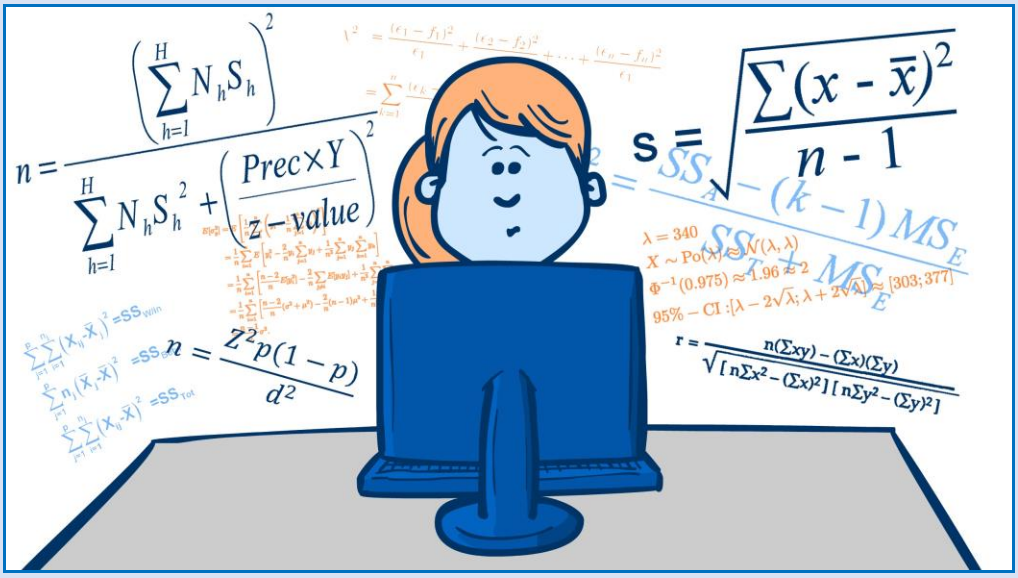
Міністерство освіти і науки України
Кам'янець-Подільський національний університет імені Івана Огієнка

У. В. ГУДИМА,
О. В. ГУДИМА

МАТЕМАТИЧНІ МЕТОДИ В ПСИХОЛОГІЇ: ОСНОВНІ ПОНЯТТЯ ТА ПРИКЛАДИ

НАВЧАЛЬНИЙ ПОСІБНИК



ЕЛЕКТРОННЕ ВИДАННЯ

Кам’янець-Подільський
2023

УДК 159.9.018:311.21(075.8)
ББК 88+60.6я73
Г93

Рекомендувала вчена рада Кам'янець-Подільського національного університету імені Івана Огієнка (протокол № 12 від 29.12.2022 року)

РЕЦЕНЗЕНТИ:

Ліана ОНУФРІЄВА, доктор психологічних наук, професор, професор кафедри загальної та практичної психології Кам'янець-Подільського національного університету імені Івана Огієнка;
Ірина СЕМЕНИШИНА, кандидат фізико-математичних наук, доцент, асистент кафедри математики, інформатики та академічного письма Закладу вищої освіти «Подільський державний університет»;
Оксана ЛЯСКА, кандидат психологічних наук, доцент, доцент кафедри права, професійної і соціально-гуманітарної освіти Закладу вищої освіти «Подільський державний університет».

Гудима У. В., Гудима О. В.
Г93 Математичні методи в психології: основні поняття та приклади: навчальний посібник [Електронний ресурс]. Кам'янець-Подільський: Кам'янець-Подільський національний університет імені Івана Огієнка, 2023. 150 с.

Електронна версія посібника доступна за покликаннями:
URL: http://elar.kpnu.edu.ua:8081/xmlui/handle/123456789/6970

Метою навчального посібника є сформувати у здобувачів вищої освіти, що навчаються за спеціальністю 053 Психологія, навички використання методів математичної статистики в психології, переконання у необхідності їх використання під час обробки експериментальних даних.

У посібнику розглядають основні поняття математичних методів у психології, алгоритми дослідження сукупностей даних, які продемонстровано на прикладах.

УДК 159.9.018:311.21(075.8)
ББК 88+60.6я73

ПЕРЕДМОВА

В історії застосування психологією математичних методів були різні періоди: від абсолютизації їх можливостей і вимоги обов'язкового їх застосування в дослідженнях психологічних явищ, до повного вилучення їх з психологічної практики.

Сьогодні математичний апарат у психології використовують для статистичної обробки результатів дослідження, планування експериментів і прогнозування очікуваних результатів, розробки і побудови математичних моделей різних явищ, процесів і станів.

У психології використовується:

- математичне моделювання – експеримент з ідеальними моделями, які описуються математичними засобами;
- статистика – опис, реєстрація і аналіз кількісних даних результатів психологічного дослідження;
- психометрія – психологічний тест, розроблений та перевірений за допомогою математики;
- оцінка кількісних даних.

Проведення дослідження у психології передбачає збір даних (кількісних та якісних), їх обробку та інтерпретацію, що вимагає знання статичних методів обробки експериментальних даних. Адже в результаті експерименту отримуємо вибіркові сукупності вимірюваних величин, які мають випадковий характер. Тому збір та обробка статистичних даних повинні ґрунтуватися на строгих методах математичної статистики, методологічною основою яких є теорія ймовірностей.

Метою математичної обробки експериментальних даних є побудова корисної аналітичної моделі досліджуваного явища або процесу на основі скінченної вибіркової сукупності експериментальних даних.

Найпоширенішими статистичними методами, що використовують в психологічному дослідженні, є методи описової статистики, методи перевірки статистичних гіпотез, кореляційний аналіз, регресійний аналіз тощо.

Навчальний посібник має на меті ознайомити здобувачів вищої освіти з основними розділами навчальної дисципліни «Математичні методи в психології», зокрема, він містить основні означення, методи та алгоритми обробки експериментальних даних, а також зразки розв’язування задач.

Навчальне видання розраховане для здобувачів вищої освіти зі спеціальності 053 Психологія.

1. ОСНОВНІ ПОНЯТТЯ

1.1. ЗМІННІ ТА ЇХ ВИМІРЮВАННЯ. ШКАЛИ ВИМІРЮВАННЯ

До базових понять, які стосуються використання математичної статистики у психології, можна перш за все віднести поняття «змінні» та поняття «вимірювання».

Змінні – це все, що можна вимірювати, контролювати або змінювати у дослідженнях. Змінною може бути рівень особистісної тривожності, інтелект, рівень агресії, мотивація досягнення, тип темпераменту тощо.

Що буде змінною у дослідженні і скільки їх буде, залежить від теми, мети, гіпотези, методів, завдань та структури дослідження.

У структурі будь-якого дослідження, включаючи психологічне, виділяють незалежні, залежні та додаткові змінні.

Незалежні змінні – активно змінюванні експериментатором умови.

У психології як незалежні змінні можуть використовуватися, наприклад:

- якісні та кількісні характеристики стимулів;
- навколишні фізичні умови;
- фізіологічні фактори: режим сну, вживання лікарських препаратів та ін.;
- фактори діяльності: зміст та складність завдань, навантаження, перешкоди, рівень сформованості навичок;
- фактори мотивації: попередні успіхи чи невдачі, нагороди та покарання тощо.

Залежні змінні – показники, що реєструються та вимірюються експериментатором, які змінюються під впливом незалежних змінних, наприклад:

- параметри фізіологічних, рухових та мовних реакцій, сенсорноперцептивних, мнемічних та розумових процесів;
- швидкість і точність діяльності, тривалість, темп і частота різних дій;

- психічні стани;
- суб’єктивні оцінки.

Додаткові змінні – інші умови, які можуть впливати на залежні змінні, і які також необхідно враховувати, наприклад:

- демографічні характеристики досліджуваних: стать, вік, національність, освіта, професія, сімейний стан, рід занять тощо;
- індивідуальні особливості досліджуваних: інтелект, риси особистості;
- наявний досвід, знання та навички;
- стан здоров’я;
- психічний стан;
- зміст та умови діяльності, параметри завдань.

Будь-яке наукове дослідження передбачає вимірювання та фіксацію вираженості діагностованих властивостей у досліджуваного, як правило, за допомогою чисел.

Вимірювання – процедура присвоєння числових характеристик властивостям явищ, що вивчаються в психології, наприклад, моторним і мовним реакціям, відчуттям, здібностям, мотивам, установкам і вчинкам особистості, її статусу в групі тощо.

