні тенденції виразності та мінливості ознаки у досліджуваній вибірці.

Таблиця 1.4 – Вибір описових параметрів вибірки в залежності від вимірювальної шкали

Вимірювальна шкала	Міри центральної тенденції	Міри мінливості
Шкала номінальна	*Mo* – мода	f_i – абсолютна частота p_i – відносна частота $p_{\%i}$ – процентна частота
Шкала порядку	*Mo* – мода *Md* – медіана Q_1, Q_2, Q_3 – квартилі D_1…….D_9 – децилі P_1…….P_{99} – процентилі	f_i – абсолютна частота F_i – накопичена абсолютна частота p_i – відносна частота P_i – накопичена відносна частота $p_{\%i}$ – процентна частота $P_{\%i}$ – накопичена процентна частота *IQR* – міжквартильний розмах A_{SQ} – квартильний коефіцієнт асиметрії
	У разі відповідності даних закону нормального розподілу, «довгих» шкал і/або великого обсягу вибірки	
	$\bar{x}$ – середнє арифметичне значення	*s* – стандартне відхилення

Продовження таблиці 1.4.

Шкала інтервалів	*Mo* – мода *Md* – медіана Q_1, Q_2, Q_3 – квартилі D_1…….D_9 – децилі P_1……P_{99} – процентилі $\bar{x}$ – середнє арифметичне значення	Усі види частот: f_i, F_i, p_i, P_i, $p_{\%i}$, $P_{\%i}$ *IQR* – міжквартильний розмах A_{SQ} – квартильний коефіцієнт асиметрії s^2 – дисперсія *s* – стандартне відхилення *A* – коефіцієнт асиметрії *E* – коефіцієнт ексцесу *Cv* – коефіцієнт варіації
Шкала відношень	*Mo* – мода *Md* – медіана Q_1, Q_2, Q_3 – квартилі D_1…….D_9 – децилі P_1……P_{99} – процентилі $\bar{x}$ – середнє арифметичне значення *G* – середнє геометричне значення *H* – середнє гармонійне значення	Усі види частот: f_i, F_i, p_i, P_i, $p_{\%i}$, $P_{\%i}$ *IQR* – міжквартильний розмах A_{SQ} – квартильний коефіцієнт асиметрії s^2 – дисперсія *s* – стандартне відхилення *A* – коефіцієнт асиметрії *E* – коефіцієнт ексцесу *Cv* – коефіцієнт варіації

Завдання на самостійну підготовку

1. Історія розвитку статистичного аналізу даних.
2. Розкрийте відмінності між вибіркою та генеральною сукупністю.
3. Дайте порівняльну характеристику основним стратегіям формування репрезентативних вибірок дослідження.
4. Розкрийте особливості вимірювання психологічних явищ.
5. Поясніть основні труднощі кількісного вимірювання у психології.
6. Основні властивості метричних і неметричних шкал вимірювання.
7. Розкрийте характеристики різних шкал вимірювання.
8. Обґрунтуйте значення правильного вибору шкали вимірювання для аналізу даних.
9. Наведіть приклади використання різних типів шкал у психологічних дослідженнях.
10. Поясніть методи первинного опису даних у психології.
11. Дайте характеристику середньому арифметичному, моді та медіані.
12. Охарактеризуйте основні показники варіативності даних.
13. Опишіть способи графічного представлення результатів дослідження.
14. Стандартизація даних і стандартизовані шкали в психології.

РОЗДІЛ 2
СТАТИСТИЧНІ ГІПОТЕЗИ ТА ОСОБЛИВОСТІ ЇХ ПЕРЕВІРКИ

2.1. Статистичні гіпотези та критерії

Отримані у дослідженнях вибіркові дані завжди є обмеженими та значною мірою випадковими. Саме з цієї причини для їх аналізу застосовується математична статистика, яка дозволяє узагальнювати закономірності, виявлені у вибірці, та поширювати їх на генеральну сукупність. Важливо підкреслити, що дані, отримані на будь-якій вибірці, слугують лише підставою для формування суджень про властивості генеральної сукупності. Проте внаслідок дії випадкових чинників оцінки параметрів генеральної сукупності, отримані на основі вибіркових даних, завжди супроводжуються певною похибкою. Відповідно, подібні оцінки слід розглядати не як остаточні твердження, а як наближені.

Судження про властивості та параметри генеральної сукупності на основі вибіркових оцінок отримали назву *статистичних гіпотез*. Суть перевірки статистичної гіпотези полягає у визначенні того, чи узгоджуються емпіричні дані з висунутою гіпотезою, та чи можна вважати, що виявлені розбіжності між нею і результатами статистичного аналізу зумовлені випадковими факторами. Формулювання статистичних гіпотез має формалізований характер: вони завжди розглядаються парами, де одна виступає як основна (нульова), а інша – як альтернативна.

Нульова гіпотеза (H_0) – це статистичне припущення, яке формулюється як твердження про відсутність відмінностей або про рівність значень показників, що порівнюються.

Альтернативна гіпотеза (H_1) – це протилежне твердження, яке заперечує нульову гіпотезу та передбачає наявність відмінностей або нерівність порівнюваних показників.

Нульова та альтернативна гіпотези утворюють повну групу несумісних подій: істинність однієї з них автоматично означає хибність іншої, і навпаки. Відхилення нульової гіпотези, таким чином, еквівалентне прийняттю альтернативної, а прийняття нульової – відкиданню альтернативної.

Таблиця 2.1 – Характеристика статистичних гіпотез

	Статистичні гіпотези	
	нульова гіпотеза	альтернативна гіпотеза
Позначається	H_0	H_1
Призначення	гіпотеза про відсутність відмінностей	гіпотеза про існування значущих відмінностей
Передбачуваний результат	це те, що ми хочемо спростувати	це те, що ми хочемо довести
Математична модель	$\bar{x}_1 - \bar{x}_2 = 0$	$\bar{x}_1 - \bar{x}_2 \neq 0$

У більшості випадків альтернативна гіпотеза може розглядатися як експериментальна, оскільки саме вона формулює припущення про наявність відмінностей між вибірками або про існування взаємозв'язку між змінними. Водночас, якщо завдання дослідника полягає у підтвердженні відсутності відмінностей, то експериментальною виступає нульова гіпотеза.

Нульова та альтернативна гіпотези можуть мати різну спрямованість: бути направленими (однобічними) або ненаправленими (двобічними). У направлених гіпотезах уточнюється очікуваний напрям відмінностей чи зв'язків, тоді як у ненаправлених вказується лише факт їх наявності без конкретизації напрямку.

Таблиця 2.2 – Види статистичних гіпотез

	Статистичні гіпотези	
	нульова гіпотеза	альтернативна гіпотеза
Направлені гіпотези	$\bar{x}_1 \le \bar{x}_2$ ($\bar{x}_1$ не перевищує $\bar{x}_2$)	$\bar{x}_1 > \bar{x}_2$ ($\bar{x}_1$ перевищує $\bar{x}_2$)
Ненаправлені гіпотези	$\bar{x}_1 = \bar{x}_2$ ($\bar{x}_1$ не відрізняється $\bar{x}_2$)	$\bar{x}_1 \ne \bar{x}_2$ ($\bar{x}_1$ відрізняється $\bar{x}_2$)

Направлені гіпотези формулюються у тих випадках, коли дослідник прагне довести, що:

1) в одній із груп індивідуальні значення досліджуваної ознаки є вищими, а в іншій – нижчими;

2) під впливом експериментальних умов в одній із груп відбулися більш виражені зміни, ніж у групі порівняння.

Ненаправлені гіпотези застосовуються тоді, коли завдання полягає у підтвердженні факту відмінностей між групами без уточнення їхнього напрямку, наприклад, якщо потрібно довести, що форми розподілу певної ознаки у двох групах досліджуваних відрізняються.

Статистична перевірка гіпотез ґрунтується на вибіркових даних і завжди пов'язана з ризиком (ймовірністю) прийняття хибного рішення стосовно генеральної сукупності. У цьому контексті розрізняють два типи помилок – помилки першого та другого роду.

Таблиця 2.3 – Статистичні рішення при оцінці гіпотез

Рішення дослідника	В дійсності	
	Істинна гіпотеза H_0	Істинна гіпотеза H_1
Гіпотеза H_0 відхиляється	Хибне рішення. Помилка 1-го роду. Ймовірність = α	Вірне рішення
Гіпотеза H_0 приймається	Вірне рішення	Хибне рішення. Помилка 2-го роду. Ймовірність = β

Максимально допустима ймовірність помилки першого роду позначається символом α. Чим нижчим є її рівень, тим більший обсяг вибірки необхідно залучати до дослідження. Величину α експериментатор визначає до початку збору даних як статистичну межу готовності припуститися помилки під час прийняття рішення. Інакше кажучи, це ймовірність отримання хибнопозитивного результату – відхилення нульової гіпотези тоді, коли у генеральній сукупності вона є істинною.

У психологічних дослідженнях найчастіше рівень α встановлюють на позначці 0,05. Проте така практика здебільшого зумовлена традицією, адже кожен дослідник має право обирати рівень значущості, виходячи зі специфіки предмета дослідження. Порогове значення 0,05 означає, що отриманий результат може виявитися випадковим приблизно в одному випадку з двадцяти. Це не є незначною ймовірністю, і тому її вплив на результати дослідження часто недооцінюється, особливо у випадках, коли аналіз охоплює велику кількість діагностичних параметрів.

Слід зауважити, що зменшення ймовірності помилки першого роду не завжди є безумовно корисним, оскільки водночас зростає ризик припущення помилки другого роду (β). Помилка другого роду означає отримання хибнонегативного результату, тобто ситуацію, коли приймається рішення не відхиляти нульову гіпотезу, хоча насправді вона є хибною.

У практиці психологічних досліджень величина цієї помилки зазвичай прямо не вказується, однак про неї необхідно пам’ятати під час інтерпретації результатів, особливо при роботі з вибірками малого обсягу. З методологічної точки зору рекомендується, щоб імовірність помилки другого роду не перевищувала 0,2.

Помилка другого роду (β) перебуває у зворотному співвідношенні зі статистичною потужністю дослідження, яка визначається як 1 - β. Статистична потужність відображає здатність виявити статистично значущі відмінності у разі їхньої реальної наявності. Інакше кажучи, це ймовірність відхилення нульової гіпотези тоді, коли вона є хибною.

У більшості емпіричних досліджень рівень статистичної потужності приймається в межах 0,8 - 0,9 (80 - 90 %). Такий діапазон вважається оптимальним компромісом між точністю висновків та ресурсними обмеженнями. У випадках, коли йдеться про вивчення рідкісних явищ і сформувати достатній обсяг вибірки є вкрай складно, допустимим вважається зниження потужності до 0,7 (70 %), що дає змогу все ж таки здійснити дослідження.

Важливо враховувати, що зі зростанням статистичної потужності зростає й необхідний обсяг вибірки. Причому це зростання має експоненційний характер: кожне додаткове підвищення потужності на 1 % вимагає дедалі більшої кількості досліджуваних.

Перед початком емпіричного дослідження критично важливо визначити його статистичну потужність. Розпочинати роботу доцільно лише

тоді, коли ймовірність виявлення статистично значущого ефекту є достатньо високою. Інакше витрати ресурсів втрачають сенс, оскільки дослідження опиняється приреченим на невдачу.

Якщо, наприклад, при статистичній потужності 40 % не було виявлено значущих відмінностей, отриманий результат залишається невизначеним. Це може бути зумовлено як реальною відсутністю ефекту, так і недостатнім обсягом вибірки для його виявлення. У наведеній ситуації ймовірність пропустити існуючий ефект становить 60 %. Таким чином, сам факт невиявлення відмінностей не дає підстав стверджувати, що вони справді відсутні.

Отже, перш ніж робити висновок про відсутність відмінностей, дослідник має пересвідчитися, що статистична потужність дослідження була достатньою для їхнього виявлення.

Під час інтерпретації результатів слід уважно збалансовувати ризики помилок першого (α) та другого (β) роду, визначаючи прийнятний рівень довіри та можливість практичного застосування висновків. Для більшості досліджень типовим вважається поєднання $\alpha = 0{,}05$ та статистичної потужності $1 - \beta \approx 0{,}80$. Втім, оптимальні значення цих параметрів визначаються індивідуально з огляду на предмет дослідження, «вартість» потенційних помилок і ресурси для формування необхідного обсягу вибірки; остаточне рішення ухвалює дослідник, який найкраще розуміє важливість очікуваних висновків і наявні обмеження.

Усі зазначені характеристики мають бути враховані дослідником на етапі планування, зокрема під час розрахунку необхідного обсягу вибірки. У психологічних дослідженнях визначення достатнього обсягу вибірки часто становить складність, і належне обґрунтування цього параметра подається далеко не завжди. У більшості випадків дослідники спираються на умовні рекомендації або на власні можливості щодо збору первинних даних. Подібний підхід може істотно знизити наукову цінність результатів, поставити під сумнів коректність статистичної інтерпретації та, зрештою, нівелювати значущість дослідження.

Насправді, ще на етапі планування дослідник має поставити перед собою принципові запитання: «Я маю змогу обстежити X осіб. Чи забезпечить така вибірка достатню потужність для отримання достовірних результатів?» або «Якого обсягу вибірку необхідно сформувати, щоб виявити очікуваний ефект?». Відповіді на ці запитання надає аналіз статистичної потужності – сукупність методів, що дозволяють оптимізувати дослідницький дизайн і раціонально використати наявні ресурси.

Не всі усвідомлюють, що визначення розміру вибірки – це не просто застосування однієї формули, а складний процес, який вимагає глибокого аналізу цільової популяції, на яку будуть поширюватися результати дослідження. Цей процес передбачає чітке розуміння дослідницьких гіпотез і завдань, специфіки дизайну дослідження, типів вимірювальних шкал, у яких

представлені дані, направленості гіпотез, взаємозв'язків між вибірками, чутливості статистичних критеріїв, визначення мінімального ефекту, що має практичне значення, а також прогнозування ймовірного вибування учасників у довготривалих проспективних дослідженнях тощо.

Недостатньо просто обрати, наприклад, 20 найбільш доступних осіб, провести з ними дослідження та проаналізувати отримані дані. Проблема полягає не лише в можливій недостатній кількості об'єктів для проведення якісного статистичного аналізу. Основне питання – у відсутності репрезентативності вибірки: за такого підходу результати відображають характеристики виключно цієї конкретної групи з 20 осіб. Отже, отримані висновки є випадковими і не можуть бути коректно екстрапольовані на генеральну сукупність.

Формування малої вибірки в дослідженні, збільшує ймовірність прийняття помилкової гіпотези, результат часто буває недооціненим. А це означає, що люди були марно піддані опитуванню та/або експериментальному впливові. Крім того, фінансові та часові ресурси були витрачені даремно, оскільки в кінцевому рахунку дослідження не покращило практику надання психологічної допомоги.

З іншого боку, не завжди коректна широко поширена думка, що великі вибірки ідеально підходять для досліджень та статистичного аналізу. Формування значної вибірки досліджуваних пов'язана з різними труднощами. Не етично залучати до експерименту більше осіб, ніж це необхідно, тому що вони піддаються впливові, результат якого невідомий. Використання великої кількості досліджуваних також вимагає додаткових фінансових, організаційних та людських затрат, ніж це вимагають поставлені завдання дослідження. Крім того, підвищується ймовірність виявлення статистично значущих результатів, які мають низьку практичну цінністю.

Вирішенням цієї дилеми є компромісний і водночас науково обґрунтований підхід – ретельне планування психологічного дослідження з попереднім розрахунком необхідного обсягу вибірки. Це є обов'язковою умовою якісного дослідження в межах методології доказової психології. Такий підхід гарантує, що у разі наявності практично значущого ефекту, він буде з високою ймовірністю виявлений засобами статистичного аналізу. Планування дослідження на основі обґрунтованого розміру вибірки дозволяє ухвалювати дослідницькі рішення більш раціонально й аргументовано, а також сприяє критичному осмисленню результатів, отриманих з наукової літератури.

Загалом висновки, які можна зробити на основі емпіричних даних, є асиметричними: гіпотезу можна відхилити, але ніколи не можна остаточно підтвердити. Будь-яка наукова гіпотеза завжди залишається відкритою для подальшої перевірки та уточнення. Водночас відхилення гіпотези не є автоматичним спростуванням теорії, з якої вона випливає. На жаль, жодна

експериментальна процедура не здатна забезпечити абсолютно достовірне психологічне знання – результати завжди залишаються ймовірнісними.

Перевірка статистичних гіпотез здійснюється на основі статистичних критеріїв – інструментів для оцінки рівня статистичної значущості. За їх допомогою приймають істинні гіпотези та відхиляють хибні. Кожен критерій має свої переваги, недоліки, можливості й обмеження застосування.

Статистичні критерії бувають: *параметричні та непараметричні*.

Загалом параметричні критерії вважаються більш потужними порівняно з непараметричними за умови, що ознака виміряна за кількісною (метричною) шкалою та має нормальний розподіл. Якщо ж розподіл ознаки суттєво відрізняється від нормального, застосовують непараметричні критерії.

Статистичний критерій складається з:

1) формули розрахунку емпіричного значення критерію за вибірковими статистиками;

2) правила (формули) визначення числа ступенів свободи *df*;

3) теоретичного розподілу для даного числа ступенів свободи;

4) правила співвідношення емпіричного значення критерію з критичним значенням теоретичного розподілу для визначеного α-рівня, яке дозволяє визначити статистичну значущість.

Критичні значення критерію визначаються за відповідними таблицями критичних значень залежно від числа ступенів свободи. Позначається буквою ν або *df*. Ми будемо використовувати – *df* (degrees of freedom).

Число ступенів свободи – кількість можливих напрямків мінливості ознаки. Дорівнює числу класів варіаційного ряду мінус число умов, за яких він був сформований. Поняття «число ступенів свободи» можна пояснити на прикладі. Нехай у нас є рівняння X1 + X2 + X3 = 10. Дана сума може бути одержана при різних значеннях змінних, наприклад: 2 + 5 + 3 = 10; 3 + 3 + 4 = 10; 1 + 0 + 9 = 10; 6 + 7 + (-3) = 10 тощо. Два доданки можуть бути будь-якими числами, а останній має доповнювати суму перших двох до десяти, тобто він не є вільним, а «зв'язаним». Отже, у даному прикладі можливе число змін дорівнює двом, а в загальному випадку для *n* доданків це число дорівнює *n* – 1.

Для кожного статистичного критерію визначення числа ступенів свободи має свою специфіку. Тому в кожному алгоритмі розрахунку статистичного критерію наводиться відповідна формула для їх обчислення. Зазвичай число ступенів свободи лінійно залежить від обсягу вибірки або від кількості ознак (k) чи їхніх градацій – чим більші ці показники, тим більше число ступенів свободи.

Статистична значущість (p-value) – це розрахована ймовірність отримання даного або більш екстремального результату вибірки певного обсягу за умови, що нульова гіпотеза є вірною для генеральної сукупності.

Якщо *p*-значення менше за заданий рівень значущості α, нульова гіпотеза вважається відхиленою як малоймовірна щодо отриманих результатів.

Найпоширенішими помилками в інтерпретації значення статистичної значущості є розуміння її як:

1) показника істинності нуль-гіпотези;
2) міри ймовірності випадкового отримання результатів;
3) об'єктивного засобу перевірки нуль-гіпотези;
4) міри доказу істинності альтернативної гіпотези;
5) міри відтворення результатів;
6) показника практичної важливості отриманих результатів.

Історично склалося, що при дослідженні психологічних явищ виокремлюють три рівні статистичної значущості критерію, які визначаються співвідношенням його емпіричного та критичного значень:

- нижчим рівнем статистичної значущості вважають 5 %-й рівень ($p \leq 0{,}05$);
- достатнім рівнем статистичної значущості вважають 1 %-й рівень ($p \leq 0{,}01$);
- вищим – 0,1%-й рівень ($p \leq 0{,}001$).

Тому в таблицях критичних значень вказують значення критеріїв, що відповідають рівням статистичної значущості $p \leq 0{,}05$, $p \leq 0{,}01$ та $p \leq 0{,}001$. Чим менше *p*-значення, тим більше підстав для того, щоби відхилити H_0, на користь H_1, і підтвердити експериментальну гіпотезу.

Припустимо, що при порівнянні двох середніх значень було отримано $p = 0{,}05$. Це означає, що ймовірність виявлених таких або більш виражених відмінностей становить не більше 5 % в ситуації вірності нульової гіпотези.

Слід пам'ятати, що зі збільшенням кількості перевірених гіпотез зростає ймовірність отримання випадкових результатів. Ця проблема відома як множинне порівняння, при якій *p*-значення зростає пропорційно числу перевірених гіпотез.

2.2. Розмір ефекту

Проблему вимірювання розміру ефекту актуалізував Джейкоб Коен у рамках розробленої ним парадигми метааналізу. Він чітко доніс сенс та важливість поняття «розмір ефекту» до фахівців соціальних наук і розробив найбільш поширені індекси для його оцінки. Розмір ефекту позначається абревіатурою *ES* (effect size).

Розмір ефекту – це кількісне відображення величини (ступеня прояву) явища, що становить науковий інтерес. Він характеризує силу взаємозв'язку між явищами, величину різниці ознаки між групами, відповідність теоретичної моделі емпіричним даним, а також дає змогу оцінити корисність, важливість і практичну цінність отриманих результатів. За своєю суттю

розмір ефекту близький до поняття практичної значущості дослідження – ключового орієнтира наукової роботи.

Розміри ефекту дозволяють дослідникам перейти від дихотомічного мислення – «чи існує відмінність (взаємозв'язок) чи ні?» – до більш інформативного питання: «Яка величина цих відмінностей (сила взаємозв'язку)?» Це сприяє відходу від надмірної уваги до статистичної значущості на користь більш цінного та зрозумілого кількісного опису розміру ефекту, що є більш науковим підходом до накопичення знань. Крім того, класичні показники розміру ефекту не залежать від обсягу вибірки, на відміну від статистичної значущості, яку саме за це часто критикують. Статистична значущість у експериментах із різним числом досліджуваних, але однаковими базовими описовими характеристиками (наприклад, середніми арифметичними, стандартними відхиленнями, довірчими інтервалами) може суттєво відрізнятися, тоді як оцінка розміру ефекту залишатиметься стабільною.

В якості показників розміру ефекту можуть використовуватися як нестандартизовані величини (різниця між двома середніми арифметичними, відсотками, нестандартизовані регресійні коефіцієнти тощо), так і стандартизовані індекси, що дозволяють перевести ефект у зрозумілий і порівнянний масштаб.

Нестандартизовані величини розміру ефекту є корисними, коли досліджувані змінні мають безпосереднє і зрозуміле значення (наприклад, швидкість прийняття рішення). Натомість стандартизовані індекси розміру ефекту необхідно застосовувати у випадках, коли вимірювання не має внутрішнього значення (наприклад, числові значення у шкалі Лікерта); коли в дослідженнях використовувалися різні шкали, що ускладнює пряме порівняння результатів; або коли розмір ефекту розглядається в контексті варіабельності популяції.

Вибір конкретного індексу розміру ефекту у дослідженні залежить від шкали вимірювання, застосованих статистичних критеріїв та дизайну дослідження (див. табл. 2.4).

Таблиця 2.4 – Індекси кількісної оцінки розміру ефекту

Індекс ES	Статистичний критерій	Формула розрахунку індексу ES
d-Коена	*t*-Стьюдента для незалежних вибірок (за умови рівності обсягів вибірок та гомогенності дисперсій).	$d = \frac{\left\lvert \overline{X}_1 - \overline{X}_2 \right\rvert}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$
g-Хеджесса	*t*-Стьюдента для незалежних вибірок (за умови різної кількості досліджуваних у групах або/та малі обсяги вибірок).	$g = \frac{\left\lvert \overline{X}_1 - \overline{X}_2 \right\rvert}{\sqrt{\frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}}}$

Продовження таблиці 2.4

Δs-Гласса	t-Стьюдента для незалежних вибірок (за умови гетерогенності дисперсій у вибірках).	$\Delta s = \frac{\left\|\overline{X}_1 - \overline{X}_2\right\|}{\sqrt{s^2_{контрол.}}}$
d_z-Коена	t-Стьюдента для залежних вибірок	$d_z = \frac{\left\|\overline{X}_d\right\|}{s_d}$
d_{os}-Коена	t-Стьюдента для однієї вибірки	$d_{os} = \frac{\left\|\overline{X} - A\right\|}{s}$
η^2	Дисперсійний аналіз	$\eta^2 = R^2 = \frac{SS_{BG}}{SS_T}$
r	U-Манна-Уітні, T-Вілкоксона	$r = \frac{z}{\sqrt{n}}$
ε^2	H-Крускала-Уолліса	$\varepsilon^2 = H \cdot \frac{n+1}{n^2-1}$
W-Кендалла	χ^2_r-Фрідмана	$W = \frac{\chi^2_r}{n(k-1)}$
w-Коена	χ^2-Пірсона для однієї вибірки	$w = \sqrt{\sum_{i=1}^{k} \frac{(p_{емп.} - p_{теор.})^2}{p_{теор.}}}$
φ	χ^2-Пірсона (таблиця крос-табуляції ознак має розмір 2×2)	$\varphi = \sqrt{\frac{\chi^2}{n}}$
V-Крамера	χ^2-Пірсона (таблиця крос-табуляції ознак має відмінний від 2×2 розмір)	$V = \sqrt{\frac{\chi^2}{n \cdot (c-1)}}$
g-Коена	Біноміальний критерій m	$g = \left\|P - 0,5\right\|$
r^2	Коефіцієнт лінійної кореляції r_{xy} Пірсона	$r^2 = r_{xy}^{\ 2}$
r	Коефіцієнт контингенції К. Пірсона φ; Коефіцієнт асоціації Д. Юла Q; Коефіцієнти взаємної зв'язаності ознак Пірсона C, Чупрова K та Крамера V; Точково-бісеріальний коефіцієнт кореляції r_{pb}; Рангово-бісеріальний коефіцієнт кореляції r_{rb}; Коефіцієнти рангової кореляції r_s Спірмена, τ Кендалла; Коефіцієнт лінійної кореляції r_{xy} Пірсона.	Їхні формули розрахунку коефіцієнтів кореляції

Необхідно зазначити, що наведені стандартизовані індекси розміру ефекту (ES) не є єдиними можливими для відповідних статистичних критеріїв. Звернення до спеціалізованої літератури дозволяє виявити альтернативні міри оцінки розміру ефекту. Більшість із них є взаємопов'язаними і, за потреби, можуть трансформуватися один в інший. Така універсальність дає змогу порівнювати результати різних досліджень у

межах метааналізу, що сприяє отриманню більш надійних оцінок параметрів генеральної сукупності.

Оскільки стандартизовані величини ефекту не мають одиниць виміру, їх інтерпретація для кожного індексу зводиться до визначення, який ефект слід вважати малим, середнім або великим. Традиційно для розв'язання цього питання застосовуються два підходи. Перший полягає у порівнянні виявленого ефекту з типовими ефектами, отриманими іншими дослідниками у відповідній галузі. Другий підхід – орієнтація на «еталонні» значення, серед яких найбільш визнаними у поведінкових науках є методичні рекомендації Дж. Коена (див. табл. 2.5).

Таблиця 2.5 – Узагальнена інтерпретація розмірів ефекту

Індекси ES	Інтерпретація величини розміру ефекту
d-Коена, *g*-Хеджесса, Δ*s*-Гласса, d_z-Коена, d_{os}-Коена	0,00 ≤ ES < 0,20 – несуттєвий; 0,20 ≤ ES < 0,50 – малий; 0,50 ≤ ES < 0,80 – середній; 0,80 ≤ ES – великий.
η^2	0,00 ≤ r^2 < 0,01 – несуттєвий; 0,01 ≤ r^2 < 0,06 – малий; 0,06 ≤ r^2 < 0,14– середній; 0,14 ≤ r^2 < 1,00 – великий.
r, *φ*, *W*-Кендалла, *w*-Коена,	0,00 ≤ | ES | < 0,10 – несуттєвий; 0,10 ≤ | ES | < 0,30 – малий; 0,30 ≤ | ES | < 0,50– середній; 0,50 ≤ | ES | < 1,00 – великий.
ε^2	0,00 ≤ ε^2 < 0,01 – несуттєвий; 0,01 ≤ ε^2 < 0,08 – малий; 0,08 ≤ ε^2 < 0,26 – середній; 0,26 ≤ ε^2 – великий.
g-Коена	0,00 ≤ *g* < 0,05 – несуттєвий; 0,05 ≤ *g* < 0,15 – малий; 0,15 ≤ *g* < 0,25 – середній; 0,25 ≤ *g* – великий.
r^2	0,00 ≤ r^2 < 0,01 – несуттєвий; 0,01 ≤ r^2 < 0,09 – малий; 0,09 ≤ r^2 < 0,25– середній; 0,25 ≤ r^2 < 1,00 – великий.

Тим не менше, жоден із традиційних підходів до інтерпретації розміру ефекту в психології не позбавлений суттєвих проблем. Порівняння виявленого ефекту з результатами інших дослідників ускладнюється низькою відтворюваністю результатів досліджень. Так, у проекті, в якому брали

участь 270 дослідників, було здійснено повторне відтворення 100 досліджень, опублікованих у провідних психологічних журналах. Результати показали, що понад 60 % реплікацій не підтвердили статистичної значущості висновків на рівні 0,05. Середній розмір ефекту у реплікаціях виявився удвічі меншим, ніж у початкових дослідженнях, що свідчить про суттєве зниження.

З іншого боку, використання еталонних стандартів для інтерпретації величини розміру ефекту в психології втрачає сенс через значну різнорідність наукових напрямів у цій галузі. Т. Шефер та М. Шварц виявили суттєві відмінності у медіанах розподілів розмірів ефекту в різних підгалузях психології. Найвищі значення розміру ефекту спостерігаються в дослідженнях експериментальної та біологічної психології, тоді як у соціальній психології та психології розвитку вони значно менші.

Одне з прагматичних рішень проблеми інтерпретації розміру ефекту запропонував сам Дж. Коен: подавати ефект у вигляді нестандартизованих величин і тлумачити його практичне значення у контексті конкретного дослідження.

Таким чином, запропоновані узагальнені градації розміру ефекту слід розглядати лише як орієнтири, адже для кожного окремого дослідження має застосовуватися індивідуальний підхід. Під час інтерпретації розміру ефекту необхідно враховувати теоретичні, економічні, етичні та практичні аспекти.

Існує кілька типових ситуацій інтерпретації результатів дослідження з урахуванням розміру ефекту після перевірки статистичної гіпотези:

1) Якщо немає достатніх підстав для відхилення нульової гіпотези (тобто відмінності не виявлено), але розмір ефекту (*ES*) є середнім або великим, це не дає підстав стверджувати про відсутність відмінностей. Така ситуація свідчить про можливу нестачу статистичної потужності дослідження й може бути аргументом на користь повторного дослідження з більшим обсягом вибірки;

2) Якщо нульову гіпотезу не відхилено, а розмір ефекту є несуттєвим або малим, можна зробити обґрунтоване припущення, що відмінностей, ймовірно, дійсно немає;

3) Якщо є підстави для відхилення нульової гіпотези (виявлено статистично значущі відмінності), але розмір ефекту несуттєвий або малий, варто переосмислити практичне значення таких відмінностей – можливо, вони не виходять за межі похибки вимірювання й не мають реальної психологічної важливості;

4) Якщо нульову гіпотезу відхилено, а розмір ефекту великий, це є підставою для впевненого ствердження, що відмінності дійсно існують і мають психологічне значення.

Оцінка розміру ефекту до тестування гіпотези є особливо актуальною в експериментальних дослідженнях. Вона дозволяє розрахувати необхідний обсяг вибірки ще на етапі планування, що має принципове значення в умовах обмеженого доступу до досліджуваних. У таких випадках необхідно також визначити допустимий рівень помилки першого роду (α) та бажану

статистичну потужність дослідження (1−β). Для виконання таких розрахунків можна скористатися безкоштовним програмним забезпеченням, зокрема пакетом *GPower* (https://surl.li/aqfgxp) або програмно-статистичним середовищем із відкритим кодом *R* (https://www.r-project.org/).

Окрім інтересу до результатів окремого дослідження, увага до розміру ефекту має ще один важливий аспект − забезпечення уніфікації звітів про результати емпіричних робіт з метою їх подальшого використання в метааналітичних дослідженнях. Для цього наукова спільнота має орієнтуватися на стандарти публікації емпіричних даних, розроблені Американською психологічною асоціацією (APA).

Отже, існує низка вагомих причин, з яких дослідникам доцільно доповнювати свої звіти про перевірку нульової гіпотези інформацією про розмір ефекту та його довірчі інтервали. Серед основних аргументів на користь такої практики можна виокремити наступні:

1. На відміну від p-значення, розмір ефекту відображає наукову та практичну цінність отриманих результатів.
2. Дає змогу оцінити ступінь прояву явища незалежно від обсягу вибірки.
3. Стандартизоване представлення результатів дозволяє порівнювати між собою показники, виміряні в різних шкалах.
4. Використовується для розрахунку необхідного обсягу вибірки під час планування майбутніх досліджень.
5. Дозволяє оцінити статистичну потужність уже проведених досліджень.
6. Є ключовим компонентом метааналітичних досліджень, забезпечуючи можливість порівняння результатів, отриманих різними авторами.
7. У поєднанні з довірчим інтервалом, розмір ефекту може бути використаний для оцінювання параметрів генеральної сукупності.

2.3. Параметри нормального розподілу даних

При прийнятті рішення щодо вибору статистичного критерію для аналізу результатів дослідження важливо враховувати характер розподілу даних. Якщо ознаки мають нормальний або близький до нормального розподіл, доцільно застосовувати методи параметричної статистики. У разі значного відхилення від нормальності слід віддавати перевагу менш чутливим до розподілу даних методам − непараметричним критеріям. У психології часто перевіряють відповідність даних закону нормального розподілу для визначення типу вимірювальної шкали, у якій представлені ознаки − інтервальної або порядкової.

Параметри розподілу – це числові характеристики, які описують центральне положення значень ознаки та ступінь їхньої мінливості. Найбільш практично важливими для оцінки нормальності розподілу даних є

середнє арифметичне ($\bar{x}$), середньоквадратичне відхилення (s), асиметрія (A) та ексцес (E).

У реальних психологічних дослідженнях ми працюємо не з параметрами генеральної сукупності безпосередньо, а з їхніми наближеними значеннями – так званими оцінками параметрів. Це пов'язано з обмеженістю розміру вибірки. Чим більша вибірка, тим точніша оцінка параметра і ближче вона до його істинного значення в популяції. Надалі, говорячи про параметри, ми будемо мати на увазі саме їхні оцінки.

У статистиці розрізняють різні види розподілів випадкових величин, зокрема: нормальний, рівномірний, біноміальний, хі-квадрат, Пуассона, Стьюдента, Фішера та інші.

Найчастіше в психологічних дослідженнях застосовують нормальний розподіл.

Нормальний розподіл отримав таку назву через свою поширеність у природничо-наукових дослідженнях, де він виступає «нормою» для масового випадкового прояву ознак. Цей тип розподілу було відкрито трьома вченими в різний час: Муавром у 1733 році в Англії, Гауссом у 1809 році в Німеччині та Лапласом у 1812 році у Франції. Нормальний розподіл є найбільш теоретично дослідженим і зручним для математичного аналізу, на основі якого розроблено потужні методи та прийоми статистичного аналізу даних.

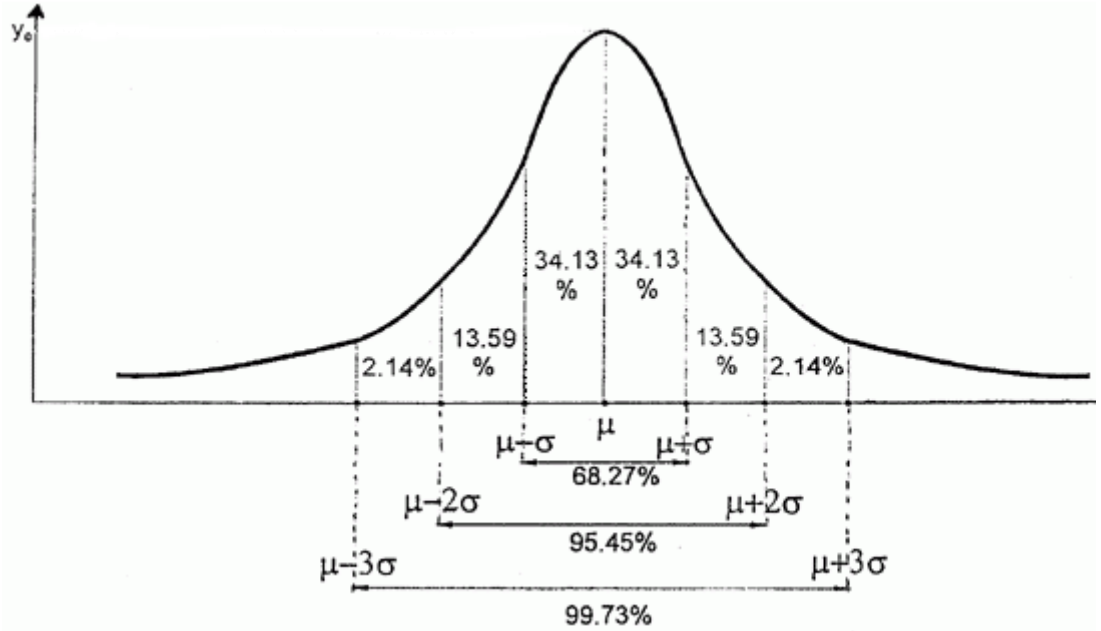


Рис. 2.1. Крива нормального розподілу

Нормальний розподіл характеризується наступними властивостями:

1. Крайні значення ознаки трапляються досить рідко, тоді як значення, близькі до середнього, є найбільш частими.
2. Графічне подання нормального розподілу має форму дзвоноподібної кривої.
3. Крива розподілу є симетричною відносно свого центру.

4. У нормальному розподілі середнє арифметичне, мода та медіана збігаються.

5. Форму розподілу повністю визначають два параметри: середнє арифметичне значення та стандартне відхилення.

6. Слід запам'ятати, що $\pm 1\sigma$ = 68,27 % площі розподілу; $\pm 2\sigma$ = 95,45 %; $\pm 3\sigma$ = 99,73 %.

Якщо стандартне відхилення стале, а величина середнього значення змінюється, то форма кривої нормального розподілу залишається незмінною, а лише її графік зміщується вправо або вліво на осі абсцис. За умови сталої величини середнього значення зміна стандартного відхилення провокує зміну тільки ширини кривої, зі зменшенням сигми крива стає більш вузькою та піднімається водночас вверх, а зі збільшенням сигми крива розширюється, але опускається вниз. Проте у всіх випадках крива нормального розподілу залишається симетричною щодо середнього значення, зберігаючи правильну дзвоноподібну форму.

Незважаючи на те, що теоретично нормальний закон розподілу передбачає можливість існування як нескінченно малих, так і нескінченно великих значень, у психологічних дослідженнях випадкові змінні завжди характеризуються кінцевими діапазонами. На практиці зазвичай застосовуються функції нормального розподілу, обмежені в межах відхилення, що не перевищує ±3σ від середнього значення.

У випадках, коли певні чинники зумовлюють підвищену частоту появи значень, що є більшими або меншими за середнє, формується асиметричний розподіл.

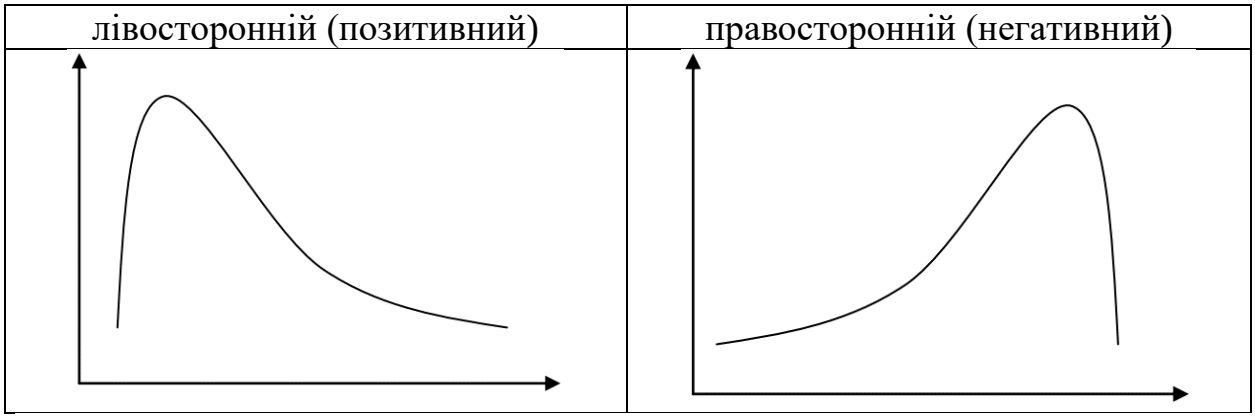


Рис. 2.2. Асиметричні розподіли

У разі лівосторонньої (позитивної) асиметрії в розподілі частіше трапляються більш низькі значення ознаки, а в ситуації правосторонньої (негативної) – більш високі.

Про симетрію можна судити за показником асиметрії.

Асиметрія (А) – це статистичний показник, що відображає ступінь відхилення емпіричного розподілу від симетричного відносно його середнього арифметичного значення. Вона дає змогу оцінити, у який бік

(праворуч чи ліворуч) зміщена більшість значень вибірки відносно центру розподілу.

Асиметрія розраховується за такою формулою:

$$A = \frac{\Sigma\left(x_i - \bar{x}\right)^3}{n \cdot s^3}.$$

Якщо обсяг вибірки невеликий (n < 30) для визначення вибіркового значення асиметрії необхідно застосовувати точну розрахункову формулу:

$$A = \frac{n \cdot \Sigma\left(x_i - \bar{x}\right)^3}{(n-1)\cdot(n-2)\cdot s^3}, \text{ де}$$

x_i – значення ознаки x у конкретного досліджуваного;

$\bar{x}$ – середнє арифметичне значення;

s – середньоквадратичне відхилення;

n – обсяг вибірки досліджуваних.

Для лівосторонніх розподілів $A > 0$; для правосторонніх $A < 0$; для симетричних розподілів $A = 0$.

У випадках, коли наявні чинники зумовлюють підвищену частоту появи середніх або близьких до середніх значень ознаки, формується розподіл із позитивним ексцесом. Якщо ж у вибірці переважають крайні значення (як нижчі, так і вищі за середнє), то такий розподіл характеризується негативним ексцесом, у цьому разі в центральній частині розподілу може спостерігатися впадина.

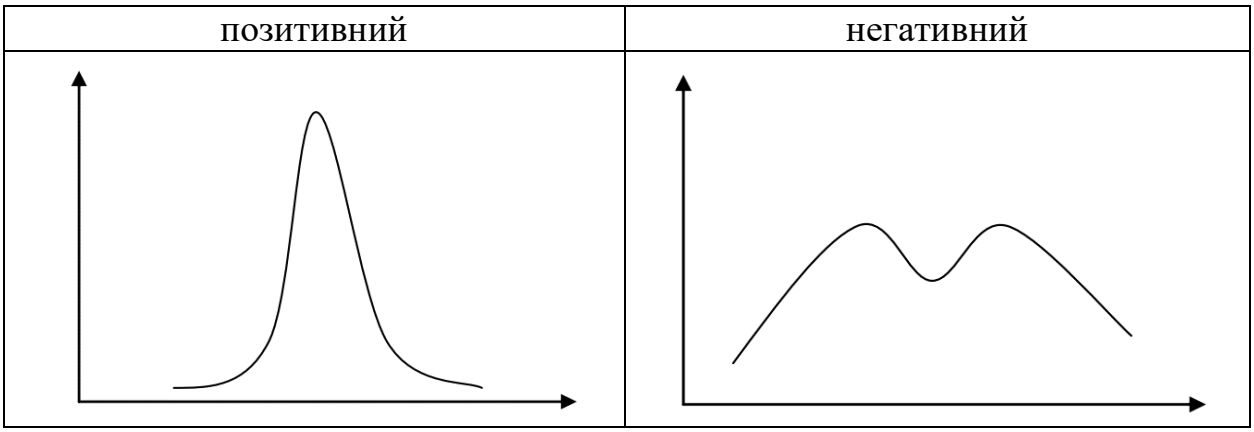


Рис. 2.3. Розподіли з вираженим ексцесом

Ексцес (Е) – це статистичний показник, що характеризує ступінь опуклості або, навпаки, згладженості розподілу вибірки порівняно з нормальним розподілом. Позитивне значення ексцесу свідчить про більш загострений розподіл, за якого більшість значень зосереджена поблизу середнього, а частота крайніх відхилень є нижчою. Негативне значення ексцесу вказує на більш згладжений розподіл, для якого характерна підвищена частота появи крайніх значень і відносно менша концентрація даних біля середнього значення.

Показник ексцесу визначається за формулою:

$$E = \frac{\Sigma\left(x_i - \bar{x}\right)^4}{n \cdot s^4} - 3.$$

Якщо обсяг вибірки невеликий ($n < 30$) для визначення вибіркового значення ексцесу необхідно застосовувати точну розрахункову формулу:

$$E = \frac{n \cdot (n+1) \cdot \Sigma\left(x_i - \bar{x}\right)^4}{(n-1) \cdot (n-2) \cdot (n-3) \cdot s^4} - \frac{3 \cdot (n-1)^2}{(n-2) \cdot (n-3)}, \text{ де}$$

x_i – значення ознаки x у конкретного досліджуваного;

$\bar{x}$ – середнє арифметичне значення;

s – середньоквадратичне відхилення;

n – обсяг вибірки досліджуваних.

У розподілах із позитивним ексцесом $E > 0$, з негативним – $E < 0$. У нормальному розподілі $E = 0$.

Нормальний розподіл характеризується нульовою асиметрією й ексцесом.

2.4. Перевірка даних на відповідність закону нормального розподілу

Нормальний розподіл розглядається як один із фундаментальних законів природи та підтверджений численними емпіричними дослідженнями. Він є характерним для великої кількості ознак генеральної сукупності. При цьому зі збільшенням обсягу вибірки зростає ймовірність того, що емпіричні дані відповідатимуть нормальному закону розподілу.

Завдяки своїм властивостям нормальний розподіл є базовою передумовою для застосування багатьох параметричних методів статистичного аналізу. Саме тому перевірка відповідності даних нормальному розподілу є важливим етапом у наукових дослідженнях.

Проблема виникає у випадках застосування статистичних критеріїв, що ґрунтуються на припущенні нормальності, до даних, які фактично не підпорядковуються цьому закону. Подібна ситуація є типовою для психологічних досліджень, оскільки вони здебільшого здійснюються на вибірках обмеженого обсягу. У зв’язку з цим постає необхідність попередньої перевірки даних на відповідність нормальному розподілу. Така перевірка дає змогу обґрунтовано обирати між параметричними та непараметричними методами статистичного аналізу, забезпечуючи більшу надійність отриманих висновків.

У разі відсутності спеціалізованих статистичних пакетів така перевірка може здійснюватися у «ручному» режимі різними способами.

1. *Візуальне визначення типу розподілу.* Для цього необхідно побудувати гістограму отриманих даних і порівняти її форму з теоретичною кривою нормального розподілу. Якщо форма гістограми наближена до

симетричної дзвоноподібної, можна припустити відповідність даних нормальному закону.

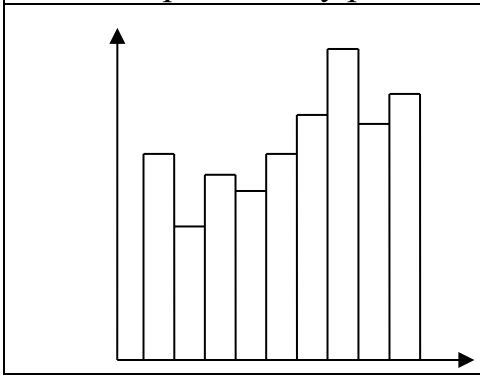
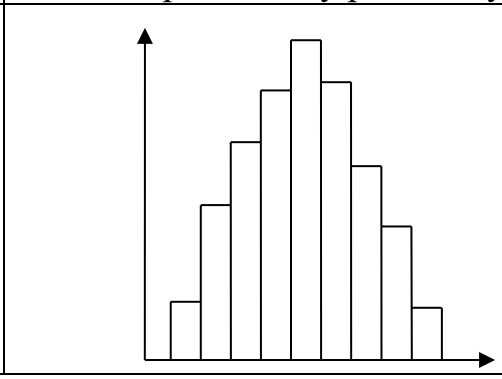
Результати не відповідають нормальному розподілу	Результати відповідають нормальному розподілу
	

Рис. 2.4. Гістограми розподілів даних

Кількість інтервалів, на які необхідно розділити вибірку досліджуваних для конструювання гістограми розподілу, можна визначити за допомогою формули Стерджеса:

$m = 1 + 3{,}32 \cdot lg\, n$, де

lg – десятинний логарифм;

n – обсяг вибірки досліджуваних.

Для визначення кількості інтервалів можна скористатися таблицею 2.6.:

Таблиця 2.6 – Кількість інтервалів в гістограмі залежно від вибірки досліджуваних

Обсяг вибірки (від - до)	6-11	12-22	23-45	46-90	91-181	>182
Кількість інтервалів	4	5	6	7	8	9

Вважається, що дана формула забезпечує побудову достатньо інформативних гістограм у разі, коли обсяг вибірки не перевищує 200 спостережень. За більшої кількості досліджуваних доцільно використовувати формули Скотта або Фрідмана-Діаконіса, які дають змогу точніше визначити оптимальну ширину інтервалів для групування даних.

Величина інтервалів визначається за формулою:

$$h = \frac{x_{\max} - x_{\min}}{n}$$, де

$x_{\max}$ – максимальне значення;

$x_{\min}$ – мінімальне значення;

n – обсяг вибірки досліджуваних.

2. *Необхідно пам'ятати, що в нормальному розподілі даних однакові середнє арифметичне значення ($\bar{x}$), мода (Мо) і медіана (Ме)*. Порядок і особливості їх розрахунку зазначено в матеріалі розділу № 1.

3. *Наближена кількісна оцінка нормальності розподілу* здійснюється на основі розрахунку коефіцієнтів асиметрії та ексцесу, а також помилок цих параметрів. Вказані критерії рекомендується застосовувати лише для малих вибірок.

Розглянемо критерії М.А. Плохінського та Є.І. Пустильника.

Згідно з критерієм М.А. Плохінського розподіл вважається нормальним, якщо показники асиметрії (A) й ексцесу (E) не перевищують у 3 рази свої стандартні помилки вимірювання (m_A і m_E).

Тобто, дані відповідають закону нормального розподілу, якщо:

$$t_A = \frac{|A|}{m_A} \le 3 \quad \text{та} \quad t_E = \frac{|E|}{m_E} \le 3, \text{ де}$$

$$m_A = \sqrt{\frac{6}{n+3}}, \quad m_E = 2 \cdot \sqrt{\frac{6}{n+3}}.$$

У статистиці під «помилкою» слід розуміти не помилку дослідження, а міру представництва даної величини, тобто наскільки величина отримана з вибіркової сукупності відрізняється від істинної в генеральній сукупності.

Згідно з критерієм Є.І. Пустильника розподіл вважається нормальним, якщо показники асиметрії (A) й ексцесу (E) менші своїх критичних значень ($A_{крит.}$ і $E_{крит.}$).

Тобто, дані відповідають закону нормального розподілу, якщо:

$|A| \le A_{крит.}$ та $|E| \le E_{крит.}$, де

$$A_{крит.} = 3 \cdot \sqrt{\frac{6 \cdot (n-1)}{(n+1)(n+3)}},$$

$$E_{крит.} = 5 \cdot \sqrt{\frac{24 \cdot n \cdot (n-2) \cdot (n-3)}{(n+1)^2 \cdot (n+3) \cdot (n+5)}}.$$

Приклад 1. Чи можна стверджувати, що отримані в досліджені емпіричні дані відповідають закону нормального розподілу?

Таблиця 2.7 – Таблиця первинних даних та проміжних розрахунків

№	X	$\left(x_i - \bar{x}\right)$	$\left(x_i - \bar{x}\right)^2$	$\left(x_i - \bar{x}\right)^3$	$\left(x_i - \bar{x}\right)^4$
1	4	-4,43	19,62	-86,94	385,14
2	5	-3,43	11,76	-40,35	138,41
3	8	-0,43	0,18	-0,08	0,03
4	7	-1,43	2,04	-2,92	4,18
5	9	0,57	0,32	0,19	0,11
6	6	-2,43	5,90	-14,35	34,87
7	8	-0,43	0,18	-0,08	0,03
8	12	3,57	12,74	45,50	162,43
9	10	1,57	2,46	3,87	6,08
10	9	0,57	0,32	0,19	0,11

Продовження таблиці 2.7

11	6	-2,43	5,90	-14,35	34,87
12	11	2,57	6,60	16,97	43,62
13	10	1,57	2,46	3,87	6,08
14	13	4,57	20,88	95,44	436,18
	$\bar{x}$ = 8,43	Σ = -0,02	Σ =91,43	Σ = 6,96	Σ = 1252,13

Формулюємо нульову і альтернативну гіпотези:

H_0: розподіл емпіричних даних відрізняється від закону нормального розподілу;

H_1: розподіл емпіричних даних відповідає закону нормального розподілу.

Застосуємо *критерій М.А. Плохинського*.

Для розрахунку показників асиметрії та ексцесу необхідно розрахувати середнє арифметичне значення ($\bar{x}$) та середньоквадратичне відхилення (s).

1. Визначаємо середнє арифметичне значення змінної X.

$\bar{x} = 8{,}43$.

2. Для кожного значення змінної x_i знаходимо різницю $\left(x_i - \bar{x}\right)$. Сума відхилень від середнього значення повинна бути рівна нулю (з точністю до похибки обчислень).

3. Кожне відхилення від середнього значення підносимо до квадрату і фіксуємо результат у відповідний стовпчик $\left(x_i - \bar{x}\right)^2$.

Сума квадратів відхилень необхідна для обчислення стандартного відхилення.

$$s = \sqrt{\frac{\Sigma\left(x_i - \bar{x}\right)^2}{n-1}} = \sqrt{\frac{91{,}43}{14-1}} = 2{,}65.$$

4. Обчислюємо суми результатів стовпчиків $\left(x_i - \bar{x}\right)^3$ та $\left(x_i - \bar{x}\right)^4$.

5. Розрахуємо показник асиметрії. Застосуємо точну розрахункову формулу.

$$A = \frac{n \cdot \Sigma\left(x_i - \bar{x}\right)^3}{(n-1)\cdot(n-2)\cdot s^3} = \frac{14 \cdot 6{,}96}{(14-1)\cdot(14-2)\cdot 2{,}65^3} = 0{,}035.$$

6. Розрахуємо показник ексцесу за точною розрахунковою формулою.

$$E = \frac{n\cdot(n+1)\cdot\Sigma\left(x_i - \bar{x}\right)^4}{(n-1)\cdot(n-2)\cdot(n-3)\cdot s^4} - \frac{3\cdot(n-1)^2}{(n-2)\cdot(n-3)} = \frac{14\cdot(14+1)\cdot 1252{,}13}{(14-1)\cdot(14-2)\cdot(14-3)\cdot 2{,}65^4} -$$

$$- \frac{3\cdot(14-1)^2}{(14-2)\cdot(14-3)} = 3{,}107 - 3{,}841 = -0{,}734.$$

7. Визначаємо стандартні помилки вимірювання асиметрії та ексцесу.

$$m_A = \sqrt{\frac{6}{n+3}} = \sqrt{\frac{6}{14+3}} = 0{,}59;$$

$$m_E = 2 \cdot \sqrt{\frac{6}{n+3}} = 2 \cdot \sqrt{\frac{6}{14+3}} = 1{,}18.$$

8. Згідно з критерієм М.А. Плохинського показники асиметрії й ексцесу свідчать про статистичну відмінність емпіричного розподілу від нормального в тому випадку, якщо вони перевищують за абсолютною величиною свої стандартні помилки вимірювання в 3 рази.

$$t_A = \frac{|A|}{m_A} = \frac{|0{,}035|}{0{,}59} = 0{,}06;$$

$$t_E = \frac{|E|}{m_E} = \frac{|-0{,}734|}{1{,}18} = 0{,}62.$$

За результатами проведених розрахунків обидва показники не перевищують утричі значення власних стандартних помилок вимірювання. Це дає підстави зробити висновок, що розподіл статистичних даних не суперечить закону нормального розподілу.

Тепер здійснимо перевірку за *критерієм Є.І. Пустильника*.

1. Розрахуємо критичні значення для показників асиметрії та ексцесу:

$$A_{крит.} = 3 \cdot \sqrt{\frac{6 \cdot (n-1)}{(n+1)(n+3)}} = 3 \cdot \sqrt{\frac{6 \cdot (14-1)}{(14+1) \cdot (14+3)}} = 3 \cdot 0{,}55 = 1{,}65;$$

$$E_{крит.} = 5 \cdot \sqrt{\frac{24 \cdot n \cdot (n-2) \cdot (n-3)}{(n+1)^2 \cdot (n+3) \cdot (n+5)}} = 5 \cdot \sqrt{\frac{24 \cdot 14 \cdot (14-2) \cdot (14-3)}{(14+1)^2 \cdot (14+3) \cdot (14+5)}} = 5 \cdot 0{,}78 = 3{,}9.$$

2. Згідно з критерієм Є.І. Пустильника дані відповідають закону нормального розподілу, якщо:

$|A| \le A_{крит.}$ та $|E| \le E_{крит.}$

У нашому випадку:

|0,035| < 1,65,

|-0,734| < 3,9.

Таким чином, можна зробити висновок, що дані відповідають закону нормального розподілу. Обидва варіанти перевірки – за критеріями М.А. Плохинського та Є.І. Пустильника – засвідчили однаковий результат: розподіл ознаки не відрізняється від нормального. Це означає, що до змінної *X* можна без обмежень застосовувати параметричні критерії статистичного аналізу. У практичних дослідженнях допустимо обрати будь-який із зазначених варіантів перевірки та послідовно дотримуватися його.

У випадку значного обсягу вибірки доцільно виконувати обчислення оцінок параметрів за допомогою комп'ютерної техніки із використанням спеціалізованих статистичних програмних пакетів. Рекомендованими для цього є Jamovi, JASP, SPSS Statistics та інші.

У статистичних програмах аналізу даних порівняння емпіричного розподілу даних із теоретично нормальним розподілом здійснюється за допомогою таких взаємодоповнювальних методів:

1. *Визначення середнього арифметичного, моди та медіани.* Співпадіння цих показників слугує індикатором відповідності емпіричних даних закону нормального розподілу.

2. *Побудова гістограми розподілу.* Гістограма дозволяє візуально зіставити емпіричний розподіл вибірки з теоретичною кривою нормального розподілу та оцінити відхилення даних від нього.

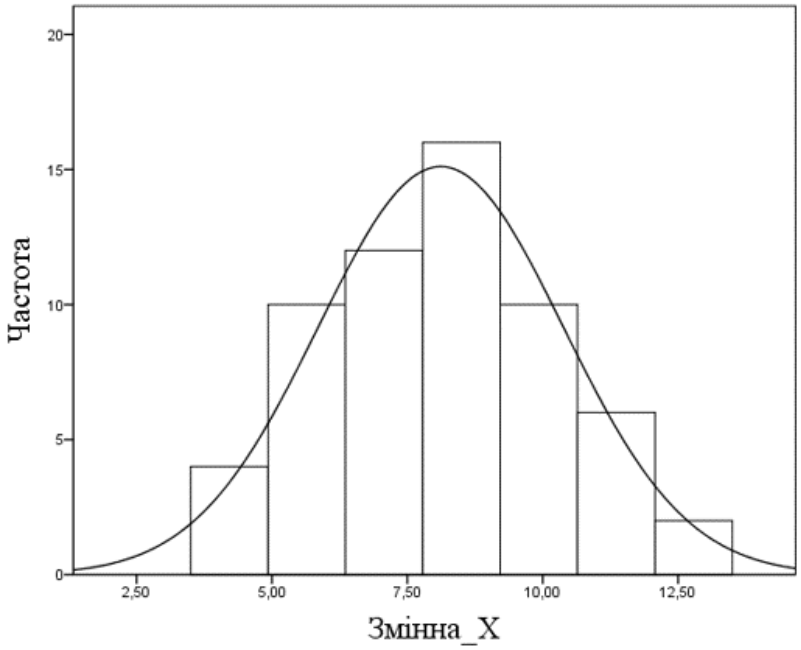


Рис. 2.2. Гістограма розподілу даних

У наведеному прикладі n = 60. Відповідно до формули Стерджеса побудовано гістограму, що складається із семи інтервалів розподілу. Отримана дзвоноподібна форма гістограми є візуальним підтвердженням відповідності емпіричних даних закону нормального розподілу.

3. *Побудова квантильного графіка* (*Q-Q plot)*. Такий графік дозволяє зіставити емпіричні квантилі вибірки з теоретичними квантилями нормального розподілу. Чим ближче точки розташовуються до діагональної лінії, тим вищою є відповідність розподілу емпіричних даних нормальному закону.

Квантильний графік будується за такою процедурою. Спочатку всі емпіричні значення впорядковуються за рангами. Відповідно до цих рангів розраховуються z-значення (стандартизовані значення нормального розподілу) за припущенням про нормальність даних. Обчислені z-значення відкладаються на осі Y, тоді як упорядковані емпіричні значення розміщуються на осі X.

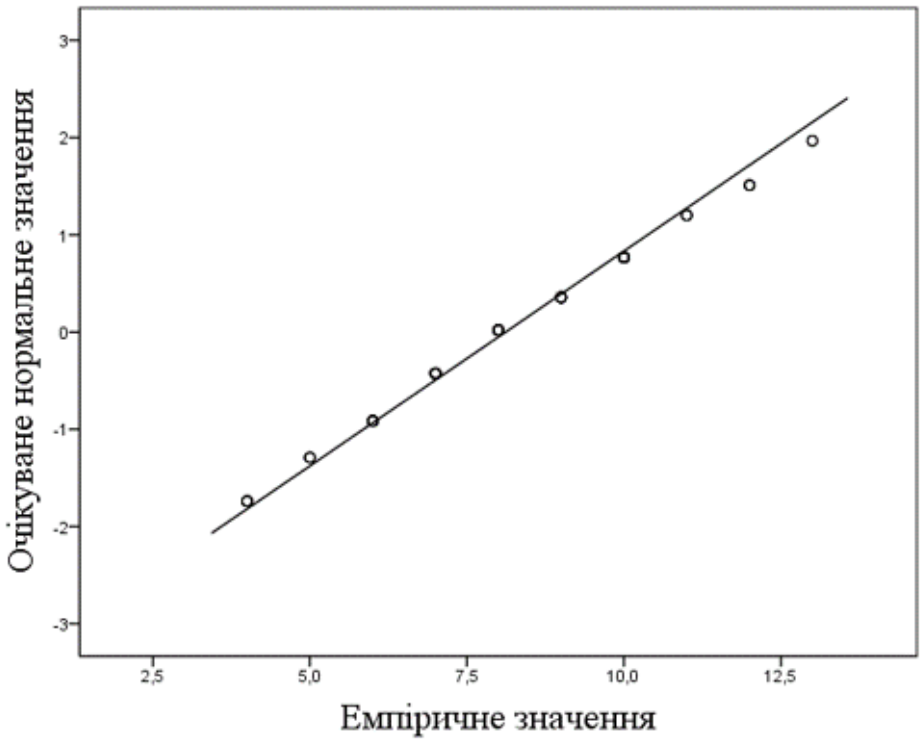


Рис. 2.3. Квантильний графік

За умови нормального розподілу емпіричних даних усі точки на графіку тяжіють до прямої лінії. У разі відхилення від нормальності спостерігається систематичне відхилення точок від діагоналі. Додатковою перевагою даного графічного методу є можливість виявлення викидів (аутлаєрів).

У наведеному прикладі точки лише незначною мірою відхиляються від прямої, що свідчить на користь відповідності емпіричних даних закону нормального розподілу.

4. *Побудова діаграми типу Box Plot* («ящик з вусами», коробчаста діаграма).

Діаграма Box Plot є наочним інструментом для стислого відображення основних характеристик розподілу даних. Вона відображає медіану, нижній та верхній квартилі, а також мінімальне й максимальне значення вибірки у межах 1,5 міжквартильного розмаху. Значення, що виходять за ці межі, позначаються як викиди (аутлаєри).

Відстань між межами «коробки» та розташування «вусів» дозволяють оцінити ступінь варіативності та наявність асиметрії в розподілі даних.

На діаграмі типу Box Plot висота прямокутника відповідає ширині міжквартильного розмаху (*IQR*), тобто інтервалу від 25 до 75 процентиля ($P_{25} – P_{75}$) або першого та третього квартилю ($Q_1 – Q_3$). У контексті наведених даних це означає, що в діапазоні від 14 до 20 зосереджено 50 % результатів: у 25 % досліджуваних показник змінної X має значення нижче 14, тоді як у 25 % – перевищує 20.

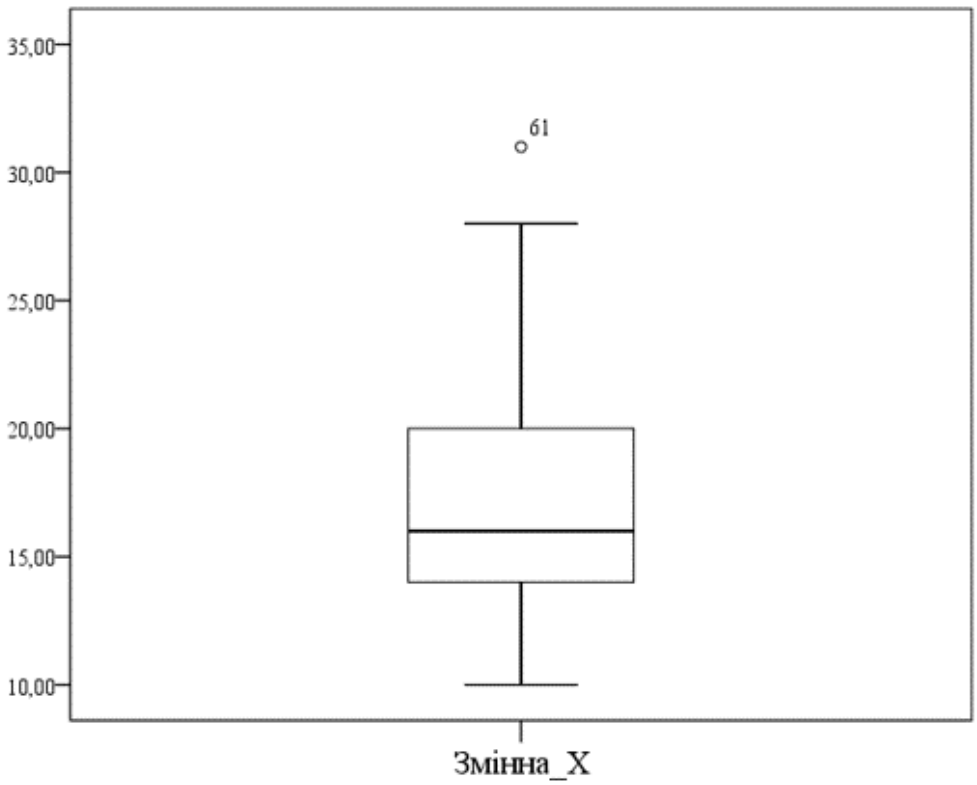


Рис. 2.4 Діаграма типу Box Plot

Горизонтальна лінія в середині прямокутника відповідає медіані (*Md*) або другому квартилю (Q_2).

Межі «вусів» визначаються за такою формулою:

$X_{нижній} = Q_1 - 1{,}5 \cdot (Q_3 - Q_1)$;

$X_{верхній} = Q_3 + 1{,}5 \cdot (Q_3 - Q_1)$.

Межі «вусів» діаграми визначаються на основі міжквартильного розмаху. Зокрема, від першого квартиля віднімається 1,5 величини міжквартильного розмаху; якщо у вибірці є значення, що менше цього порогу, воно розглядається як викид (аутлаєр). Аналогічно верхня межа визначається як третій квартиль плюс 1,5 міжквартильного розмаху; усі значення, що перевищують цю межу, також класифікуються як аутлаєри. Останнє значення даних, яке потрапляє в зазначений інтервал, і визначає межу «вуса», тоді як аутлаєри до розрахунків не включаються.

У нашому прикладі «вусами» позначено мінімальне значення (10) та максимальне значення (28), що належать до діапазону 1,5 міжквартильного розмаху, тоді як аутлаєр (результат досліджуваного № 61) відображено окремим маркером (кружечком).

Інтерпретація діаграми свідчить про те, що медіана не розташована в центрі міжквартильного розмаху, «вуса» є асиметричними (нижній коротший за верхній), а також наявний аутлаєр у даних. Сукупність цих ознак вказує на відхилення емпіричного розподілу від нормального.

5. *Застосування статистичних критеріїв Колмогорова-Смирнова з поправкою Ліллієфорса та Шапіро-Уїлка*.

Якщо отримане в результаті розрахунку значення статистичної значущості перевищує 0,05 ($p > 0,05$), то емпіричний розподіл можна вважати таким, що відповідає нормальному закону. Критерій Колмогорова-Смирнова з поправкою Лілліефорса доцільно застосовувати для вибірок великого обсягу, тоді як критерій Шапіро-Уїлка рекомендується для вибірок порівняно невеликого розміру.

Результати перевірки нормальності розподілу слід інтерпретувати комплексно. Використання лише критеріїв Колмогорова-Смирнова або Шапіро-Уїлка не завжди є достатнім, оскільки їхня чутливість істотно залежить від обсягу вибірки. Так, за однакових відхилень емпіричного розподілу від нормального ймовірність отримання статистично значущих відмінностей при $n = 1000$ значно вища, ніж при $n = 30$. Деякі дослідники пропонують завжди вважати розподіл відмінним від нормального у випадку малих вибірок ($n < 30$).

В науковій літературі надаються наступні практичні рекомендації: при кількості спостережень від 30 до 100, якщо статистичні критерії покажуть відхилення розподілу від нормального ($p < 0,05$), слід вважати, що розподіл відрізняється від нормального, якщо графіки і значення асиметрії та ексцесу не свідчать про зворотне. Якщо кількість спостережень перевищує 100, і статистична значущість критеріїв перевірки розподілу на нормальність перевищує 0,05 ($p > 0,05$), то розподіл вважають нормальним, якщо графіки і значення асиметрії та ексцесу не говорять про зворотне.

Для умовної відповідності емпіричного розподілу нормальному вважається, що показники асиметрії та ексцесу мають перебувати у межах від –1 до +1.

Завдання на самостійну підготовку

1. Характеристика статистичних гіпотез.
2. Ідея процедури перевірки статистичної значущості нульової гіпотези.
3. Навіщо здійснювати статистичний аналіз даних?
4. Розповсюджені помилки в інтерпретації p-значення статистичної значущості при перевірці гіпотез.
5. Характеристика помилок I та II роду в дослідженні.
6. Розмір ефекту як показник практичної значущості дослідження.
7. Важливість розрахунку обсягу вибірки в дослідженні.
8. Нормальний закон розподілення даних і його застосування. Характеристики параметрів нормального закону розподілення.
9. Методи перевірки даних на відповідність закону нормального розподілу.
10. Які основні параметри z-розподілу?

РОЗДІЛ 3
МЕТОДИ АНАЛІЗУ СТАТИСТИЧНОГО ЗВ'ЯЗКУ МІЖ ЗМІННИМИ

3.1. Сутність методів встановлення статистичних зв'язків

Важливим завданням математико-статистичних досліджень у психології є виявлення специфічних зв'язків між різними змінними, ознаками та факторами. Особливість цих зв'язків полягає в тому, що залежність між змінними зазвичай неможливо описати у вигляді функціональної залежності, тобто через однозначно визначену функцію. Натомість зв'язок проявляється інакше: від значення однієї змінної залежить імовірний характер розподілу іншої. Отже, навіть якщо неможливо точно передбачити значення залежної ознаки, завжди можна отримати інформацію про ймовірні межі її варіювання, середнє значення та ступінь мінливості. Такі залежності у статистиці отримали назву кореляцій, коли йдеться про кількісні або порядкові показники, і асоціацій, коли мова йде про номінальні показники. Для узагальненої характеристики методів виявлення статистичних зв'язків у подальшому використовуватимемо термін «кореляція».

Кореляція – це статистична залежність, за якої кожному значенню однієї змінної x відповідає певне очікуване значення іншої змінної y. Така залежність не обов'язково є причинно-наслідковою, проте свідчить про наявність взаємозв'язку між змінними.

Термін «кореляція» уперше був застосований французьким палеонтологом Ж. Кюв'є, який сформулював закон кореляції частин і органів тварин. Згідно з цим законом, за окремими частинами тіла можна реконструювати загальний вигляд тварини, оскільки між окремими органами існує взаємозумовленість.

У статистичну термінологію поняття кореляції було введено англійським біологом і статистиком Ф. Гальтоном. Він досліджував статистичні зв'язки між ознаками, зокрема в контексті спадковості та антропометрії.

У статистиці міри зв'язку (коефіцієнти кореляції) використовуються для кількісного оцінювання сили та напряму зв'язку між двома змінними, які спостерігаються в межах однієї вибірки.

Кореляційний аналіз для двох випадкових величин включає такі основні етапи:

- графічне представлення залежності між змінними (наприклад, за допомогою діаграми розсіювання);
- обчислення коефіцієнта кореляції;
- перевірка статистичної значущості виявленого зв'язку.

У якості числового показника тісноти зв'язку між змінними застосовується коефіцієнт кореляції (r).

Коефіцієнт кореляції – це числова характеристика, яка відображає силу

та напрямок імовірного зв’язку між двома змінними. Його значення лежить у межах від –1 до +1.

Кореляційні зв’язки розрізняються за силою, формою та напрямком.

Показником сили зв’язку є абсолютна величина коефіцієнта кореляції. Сила зв’язку безпосередньо вказує, наскільки синхронно проявляється спільна мінливість досліджуваних змінних. Силу зв’язку можна оцінити за допомогою шкали Чедока.

Таблиця 3.1 – Класифікація зв’язку за силою (шкала Чедока)

Якісна характеристика сили зв’язку	Кількісна міра сили зв’язку
дуже слабка	$0{,}00 < \lvert r \rvert \leq 0{,}10$
слабка	$0{,}10 < \lvert r \rvert \leq 0{,}30$
помірна	$0{,}30 < \lvert r \rvert \leq 0{,}50$
середня	$0{,}50 < \lvert r \rvert \leq 0{,}70$
сильна	$0{,}70 < \lvert r \rvert \leq 0{,}90$
дуже сильна	$0{,}90 < \lvert r \rvert \leq 1{,}00$

Наочне уявлення про характер зв’язку між змінними надає діаграма розсіювання – графічне зображення, у якому осі координат відповідають значенням двох змінних, а кожне спостереження представлено у вигляді точки. Такий графік дозволяє візуально оцінити наявність, напрям і форму зв’язку.

У більшості випадків статистичні кореляційні методи орієнтовані на виявлення та інтерпретацію зв’язків лінійної форми, тобто таких, які можна описати рівнянням прямої. Саме тому перед застосуванням відповідних критеріїв доцільно провести попередній візуальний аналіз за допомогою діаграми розсіювання.

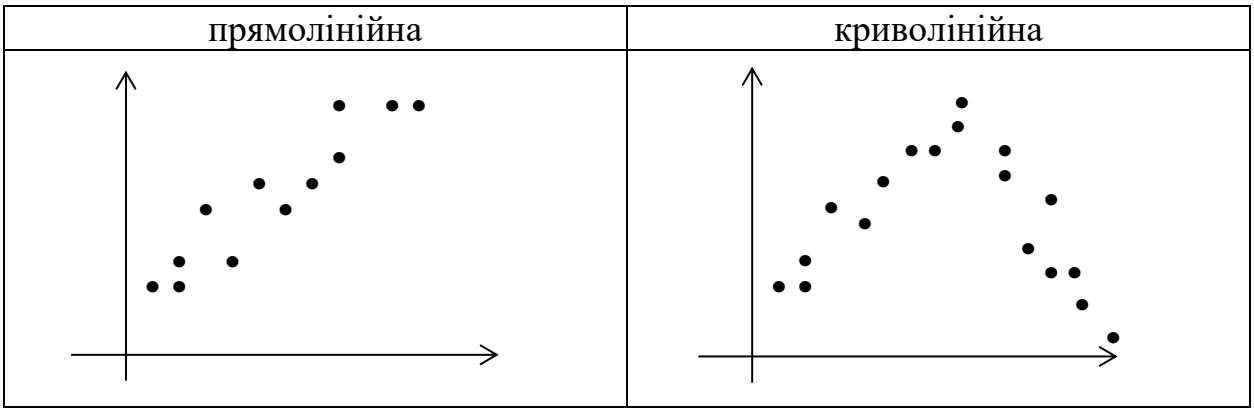


Рис. 3.1. Класифікація зв’язку за формою

Показником напрямку зв’язку між змінними є знак коефіцієнта кореляції.

Позитивне значення коефіцієнта вказує на прямий зв'язок (зі зростанням однієї змінної зростає й інша), тоді як від'ємне – на обернений зв'язок (зі зростанням однієї змінної інша зменшується).

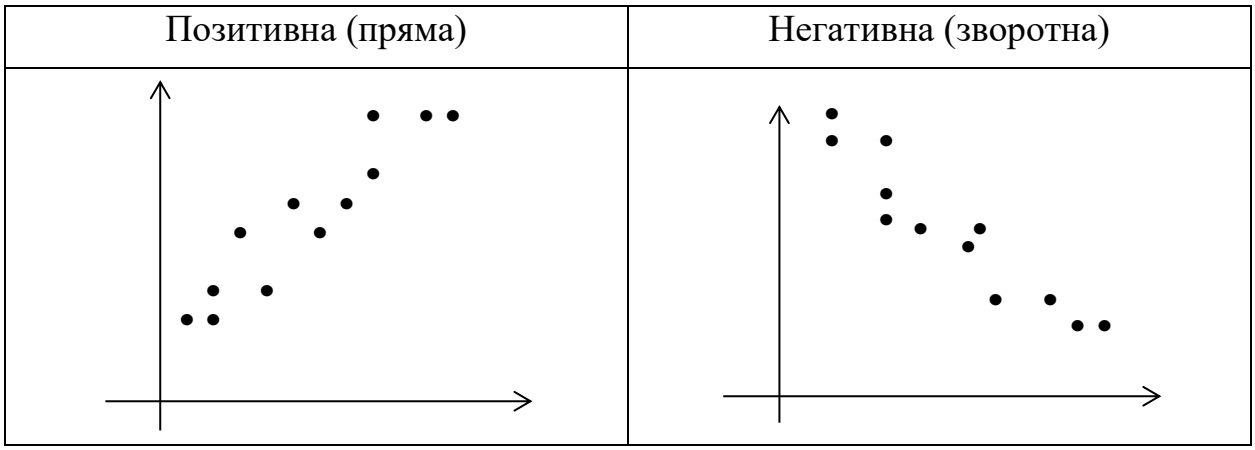


Рис. 3.2. Класифікація зв'язку за напрямком

Коефіцієнти кореляції характеризуються не лише напрямком і силою зв'язку, але й статистичною значущістю. Зокрема, сильний кореляційний зв'язок може виявитися статистично незначущим у разі невеликого обсягу вибірки, тоді як слабкий зв'язок може бути статистично значущим за наявності великої кількості спостережень. Тому під час інтерпретації результатів кореляційного аналізу важливо враховувати величину розміру ефекту.