Шкала – інструмент для вимірювання неперервних властивостей об’єктів, є числовою системою, в якій відношення між різними властивостями об'єктів виражені властивостями числового ряду. Застосування шкал пов’язане з необхідністю якісної та кількісної оцінки певних ознак та змінних досліджуваних об’єктів.

Шкала вимірювання формується на основі двох ключових понять – вимірювання та масштабування.

Шкалювання є одним із найважливіших засобів математичного аналізу досліджуваного явища, а також способом організації емпіричних даних, одержуваних за допомогою спостереження, вивчення документів, анкетного опитування, експериментів, тестування. Більшість соціальних та психологічних об'єктів не може бути строго фіксованим щодо місця та часу свого існування, і тому виникає питання про специфіку числової системи, яка могла б співвідноситися з емпіричними даними такого роду.

Різні методи шкалювання і є особливими прийомами трансформації якісних характеристик в деяку кількісну змінну.

Найпоширенішою у психології є типологія шкал С. Стівенса, в основу якої покладено точність градуювання шкал та операції, які можна виконувати з числами. В межах цієї типології виділяють такі типи вимірювальних шкал:

- шкала найменувань (номінальна);
- шкала порядку (рангова чи ординальна);
- шкала інтервалів (інтервальна);
- шкала відношень (пропорційна).

Характеристика вимірювальних шкал С. Стівенса

Номінальна шкала. Номінальна шкала, або, як її ще називають шкала найменувань, вимірювань, або класифікацій – це найпростіший тип шкал, заснований на приписуванні чисел якісним властивостям об’єктів. У межах цієї шкали об’єкти класифікуються, а класи позначаються номерами. Таким чином, число тут слугує лише назвою певного класу, а тому нічого не говорить про властивості об’єкту, крім того, що він належить до певного класу. Єдиний вид статистичного аналізу, який можна виконати з використанням номінальної шкали, це обчислення процентних часток.

Частковим випадком шкали найменувань є дихотомічна шкала. При вимірюванні у дихотомічній шкалі ознаки можна кодувати двома символами або цифрами, наприклад 0 і 1, або буквами А і Б, а також будь-якими двома відмінними один від одного символами.

Наприклад типам темпераменту приписуємо номери: холерик – 1, сангвінік – 2, меланхолік – 3, флегматик – 4; поділ за статтю: жіноча – 0, чоловіча – 1.

Шкала порядку. Шкала порядку (рангова чи ординальна) передбачає ранжування (упорядкування) значень змінної залежно від масштабування. Елементи в порядковій шкалі зазвичай розташовуються в порядку зростання або спадання у них певної якості. При цьому кожній градації якості приписується свій порядковий номер (ранг). Фактично, об’єкти лише впорядковуються.

Особливістю цієї шкали є те, що однакові різниці між сусідніми рангами не означають однакової різниці між ступенями прояву виміряної якості.

Наприклад: розподіл місць на змаганнях: найкращому приписуємо перше місце, далі друге, третє.

Шкала інтервалів. Шкала інтервалів (інтервальна) – це шкала, у якій рівні упорядковані, а інтервали між ними рівні. Її можна розглядати, як розширення порядкової шкали. Основною відмінністю є властивість рівних інтервалів. Інтервальна шкала не тільки дозволяє однозначно визначити, яке значення більше (менше), а й на скільки. Крім того, на відміну від порядкової та номінальної шкал, в інтервальній можуть виконуватися арифметичні операції. Однакові різниці між числами виражають однакові відмінності у проявах вимірюваної якості. Особливістю шкали є довільний нуль, який не говорить про відсутність якості у об’єкта.

Наприклад: календарний час, шкала температур за Цельсієм.

Шкала відношень. Шкала відношень (пропорційна) може розглядатися як розширення інтервальної шкали. У межах цієї шкали числа мають такі ж властивості, як і в шкалі інтервалів, але, крім того, відношення чисел виражають кількісні відношення ступенів прояву явища.

Особливість шкали є наявність абсолютного нуля, який означає відсутність явища.

Наприклад: зріст, вага, температура за Кельвіном, рівень інтелекту, мотивація тощо.

В залежності від вибору шкали в психології виокремлюють два підходи до психологічних вимірювань: метричний (більш точний) і неметричний (менш точний). До неметричних відносять номінальну та рангову шкали вимірювань, а до метричних – шкалу інтервалів та шкалу відношень.

1.2. ГЕНЕРАЛЬНА СУКУПНІСТЬ ТА ВИБІРКА. РЕПРЕЗЕНТАТИВНІСТЬ ВИБІРКИ

Генеральна сукупність – це сукупність усіх об'єктів, що володіють якостями та властивостями, які цікавлять дослідника та відповідають меті дослідження.

Вибірка – це необхідний для дослідження мінімум результатів (випадків, досліджуваних об'єктів, подій), відібраних за допомогою визначеної процедури з генеральної сукупності.

На основі аналізу вибірки можна отримати досить повне уявлення про закономірності, властиві всій генеральній сукупності.

Визначальним фактором для подальшого застосування суб'єктів вибіркової сукупності є оцінювання її якості за показниками репрезентативності та надійності.

Надійність визначається за такими параметрами:

- *повнота вибірки* (в ній представлені всі елементи генеральної сукупності);
- *точність інформації* (у ній немає неіснуючих одиниць спостереження);
- *адекватність* (вибірка співвідноситься з розв'язанням поставлених дослідженням завдань).

Вибіркова сукупність має бути репрезентативною, тобто у відібраній частині мають бути представлені всі елементи й у такому співвідношенні, як і у генеральній сукупності.

Репрезентативність – це відповідність характеристик вибірки генеральній сукупності загалом.

Інакше кажучи, вибіркова сукупність має відображати властивості генеральної сукупності, правильно її представляти.

Репрезентативність має бути кількісною та якісною.

Кількісна – заснована на законі великих чисел і означає достатню чисельність елементів вибіркової сукупності, що розраховується за спеціальними формулами та таблицями.

Якісна – заснована на законі ймовірності і означає відповідність (однотипність) ознак, що характеризують елементи вибіркової сукупності по відношенню до генеральної.

Можна виокремити такі вибірки.

За способом відбору.

Випадкова – якщо елементи відбираються випадковим чином. Так як більшість методів математичної статистики ґрунтується на понятті випадкової вибірки, то природно, що вибірка повинна бути випадковою.

Невипадкова вибірка:

- механічний відбір – сукупність ділиться на стільки частин, скільки одиниць планується у вибірці і потім з кожної частини відбирається один елемент;
- типовий відбір – сукупність ділиться на гомогенні частини, і з кожної здійснюється випадкова вибірка;
- серійний відбір, при якому у вибіркову сукупність відбираються групи (серії), які надалі досліджуються суцільно.
- комбінований відбір – поєднуються представлені види відбору, на різних етапах.

За схемою випробувань – вибірки можуть бути незалежні і залежні.

Незалежні вибірки характеризуються тим, що ймовірність відбору будь-якого досліджуваного однієї вибірки не залежить від відбору будь-якого досліджуваного з іншої вибірки. Навпаки, залежні вибірки характеризуються тим, що кожному досліджуваному однієї вибірки поставлений у відповідність за певним критерієм досліджуваний з іншої вибірки.

За обсягом вибірки ділять на малі і великі. До малих відносять вибірки, в яких число елементів $n \leq 30$. Поняття великої вибірки не визначено, але великою вважається вибірка в якій число елементів $n > 200$ і середня вибірка задовольняє умові $30 \leq n \leq 200$. Цей поділ є умовним.

Малі вибірки використовуються при статистичному контролі відомих властивостей. Великі вибірки використовуються для дослідження невідомих властивостей і параметрів.

1.3. РІВНІ ЗНАЧУЩОСТІ

Рівень значущості (чи інакше, поріг достовірності, β) є показником ймовірності безпомилкових висновків та прогнозів. Найчастіше у статистиці використовуються чотири стандартні рівні значущості – нульовий ($\beta_0 = 0,90$), перший ($\beta_1 = 0,95$), другий ($\beta_2 = 0,99$) і третій ($\beta_3 = 0,999$). Іншими словами, якщо дослідник задає нульовий рівень значущості, то його висновки та прогнози справедливі у 90 % випадків (ймовірність дорівнює 0,90), якщо перший рівень – у 95 % випадків і т. д.

Більшість існуючих статистичних таблиць засновані саме на цих стандартних рівнях.

Слід зазначити, що у психологічних дослідженнях рівень значущості 0,95, як правило, цілком достатній для формулювання тих чи інших висновків та прогнозів. Вищі рівні у низці психологічних досліджень майже недосяжні і використовуються тоді, коли до дослідження ставляться підвищені вимоги.

Робота на кожному рівні значущості передбачає мінімальний обсяг вибіркової сукупності, на якій проводиться дослідження. Так, якщо обсяг вибірки (n) – від 20 до 30 досліджуваних, тоді використовують тільки нульовий рівень значущості, при $n > 30$ – нульовий і перший рівень, при $n > 100$ – рівні β_0, β_1, β_2, і, нарешті, при $n > 200$ – всі чотири рівні.

Деякі дослідники в якості рівня значущості використовують величину α (p), що дорівнює $1 - \beta$. У цьому випадку рівні значущості набувають наступного вигляду: $\alpha_0 \leq 0,10$, $\alpha_1 \leq 0,05$, $\alpha_2 \leq 0,01$, $\alpha_3 \leq 0,001$. Логічний зміст цих величин у тому, що вони характеризують ймовірність випадковості (ймовірність хибних прогнозів). Іншими словами, це та ймовірність, яка припадає на долю випадкових (як правило, непередбачуваних) факторів.

1.4. ФОРМИ ОБЛІКУ РЕЗУЛЬТАТІВ ВИМІРЮВАНЬ. ТАБЛИЦІ, ГРАФІКИ, ДІАГРАМИ

При описанні результатів емпіричного дослідження необхідно обробити велику кількість фактичного та цифрового матеріалу.

Для впорядкування даних, представлення їх в найбільш компактній, зручній для обробки формі використовують таблиці, графіки, діаграми.

Статистична таблиця – це форма наочного і систематизованого зображення числових результатів зведення і обробки статистичних даних.

Таблиці складають як на заключному етапі дослідження, так і в процесі обробки даних (допоміжні і робочі таблиці).

У статистичній таблиці розрізняють підмет і присудок.

Підметом статистичної таблиці є об'єкти, що характеризуються числовими показниками, тобто те, про що йдеться в таблиці.

Присудком статистичної таблиці є числові показники, що характеризують статистичну сукупність.

За побудовою підмета таблиці поділяють на три види: прості, групові, комбінаційні.

Проста таблиця містить лише перелік одиниць статистичної сукупності.

Групова таблиця містить у підметі зведення про сукупність, розподілену на окремі групи за однією ознакою. При цьому кожна група може бути охарактеризована рядом показників.

Комбінаційна таблиця містить дані, згруповані за двома і більше ознаками. Інколи у комбінаційних таблицях групи за однією ознакою розміщують у підметі, а за другою – у присудку.

Використання графіків для зображення статистичних показників дозволяє надати останнім наочність і виразність, полегшити їх сприйняття, а в багатьох випадках допомагає усвідомити суть досліджуваного явища, його закономірності та особливості, побачити тенденції змін, взаємозв'язок показників, що його характеризують.

Графік є лінією, яка зображає залежність між змінними. На осях відкладаються значення досліджуваних кількісних показників. При використанні двовимірного графіка на осі абсцис зазвичай розміщують незалежну змінну, по осі ординат – залежну змінну.

Діаграми використовуються головним чином для відображення співвідношень між величинами.

Це спосіб графічного зображення величин за допомогою фігур (секторів, стовпців тощо), площі яких пропорційні величинам.

Секторна діаграма – діаграма, де числа (зазвичай відсотки) зображені у вигляді кругових секторів.

Гістограма – це стовпчаста діаграма, що складається з вертикальних прямокутників, розташованих основами на одній прямій (наприклад, осі абсцис).

Гістограми часто використовуються для графічного представлення щільності розподілу (частотного розподілу), при якому кількість випадків у класі зображується у вигляді вертикальних смуг (стовпчиків).

Аналог діаграми – полігон. Цей графічний спосіб відображення даних використовується переважно для зображення дискретних рядів.

1.5. СТАТИСТИЧНІ ГІПОТЕЗИ

Одне з основних завдань, що стоїть перед психологом, який проводить емпіричне дослідження, – це з'ясування того, які висновки про властивості генеральної сукупності можна зробити за вибірковим спостереженням.

Оцінка параметрів генеральної сукупності, зроблена на підставі вибіркових даних, неминуче супроводжується похибкою і тому розглядається як ймовірне, а не остаточне твердження.

Подібні припущення про властивості та параметри генеральної сукупності називаються статистичними гіпотезами.

Статистична гіпотеза – наукова гіпотеза, яка допускає статистичну перевірку.

Статистичні гіпотези поділяють на два типи: **нульову** та **альтернативну.**

Нульова гіпотеза – це гіпотеза про відсутність відмінностей.

Вона позначається H_0 і називається нульовою тому, що містить 0: $x_1 - x_2 = 0$, де x_1 та x_2 – показники, між якими з'ясовують відмінність.

Альтернативна гіпотеза – це гіпотеза про існування та значимість відмінностей. Позначається H_1.

Висунута гіпотеза може бути правильною чи неправильною, тому виникає необхідність перевірити її. Оскільки перевірку роблять статистичними методами, то ця перевірка називається статистичною.

При перевірці статистичних гіпотез можливі помилки (помилкові судження) двох видів:

- можна відкинути нульову гіпотезу, коли вона насправді вірна (так звана помилка першого роду);
- можна прийняти нульову гіпотезу, коли вона насправді не вірна (так звана помилка другого роду).

Перевірка статистичних гіпотез здійснюється за допомогою статистичних критеріїв.

Статистичний критерій – це метод обчислення, який дозволяє прийняти істинну та відхилити хибну статистичну гіпотезу.

Усі критерії використовуються з однією головною метою: визначити рівень значимості проаналізованих з їх допомогою даних (тобто, ймовірність того, що ці дані відображають справжній ефект, правильно представляють популяцію, з якої сформовано вибірку).

Усі критерії різняться за потужністю.

Потужність критерію – це його здатність виявляти відмінності чи відхиляти нульову гіпотезу, якщо вона невірна.

Велика різноманітність критеріїв надає наступні можливості:

- вибирати критерії, адекватні типу шкали, у якій отримано експериментальні дані;

- працювати зі зв’язними (залежними) та незв'язними (незалежними) вибірками;
- працювати з нерівними за обсягом вибірками;
- вибирати із критеріїв різні за потужністю (залежно від цілей дослідження).

Критична область – сукупність значень критерію, у якому нульову гіпотезу відхиляють.

Область прийняття нульової гіпотези (область допустимих значень) – сукупність значень критерію, у якому нульову гіпотезу приймають. При справедливості нульової гіпотези ймовірність того, що статистика критерію потрапляє в область прийняття нульової гіпотези, повинна дорівнювати $(1 - \alpha)$. Число ступенів свободи у будь-якого параметра визначають як кількість досліджень, за якими розрахований даний параметр, мінус кількість однакових значень, знайдених за цими дослідженнями незалежно один від одного.

Позначається буквою ν або df .

Усі критерії перевірки нульових гіпотез побудовані за єдиним принципом. Вихідними даними для перевірки зазвичай є дві випадкові вибірки, отримані в результаті вимірювань. Для різних критеріїв на основі виконаних вимірювань запропоновані процедури обчислення спеціальних величин, які називаються статистиками (значення статистики, що спостерігаються). Вони різні для різних критеріїв і залежать від обсягів випадкових вибірок.

Перевірка статистичних гіпотез складається з наступних кроків:

- формулюється завдання дослідження у вигляді статистичної гіпотези;
- вибирається статистична характеристика гіпотези;
- аналізуються можливі помилкові рішення та оцінюються їх наслідки;
- формулюються нульова (H_0) та альтернативна (H_1) гіпотези;
- визначається рівень значущості α, і визначається критичне значення статистичної характеристики критерію;

- обчислюється фактичне (експериментальне) значення статистичної характеристики, порівнюється з критичним значенням, приймається рішення щодо сформульованих гіпотез.

Статистичні критерії є **параметричними** та **непараметричними.**

Параметричні критерії включають в себе параметри розподілу (середнє арифметичне та дисперсію) і дають достовірні результати, коли дані розподілені нормально. Вони можуть бути використані до великих вибірок.