Крім того, слід мати на увазі, що наявність кореляції не є підтвердженням причинно-наслідкового зв'язку. Іншими словами, встановлення статистичної залежності між двома змінними не дає підстав стверджувати, що одна змінна є причиною змін в іншій. Кореляція лише вказує на наявність асоціації, яка може бути зумовлена як прямими, так і опосередкованими чинниками або спільним впливом сторонньої змінної.

Якщо статистично значущий зв'язок між змінними не виявлено, але існують теоретичні або емпіричні підстави припускати його наявність, доцільно перевірити можливі причини недостовірності результату. Основними з них можуть бути:

1. *Нелінійність зв'язку*. Для виявлення такої ситуації слід проаналізувати діаграму розсіювання. Якщо залежність між змінними виявляється нелінійною, але зберігає монотонність, доцільно застосувати рангові коефіцієнти кореляції (наприклад, коефіцієнт Спірмена або Кендалла). Якщо ж зв'язок не є монотонним, рекомендується розділити вибірку на підгрупи, у межах яких зв'язок має монотонний характер, і обчислити кореляцію окремо для кожної підгрупи або сформувати контрастні групи (наприклад, за крайніми значеннями ознаки) та порівняти їх за рівнем вираженості іншої змінної.

2. *Наявність аномальних спостережень (аутлаєрів) або виражена асиметрія розподілу однієї чи обох змінних.* Для цього слід проаналізувати гістограми розподілу частот. У разі наявності аутлаєрів або асиметрії необхідно видалити аутлаєри або перейти до рангової кореляції.

3. *Неоднорідність вибірки.* Варто дослідити діаграму розсіювання на предмет наявності структурної неоднорідності (кластерів), яка може маскувати наявний зв'язок. У такому випадку доцільно розділити вибірку на окремі підгрупи, в яких зв'язок може бути наявним, але мати різний напрямок або силу.

Якщо зв'язок між змінними є статистично значущим, то перед формулюванням змістовних висновків необхідно виключити ймовірність помилкової (спотвореної) кореляції. До основних можливих причин такої кореляції належать:

1. *Наявність аномальних спостережень (аутлаєрів).* Аутлаєри можуть суттєво впливати на значення коефіцієнта кореляції, викривляючи реальний характер зв'язку. У разі їх виявлення доцільно або видалити такі спостереження з вибірки (якщо їх наявність не є теоретично обґрунтованою), або перейти до використання рангових коефіцієнтів кореляції, які є менш чутливими до впливу таких значень.

2. *Вплив третьої (контрольної або латентної) змінної.* Кореляція між двома змінними може бути зумовлена дією третього фактора, який одночасно впливає на обидві змінні. У такому випадку рекомендовано: розбити вибірку на підгрупи відповідно до рівнів третьої змінної; обчислити коефіцієнти кореляції окремо для кожної групи; або застосувати часткову кореляцію, що дозволяє оцінити зв'язок між двома змінними при фіксованому рівні третьої.

При оцінці кореляційного зв'язку необхідно враховувати тип шкал, у яких вимірюються змінні, характер розподілу первинних даних та форму залежності між змінними.

Розмір ефекту безпосередньо характеризується емпіричним значенням коефіцієнта кореляції r, тому додаткові розрахунки для його визначення, як правило, не потрібні.

Таблиця 3.2 – Методичні рекомендації інтерпретації величину розміру ефекту кореляційної залежності запропоновані Дж. Коеном

Рівень розміру ефекту	Величина коефіцієнта кореляції
несуттєвий	$0{,}00 \le \lvert r \rvert < 0{,}10$
малий	$0{,}10 \le \lvert r \rvert < 0{,}30$
середній	$0{,}30 \le \lvert r \rvert < 0{,}50$
великий	$0{,}50 \le \lvert r \rvert < 1{,}00$

Для спрощення вибору відповідного коефіцієнта кореляції залежно від типу вимірювальної шкали змінних доцільно скористатися таблицею 3.3.

Таблиця 3.3 – Статистичні методи знаходження взаємозв’язку між змінними

<table>
<tr><td colspan="2" rowspan="2">Вимірювальні шкали</td><td colspan="2">Шкала номінальна</td><td rowspan="2">Шкала порядку</td><td rowspan="2">Шкали інтервалів та відношень</td></tr>
<tr><td>k = 2</td><td>k > 2</td></tr>
<tr><td rowspan="2">Шкала номінальна</td><td>k = 2</td><td>Коефіцієнт контингенції Пірсона φ

Коефіцієнт асоціації Юла Q</td><td></td><td></td><td></td></tr>
<tr><td>k > 2</td><td>Коефіцієнт взаємної зв’язаності Чупрова K

Коефіцієнт взаємної зв’язаності Крамера V</td><td>Коефіцієнт взаємної зв’язаності Пірсона C

Коефіцієнт взаємної зв’язаності Чупрова K

Коефіцієнт взаємної зв’язаності Крамера V</td><td></td><td></td></tr>
<tr><td colspan="2">Шкала порядку</td><td>Рангово-бісеріальний коефіцієнт кореляції r_{rb}</td><td></td><td>Коефіцієнт кореляції r_s Спірмена

Коефіцієнт кореляції τ Кендалла</td><td></td></tr>
<tr><td colspan="2">Шкали інтервалів та відношень</td><td>Точково-бісеріальний коефіцієнт кореляції r_{pb}</td><td></td><td></td><td>Коефіцієнт кореляції Пірсона r_{xy}</td></tr>
</table>

Примітка: k – кількість градацій номінальної змінної.

3.2. Лінійна кореляція

3.2.1. Коефіцієнт лінійної кореляції Пірсона r_{xy}

Для дослідження лінійного взаємозв’язку між двома кількісними змінними, виміряними у інтервальній або шкалі відношень, застосовується коефіцієнт кореляції Пірсона. У науковій літературі, коли йдеться про кореляцію без додаткових уточнень, зазвичай мається на увазі саме цей метод.

Коефіцієнт кореляції Пірсона фіксує лінійний зв’язок між змінними та приймає значення в діапазоні від –1 до +1. Знак коефіцієнта відображає напрямок взаємозв’язку (позитивний або негативний), а абсолютне значення характеризує силу зв’язку.

Коефіцієнт кореляції Пірсона є параметричним методом, коректне застосування якого можливе за умови, що розподіли двох змінних наближені до нормального. Його значення обчислюється за формулою:

$$r_{xy} = \frac{\Sigma\left(x_i - \bar{x}\right)\cdot\left(y_i - \bar{y}\right)}{\left(n-1\right)\cdot s_x \cdot s_y}$$, де

x_i, y_i – первинні значення змінних *X* та *Y*;

$\bar{x}$, $\bar{y}$ – середнє арифметичне змінних *X* та *Y*;

s_x, s_y – стандартне відхилення змінних *X* та *Y*;

n – обсяг вибірки досліджуваних.

Якщо емпіричне значення r_{xy} дорівнює або перевищує критичне значення для заданого α-рівня, то H_0 відхиляється і формулюється висновок, що коефіцієнт кореляції статистично значуще відрізняється від нуля.

Показником розміру ефекту зв'язку між змінними окрім безпосередньо самого емпіричного значення коефіцієнта кореляції Пірсона r_{xy} може виступати й коефіцієнт детермінації r^2, тобто коефіцієнт кореляції в квадраті. Він змінюється від 0 до 1 (або від 0 до 100 %) і показує в якій мірі мінливість однієї змінної обумовлена (детермінована) впливом іншої змінної. На підставі показника r^2 Дж. Коен виділив наступні градації величини розміру ефекту:

- $0{,}00 \le r^2 < 0{,}01$ – несуттєвий;
- $0{,}01 \le r^2 < 0{,}09$ – малий;
- $0{,}09 \le r^2 < 0{,}25$– середній;
- $0{,}25 \le r^2 < 1{,}00$ – великий.

Приклад 3.1. У вибірці з 20 досліджуваних було виміряно рівень інтернет-залежності та рівень тривожності особистості. Поставлено завдання визначити наявність та статистичну значущість кореляційного зв'язку між цими двома параметрами.

Таблиця 3.4 – Таблиця первинних даних та проміжних розрахунків

№	Інтернет-залежність (*X*)	Рівень тривожності (*Y*)	$\left(x_i - \bar{x}\right)$	$\left(y_i - \bar{y}\right)$	$\left(x_i - \bar{x}\right)^2$	$\left(y_i - \bar{y}\right)^2$	$\left(x_i - \bar{x}\right)\cdot\left(y_i - \bar{y}\right)$
1	84	20	32,1	3,5	1030,41	12,25	112,35
2	57	16	5,1	-0,5	26,01	0,25	-2,55
3	21	7	-30,9	-9,5	954,81	90,25	293,55
4	32	15	-19,9	-1,5	396,01	2,25	29,85
5	69	17	17,1	0,5	292,41	0,25	8,55
6	37	14	-14,9	-2,5	222,01	6,25	37,25
7	49	17	-2,9	0,5	8,41	0,25	-1,45
8	47	16	-4,9	-0,5	24,01	0,25	2,45
9	82	29	30,1	12,5	906,01	156,25	376,25
10	65	22	13,1	5,5	171,61	30,25	72,05
11	23	11	-28,9	-5,5	835,21	30,25	158,95

Продовження таблиці 3.4.

12	48	14	-3,9	-2,5	15,21	6,25	9,75
13	77	25	25,1	8,5	630,01	72,25	213,35
14	39	7	-12,9	-9,5	166,41	90,25	122,55
15	45	16	-6,9	-0,5	47,61	0,25	3,45
16	68	19	16,1	2,5	259,21	6,25	40,25
17	68	21	16,1	4,5	259,21	20,25	72,45
18	27	10	-24,9	-6,5	620,01	42,25	161,85
19	41	16	-10,9	-0,5	118,81	0,25	5,45
20	59	18	7,1	1,5	50,41	2,25	10,65
	$\bar{x} = 51,9$	$\bar{y} = 16,5$	$\Sigma = 0$	$\Sigma = 0$	$\Sigma = 7033,8$	$\Sigma = 569$	$\Sigma = 1727$

Будемо перевіряти первинні дані на можливість використання параметричного коефіцієнта лінійної кореляції r_{xy} Пірсона.

1. З метою перевірки даних у обох змінних на відповідність закону нормального розподілу використаємо критерій М.А. Плохінського.

Розподіл емпіричних даних вважається таким, що відповідає нормальному, якщо виконуються такі умови:

$$t_A = \frac{|A|}{m_A} \le 3 \quad \text{та} \quad t_E = \frac{|E|}{m_E} \le 3.$$

Для розрахунку показників асиметрії та ексцесу необхідно розрахувати середнє арифметичне значення та стандартне відхилення.

$\bar{x} = 51,9$; $\bar{y} = 16,5$.

Для кожного значення змінних x_i та y_i знаходимо різницю $\left(x_i - \bar{x}\right)$ та $\left(y_i - \bar{y}\right)$, результат записуємо у відповідні стовпчики. Суми відхилень від середнього значення для кожної змінної повинні бути рівні нулю (з точністю до похибки обчислень).

Кожне відхилення від середнього значення підносимо до квадрату і фіксуємо результат у відповідні стовпчики $\left(x_i - \bar{x}\right)^2$ та $\left(y_i - \bar{y}\right)^2$.

Суми квадратів відхилень необхідні для обчислення стандартних відхилень.

$$s_x = \sqrt{\frac{\Sigma\left(x_i - \bar{x}\right)^2}{n-1}} = \sqrt{\frac{7033,8}{20-1}} = 19,24;$$

$$s_y = \sqrt{\frac{\Sigma\left(y_i - \bar{y}\right)^2}{n-1}} = \sqrt{\frac{569}{20-1}} = 5,47.$$

Додатково розрахували:

$\Sigma \left(x_i - \bar{x}\right)^3 = 8006,16;$

$\Sigma \left(x_i - \bar{x}\right)^4 = 4776778,43;$

$\Sigma \left(y_i - \bar{y}\right)^3 = 696;$

$\Sigma \left(y_i - \bar{y}\right)^4 = 50227{,}25.$

Для визначення вибіркового значення асиметрії та ексцесу в обох вибірках дослідження застосуємо точні розрахункові формули.

$$A_x = \frac{n \cdot \Sigma\left(x_i - \bar{x}\right)^3}{(n-1)\cdot(n-2)\cdot s_x^3} = \frac{20 \cdot 8006{,}16}{(20-1)\cdot(20-2)\cdot 19{,}24^3} = 0{,}066.$$

$$E_x = \frac{n\cdot(n+1)\cdot\Sigma\left(x_i - \bar{x}\right)^4}{(n-1)\cdot(n-2)\cdot(n-3)\cdot s_x^4} - \frac{3\cdot(n-1)^2}{(n-2)\cdot(n-3)} = \frac{20\cdot(20+1)\cdot 4776778{,}43}{(20-1)\cdot(20-2)\cdot(20-3)\cdot 19{,}24^4} -$$

$$- \frac{3\cdot(20-1)^2}{(20-2)\cdot(20-3)} = 2{,}518 - 3{,}539 = -1{,}021.$$

$$A_y = \frac{n \cdot \Sigma\left(y_i - \bar{y}\right)^3}{(n-1)\cdot(n-2)\cdot s_y^3} = \frac{20 \cdot 696}{(20-1)\cdot(20-2)\cdot 5{,}47^3} = 0{,}248.$$

$$E_y = \frac{n\cdot(n+1)\cdot\Sigma\left(y_i - \bar{y}\right)^4}{(n-1)\cdot(n-2)\cdot(n-3)\cdot s_y^4} - \frac{3\cdot(n-1)^2}{(n-2)\cdot(n-3)} = \frac{20\cdot(20+1)\cdot 50227{,}25}{(20-1)\cdot(20-2)\cdot(20-3)\cdot 5{,}47^4} -$$

$$- \frac{3\cdot(20-1)^2}{(20-2)\cdot(20-3)} = 4{,}053 - 3{,}539 = 0{,}514.$$

Визначаємо стандартні помилки вимірювання асиметрії та ексцесу для $n = 20$.

$$m_A = \sqrt{\frac{6}{n+3}} = \sqrt{\frac{6}{20+3}} = 0{,}511;$$

$$m_E = 2\cdot\sqrt{\frac{6}{n+3}} = 2\cdot\sqrt{\frac{6}{20+3}} = 1{,}022.$$

Згідно з критерієм М.А. Плохинського показники асиметрії й ексцесу свідчать про статистичну відмінність емпіричного розподілу від нормального в тому випадку, якщо вони перевищують за абсолютною величиною свої стандартні помилки вимірювання в 3 рази.

$$t_{A_x} = \frac{|A_x|}{m_A} = \frac{|0{,}066|}{0{,}511} = 0{,}13;\ t_{E_x} = \frac{|E_x|}{m_E} = \frac{|\text{-}1{,}021|}{1{,}022} = 1.$$

$$t_{A_y} = \frac{|A_y|}{m_A} = \frac{|0{,}248|}{0{,}511} = 0{,}49;\ t_{E_y} = \frac{|E_y|}{m_E} = \frac{|0{,}514|}{1{,}022} = 0{,}5.$$

За результатами проведених розрахунків встановлено, що значення асиметрії та ексцесу для обох змінних не перевищують утричі відповідні стандартні помилки їх вимірювання. Це дає підстави вважати, що розподіл досліджуваних даних відповідає закону нормального розподілу. Відповідно,

для оцінки сили та напряму зв'язку між змінними доцільним є застосування параметричного коефіцієнта кореляції r_{xy} Пірсона.

2. Встановлюємо прийнятний рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: інтернет-залежність й тривожність особистості не корелюють ($r_{xy} = 0$);

H_1: інтернет-залежність й тривожність особистості корелюють ($r_{xy} \neq 0$).

3. Розраховуємо емпіричне значення коефіцієнта кореляції r_{xy} Пірсона:

$$r_{xy} = \frac{\Sigma\left(x_i - \bar{x}\right)\cdot\left(y_i - \bar{y}\right)}{\left(n-1\right)\cdot s_x \cdot s_y} = \frac{1727}{\left(20-1\right)\cdot 19{,}24 \cdot 5{,}47} = 0{,}863.$$

Отримане значення свідчить про пряму сильну кореляцію між змінними. Перевіримо її статистичну значущість.

4. За таблицею критичних значень коефіцієнта кореляції r_{xy} Пірсона (додаток 1) для $df = n = 20$ та заданого $\alpha = 0{,}05$ знаходимо $r_{крит.} = 0{,}444$..

Оскільки $|r_{емп.}| > r_{крит.}$ (0,863 > 0,444), гіпотеза H_0 відхиляється.

5. Оцінку розміру ефекту здійснюємо за показником коефіцієнта детермінації r^2.

$$r^2 = r_{xy}{}^2 = 0{,}863^2 = 0{,}745.$$

Розмір ефекту відповідає великому рівню згідно з інтерпретацією Дж. Коена. Інакше кажучи, спільна дисперсія змінних інтернет-залежності та тривожності становить 74,5 %.

Висновок. Коефіцієнт кореляції Пірсона дозволив відхилити нульову гіпотезу для заданого набору даних (r_{xy} = 0,863; $p < 0{,}05$; n = 20). Існує статистично значуща сильна позитивна кореляція між інтернет-залежністю й тривожністю особистості. Розміру ефекту відповідає великому рівню згідно інтерпретації Дж. Коена ($r^2 = 0{,}75$).

– відео-огляд процедури аналізу даних Прикладу 3.1. у статистичній програмі SPSS Statistics 23.0. та інтерпретація результатів дослідження.

3.3. Рангові коефіцієнти кореляції

На практиці дослідники проявів психіки переважно мають справу з неметричними вимірювальними шкалами. Якщо дані у дослідженні не відповідають закону нормального розподілу або представлені в порядковій шкалі, для вивчення зв'язку між ознаками застосовують коефіцієнти рангової

кореляції – r_s Спірмена або τ Кендалла. Їхні переваги – простота, широкі можливості та універсальність.

Принципової відмінності між цими критеріями немає, однак загальноприйнято вважати, що коефіцієнт Кендалла є більш «змістовним», оскільки детальніше аналізує взаємозв'язок між змінними, порівнюючи всі можливі відповідності між парами значень. Натомість коефіцієнт Спірмена точніше відображає кількісну ступінь зв'язку між змінними. За результатами обчислень для одного й того самого набору даних зазвичай $r_s > \tau$.

Значення коефіцієнтів рангової кореляції можуть змінюватися в межах від –1 до +1. Позитивний знак свідчить про прямий взаємозв'язок між змінними, від'ємний – про зворотний.

3.3.1. Коефіцієнт рангової кореляції r_s Спірмена

Вперше завдання обчислення коефіцієнта кореляції між двома змінними, представленими не власними значеннями, а рангами цих значень, було вирішено Френсісом Гальтоном наприкінці XIX століття. Проте такий підхід до обчислення коефіцієнта кореляції набув популярності лише після його використання Чарльзом Спірменом на початку XX століття. Саме ця обставина зумовила те, що коефіцієнт рангової кореляції отримав назву «коефіцієнт Спірмена».

Метод рангової кореляції Спірмена дає змогу визначити силу та напрямок зв'язку між двома ознаками або профілями (ієрархіями) ознак, представленими у порядкових шкалах, або у випадку, коли одна змінна вимірюється у порядковій шкалі, а інша – у кількісній (інтервальній чи шкалі відношень).

Для обчислення коефіцієнта рангової кореляції r_s Спірмена (інше позначення ρ) необхідно мати два ряди значень, які можна прорангувати. Такими рядами значень можуть бути:

1) дві ознаки однієї і тієї самої групи досліджуваних;

2) дві індивідуальні ієрархії ознак двох досліджуваних за одним і тим самим набором ознак;

3) дві групові ієрархії ознак;

4) індивідуальна та групова ієрархія ознак.

Емпіричне значення коефіцієнта рангової кореляції r_s Спірмена визначається за формулою:

$$r_s = 1 - \frac{6 \cdot \Sigma d_i^2}{n \cdot (n^2 - 1)}, \text{ де}$$

d – різниця рангів;

n – обсяг вибірки досліджуваних або кількість ознак.

У разі наявності зв'язаних рангів (однакових значень) в однієї або обох змінних загальна формула коефіцієнта рангової кореляції r_s Спірмена може давати дещо неточне значення. У такій ситуації необхідно здійснити поправку на зв'язані ранги. Скоригована формула має такий вигляд:

$$r_s = \frac{\Sigma x^2 + \Sigma y^2 - \Sigma d_i^2}{2 \cdot \sqrt{\Sigma x^2 \cdot \Sigma y^2}}$$, де

$$\Sigma x^2 = \frac{n^3 - n}{12} - T_x;\ \Sigma y^2 = \frac{n^3 - n}{12} - T_y;$$

$$T_x = \frac{\Sigma\left(a^3 - a\right)}{12};\ T_y = \frac{\Sigma\left(b^3 - b\right)}{12};$$

d – різниця рангів;

n – обсяг вибірки досліджуваних або кількість ознак.

a – обсяг кожної групи однакових рангів у першому ранговому ряду X;

b – обсяг кожної групи однакових рангів у другому ранговому ряду Y.

Якщо емпіричне значення r_s дорівнює або більше критичного значення для заданого α-рівня, то H_0 відхиляється і формулюється висновок, що коефіцієнт кореляції статистично значуще відрізняється від нуля.

Показником розміру ефекту виступає безпосередньо саме емпіричне значення коефіцієнта r_s. При інтерпретації необхідно орієнтуватися на табл. 3.2.

Приклад 3.2. Дослідницьке завдання полягає у з'ясуванні, чи існує статистично значущий взаємозв'язок між показниками значущості типів цінностей на рівні нормативних ідеалів у студентів хімічного та психологічного факультетів.

Таблиця 3.5 – Таблиця первинних даних та проміжних розрахунків

№	Цінності	Хіміки (X)	Психологи (Y)	*d*	d^2
1	Конформність	8	8	0	0
2	Традиції	9	10	-1	1
3	Доброта	7	4	3	9
4	Універсалізм	10	9	1	1
5	Самостійність	3	2	1	1
6	Стимуляція	4	6	-2	4
7	Гедонізм	2	3	-1	1
8	Досягнення	5	1	4	16
9	Влада	6	5	1	1
10	Безпека	1	7	-6	36
				Σ = 0	Σ = 70

Первинні дані за двома параметрами представлені в порядковій вимірювальній шкалі. Оскільки ранги не повторюються, емпіричне значення коефіцієнта рангової кореляції r_s Спірмена доцільно обчислювати за загальною формулою.

1. Встановлюємо рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: типи цінностей у студентів хімічного та психологічного факультетів не взаємозв'язані ($r_s = 0$);

H_1: типи цінностей у студентів хімічного та психологічного факультетів взаємозв'язані ($r_s \neq 0$).

2. Обчислюємо d – різницю рангів за формулою: $d_i = R_{Xi} - R_{Yi}$;

3. Визначаємо d^2 – квадрат різниці рангів.

4. Знаходимо суму d^2 за вибіркою досліджуваних.

5. Розраховуємо емпіричне значення коефіцієнта рангової кореляції Спірмена за формулою r_s:

$$r_s = 1 - \frac{6 \cdot \Sigma d_i^2}{n \cdot (n^2 - 1)} = 1 - \frac{6 \cdot 70}{10 \cdot \left(10^2 - 1\right)} = 0{,}576.$$

Отримане значення свідчить про пряму кореляцію середньої сили між змінними. Перевіримо її статистичну значущість.

6. За таблицею критичних значень коефіцієнта кореляції r_s Спірмена (додаток 1) для $df = n = 10$ та заданого $\alpha = 0{,}05$ знаходимо критичне значення $r_{крит.} = 0{,}632$..

Оскільки $|r_{емп.}| < r_{крит.}$ (0,576 < 0,632), гіпотеза H_0 приймається.

7. Оцінюємо розмір ефекту.

Розмір стандартизованого ефекту, згідно з класифікацією Дж. Коена, відповідає великому рівню ($r_s = 0{,}576$).

Висновок. Статистично значущого зв'язку між типами цінностей на рівні нормативних ідеалів у студентів хімічного та психологічного факультетів не виявлено. Згідно з результатами перевірки за критерієм рангової кореляції Спірмена ($r_s = 0{,}576$; $p > 0{,}05$; $n = 10$), підстав для відхилення нульової гіпотези немає. Водночас великий розмір ефекту виявленої кореляції свідчить про доцільність продовження дослідження із забезпеченням більшої статистичної потужності шляхом збільшення вибірки.

– відео-огляд процедури аналізу даних Прикладу 3.2. у статистичній програмі SPSS Statistics 23.0. та інтерпретація результатів дослідження.

Приклад 3.3. Дослідника цікавить питання: чи можна стверджувати, що різні види пам'яті статистично значуще взаємопов'язані між собою в структурі особистості? Дослідження проводилось на рандомно створеній вибірці обсягом 15 осіб.

Таблиця 3.6 – Таблиця первинних даних та проміжних розрахунків

№	Логічна пам’ять (X)	Механічна пам’ять (Y)	R_{Xi}	R_{Yi}	d	d^2
1	8	46	3,5	14	-10,5	110,25
2	5	34	1	2	-1	1
3	13	39	11,5	7	4,5	20,25
4	10	32	6	1	5	25
5	11	45	7,5	12,5	-5	25
6	8	36	3,5	4	-0,5	0,25
7	9	40	5	8	-3	9
8	13	38	11,5	6	5,5	30,25
9	13	41	11,5	9	2,5	6,25
10	7	42	2	10	-8	64
11	13	44	11,5	11	0,5	0,25
12	11	37	7,5	5	2,5	6,25
13	15	45	14	12,5	1,5	2,25
14	12	35	9	3	6	36
15	17	47	15	15	0	0
			$\Sigma_R = 120$	$\Sigma_R = 120$	$\Sigma = 0$	$\Sigma = 336$

Підстав стверджувати, що первинні дані за обома змінними відповідають закону нормального розподілу, немає. Крім того, обсяг вибірки є недостатнім для формування надійного висновку при перевірці даних на нормальність розподілу. У зв’язку з цим застосування параметричних методів аналізу є недоцільним. Для виявлення взаємозв’язку між досліджуваними змінними доцільно використовувати непараметричний коефіцієнт рангової кореляції r_s Спірмена, який не потребує виконання припущень щодо нормальності розподілу.

1. Встановлюємо прийнятний рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: логічна та механічна пам’ять особистості не взаємозв’язані між собою ($r_s = 0$);

H_1: логічна та механічна пам’ять особистості взаємозв’язані між собою ($r_s \neq 0$).

1. Прорангуємо обидва показники від найменшого до найбільшого. Для перевірки правильності ранжування можна використовувати той факт, що $\Sigma_R = n\cdot(n+1)/2$. У нашому випадку сума рангів рівна 120 ($\Sigma_R = 120$), $n\cdot(n+1)/2 = 15\cdot(15+1)/2 = 120$. Тотожність підтвердилась (120 =120).

2. Обчислюємо d – різницю рангів за формулою: $d_i = R_{Xi} - R_{Yi}$;
3. Визначаємо d^2 – квадрат різниці рангів.
4. Знаходимо суму d^2 за вибіркою досліджуваних.

5. Будемо розраховувати емпіричне значення коефіцієнта рангової кореляції r_s Спірмена за спеціальною формулою через існування зв'язаних рангів. Визначаємо поправки на однакові ранги.

$$T_x = \frac{\Sigma\left(a^3 - a\right)}{12} = \frac{(2^3 - 2) + (2^3 - 2) + (4^3 - 4)}{12} = 6;$$

$$T_y = \frac{\Sigma\left(b^3 - b\right)}{12} = \frac{\left(2^2 - 2\right)}{12} = 0{,}5;$$

$$\Sigma x^2 = \frac{n^3 - n}{12} - T_x = \frac{15^3 - 15}{12} - 6 = 274;$$

$$\Sigma y^2 = \frac{n^3 - n}{12} - T_y = \frac{15^3 - 15}{12} - 0{,}5 = 279{,}5;$$

6. Розраховуємо емпіричне значення коефіцієнта рангової кореляції r_s Спірмена:

$$r_s = \frac{\Sigma x^2 + \Sigma y^2 - \Sigma d_i^2}{2 \cdot \sqrt{\Sigma x^2 \cdot \Sigma y^2}} = \frac{274 + 279{,}5 - 336}{2 \cdot \sqrt{274 \cdot 279{,}5}} = 0{,}393.$$

Отримане значення свідчить про пряму кореляцію помірної сили між змінними. Перевіримо її статистичну значущість.

8. За таблицею критичних значень коефіцієнта кореляції r_s Спірмена (додаток 1) для $df = n = 15$ та заданого $\alpha = 0{,}05$ знаходимо $r_{крит.} = 0{,}514$..

Оскільки $|r_{емп.}| < r_{крит.}$ (0,393 < 0,514), гіпотеза H_0 приймається.

9. Оцінюємо розмір ефекту.

Розмір стандартизованого ефекту згідно класифікації Дж. Коена – середній ($r_s = 0{,}393$).

Висновок. Застосування критерію r_s Спірмена не дозволило виявити статистично значущого зв'язку між логічною та механічною пам'яттю ($r_s = 0{,}393$; $p > 0{,}05$; $n = 15$), що свідчить про прийняття нульової гіпотези. Водночас, встановлений середній розмір ефекту, згідно з критеріями інтерпретації Дж. Коена, вказує на доцільність підвищення статистичної потужності дослідження.

3.3.2. Коефіцієнт рангової кореляції τ Кендалла

У середині 1940-х років Моріс Кендалл запропонував альтернативний підхід до оцінювання кореляційного зв'язку між ранговими змінними.

Суть методу Кендалла полягає в наступному. Припустимо, що групу людей впорядковано за зростом – від найвищого до найнижчого – та відомі показники їхньої ваги. Таким чином, отримано два впорядковані ряди: один – за зростом, інший – за вагою. Якщо між цими змінними існує позитивна залежність (тобто зі збільшенням зросту зростає і вага), тоді порядок рангів в обох рядах буде узгодженим: першому рангу за зростом відповідатиме перший ранг за вагою, другому – другий і так далі.

Однак в емпіричних даних така узгодженість часто порушується: можливі ситуації, коли висока людина має низьку вагу, а людина невеликого зросту – високу. У таких випадках спостерігається інверсія (перестановка рангів), тобто зміна напрямку рангових значень.

Якщо ранги в обох рядах повністю узгоджені, коефіцієнт рангової кореляції Кендалла дорівнює +1. У разі повної інверсії (усі ранги змінюються в протилежному напрямі) значення коефіцієнта дорівнює –1. У більшості практичних випадків серед пар спостережень наявні як узгодження, так і інверсії.

Для кількісного вимірювання ступеня відповідності рангів М. Кендалл запропонував коефіцієнт τ (тау), який розраховується на основі кількості пар, що збігаються, та кількості пар з інверсіями.

Емпіричне значення коефіцієнта τ Кендалла визначається за формулою:

$$\tau = \frac{P - Q}{n \cdot (n-1)/2}, \text{ де}$$

P – кількість збігів (погоджені пари);

Q – кількість інверсій (непогоджені пари);

n – обсяг вибірки досліджуваних або кількість ознак.

Існує кілька модифікацій формули для обчислення коефіцієнта рангової кореляції τ Кендалла, у яких використовуються лише значення кількості узгоджених пар (збігів), лише кількість неузгоджених пар (інверсій), або ж – як у нашому випадку – і ті, й інші. Незважаючи на відмінності у підходах до розрахунку, усі ці формули приводять до одного й того самого результату.

У разі наявності зв’язаних рангів (тобто однакових значень рангів) в одній або обох змінних, базове значення коефіцієнта τ Кендалла може бути неточним. Це зумовлено тим, що класична формула не враховує порушення припущення про унікальність рангів, що може спричинити зміщення результатів.

У таких випадках застосовується корекція (поправка) на зв’язані ранги, яка дозволяє компенсувати викривлення, зумовлене однаковими значеннями. Внаслідок цього модифікована формула набуває такого вигляду:

$$\tau = \frac{P - Q}{\left(\sqrt{\frac{n \cdot (n-1)}{2} - T_x}\right) \cdot \left(\sqrt{\frac{n \cdot (n-1)}{2} - T_y}\right)}, \text{ де}$$

$$T_x = \frac{\Sigma(a_x - 1)}{2};\ T_y = \frac{\Sigma(b_y - 1)}{2}$$

P – кількість збігів (погоджені пари);

Q – кількість інверсій (непогоджені пари);

n – обсяг вибірки досліджуваних або кількість ознак.

a_x – обсяг кожної групи однакових рангів у першому ранговому ряду X;

b_y – обсяг кожної групи однакових рангів у другому ранговому ряду Y.

Коефіцієнт рангової кореляції τ Кендалла не має окремої таблиці критичних значень. Оскільки його статистика наближено підпорядковується розподілу Стьюдента, для перевірки статистичної значущості отриманого значення використовується таблиця критичних значень t-критерію Стьюдента (додаток 3). Обчислюється емпіричне значення t-статистики за формулою:

$$t = |\tau| \cdot \sqrt{\frac{n-2}{1-\tau^2}}$$, де

τ – емпіричне значення коефіцієнта кореляції τ Кендалла;
n – обсяг вибірки досліджуваних.

У разі, якщо емпіричне значення t-статистики дорівнює або перевищує критичне значення для заданого рівня значущості α, нульова гіпотеза (H_0) відхиляється, а натомість приймається альтернативна гіпотеза (H_1). Це дає підстави зробити висновок про те, що коефіцієнт рангової кореляції Кендалла є статистично значущим, тобто його значення відрізняється від нуля на обраному рівні α.

Показником розміру ефекту у даному випадку слугує саме емпіричне значення коефіцієнта τ Кендалла, яке відображає силу зв'язку між двома ранговими змінними. Для інтерпретації величини цього коефіцієнта доцільно орієнтуватися на методичні рекомендації, запропоновані Джейкобом Коеном (див. таблицю 3.2), які дозволяють класифікувати ефект як несуттєвий, малий, середній або великий залежно від значення τ.

При інтерпретації коефіцієнта кореляції τ Кендалла можна врахувати, що для двох рангових рядів ймовірність збігів дорівнює (1 + τ)/2, а ймовірність інверсій дорівнює (1 – τ)/2. Наприклад, якщо взаємозв'язок оцінок жіночої привабливості двох здобувачів вищої освіти дорівнює τ = 0,6, то це означає 80 % збігів і 20 % інверсій. Іншими словами, думки студентів щодо жіночої привабливості співпадають вчетверо частіше ніж розходяться.

Інтерпретація коефіцієнтів кореляції Пірсона та Спірмена не є однозначною у контексті визначення сили зв'язку між змінними. На відміну від τ Кендалла, де саме значення коефіцієнта може безпосередньо трактуватися як показник розміру ефекту, у випадку кореляцій Пірсона та Спірмена доцільно здійснювати додаткове перетворення.

Зокрема, отримане значення коефіцієнта кореляції рекомендується звести до квадрату, що дозволяє інтерпретувати результат як частку дисперсії однієї змінної, яка пояснюється варіацією іншої змінної в межах досліджуваної вибірки.

Приклад 3.4. Необхідно встановити, чи існує статистично значущий взаємозв'язок між відповідальністю особистості та успішністю в професійній діяльності.

Таблиця 3.7 – Таблиця первинних даних та проміжних розрахунків

№	Відповідальність особистості (X)	Професійна успішність (Y)	Відповідальність особистості ($X_{ранг}$)	Професійна успішність ($Y_{ранг}$)
1	40	37	7	8
2	49	42	12	11
3	44	25	9	5
4	42	40	8	10
5	24	19	1	3
6	48	39	11	9
7	36	27	5	6
8	25	14	2	1
9	45	43	10	12
10	28	16	3	2
11	31	20	4	4
12	39	35	6	7
			$\Sigma_R = 78$	$\Sigma_R = 78$

Відсутні підстави вважати, що первинні дані за обома змінними представлені в інтервальній шкалі вимірювання та відповідають закону нормального розподілу. Крім того, розмір вибірки є недостатнім для надійної перевірки нормальності розподілу обох змінних. Враховуючи це, припускаємо, що досліджувані змінні мають порядкову шкалу вимірювання. З огляду на зазначене, для виявлення взаємозв'язку між змінними доцільно застосувати непараметричний коефіцієнт рангової кореляції τ Кендалла.

1. Встановлюємо прийнятний рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: відповідальність особистості та професійна успішність не взаємозв'язані ($\tau = 0$);

H_1: відповідальність особистості та професійна успішність взаємозв'язані ($\tau \neq 0$).

2. Прорангуємо обидва показники за зростанням – від найменшого до найбільшого значення. Для перевірки правильності ранжування можна використовувати той факт, що $\Sigma_R = n \cdot (n+1)/2$. У нашому випадку сума рангів рівна 78 ($\Sigma_R = 78$), $n \cdot (n+1)/2 = 12 \cdot (12+1)/2 = 78$. Тотожність підтвердилась (78 =78).

3. Для обчислення коефіцієнта рангової кореляції τ Кендалла необхідно впорядкувати значення другої змінної (відповідальність особистості) за зростанням рангів. Відповідно до цього зміниться порядок рангових значень показника професійної успішності (див. табл. 3.10).

4. Підраховуємо кількість збігів та інверсій.

Підрахунок кількості збігів (*P*) здійснюється в такий спосіб. Беремо саме верхнє число третього стовпчика – 3. Рахуємо скільки всього чисел більших за 3 трапляються нижче в цьому ж стовпці. Їх – 9 (4, 5, 6, 7, 8, 9, 10, 11, 12). Число дев'ять записуємо навпроти 5 досліджуваного (див. № досліджуваного) у колонці *P*. Потім аналізуємо наступне число з третього стовпчика. Це – одиниця, більше неї нижче за стовпцем розташовані 10 чисел. Навпроти 8 досліджуваного записуємо 10. І так далі. $\Sigma_P = 57$.

Таблиця 3.8 – Таблиця проміжних розрахунків

№	Відповідальність особистості ($X_{ранг}$)	Професійна успішність ($Y_{ранг}$)	*P*	*Q*
5	1	3	9	2
8	2	1	10	0
10	3	2	9	0
11	4	4	8	0
7	5	6	6	1
12	6	7	5	1
1	7	8	4	1
4	8	10	2	2
3	9	5	3	0
9	10	12	0	2
6	11	9	1	0
2	12	11	0	0
			$\Sigma_P = 57$	$\Sigma_Q = 9$

Для визначення кількості інверсій знову беремо саме верхнє число третього стовпчика – 3. Підраховуємо скільки усього чисел менших за 3 трапляються нижче в цьому ж стовпці. Це числа 2 і 1, тобто їх два. Двійку записуємо навпроти 5 досліджуваного в колонці *Q*. Потім аналізуємо наступне число з третього стовпчика. Це – одиниця, менше неї нижче за стовпцем немає жодного числа. Навпроти досліджуваного під № 8 записуємо 0. І так далі. $\Sigma_Q = 9$.

Для перевірки правильності підрахунку кількості збігів та інверсій можна використати формулу:

$$\Sigma_P + \Sigma_Q = \frac{(n-1)\cdot n}{2};$$

У нашому випадку $\Sigma_P + \Sigma_Q = 66$; $\frac{(n-1)\cdot n}{2} = \frac{(12-1)\cdot 12}{2} = 66$;

Тотожність підтвердилась (66 = 66).

5. Розраховуємо емпіричне значення за формулою кореляційного критерія τ Кендалла.

$$\tau = \frac{P - Q}{n \cdot (n-1)/2} = \frac{57 - 9}{12 \cdot (12-1)/2} = 0{,}727 .$$

Отримане значення свідчить про сильну пряму кореляцію між змінними. Перевіримо її статистичну значущість.