Непараметричні критерії не включають параметрів розподілу, а ґрунтуються на оперуванні частотами чи рангами, а тому можуть бути використані до малих вибірок. Крім того, їх простіше обчислювати.

2. ОСНОВНІ СТАТИСТИЧНІ ХАРАКТЕРИСТИКИ ВИБІРКИ. НОРМАЛЬНИЙ ЗАКОН РОЗПОДІЛУ ВИПАДКОВИХ ВЕЛИЧИН

2.1. МІРИ ЦЕНТРАЛЬНОЇ ТЕНДЕНЦІЇ

Міра центральної тенденції – це число, що характеризує вибірку за рівнем вираженості вимірюваної ознаки. Вони показують, навколо яких значень групується більшість експериментальних даних.

Існує три способи визначення «центральної тенденції», кожному з яких відповідає своя міра: мода, медіана, вибіркове середнє.

Мода **–** це значення у множині спостережень, яке зустрічається найчастіше (Mo).

Наприклад, для множини даних спостереження 1, 1, 2, 3, 3, 4, 4, 4, 7, 7, 7, 7, 9, 9 мода дорівнює 7:

$$Mo = Mo(1,\ 1,\ 2,\ 3,\ 3,\ 4,\ 4,\ 4,\ 7,\ 7,\ 7,\ 7,\ 9,\ 9) = 7\,.$$

Правила обрахування моди:

1. Якщо значення даних у вибірці зустрічається однаково часто, то кажуть, що цей ряд немає моди.
   Наприклад ряд $\{1,\ 1,\ 2,\ 2,\ 4,\ 4,\ 7{,}7\}$ – немає моди.
2. Якщо два сусідні значення мають однакову частоту, то модою називають їх середнє:
   $$Mo(1, 2, 2, 3, 3, 4) = 2{,}5.$$
3. Якщо два несусідні значення мають однакову частоту, то кажуть, що дані мають дві моди, а ряд даних називається бімодальним:
   $$Mo(1, 1, 1, 2, 3, 4, 5, 5, 5) = 1 \text{ та } 5.$$

Мода використовується головним чином для того, щоб дати загальну уяву про розподіл.

Медіана – це значення, яке ділить **упорядковану** множину даних навпіл, так що одна половина даних виявляється меншою за медіану, а друга – більшою (Md).

Правила обчислення:

При обчислені медіани слід пам'ятати, що дані ряду повинні бути ранжовані (впорядковані) у порядку зростання або спадання.

Розглядають два випадки: 1. ряд містить непарну кількість значень; 2. ряд містить парну кількість значень.

1. Якщо ряд містить непарну кількість значень, то медіана є центральним значенням: $Md(11, 13, \mathbf{25}, 48, 49) = 25$.
2. Якщо ряд містить парну кількість значень, то медіана обраховується як середнє двох центральних значень:

$$Md(11,\mathbf{13},\mathbf{25},48) = \frac{13+25}{2} = 19.$$

Середнім арифметичним ряду називають частку від ділення суми даних ряду на кількість доданків ($\overline{x}$).

Для ряду $x_1, ..., x_n$ середнє арифметичне обраховується за формулою:

$$\overline{x} = \frac{x_1 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

Наприклад, середнє арифметичне ряду: 2, 5, 5, 7, 8, 9, 11, 12, 13, 13, обчислюється таким чином:

$$\overline{x} = \frac{2+5+5+7+8+9+11+12+13+13}{10} = 8,5.$$

Якщо сукупність даних поділена на групи з однаковим значенням ознаки (частотою ознаки), тоді для обчислення середнього арифметичного використовують формулу:

$$\overline{x} = \frac{\sum_{i=1}^{m} n_i x_i}{\sum_{i=1}^{m} n_i},$$

де n_i – частота елемента x_i, тобто кількість об'єктів, що мають ознаку x_i.

<u>Приклад 2.1.</u> Знайти середнє арифметичне ряду:

2, 2, 2, 5, 5, 6, 6, 8, 9, 11, 12, 13, 13, 17, 17, 17.

Розв'язування. Поділимо данні на групи з однаковим значенням ознаки:

x_i	2	5	6	8	9	11	12	13	17
n_i	3	2	2	1	1	1	1	2	3

Обрахуємо середнє арифметичне:

$$\bar{x} = \frac{2\cdot 3 + 5\cdot 2 + 6\cdot 2 + 8\cdot 1 + 9\cdot 1 + 11\cdot 1 + 12\cdot 1 + 13\cdot 2 + 17\cdot 3}{3+2+2+1+1+1+1+2+3} \approx 9{,}06 .$$

Властивості середнього арифметичного:

1. Сума відхилень індивідуальних значень ознак від середнього арифметичного дорівнює нулю: $\sum_{i=1}^{n}\left(x_i - \bar{x}\right) = 0$.
2. Якщо кожен член ряду помножити (чи поділити) на константу c (у випадку ділення $c \neq 0$), то середнє арифметичне утвореного ряду буде рівне середньому арифметичному початкового ряду помноженого (поділеного) на константу c.

   Наприклад, якщо для ряду $x_1,...,x_n$ середнє арифметичне дорівнює $\bar{x}$, то для ряду $cx_1,...,cx_n$ середнє арифметичне дорівнює $c\bar{x}$.
3. Якщо до кожного члену ряду додати (відняти) константу c, то середнє арифметичне утвореного ряду буде дорівнювати сумі (різниці) середнього арифметичного початкового ряду та константи c.

   Наприклад, якщо для ряду $x_1,...,x_n$ середнє арифметичне дорівнює $\bar{x}$, то для ряду $x_1 + c,...,x_n + c$ середнє арифметичне дорівнює $\bar{x} + c$.

2.2. МІРИ МІНЛИВОСТІ

Міри центральної тенденції відображають рівень вираженості вимірюваної ознаки. Разом з тим, не менш важливою характеристикою є вираженість індивідуальних відмінностей досліджуваних за вимірюваною ознакою. Міри мінливості застосовують у психології для чисельного вираження величини міжіндивідуальної варіації ознаки.

Основними показниками міри мінливості (розсіювання варіант) є **розмах**, **дисперсія**, **стандартне відхилення**.

Розмах (R) – різниця максимального і мінімального значення в групі:

$$R = x_{\max} - x_{\min}.$$

<u>Приклад 2.2.</u> Обчислити розмах ряду даних:

$$R(2,2,2,5,5,6,6,8,9,11,12,13,13,17,17,17) = 17 - 2 = 15.$$

На розмах, однак, не впливають дані, що лежать між максимальним та мінімальним показниками.

Отже, розмах є досить грубою мірою мінливості. Однак, він є досить поширеною мірою, оскільки дозволяє обчислити коефіцієнт осциляції.

Коефіцієнт осциляції відображає відносні коливання крайніх значень ряду відносно середнього показника: $K_0 = \frac{R}{\bar{x}} \cdot 100$.

<u>Приклад 2.3.</u> Обчислити коефіцієнт осциляції для ряду даних:

$$K_0(2,2,2,5,5,6,6,8,9,11,12,13,13,17,17,17) = \frac{15}{8,9} \cdot 100 = 168,5\%.$$

Дисперсія – це середнє арифметичне квадратів відхилень значень змінних ряду від його середнього арифметичного. Її обчислюють за такою формулою:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

Вибіркова дисперсія є зміщеною оцінкою дисперсії генеральної сукупності (дає занижені значення для генеральної дисперсії). Тому вибіркову дисперсію «виправляють» таким чином, щоб вона стала незміщеною оцінкою. Для цього вибіркову дисперсію множать на коефіцієнт Бесселя $\frac{n}{n-1}$:

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 .$$

Дисперсія – це параметр, що характеризує ступінь розкиду елементів вибірки щодо середнього значення. Чим більше дисперсія тим більше відхиляються значення елементів вибірки від середнього арифметичного.

Властивості дисперсії:

1. Якщо кожен член ряду помножити на константу c, то дисперсія утвореного ряду збільшиться в c^2.
2. Якщо до кожного члену ряду додати константу c, то дисперсія утвореного ряду не зміниться.
3. При об'єднанні двох вибірок з однаковою дисперсією, але з різними середніми арифметичними дисперсія збільшується.

Дуже тісно пов'язане з дисперсією стандартне відхилення (середнє квадратичне відхилення).