6. Пошук критичних значень здійснюється за таблицею критичних значень критерію *t*-Стьюдента. Для обчислення емпіричного значення використовується формула:

$$t = |\tau| \cdot \sqrt{\frac{n-2}{1-\tau^2}} = |0{,}727| \cdot \sqrt{\frac{12-2}{1-0{,}727^2}} = 3{,}348.$$

7. За таблицею критичних значень критерію *t*-Стьюдента (додаток 3) для $df = n - 2 = 12 - 2 = 10$ та заданого $\alpha = 0{,}05$ знаходимо $t_{крит.} = 2{,}23$..

Оскільки $t_{емп.} > t_{крит.}$ (3,348 > 2,23), гіпотеза H_0 відхиляється.

8. Оцінюємо розмір ефекту.

Розмір стандартизованого ефекту згідно Дж. Коена – великий ($\tau = 0{,}727$).

Висновок. Коефіцієнт кореляції Кендалла дозволив відхилити нульову гіпотезу ($\tau = 0{,}727$; $p < 0{,}05$; $n = 12$). Виявлено статистично значущий сильний позитивний кореляційний зв'язок між відповідальністю особистості та професійною успішністю. Розмір ефекту відповідає великому рівню відповідно до інтерпретації Дж. Коена.

– відео-огляд процедури аналізу даних Прикладу 3.4. у статистичній програмі SPSS Statistics 23.0. та інтерпретація результатів дослідження.

3.4. Коефіцієнти асоціативної залежності номінальних даних

У психологічних дослідженнях досить часто реєструються та аналізуються характеристики, що не мають безпосереднього кількісного виміру. Такі дані подаються в номінальній (номінативній, або шкалі найменувань) вимірювальній шкалі. У межах цієї шкали здійснюється класифікація об'єктів, а належність до певного класу позначається номером чи іншим умовним маркером, який не відображає жодних властивостей самого об'єкта, окрім його групової приналежності.

У випадках роботи з номінативними даними для оцінювання статистичної залежності між змінними застосовуються коефіцієнти, розроблені на основі критерію χ^2-Пірсона. Значення коефіцієнта, що дорівнює 0, свідчить про повну незалежність змінних, а значення, наближене до 1, – про їх максимальну залежність. При цьому знак коефіцієнта не

інтерпретується, оскільки він залежить від умовного кодування (позначення) градацій ознак і не відображає напрямку залежності.

3.4.1. Коефіцієнт контингенції Пірсона *φ*

Коефіцієнт контингенції Карла Пірсона φ використовується для кількісної оцінки асоціації між якісними характеристиками об'єктів. Його застосування є доцільним у випадках аналізу змінних, виміряних у дихотомічній номінативній шкалі, тобто коли ознака може набувати лише двох значень. Значення коефіцієнта φ варіює в межах від –1 до +1. При цьому його крайні значення вказують на максимальний рівень залежності між змінними, тоді як значення, наближене до нуля, – на їх незалежність.

Існує декілька способів обчислення коефіцієнта контингенції Пірсона *φ*. Найбільш поширеним і водночас простим є розрахунок на основі таблиці зв'язку ознак (таблиці крос-табуляції), що дає змогу відобразити співвідношення частот для комбінацій значень аналізованих змінних (табл. 3.9).

Таблиця 3.9 – Таблиця зв'язку ознак розміром 2×2

Ознаки		Змінна X		Усього
		1	0	
Змінна Y	1	a	b	a+b
	0	c	d	c+d
Усього		a+c	b+d	n

У такому випадку розрахункове значення коефіцієнта контингенції обчислюється за формулою:

$$\varphi = \frac{a \cdot d - b \cdot c}{\sqrt{(a+c)\cdot(b+d)\cdot(a+b)\cdot(c+d)}}$$

Коефіцієнт контингенції не має власної таблиці критичних значень. Статистична значущість встановленої залежності визначається за допомогою критерію χ^2-Пірсона..

$\chi^2 = \varphi^2 \cdot n$, де

φ – емпіричне значення коефіцієнта контингенції;

n – обсяг вибірки досліджуваних.

Для прийняття рішення про рівень статистичної значущості необхідно емпіричне (розрахункове) значення критерію χ^2-Пірсона порівняти з критичним (табличним) значенням. Критичне значення визначається залежно від числа ступенів свободи та заданого α-рівня. Зважаючи на те, що для таблиці розміром 2×2 число ступенів свободи завжди однакове та дорівнює 1 (df = 1), то ці значення завжди сталі (додаток 2): $\chi^2_{0,05.}$ = 3,841; $\chi^2_{0,01.}$= 6,635; $\chi^2_{0,001.}$= 10,829.

Якщо $\chi^2_{емп.} \geq \chi^2_{крит.}$, то залежність між ознаками статистично значуща, тобто ознаки змінюються узгоджено.

Якщо $\chi^2_{емп.} < \chi^2_{крит.}$, то статистично значущої залежності між ознаками немає.

Показником розміру ефекту в даному випадку виступає емпіричне значення коефіцієнта контингенції φ. Для інтерпретації отриманих результатів доцільно використовувати орієнтовні критерії, запропоновані Дж. Коеном (див. табл. 3.2).

Приклад 3.5. Необхідно перевірити наявність статистично значущого зв'язку між статтю респондентів та рівнем депресивності особистості. Первинні результати дослідження наведено в таблиці 3.10.

Ознака X (стать): 1 – жінка; 2 – чоловік.

Ознака У (депресія): 1 – є ознаки; 2 – відсутні ознаки.

Таблиця 3.10 – Таблиця первинних даних

№п/п	Стать	Депресія	№п/п	Стать	Депресія
1	2	2	21	1	1
2	1	1	22	2	2
3	2	1	23	1	1
4	1	1	24	2	2
5	1	2	25	2	2
6	1	2	26	2	2
7	2	1	27	2	2
8	1	1	28	1	2
9	1	2	29	2	2
10	1	1	30	1	2
11	1	2	31	2	2
12	1	2	32	1	1
13	1	2	33	1	1
14	2	2	34	2	2
15	2	1	35	2	2
16	1	2	36	1	2
17	1	2	37	2	2
18	2	1	38	2	2
19	1	2	39	2	2
20	1	1	40	1	2

Оскільки обидві ознаки представлені в дихотомічній номінальній шкалі, то для перевірки наявності залежності між ними можна використати коефіцієнт контингенції Пірсона φ.

1. Встановлюємо рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: між статтю та депресивністю особистості відсутня залежність ($\varphi = 0$);

H_1: між статтю та депресивністю особистості існує статистично значуща залежність ($\varphi \neq 0$).

1. За даними таблиці 3.10. сформуємо таблицю зв'язку ознак і підрахуємо абсолютні частоти всіх можливих поєднань значень (табл. 3.11.).

Таблиця 3.11 – Таблиця зв'язку ознак

Ознаки		Депресія		Усього
		Є ознаки	Відсутні ознаки	
Стать	Жінки	9	13	22
	Чоловіки	4	14	18
Усього		13	27	40

2. У даній таблиці: $a = 9$, $b = 13$, $c = 4$, $d = 14$.

Підставами ці значення у формулу коефіцієнта контингенції φ:

$$\varphi = \frac{9 \cdot 14 - 13 \cdot 4}{\sqrt{(9+4)\cdot(13+14)\cdot(9+13)\cdot(4+14)}} = 0{,}198.$$

Отримане значення свідчить про залежність слабкої сили між ознаками. Перевіримо її статистичну значущість.

4. Для оцінки статистичної значущості залежності необхідно розрахувати емпіричне значення критерію χ^2-Пірсона.

$$\chi^2 = \varphi^2 \cdot n = (0{,}198)^2 \cdot 40 = 1{,}57.$$

За таблицею критичних значень критерію χ^2-Пірсона (додаток 2) для $df = 1$ та заданого $\alpha = 0{,}05$ знаходимо $\chi^2_{крит.} = 3{,}841$..

Оскільки $\chi^2_{емп.} < \chi^2_{крит.}$ (1,57 < 3,841), гіпотеза H_0 приймається.

5. Оцінюємо розмір ефекту.

Показником розміру ефекту є безпосередньо саме емпіричне значення коефіцієнта φ. Згідно Джейкоба Коена 0,198 відповідає малому рівню.

Для таблиці розміром 2×2 можна визначити ще один показник величини розміру ефекту – відношення шансів (odds ratio – OR).

$$OR = \frac{a \cdot d}{b \cdot c} = \frac{9 \cdot 14}{13 \cdot 4} = 2{,}42.$$

Іншими словами, ймовірність виявлення ознак депресивності у жінок є у 2,4 раза вищою порівняно з чоловіками.

Висновок. Емпіричні дані не дають підстав для відхилення нульової гіпотези за результатами розрахунку коефіцієнта контингенції (φ = 0,198; $p > 0{,}05$; n = 40). Отже, статистично значущої асоціативної залежності між статтю та депресивністю особистості не виявлено. Водночас розмір ефекту відповідає малому рівню згідно з інтерпретацією Дж. Коена. Варто

зазначити, що ймовірність наявності ознак депресивності у жінок є у 2,4 раза вищою, ніж у чоловіків.

– відео-огляд процедури аналізу даних Прикладу 3.5. у статистичній програмі SPSS Statistics 23.0. та інтерпретація результатів дослідження.

3.4.2. Коефіцієнт асоціації Юла *Q*

Коефіцієнт асоціації Юла Q застосовується для виявлення залежності між змінними, виміряними у номінальній дихотомічній шкалі. Значення цього коефіцієнта можуть змінюватися в інтервалі від –1 до +1. Розрахунок здійснюється за формулою:

$$Q = \frac{ad - bc}{ad + bc}$$, де

a, b, c, d – дані, що визначаються згідно з таблицею зв'язку ознак розміром 2×2 (див. табл. 3.9.).

Коефіцієнт асоціації як і коефіцієнт контингенції немає своєї таблиці з критичними значеннями. Статистична значущість залежності оцінюється за допомогою критерію χ^2-Пірсона.

$$\chi^2 = Q^2 \cdot n$$, де

Q – емпіричне значення коефіцієнта асоціації;
n – обсяг вибірки досліджуваних.

Для прийняття рішення про рівень статистичної значущості необхідно емпіричне (розрахункове) значення критерію χ-Пірсона.[2] порівняти з критичним (табличним) значенням. Критичне значення визначається залежно від числа ступенів свободи та заданого α-рівня. Зважаючи на те, що для таблиці розміром 2×2 число ступенів свободи завжди дорівнює 1 (*df* = 1), то ці значення завжди сталі (додаток 2): $\chi^2_{0,05.} = 3{,}841$; $\chi^2_{0,01.} = 6{,}635$; $\chi^2_{0,001.} = 10{,}829$.

Якщо $\chi^2_{емп.} \geq \chi^2_{крит.}$, то залежність між ознаками статистично значуща, тобто ознаки змінюються узгоджено.

Якщо $\chi^2_{емп.} < \chi^2_{крит.}$, то статистично значущої залежності між ознаками немає.

Показником розміру ефекту виступає безпосередньо саме емпіричне значення коефіцієнта контингенції *Q*. При інтерпретації необхідно орієнтуватися на рекомендації запропоновані Дж. Коеном (див. таблицю 3.2.).

Незважаючи на те, що коефіцієнт асоціації Юла та коефіцієнт контингенції Пірсона застосовуються для вирішення подібних завдань,

кожен із них має власні особливості. Коефіцієнт контингенції відображає двобічну залежність між змінними, тоді як коефіцієнт асоціації Юла характеризує однобічну залежність. Унаслідок цього значення коефіцієнта асоціації зазвичай є більшим, ніж коефіцієнта контингенції, якщо вони обчислені за одними й тими самими даними. При цьому коефіцієнт контингенції забезпечує більш обережну оцінку ступеня тісноти зв'язку.

Варто враховувати, що якщо хоча б одне з чотирьох значень у таблиці крос-табуляції 2×2 відсутнє, коефіцієнт асоціації Юла набуває значення, рівного одиниці, що призводить до перебільшеної оцінки сили залежності між ознаками. У таких випадках доцільніше використовувати коефіцієнт контингенції.

Приклад 3.6. Необхідно оцінити вплив участі студентів у заняттях науково-дослідної лабораторії на рівень їхньої впевненості під час захисту кваліфікаційних робіт. Для цього дослідник використовує дані, зібрані в ході дослідження 320 студентів, серед яких 240 є активними учасниками занять науково-дослідної лабораторії.

Таблиця 3.12 – Таблиця зв'язку ознак

Ознаки		Впевненість під час захисту кваліфікаційних робіт		Усього
		Впевнені	Невпевнені	
Групи студентів	Займаються в науково-дослідній лабораторії	163	77	240
	Не займаються в науково-дослідній лабораторії	46	34	80
Усього		209	111	320

1. Встановлюємо рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: впевненість під час захисту кваліфікаційних робіт не залежить від занять у науково-дослідній лабораторії ($Q = 0$);

H_1: впевненість під час захисту кваліфікаційних робіт залежить від занять у науково-дослідній лабораторії ($Q \neq 0$).

2. У даній таблиці: $a = 163$, $b = 77$, $c = 46$, $d = 34$.

Підставами ці значення у формулу коефіцієнта асоціації Q.

$$Q = \frac{ad - bc}{ad + bc} = \frac{163 \cdot 34 - 77 \cdot 46}{163 \cdot 34 + 77 \cdot 34} = 0{,}215.$$

Отримане значення свідчить про залежність слабкої сили між ознаками. Перевіримо її статистичну значущість.

2. Для оцінки статистичної значущості залежності необхідно знайти емпіричне значення критерію χ^2-Пірсона.

$$\chi^2 = Q^2 \cdot n = 0{,}215^2 \cdot 320 = 14{,}792.$$

За таблицею критичних значень критерію χ^2-Пірсона (додаток 2) для $df = 1$ та заданого $\alpha = 0{,}05$ знаходимо $\chi^2_{крит.} = 3{,}841$..

Оскільки $\chi^2_{емп.} > \chi^2_{крит.}$ (14,792 > 3,841), гіпотеза H_0 відхиляється.

5. Оцінюємо розмір ефекту.

Показником розміру ефекту є безпосередньо саме емпіричне значення коефіцієнта *Q*. Згідно Дж. Коена 0,215 відповідає малому рівню.

Висновок. Коефіцієнт асоціації *Q* дозволив відхилити нульову гіпотезу для досліджуваного набору даних ($Q = 0{,}215$; $p < 0{,}05$; $n = 320$). Виявлено статистично значущу асоціацію слабкої сили між участю студентів у науково-дослідній лабораторії та рівнем їхньої впевненості під час захисту кваліфікаційних робіт. Водночас невеликий розмір ефекту свідчить про обмежену практичну значущість виявленої залежності.

3.4.3. Коефіцієнти взаємної зв'язаності ознак Пірсона *C*, Чупрова *K* та Крамера *V*

Коефіцієнти взаємної зв'язаності ознак Пірсона *C*, Чупрова *K* та Крамера *V* застосовуються для оцінки залежності у випадках, коли кожна змінна представлена номінальною шкалою, і при цьому хоча б одна зі змінних має більше двох градацій. Інакше кажучи, таблиця зв'язку ознак повинна мати розмірність більшу за 2×2, на відміну від коефіцієнтів асоціації та контингенції, які застосовуються для двовимірних таблиць.

Зазначені критерії базуються на безпосередньому розрахунку статистики χ^2-Пірсона. Слід зауважити, що зі зростанням значення χ^2 підвищується сила залежності, яку характеризують коефіцієнти взаємної зв'язаності.

Розрахункова формула χ^2-Пірсона має наступний вигляд:

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(f_{емп.} - f_{теор.}\right)^2}{f_{теор.}}, \text{ де}$$

k – кількість класових інтервалів (градацій) досліджуваної ознаки;

$f_{емп.}$ та $f_{теор.}$ – відповідно емпіричні та теоретичні (очікувані) частоти, що відповідають визначеним градаціям змінної.

Для розрахунку теоретичної (очікуваної) частоти кожної комірки таблиці необхідно враховувати наступне:

$$f_{ij} = \frac{f_i \cdot f_j}{n}, \text{ де}$$

f_i – сума частот у всіх комірках *i*-рядка;

f_j – сума частот у всіх комірках *j*- стовпчика;

n – сума частот у всій таблиці зв'язаності ознак.

Значення коефіцієнтів Пірсона *C*, Чупрова *K* та Крамера *V* варіюються в межах від 0 до +1, проте кожен із цих критеріїв має свої специфічні особливості.

Значення коефіцієнта Пірсона *C* при повній відсутності асоціації між змінними дорівнює нулю. Проте в умовах абсолютної залежності він не набуває значення одиниці, а лише прагне до нього зі збільшенням кількості градацій ознаки. Максимальне значення коефіцієнта *C* досягається за умови рівності кількості рядків і стовпчиків у таблиці зв'язку ознак ($k = m$, де k і m – відповідно кількість рядків і стовпчиків). Водночас максимальне значення коефіцієнта *C* змінюється залежно від кількості категорій: при k (m) = 3 значення *C* не перевищує 0,8; при k (m) = 5 – максимальне значення становить 0,89 і так далі. Це ускладнює порівняння результатів, отриманих на основі таблиць різного розміру. Тому застосування коефіцієнта Пірсона *C* рекомендується лише у випадках, коли таблиця зв'язку має однакову кількість рядків і стовпчиків ($k = m$).

Для подолання вказаного недоліку коефіцієнта Пірсона Олександр Чупров запропонував альтернативний показник – коефіцієнт Чупрова *K*. Максимальне значення коефіцієнта Чупрова дорівнює одиниці лише за умови рівності кількості градацій досліджуваних ознак ($k = m$). У випадках, коли $k \neq m$, цей коефіцієнт не може досягати одиниці.

З-поміж перелічених критеріїв лише коефіцієнт кореляції Крамера *V* може досягати одиниці незалежно від розміру таблиці зв'язку. Для квадратних таблиць зв'язку ($k = m$) коефіцієнти Чупрова *K* та Крамера *V* співпадають, тоді як у всіх інших випадках $V < K$. При цьому коефіцієнт Чупрова *K* завжди менший за коефіцієнт Пірсона *C*.

Коефіцієнти взаємної зв'язаності Пірсона *C*, Чупрова *K* та Крамера *V* визначаються за такими формулами:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}};$$

$$K = \sqrt{\frac{\chi^2}{n \cdot \sqrt{(k-1)\cdot(m-1)}}};$$

$$V = \sqrt{\frac{\chi^2}{n \cdot (c-1)}}, \text{ де}$$

χ^2 – емпіричне значення критерію χ^2-Пірсона (у всіх формулах);
n – обсяг вибірки досліджуваних (у всіх формулах);
k, m – кількість рядків і стовпчиків у таблиці зв'язку ознак (Чупрова *K*);
c – найменша кількість градацій із двох змінних (Крамера *V*).

Обмеження на використання зазначених коефіцієнтів відповідають обмеженням критерію χ^2 Пірсона, а саме: обсяг вибірки повинен становити не менше ніж $n \geq 30$; не більше 20 % теоретичних (очікуваних) частот можуть бути меншими за 5; при цьому жодна теоретична частота не повинна бути меншою за 1.

Таблиць із критичними значеннями для коефіцієнтів зв'язаності немає. Тому для прийняття статистичного рішення необхідно:

1. Вибрати рівень значущості та сформулювати нульову і альтернативну гіпотези.

2. Обчислити емпіричне значення критерію χ^2-Пірсона.

3. Розрахувати значення коефіцієнта взаємної зв'язаності.

4. Порівняти емпіричне значення критерію χ^2 з критичним значенням (додаток 2) для відповідного числа ступенів свободи.

$df = (k-1)\cdot(m-1)$, де k, m – кількість рядків і стовпчиків таблиці зв'язку ознак.

5. Якщо $\chi^2_{емп.} \geq \chi^2_{крит.}$, то залежність між ознаками статистично значуща, тобто ознаки змінюються узгоджено.

Якщо $\chi^2_{емп.} < \chi^2_{крит.}$, то формулюється висновок про відсутність статистично значущої залежності.

Показником розміру ефекту виступає безпосередньо саме емпіричне значення коефіцієнтів взаємної зв'язаності ознак. При інтерпретації необхідно орієнтуватися на рекомендації запропоновані Дж. Коеном (див. таблицю 3.2.).

Приклад 3.7. Необхідно визначити, чи існує статистично значуща залежність між рівнями розвитку загальних пізнавальних здібностей (ЗПЗ) та показниками професійної успішності.

Таблиця 3.13 – Таблиця зв'язку ознак

Параметри		Рівень професійної успішності		
		Високий	Середній	Низький
Рівень ЗПЗ	Високий	20	15	6
	Середній	12	13	8
	Низький	5	10	13

Первинні дані представлені в номінальній шкалі. Таблиця зв'язку ознак має розмір 3×3. У такому випадку для визначення асоціативної залежності доцільно застосувати коефіцієнти взаємної зв'язаності Пірсона *C*, Чупрова *K* та Крамера *V*.

1. Встановлюємо рівень ймовірності помилки першого роду α = 0,05 і формулюємо нульову та альтернативну гіпотези:

H_0: між рівнями розвитку загальних пізнавальних здібностей і професійної успішністю відсутня асоціативна залежність (C, K, $V = 0$);

H_1: між рівнями розвитку загальних пізнавальних здібностей і професійної успішністю існує асоціативна залежність (C, K, $V \neq 0$).

2. Для розрахунку емпіричного значення χ^2 необхідно спочатку визначити теоретичні (очікувані) частоти для кожної комірки таблиці.

$$f_{ij} = \frac{f_i \cdot f_j}{n}, \text{ де}$$

f_i – сума частот у всіх комірках i-рядка;

f_j – сума частот у всіх комірках j- стовпчика;

n – сума частот у всій таблиці зв'язаності ознак.

Таблиця 3.14 – Значення очікуваних частот

ЗПЗ	Рівень професійної успішності			Σ
	Високий	Середній	Низький	
Високі	$f_{11}=\frac{41\cdot 37}{102}=14,87$	$f_{12}=\frac{41\cdot 38}{102}=15,28$	$f_{13}=\frac{41\cdot 27}{102}=10,85$	41
Середні	$f_{21}=\frac{33\cdot 37}{102}=11,97$	$f_{22}=\frac{33\cdot 38}{102}=12,29$	$f_{23}=\frac{33\cdot 27}{102}=8,74$	33
Низькі	$f_{31}=\frac{28\cdot 37}{102}=10,16$	$f_{32}=\frac{28\cdot 38}{102}=10,43$	$f_{33}=\frac{28\cdot 27}{102}=7,41$	28
Σ	37	38	27	102

3. Підставимо значення емпіричних і очікуваних частот у формулу для розрахунку χ^2.

$$\chi^2=\sum_{i=1}^{k}\frac{\left(f_{емп.}-f_{теор.}\right)^2}{f_{теор.}}=\frac{(20-14,87)^2}{14,87}+\frac{(15-15,28)^2}{15,28}+\frac{(6-10,85)^2}{10,85}+$$

$$+\frac{(12-11,97)^2}{11,97}+\frac{(13-12,29)^2}{12,29}+\frac{(8-8,74)^2}{8,74}+\frac{(5-10,16)^2}{10,16}+\frac{(10-10,43)^2}{10,43}+$$

$$+\frac{(13-7,41)^2}{7,41}=10,86.$$

4. Обчислюємо значення коефіцієнтів взаємної зв’язаності за відповідними формулами:

$$C=\sqrt{\frac{\chi^2}{\chi^2+n}}=\sqrt{\frac{10,86}{10,86+102}}=0,31;$$

$$K=\sqrt{\frac{\chi^2}{n\cdot\sqrt{(k-1)\cdot(m-1)}}}=\sqrt{\frac{10,86}{102\cdot\sqrt{(3-1)\cdot(3-1)}}}=0,23;$$

$$V=\sqrt{\frac{\chi^2}{n\cdot(c-1)}}=\sqrt{\frac{10,86}{102\cdot(3\text{-}1)}}=0,23.$$

Отримані значення свідчать про помірний та низький рівень сили залежності між ознаками. Перевіримо їхню статистичну значущість.

5. За таблицею критичних значень критерію χ^2-Пірсона (додаток 2) для $df=(k-1)\cdot(m-1)=(3-1)\cdot(3-1)=4$ та заданого $\alpha = 0,05$ визначаємо $\chi^2_{крит.} = 9,488$..

Оскільки $\chi^2_{емп.} > \chi^2_{крит.}$ (10,86 > 9,488), гіпотеза H_0 відхиляється.

5. Оцінюємо розмір ефекту.

Показником розміру ефекту є безпосередньо саме емпіричне значення коефіцієнтів взаємної зв’язаності ознак. Згідно інтерпретації Дж. Коена 0,231 відповідає малому рівню, а 0,311 – середньому.

Висновок. Коефіцієнти взаємної зв’язаності між ознаками дозволили відхилити нульову гіпотезу про відсутність асоціації (C = 0,31; K = 0,23; V = 0,23; $p < 0,05$ df = 4; n = 102). Це свідчить про наявність статистично

значущого асоціативного зв'язку між рівнем розвитку загальних пізнавальних здібностей і показниками професійної успішності. Отримані значення коефіцієнтів ефекту відповідають середньому або малому рівню залежно від використаного статистичного критерію.

– відео-огляд процедури аналізу даних Прикладу 3.7. у статистичній програмі SPSS Statistics 23.0. та інтерпретація результатів дослідження.

3.5. Бісеріальні коефіцієнти кореляції

Бісеріальні коефіцієнти кореляції застосовуються для оцінювання взаємозв'язку між двома змінними, одна з яких представлена у номінальній дихотомічній шкалі, а інша – в порядковій або кількісній шкалі (інтервальній чи шкалі відношень). Значення бісеріального коефіцієнта змінюється в межах від –1 до +1. При цьому знак коефіцієнта не має істотного значення для інтерпретації результатів, оскільки він залежить винятково від порядку подання градацій номінальної змінної.

3.5.1. Точково-бісеріальний коефіцієнт кореляції r_{pb}

Точково-бісеріальний коефіцієнт кореляції (r_{pb}) використовується у випадках, коли одна з досліджуваних змінних вимірюється в дихотомічній номінальній шкалі, а інша – в інтервальній або шкалі відношень. Емпіричне значення точково-бісеріального коефіцієнта кореляції обчислюється за такою формулою:

$$r_{pb} = \frac{\overline{y_1} - \overline{y_0}}{s_y} \cdot \sqrt{\frac{n_1 \cdot n_0}{n \cdot (n-1)}}$$, де

$\overline{y_1}$ – середнє арифметичне змінної У для об'єктів, що мають за Х одиницю;

$\overline{y_0}$ – середнє арифметичне змінної У для об'єктів, що мають за Х нуль;

s_y – стандартне відхилення Y;

n_1, n_0 – чисельність підгруп із відповідними значеннями номінальної змінної;

n – обсяг вибірки досліджуваних.

Точково-бісеріальний коефіцієнт кореляції не має власної таблиці критичних значень. Оскільки його статистичний розподіл відповідає розподілу Стьюдента, перевірка статистичної значущості здійснюється з використанням таблиці критичних значень *t*-критерію Стьюдента. Для цього розраховується емпіричне значення *t*-статистики за наступною формулою:

$$t = \left|r_{pb}\right| \cdot \sqrt{\frac{n-2}{1-r_{pb}^2}}$$, де

r_{pb} – емпіричне значення точково-бісеріального коефіцієнта кореляції;
n – обсяг вибірки досліджуваних.

Якщо емпіричне значення t дорівнює або більше критичного значення для заданого α-рівня, то H_0 відхиляється і формулюється висновок, що коефіцієнт кореляції статистично значуще відрізняється від нуля.

Показником розміру ефекту виступає безпосередньо саме емпіричне значення точково-бісеріального коефіцієнта.

Приклад 3.8. Необхідно перевірити, чи існує взаємозв'язок між статтю людини та її самооцінкою? Параметр «самооцінка» в популяції відповідає закону нормального розподілу, жіноча стать досліджуваного закодована – 0, чоловіча – 1. Результати дослідження представлені в таблиці 3.15.

Таблиця 3.15 – Таблиця первинних даних та проміжних розрахунків

№	Стать (X)	Самооцінка (Y)	$y_i - \bar{y}$	$\left(y_i - \bar{y}\right)^2$
1	1	95,7	15,9	252,81
2	1	97	17,2	295,84
3	1	83	3,2	10,24
4	1	98,6	18,8	353,44
5	1	88,6	8,8	77,44
6	1	83	3,2	10,24
7	1	71	-8,8	77,44
8	0	90	10,2	104,04
9	0	70	-9,8	96,04
10	0	70	-9,8	96,04
11	0	70	-9,8	96,04
12	0	68,6	-11,2	125,44
13	0	70	-9,8	96,04
14	0	73	-6,8	46,24
15	0	68,5	-11,3	127,69
		$\bar{y}$ = 79,8	Σ = 0	Σ = 1865,02

Змінна X представлена в дихотомічній номінальній шкалі, а змінна Y вимірюється в інтервальній шкалі та відповідає закону нормального розподілу (згідно з умовою завдання), за таких умов для оцінювання залежності між цими змінними необхідно застосовувати точково-бісеріальний коефіцієнт кореляції.

1. Встановлюємо рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

α = 0,05;

H_0: самооцінка не залежить від статі людини ($r_{pb} = 0$);

H_1: самооцінка залежить від статі людини ($r_{pb} \neq 0$).

2. Визначаємо середнє арифметичне значення самооцінки для дівчат та хлопців:

$$\bar{y}_1 = \frac{95{,}7 + 97 + 83 + 98{,}6 + 88{,}6 + 83 + 71}{7} = 88{,}1;$$

$$\bar{y}_0 = \frac{90 + 70 + 70 + 70 + 68{,}6 + 70 + 73 + 68{,}5}{8} = 72{,}5.$$

3. Обчислюємо стандартне відхилення для Y:

$$s = \sqrt{\frac{\Sigma\left(y_i - \bar{y}\right)^2}{n-1}} = \sqrt{\frac{1865{,}02}{15-1}} = 11{,}54.$$

4. Розраховуємо емпіричне значення точково-бісеріального коефіцієнту кореляції за формулою r_{pb}:

$$r_{pb} = \frac{\overline{y_1} - \overline{y_0}}{s_y} \cdot \sqrt{\frac{n_1 \cdot n_0}{n \cdot (n-1)}} = \frac{88{,}1 - 72{,}5}{11{,}54} \cdot \sqrt{\frac{7 \cdot 8}{15 \cdot (15-1)}} = 0{,}699.$$

Отримане значення свідчить про кореляцію середньої сили між змінними. Перевіримо її статистичну значущість.

5. Пошук критичних значень здійснюється за таблицею критичних значень критерію *t*-Стьюдента. Для обчислення емпіричного значення використовується формула:

$$t = \left|r_{pb}\right| \cdot \sqrt{\frac{n-2}{1-r_{pb}^2}} = \left|0{,}699\right| \cdot \sqrt{\frac{15-2}{1-0{,}699^2}} = 3{,}53.$$

6. За таблицею критичних значень критерію *t*-Стьюдента (додаток 3) для $df = n - 2 = 15 - 2 = 13$ та заданого $\alpha = 0{,}05$ знаходимо $t_{крит.} = 2{,}16$.

Оскільки $t_{емп.} > t_{крит.}$ (3,53 > 2,16), гіпотеза H_0 відхиляється.

7. Оцінюємо розмір ефекту.

Згідно Дж. Коена значення $r_{pb} = 0{,}699$ відповідає великому розміру ефекту.

Висновок. Застосування точково-бісеріального коефіцієнта кореляції дозволяє стверджувати про наявність статистично значущої залежності середньої сили між рівнем самооцінки та статтю респондентів ($r_{pb} = 0{,}699$; $p < 0{,}05$; $n = 15$). Розмір стандартизованого ефекту – великий.

– відео-огляд процедури аналізу даних Прикладу 3.8. у статистичній програмі SPSS Statistics 23.0. та інтерпретація результатів дослідження.

3.5.2. Рангово-бісеріальний коефіцієнт кореляції r_{rb}

Рангово-бісеріальний коефіцієнт кореляції застосовується у випадках, коли одна змінна представлена в дихотомічній номінальній шкалі, а інша – у порядковій шкалі. Емпіричне значення розраховується за формулою:

$$r_{rb} = \frac{(\overline{y_1} - \overline{y_0}) \cdot 2}{n}$$, де

$\overline{y_1}$ – середній ранг змінної У для об'єктів, що мають за Х одиницю;

$\overline{y_0}$ – середній ранг змінної У для об'єктів, що мають за Х нуль;

n – обсяг вибірки досліджуваних.

Рангово-бісеріальний коефіцієнт кореляції не має власної таблиці критичних значень. Оскільки статистика цього коефіцієнта має розподіл Стьюдента, визначення критичних значень здійснюється за таблицею критичних значень *t*-критерію Стьюдента. Для обчислення емпіричного значення *t*-статистики використовується формула:

$$t = |r_{rb}| \cdot \sqrt{\frac{n-2}{1-r_{rb}^2}}$$, де

r_{rb} – емпіричне значення рангово-бісеріального коефіцієнта кореляції;
n – обсяг вибірки досліджуваних.

Якщо емпіричне значення *t* дорівнює або більше критичного значення для заданого α-рівня, то H_0 відхиляється і формулюється висновок, що коефіцієнт кореляції статистично значуще відрізняється від нуля.

Показником розміру ефекту виступає безпосередньо саме емпіричне значення рангово-бісеріального коефіцієнта. При інтерпретації необхідно орієнтуватися на рекомендації запропоновані Дж. Коеном (див. таблицю 3.2.).

Приклад 3.9. Необхідно перевірити наявність статистично значущого зв'язку між статтю досліджуваних та рівнем їх комунікативних здібностей. Змінна «стать» закодована дихотомічно: дівчата – 0, хлопці – 1. Результати дослідження наведені у таблиці 3.16.

Таблиця 3.16 – Таблиця первинних даних

№	Стать (Х)	Ранг вербальних здібностей (Y)	№	Стать (Х)	Ранг вербальних здібностей (Y)
1	1	1	9	1	4
2	0	10	10	1	3
3	1	6	11	1	5
4	1	9	12	0	11
5	0	15	13	1	12
6	1	7	14	1	2
7	0	8	15	0	14
8	0	13			

Змінна X представлена в дихотомічній номінальній шкалі, а змінна Y в шкалі порядку. Для того, щоб визначити зв'язок між змінними необхідно використати рангово-бісеріальний коефіцієнт кореляції.

1. Встановлюємо рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: комунікативні здібності особистості не залежать від статі ($r_{rb} = 0$);

H_1: комунікативні здібності особистості залежать від статі ($r_{rb} \neq 0$).

1. Визначаємо середній ранг комунікативних здібностей для дівчат та хлопців:

$$\bar{y}_1 = \frac{1+6+9+7+4+3+5+12+2}{9} = 5{,}44;$$

$$\bar{y}_0 = \frac{10+15+8+13+11+14}{6} = 11{,}83.$$

2. Розраховуємо емпіричне значення рангово-бісеріального коефіцієнту кореляції за формулою r_{rb}:

$$r_{rb} = \frac{(\bar{y}_1 - \bar{y}_0)\cdot 2}{n} = \frac{(5{,}44 - 11{,}83)\cdot 2}{15} = -0{,}852.$$

Отримане значення свідчить про сильну кореляцію між змінними (знак у бісеріальних коефіцієнтах кореляції не інтерпретується). Перевіримо її статистичну значущість.

3. Пошук критичних значень здійснюється за таблицею критичних значень критерію t-Стьюдента. Для обчислення емпіричного значення використовується формула:

$$t = |r_{rb}| \cdot \sqrt{\frac{n-2}{1-r_{rb}^2}} = |-0{,}852| \cdot \sqrt{\frac{15-2}{1-(-0{,}852)^2}} = 5{,}869.$$

5. За таблицею критичних значень критерію *t*-Стьюдента (додаток 3) для $df = n - 2 = 15 - 2 = 13$ та заданого $\alpha = 0{,}05$ знаходимо $t_{крит.} = 2{,}16$.

Оскільки $t_{емп.} > t_{крит.}$ (5,869 > 2,16), гіпотеза H_0 відхиляється.

7. Оцінюємо розмір ефекту.

Згідно Дж. Коена значення 0,852 відповідає великому розміру ефекту.

Висновок. Рангово-бісеріальний коефіцієнт кореляції дозволив відхилити нульову гіпотезу ($r_{rb} = 0{,}852$; $p < 0{,}05$; $n = 15$). Тобто, існує статистично значуща залежність між комунікативними здібностями особистості та статтю. Розмір стандартизованого ефекту – великий.

Завдання на самостійну підготовку

1. Поясніть сутність кореляційного аналізу та його роль у психологічних дослідженнях.

2. Розкрийте поняття коефіцієнта кореляції та його математичний зміст.

3. Поясніть відмінності між сильною, середньою та слабкою кореляцією.

4. Охарактеризуйте поняття функціонального та кореляційного зв’язку між змінними.

5. Вкажіть причини та помилки, які можуть виникати під час інтерпретації коефіцієнтів кореляції.

6. За яких умов для встановлення зв’язку між змінними використовується тетрахоричний коефіцієнт кореляції?

7. В яких ситуаціях використовується кореляційне відношення η?

8. Сутність методу регресійного аналізу даних

9. Як графічно можна зобразити зв’язок між двома змінними? Кореляційні плеяди.

10. Проаналізуйте схему вибору кореляційних критеріїв залежно від характеру сукупності і досліджуваних завдань.

11. Особливості кореляційного аналізу номінативних даних.

12. Визначте переваги та обмеження рангових методів кореляційного аналізу.

13. Характеристика бісеріальних коефіцієнтів кореляції.

14. Проаналізуйте приклади практичного застосування кореляційного аналізу у психологічних дослідженнях.

РОЗДІЛ 4
ПАРАМЕТРИЧНІ МЕТОДИ СТАТИСТИЧНОГО ПОРІВНЯННЯ ВИБІРОК ДОСЛІДЖУВАНИХ

4.1. Теоретичні засади та сфера застосування критерію *t*-Стьюдента

Критерій t-Стьюдента – загальна назва групи статистичних методів, що базуються на порівнянні вибіркових даних із розподілом Стьюдента. Найчастіше цей критерій використовується для перевірки гіпотези про рівність середніх значень у двох вибірках.

Критерій був розроблений англійським статистиком Вільямом Госсетом (1876–1937). Працюючи на пивоварні Guinness у Дубліні, він через умови контракту не мав права публікувати результати своїх досліджень під власним ім'ям. Щоб обійти цю заборону, Госсет публікував роботи під псевдонімом «Student», що і зумовило назву критерію.

У вітчизняній науковій літературі цей метод часто називають «t-критерієм Стьюдента», тоді як у англомовних джерелах і статистичних програмних пакетах зазвичай використовується термін «t-test».

Критерій t-Стьюдента застосовується у трьох основних випадках:

1) для порівняння середніх значень двох незалежних вибірок (*t*-критерій для незалежних вибірок);

2) для порівняння середніх значень двох залежних вибірок (*t*-критерій для залежних вибірок);

3) для порівняння середнього значення однієї вибірки із заданою константою (*t*-критерій для однієї вибірки, одновибірковий *t*-критерій).

Вимоги до застосування.

Загальною вимогою для застосування всіх видів *t*-критеріїв Стьюдента є нормальність розподілу даних у вибірках. У разі відсутності нормального розподілу використання цього критерію вважається некоректним. Крім того, первинні дані повинні бути представлені у метричній шкалі вимірювання (інтервальній або шкалі відношень).

Особливою вимогою, що стосується *t*-критерію для незалежних вибірок, є гомогенність дисперсій порівнюваних вибірок. У випадку гетерогенності дисперсій доцільно застосовувати альтернативний метод – *t*-критерій Уелша.

Для застосування *t*-критерію Стьюдента у випадку залежних вибірок формальною умовою є наявність статистично значущого прямого кореляційного зв'язку між парними спостереженнями.

Щодо обсягу вибірок, формальних обмежень не встановлено. Проте на малих вибірках (менше 30 спостережень) досить складно об'єктивно оцінити відповідність емпіричного розподілу нормальному закону. У таких випадках дослідники часто відмовляються від перевірки нормальності та безпосередньо застосовують непараметричні методи аналізу, такі як критерії Манна-Уітні, Вілкоксона тощо.

Загальний алгоритм застосування t-критерію Стьюдента.

1. *Перевірка даних у обох вибірках досліджуваних на відповідність закону нормального розподілу.*

Без використання спеціалізованих статистичних програм кількісна оцінка нормальності розподілу даних може здійснюватися шляхом розрахунку коефіцієнтів асиметрії, ексцесу та відповідних стандартних помилок вимірювання цих показників. Детальніше це питання розглядається у розділі №2.

2А. Перевірка рівності (гомогенності) дисперсій для незалежних вибірок.

Використовується *критерій F-Фішера для перевірки гомогенності дисперсій*. Емпіричне значення визначається за формулою:

$$F = \frac{s_1^2}{s_2^2};\ \ df_{чис.} = n_1 - 1;\ df_{знам.} = n_2 - 1, \text{ де}$$

s_1^2 – більша дисперсія із двох вибірок;

s_2^2 – менша дисперсія із двох вибірок;

n_1, n_2 – обсяг досліджуваних у двох вибірках.

Критичні значення ($F_{крит.}$) визначаємо за таблицею критичних значень *F*-критерію Фішера для перевірки ненаправлених альтернатив (додаток 4).

Якщо $F_{емп.} \geq F_{крит.}$ для відповідно числа ступенів свободи, то приймаємо рішення про підтвердження гіпотези про відмінність дисперсій на заданому α-рівні, та відповідно, слід застосовувати не критерій Стьюдента в стандартній формі, а *t-критерій Уелша* (модифікований критерій Стьюдента для незалежних вибірок із гетерогенними дисперсіями).

2Б. Перевірка наявності кореляційного зв'язку для залежних вибірок дослідження.

Дуже часто залежність вибірок визначається дизайном дослідження, хоча формальним критерієм є наявність кореляційного зв'язку.

Використовується лінійний коефіцієнт кореляції r_{xy} Пірсона, який визначається за формулою:

$$r_{xy} = \frac{\Sigma\left(x_i - \bar{x}_1\right)\cdot\left(x_i - \bar{x}_2\right)}{\left(n-1\right)\cdot s_1 \cdot s_2}, \text{ де}$$

x_i – первинні значення;

$\bar{x}_1$, $\bar{x}_2$ – середнє арифметичне змінних X_1 та X_2;

s_1, s_2 – стандартне відхилення змінних X_1 та X_2;

n – обсяг вибірки досліджуваних.

Отримане емпіричне значення коефіцієнта кореляції r_{xy} Пірсона необхідно порівняти із критичним значенням. Кількість ступенів свободи дорівнює обсягу вибірки досліджуваних ($df = n$).

3. Власне обчислення t-критерію Стьюдента.

Обчислення критерію залежить від:

- вибірки залежні чи незалежні;

- дисперсії гомогенні чи гетерогенні;
- однаковий чи різний обсяг вибірок.

4. Порівняння отриманого емпіричного значення t-критерію із критичним табличним значенням для формулювання статистичного висновку.

За таблицею критичних значень критерію *t*-Стьюдента (додаток 3) для заданого α-рівня та визначеного число ступенів свободи (*df*), знаходимо $t_{крит.}$:

- якщо $|t_{емп.}| < t_{крит.}$ – приймаємо гіпотезу H_0. Відмінності між вибірками статистично незначущі.

- якщо $|t_{емп.}| \geq t_{крит.}$ – приймаємо гіпотезу H_1. Відмінності між вибірками статистично значущі.

5. Оцінка розміру ефекту.

Для оцінки розміру ефекту критерію *t*-Стьюдента використовується індекс *d*-Коена та його модифікації:

1) у випадку незалежності вибірок, рівності обсягів вибірок та гомогенності дисперсій:

$$d = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{s_{pooled.}}} = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}, \text{ де}$$

$\bar{x}_1$, $\bar{x}_2$ – середнє арифметичне значення в першій та другій вибірках;
s_1, s_2 – стандартне відхилення в першій та другій вибірках.

2) в ситуації незалежності вибірок, коли у них різна кількість досліджуваних у групах або/та малі обсяги вибірок для уникнення переоцінки розміру ефекту у формулу розрахунку *d*-Коена вноситься поправка *g*-Хеджесса:

$$g = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{\frac{(n_1 - 1)\cdot s_1^2 + (n_2 - 1)\cdot s_2^2}{n_1 + n_2 - 2}}}, \text{ де}$$

$\bar{x}_1$, $\bar{x}_2$ – середнє арифметичне значення в першій та другій вибірках;
s_1, s_2 – стандартне відхилення в першій та другій вибірках;
n_1, n_2 – обсяг досліджуваних у першій та другій вибірках.

3) за умови гетерогенності дисперсій у незалежних вибірках до індексу *d*-Коена вноситься поправка Δ*s*-Гласса:

$$\Delta s = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{s^2_{контрол.}}}, \text{ де}$$

$\bar{x}_1$, $\bar{x}_2$ – середнє арифметичне значення в першій та другій вибірках;
$s^2_{контрол.}$ – стандартне відхилення в контрольній групі.

4) індекс розміру ефекту для залежних вибірок має позначення d_z-Коена:

$$d_z = \frac{\left|\bar{x}_d\right|}{s_d}$$, де

$\left|\bar{x}_d\right|$ – середнє арифметичне різниці пар значень;

s_d – стандартне відхилення різниці пар значень.

5) для критерія *t*-Стьюдента для однієї вибірки використовується індекс d_{os}-Коена:

$$d_{os} = \frac{\left|\bar{x} - A\right|}{s}$$, де

$\bar{x}$ – середнє арифметичне значення;

A – величина, з якою середнє арифметичне порівнюється;

s – стандартне відхилення.

На сьогодні не існує єдиних узгоджених стандартів для інтерпретації величини розміру ефекту. Дж. Коеном було запропоновано методичні рекомендації щодо градації значень коефіцієнта *d* та його модифікацій, які широко використовуються в практиці. Зокрема, розмір ефекту поділяється на такі категорії:

$0{,}00 \le d < 0{,}20$ – несуттєвий;

$0{,}20 \le d < 0{,}50$ – малий;

$0{,}50 \le d < 0{,}80$ – середній;

$0{,}80 \le d$ – великий.

Ці діапазони необхідно сприймати тільки в якості орієнтирів, так як для кожного окремого дослідження повинен застосовуватися свій підхід до інтерпретації. При оцінці розміру ефекту необхідно враховувати теоретичні, економічні, етичні та практичні аспекти.

4.1.1. Критерій *t*-Стьюдента для незалежних вибірок

Критерій *t*-Стьюдента призначений для порівняння середніх значень ознак у двох незалежних групах.

Формальною умовою незалежності вибірок є відсутність кореляції між ними. З точки зору змісту, незалежними вважаються вибірки, між якими не існує жодних зв'язків. Прикладами таких вибірок можуть бути рандомізовані (сформовані випадковим чином) контрольна та експериментальна групи, дві різні професійні групи тощо.

У разі рівності обсягів вибірок емпіричне значення *t*-критерію Стьюдента обчислюється за формулою:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2 + s_2^2}{n}}}; \; df = 2n - 2.$$

Якщо обсяги вибірок різні:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)\cdot s_1^2 + (n_2 - 1)\cdot s_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}; \; df = n_1 + n_2 - 2, \text{ де}$$

$\bar{x}_1$, $\bar{x}_2$ – середнє арифметичне значення в першій та другій вибірках;
s_1, s_2 – стандартне відхилення в першій та другій вибірках;
n_1, n_2 – обсяг досліджуваних у першій та другій вибірках;
n – загальний обсяг досліджуваних.

Якщо дисперсії гетерогенні, то застосовується *t*-критерій Уелша для незалежних вибірок:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}; \; df = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\left(s_1^2/n_1\right)^2/(n_1 - 1) + \left(s_2^2/n_2\right)^2/(n_2 - 1)}, \text{ де}$$

$\bar{x}_1$, $\bar{x}_2$ – середнє арифметичне значення в першій та другій вибірках;
s_1, s_2 – стандартне відхилення в першій та другій вибірках;
n_1, n_2 – обсяг досліджуваних у першій та другій вибірках;
n – загальний обсяг досліджуваних.

Приклад 4.1. Психолог досліджував час прийняття рішення (у секундах) в умовах стресу. До першої групи (X_1) увійшло 15 осіб, які регулярно займаються спортом, до другої групи (X_2) – 13 осіб, які спортом не займаються. Поставлено завдання визначити, чи існує статистично значуща різниця у швидкості прийняття рішення між цими двома групами з різним рівнем фізичної активності.

Таблиця 4.1 – Таблиця первинних даних та проміжних розрахунків

№	Група №1 (X_1)	Група №2 (X_2)	$(x_i - \bar{x}_1)$	$(x_i - \bar{x}_2)$	$(x_i - \bar{x}_1)^2$	$(x_i - \bar{x}_2)^2$	$(x_i - \bar{x}_1)^3$	$(x_i - \bar{x}_2)^3$	$(x_i - \bar{x}_1)^4$	$(x_i - \bar{x}_2)^4$
1	7	4	1,73	-2,15	2,99	4,62	5,18	-9,94	8,96	21,37
2	5	6	-0,27	-0,15	0,07	0,02	-0,02	0,00	0,01	0,00
3	8	9	2,73	2,85	7,45	8,12	20,35	23,15	55,55	65,98
4	6	8	0,73	1,85	0,53	3,42	0,39	6,33	0,28	11,71
5	4	4	-1,27	-2,15	1,61	4,62	-2,05	-9,94	2,60	21,37
6	5	6	-0,27	-0,15	0,07	0,02	-0,02	0,00	0,01	0,00
7	7	6	1,73	-0,15	2,99	0,02	5,18	0,00	8,96	0,00
8	5	5	-0,27	-1,15	0,07	1,32	-0,02	-1,52	0,01	1,75
9	4	7	-1,27	0,85	1,61	0,72	-2,05	0,61	2,60	0,52
10	5	4	-0,27	-2,15	0,07	4,62	-0,02	-9,94	0,01	21,37
11	4	7	-1,27	0,85	1,61	0,72	-2,05	0,61	2,60	0,52
12	5	9	-0,27	2,85	0,07	8,12	-0,02	23,15	0,01	65,98
13	5	5	-0,27	-1,15	0,07	1,32	-0,02	-1,52	0,01	1,75
14	6		0,73		0,53		0,39		0,28	
15	3		-2,27		5,15		-11,70		26,55	
$\bar{x}_1$ = 5,27; $\bar{x}_2$ = 6,15.		Σ =			24,93	37,69	13,52	20,99	108,42	212,31

У таблиці наведено результати тестування двох незалежних груп досліджуваних. Первинні дані представлені у шкалі відношень. Для виявлення відмінностей між групами існують передумови до застосування параметричного *t*-критерію Стьюдента для незалежних вибірок. Наступним кроком є перевірка первинних даних на відповідність обмеженням, необхідним для коректного застосування цього статистичного критерію.

1. У наведеному прикладі для перевірки відповідності даних у обох вибірках закону нормального розподілу застосуємо критерій Є.І. Пустильника.

Розподіл вважається нормальним, якщо показники асиметрії (A) й ексцесу (E) менші своїх критичних значень ($A_{крит.}$ і $E_{крит.}$).

Спочатку розрахуємо показники асиметрії та ексцесу в обох вибірках дослідження. Для цього необхідно визначити середнє арифметичне значення та стандартне відхилення.

$\bar{x}_1 = 5{,}27;\ \bar{x}_2 = 6{,}15.$

$$s_1 = \sqrt{\frac{\Sigma\left(x_i - \bar{x}_1\right)^2}{n-1}} = \sqrt{\frac{24{,}93}{15-1}} = 1{,}33;$$

$$s_2 = \sqrt{\frac{\Sigma\left(x_i - \bar{x}_2\right)^2}{n-1}} = \sqrt{\frac{37{,}69}{13-1}} = 1{,}77.$$

Для визначення асиметрії та ексцесу в обох вибірках дослідження використаємо точні розрахункові формули:

$$A_1 = \frac{n \cdot \Sigma\left(x_i - \bar{x}_1\right)^3}{(n-1)\cdot(n-2)\cdot s_1^3} = \frac{15 \cdot 13{,}52}{(15-1)\cdot(15-2)\cdot 1{,}33^3} = 0{,}474.$$

$$E_1 = \frac{n\cdot(n+1)\cdot\Sigma\left(x_i - x_1\right)^4}{(n-1)\cdot(n-2)\cdot(n-3)\cdot s_1^4} - \frac{3\cdot(n-1)^2}{(n-2)\cdot(n-3)} = \frac{15\cdot(15+1)\cdot 108{,}42}{(15-1)\cdot(15-2)\cdot(15-3)\cdot 1{,}33^4} -$$

$$- \frac{3\cdot(15-1)^2}{(15-2)\cdot(15-3)} = 3{,}808 - 3{,}769 = 0{,}039.$$

$$A_2 = \frac{n \cdot \Sigma\left(x_i - \bar{x}_2\right)^3}{(n-1)\cdot(n-2)\cdot s_2^3} = \frac{13 \cdot 20{,}99}{(13-1)\cdot(13-2)\cdot 1{,}77^3} = 0{,}37.$$

$$E_2 = \frac{n\cdot(n+1)\cdot\Sigma\left(x_i - \bar{x}_2\right)^4}{(n-1)\cdot(n-2)\cdot(n-3)\cdot s_2^4} - \frac{3\cdot(n-1)^2}{(n-2)\cdot(n-3)} = \frac{13\cdot(13+1)\cdot 212{,}31}{(13-1)\cdot(13-2)\cdot(13-3)\cdot 1{,}77^4} -$$

$$- \frac{3\cdot(13-1)^2}{(13-2)\cdot(13-3)} = 2{,}982 - 3{,}927 = -0{,}945.$$

Розрахуємо критичні значення для показників асиметрії та ексцесу для обох вибірок дослідження:

$$A_{крит.1} = 3\cdot\sqrt{\frac{6\cdot(n-1)}{(n+1)(n+3)}} = 3\cdot\sqrt{\frac{6\cdot(15-1)}{(15+1)\cdot(15+3)}} = 3\cdot 0{,}54 = 1{,}62;$$

$$E_{крит.1} = 5 \cdot \sqrt{\frac{24 \cdot n \cdot (n-2) \cdot (n-3)}{(n+1)^2 \cdot (n+3) \cdot (n+5)}} = 5 \cdot \sqrt{\frac{24 \cdot 15 \cdot (15-2) \cdot (15-3)}{(15+1)^2 \cdot (15+3) \cdot (15+5)}} = 5 \cdot 0{,}78 = 3{,}9.$$

$$A_{крит.2} = 3 \cdot \sqrt{\frac{6 \cdot (n-1)}{(n+1)(n+3)}} = 3 \cdot \sqrt{\frac{6 \cdot (13-1)}{(13+1) \cdot (13+3)}} = 3 \cdot 0{,}57 = 1{,}71;$$

$$E_{крит.2} = 5 \cdot \sqrt{\frac{24 \cdot n \cdot (n-2) \cdot (n-3)}{(n+1)^2 \cdot (n+3) \cdot (n+5)}} = 5 \cdot \sqrt{\frac{24 \cdot 13 \cdot (13-2) \cdot (13-3)}{(13+1)^2 \cdot (13+3) \cdot (13+5)}} = 5 \cdot 0{,}78 = 3{,}9.$$

Згідно з критерієм Є.І. Пустильника дані відповідають закону нормального розподілу, якщо:

$|A| \le A_{крит.}$ та $|E| \le E_{крит.}$

У нашому випадку:

|0,474| < 1,62 та |0,039| < 3,9 – для першої вибірки;

|0,37| < 1,71 та |-0,945| < 3,9 – для другої вибірки;

Можемо зробити висновок, що розподіл даних у обох вибірках відповідає нормальному закону розподілу. Отже, для подальшого аналізу доцільно застосувати параметричний *t*-критерій Стьюдента для незалежних вибірок.

2. Здійснюємо перевірку рівності (гомогенності) дисперсій для незалежних вибірок.

Використовуємо критерій *F*-Фішера для визначення гомогенності дисперсій двох вибірок. Обчислюємо емпіричне значення статистики за формулою:

$$F = \frac{s_1^2}{s_2^2} = \frac{1{,}77^2}{1{,}33^2} = 1{,}77.$$

s_1^2 – більша дисперсія із двох вибірок;

s_2^2 – менша дисперсія із двох вибірок.

Обчислюємо кількість ступенів свободи:

$df_{чис.} = n_1 - 1 = 13 - 1 = 12$;

$df_{знам.} = n_2 - 1 = 15 - 1 = 14$.

Визначаємо $F_{0,05}$, за таблицею критичних значень *F*-критерію Фішера для перевірки ненаправлених альтернатив (додаток 4).

$F_{0,05} = 3{,}05$.

Оскільки $F_{емп.} < F_{крит.}$ (1,77 < 3,05), гіпотеза H_0 підтверджуються. Дисперсії вибірок не відрізняються, тобто гомогенні ($F_{(12;\ 14)} = 1{,}77$; $p > 0{,}05$). Ми можемо застосувати критерій *t*-Стьюдента для незалежних вибірок

3. Встановлюємо рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: швидкість прийняття рішення у стресових умовах в групах з різним відношенням до спорту не відрізняється ($\mu_1 = \mu_2$);

H_1: швидкість прийняття рішення у стресових умовах в групах з різним відношенням до спорту відрізняється ($\mu_1 \neq \mu_2$).

4. Розраховуємо емпіричне значення *t*-критерію Стьюдента для незалежних вибірок.

У нашому випадку обсяги двох вибірок різні, тому емпіричне значення *t*-критерію розраховується за уточненою формулою:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)\cdot s_1^2 + (n_2-1)\cdot s_2^2}{n_1+n_2-2}\cdot\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{5,27-6,15}{\sqrt{\frac{(15-1)\cdot 1,33^2 + (13-1)\cdot 1,77^2}{15+13-2}\cdot\left(\frac{1}{15}+\frac{1}{13}\right)}} =$$

$$= \frac{-0,88}{\sqrt{2,4\cdot 0,14}} = -1,51.$$

5. За таблицею критичних значень критерію *t*-Стьюдента (додаток 3) для $df = n_1 + n_2 - 2 = 15 + 13 - 2 = 26$ та заданого $\alpha = 0,05$ знаходимо $t_{крит.} = 2,06$ (двобічна критична область).

Оскільки $|t_{емп.}| > t_{крит.}$ (1,51 < 2,06), гіпотеза H_0 приймається.

5. Оцінка розміру ефекту здійснюється за індексом *g*-Хеджесса, тому що у групах різна кількість досліджуваних:

$$g = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{\frac{(n_1-1)\cdot s_1^2 + (n_2-1)\cdot s_2^2}{n_1+n_2-2}}} = \frac{\left|5,27-6,15\right|}{\sqrt{\frac{(15-1)\cdot 1,33^2 + (13-1)\cdot 1,77^2}{15+13-2}}} = \frac{0,88}{\sqrt{2,4}} = 0,57.$$

Згідно рекомендацій до інтерпретації Дж. Коена розмір ефекту відповідає середньому рівню.

Висновок. Емпіричні дані не дають підстав для відхилення нульової гіпотези за результатами *t*-критерію Стьюдента для незалежних вибірок ($t_{(26)} = -1,51$; $p > 0,05$; двобічна критична область). Відсутня статистично значуща різниця між середніми значеннями швидкості прийняття рішення в стресових умовах у групах спортсменів ($\bar{x} = 5,27$; $s = 1,33$) та осіб, які не займаються спортом ($\bar{x} = 6,15$; $s = 1,77$). Водночас виявлений середній розмір стандартизованого ефекту ($g = 0,57$) свідчить про доцільність збільшення статистичної потужності дослідження шляхом розширення вибірки.

– відео-огляд процедури аналізу даних Прикладу 4.1. у статистичній програмі SPSS Statistics 23.0. та інтерпретація результатів дослідження.

4.1.2. Критерій *t*-Стьюдента для залежних вибірок

Критерій *t*-Стьюдента для залежних вибірок (або *t*-критерій для парних вибірок) дозволяє перевірити гіпотезу про наявність статистично значущої різниці між середніми значеннями двох залежних (парних) вибірок.

Залежність вибірок, як правило, визначається дизайном дослідження. Прикладами залежних вибірок можуть бути група досліджуваних до та після експериментального впливу або ситуації, коли ознака вимірюється у одних і тих самих учасників у два різні моменти часу. До певної міри залежними також можна вважати вибірки, сформовані, наприклад, із чоловіків і їхніх дружин або братів і сестер.

Оскільки дослідження передбачає залежні вибірки, фактично – пари спостережень, одиницею аналізу виступає різниця між значеннями у відповідних парах.

Емпіричне значення *t*-критерію Стьюдента для залежних вибірок обчислюється за формулою:

$$t = \frac{\bar{x}_d}{s_d / \sqrt{n}}, \; df = n - 1, \text{ де}$$

$\bar{x}_d$ – середнє арифметичне різниці пар значень;
s_d – стандартне відхилення різниці пар значень;
n – обсяг вибірки досліджуваних.

Приклад 4.2. Чи відбулися статистично значущі зміни у показниках агресивності підлітків після проведення спеціальних психологічних корекційних вправ?

Таблиця 4.2 – Таблиця первинних даних та проміжних розрахунків

№	До вправ (X_1)	Після вправ (X_2)	$(x_i - \bar{x}_1)$	$(x_i - \bar{x}_2)$	$(x_i - \bar{x}_1)^2$	$(x_i - \bar{x}_2)^2$	$(x_i - \bar{x}_1)^3$	$(x_i - \bar{x}_2)^3$	$(x_i - \bar{x}_1)^4$	$(x_i - \bar{x}_2)^4$
1	24	22	-9,5	-7,4	90,25	54,76	-857,38	-405,22	8145,06	2998,66
2	12	12	-21,5	-17,4	462,25	302,76	-9938,38	-5268,02	213675,06	91663,62
3	42	41	8,5	11,6	72,25	134,56	614,13	1560,90	5220,06	18106,39
4	34	31	0,5	1,6	0,25	2,56	0,13	4,10	0,06	6,55
5	28	32	-5,5	2,6	30,25	6,76	-166,38	17,58	915,06	45,70
6	55	44	21,5	14,6	462,25	213,16	9938,38	3112,14	213675,06	45437,19
7	50	50	16,5	20,6	272,25	424,36	4492,13	8741,82	74120,06	180081,41
8	52	32	18,5	2,6	342,25	6,76	6331,63	17,58	117135,06	45,70
9	50	42	16,5	12,6	272,25	158,76	4492,13	2000,38	74120,06	25204,74
10	32	21	-1,5	-8,4	2,25	70,56	-3,38	-592,70	5,06	4978,71
11	33	34	-0,5	4,6	0,25	21,16	-0,13	97,34	0,06	447,75
12	31	36	-2,5	6,6	6,25	43,56	-15,63	287,50	39,06	1897,47
13	60	38	26,5	8,6	702,25	73,96	18609,63	636,06	493155,06	5470,08
14	25	23	-8,5	-6,4	72,25	40,96	-614,13	-262,14	5220,06	1677,72
15	33	33	-0,5	3,6	0,25	12,96	-0,13	46,66	0,06	167,96
16	29	26	-4,5	-3,4	20,25	11,56	-91,13	-39,30	410,06	133,63
17	17	16	-16,5	-13,4	272,25	179,56	-4492,13	-2406,10	74120,06	32241,79
18	12	18	-21,5	-11,4	462,25	129,96	-9938,38	-1481,54	213675,06	16889,60
19	25	25	-8,5	-4,4	72,25	19,36	-614,13	-85,18	5220,06	374,81
20	26	12	-7,5	-17,4	56,25	302,76	-421,88	-5268,02	3164,06	91663,62
$\bar{x}_1$=33,5; $\bar{x}_2$=29,4. Σ =			0	0	3671	2210,8	17325	713,76	1502014,25	519533,1

У таблиці подано результати психологічного тестування однієї й тієї ж групи досліджуваних, проведеного двічі – до та після застосування корекційних психологічних вправ. Змістовно такі вибірки є залежними (зв'язаними, парними), оскільки спостереження здійснювалися над одними й тими самими респондентами.

Перед застосуванням параметричного *t*-критерію Стьюдента для залежних вибірок необхідно перевірити первинні дані на відповідність вимогам цього критерію.

1. Для перевірки даних на відповідність закону нормального розподілу використаємо критерій М.А. Плохінського. Дані відповідають закону нормального розподілу, якщо:

$$t_A = \frac{|A|}{m_A} \le 3 \quad \text{та} \quad t_E = \frac{|E|}{m_E} \le 3.$$

Для розрахунку показників асиметрії та ексцесу необхідно розрахувати середнє арифметичне значення та стандартне відхилення.

$\bar{x}_1 = 33{,}5;\ \bar{x}_2 = 29{,}4.$

$$s_1 = \sqrt{\frac{\Sigma\left(x_i - \bar{x}_1\right)^2}{n-1}} = \sqrt{\frac{3671}{20-1}} = 13{,}9;$$

$$s_2 = \sqrt{\frac{\Sigma\left(x_i - \bar{x}_2\right)^2}{n-1}} = \sqrt{\frac{2210{,}8}{20-1}} = 10{,}79.$$

Для визначення асиметрії та ексцесу в обох вибірках дослідження застосуємо точні розрахункові формули:

$$A_1 = \frac{n \cdot \Sigma\left(x_i - \bar{x}_1\right)^3}{(n-1)\cdot(n-2)\cdot s_1^3} = \frac{20 \cdot 17325}{(20-1)\cdot(20-2)\cdot 13{,}9^3} = 0{,}377.$$

$$E_1 = \frac{n\cdot(n+1)\cdot\Sigma\left(x_i - \bar{x}_1\right)^4}{(n-1)\cdot(n-2)\cdot(n-3)\cdot s_1^4} - \frac{3\cdot(n-1)^2}{(n-2)\cdot(n-3)} = \frac{20\cdot(20+1)\cdot 1502014{,}25}{(20-1)\cdot(20-2)\cdot(20-3)\cdot 13{,}9^4} -$$

$$- \frac{3\cdot(20-1)^2}{(20-2)\cdot(20-3)} = 2{,}907 - 3{,}539 = -0{,}632.$$

$$A_2 = \frac{n \cdot \Sigma\left(x_i - \bar{x}_2\right)^3}{(n-1)\cdot(n-2)\cdot s_2^3} = \frac{20 \cdot 713{,}76}{(20-1)\cdot(20-2)\cdot 10{,}79^3} = 0{,}033.$$

$$E_2 = \frac{n\cdot(n+1)\cdot\Sigma\left(x_i - \bar{x}_2\right)^4}{(n-1)\cdot(n-2)\cdot(n-3)\cdot s_2^4} - \frac{3\cdot(n-1)^2}{(n-2)\cdot(n-3)} = \frac{20\cdot(20+1)\cdot 519533{,}10}{(20-1)\cdot(20-2)\cdot(20-3)\cdot 10{,}79^4} -$$

$$- \frac{3\cdot(20-1)^2}{(20-2)\cdot(20-3)} = 2{,}769 - 3{,}539 = -0{,}77.$$

Визначаємо стандартні помилки вимірювання асиметрії та ексцесу для $n = 20$.

$$m_A = \sqrt{\frac{6}{n+3}} = \sqrt{\frac{6}{20+3}} = 0{,}511;$$

$$m_E = 2 \cdot \sqrt{\frac{6}{n+3}} = 2 \cdot \sqrt{\frac{6}{20+3}} = 1{,}022.$$

Згідно з критерієм М.А. Плохинського показники асиметрії й ексцесу свідчать про статистичну відмінність емпіричного розподілу від нормального в тому випадку, якщо вони перевищують за абсолютною величиною свої стандартні помилки вимірювання в 3 рази.

$$t_{A_1} = \frac{|A_1|}{m_A} = \frac{|0{,}377|}{0{,}511} = 0{,}74;\ t_{E1} = \frac{|E_1|}{m_E} = \frac{|-0{,}632|}{1{,}022} = 0{,}62.$$

$$t_{A_2} = \frac{|A_2|}{m_A} = \frac{|0{,}033|}{0{,}511} = 0{,}06;\ t_{E2} = \frac{|E_2|}{m_E} = \frac{|-0{,}77|}{1{,}022} = 0{,}75.$$

За результатами розрахунків показники асиметрії та ексцесу в обох вибірках не перевищують утричі свої стандартні помилки вимірювання, тому можемо зробити висновок, що розподіли відповідають закону нормального розподілу. Тобто, можна використовувати параметричний *t*-критерій Стьюдента для залежних вибірок.

2. Встановлюємо рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: агресивність підлітків не змінилася після спеціальних корекційних вправ ($\mu_1 = \mu_2$);

H_1: агресивність підлітків змінилася після спеціальних корекційних вправ ($\mu_1 \neq \mu_2$).

3. Розраховуємо емпіричне значення *t*-критерію Стьюдента для залежних вибірок.

Таблиця 4.3 – Таблиця первинних даних та проміжних розрахунків

№	До вправ (X_1)	Після вправ (X_2)	$d_i = X_1 - X_2$	$(d_i - \bar{x}_d)$	$(d_i - \bar{x}_d)^2$
1	24	22	2	-2,1	4,41
2	12	12	0	-4,1	16,81
3	42	41	1	-3,1	9,61
4	34	31	3	-1,1	1,21
5	28	32	-4	-8,1	65,61
6	55	44	11	6,9	47,61
7	50	50	0	-4,1	16,81
8	52	32	20	15,9	252,81
9	50	42	8	3,9	15,21
10	32	21	11	6,9	47,61
11	33	34	-1	-5,1	26,01
12	31	36	-5	-9,1	82,81

Продовження таблиці 4.3

13	60	38	22	17,9	320,41
14	25	23	2	-2,1	4,41
15	33	33	0	-4,1	16,81
16	29	26	3	-1,1	1,21
17	17	16	1	-3,1	9,61
18	12	18	-6	-10,1	102,01
19	25	25	0	-4,1	16,81
20	26	12	14	9,9	98,01
			$\bar{x}_d$ = 4,1	Σ = 0	Σ = 1155,8

Для розрахунку *t*-критерію необхідно розрахувати середнє арифметичне різниці пар значень ($\bar{x}_d$) та стандартне відхилення різниці пар значень (s_d).

$\bar{x}_d$ = 4,1.

$$s_d = \sqrt{\frac{\Sigma\left(d_i - \bar{x}_d\right)^2}{n-1}} = \sqrt{\frac{1155,8}{20-1}} = 7,8.$$

$$t = \frac{\bar{x}_d}{s_d / \sqrt{n}} = \frac{4,1}{7,8 / \sqrt{20}} = 2,35.$$

4. За таблицею критичних значень критерію *t*-Стьюдента (додаток 3) для $df = n - 1 = 20 - 1 = 19$ та заданого $\alpha = 0,05$ знаходимо $t_{крит.} = 2,09$ (двобічна критична область).

Оскільки $|t_{емп.}| > t_{крит.}$ (2,35 > 2,09), гіпотеза H_0 відхиляється.

5. Оцінка розміру ефекту здійснюється за індексом d_z-Коена:

$$d_z = \frac{\left|\bar{x}_d\right|}{s_d} = \frac{4,1}{7,8} = 0,53.$$

Згідно Дж. Коена значення 0,53 відповідає середньому рівню.

Висновок. Застосування *t*-критерію Стьюдента для залежних вибірок дало підстави для відхилення нульової гіпотези на користь альтернативної ($t_{(19)}$ = 2,35; $p < 0,05$; двобічна критична область). Це свідчить про наявність статистично значущих змін у середньому рівні агресивності підлітків після проведення спеціальних психологічних корекційних вправ ($\bar{x}_d$ = 4,1; s_d = 7,8). Розмір ефекту є середнім за шкалою інтерпретації (d_z = 0,53).

– відео-огляд процедури аналізу даних Прикладу 4.2. у статистичній програмі SPSS Statistics 23.0. та інтерпретація результатів дослідження.

4.1.3. Критерій *t*-Стьюдента для однієї вибірки

Критерій *t*-Стьюдента для однієї вибірки (одновибірковий *t*-критерій) застосовується для перевірки статистичної гіпотези про те, що середнє значення досліджуваної ознаки в генеральній сукупності відрізняється від деякого заданого (теоретичного або нормативного) значення.

Емпіричне значення розраховується за формулою:

$t = \frac{\bar{x} - A}{s/\sqrt{n}}$, $df = n - 1$, де

$\bar{x}$ – середнє арифметичне значення;

A – величина, з якою середнє арифметичне порівнюється;

s – стандартне відхилення;

n – обсяг вибірки досліджуваних.

Приклад 4.3. Чи є статистично значущим перевищення середнього рівня розвитку посттравматичного стресового розладу (ПТСР) у ліквідаторів аварії на Чорнобильській АЕС порівняно з нормативним пороговим показником у 35 балів?

Таблиця 4.4 – Таблиця первинних даних та проміжних розрахунків

№	ПТСР (X)	$\left(x_i - \bar{x}\right)$	$\left(x_i - \bar{x}\right)^2$	$\left(x_i - \bar{x}\right)^3$	$\left(x_i - \bar{x}\right)^4$
1	33	-3	9	-27	81
2	34	-2	4	-8	16
3	39	3	9	27	81
4	36	0	0	0	0
5	41	5	25	125	625
6	42	6	36	216	1296
7	29	-7	49	-343	2401
8	34	-2	4	-8	16
9	38	2	4	8	16
10	32	-4	16	-64	256
11	29	-7	49	-343	2401
12	31	-5	25	-125	625
13	31	-5	25	-125	625
14	38	2	4	8	16
15	40	4	16	64	256
16	42	6	36	216	1296
17	33	-3	9	-27	81
18	35	-1	1	-1	1
19	35	-1	1	-1	1
20	41	5	25	125	625
21	40	4	16	64	256
22	38	2	4	8	16
23	37	1	1	1	1
	$\bar{x} = 36$	$\Sigma = 0$	$\Sigma = 368$	$\Sigma = -210$	$\Sigma = 10988$

Для того щоб мати підстави застосовувати критерій *t*-Стьюдента для однієї вибірки, необхідно перевірити, чи відповідають результати тестування закону нормального розподілу.

1. З метою перевірки даних на нормальність розподілу використаємо критерій М.А. Плохінського, згідно з яким розподіл вважається нормальним, якщо:

$$t_A = \frac{|A|}{m_A} \le 3 \quad \text{та} \quad t_E = \frac{|E|}{m_E} \le 3.$$

Для розрахунку показників асиметрії та ексцесу необхідно розрахувати середнє арифметичне значення та стандартне відхилення.

$\bar{x} = 36.$

$$s = \sqrt{\frac{\Sigma\left(x_i - \bar{x}\right)^2}{n-1}} = \sqrt{\frac{368}{23-1}} = 4,09.$$

Для визначення значення асиметрії та ексцесу застосуємо точні розрахункові формули:

$$A = \frac{n \cdot \Sigma\left(x_i - \bar{x}\right)^3}{(n-1)\cdot(n-2)\cdot s^3} = \frac{23\cdot(-210)}{(23-1)\cdot(23-2)\cdot 4,09^3} = -0,153.$$

$$E = \frac{n\cdot(n+1)\cdot\Sigma\left(x_i - \bar{x}\right)^4}{(n-1)\cdot(n-2)\cdot(n-3)\cdot s^4} - \frac{3\cdot(n-1)^2}{(n-2)\cdot(n-3)} = \frac{23\cdot(23+1)\cdot 10988}{(23-1)\cdot(23-2)\cdot(23-3)\cdot 4,09^4} -$$

$$- \frac{3\cdot(23-1)^2}{(23-2)\cdot(23-3)} = 2,346 - 3,457 = -1,111.$$

Визначаємо стандартні помилки вимірювання асиметрії та ексцесу для $n = 23$.

$$m_A = \sqrt{\frac{6}{n+3}} = \sqrt{\frac{6}{23+3}} = 0,48;$$

$$m_E = 2\cdot\sqrt{\frac{6}{n+3}} = 2\cdot\sqrt{\frac{6}{23+3}} = 0,96.$$

Згідно з критерієм М.А. Плохинського показники асиметрії й ексцесу свідчать про статистичну відмінність емпіричного розподілу від нормального в тому випадку, якщо вони перевищують за абсолютною величиною свої стандартні помилки вимірювання в 3 рази.

$$t_A = \frac{|A|}{m_A} = \frac{|-0,153|}{0,48} = 0,32;$$

$$t_E = \frac{|E|}{m_E} = \frac{|-1,111|}{0,96} = 1,16.$$

За результатами розрахунків встановлено, що значення коефіцієнтів асиметрії та ексцесу не перевищують утричі свої стандартні помилки вимірювання. Це дає підстави вважати, що емпіричний розподіл не

відхиляється від нормального. Таким чином, є підстави для застосування параметричного критерію *t*-Стьюдента для однієї вибірки.

2. Встановлюємо рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: рівень розвитку ПТСР у ліквідаторів Чорнобильської катастрофи не перевищує нормативний показник в 35 балів ($\mu \leq A$);

H_1: рівень розвитку ПТСР у ліквідаторів Чорнобильської катастрофи перевищує нормативний показник в 35 балів ($\mu > A$).

3. Обчислюємо емпіричне значення критерію *t*-Стьюдента для однієї вибірки.

$$t = \frac{\bar{x} - A}{s/\sqrt{n}} = \frac{36 - 35}{4{,}09/\sqrt{23}} = 1{,}173.$$

4. За таблицею критичних значень критерію Стьюдента (додаток 3) для $df = n - 1 = 23 - 1 = 22$ та заданого $\alpha = 0{,}05$ знаходимо $t_{крит.} = 1{,}72$ (однобічна критична область).

Оскільки $|t_{емп.}| < t_{крит.}$ (1,173 < 1,72), гіпотеза H_0 приймається.

5. Визначаємо розмір ефекту за індексом d_{os}-Коена:

$$d_{os} = \frac{|\bar{x} - A|}{s} = \frac{|36 - 35|}{4{,}09} = 0{,}24.$$

Розмір стандартизованого ефекту – малий.

Висновок. Критерій *t*-Стьюдента для однієї вибірки не виявив достатніх підстав для ствердження про статистично значуще перевищення середнього рівня розвитку посттравматичного стресового розладу (ПТСР) у ліквідаторів Чорнобильської катастрофи ($\bar{x} = 36$; $s = 4{,}09$) порівняно з нормативним значенням у 35 балів ($t_{(22)} = 1{,}173$; $p > 0{,}05$; однобічна критична область). Згідно з інтерпретацією Дж. Коена, розмір ефекту є малим ($d_{os} = 0{,}24$).

– відео-огляд процедури аналізу даних Прикладу 4.3. у статистичній програмі SPSS Statistics 23.0. та інтерпретація результатів дослідження.

4.2. Теоретичні засади та сфера застосування дисперсійного аналіз даних (ANOVA)

Дисперсійний аналіз (ANOVA, від англ. *Analysis of Variance*) є одним із найбільш поширених методів статистичного аналізу в психології. Подібно до *t*-критерію Стьюдента, він дозволяє оцінити відмінності між середніми

значеннями двох груп. Водночас його перевагою є можливість порівнювати не лише дві, а й три та більше груп досліджуваних.

Вперше дисперсійний аналіз був запропонований американським статистиком Рональдом Фішером у 1925 році для обробки результатів експериментальних досліджень. Відповідно, різні модифікації дисперсійного аналізу відображають найбільш поширені плани організації експериментів.