Для того щоб отримати показник, який можна співставити з середнім відхиленням, добувають із дисперсії квадратний корінь. При цьому отримують стандартне відхилення:

$$S_x = \sigma_x = \sqrt{\sigma^2} .$$

Коефіцієнт варіації відображає відносні коливання всіх значень ряду відносно його середнього арифметичного:

$$V = \frac{S_x}{\overline{x}} \cdot 100\% .$$

Приклад 2.4. Знайти дисперсію, стандартне відхилення та коефіцієнт варіації ряду даних: 2, 2, 2, 5, 5, 6, 6, 8, 9, 11, 12, 13, 13, 17, 17, 17.

Розв'язування. Для зручного обчислення дисперсії використовують таблицю даних та проміжних обрахунків:

x_i	$x_i - \overline{x}$	$\left(x_i - \overline{x}\right)^2$
2	-7,06	49,8436
2	-7,06	49,8436
2	-7,06	49,8436
5	-4,06	16,4836
5	-4,06	16,4836
6	-3,06	9,3636
6	-3,06	9,3636
8	-1,06	1,1236
9	-0,06	0,0036
11	1,94	3,7636
12	2,94	8,6436
13	3,94	15,5236
13	3,94	15,5236
17	7,94	63,0436
17	7,94	63,0436
17	7,94	63,0436
$\overline{x}$ =9,06	$\sum_{i=1}^{16}\left(x_i - \overline{x}\right)$ =0	$\sum_{i=1}^{16}\left(x_i - \overline{x}\right)^2$ =434,94

Обрахуємо дисперсію та стандартне відхилення:

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \frac{434,94}{16-1} \approx 29\,,\ S_x = \sqrt{\sigma^2} = \sqrt{29} \approx 5,39\,.$$

Тоді коефіцієнт варіації буде дорівнювати:

$$V = \frac{S_x}{\overline{x}} \cdot 100 = \frac{5,39}{9,06} \cdot 100 = 59,49\%\,.$$

2.3. НОРМАЛЬНИЙ ЗАКОН РОЗПОДІЛУ ВИПАДКОВИХ ВЕЛИЧИН

Серед розподілів випадкових величин особливо виділяється так званий нормальний розподіл. Його назва пов'язана з тим, що нормальний розподіл є граничним випадком розподілу найрізноманітніших масових процесів, зокрема і в психології. Він лежить в основі вимірювань, розробці тестових шкал, методів перевірки гіпотез.

Нормальний розподіл (розподіл Гауса) характеризується тим, що крайні значення ознаки в ньому зустрічаються досить рідко, а значення, близькі до середнього, зустрічаються досить часто.

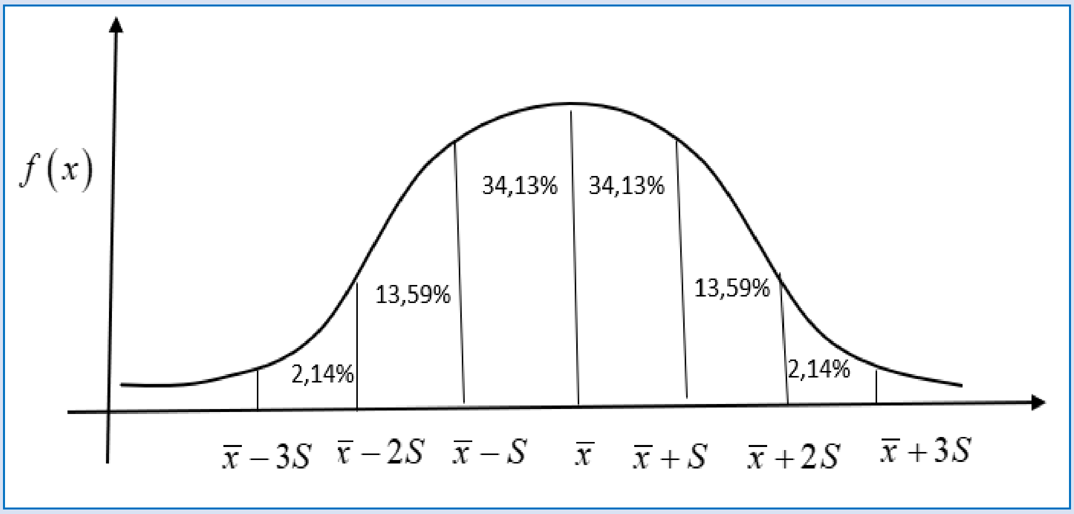
Випадкова величина X вважається розподіленою за нормальним розподілом (або розподілом Гауса), якщо щільність розподілу її ймовірностей описується функцією:

$$f = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\bar{x})^2}{2\sigma^2}}.$$

За цією формулою для різних значень середнього арифметичного $\bar{x}$ та квадратичного відхилення σ отримується сім'я нормальних кривих.

Теоретичний нормальний розподіл має вигляд симетричної кривої дзвоноподібного типу, яка володіє такими закономірностями:

1. Права і ліва гілки теоретичного нормального розподілу абсолютно симетричні і немов віддзеркалюють одна одну.
2. У нормальному розподілі основні показники центральної тенденції (мода, медіана і середнє арифметичне) співпадають і відповідають найвищій точці (вершині) розподілу.
3. Крайні значення ознаки в ньому досить рідкісні, а значення, близькі до середньої величини – досить часті.
4. Форму розподілу можна описати двома параметрами: середнім арифметичним значенням і стандартним відхиленням.
5. Права і ліва гілки розподілу прямують до безмежності, наближаючись до вісі абсцис, але не перетинаючи її.



При нормальному розподілі «більша частина» результатів розташовані в межах одного середнього квадратичного відхилення по обидві сторони від середнього арифметичного та в процентному співвідношенні складає 68,27% площі розподілу; 95,45% – не виходять за межі двох середніх квадратичних відхилень від середнього арифметичного; в межах трьох середніх квадратичних відхилень – 99,73%.

Параметри розподілу – це його числові характеристики, що вказують, де «в середньому» розташовуються значення ознаки, наскільки ці значення є мінливими і чи спостерігається переважна поява певних значень ознаки. В реальних психологічних дослідженнях оперують не параметрами, а їх наближеними значеннями – оцінками параметрів. Чим більша вибірка, тим ближче оцінка параметру до його дійсного значення.

У ситуаціях, коли які-небудь причини сприяють більш частій появі значень, які вищі або, навпаки, нижчі середнього значення, утворюються 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 (A) характеризує степінь асиметричності розподілу. Коефіцієнт асиметрії може приймати значення від $-\infty$ до $+\infty$ та розраховується за формулою:

$$A = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{ns^3},$$

а у випадку невеликої вибірки $n \le 30$:

$$A = \frac{n\sum_{i=1}^{n}(x_i - \bar{x})^3}{(n-1)(n-2)s^3}.$$

Для симетричних розподілів $A = 0$. При лівосторонній асиметрії ($A > 0$) в розподілі частіше зустрічаються більш низькі значення ознаки, а при правосторонній ($A < 0$) – високі.

Ексцес (E) характеризує відносну опуклість або згладженість розподілу вибірки в порівнянні з нормальним розподілом. Позитивний ексцес ($E > 0$) характеризує порівняно загострений розподіл, негативний ($E < 0$) – порівняно згладжений. В розподілі з нормальною випуклістю $E = 0$.

Показник ексцесу визначається за формулою:

$$E = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{ns^4} - 3,$$

а у випадку невеликої вибірки $n \le 30$:

$$E = \frac{n(n+1)\sum_{i=1}^{n}(x_i - \bar{x})^4}{(n-1)(n-2)(n-3)s^4} - \frac{3(n-1)^2}{(n-2)(n-3)}.$$

Нормальність розподілу ознаки можна перевірити шляхом розрахунку показників асиметрії та ексцесу і порівнюючи їх з критичними значеннями (критерії М.А. Плохінського або Є.І. Пустильника).

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

Тобто, дані відповідають закону нормального розподілу, якщо:

$$t_A = \frac{|A|}{m_A} \le 3, \; t_E = \frac{|E|}{m_E} \le 3,$$

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

Критерій *Є.І. Пустильника.* 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:

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

де

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

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

<u>**Приклад 2.5.**</u> Чи можна стверджувати, що отримані в досліджені емпіричні дані відповідають закону нормального розподілу?