Суть дисперсійного аналізу полягає у поділі (аналізі) дисперсії однієї або кількох змінних на складові компоненти. Порівнюючи ці компоненти між собою за допомогою *F*-критерію, можна визначити їхній внесок у загальну варіацію даних. Оскільки дисперсійний аналіз дає змогу оцінювати різні експериментальні взаємодії, його застосування потребує ретельного планування емпіричного дослідження.

Типова схема експерименту зводиться до вивчення впливу незалежної змінної (однієї або кількох) на залежну змінну. *Незалежна змінна* (Independent Variable) – це показник, що представлений в номінативній вимірювальній шкалі, що має дві або більше градацій (також називається фактором). Кожній градації незалежної змінної відповідає вибірка об'єктів (досліджуваних), для яких визначаються значення залежної змінної. *Залежна змінна* (Dependent Variable) в експериментальному дослідженні розглядається як така, що змінюється під впливом незалежної змінної або змінних. У моделі ANOVA залежна змінна має бути представлена у метричній вимірювальній шкалі.

Залежно від співвідношення вибірок, що відповідають різним градаціям (рівням) фактора, розрізняють два типи незалежних змінних (факторів). Міжгруповий фактор: його рівням відповідають незалежні вибірки об'єктів. Внутрішньогруповий фактор: його рівням відповідають залежні вибірки, найчастіше – повторні вимірювання залежної змінної на одній і тій самій вибірці.

Залежно від типу експериментального плану виділяють чотири основні варіанти дисперсійного аналізу (ANOVA): однофакторний, багатофакторний, дисперсійний аналіз із повторними вимірюваннями та багатовимірний дисперсійний аналіз (MANOVA).

Однофакторний дисперсійний аналіз (One-Way ANOVA) застосовується для дослідження впливу одного фактора на залежну змінну. У цьому випадку перевіряється одна гіпотеза щодо наявності впливу фактора на залежну змінну.

Багатофакторний дисперсійний аналіз (дво-, три- і більше факторний ANOVA; 2-Way, 3-Way тощо) використовується для аналізу впливу двох і більше незалежних змінних (факторів) на одну залежну змінну. Багатофакторний ANOVA дозволяє перевіряти не лише гіпотези щодо впливу кожного фактора окремо, але й гіпотези щодо взаємодії факторів. Наприклад, у двофакторному ANOVA перевіряються три гіпотези: а) про вплив першого фактора; б) про вплив другого фактора; в) про взаємодію між

факторами (тобто про залежність ступеня впливу одного фактора від рівнів іншого фактора).

Дисперсійний аналіз із повторними вимірюваннями (Repeated Measures ANOVA) застосовується в тих випадках, коли принаймні один із факторів варіюється за внутрішньогруповим планом, тобто різним рівням цього фактора відповідає одна й та сама вибірка об'єктів (досліджуваних). У моделі дисперсійного аналізу з повторними вимірюваннями розрізняють внутрішньогрупові та міжгрупові фактори. У разі двофакторного ANOVA з повторними вимірюваннями за одним із факторів перевіряються три гіпотези: а) щодо впливу внутрішньогрупового фактора; б) щодо впливу міжгрупового фактора; в) щодо взаємодії між внутрішньогруповим і міжгруповим факторами.

Багатовимірний дисперсійний аналіз (Multivariate ANOVA, MANOVA) використовується тоді, коли залежна змінна є багатовимірною, тобто представлена декількома (або багатьма) вимірами досліджуваного явища (властивості).

Вимоги до даних при застосуванні дисперсійного аналізу:

1. Розподіл залежної змінної для кожної градації фактора має відповідати нормальному закону;
2. Дисперсії вибірок, що відповідають різним градаціям фактора, мають бути однаковими (гомогенними);
3. Вибірки, що відповідають різним градаціям фактора, повинні бути незалежними (для міжгрупового фактора).

Якщо дані не відповідають закону нормального розподілу, то необхідно застосовувати непараметричні аналоги дисперсійного аналізу: критерій *H*-Краскела-Уолліса в ситуації незалежних вибірок та χ_r^2-Фрідмана для повторних вимірювань.

Якщо дисперсії вибірок гетерогенні (не однакові), то необхідно застосовувати критерій *Уелша* (Welch's ANOVA) або критерій Браун-Форсайта (Brown-Forsythe test). Вони дозволяють проводити дисперсійний аналіз навіть при значній різниці внутрішньогрупових дисперсій.

Основна гіпотеза дисперсійного аналізу стверджує, що всі вибірки сформовано з однієї генеральної сукупності, тобто їхні середні значення однакові. Вона виражається наступним чином:

H_0: $\mu_{заг.} = \mu_1 = \mu_2 = \mu_3$

Альтернативною гіпотезою (H_1) буде припущення про те, що принаймні одне середнє відрізняється від інших.

Необхідно враховувати, що дисперсійний аналіз дає можливість встановити наявність загальних відмінностей між середніми значеннями досліджуваних груп, проте не визначає, між якими саме існують статистично значущі відмінності. З цією метою застосовувати попарні порівняння груп критерієм *t*-Стьюдента некоректно, оскільки це призводить до зростання ймовірності помилки першого роду через багаторазове тестування (*p*-значення збільшується пропорційно кількості перевірених гіпотез).

Для цього необхідно використовувати методи множинного порівняння. Вони застосовуються після дисперсійного аналізу, коли встановлено, що існують статистично значущі відмінності між групами. Існують такі методи перевірки апостеріорних гіпотез (після аналізу даних):

1) Метод Тьюкі (Tukey HSD) – дозволяє робити попарні порівняння середніх для всіх груп, контролюючи ймовірність помилки першого роду (α).

2) Метод Шеффе (Scheffé test) – вважається одним з найбільш «консервативних» методів, знижує ймовірність хибнопозитивних результатів, але менш чутливий.

3) Метод Данкана (Duncan's multiple range test) – менш строгий, ніж Тьюкі, виявляє більше відмінностей, але з вищим ризиком помилки.

4) Метод Ньюмана-Кеулса (Newman–Keuls test) – схожий на метод Тьюкі, але менш строгий.

5) Метод Бонферроні (Bonferroni correction) – корекція рівня значущості шляхом поділу α на кількість порівнянь. Простий у використанні, але зменшує статистичну потужність.

6) Метод Холма-Бонферроні (Holm-Bonferroni) – більш «м'яка» модифікація Бонферроні, яка зберігає потужність тесту.

Щоб дізнатися, наскільки сильний вплив рівнів прояву ознак на залежну змінну корисно визначити розмір статистичного ефекту. Для цього існує декілька індексів розміру ефекту: η^2 (ета-квадрат) або ω^2 (омега-квадрат). На практиці частіше використовують η^2, його простіше інтерпретувати, він відображає частку дисперсії залежної змінної, яка пов'язана з відмінностями незалежної змінної. Пропонуються наступні орієнтовні межі для оцінки величини розміру ефекту:

- $0{,}00 \le \eta^2 < 0{,}01$ – несуттєвий;
- $0{,}01 \le \eta^2 < 0{,}06$ – малий;
- $0{,}06 \le \eta^2 < 0{,}14$ – середній;
- $0{,}14 \le \eta^2$ – великий.

4.2.1. Однофакторний дисперсійний аналіз

Розглянемо загальні принципи та послідовність обчислень для однофакторного дисперсійного аналізу у випадку рівної чисельності порівнюваних вибірок.

Вихідна ідея дисперсійного аналізу полягає у можливості поділу варіативності ознаки на дві складові: внутрішньогрупову та міжгрупову.

В якості показника варіативності використовується сума квадратів відхилень значень ознаки від середнього, яка позначається *SS* (Sum of Squares).

Загальна (Total) сума квадратів (SS_{total}) є показником загальної варіативності залежної змінної і представляє чисельник дисперсії:

$SS_{total} = \sum_{i=1}^{N} (x_i - \bar{x})^2$, де

x_i – первинні значення;

$\bar{x}_1$, – середнє арифметичне змінних.

Відповідно, загальна сума квадратів дорівнює сумі міжгрупової та внутрішньогрупової сум квадратів**:**

$$SS_{total} = SS_{wg} + SS_{bg}$$

Міжгрупова (Between-Group) сума квадратів (SS_{bg}) – показник варіативності між *k* групами (кожна чисельністю *n* об'єктів):

$$SS_{bg} = \sum_{j=1}^{k} n_j (x_j - \bar{x})^2, \text{ де}$$

x_j – середнє значення для групи j;

n_j – чисельність групи j.

Відношення міжгрупової суми квадратів до загальної суми квадратів показує частку загальної дисперсії залежної змінної, зумовлену впливом фактора. Це і є показником розміру ефекту – η^2 (ета-квадрат).За змістом цей показник ідентичний квадрату коефіцієнта кореляції, тому його також називають коефіцієнтом детермінації (R^2):

$$\eta^2 = R^2 = \frac{SS_{bg}}{SS_{total}}$$

Коефіцієнт детермінації може набувати значень від 0 до 1. Чим більше цей показник, тим сильніший вплив досліджуваного фактора на дисперсію залежної змінної. Помножений на 100, він виражає відсоток поясненої дисперсії**.**

Внутрішньогрупова (Within-Group) сума квадратів (SS_{wg}) — показник випадкової варіативності всередині груп**:**

$$SS_{wg} = SS_{total} - SS_{bg} = \sum_{j=1,i=1}^{k,n} (x_i - \bar{x}_j)^2$$

На величину сум сукупних квадратів впливає чисельність та кількість порівнюваних груп**.** Тому для порівняння міжгрупової та внутрішньогрупової варіативності використовують середні квадрати (позначається *MS* – від англ. Mean of Squares). Середній квадрат обчислюється як відношення суми квадратів до відповідного числа ступенів свободи**.**

Кожна сума квадратів характеризується своїм числом ступенів свободи (*df*)**.** Так, загальне число ступенів свободи**,** що відповідає загальній сумі квадратів, дорівнює: $df_{total} = N - 1$.

Число ступенів свободи для міжгрупової суми квадратів дорівнює кількості складових мінус один (кількість груп мінус 1): $df_{bg} = k - 1$.

Число ступенів свободи для внутрішньогрупової суми квадратів обчислюється як: $df_{wg} = df_{total} - df_{bg} = N - k$. (*N* – загальна кількість досліджуваних; *k* – кількість груп).

Після визначення числа ступенів свободи обчислюють середні квадрати (*MS*) за формулами:

$MS_{bg} = \frac{SS_{bg}}{df_{bg}}$ – нутрішньогруповий середній квадрат;

$MS_{wg} = \frac{SS_{wg}}{df_{wg}}$ – міжгруповий середній квадрат.

Чим більшим є співвідношення міжгрупового до внутрішньогрупового середнього квадрату, тим більше підстав вважати, що порівнювані середні значення статистично відрізняються. Відповідно, основним показником дисперсійного аналізу є *F*-співвідношення – емпіричне значення критерію Фішера:

$$F = \frac{MS_{bg}}{MS_{wg}}.$$

Для визначення рівня статистичної значущості при обчисленнях «вручну» використовуються таблиці критичних значень *F*-розподілу для направлених альтернатив **(**Додаток 4).

Якщо вплив фактора відсутній, то обчислене співвідношення не перевищить критичне значення ($F_{емп.} \leq F_{крит.}$), і в цьому випадку слід прийняти нульову гіпотезу H_0 про відсутність впливу фактора на залежну змінну. І навпаки, якщо вплив фактора статистично значущий, то ($F_{емп.} > F_{крит.}$), у такому випадку слід прийняти альтернативну гіпотезу H_1.

Приклад 4.4. Дослідник вивчає вплив різних методів навчання на рівень запам'ятовування інформації. Для цього здобувачі вищої освіти були розподілені на три групи по 5 осіб: перша навчалася у форматі лекції, друга – тренінгу, третя – шляхом самостійного опрацювання матеріалу. Після навчання рівень засвоєння знань оцінювався у балах. Необхідно за отриманими результатами визначити, чи існують статистично значущі відмінності між групами, первинні дані відповідають закону нормального розподілу.

Таблиця 4.5 – Таблиця первинних даних

№	Метод навчання		
	Лекція (X_1)	Тренінг (X_2)	Самостійне (X_3)
1	60	75	50
2	65	80	55
3	70	85	45
4	75	90	60
5	68	78	52

Згідно з умовами завдання, дані відповідають закону нормального розподілу. Необхідно порівняти результати трьох груп досліджуваних. За таких умов для аналізу даних слід застосовувати однофакторний дисперсійний аналіз.

1. Встановлюємо рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: метод навчання не впливає на рівень запам'ятовування інформації ($\mu_1 = \mu_2 = \mu_3$);

H_1: метод навчання впливає на рівень запам'ятовування інформації ($\mu_1 \neq \mu_2 \neq \mu_3$).

2. Обчислюємо середні значення запам'ятовування інформації у груп з різними методами навчання:

$\bar{x}_1 = 67{,}6$; $\bar{x}_2 = 81{,}6$; $\bar{x}_3 = 52{,}4$.

Загальне середнє: $\bar{x} = 67{,}2$.

3. Обчислюємо сими квадратів:

$$SS_{total} = \sum_{i=1}^{N}(x_i - \bar{x})^2 = (60-67{,}2)^2 + (65-67{,}2)^2 + ... + (52-67{,}2)^2 = 2524{,}4$$

$$SS_{bg} = \sum_{j=1}^{k} n(x_j - \bar{x})^2 = 5 \cdot (67{,}6-67{,}2)^2 + (81{,}6-67{,}2)^2 + (52{,}4-67{,}2)^2 = 2132{,}8$$

$$SS_{wg} = SS_{total} - SS_{bg} = 2524{,}4 - 2132{,}8 = 391{,}6$$

4. Визначаємо ступені свободи:

df_{bg} = k – 1 = 3 – 1 = 2

df_{wg} = N – k = 15 – 3 =12

5. Обчислюємо середні квадрати:

$$MS_{bg} = \frac{SS_{bg}}{df_{bg}} = \frac{2132{,}8}{2} = 1066{,}4$$

$$MS_{wg} = \frac{SS_{wg}}{df_{wg}} = \frac{391{,}6}{12} = 32{,}63$$

6. Розраховуємо емпіричне значення F:

$$F = \frac{MS_{bg}}{MS_{wg}} = \frac{1066{,}4}{32{,}63} = 32{,}68$$

7. За таблицею критичних значень критерію F-Фішера (додаток 4) для $df_{чис.} = 2$, $df_{знам.} = 12$ та заданого $\alpha = 0{,}05$ знаходимо $F_{крит.} = 5{,}10$.

Оскільки $F_{емп.} > F_{крит.}$ (32,68 > 5,10), гіпотеза H_0 відхиляється.

8. Визначаємо розмір ефекту за індексом η^2:

$$\eta^2 = R^2 = \frac{SS_{bg}}{SS_{total}} = \frac{2132{,}8}{2524{,}4} = 0{,}84$$

Значення 0,84 для індексу η^2 відповідає дуже великому рівню розміру ефекту.

Висновок. Застосування однофакторного дисперсійного аналізу дало підстави для відхилення нульової гіпотези ($F_{(2,\ 12)} = 32{,}68$; $p < 0{,}05$). Це свідчить, шо рівень запам'ятовування інформації здобувачами вищої освіти статистично значуще відрізняється залежно від методу навчання. Розмір ефекту є великим ($\eta^2 = 0{,}84$).

Завдання на самостійну підготовку

1. Вимоги до даних при застосуванні параметричних критеріїв статистичного порівняння вибірок досліджуваних.
2. Поясніть принцип *t*-Стьюдента для незалежних вибірок.
3. Охарактеризуйте *t*-Стьюдента для залежних (парних) вибірок.
4. Визначте сфери застосування *t*-Стьюдента для однієї вибірки.
5. Поясніть, чому індекс розміру ефекту *d*-Коена вважають показником практичної значущості результатів.
6. Розкрийте основні можливості методу дисперсійного аналізу даних.
7. Проаналізуйте види дисперсійного аналізу даних.
8. Опишіть проблему множинного порівняння вибірок досліджуваних.
9. Особливості інтерпретації та представлення результатів дисперсійного аналізу.
10. Характеристика однофакторного дисперсійного аналізу.
11. Характеристика двофакторного дисперсійного аналізу.
12. Характеристика ANCOVA.
13. Характеристика дисперсійного аналізу з повторними вимірюваннями.
14. Характеристика багатовимірного дисперсійного аналізу.

РОЗДІЛ 5
НЕПАРАМЕТРИЧНІ МЕТОДИ СТАТИСТИЧНОГО ПОРІВНЯННЯ ВИБІРОК ДОСЛІДЖУВАНИХ

5.1. Непараметричні критерії порівняння ознак

Критерії порівняння ознак – це група статистичних методів, що застосовуються для перевірки, чи існують статистично значущі відмінності між двома або більше вибірками за певною ознакою (змінною). Вони дають змогу визначити, чи відмінності, виявлені в емпіричних даних, зумовлені випадковими коливаннями, чи мають систематичний характер.

Непараметричні критерії – це статистичні методи аналізу, які не висувають жорстких вимог до розподілу даних (наприклад, нормальності чи однаковості дисперсій). Їх ще називають ранговими критеріями, оскільки вони базуються на порівнянні порядкових характеристик (рангів), а не на самих числових значеннях. Вони особливо корисні, коли: дані представлені у номінальній чи порядковій шкалі; вибірки мають малий обсяг; дані суттєво відхиляються від нормального розподілу.

5.1.1. *U*-критерій Манна-Уітні

U-критерій вперше був запропонований Ф. Вілкоксоном у 1945 році для аналізу статистичної значущості відмінностей між двома незалежними вибірками однакового обсягу. Згодом, у 1947 році, Г.Б. Манн і Д.Р. Уітні модифікували цей критерій для застосування до вибірок різної чисельності. У сучасній статистичній практиці існує кілька підходів до обчислення *U*-критерію, що зумовлює наявність різних варіантів таблиць критичних значень. Тому під час аналізу результатів слід уважно враховувати спосіб розрахунку, який використовується. Також у літературі даний критерій зустрічається під різними назвами: критерій Манна-Уітні-Вілкоксона (скорочено – MWW), критерій Вілкоксона-Манна-Уітні, критерій рангових сум Вілкоксона тощо.

Критерій Манна-Уітні є непараметричним, тобто не вимагає дотримання припущень щодо нормального розподілу даних та гомогенності дисперсій у вибірках. Завдяки цьому він є менш вимогливим у порівнянні з параметричним аналогом – критерієм t-Стьюдента для незалежних вибірок. Водночас слід зазначити, що *U*-критерій є менш чутливим, тобто має нижчу статистичну потужність, особливо при малих вибірках або слабкому ефекті.

Вимоги до застосування критерію:

1. Вибірки дослідження повинні бути незалежними.
2. Дані повинні бути виміряні принаймні на рівні порядкової шкали та варіюватися в достатньо широкому діапазоні значень.
3. У кожній вибірці повинно бути щонайменше три спостереження. Якщо в одній вибірці 2 особи, то в іншій має бути мінімум 5.

Емпіричне значення *U*-критерію Манна-Уітні визначається за формулою:

$$U = (n_1 \cdot n_2) + \frac{n_x \cdot (n_x + 1)}{2} - T_x, \text{ де}$$

n_1 – кількість досліджуваних у першій групі;
n_2 – кількість досліджуваних у другій групі;
n_x – кількість досліджуваних у групі з більшою сумою рангів;
T_x – більша із двох рангових сум.

Якщо емпіричне значення *U* дорівнює або менше критичного значення для заданого α-рівня, то H_0 відхиляється і формулюється висновок, що відмінності між вибірками статистично значущі.

Для оцінки розміру ефекту для критерію *U*-Манна-Уітні існує декілька формул.

Найбільш часто в спеціальній літературі вказується, що значення розміру ефекту оцінюється через *z*-показник. Стандартизована статистика критерію ділиться на квадратний корінь розміру вибірки.

$$r_{effect} = \frac{z}{\sqrt{n}}, \text{ де}$$

z – стандартизований показник для *U*-значення;
n – загальний обсяг досліджуваних.

Б. Кінг та Е. Мініум запропонували використовувати вираз, що базується на рангово-бісеріальній кореляції:

$$r_{effect} = \frac{2 \cdot (\overline{R}_1 - \overline{R}_2)}{n_1 + n_2}, \text{ де}$$

$\overline{R}_1$, $\overline{R}_2$ – середній ранг в першій та другій вибірках;
n_1, n_2 – обсяг досліджуваних у першій та другій вибірках.

За необхідності існує можливість застосовувати альтернативну формулу через емпіричне значення критерію *U*-Манна-Уітні, що приводить до того самого результату:

$$r_{effect} = 1 - \frac{2 \cdot U}{n_1 \cdot n_2}, \text{ де}$$

U – емпіричне значення *U*-критерію Манна-Уітні;
n_1, n_2 – обсяг досліджуваних у першій та другій вибірках.
n – загальний обсяг досліджуваних в обох вибірках.

Інтерпретація рівня розміру ефекту за *r*-індексом (відповідно до рекомендацій Дж. Коена):

$0{,}00 \le |r| < 0{,}10$ – несуттєвий;
$0{,}10 \le |r| < 0{,}30$ – малий;
$0{,}30 \le |r| < 0{,}50$– середній;
$0{,}50 \le |r| < 1{,}00$ – великий.

Приклад 5.1. У трудовому колективі було здійснено оцінку рівня мотивації досягнення в кожного працівника. Постає питання: чи існують

статистично значущі відмінності між чоловіками та жінками щодо прояву мотивації досягнення?

Таблиця 5.1 – Таблиця первинних даних та проміжних розрахунків

№	Чоловіки (X_1)	Жінки (X_2)	Rx_1	Rx_2
1	8	10	11,5	14,5
2	4	5	3	6
3	11	12	17,5	20,5
4	5	7	6	9,5
5	6	8	8	11,5
6	7	10	9,5	14,5
7	3	4	1	3
8	9	11	13	17,5
9	11	11	17,5	17,5
10	4	5	3	6
11	12		20,5	
			$\Sigma = 110,5$	$\Sigma = 120,5$

Дві вибірки в дослідженні є незалежними. Оскільки відсутні підстави вважати, що первинні дані за обома групами відповідають закону нормального розподілу, а також через недостатній обсяг вибірок, який не дозволяє достовірно оцінити нормальність розподілу, застосування параметричних критеріїв є недоцільним.

У зв'язку з цим для виявлення статистично значущих відмінностей між групами буде використано непараметричний U-критерій Манна-Уітні, який не потребує дотримання припущення про нормальність розподілу.

1. Встановлюємо рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0,05$;

H_0: рівень мотивації чоловіків не відрізняється від рівня мотивації в жінок;

H_1: рівень мотивації чоловіків відрізняється від рівня мотивації в жінок.

2. Дані обох груп об'єднуємо в один ряд та присвоюємо кожному значенню ранг за порядком зростання. Якщо є декілька однакових числових значень, то використовується правило зв'язаних рангів.

3. Підраховуємо суми рангів для обох груп. $\Sigma Rx_1 = 110,5$; $\Sigma Rx_2 = 120,5$.

Для перевірки правильності ранжування можна використовувати той факт, що $\Sigma_R = n \cdot (n+1)/2$. У нашому випадку сума рангів усієї вибірки досліджуваних дорівнює 231 ($\Sigma_R = 231$), $n \cdot (n+1)/2 = 21 \cdot (21+1)/2 = 231$. Тотожність підтвердилась (231 =231).

4. Визначаємо більшу із двох рангових сум. T_x =120,5.

5. Обчислюємо емпіричне значення критерію *U*-Манна-Уітні за формулою:

$$U = (n_1 \cdot n_2) + \frac{n_x \cdot (n_x + 1)}{2} - T_x = (11 \cdot 10) + \frac{10 \cdot (10 + 1)}{2} - 120{,}5 = 44{,}5.$$

6. За таблицею критичних значень *U*-критерію Манна-Уітні для ненаправлених альтернатив (додаток 5) для $df_1 = n_1 = 11$, $df_2 = n_2 = 10$ та заданого $\alpha = 0{,}05$ знаходимо $U_{крит.} = 26$.

Оскільки $U_{емп.} > U_{крит.}$ (44,5 > 26), гіпотеза H_0 приймається.

1. Визначаємо розмір ефекту.

$$r_{effect} = \frac{z}{\sqrt{n}}$$

$$z = \frac{U - \frac{n_1 \cdot n_2}{2}}{\sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}}} = \frac{44{,}5 - \frac{11 \cdot 10}{2}}{\sqrt{\frac{10 \cdot 11(11 + 10 + 1)}{12}}} = -0{,}74.$$

$$r_{effect} = \frac{z}{\sqrt{n}} = \frac{-0{,}74}{\sqrt{21}} = -0{,}16.$$ (знак не інтерпретується)

Згідно методичних рекомендацій Дж. Коена розмір ефекту відповідає малому рівню.

Висновок. Представлені дані не дають підстав відхилити нульову гіпотезу критерієм *U*-Манна-Уітні ($U_{(11,\ 10)} = 44{,}5$; $p > 0{,}05$; двобічна критична область). Отже, відсутні статистично значущі відмінності в середніх рангах прояву мотивації досягнення між чоловіками ($\overline{R} = 10{,}05$) та жінками ($\overline{R} = 12{,}05$). Розмір стандартизованого ефекту відповідає малому рівню ($r_{effect} = 0{,}16.$).

– відео-огляд процедури аналізу даних Прикладу 5.1. у статистичній програмі SPSS Statistics 23.0. та інтерпретація результатів дослідження.

5.1.2. *H*-критерій Краскела-Уолліса

Непараметричний *H*-критерій Краскела-Уолліса застосовується для оцінювання наявності статистично значущих відмінностей між трьома і більше незалежними вибірками за рівнем прояву певної ознаки. У науковій літературі цей критерій також називають непараметричним аналогом дисперсійного аналізу (ANOVA).

H-критерій дозволяє з'ясувати, чи змінюється рівень досліджуваної ознаки при переході від однієї групи до іншої. Однак він не дає інформації про напрямок та локалізацію виявлених відмінностей, тобто не дозволяє визначити, яка саме група статистично значуще відрізняється від інших.

У випадку виявлення загальної статистичної значущості застосовувати попарні порівняння за допомогою *U*-критерію Манна-Уітні некоректно, оскільки це призводить до зростання ймовірності помилки першого роду через багаторазове тестування. Для контролю рівня цієї помилки використовуються непараметричні методи множинного порівняння. Зокрема за умови однакових обсягів вибірок застосовується непараметричний варіант критерію Ньюмена-Кейлса; за умови відмінностей у чисельності груп – критерій Данна.

Якщо є необхідність виявити тенденції зміни величин ознаки при переході від однієї вибірки до іншої порівнюючи три та більше вибірок необхідно використати критерій тенденцій *S*-Джонкхієра-Терпстри. Цей критерій можна розглядати як продовження тесту *H*-Краскела-Уолліса, оскільки він не лише констатує відмінності, а й зазначає напрямок змін.

Вимоги до застосування критерію:

1. Вибірки мають бути незалежними.

2. Первинні дані мають бути виміряні щонайменше на рівні порядкової шкали.

Емпіричне значення *H*-критерію Краскела-Уолліса обчислюється за наступною формулою:

$$H = \frac{12}{n(n+1)} \cdot \Sigma \frac{R_i^2}{n_i} - 3(n+1), \text{ де}$$

n – загальна кількість досліджуваних у всіх вибірках;

n_i – обсяг вибірки i;

R_i – сума рангів для вибірки i.

У разі наявності зв’язаних (однакових) рангів загальна формула *H*-критерію Краскела-Уолліса дає дещо неточне значення. Тому для корекції необхідно застосувати поправку на однакові ранги. Скоригована формула має вигляд:

$$H = \frac{\frac{12}{n(n+1)} \cdot \Sigma \frac{R_i^2}{n_i} - 3(n+1)}{1 - \frac{\Sigma T_i}{n^3 - n}}, \text{ де}$$

$$T_i = \left(t_i^3 - t_i\right);$$

t_i – розмір i-групи зв’язаних рангів.

Для прийняття рішення про рівень статистичної значущості необхідно емпіричне (розрахункове) значення критерію H порівняти з критичним (табличним) значенням.

Якщо кількість вибірок $k \geq 3$, $n \geq 6$, то користуємося таблицею критичних значень χ^2, $df = k - 1$, (додаток 2).

Якщо $k = 3$, $n \leq 5$, то користуємося таблицею критичних значень *H*-Краскела-Уолліса (додаток 7).

Якщо емпіричне значення дорівнює або більше критичного значення для заданого α-рівня, то H_0 відхиляється і формулюється висновок, що відмінності між вибірками статистично значущі.

Для оцінки сили впливу рівня прояву ознаки на залежну змінну корисно визначати розмір статистичного ефекту. Однак для *H*-критерію Краскела-Уолліса відсутній єдиний узгоджений метод розрахунку. У науковій практиці використовують два основні індекси ефекту: ε^2 (епсилон-квадрат) та η^2 (ета-квадрат). Індекс η^2 є більш консервативним.

$$\varepsilon^2 = \frac{H}{\left(n^2-1\right)/\left(n+1\right)}, \text{ де}$$

H – емпіричне значення *H*-критерію Краскела-Уолліса;

n – загальна кількість досліджуваних у всіх вибірках.

$$\eta^2 = \frac{H-k+1}{n-k}, \text{ де}$$

H – емпіричне значення *H*-критерію Краскела-Уолліса;

k – кількість груп;

n – загальна кількість досліджуваних у всіх вибірках.

В літературних джерелах пропонуються наступні орієнтовні межі для оцінки величини статистичного ефекту для індексу ε^2:

- $0{,}00 \le \varepsilon^2 < 0{,}01$ – несуттєвий;
- $0{,}01 \le \varepsilon^2 < 0{,}08$ – малий;
- $0{,}08 \le \varepsilon^2 < 0{,}26$ – середній;
- $0{,}26 \le \varepsilon^2$ – великий.

Приклад 5.2. У здобувачів вищої освіти різних навчальних спеціальностей на 4 курсі навчання визначили інтегральну оцінку професійної ідентичності. Чи відрізняється професійна ідентичність (ПІ) у студентів різних спеціальностей?

Таблиця 5.2 – Таблиця первинних даних та проміжних розрахунків

№	Психологи		Політологи		Соціологи		Журналісти	
	ПІ	Ранг	ПІ	Ранг	ПІ	Ранг	ПІ	Ранг
1	69	16	62	14	29	2	71	18
2	89	29	45	6	76	21	94	32
3	91	31	77	22	70	17	100	35
4	83	27	44	5	42	4	98	33
5	72	19	54	12	51	9	84	28
6	46	7	31	3	25	1	82	26
7	52	10	73	20	53	11	79	23
8	99	34	81	25	67	15		
9	90	30			48	8		
10	58	13						
11	80	24						
		Σ = 240		Σ = 107		Σ = 88		Σ = 195

Первинні дані представлені в шкалі порядку. Вибірки в дослідження незалежні. Необхідно оцінити відмінність одночасно між чотирма вибірками за рівнем досліджуваної ознаки. Для вирішення даного завдання розроблений критерій *H*-Краскела-Уолліса.

1. Встановлюємо рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: здобувачі вищої освіти різних спеціальностей не відрізняються за параметром професійної ідентичності;

H_1: здобувачі вищої освіти різних спеціальностей відрізняються за параметром професійної ідентичності.

2. Дані всіх груп досліджуваних об'єднуємо у один ряд та присвоюємо кожному числу ранг за порядком зростання. Якщо є декілька однакових числових значень, то використовується правило зв'язаних рангів.

3. Підраховуємо суми рангів для всіх груп. $R_1 = 240$; $R_2 = 107$; $R_3 = 88$; $R_4 = 195$.

Для перевірки правильності ранжування можна використовувати той факт, що $\Sigma_R = n\cdot(n+1)/2$. У нашому випадку сума рангів всієї вибірки досліджуваних дорівнює 630 ($\Sigma_R = 630$), $n\cdot(n+1)/2 = 35\cdot(35+1)/2 = 630$. Тотожність підтвердилась (630 = 630).

4. У нашому прикладі відсутні зв'язані ранги. Емпіричне значення критерію *H*-Краскела-Уолліса будемо визначати за загальною формулою:

$$H = \frac{12}{n(n+1)}\cdot\Sigma\frac{R_i^2}{n_i} - 3(n+1).$$

$$H = \frac{12}{35\cdot(35+1)}\cdot\left[\frac{240^2}{11}+\frac{107^2}{8}+\frac{88^2}{9}+\frac{195^2}{7}\right] - 3\cdot(35+1) = 15{,}43.$$

5. Так як, $k \geq 3$, $n \geq 6$ при прийнятті рішення про рівень статистичної значущості використовуємо таблицю критичних значень χ^2 (додаток 2). За даною таблицею для $df = k - 1 = 4 - 1 = 3$ та заданого $\alpha = 0{,}05$ знаходимо $\chi^2_{крит.} = 7{,}815$.

Оскільки $\chi^2_{емп.} > \chi^2_{крит.}$ (15,43 > 7,815), гіпотеза H_0 відхиляється.

6. Оцінюємо величину розміру статистичного ефекту за індексом ε^2:

$$\varepsilon^2 = \frac{H}{(n^2-1)/(n+1)} = \frac{15{,}43}{(35^2-1)/(35+1)} = 0{,}454.$$

Розмір стандартизованого ефекту – великий.

Висновок. Критерій *H*-Краскела-Уолліса дозволив відхилити нульову гіпотезу ($H_{(3)} = 15{,}43$; $p < 0{,}05$). Існує статистично значуща відмінність в середніх рангах параметра професійної ідентичності у здобувачі вищої освіти різних спеціальностей. Розмір ефекту – великий ($\varepsilon^2 = 0{,}454$).

– відео-огляд процедури аналізу даних Прикладу 5.2. у статистичній програмі SPSS Statistics 23.0. та інтерпретація результатів дослідження.

Приклад 5.3. У колективах з різним стилем керівництва було виміряно рівень особистісної відповідальності у працівників. Чи залежить рівень особистісної відповідальності від стилю керівництва?

Таблиця 5.3 – Таблиця первинних даних та проміжних розрахунків

№	Авторитарний стиль керівництва		Демократичний стиль керівництва		Ліберальний стиль керівництва	
	Рівень відповідал.	Ранг	Рівень відповідал.	Ранг	Рівень відповідал.	Ранг
1	12	18,5	11	16	10	13
2	10	13	8	6	9	9,5
3	12	18,5	8	6	9	9,5
4	9	9,5	14	20	11	16
5	8	6	7	4	6	2,5
6	11	16	10	13	6	2,5
7	15	21	5	1		
8			9	9,5		
		Σ = 102,5		Σ = 75,5		Σ = 53

Первинні дані представлені у порядковій шкалі. Завдання полягає у порівнянні рівня розвитку ознаки між трьома незалежними вибірками. У такому випадку доцільно застосувати *H*-критерій Краскела-Уолліса.

1. Встановлюємо рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

α = 0,05;

H_0: рівень особистісної відповідальності не залежить від стилю керівництва;

H_1: рівень особистісної відповідальності залежить від стилю керівництва.

2. Дані всіх груп досліджуваних об’єднуємо в один ряд та присвоюємо кожному значенню ранг за порядком зростання. Якщо є декілька однакових числових значень, то використовується правило зв’язаних рангів.

3. Підраховуємо суми рангів для всіх груп. R_1 = 102,5; R_2 = 75,5; R_3 = 53.

Для перевірки правильності ранжування можна використовувати той факт, що $\Sigma_R = n\cdot(n+1)/2$. У нашому випадку сума рангів усієї вибірки

досліджуваних дорівнює 231 (Σ_R = 229), $n\cdot(n+1)/2 = 21\cdot(21+1)/2 = 231$. Тотожність підтвердилась (231 = 231).

4. У наведеному прикладі наявні зв'язані ранги. Емпіричне значення статистики критерію Краскела-Уолліса обчислюється за скоригованою формулою з урахуванням зв'язаних рангів:

$$H = \frac{\frac{12}{n(n+1)}\cdot\Sigma\frac{R_i^2}{n_i} - 3(n+1)}{1-\frac{\Sigma T_i}{n^3-n}},\ \text{де}$$

$T_i = \left(t_i^3 - t_i\right)$;

t_i – розмір i-групи зв'язаних рангів.

Таблиця 5.4 – Розмір груп зв'язаних рангів

Рівень емпатії	6	8	9	10	11	12
t_i	2	3	4	3	3	2
$T_i = \left(t_i^3 - t_i\right)$	6	24	60	24	24	6

$$H = \frac{\frac{12}{21\cdot(21+1)}\cdot\left[\frac{102{,}5^2}{7}+\frac{75{,}5^2}{8}+\frac{53^2}{6}\right]-3\cdot(21+1)}{1-\frac{(6+24+60+24+24+6)}{21^3-21}} = 3{,}71.$$

5. Так як, $k \geq 3$, $n \geq 6$ при прийнятті рішення про рівень статистичної значущості використовуємо таблицю критичних значень χ^2 (додаток 2). За даною таблицею для $df = k - 1 = 3 - 1 = 2$ та заданого $\alpha = 0{,}05$ знаходимо $\chi^2_{крит.} = 5{,}991$.

Оскільки $\chi^2_{емп.} < \chi^2_{крит.}$ (3,71 < 5,991), гіпотеза H_0 приймається.

6. Оцінюємо величину розміру статистичного ефекту за індексом ε^2:

$$\varepsilon^2 = H\cdot\frac{n+1}{n^2-1} = 3{,}71\cdot\frac{21+1}{21^2-1} = 0{,}186.$$

Розмір стандартизованого ефекту – середній.

Висновок. Рівень особистісної відповідальності у працівників колективу не залежить від стилю керівництва. Згідно статистики критерію H-Краскела-Уолліса ($H_{(2)}$ = 3,71; p > 0,05) для відхилення нульової гіпотези немає достатніх підстав, проте середній розмір ефекту (ε^2 = 0,186) свідчить про доцільність збільшення статистичної потужності дослідження.

5.2. Критерії розпізнавання зсувів

Критерії розпізнавання зсувів – це група непараметричних статистичних методів, призначених для перевірки, чи існує систематичне зміщення (зсув) між двома або кількома пов'язаними вибірками (наприклад, до і після експериментального впливу). Інакше кажучи, ці критерії

допомагають визначити, чи має місце тенденція до збільшення або зменшення значень у парних спостереженнях.

5.2.1. Критерій *T*-Вілкоксона

Критерій *T*-Вілкоксона використовують для порівняння показників, виміряних у двох різних умовах на одній і тій самій вибірці досліджуваних. Він дозволяє встановити не лише спрямованість змін, але й їхню виразність, тобто довести, що зсув показників у одному напрямку є більш інтенсивним, ніж у іншому.

Критерій *T*-Вілкоксона належить до непараметричних методів статистичного аналізу, а отже, його застосування не передбачає дотримання вимог щодо нормальності розподілу та рівності дисперсій. У цьому аспекті він є менш вимогливим порівняно з параметричним аналогом – критерій *t*-Стьюдента для залежних вибірок. Водночас критерій *T*-Вілкоксона характеризується нижчою чутливістю, ніж *t*-критерій, що слід враховувати під час вибору відповідного статистичного методу для аналізу даних.

Сутність методу полягає у впорядкуванні змін між двома вимірюваннями за їхніми абсолютними величинами. У разі, якщо зсуви в бік збільшення чи зменшення показників відбуваються випадково, суми рангів абсолютних значень у кожному з напрямів будуть приблизно однаковими. Якщо ж інтенсивність змін у певному напрямку є більш вираженою, то сума рангів абсолютних значень у протилежному напрямку виявиться суттєво меншою, ніж це можна було б очікувати за умови випадкового характеру зсувів.