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x_i	11	13	12	9	10	11	8	10	15	14	8	7	10	10	5	8

Розв'язування. Складемо таблицю проміжних розрахунків:

№	x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})^3$	$(x_i - \overline{x})^4$
1	11	0,937	0,878	0,823	0,771
2	13	2,937	8,626	25,334	74,407
3	12	1,937	3,752	7,268	14,077
4	9	-1,063	1,13	-1,201	1,277
5	10	-0,063	0,004	0,000	0,000
6	11	0,937	0,878	0,823	0,771
7	8	-2,063	4,256	-8,780	18,113
8	10	-0,063	0,004	0,000	0,000

9	15	4,937	24,374	120,334	594,090
10	14	3,937	15,5	61,023	240,249
11	8	-2,063	4,256	-8,780	18,113
12	7	-3,063	9,382	-28,737	88,021
13	10	-0,063	0,004	0,000	0,000
14	10	-0,063	0,004	0,000	0,000
15	5	-5,063	25,634	-129,785	657,100
16	8	-2,063	4,256	-8,780	18,113
Сума	161		102,938	29,541	1725,104
Сер. знач.	10,063				

Скористаємося критерієм М.А. Плохінського. Для цього спочатку обрахуємо значення середнього арифметичного ряду, стандартного відхилення, показники асиметрії та ексцесу і їх помилки репрезентативності:

$$\bar{x} = 10{,}063\,,\ s = \sigma = \sqrt{\frac{\left(x_i - \bar{x}\right)^2}{n-1}} = \sqrt{\frac{102{,}938}{16-1}} = 2{,}62\,,$$

$$A = \frac{n\sum\limits_{i=1}^{n}\left(x_i - \bar{x}\right)^3}{(n-1)(n-2)\sigma^3} = \frac{16\cdot 29{,}541}{15\cdot 14\cdot (2{,}62)^3} = 0{,}125\,,$$

$$m_A = \sqrt{\frac{6}{n+3}} = \sqrt{\frac{6}{19}} = 0{,}562\,,\ t_A = \frac{|A|}{m_A} = \frac{0{,}125}{0{,}562} = 0{,}222 \le 3\,.$$

$$E = \frac{n(n+1)\sum\limits_{i=1}^{n}\left(x_i - \bar{x}\right)^4}{(n-1)(n-2)(n-3)\sigma^4} - \frac{3(n-1)^2}{(n-2)(n-3)} =$$

$$= \frac{16\cdot 17\cdot 1725{,}104}{15\cdot 14\cdot 13\cdot (2{,}62)^4} - \frac{3\cdot 15^2}{14\cdot 13} = -0{,}061\,.$$

Оскільки $m_E = 2\sqrt{\dfrac{6}{n+3}} = 1{,}124$, $t_E = \dfrac{|E|}{m_E} = \dfrac{0{,}061}{1{,}124} = 0{,}054 \le 3$, то можемо зробити висновок, що розподіл відповідає закону нормального розподілу.

Скористаємося критерієм Є.І. Пустильника. Обрахуємо критичні значення:

$$A_{крит} = 3\sqrt{\frac{6(n-1)}{(n+1)(n+3)}} = 3\cdot\sqrt{\frac{6\cdot 15}{17\cdot 19}} = 1{,}584,$$

$$E_{крит} = 5\sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}} = 5\sqrt{\frac{24\cdot 16\cdot 14\cdot 13}{17^2\cdot 19\cdot 21}} = 3{,}89.$$

Оскільки $|A| = 0{,}125 \le A_{крит} = 1{,}584$, $|E| = 0{,}061 \le E_{крит} = 3{,}89$, то можемо зробити висновок, що розподіл відповідає закону нормального розподілу.

3. МІРИ ЗВ'ЯЗКУ В ПАРАМЕТРИЧНІЙ СТАТИСТИЦІ

3.1. ПОНЯТТЯ ПРО КОРЕЛЯЦІЙНИЙ ЗВ'ЯЗОК

Аналіз зв'язків між ознаками – одне із головних завдань, що зустрічається практично у будь-якому емпіричному дослідженні.

При вивченні кореляцій намагаються встановити, чи існує якийсь зв'язок між двома показниками в одній вибірці (наприклад, між ростом та вагою дітей; між рівнем IQ та шкільною успішністю) або між двома різними вибірками (наприклад, при порівнянні пар близнюків), і якщо цей зв'язок існує, то чи супроводжується збільшення одного показника зростанням (позитивна кореляція) або зменшенням (негативна кореляція) іншого. Іншими словами, кореляційний аналіз допомагає встановити, чи можна передбачити можливі значення одного показника, знаючи величину іншого.

Кореляційний аналіз для двох випадкових величин складається з:

- графічного представлення залежності;
- обчислення коефіцієнта кореляції;
- перевірки статистичної значущості зв'язку.

Дослідження характеру взаємозв'язку починається з побудови графічного зображення результатів вимірювання в прямокутній системі координат, де кожна пара результатів буде відображатись точкою. Така графічна залежність називається діаграмою розсіювання або кореляційним полем.

Кореляційне поле відображає статистичний взаємозв'язок між результатами вимірювань. Візуальний аналіз кореляційного поля дозволяє якісно оцінити форму, спрямованість і тісноту взаємозв'язку.

Ступінь, сила чи тіснота кореляційного зв'язку визначається за величиною коефіцієнта кореляції.

Сила зв'язку не залежить від її спрямованості та визначається за абсолютним значенням коефіцієнта кореляції. Коефіцієнт кореляції – це величина, яка може змінюватись у межах від -1 до $+1$. У разі повної позитивної кореляції цей коефіцієнт дорівнює плюс одиниці, а за повної негативної – мінус одиниці. Якщо коефіцієнт кореляції дорівнює 0, обидві змінні повністю незалежні одна від одної.

Силу зв'язку можна оцінити за допомогою шкали Чедока:

Сила зв'язку	Значення коефіцієнта кореляції
дуже слабкий	$0 < \lvert r \rvert \le 0{,}1$
слабкий	$0{,}1 < \lvert r \rvert \le 0{,}3$
помірний	$0{,}3 < \lvert r \rvert \le 0{,}5$
середній	$0{,}5 < \lvert r \rvert \le 0{,}7$
сильний	$0{,}7 < \lvert r \rvert \le 0{,}9$
дуже сильний	$0{,}9 < \lvert r \rvert \le 1$

Кореляційний зв'язок розрізняють за напрямком зв'язку (прямий, обернений), за формою аналітичного вираження (прямолінійний, криволінійний), за числом факторних показників, що враховуються для оцінки ступеня їх впливу на результативний показник (парний, множинний).

3.2. КОЕФІЦІЄНТ КОРЕЛЯЦІЇ ПІРСОНА

Для пошуку міри лінійного зв'язку між двома змінними розглянемо відхилення кожної змінної від її середнього арифметичного: $(x_i - \bar{x})$ і $(y_i - \bar{y})$. Було встановлено, якщо змінні X та Y в основному пов'язані прямо, то більшість добутків $(x_i - \bar{x})(y_i - \bar{y})$ будуть додатні, а значить і сума $\sum (x_i - \bar{x})(y_i - \bar{y})$ також буде додатною.

Якщо змінні в основному пов'язані обернено, то більшість добутків $(x_i - \bar{x})(y_i - \bar{y})$ будуть від'ємними, а значить і сума $\sum (x_i - \bar{x})(y_i - \bar{y})$ теж буде від'ємною.

Таким чином, ми отримали досить ефективну міру зв'язку. Однак, вона має недолік, пов'язаний із впливом чисельності вибірки. Тому отриману суму ділять на $n-1$.

У результаті отримують величину, яка називається **коваріацією**, і позначається S_{xy}:

$$S_{xy}=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n-1}.$$

Коваріація є досить задовільною мірою зв'язку для багатьох задач, однак, залежить від впливу стандартних відхилень обох груп.

У результаті ділення коваріації на стандартні відхилення цих груп отримується **коефіцієнт кореляції Пірсона**, який позначається r_{xy} та обраховується за формулою:

$$r_{xy}=\frac{S_{xy}}{S_xS_y}=\frac{\overline{xy}-\bar{x}\,\bar{y}}{\sqrt{\overline{x^2}-\bar{x}^2}\sqrt{\overline{y^2}-\bar{y}^2}}.$$

Слід зауважити, що коефіцієнт кореляції r_{xy} Пірсона є параметричним методом і його коректне застосування можливе тільки у випадку, якщо розподіл даних двох змінних не відрізняється від нормального.

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

Приклад 3.1. У 12 досліджуваних виміряли рівень двох ознак. Необхідно вияснити чи є статистично значущий зв'язок між цими ознаками. Результати дослідження подано в таблиці:

№	1	2	3	4	5	6	7	8	9	10	11	12
x_i	10	12	10	9	8	7	10	9	11	10	12	8
y_i	48	49	45	38	34	25	42	41	43	46	48	23

Розв'язування. 1. З метою перевірки обох рядів змінних на відповідність закону нормального розподілу використаємо критерій М.А. Плохинського. Побудуємо таблицю проміжних розрахунків:

№	x_i	$(x_i-\bar{x})^2$	$(x_i-\bar{x})^3$	$(x_i-\bar{x})^4$
1	10	0,1089	0,0359	0,01186
2	12	5,4289	12,6493	29,4730
3	10	0,1089	0,0359	0,01186
4	9	0,4489	-0,3008	0,2015
5	8	2,7889	-4,6575	7,7780
6	7	7,1289	-19,0342	50,8212
7	10	0,1089	0,0359	0,01186
8	9	0,4489	-0,3008	0,2015
9	11	1,7689	2,3526	3,1290
10	10	0,1089	0,0359	0,01186
11	12	5,4289	12,6493	29,4730
12	8	2,7889	-4,6575	7,7780
	9,67	26,67	-1,156	128,9026

Обрахуємо значення стандартного відхилення, показники асиметрії та ексцесу і їх помилки репрезентативності для першого ряду:

$$S_x=\sqrt{\sigma^2}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}=\sqrt{\frac{26,67}{11}}\approx 1,56,$$

$$A_{sx}=\frac{n}{(n-1)(n-2)}\frac{\sum_{i=1}^{n}(x_i-\bar{x})^3}{S_x^{\ 3}}=\frac{12}{11\cdot 10}\frac{-1,156}{(1,56)^3}\approx -0,033,$$

$$m_A=\sqrt{\frac{6}{n+3}}=\sqrt{0,4}=0,63,$$

$$t_{A_x}=\frac{|A_{sx}|}{m_{A_x}}=\frac{0,033}{0,63}=0,052\le 3.$$

$$E_x = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{S_x^{\ 4}} - \frac{3(n-1)^2}{(n-2)(n-3)} =$$

$$= \frac{12 \cdot 13}{11 \times 10 \times 9} \frac{128,9026}{(1,56)^4} - \frac{3 \cdot 11^2}{10 \times 9} \approx 3,43 - 4,03 = -0,6 .$$

Оскільки $t_{A_x} = \frac{|A_{sx}|}{m_{A_x}} = \frac{0,033}{0,63} = 0,052 \le 3$, $t_{E_x} = \frac{|E_x|}{m_{E_x}} = \frac{0,6}{1,26} = 0,48 \le 3$, то можемо зробити висновок, що розподіл даних першого ряду відповідає закону нормального розподілу.

Аналогічно для другого ряду даних побудуємо таблицю проміжних розрахунків та обрахуємо значення стандартного відхилення, показники асиметрії та ексцесу і їх помилки репрезентативності:

№	y_i	$(y_i - \bar{y})^2$	$(y_i - \bar{y})^3$	$(y_i - \bar{y})^4$
1	48	61,3089	480,04869	3758,7812
2	49	77,9689	688,4654	6079,1493
3	45	23,3289	112,6786	544,2376
4	38	4,7089	-10,2183	22,1737
5	34	38,0689	-234,8851	1449,2411
6	25	230,1289	-3491,0554	52959,3106
7	42	3,3489	6,1285	11,2151
8	41	0,6889	0,5718	0,4746
9	43	8,0089	22,6652	64,1425
10	46	33,9889	198,1553	1155,2453
11	48	61,3089	480,0487	3758,7812
12	23	294,8089	-5061,8688	86912,2875
	482	837,6668	-6809,27	156715
	$\bar{y} = 40,17$			

$$S_y \approx 8{,}73\,,$$

$$A_{sy} = \frac{n}{(n-1)(n-2)} \frac{\sum_{i=1}^{n}(y_i - \bar{y})^3}{S_y^{\,3}} \approx -1{,}12\,,\; t_{A_y} = 1{,}78 \le 3\,.$$

$$E_y = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \frac{\sum_{i=1}^{n}(y_i - \bar{y})^4}{S_y^{\,4}} - \frac{3(n-1)^2}{(n-2)(n-3)} = 0{,}22\,,\; t_{E_Y} = 0{,}17 \le 3\,.$$

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 помилки репрезентативності в 3 рази.

За результатами розрахунків показники асиметрії та ексцесу для другого ряду даних не перевищують утричі свої помилки репрезентативності, тому можемо зробити висновок, що дані мають нормальний розподіл. Відповідно, ми можемо використати параметричний коефіцієнт кореляції Пірсона r_{xy}.

2. Складемо таблицю первинних даних та проміжних розрахунків.

№	x_i	y_i	$x_i^{\,2}$	$y_i^{\,2}$	$x_i \cdot y_i$
1	10	48	100	2304	480
2	12	49	144	2401	588
3	10	45	100	2025	450
4	9	38	81	1444	342
5	8	34	64	1156	272
6	7	25	49	625	175
7	10	42	100	1764	420
8	9	41	81	1681	369
9	11	43	121	1849	473
10	10	46	100	2116	460
11	12	48	144	2304	576

12	8	23	64	529	184
	$\overline{x}=9{,}67$	$\overline{y}=40{,}17$	$\overline{x^2}=95{,}67$	$\overline{y^2}=1683{,}17$	$\overline{xy}$=399,08

Використовуючи одержанні проміжні результати можна порахувати значення коефіцієнта кореляції r_{xy} Пірсона:

$$r_{xy}=\frac{\overline{xy}-\overline{x}\,\overline{y}}{\sqrt{\overline{x^2}-\overline{x}^2}\sqrt{\overline{y^2}-\overline{y}^2}}=$$

$$=\frac{399{,}08-9{,}67\cdot 40{,}17}{\sqrt{95{,}67-(9{,}67)^2}\sqrt{1683{,}17-(40{,}17)^2}}\approx\frac{10{,}64}{1{,}47\cdot 8.34}\approx 0{,}87.$$

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

3. Встановлюємо прийнятний рівень ймовірності $\alpha=0{,}01$. За таблицею критичних значень коефіцієнта кореляції r_{xy} Пірсона для $n=12$ та заданого $\alpha=0{,}01$ знаходимо $r_{крит}=0{,}708$. Оскільки $|r_{xy}|>r_{крит}$, тому існує статистично значуща сильна позитивна кореляція між показниками x та y.

4. МІРИ ЗВ'ЯЗКУ В НЕПАРАМЕТРИЧНІЙ СТАТИСТИЦІ. КОЕФІЦІЄНТ КОРЕЛЯЦІЇ. ЗАЛЕЖНІСТЬ ТИПУ КОЕФІЦІЄНТА КОРЕЛЯЦІЇ ВІД СПОСОБУ ВИМІРЮВАННЯ

4.1. КОЕФІЦІЄНТ КОРЕЛЯЦІЇ φ

Коефіцієнт кореляції φ дозволяє обчислити кореляцію між якісними характеристиками об'єктів. Використовується для пошуку зв'язку між змінними, виміряними у дихотомічній шкалі найменувань та коли одна зі змінних виражена у цій шкалі, а друга – у цій же шкалі, але з припущенням нормальності.

Обчислюється

$$\varphi = \frac{p_{xy} - p_x p_y}{\sqrt{p_x q_x p_y q_y}},$$

де p_{xy} – відсоток людей, що мають одиницю по X та по Y одночасно, p_x – відсоток людей, що мають одиницю по X, p_y – відсоток людей, що мають одиницю по Y, q_x – відсоток людей, що мають нуль по X, q_y – відсоток людей, що мають нуль по Y.