Доцільність застосування критерію Т-Вілкоксона визначається у випадках, коли величини зсувів варіюють у достатньо широкому діапазоні. Якщо ж зміни є незначними та мало відрізняються між собою (наприклад, +1, –1 та 0), то через велику кількість однакових рангів процедура рангування втрачає свою ефективність. У подібних ситуаціях більш раціональним є використання критерію *G*-знаків.

Вимоги до застосування та особливості критерію:

1. Вибірка повинна складатися зі зв’язаних (залежних, парних) спостережень, тобто таких, що отримані від одних і тих самих досліджуваних у різні моменти часу або за різних умов.

2. Дані мають бути виміряні принаймні на рівні порядкової шкали та варіювати у достатньо широкому діапазоні.

3. Мінімальний обсяг вибірки становить не менше ніж 5 досліджуваних.

4. Нульові зсуви (різниця між значеннями, що дорівнює нулю) вилучаються з розгляду, при цьому фактичний обсяг вибірки n зменшується на кількість таких зсувів.

Для обчислення критерію *T*-Вілкоксона не використовуються спеціальні формули. Достатньо визначити різниці між парними

спостереженнями, проставити їм ранги за абсолютними значеннями та підрахувати суми рангів для позитивних і негативних зсувів. За емпіричне значення критерію приймається менша з отриманих сум рангів.

$T_{емп.} = \min(R_1, R_2)$, де

R_1 – сума рангів для позитивних величин різниці;

R_2 – сума рангів для негативних величин різниці.

Якщо емпіричне значення T дорівнює або менше критичного значення для заданого α-рівня, то H_0 відхиляється і формулюється висновок, що відмінності між вибірками статистично значущі.

Для оцінки розміру стандартизованого ефекту після використання критерію T-Вілкоксона рекомендується використати одну із запропонованих формул:

$$r_{effect} = \frac{4 \cdot \left| T - \left(\frac{R_1 + R_2}{2} \right) \right|}{n \cdot (n+1)}, \text{ де}$$

R_1 – сума рангів для позитивних величин різниці;

R_2 – сума рангів для негативних величин різниці;

T – менша сума рангів (R_1 або R_2);

n – обсяг досліджуваних (обсяг вибірки зменшився на кількість нульових зсувів).

Розмір ефекту для даного критерію найбільш часто оцінюється через z-показник. Стандартизована статистика критерію ділиться на квадратний корінь розміру вибірки.

$$r_{effect} = \frac{z}{\sqrt{n}}, \text{ де}$$

z – стандартизована статистика критерію T- Вілкоксона;

n – загальний обсяг досліджуваних в обох вибірках.

Дейв Керби запропонував ще один варіант розрахунку розміру стандартизованого ефекту:

$$r_{effect} = \frac{W}{\Sigma R}, \text{ де}$$

W – сума рангів з врахуванням знаку величин різниці;

ΣR – сума рангів абсолютних величин різниці.

Інтерпретація рівня розміру ефекту для запропонованих r-індексів згідно з підходом Дж. Коена:

$0{,}00 \le |r| < 0{,}10$ – несуттєвий;

$0{,}10 \le |r| < 0{,}30$ – малий;

$0{,}30 \le |r| < 0{,}50$– середній;

$0{,}50 \le |r| < 1{,}00$ – великий.

Приклад 5.4. Метою дослідження було з’ясувати, чи відбуваються зміни у стійкості уваги студентів протягом навчального дня. Оцінювання відповідних показників проводилося в одній і тій самій групі студентів двічі: перед початком першої пари та після завершення останньої.

Таблиця 5.5 – Таблиця первинних даних та проміжних розрахунків

№	До занять (X_1)	Після занять (X_2)	Різниця d_i	Ранги абсолютних величин різниці $\|d_i\|$	Ранги позитивних величин різниці d_i+	Ранги негативних величин різниці d_i-
1	23	21	2	3	3	
2	17	17	0	-		
3	23	25	-2	3		3
4	15	11	4	7	7	
5	28	21	7	11	11	
6	24	24	0	-		
7	21	16	5	9,5	9,5	
8	22	26	-4	7		7
9	17	18	-1	1		1
10	26	22	4	7	7	
11	25	20	5	9,5	9,5	
12	19	21	-2	3		3
13	13	16	-3	5		5
14	19	8	11	12	12	
				$\Sigma_R = 78$	$T_1 = 59$	$T_2 = 19$

У таблиці подано результати психологічного тестування однієї й тієї самої групи досліджуваних, яке проводилося двічі. Такий тип вибірки в дослідницькій практиці називають зв'язаною або парною. Наявних підстав вважати, що первинні дані відповідають нормальному закону розподілу, немає. Крім того, обсяг вибірки є недостатнім для отримання достовірного висновку за результатами перевірки на нормальність розподілу. У зв'язку з цим, для оцінки змін, що відбулися у показниках між першим і другим вимірюванням, доцільно застосувати непараметричний критерій *T*-Вілкоксона.

1. Встановлюємо рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: у студентів рівень стійкості уваги не змінюється впродовж навчального дня;

H_1: у студентів рівень стійкості уваги змінюється впродовж навчального дня.

2. Обчислюємо різницю між індивідуальними значеннями першого та другого вимірювання («до – після»).

3. Рангуємо абсолютні величини різниці ($|d_i|$) за порядком зростання. Якщо є декілька однакових числових значень, то використовується правило зв'язаних рангів.

Для перевірки правильності ранжування можна використовувати той факт, що $\Sigma_R = n \cdot (n+1)/2$. У нашому випадку сума рангів рівна 78 ($\Sigma R = 78$), $n = 12$ (обсяг вибірки зменшився на кількість нульових зсувів). $n \cdot (n+1)/2 = 12 \cdot (12+1)/2 = 78$. Отже, рівність реальної та розрахункової сум рангів підтверджено (78 =78).

4. Підраховуємо суми рангів окремо для позитивних та негативних величин різниці (d_i+; d_i-):

$R_1 = 59$;

$R_2 = 19$.

5. За емпіричне значення критерію $T_{емп.}$ приймається менша сума:

$T_{емп.} = 19$.

6. За таблицею критичних значень критерію *T*-Вілкоксона (додаток 8) для $df = n = 12$ (обсяг вибірки зменшився на кількість нульових зсувів) та заданого $\alpha = 0{,}05$ знаходимо $T_{крит.} = 13$ (двобічна критична область).

Оскільки $T_{емп.} > T_{крит.}$ (19 > 13), гіпотеза H_0 приймається.

7. Визначаємо розмір ефекту:

$$r_{effect} = \frac{z}{\sqrt{n}}$$

$$z = \frac{\min(T_1, T_2) - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} = \frac{19 - \frac{12(12+1)}{4}}{\sqrt{\frac{12(12+1)(2 \cdot 12+1)}{24}}} = -1{,}57.$$

$$r_{effect} = \frac{z}{\sqrt{n}} = \frac{-1{,}57}{\sqrt{12}} = -0{,}45.$$ (знак не інтерпретується)

Розмір стандартизованого ефекту – середній.

Висновок. Отримані емпіричні результати не дозволяють відхилити нульову гіпотезу за критерієм *T*-Вілкоксона ($T_{(12)} = 19$; $p > 0{,}05$; двобічна критична область), що свідчить про відсутність статистично значущих змін у рівні стійкості уваги студентів протягом навчального дня. Разом з тим, значення коефіцієнта розміру ефекту ($r_{effect} = 0{,}45$), яке перевищує середній рівень, свідчить про потенційну наявність ефекту та обґрунтовує доцільність збільшення статистичної потужності дослідження (наприклад, за рахунок розширення вибірки).

– відео-огляд процедури аналізу даних Прикладу 5.4. у статистичній програмі SPSS Statistics 23.0. та інтерпретація результатів дослідження.

5.2.2. Критерій χ_r^2-Фрідмана

Непараметричний критерій χ_r^2-Фрідмана (читається як «хі-ар-квадрат») застосовується для порівняння показників, виміряних у трьох і більше умовах на одній і тій самій вибірці досліджуваних. Цей критерій дозволяє виявити наявність загальних відмінностей між умовами, проте не дає змоги визначити, між якими саме умовами існують статистично значущі розбіжності.

Здійснювати попарні порівняння за допомогою критерію *T*-Вілкоксона у цьому випадку є некоректним, оскільки це призводить до завищення рівня статистичної значущості, тобто підвищення ймовірності помилки першого роду. З метою коректного аналізу між умовами застосовують спеціальні непараметричні методи множинного порівняння, наприклад критерій Данна.

У випадках, коли виникає потреба виявити тенденції зміни величин ознаки під час переходу від однієї дослідницької умови до іншої, доцільно застосовувати критерій тенденцій *L*-Пейджа. Він розглядається як продовження критерію Фрідмана, оскільки він не лише фіксує наявність статистично значущих відмінностей між умовами, але й дозволяє визначити напрямок виявлених змін.

Критерій Фрідмана, подібно до критерію Вілкоксона, ґрунтується на процедурі ранжування результатів вимірювань. Водночас принцип ранжування має відмінність: якщо у критерії Вілкоксона ранжування здійснюється вертикально (за різницею між умовами), то у випадку критерію Фрідмана ранги присвоюються горизонтально – від одного вимірювання до іншого.

Застосування даного критерію не потребує виконання вимоги нормальності розподілу варіаційних рядів, що робить його зручним для аналізу даних у психологічних дослідженнях із невеликими вибірками та неметричними шкалами вимірювання.

Вимоги до застосування критерію:

1. Вибірка повинна бути зв’язаною, тобто одні й ті самі досліджувані беруть участь у всіх умовах експерименту.
2. Дані повинні бути виміряні не нижче порядкової шкали.
3. Мінімальна кількість досліджуваних двоє ($n \geq 2$), в яких ознаки вимірювалися в щонайменше трьох умовах ($k \geq 3$).

Емпіричне значення критерію χ_r^2-Фрідмана визначається за допомогою формули:

$$\chi_r^2 = \frac{12}{nk(k+1)} \cdot \Sigma R_i^2 - 3n(k+1), \text{ де}$$

n – обсяг вибірки досліджуваних;
k – кількість вимірювань;
R_i – сума рангів для умови i.

У випадках, коли в даних наявні зв'язані (однакові) ранги, загальна формула критерію χ_r^2-Фрідмана може давати дещо неточне значення. Для усунення цього недоліку необхідно здійснити поправку на однакові ранги. У такому разі формула розрахунку критерію набуває наступного вигляду:

$$\chi_r^2 = \frac{12 \cdot \Sigma R_i^2 - 3n^2 k(k+1)^2}{nk(k+1) + \left(nk - \sum_{j=1}^{g} t_j^3 \right) \Big/ (k-1)}, \text{ де}$$

g – число груп зв'язаних рангів у кожному рядку таблиці;
t – розмір j-групи зв'язаних рангів (скільки однакових рангів у неї входить).

Зверніть увагу! Навіть у випадку відсутності зв'язаних рангів кожен окремий (незв'язаний) ранг розглядається як група зв'язаних рангів, що має одне значення ($t = 1$). Це положення є необхідним для коректного застосування поправки на зв'язані ранги під час обчислення критерію.

Для прийняття рішення про рівень статистичної значущості необхідно емпіричне (розрахункове) значення критерію χ_r^2 порівняти з критичним (табличним) значенням.

Якщо $k = 3$, $n > 9$ або $k > 3$, $n > 4$, то користуємося таблицею критичних значень χ^2, $df = k - 1$ (додаток 2).

Якщо $k = 3$, $n < 10$ або $k = 4$, $n < 5$, то користуємося спеціальними таблицями критичних значень χ_r^2-Фрідмана (додатки 10 та 11).

Якщо емпіричне значення дорівнює або більше критичного значення для заданого α-рівня, то H_0 відхиляється і формулюється висновок, що відмінності статистично значущі.

Рекомендованим індексом оцінки розміру ефекту для критерію χ_r^2-Фрідмана є W-Кендалла.

$$W = \frac{\chi_r^2}{n(k-1)}, \text{ де}$$

χ_r^2 – емпіричне значення χ_r^2-Фрідмана;
n – обсяг вибірки досліджуваних;
k – кількість вимірювань.

Цей показник відображає ступінь погодженості між досліджуваними: значення, близьке до 0, свідчить про повну відсутність згоди, тоді як значення, що наближається до 1, вказує на цілковиту узгодженість. Інтерпретація величини ефекту залежить від галузі застосування. У психологічних дослідженнях зазвичай орієнтуються на такі межі:

$0{,}00 \le W < 0{,}10$ – несуттєвий;
$0{,}10 \le W < 0{,}30$ – малий;
$0{,}30 \le W < 0{,}50$– середній;
$0{,}50 \le W < 1{,}00$ – великий.

Приклад 5.5. У рамках дослідження десяти студентам було запропоновано виконати завдання чотирьох різних варіантів самостійної роботи з навчальної дисципліни. Метою є перевірка гіпотези про те, чи

можна вважати розроблені викладачем варіанти завдань еквівалентними за рівнем складності. Результати виконання кожного варіанта були оцінені за стобальною шкалою.

Таблиця 5.6 – Таблиця первинних даних та проміжних розрахунків

№	Варіант № 1		Варіант № 2		Варіант № 3		Варіант № 4	
	Оцінка	Ранг	Оцінка	Ранг	Оцінка	Ранг	Оцінка	Ранг
1	92	3	85	1	96	4	90	2
2	65	1	73	3	77	4	69	2
3	59	2	53	1	61	3	63	4
4	84	2	92	4	86	3	72	1
5	65	2	68	3	62	1	72	4
6	71	4	57	1	70	3	63	2
7	86	3	75	1	84	2	91	4
8	79	2	82	3	72	1	84	4
9	85	2	86	3	80	1	87	4
10	100	4	94	2	92	1	97	3
		$\Sigma = 25$		$\Sigma = 22$		$\Sigma = 23$		$\Sigma = 30$

Оскільки первинні дані представлені у порядковій шкалі, для порівняння показників, виміряних у чотирьох умовах на одній і тій самій вибірці досліджуваних, доцільно застосувати критерій χ_r^2-Фрідмана.

1. Встановлюємо рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: за складністю розроблені викладачем варіанти завдань самостійної роботи не відрізняються;

H_1: за складністю розроблені викладачем варіанти завдань самостійної роботи відрізняються.

2. Рангуємо індивідуальні результати першого досліджуваного, отримані ним у першому, другому, третьому і четвертому варіантах завдань. У такий же спосіб необхідно проранжувати індивідуальні значення інших досліджуваних.

3. Розраховуємо суми рангів для кожного варіанту завдань.

4. У нашому прикладі відсутні зв'язані ранги. Емпіричне значення критерію χ_r^2-Фрідмана будемо визначати за загальною формулою:

$$\chi_r^2 = \frac{12}{n \cdot k \cdot (k+1)} \cdot \sum R_i^2 - 3n \cdot (k+1).$$

$$\chi_r^2 = \frac{12}{10 \cdot 4 \cdot (4+1)} \cdot \left[25^2 + 22^2 + 23^2 + 30^2\right] - 3 \cdot 10 \cdot (4+1) = 2{,}28.$$

5. Так як, в нашому прикладі $k > 3$, $n > 4$, то для визначення рівня статистичної значущості користуємося таблицею критичних значень χ^2

(додаток 2). За даною таблицею для $df = k - 1 = 4 - 1 = 3$ та заданого $\alpha = 0{,}05$ знаходимо $\chi^2_{крит.} = 7{,}815$.

Оскільки $\chi^2_{емп.} > \chi^2_{крит.}$ (2,28 < 7,815), гіпотеза H_0 приймається.

6. Оцінюємо величину розміру стандартизованого ефекту:

$$W = \frac{\chi_r^2}{n(k-1)} = \frac{2,28}{10 \cdot (4-1)} = 0,076.$$

Розмір стандартизованого ефекту дуже низький (несуттєвий).

Висновок. Результати аналізу за критерієм χ_r^2-Фрідмана свідчать про відсутність достатніх підстав для відхилення нульової гіпотези ($\chi_{r\,(3)}^2 = 2{,}28$; $p > 0{,}05$). Це дає змогу зробити висновок, що розроблені викладачем варіанти завдань самостійної роботи не відрізняються за рівнем складності. Розмір стандартизованого ефекту ($W = 0{,}076$) відповідає низькому рівню.

– відео-огляд процедури аналізу даних Прикладу 5.5. у статистичній програмі SPSS Statistics 23.0. та інтерпретація результатів дослідження.

Приклад 5.6. У дев'яти учасників було виміряно рівень конформізму за різних експериментальних умов. Досліджується, чи мають експериментальні умови вплив на рівень конформізму учасників.

Таблиця 5.7 – Таблиця первинних даних та проміжних розрахунків

№	Експеримент № 1		Експеримент № 2		Експеримент № 3	
	Рівень конформізму	Ранг	Рівень конформізму	Ранг	Рівень конформізму	Ранг
1	6	1	12	3	8	2
2	4	1	15	3	9	2
3	6	1,5	11	3	6	1,5
4	6	2	14	3	4	1
5	5	1	15	3	7	2
6	8	2	7	1	11	3
7	10	1,5	11	3	10	1,5
8	7	1	13	3	12	2
9	9	2,5	9	2,5	8	1
	Σ =	13,5		24,5		16

Первинні дані подані в порядковій шкалі. Метою дослідження є порівняння показників, виміряних у трьох експериментальних умовах на

одній і тій самій вибірці досліджуваних. За таких умов доцільним є застосування критерію χ_r^2-Фрідмана.

1. Встановлюємо рівень ймовірності помилки першого роду та формулюємо нульову і альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: умови експерименту не впливають на рівень конформізму досліджуваних;

H_1: умови експерименту впливають на рівень конформізму досліджуваних.

2. Рангуємо індивідуальні результати першого досліджуваного, отримані ним у різних умовах дослідження. У такий же спосіб необхідно проранжувати індивідуальні значення інших досліджуваних.

3. Розраховуємо суми рангів конформізму для кожної експериментальної ситуації.

4. Оскільки в нашому прикладі є зв'язані ранги емпіричне значення критерію χ_r^2-Фрідмана будемо визначати за скоригованою формулою:

$$\chi_r^2 = \frac{12 \cdot \Sigma R_i^2 - 3n^2 k(k+1)^2}{nk(k+1) + \left(nk - \sum_{j=1}^{g} t_j^3 \right) \Big/ (k-1)}.$$

Спочатку розрахуємо значення $\sum_{j=1}^{g} t_j^3$:

$$\sum_{j=1}^{g} t_j^3 = (1^3+1^3+1^3) + (1^3+1^3+1^3) + (1^3+2^3) + (1^3+1^3+1^3) + (1^3+1^3+1^3) +$$
$$+ (1^3+1^3+1^3) + (1^3+2^3) + (1^3+1^3+1^3) + (1^3+2^3) = 45.$$

$$\chi_r^2 = \frac{12 \cdot \left[13{,}5^2 + 24{,}5^2 + 16^2 \right] - 3 \cdot 9^2 \cdot 3 \cdot (3+1)^2}{9 \cdot 3 \cdot (3+1) + (9 \cdot 3 - 45)/(3-1)} = \frac{798}{99} = 8{,}06.$$

5. Так як, в нашому прикладі $k = 3$, $n < 10$, то для визначення рівня статистичної значущості користуємося спеціальною таблицею критичних значень χ_r^2-Фрідмана (додаток 10). За даною таблицею знаходимо критичне значення близьке до заданого $\alpha = 0{,}05$, $\chi^2_{0{,}048.} = 6{,}222$.

Оскільки $\chi^2_{емп.} > \chi^2_{крит.}$ (8,06 > 6,222), гіпотеза H_0 відхиляється.

6. Оцінюємо величину розміру статистичного ефекту:

$$W = \frac{\chi_r^2}{n(k-1)} = \frac{8{,}06}{9 \cdot (3-1)} = 0{,}448.$$

Розмір стандартизованого ефекту – середній.

Висновок. Результати застосування критерію χ_r^2-Фрідмана свідчать про наявність статистично значущих відмінностей між середніми рангами рівня конформізму досліджуваних за різних експериментальних умов (χ_r^2 =

8,06; $p < 0,05$; $n = 9$). Розмір стандартизованого ефекту ($W = 0,448$) відповідає середньому рівню ефекту.

5.3. Критерії порівняння розподілів

Критерії порівняння розподілів для номінальних (категоріальних) даних використовуються для оцінки того, чи однаково розподілені частоти спостережень у різних групах за певною ознакою. Їх суть полягає в наступному: кожна ознака поділяється на дискретні категорії (наприклад, «так», «ні», «не визначився»); підраховуються спостережувані частоти (кількість випадків у кожній категорії); розраховуються очікувані частоти за умови нульової гіпотези (що категорії розподілені однаково або не залежать від групи); порівнюються спостережувані та очікувані значення для оцінки відмінностей або взаємозв'язків. Таким чином, критерії дозволяють оцінити не конкретні числові значення, а структуру розподілу ознаки, чи існують статистично значущі відмінності у частотах категорій між групами або від очікуваного розподілу.

5.3.1. Критерій χ^2-Пірсона

Критерій χ^2-Пірсона (критерій узгодженості, критерій χ^2) є одним із найпоширеніших і найпотужніших непараметричних статистичних методів. Його застосування можливе для аналізу числових, рангових і номінальних даних. До того ж, цей критерій не обмежує кількість порівнюваних розподілів, що робить його універсальним інструментом для перевірки статистичних гіпотез щодо відповідності емпіричних і теоретичних розподілів.

Критерій χ^2 дає змогу з'ясувати, чи спостерігаються статистично значущі відмінності у частотах появи різних значень ознаки у двох або більше розподілах.

Після виявлення загальних відмінностей між розподілами за допомогою критерію χ^2-Пірсона для уточнення, за якими саме категоріями (ознаками) спостерігаються значущі розбіжності, доцільно застосовувати *z*-критерій для порівняння пропорцій або біноміальний критерій *m*. Додаткову цінну інформацію можна отримати шляхом аналізу скоригованих стандартизованих залишків, що дозволяє виявити конкретні осередки значущих відхилень у таблиці крос-табуляції.

Вимоги до застосування критерію:

1. Обсяг вибірки має бути достатньо великим ($n \geq 30$), оскільки за меншої кількості спостережень критерій χ^2 забезпечує лише приблизні результати. Точність оцінювання за допомогою цього критерію зростає зі збільшенням вибірки.

2. Якщо $df = 1$ (або $k = 2$), усі теоретичні (очікувані) частоти мають перевищувати значення 5. Інакше кажучи, якщо кількість категорій

(градацій) задана наперед і не підлягає зміні, критерій χ^2 не може бути застосований без накопичення мінімально необхідної кількості спостережень. Недотримання цієї умови призводить до втрати надійності результатів і спотворення рівня значущості.

3. Якщо $df > 1$ (або $k > 2$), допускається, щоб не більше 20 % теоретичних (очікуваних) частот були меншими за 5. Водночас жодна теоретична частота не повинна бути меншою за 1. У разі невідповідності цим вимогам часто необхідно об'єднувати деякі класові інтервали для забезпечення коректності застосування критерію χ^2.

4. Категорії мають бути взаємовиключними, тобто кожне спостереження повинно належати лише до однієї категорії та не може одночасно належати до будь-якої іншої.

5. У разі порівняння розподілів ознак, які набувають лише двох значень, необхідно застосовувати поправку на неперервність.

6. Критерій χ^2 дозволяє виявляти статистично значущі відмінності між емпіричним розподілом та теоретичним, а також між кількома емпіричними розподілами. Водночас він не дає інформації щодо спрямованості або характеру цих відмінностей.

Критерій χ^2-Пірсона застосовують у двох основних випадках:

1. Для співставлення двох і більше емпіричних розподілів.

2. Для порівняння емпіричного розподілу з теоретичним (наприклад, нормальним, рівномірним тощо).

Ще одним важливим застосуванням критерію χ^2-Пірсона є перевірка гіпотез щодо наявності зв'язку між ознаками (змінними). На основі значення χ^2 сбуло розроблено коефіцієнти взаємної зв'язаності, зокрема коефіцієнти Пірсона *C*, Чупрова *K* та Крамера *V*.

Основна розрахункова формула має такий вигляд:

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(f_{емп.} - f_{теор.}\right)^2}{f_{теор.}}, \text{ де}$$

k – кількість класових інтервалів (градацій) досліджуваної ознаки;

$f_{емп.}$ та $f_{теор.}$ – відповідно емпіричні та теоретичні (очікувані) частоти, що відповідають визначеним градаціям змінної.

У випадку порівняння двох або більше емпіричних розподілів якісних ознак, отриманих на незалежних вибірках, для розрахунку теоретичної (очікуваної) частоти кожної комірки таблиці слід враховувати, що:

$$f_{ij} = \frac{f_i \cdot f_j}{n}, \text{ де}$$

f_i – сума частот у всіх комірках i-рядка;

f_j – сума частот у всіх комірках j- стовпчика;

n – сума частот у всій таблиці зв'язаності ознак.

Для таблиці крос-табуляції ознак розміром 2×2 застосовується критерій χ^2-Пірсона з поправкою на неперервність.

$$\chi^2 = \frac{n \cdot \left(|ad - bc| - \frac{n}{2} \right)^2}{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)}$$, де

a, *b*, *c*, *d* – дані, що визначаються згідно з таблицею зв'язаності ознак розміром 2×2;

n – обсяг вибірки досліджуваних.

Обчислене емпіричне значення критерію порівнюють із критичним для заданого α-рівня, що визначається за таблицею критичних значень критерію χ^2 (додаток 2), для кількості ступенів свободи:

$df = (k - 1) \cdot (m - 1)$, де

k – кількість розрядів (класових інтервалів або градацій ознаки);

m – кількість розподілів, що порівнюються.

Якщо $\chi^2_{емп.}$ менше $\chi^2_{крит.}$, то формулюється висновок про відсутні статистично значущі відмінності між розподілами (приймаємо гіпотезу H_0).

Якщо $\chi^2_{емп.}$ більше або дорівнює $\chi^2_{крит.}$, то нульова гіпотеза відхиляється та приймається альтернативна.

Для оцінки розміру стандартизованого ефекту для критерію χ^2-Пірсона в залежності від умов застосовуються різні індекси:

1) φ – таблиця крос-табуляції ознак має розмір 2×2:

$$\varphi = \sqrt{\frac{\chi^2}{n}}$$, де

χ^2 – емпіричне значення критерію χ^2-Пірсона;

n – обсяг вибірки досліджуваних.

2) *V*-Крамера в ситуації, коли таблиця крос-табуляції ознак має відмінний від 2×2 розмір:

$$V = \sqrt{\frac{\chi^2}{n \cdot (c-1)}}$$, де

χ^2 – емпіричне значення критерію $\chi 2$-Пірсона;

n – обсяг вибірки досліджуваних;

c – найменша кількість градацій із двох змінних.

3) *w*-Коена використовується при співставленні емпіричного розподілу з теоретичним:

$$w = \sqrt{\sum_{i=1}^{k} \frac{(p_{емп.} - p_{теор.})^2}{p_{теор.}}}$$, де

k – кількість класових інтервалів (градацій) досліджуваної ознаки;

$p_{емп.}$ та $p_{теор.}$ – відповідно емпіричні та теоретичні (очікувані) відносні частоти.

При інтерпретації величини розміру ефекту для φ та *w*-Коена психологи використовують наступні рекомендації:

$0{,}00 \le \varphi, w < 0{,}10$ – несуттєвий;

$0{,}10 \le \varphi, w < 0{,}30$ – малий;

$0{,}30 \le \varphi, w < 0{,}50$ – середній;

$0{,}50 \le \varphi, w < 1{,}00$ – великий.

Що стосується інтерпретації коефіцієнта *V*-Крамера, існують різні емпіричні правила, одне з яких дозволяє інтерпретувати його значення залежно від числа ступенів свободи (*df*):

Таблиця 5.8 – Інтерпретація розміру ефекту індексу *V*-Крамера

df	Несуттєвий	Малий	Середній	Великий
1	0,00 – 0,09	0,10 – 0,29	0,30 – 0,49	> 0,50
2	0,00 – 0,06	0,07 – 0,20	0,21 – 0,34	> 0,35
3	0,00 – 0,05	0,06 – 0,16	0,17 – 0,28	> 0,29
4	0,00 – 0,04	0,05 – 0,14	0,15 – 0,24	> 0,25
5	0,00 – 0,04	0,05 – 0,12	0,13 – 0,21	> 0,22

Приклад 5.7. Чи існують статистично значущі відмінності в професійних здібностях між юнаками та дівчатами?

Таблиця 5.9 – Таблиця крос-табуляції ознак

Тип професійних здібностей	Юнаки ($f_{емп.}$)	Дівчата ($f_{емп.}$)	Усього
Людина-природа	27	41	68
Людина-техніка	29	13	42
Людина - людина	32	47	79
Людина - художній образ	18	30	48
Людина - знак	25	24	49
Усього	131	155	286

Первинні дані представлені в номінальній шкалі. Таблиця крос-табуляції ознак має розмір відмінний від 2×2, а саме 5×2. У такому випадку застосовується класичний варіант критерію χ^2-Пірсона.

1. Встановлюємо рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: між розподілами типів професійних здібностей у юнаків та дівчат відмінностей немає;

H_1: існують відмінності між розподілами типів професійних здібностей у юнаків та дівчат.

2. Для розрахунку емпіричного значення χ^2-Пірсона необхідно спочатку визначити теоретичні (очікувані) частоти для кожної комірки таблиці:

$$f_{ij} = \frac{f_i \cdot f_j}{n}, \text{ де}$$

f_i – сума частот у всіх комірках *i*-рядка;

f_j – сума частот у всіх комірках *j*-стовпчика;

n – сума частот у всій таблиці зв'язаності ознак.

Таблиця 5.10 – Значення очікуваних частот

Тип професійних здібностей	Юнаки ($f_{теор.}$)	Дівчата ($f_{теор.}$)	Усього
Людина - природа	$f_{11} = \frac{68 \cdot 131}{286} = 31,1$	$f_{12} = \frac{68 \cdot 155}{286} = 36,9$	68
Людина - техніка	$f_{21} = \frac{42 \cdot 131}{286} = 19,2$	$f_{22} = \frac{42 \cdot 155}{286} = 22,8$	42
Людина - людина	$f_{31} = \frac{79 \cdot 131}{286} = 36,2$	$f_{32} = \frac{79 \cdot 155}{286} = 42,8$	79
Людина - художній образ	$f_{41} = \frac{48 \cdot 131}{286} = 22,0$	$f_{42} = \frac{48 \cdot 155}{286} = 26,0$	48
Людина - знак	$f_{51} = \frac{49 \cdot 131}{286} = 22,4$	$f_{52} = \frac{49 \cdot 155}{286} = 26,6$	49
Усього	131	155	286

3. Підставимо значення емпіричних і очікуваних частот у формулу для розрахунку χ^2:

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(f_{емп.} - f_{теор.} \right)^2}{f_{теор.}} = \frac{(27-31,1)^2}{31,1} + \frac{(41-36,9)^2}{36,9} + \frac{(29-19,2)^2}{19,2} +$$

$$+ \frac{(13-22,8)^2}{22,8} + \frac{(32-36,2)^2}{36,2} + \frac{(47-42,8)^2}{42,8} + \frac{(18-22,0)^2}{22,0} + \frac{(30-26,0)^2}{26,0} +$$

$$+ \frac{(25-22,4)^2}{22,4} + \frac{(24-26,6)^2}{26,6} = 12,92.$$

4. За таблицею критичних значень критерію χ^2-Пірсона (додаток 2) для заданого $\alpha = 0,05$ та $df = (k-1)\cdot(m-1) = (5-1)\cdot(2-1) = 4$ знаходимо $\chi^2_{крит.} = 9,488$.

Оскільки $\chi^2_{емп.} > \chi^2_{крит.}$ (12,92 > 9,488), гіпотеза H_0 відхиляється.

1. Оцінка розміру ефекту здійснюється за індексом V-Крамера, тому що таблиця зв'язаності ознак має відмінний від 2×2 розмір:

$$V = \sqrt{\frac{\chi^2}{n \cdot (c-1)}} = \sqrt{\frac{12,92}{286 \cdot (2-1)}} = 0,21.$$

Згідно інтерпретації Дж. Коена розмір ефекту відповідає середньому рівню ($df = 4$).

Висновок. Критерій χ^2-Пірсона дозволив відхилити нульову гіпотезу про відсутність зв'язку між статтю та типами професійних здібностей ($\chi^2_{(4)} = 12,92$; $p < 0,05$). Це свідчить про наявність статистично значущих відмінностей у типах професійних здібностей між юнаками та дівчатами.

Розмір ефекту за коефіцієнтом Крамера *V* становить 0,21, що відповідає середньому рівню.

– відео-огляд процедури аналізу даних Прикладу 5.7. у статистичній програмі SPSS Statistics 23.0. та інтерпретація результатів дослідження.

Приклад 5.8. За результатами дослідження встановлено, що 97 студентів із 925 з екстраверсією після навчання у закладі вищої освіти отримали диплом з відзнакою, а із інтроверсією – 44 студенти із 661. Чи можна стверджувати, що отримання диплому з відзнакою залежить від характеристики темпераменту.

Таблиця 5.11 – Таблиця крос-табуляції ознак

Параметри	Тип диплому після навчання у ЗВО		Усього
	з відзнакою	без відзнаки	
Екстраверсія	97	828	925
Інтроверсія	44	617	661
Усього	141	1445	1586

Суть завдання полягає в порівнянні двох емпіричних розподілів номінальних ознак, отриманих на незалежних вибірках. У разі, коли таблиця крос-табуляції має розмір 2×2, застосовується критерій χ^2-Пірсона з поправкою на неперервність (корекцією Єйтса).

$$\chi^2 = \frac{n \cdot \left(|ad - bc| - \frac{n}{2} \right)^2}{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)}$$

У нашому прикладі: $a = 97$, $b = 828$, $c = 44$, $d = 617$, $n = 1586$.

1. Встановлюємо рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: отримання диплому з відзнакою не залежить від екстраверсії/інтроверсії;

H_1: отримання диплому з відзнакою залежить від екстраверсії/інтроверсії.

2. Обчислимо емпіричне значення χ^2-Пірсона з поправкою на неперервність.

$$\chi^2 = \frac{1586 \cdot \left(\left| 97 \cdot 617 - 828 \cdot 44 \right| - \frac{1586}{2} \right)^2}{(97+828) \cdot (44+617) \cdot (97+44) \cdot (828+617)} = 6{,}516.$$

3. За таблицею критичних значень критерію χ^2 (додаток 2) для заданого $\alpha = 0{,}05$ та $df = (k-1) \cdot (m-1) = (2-1) \cdot (2-1) = 1$ знаходимо $\chi^2_{крит.} = 3{,}841$.

Оскільки $\chi^2_{емп.} > \chi^2_{крит.}$ (6,516 > 3,841), гіпотеза H_0 відхиляється, приймається альтернативна.

5. Оцінка розміру ефекту здійснюється за індексом φ, тому що таблиця зв'язаності ознак має розмір 2×2:

$$\varphi = \sqrt{\frac{\chi^2}{n}} = \sqrt{\frac{6{,}516}{1586}} = 0{,}06.$$

Розмір стандартизованого ефекту дуже низький (несуттєвий).

Висновок. Критерій χ^2-Пірсона дає підстави відхилити нульову гіпотезу ($\chi^2_{(1)} = 6{,}516$; $p < 0{,}05$). Тобто, отримання диплому з відзнакою залежить від екстраверсії/інтроверсії. Водночас дуже низький розмір стандартизованого ефекту ($\varphi = 0{,}06$) свідчить про низьку практичну цінність отриманих результатів.

Приклад 5.9. Чи свідчать отримані емпіричні дані про рівномірний розподіл типів темпераменту серед студентів-психологів?

Таблиця 5.12 – Розподіл типів темпераменту в студентів-психологів

Тип темпераменту	Холерик	Флегматик	Сангвінік	Меланхолік	Усього
Кількість	17	23	32	8	80

Первинні дані подані в номінальній шкалі. Необхідно перевірити, чи відповідає емпіричний розподіл результатів у вибірці теоретичному розподілу в генеральній сукупності. Для розв'язання подібних завдань застосовується критерій χ^2-Пірсона для однієї вибірки (одновибірковий критерій χ^2-Пірсона).

1. Що стосується виду теоретичного розподілу, то в даному випадку використовується рівномірний розподіл. Його сутність полягає в тому, що всі результати вважаються рівноймовірними. За наявності чотирьох типів темпераменту, ймовірність діагностування будь-якого з них у студентів-психологів становить 1/4 = 0,25. Інакше кажучи, якби емпіричний розподіл повністю збігався з теоретичним, то в кожну комірку таблиці мало б потрапити однакове число спостережень, тобто 80 / 4 = 20 осіб.

З урахуванням зазначених обставин остаточна розрахункова таблиця даних для нашого прикладу має такий вигляд.

Таблиця 5.13 – Теоретичний та емпіричний розподіл типів темпераменту серед студентів-психологів

Тип темпераменту	Холерик	Флегматик	Сангвінік	Меланхолік	Усього
Емпіричні частоти	17	23	32	8	80
Теоретичні частоти	20	20	20	20	80

2. Встановлюємо рівень ймовірності помилки першого роду і формулюємо нульову та альтернативну гіпотези:

$\alpha = 0{,}05$;

H_0: типи темпераменту у студентів-психологів розподілені з однаковою частотою;

H_1: типи темпераменту у студентів-психологів розподілені з різною частотою.

3. Обчислюється емпіричне значення критерію :

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(f_{емп.} - f_{теор.}\right)^2}{f_{теор.}} = \frac{(17-20)^2}{20} + \frac{(23-20)^2}{20} + \frac{(32-20)^2}{20} + \frac{(8-20)^2}{20} =$$

$$= 0{,}45 + 0{,}45 + 7{,}2 + 7{,}2 = 15{,}3.$$

4. За таблицею критичних значень критерію χ^2 (додаток 2) для заданого $\alpha = 0{,}05$ та $df = (k-1)\cdot(m-1) = (4-1)\cdot(2-1) = 3$ знаходимо $\chi^2_{крит.} = 7{,}815$.

Оскільки $\chi^2_{емп.} > \chi^2_{крит.}$ (15,3 > 7,815), гіпотеза H_0 відхиляється.

5. Оцінка розміру ефекту здійснюється за індексом *w*-Коена.

$$w = \sqrt{\sum_{i=1}^{k} \frac{(p_{емп.} - p_{теор.})^2}{p_{теор.}}}$$, де

k – кількість класових інтервалів (градацій) досліджуваної ознаки;

$p_{емп.}$ та $p_{теор.}$ – емпіричні та теоретичні (очікувані) відносні частоти.

Підраховуємо відносні частоти для всіх категорій ознаки за формулою:

$$p_i = \frac{f_i}{n_i}$$, де

f_i – частота для категорії ознаки;

n_i – кількість спостережень у вибірці.

Таблиця 5.14 – Теоретичний і емпіричний розподіл відносних частот типів темпераменту серед студентів-психологів

Тип темпераменту	Холерик	Флегматик	Сангвінік	Меланхолік	Усього
Емпіричні відносні частоти ($p_{емп.}$)	17/80 = 0,2125	23/80 = 0,2875	32/80 = 0,40	8/80 = 0,10	Σ = 1
Теоретичні відносні частоти ($p_{теор.}$)	20/80 = 0,25	20/80 = 0,25	20/80 = 0,25	20/80 = 0,25	Σ = 1

$$w = \sqrt{\frac{(0,2125-0,25)^2}{0,25} + \frac{(0,2875-0,25)^2}{0,25} + \frac{(0,40-0,25)^2}{0,25} + \frac{(0,10-0,25)^2}{0,25}} = 0,44.$$

Розмір стандартизованого ефекту відповідає середньому рівню.

Висновок. Застосування критерію χ^2-Пірсона дає підстави відхилити нульову гіпотезу та прийняти альтернативну ($\chi^2_{(3)}$= 15,3; p < 0,05). Типи темпераменту розподілені не рівномірно у студентів-психологів. Розмір ефекту відповідає середньому рівню (w = 0,44).

– відео-огляд процедури аналізу даних Прикладу 5.9. у статистичній програмі SPSS Statistics 23.0. та інтерпретація результатів дослідження.

Завдання на самостійну підготовку

1. Вимоги до даних при застосуванні непараметричних критеріїв статистичного порівняння вибірок досліджуваних.
2. Поясніть сутність критерію *U*-Манна-Уітні та його використання для незалежних вибірок.
3. Поясніть призначення критерію *H*-Краскела-Уолліса.
4. Визначте, за яких умов застосовують критерій *T*-Вілкоксона у психології.
5. Поясніть призначення критерію χ_r^2 -Фрідмана.
6. Розкрийте умови використання критерію χ^2-Пірсона та інтерпретацію його результатів у психологічних дослідженнях.
7. Особливості використання та алгоритм розрахунку *G*-критерію знаків.
8. Особливості використання та алгоритм розрахунку Q-критерію Розенбаума.
9. Особливості використання та алгоритм розрахунку критерію тенденцій *L*-Пейджа
10. Особливості використання та алгоритм розрахунку критерію тенденцій *S*-Джонкхієра-Терпстри.
11. Особливості використання та алгоритм розрахунку критерію φ – кутове перетворення Фішера.
12. Складіть таблицю умов застосування статистичних критеріїв статистичного порівняння вибірок досліджуваних.

СПИСОК ВИКОРИСТАНОЇ ТА РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Боснюк В.Ф. Математичні методи в психології: курс лекцій. Мультимедійне навчальне видання. Х.: НУЦЗУ, 2020. 141 с.
2. Вдовенко В.В. Математичні методи в психології: Навчально-методичний посібник. Кіровоград: ПП «Центр оперативної поліграфії» Авангард», 2017. 112 с.
3. Гаркуша С.В. Методи математичної статистики в педагогічних дослідженнях. Навчально-методичний посібник для аспірантів. Чернігів, 2019. 72 с.
4. Герич М.С., Синявська О.О. Математична статистика: навч. посібник. Ужгород: ДВНЗ «УжНУ», 2021. 146 с.
5. Гульбс О., Кравченко О., Лантух Л., Махомета Т., Поліщук Т. Статистичний аналіз педагогічних та психологічних досліджень. Навчальний посібник. Методичні рекомендації, тести, індивідуальні завдання, задачі для самостійного розв'язування. Умань-Київ: ЦТЗІ. «Друкарський дім», 2023. 412 с.
6. Жерновникова О.А., Золотухіна С.Т. Статистичні методи в педагогічних дослідженнях у схемах і таблицях: навчальний посібник / за ред. д. пед. наук, чл.-кор. НАПН України В. І. Лозової. Харків, 2018. 108 с.
7. Кашуба М.О., Корда М.М. Основи медичної статистики та проведення комп'ютерного статистичного аналізу даних статистичними програмами. навч.-метод. посіб.: у 4 ч. Ч 1. Порівняння середніх дисперсій. Тернопіль: ТНМУ. Укрмедкнига, 2021. 120 с.
8. Кашуба М.О., Корда М.М. Основи медичної статистики та проведення комп'ютерного статистичного аналізу даних статистичними програмами. навч.-метод. посіб.: у 4 ч. Ч 2. Кореляція та регресія. Тернопіль: ТНМУ. Укрмедкнига, 2022. 212 с.
9. Кашуба М.О., Корда М.М. Біостатистика: підручник. Тернопіль: ТНМУ: Укрмедкнига, 2024. 707 с.
10. Лупан І.В., Авраменко О.В., Акбаш К.С. Комп'ютерні статистичні пакети: навчально-методичний посібник. 2-ге вид. Кіровоград: «КОД» 2015. 236 с.
11. Москальов І.О., Лисенко Д.П. Застосування методів математичної статистики у психолого-педагогічних дослідженнях: навч. посіб. Київ: НУОУ, 2023. 187 с.
12. Негрей М., Гнот Т. Аналітика з R: навчальний посібник. Київ: ФОП Ямчинський О. В., 2020. 236 с.
13. Паянок Т.М., Задорожня Т.М. Статистичний аналіз даних: навчальний посібник. Ірпінь: Університет державної фіскальної служби України, 2020. 312 с.
14. Петровська І.Р., Островська К.О. Математично-статистичні методи обробки емпіричних даних психолого-педагогічних досліджень: навч. посіб. Львів: Друкарня «Справи Кольпінга в Україні», 2021. 140 с.
15. Поліщук Т.В. Математичний апарат педагогічної науки: навчальний посібник. МОН України, Уманський держ. пед.. ун-т імені Павла Тичини.

Умань: Візаві, 2019. 109 с.

16. Руська Р.В. Теорія імовірності та математична статистика в психології: навч. посіб. Тернопіль, 2020. 112 с.

17. Шинкарук О.М., Боровик Л.В., Боровик О.В. Імовірнісно-статистичні методи в психолого-педагогічних дослідженнях: навчальний посібник. Хмельницький : Видавництво НАДПСУ, 2019. 272 с.

18. Яременко Л.І., Лупан І.В. Кількісні методи у поведінкових науках: навчальний посібник. Кропивницький: Видавець Лисенко В.Ф., 2019. 224 с.

19. American Psychological Association. Publication manual of the American Psychological Association: 7-th ed., Waschinton, DC: Author, 2020. 428 p.

20. Foster, G., Lane D. Scott D., Hebl M. and other. An Introduction to Psychological Statistics. University of Missouri, St. Louis, 2018. 271 p.

СЛОВНИК ТЕРМІНІВ

Аномалія (аутлаєр) – спостережуване значення, яке суттєво відхиляється від інших у вибірці або розподілі. Аномалії можуть бути наслідком помилок вимірювання, рідкісних подій або справжньої індивідуальної особливості, і їхня ідентифікація важлива для коректного аналізу даних.

Апостеріорна ймовірність – ймовірність настання події з урахуванням отриманих емпіричних даних, що обчислюється після проведення дослідження. В психології її застосовують у баєсових моделях для оновлення оцінок ймовірності певних когнітивних чи поведінкових явищ.

Апробація гіпотези – процес перевірки статистичної гіпотези на основі вибіркових даних. Включає обчислення статистики тесту, порівняння її з критичним значенням і формулювання висновку про достовірність гіпотези.

Асиметрія – статистичний показник, що характеризує ступінь відхилення розподілу випадкової змінної від симетричного (нормального) розподілу**.**

Вибірка – підмножина елементів генеральної сукупності, що обирається для дослідження, щоб отримати інформацію про сукупність у цілому.

Вибіркова дисперсія – числова характеристика, що відображає ступінь розсіяння значень змінної в вибірці навколо середнього.

Вибіркове середнє – арифметичне середнє значень у вибірці, що служить оцінкою середнього генеральної сукупності.

Відносна частота – частка випадків певної події від загальної кількості спостережень. Дозволяє порівнювати поширеність різних проявів поведінки або відповідей у групі.

Вірогідність (ймовірність) – числова оцінка можливості настання певної події, що варіює від 0 (неможливо) до 1 (обов’язково).

Внутрішньогрупова дисперсія – міра розсіяння даних усередині однієї групи. Вона показує, наскільки однорідні учасники щодо певної психологічної характеристики.

Гауссова крива – графічне відображення нормального розподілу ймовірностей, що характеризується симетрією відносно середнього.

Генеральна сукупність – це повний набір всіх елементів, об’єктів або індивідів, що володіють певними характеристиками і про які дослідник хоче зробити висновки.

Гіпотеза альтернативна (H_1) – припущення, що протилежне нульовій гіпотезі, тобто ефект, різниця або взаємозв’язок між змінними існує.

Гіпотеза нульова (H_0) – припущення про відсутність ефекту, різниці або взаємозв’язку між змінними.

Гістограма – графік у вигляді стовпчиків, що показує частоти значень змінної. Використовується для візуалізації розподілу.

Двофакторний дисперсійний аналіз – аналіз впливу двох факторів на залежну змінну одночасно. Дозволяє оцінити як окремий, так і взаємний ефект факторів.

Дисперсійний аналіз (ANOVA) – статистичний метод, який використовується для порівняння середніх значень трьох і більше груп. Він дозволяє визначити, чи є різниця між групами статистично значущою, оцінюючи співвідношення між варіацією всередині груп і варіацією між групами.

Дисперсія – середнє квадратів відхилень значень змінної від середнього. Використовується для кількісної оцінки варіативності психологічних показників.

Додаткова змінна (коваріата) – змінна, яку контролюють для зменшення впливу на залежну змінну, забезпечуючи точнішу оцінку основних ефектів.

Достовірність результатів – характеристика, що відображає ймовірність того, що результати дослідження не випадкові і можуть бути відтворені при повторних дослідженнях.

Ексцес – це показник, що відображає ступінь «гостровершинності» або «плосковершинності» розподілу порівняно з нормальним**.** Він характеризує концентрацію значень навколо середнього.

Емпіричний розподіл – розподіл значень змінної, отриманий безпосередньо з даних спостережень або вимірювань. Використовується для аналізу реальних психологічних показників.

Залежна змінна – змінна, значення якої вимірюють у дослідженні, щоб оцінити вплив незалежних змінних.

Зсув (біас) – систематична помилка, яка спотворює оцінку параметра і може виникати через метод збору даних або особливості вибірки.

Інтервальна шкала – шкала вимірювання з рівними інтервалами між значеннями, але без абсолютного нуля. Приклади: IQ-тест, температура у градусах Цельсія.

Інтерквартильний розмах – різниця між 75-м і 25-м процентилем розподілу, що показує розсіяння середньої половини значень і стійкий до аномалій.

Кластерний аналіз – статистичний метод групування об’єктів у кластери (групи) на основі подібності їхніх характеристик. У психології застосовується для виділення типів особистості або моделей поведінки серед учасників дослідження.

Коефіцієнт варіації – відношення стандартного відхилення до середнього значення вибірки, виражене у відсотках. Дозволяє порівнювати розсіяння різних змінних у відносних величинах.

Коефіцієнт детермінації (R^2) – частка варіації залежної змінної, пояснена незалежними змінними у регресійній моделі. Дозволяє оцінити, наскільки добре модель пояснює досліджуване явище.

Коефіцієнт кореляції – числова характеристика напрямку та сили взаємозв'язку між двома змінними. Значення варіює від −1 (сильний негативний зв'язок) до +1 (сильний позитивний зв'язок).

Коефіцієнт кореляції Пірсона – показник лінійного взаємозв'язку між двома кількісними змінними.

Коефіцієнт кореляції Спірмена – ранговий показник взаємозв'язку між двома змінними.

Критерій χ^2-Пірсона – статичстичний критерій, що використовується для оцінки залежності між номінальними змінними або для перевірки відповідності емпіричних даних теоретичному розподілу.

Критерій Колмогорова-Смірнова – тест, що дозволяє перевірити, наскільки емпіричний розподіл даних відповідає теоретичному (наприклад, нормальному).

Критерій Крускала-Уолліса – непараметричний критерію, який дозволяє порівнювати дані у трьох і більше вибірках досліджуваних.

Критерій Манна-Уітні (U-критерій) – непараметричний тест для порівняння двох незалежних вибірок на відмінності у розподілі або медіанах.

Кумулятивна частота – сума частот значень змінної до даного порогу. Використовується для оцінки розподілу та визначення процентилів у психологічних вимірюваннях.

Логістична регресія – модель, що описує ймовірність настання бінарної події залежно від однієї або кількох незалежних змінних.

Медіана – значення, яке ділить впорядковану вибірку на дві рівні частини. Використовується як стійкий показник центральної тенденції, особливо при наявності аномалій.

Межі довірчого інтервалу – верхнє та нижнє значення, у межах яких із заданою ймовірністю знаходиться параметр генеральної сукупності.

Множинна регресія – модель залежності однієї залежної змінної від кількох незалежних змінних. Використовується для комплексного прогнозування психологічних показників з урахуванням різних факторів.

Модальне значення (мода) – найпоширеніше значення змінної у вибірці. У психології використовується для опису найтиповіших відповідей учасників.

Нормалізація даних – процес перетворення даних у стандартний масштаб для порівняння або подальшого аналізу.

Нормальний розподіл – симетричний розподіл ймовірностей із середнім, медіаною та модою, що збігаються. Використовується як базова модель для багатьох психологічних характеристик.

Одновибірковий t-тест – тест для перевірки різниці середнього значення вибірки з певним теоретичним значенням.

Оцінка параметра – числова характеристика генеральної сукупності, розрахована на основі вибірки (наприклад, середнє або дисперсія).

Параметричні методи – методи статистики, що базуються на припущеннях про форму розподілу даних (зазвичай нормальний розподіл).

Перетворення даних – математичні операції над даними (логарифмування, стандартизація) для покращення їхнього розподілу або масштабування.

Порівняння парних вибірок – аналіз різниці між двома пов'язаними групами або спостереженнями (наприклад, до і після експерименту).

Порядкова шкала – шкала, що відображає порядок значень, без гарантії рівності інтервалів між ними (наприклад, рівень тривожності: низький, середній, високий).

Похибка вимірювання – різниця між фактичним і спостережуваним значенням. В психології важлива для оцінки точності тестів та методик.

Прийняття нульової гіпотези – висновок про те, що немає статистично значущих доказів для відхилення H_0.

Пропорційна частота – частка випадків певної події у загальній чисельності спостережень.

Прямий ефект – безпосередній вплив однієї змінної на іншу, не опосередкований іншими факторами.

Рангова кореляція – оцінка взаємозв'язку змінних на основі їх порядкових рангів, застосовується для непараметричних даних.

Регресійний аналіз – статистичний метод, що моделює залежність однієї змінної від однієї або кількох незалежних змінних для прогнозування результатів та оцінки впливу факторів.

Розмах (range) – різниця між максимальним і мінімальним значенням змінної; використовується для оцінки загального розсіяння даних.

Розподіл ймовірностей – математичне відображення того, як ймовірності різних значень змінної розподіляються; включає дискретні та неперервні розподіли.

Середнє арифметичне – сума всіх значень змінної, поділена на їхню кількість; найбільш поширена міра центральної тенденції.

Середнє квадратичне відхилення (s) – квадратний корінь із дисперсії; показує середнє відхилення значень від середнього, використовується для оцінки варіативності психологічних показників.

Спостережувана змінна – змінна, яка безпосередньо вимірюється в дослідженні (observable variable), наприклад, час реакції або результат тесту.

Стандартизація даних (z-перетворення) – перетворення значень змінної у стандартні бали для порівняння даних з різних шкал; середнє стає 0, стандартне відхилення – 1.

Стандартне відхилення – показник варіативності, що відображає середнє відхилення значень від середнього арифметичного; ключова характеристика розсіяння даних.

Статистична значущість (p-значення) – ймовірність отримання результату, більш крайнього, ніж спостережуваний, при правильності нульової гіпотези.

Т-критерій (тест Стьюдента) – параметричний тест для перевірки різниці середніх значень двох груп; застосовується при нормальному розподілі даних і приблизно однаковій дисперсії.

Фактор – змінна, вплив якої досліджується у експерименті; може бути незалежною або контрольованою.

Факторний аналіз – метод, що дозволяє виявляти приховані змінні (фактори), які пояснюють кореляції між спостережуваними змінними.

Частота – кількість випадків появи конкретного значення або події у вибірці.

Частотний розподіл – таблиця або графік, що відображає, як часто зустрічаються значення змінної; використовується для аналізу закономірностей у вибірці.

Числова характеристика – показник, що узагальнює властивості вибірки чи сукупності, наприклад, середнє, медіана, мода або дисперсія.

Шкала вимірювання – система, що визначає, як кількісно або якісно оцінюються ознаки: номінальна, порядкова, інтервальна або шкала відношення.

Шкала відношення – інтервальна шкала з абсолютним нулем, що дозволяє виконувати порівняння та відношення величин (наприклад, вага або час реакції).

Явна (маніфестна) змінна – змінна, яку можна безпосередньо спостерігати та вимірювати; наприклад, кількість правильних відповідей у тесті або рівень тривожності за опитувальником.

ДОДАТКИ

Додаток 1

Критичні значення коефіцієнта кореляції Пірсона r_{xy} (Спірмена r_s)

df	Рівень значущості			
	0,10	**0,05**	**0,01**	**0,001**
5	0,805	0,878	0,959	0,991
6	0,729	0,811	0,917	0,974
7	0,669	0,754	0,875	0,951
8	0,621	0,707	0,834	0,925
9	0,582	0,666	0,798	0,898
10	0,549	0,632	0,765	0,872
11	0,521	0,602	0,735	0,847
12	0,497	0,576	0,708	0,823
13	0,476	0,553	0,684	0,801
14	0,458	0,532	0,661	0,780
15	0,441	0,514	0,641	0,760
16	0,426	0,497	0,623	0,742
17	0,412	0,482	0,606	0,725
18	0,400	0,468	0,590	0,708
19	0,389	0,456	0,575	0,693
20	0,378	0,444	0,561	0,679
21	0,369	0,433	0,549	0,665
22	0,360	0,423	0,537	0,652
23	0,352	0,413	0,526	0,640
24	0,344	0,404	0,515	0,629
25	0,337	0,396	0,505	0,618
26	0,330	0,388	0,496	0,607
27	0,323	0,381	0,487	0,597
28	0,317	0,374	0,479	0,588
29	0,311	0,367	0,471	0,579
30	0,306	0,361	0,463	0,570
31	0,301	0,355	0,456	0,562
32	0,296	0,349	0,449	0,554
33	0,291	0,344	0,442	0,547
34	0,287	0,339	0,436	0,539
35	0,283	0,334	0,430	0,532
36	0,279	0,329	0,424	0,525
37	0,275	0,325	0,418	0,519
38	0,271	0,320	0,413	0,513
39	0,267	0,316	0,408	0,507
40	0,264	0,312	0,403	0,501

41	0,260	0,308	0,398	0,495
42	0,257	0,304	0,393	0,490
43	0,254	0,301	0,389	0,484
44	0,251	0,297	0,384	0,479
45	0,248	0,294	0,380	0474
46	0,246	0,291	0,376	0,469
47	0,243	0,288	0,372	0,465
48	0,240	0,285	0,368	0,460
49	0,238	0,282	0,365	0,456
50	0,235	0,279	0,361	0,451
51	0,233	0,276	0,358	0,447
52	0,231	0,273	0,354	0,443
53	0,228	0,271	0,351	0,439
54	0,226	0,268	0,348	0,435
55	0,224	0,266	0,345	0,432
56	0,222	0,263	0,341	0,428
57	0,220	0,261	0,339	0,424
58	0,218	0,259	0,336	0,421
59	0,216	0,256	0,333	0,418
60	0,214	0,254	0,330	0,414
61	0,213	0,252	0,327	0,411
62	0,211	0,250	0,325	0,408
63	0,209	0,248	0,322	0,405
64	0,207	0,246	0,320	0,402
65	0,206	0,244	0,317	0,399
66	0,204	0,242	0,315	0,396
67	0,203	0,240	0,313	0,393
68	0,201	0,239	0,310	0,390
69	0,200	0,237	0,308	0,388
70	0,198	0,235	0,306	0,385
80	0,185	0,220	0,286	0,361
90	0,174	0,207	0,270	0,341
100	0,165	0,197	0,256	0,324
110	0,158	0,187	0,245	0,310
120	0,151	0,179	0,234	0,297
130	0,145	0,172	0,225	0,285
140	0,140	0,166	0,217	0,275
150	0,135	0,160	0,210	0,266
200	0,117	0,139	0,182	0,231
250	0,104	0,124	0,163	0,207
300	0,095	0,113	0,149	0,189

Додаток 2

Критичні значення критерію χ^2

df	Рівень значущості			*df*	Рівень значущості		
	0,05	**0,01**	**0,001**		**0,05**	**0,01**	**0,001**
1	3,841	6,635	10,829	**39**	54,572	62,428	72,080
2	5,991	9,210	13,817	**40**	55,758	63,691	73,428
3	7,815	11,345	16,269	**41**	56,942	64,950	74,772
4	9,488	13,277	18,470	**42**	58,124	66,206	76,111
5	11,070	15,086	20,519	**43**	59,304	67,459	77,447
6	12,592	16,812	22,462	**44**	60,481	68,709	78,779
7	14,067	18,475	24,327	**45**	61,656	69,957	80,107
8	15,507	20,090	26,130	**46**	62,830	71,201	81,431
9	16,919	21,666	27,883	**47**	64,001	72,443	82,752
10	18,307	23,209	29,594	**48**	65,171	73,683	84,069
11	19,675	24,725	31,271	**49**	66,339	74,919	85,384
12	21,026	26,217	32,917	**50**	67,505	76,154	86,694
13	22,362	27,688	34,536	**51**	68,669	77,386	88,003
14	23,685	29,141	36,132	**52**	69,832	78,616	89,308
15	24,996	30,578	37,706	**53**	70,993	79,843	90,609
16	26,296	32,000	39,262	**54**	72,153	81,069	91,909
17	27,587	33,409	40,801	**55**	73,311	82,292	93,205
18	28,869	34,805	42,323	**56**	74,468	83,513	94,499
19	30,144	36,191	43,832	**57**	75,624	84,733	95,790
20	31,410	37,566	45,327	**58**	76,778	85,950	97,078
21	32,671	38,932	46,810	**59**	77,931	87,166	98,365
22	33,924	40,289	48,281	**60**	79,082	88,379	99,649
23	35,172	41,638	49,742	**61**	80,232	89,591	100,887
24	36,415	42,980	51,194	**62**	81,381	90,802	102,165
25	37,652	44,314	52,635	**63**	82,529	92,010	103,442
26	38,885	45,642	54,068	**64**	83,675	93,217	104,717
27	40,113	46,963	55,493	**65**	84,821	94,422	105,988
28	41,337	48,278	56,910	**66**	85,965	95,626	107,257
29	42,557	49,588	58,320	**67**	87,108	96,828	108,525
30	43,773	50,892	59,722	**68**	88,250	98,028	109,793
31	44,985	52,191	61,118	**69**	89,391	99,227	111,055
32	46,194	53,486	62,508	**70**	90,631	100,425	112,317
33	47,400	54,776	63,891	**71**	91,670	101,621	113,577
34	48,602	56,061	65,269	**72**	92,808	102,816	114,834
35	49,802	57,342	66,641	**73**	93,945	104,010	116,092
36	50,998	58,619	68,008	**74**	95,081	105,202	117,347
37	52,192	59,892	69,370	**75**	96,217	106,393	118,599
38	53,384	61,162	70,728	**76**	97,351	107,582	119,850

Додаток 3

Критичні значення критерію *t*-Стьюдента

df	**Рівень значущості (двобічна критична область)**					
	0,10	**0,05**	**0,02**	**0,01**	**0,002**	**0,001**
1	6,31	12,70	31,82	63,70	318,30	636,61
2	2,92	4,30	6,97	9,92	22,33	31,60
3	2,35	3,18	4,54	5,84	10,22	12,92
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,90	6,87
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,41
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,02
17	1,74	2,11	2,57	2,90	3,65	3,97
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,74
25	1,71	2,06	2,49	2,79	3,45	3,73
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	3,41	3,67
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29
	0,05	**0,025**	**0,01**	**0,005**	**0,001**	**0,0005**
	Рівень значущості (однобічна критична область)					

Додаток 4

Критичні значення критерію *F*-Фішера для перевірки ненаправлених альтернатив

α =0,05

		df чисельника											
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**10**	**12**	**24**	**…**
df знаменника	**3**	17,4	16,0	15,4	15,1	14,9	14,7	14,6	14,5	14,4	14,3	14,1	13,9
	5	10,0	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,62	6,53	6,28	6,02
	7	8,07	6,54	5,89	5,52	5,29	5,12	5,00	4,90	4,76	4,67	4,42	4,14
	10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,86	3,72	3,62	3,37	3,08
	11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,53	3,43	3,17	2,88
	12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,37	3,28	3,02	2,73
	13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,25	3,15	2,89	2,60
	14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,15	3,05	2,79	2,49
	15	6,20	4,77	4,15	3,80	3,58	3,42	3,29	3,20	3,06	2,96	2,70	2,40
	16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,13	2,99	2,89	2,63	2,32
	18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,87	2,77	2,50	2,19
	20	5,87	4,46	3,86	3,52	3,29	3,13	3,01	2,91	2,77	2,68	2,41	2,09
	30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,51	2,41	2,14	1,79
	40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,39	2,29	2,01	1,64
	50	5,34	3,98	3,39	3,05	2,83	2,67	2,55	2,46	2,32	2,22	1,93	1,55
	70	5,25	3,89	3,31	2,98	2,75	2,60	2,47	2,38	2,24	2,14	1,85	1,44
	100	5,18	3,83	3,25	2,92	2,70	2,54	2,42	2,32	2,18	2,08	1,78	1,35
	200	5,10	3,76	3,18	2,85	2,63	2,47	2,35	2,27	2,11	2,01	1,71	1,23
	…	5,03	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,05	1,95	1,64	

Додаток 5

Критичні значення критерію *U*-Манна-Уітні для ненаправлених альтернатив (двобічна область)

α = 0,05																		
$df_{1,2}$	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
4	-	0																
5	0	1	2															
6	1	2	3	5														
7	1	3	5	6	8													
8	2	4	6	8	10	13												
9	2	4	7	10	12	15	17											
10	3	5	8	11	14	17	20	23										
11	3	6	9	13	16	19	23	26	30									
12	4	7	11	14	18	22	26	29	33	37								
13	4	8	12	16	20	24	28	33	37	41	45							
14	5	9	13	17	22	26	31	36	40	45	50	55						
15	5	10	14	19	24	29	34	39	44	49	54	59	64					
16	6	11	15	21	26	31	37	42	47	53	59	64	70	75				
17	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87			
18	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99		
19	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	
20	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127
α = 0,01																		
5	-	-	0															
6	-	0	1	2														
7	-	0	1	3	4													
8	-	1	2	4	6	7												
9	0	1	3	5	7	9	11											
10	0	2	4	6	9	11	13	16										
11	0	2	5	7	10	13	16	18	21									
12	1	3	6	9	12	15	18	21	21	27								
13	1	3	7	10	13	17	20	24	24	31	34							
14	1	4	7	11	15	18	22	26	26	34	38	42						
15	2	5	8	12	16	20	24	29	29	37	42	46	51					
16	2	5	9	13	18	22	27	31	31	41	45	50	55	60				
17	2	6	10	15	19	24	29	34	34	44	49	54	60	65	70			
18	2	6	11	16	21	26	31	37	37	47	53	58	64	70	75	81		
19	3	7	12	17	22	28	33	39	39	51	56	63	69	74	81	87	93	
20	3	8	13	18	24	30	36	42	42	54	60	67	73	79	86	92	99	105

Додаток 6

Критичні значення критерію *U*-Манна-Уітні для направлених альтернатив (однобічна область)

α = 0,05																			
$n_{1,2}$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	-	0																	
4	-	0	1																
5	0	1	2	4															
6	0	2	3	5	7														
7	0	2	4	6	8	11													
8	1	3	5	8	10	13	15												
9	1	4	6	9	12	15	18	21											
10	1	4	7	11	14	17	20	24	27										
11	1	5	8	12	16	19	23	27	31	34									
12	2	5	9	13	17	21	26	30	34	38	42								
13	2	6	10	15	19	24	28	33	37	42	47	51							
14	3	7	11	16	21	26	31	36	41	46	51	56	61						
15	3	7	12	18	23	28	33	39	44	50	55	61	66	72					
16	3	8	14	19	25	30	36	42	48	54	60	65	71	77	83				
17	3	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96			
18	4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109		
19	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	
20	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138
α = 0,01																			
5	-	-	0	1															
6	-	-	1	2	3														
7	-	0	1	3	4	6													
8	-	0	2	4	6	7	9												
9	-	1	3	5	7	9	11	14											
10	-	1	3	6	8	11	13	16	19										
11	-	1	4	7	9	12	15	18	22	25									
12	-	2	5	8	11	14	17	21	24	28	31								
13	0	2	5	9	12	16	20	23	27	31	35	39							
14	0	2	6	10	13	17	22	26	30	34	38	43	47						
15	0	3	7	11	15	19	24	28	33	37	42	47	51	56					
16	0	3	7	12	16	21	26	31	36	41	46	51	56	61	66				
17	0	4	8	13	18	23	28	33	38	44	49	55	60	66	71	77			
18	0	4	9	14	19	24	30	36	41	47	53	59	65	70	76	82	88		
19	1	4	9	15	20	26	32	38	44	50	56	63	69	75	82	88	94	101	
20	1	5	10	16	22	28	34	40	47	53	60	67	73	80	87	93	100	107	114

Додаток 7

Критичні значення критерію *H*-Краскела-Уолліса для трьох вибірок чисельністю $n \leq 5$

Обсяги вибірок					Обсяги вибірок					Обсяги вибірок				
n_1	n_2	n_3	H	p	n_1	n_2	n_3	H	p	n_1	n_2	n_3	H	p
2	1	1	2,7000	0,500	4	4	1	6,6667	0,010	5	4	1	6,9545	0,008
2	2	1	3,6000	0,200				6,1667	0,022				6,8400	0,011
2	2	2	4,5714	0,067				4,9667	0,048				4,9855	0,044
3	1	1	3,2000	0,300				4,8667	0,054				4,8600	0,056
3	2	1	4,2857	0,100				4,1667	0,082				3,9873	0,098
			3,8571	0,133				4,0667	0,102				3,9600	0,102
3	2	2	5,3572	0,029	4	4	2	7,0364	0,006	5	4	2	7,2045	0,009
			4,7143	0,048				6,8727	0,011				7,1182	0,010
			4,5000	0,067				5,4545	0,046				5,2727	0,049
			4,4643	0,105				5,2364	0,052				5,2682	0,050
3	3	1	5,1429	0,043				4,5545	0,098				4,5409	0,098
			4,5714	0,100				4,4455	0,103				4,5182	0,101
			4,0000	0,129	4	4	3	7,1439	0,010	5	4	3	7,4449	0,010
3	3	2	6,2500	0,011				7,1364	0,011				7,3949	0,011
			5,3611	0,032				5,5985	0,049				5,6564	0,049
			5,1389	0,061				5,5758	0,051				5,6308	0,050
			4,5556	0,100				4,5455	0,099				4,5487	0,099
			4,2500	0,121				4,4773	0,102				4,5231	0,103
3	3	3	7,2000	0,004	4	4	4	7,6538	0,008	5	4	4	7,7604	0,009
			6,4889	0,011				7,5385	0,011				7,7440	0,011
			5,6889	0,029				5,6923	0,049				5,6571	0,049
			5,6000	0,050				5,6538	0,054				5,6176	0,050
			5,0667	0,086				4,6539	0,097				4,6187	0,100
			4,6222	0,100				4,5001	0,104				4,5527	0,102
4	1	1	3,5714	0,200	5	1	1	3,8571	0,143	5	5	1	7,3091	0,009
4	2	1	4,8214	0,057	5	2	1	5,2500	0,036				6,8364	0,011
			4,5000	0,076				5,0000	0,048				5,1273	0,046
			4,0179	0,114				4,4500	0,071				4,9091	0,053
4	2	2	6,0000	0,014				4,2000	0,095				4,1091	0,086
			5,3333	0,033				4,0500	0,119				4,0364	0,105
			5,1250	0,052	5	2	2	6,5333	0,008	5	5	2	7,3385	0,010
			4,4583	0,100				6,1333	0,013				7,2692	0,010
			4,1667	0,105				5,1600	0,034				5,3385	0,047
4	3	1	5,8333	0,021				5,0400	0,056				5,2462	0,051
			5,2083	0,050				4,3733	0,090				4,6231	0,097
			5,0000	0,057				4,2933	0,122				4,5077	0,100
			4,0556	0,093	5	3	1	6,4000	0,012	5	5	3	7,5780	0,010

Обсяги вибірок					Обсяги вибірок					Обсяги вибірок				
n_1	n_2	n_3	H	p	n_1	n_2	n_3	H	p	n_1	n_2	n_3	H	p
			3,8889	0,129				4,9600	0,048				7,5429	0,010
4	3	2	6,4444	0,008				4,8711	0,052				5,7055	0,046
			6,3000	0,011				4,0178	0,095				5,6264	0,051
			5,4444	0,046				3,8400	0,123				4,5451	0,100
			5,4000	0,051	5	3	2	6,9091	0,009				4,5363	0,102
			4,5111	0,098				6,8218	0,010	5	5	4	7,8229	0,010
			4,4444	0,102				5,2509	0,049				7,7914	0,010
4	3	3	6,7455	0,010				5,1055	0,052				5,6657	0,049
			6,7091	0,013				4,6509	0,091				5,6429	0,050
			5,7909	0,046				4,4945	0,101				4,5229	0,099
			5,7273	0,050	5	3	3	7,0788	0,009				4,5200	0,101
			4,7091	0,092				6,9818	0,011	5	5	5	8,0000	0,009
			4,7000	0,101				5,6485	0,049				7,9800	0,010
								5,5152	0,051				5,7800	0,049
								4,5333	0,097				5,6600	0,051
								4,4121	0,109				4,5600	0,100
													4,5000	0,102

Додаток 8

Критичні значення критерію *T*-Вілкоксона для ненаправлених альтернатив (двобічна область)

n	Рівень значущості		*n*	Рівень значущості	
	0,05	**0,01**		**0,01**	**0,05**
5	-	-	**28**	116	91
6	0	-	**29**	126	100
7	2	-	**30**	137	109
8	3	0	**31**	147	118
9	5	1	**32**	159	128
10	8	3	**33**	170	138
11	10	5	**34**	182	148
12	13	7	**35**	195	159
13	17	9	**36**	208	171
14	21	12	**37**	221	182
15	25	15	**38**	235	194
16	29	19	**39**	249	207
17	34	23	**40**	264	220
18	40	27	**41**	279	233
19	46	32	**42**	294	247
20	52	37	**43**	310	261
21	58	42	**44**	327	276
22	65	48	**45**	343	291
23	73	54	**46**	361	307
24	81	61	**47**	378	322
25	89	68	**48**	396	339
26	98	75	**49**	415	355
27	107	83	**50**	434	373

Додаток 9

Критичні значення критерію *T*-Вілкоксона для направлених альтернатив (однобічна область)

n	Рівень значущості		*n*	Рівень значущості	
	0,05	**0,01**		**0,01**	**0,05**
5	0	-	**28**	130	101
6	2	-	**29**	140	110
7	3	0	**30**	151	120
8	5	1	**31**	163	130
9	8	3	**32**	175	140
10	10	5	**33**	187	151
11	13	7	**34**	200	162
12	17	9	**35**	213	173
13	21	12	**36**	227	185
14	25	15	**37**	241	198
15	30	19	**38**	256	211
16	35	23	**39**	271	224
17	41	27	**40**	286	238
18	47	32	**41**	302	252
19	53	37	**42**	319	266
20	60	43	**43**	336	281
21	67	49	**44**	353	296
22	75	55	**45**	371	312
23	83	62	**46**	389	328
24	91	69	**47**	407	345
25	100	76	**48**	426	362
26	110	84	**49**	446	379
27	119	92	**50**	466	397

Додаток 10

Критичні значення критерію χ_r^2-Фрідмана
для кількості умов $k = 3$ та кількості досліджуваних $n < 10$

$n = 2$		$n = 3$		$n = 4$		$n = 5$	
χ_r^2	p	χ_r^2	p	χ_r^2	p	χ_r^2	p
0	1,000	0,000	1,000	0,0	1,000	0,0	1,000
1	0,833	0,667	0,944	0,5	0,931	0,4	0,954
3	0,500	2,000	0,528	1,5	0,653	1,2	0,691
4	0,167	2,667	0,361	2,0	0,431	1,6	0,522
		4,667	0,194	3,5	0,273	2,8	0,367
		6,000	0,028	4,5	0,125	3,6	0,182
				6,0	0,069	4,8	0,124
				6,5	0,042	5,2	0,093
				8,0	0,0046	6,4	0,039
						7,6	0,024
						8,4	0,0085
						10,0	0,00077

$n = 6$		$n = 7$		$n = 8$		$n = 9$	
χ_r^2	p	χ_r^2	p	χ_r^2	p	χ_r^2	p
0,00	1,000	0,000	1,000	0,00	1,000	0,000	1,000
0,33	0,956	0,286	0,964	0,25	0,967	0,222	0,971
1,00	0,740	0,857	0,768	0,75	0,794	0,667	0,814
1,33	0,570	1,143	0,620	1,00	0,654	0,889	0,865
2,33	0,430	2,000	0,486	1,75	0,531	1,556	0,569
3,00	0,252	2,571	0,305	2,25	0,355	2,000	0,398
4,00	0,184	3,429	0,237	3,00	0,285	2,667	0,328
4,33	0,142	3,714	0,192	3,25	0,236	2,889	0,278
5,33	0,072	4,571	0,112	4,00	0,149	3,556	0,187
6,33	0,052	5,429	0,085	4,75	0,120	4,222	0,154
7,00	0,029	6,000	0,052	5,25	0,079	4,667	0,107
8,33	0,012	7,143	0,027	6,25	0,047	5,556	0,069
9,00	0,0081	7,714	0,021	6,75	0,038	6,000	0,057
9,33	0,0055	8,000	0,016	7,00	0,030	6,222	0,048
10,33	0,0017	8,857	0,0084	7,75	0,018	6,889	0,031
12,00	0,00013	10,286	0,0036	9,00	0,0099	8,000	0,019
		10,571	0,0027	9,25	0,0080	8,222	0,016
		11,143	0,0012	9,75	0,0048	8,667	0,010
		12,286	0,00032	10,75	0,0024	9,556	0,0060
		14,000	0,000021	12,00	0,0011	10,667	0,0035
				12,25	0,00086	10,889	0,0029
				13,00	0,00026	11,556	0,0013
				14,25	0,000061	12,667	0,00066
				16,00	0,0000036	13,556	0,00035
						14,000	0,00020
						14,222	0,000097
						14,889	0,000054
						16,222	0,000011
						18,000	0,0000006

Додаток 11

Критичні значення критерію χ_r^2-Фрідмана для кількості умов $k = 4$ та кількості досліджуваних $n < 5$

$n = 2$		$n = 3$		$n = 4$			
χ_r^2	p	χ_r^2	p	χ_r^2	p	χ_r^2	p
0	1,000	0,0	1,000	0,0	1,000	5,7	0,141
0,6	0,958	0,6	0,958	0,3	0,992	6,0	0,105
1,2	0,834	1,0	0,910	0,6	0,928	6,3	0,094
1,8	0,792	1,8	0,727	0,9	0,900	6,6	0,077
2,4	0,625	2,2	0,608	1,2	0,800	6,9	0,068
3,0	0,542	2,6	0,524	1,5	0,754	7,2	0,054
3,6	0,458	3,4	0,446	1,8	0,677	7,5	0,052
4,2	0,375	3,8	0,342	2,1	0,649	7,8	0,036
4,8	0,208	4,2	0,300	2,4	0,524	81	0,033
5,4	0,167	5,0	0,207	2,7	0,508	8,4	0,019
6,0	0,042	5,4	0,175	3,0	0,432	8,7	0,014
		5,8	0,148	3,3	0,389	9,3	0,012
		6,6	0,075	3,6	0,355	9,6	0,0069
		7,0	0,054	3,9	0,324	9,9	0,0062
		7,4	0,033	4,5	0,242	10,2	0,0027
		8,2	0,017	4,8	0,200	10,8	0,0016
		9,0	0,0017	5,1	0,190	11,1	0,00094
				5,4	0,158	12,0	0,000072

Навчальне видання

Боснюк Валерій Федорович

МАТЕМАТИЧНІ МЕТОДИ В ПСИХОЛОГІЇ

Навчальний посібник

Формат 60х84 1/16. Ум.друк.арк. 9,44. Тираж ____ пр. Зам. №____.
Національний університет цивільного захисту України.
18034, м. Черкаси, вул. Онопрієнка, 8.