Якщо нас не цікавлять відсотки одиниць та нулів, ми користуємося **таблицею спряженості ознак**, яка має в загальному такий вигляд:

		Змінні X		
		1	0	Всього
Змінні Y	1	a	b	$a+b$
	0	c	d	$c+d$
	Всього	$a+c$	$b+d$	

Тоді обчислення проводяться за формулою:

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

Для коефіцієнта φ не існує своєї таблиці з критичними значеннями. Статистична значущість взаємозв'язку оцінюється за допомогою критерію χ^2-Пірсона:

$$\chi^2 = \varphi^2 \cdot n,$$

де φ – емпіричне значення коефіцієнта; 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, то ці значення завжди сталі: $\chi^2_{0,05} = 3,841$, $\chi^2_{0,01} = 6,635$, $\chi^2_{0,001} = 10,829$. Якщо $\chi^2_{емп} > \chi^2_{крит}$, то залежність між ознаками статистично значуща. Якщо $\chi^2_{емп} < \chi^2_{крит}$, то статистично значущої залежності між ознаками немає.

Приклад 4.1. Проаналізувати успішність студентів залежно від статі, виділивши дві групи: студенти, що склали іспит, та студенти, що не склали іспит. Для отриманих даних сформували таблицю зв'язаності ознак:

	Змінні X		
Стать	Склав	Не склав	Всього
Жіноча	$a = 23$	$b = 6$	29
Чоловіча	$c = 21$	$d = 8$	29
Всього	44	14	58

Розв'язування. Обрахуємо значення коефіцієнта φ:

$$\varphi = \frac{23 \cdot 8 - 6 \cdot 21}{\sqrt{44 \cdot 29 \cdot 14 \cdot 29}} \approx 0,081.$$

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

$$\chi^2_{емп} = (0{,}081)^2 \cdot 58 = 0{,}38 .$$

Для $\alpha = 0{,}05$ маємо, що $\chi^2_{емп} = 0{,}38 < \chi^2_{крит} = 3{,}841$, а отже статистично значущої залежності між ознаками немає.

4.2. ТОЧКОВИЙ БІСЕРІАЛЬНИЙ КОЕФІЦІЄНТ КОРЕЛЯЦІЇ

Точковий бісеріальний коефіцієнт кореляції використовується у випадку, коли одна змінна виміряна у дихотомічній шкалі найменувань, а друга – в шкалі інтервалів чи відношень. Обчислюється за формулою:

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

де $\overline{y_1}$ – середнє арифметичне змінної *Y*, пораховане для об'єктів, що мають по *X* одиницю, $\overline{y_0}$ – середнє арифметичне змінної *Y*, пораховане для об'єктів, що мають по *X* нуль, S_y – стандартне відхилення *Y*, n_1 – кількість об'єктів по *X*, які мають одиницю, n_0 – кількість об'єктів по *X*, які мають нуль, n – загальна кількість об'єктів.

Точковий бісеріальний коефіцієнт кореляції немає своєї таблиці для знаходження критичних значень. Пошук критичних значень здійснюється за таблицею критичних значень критерію *t*-Стьюдента. Для обчислення емпіричного значення використовується формула:

$$t_{емп} = |r_{pb}| \sqrt{\frac{n-2}{1 - r_{pb}^2}} .$$

Якщо $t_{емп} \geq t_{крит}$ для заданого α-рівня значущості, то формується висновок, що коефіцієнт кореляції статистично значуще відрізняється від нуля.

<u>Приклад 4.2.</u> Необхідно перевірити, чи є взаємозв’язок між статтю людини і її вольовими якостями? Результати дослідження подано в таблиці:

№	1	2	3	5	6	7	8	9	10	11	12	13	14	15	16
X	0	1	1	1	0	0	1	1	0	0	1	1	1	1	0
Y	64	77	95	105	83	73	75	101	84	82	103	77	76	92	94

Розв’язування. Складемо таблицю проміжних розрахунків:

№	**X**	**Y**	**0**	**1**	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	0	64	64		-20,69	428,0761
2	1	77		77	-7,69	59,1361
3	1	74		74	-10,69	114,2761
4	1	95		95	10,31	106,2961
5	0	105	105		20,31	412,4961
6	0	83	83		-1,69	2,8561
7	1	73		73	-11,69	136,6561
8	0	75	75		-9,69	93,8961
9	0	101	101		16,31	266,0161
10	1	84		84	-0,69	0,4761
11	1	82		82	-2,69	7,2361
12	0	103	103		18,31	335,2561
13	0	77	77		-7,69	59,1361
14	1	76		76	-8,69	75,5161
15	1	92		92	7,31	53,4361
16	0	94	94		9,31	86,6761
		$\bar{y} = 84{,}69$	$\bar{y}_0 = 87{,}75$	$\bar{y}_1 = 81{,}63$		$\sum_{i=1}^{16}(y_i - \bar{y})^2 = 2237{,}438$

З таблиці випливає, що $n_0 = 8$, $n_1 = 8$.

Обчислимо $\overline{y_1}$, $\overline{y_0}$: $\overline{y_1} = \frac{653}{8} \approx 81{,}63$, $\overline{y_0} = \frac{702}{8} = 87{,}75$.

Для обрахунку S_y пригадаємо, що $S_y = \sqrt{\sigma^2}$, $\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2$:

$$\sigma^2 = \frac{1}{15}\cdot 2237{,}438 \approx 149{,}163,\ S_y = \sqrt{149{,}163} \approx 12{,}21.$$

Знайдемо значення точкового бісеріального коефіцієнту кореляції:

$$r_{pb} = \frac{81{,}63 - 87{,}75}{12{,}21}\sqrt{\frac{8\cdot 8}{16\cdot 15}} = -0{,}26.$$

Отримане значення свідчить про слабку обернену кореляцію між змінними. Перевіримо її статистичну значущість при $\alpha = 0{,}05$. Обрахуємо емпіричне значення:

$$t_{емп} = 0{,}26\cdot\sqrt{\frac{14}{1-\left(0{,}26\right)^2}} \approx 1{,}04.$$

Для ступенів свободи $\nu = n - 2 = 16 - 2 = 14$ при рівні значимості $\alpha = 0{,}05$ $t_{крит} = 2{,}14$, тому $t_{емп} < t_{крит}$, а, отже можна зробити висновок, що між статтю людини і її вольовими якостями залежність відсутня.

4.3. КОЕФІЦІЄНТ РАНГОВОЇ КОРЕЛЯЦІЇ СПІРМЕНА

Методом рангової кореляції Спірмена можна визначити силу і напрямок зв'язку між двома ознаками чи профілями (ієрархіями) ознак, що представлені у вимірювальних шкалах порядку, або одна ознака – у шкалі порядку, а друга – у будь-якій кількісній шкалі (інтервальній або відношень). Для обчислення коефіцієнта рангової кореляції Спірмена необхідно мати два ряди значень, які можна проранжувати.

Правила ранжування

В процесі ранжування потрібно дотримуватися двох правил.

Правило порядку ранжування. Потрібно вирішити, хто одержить перший ранг: об'єкт з самим великим степенем вираженості деякої якості чи навпаки, тобто вирішити чи об'єкти ранжувати в порядку спадання якості чи у порядку її зростання.

Правило зв'язних рангів. Оскільки однаковим значенням ознаки не можна надати різні ранги, то кожному значенню, що рівні між собою, присвоюється ранг, рівний середньому арифметичному з тих рангів, які мали б ці елементи, якби вони були різними.

Приклад 4.3. Проранжувати ряд:

№	1	2	3	4	5	6	7	8	9	10	11	12	13
X	76	74	74	74	70	56	55	54	54	37	36	33	28

Розв'язування.

№	1	2	3	4	5	6	7	8	9	10	11	12	13
X	76	74	74	74	70	56	55	54	54	37	36	33	28
Ранг	1	3	3	3	5	6	7	8,5	8,5	10	11	12	13

Для впорядкованого ряду при ранжуванні значенню ознаки, що зустрічається один раз приписуємо її порядковий номер, так: для 76 ранг дорівнює 1, для 55 – 7.

У випадку, коли значення ознаки рівні між собою для декількох об'єктів, то при присвоєні рангу знаходимо середнє арифметичне порядкових номерів цих об'єктів, так:

- значення 74 властиве для 2, 3, 4 об'єктів, щоб обчислити їх ранг знаходимо $\frac{2+3+4}{3}=3$, отже кожному об'єкту зі значенням ознаки 74 приписуємо ранг 3;
- значення 54 властиве для 8 та 9 об'єктів, щоб обчислити їх ранг знаходимо $\frac{8+9}{2}=8{,}5$, отже кожному об'єкту зі значенням ознаки 54 приписуємо ранг 8,5.

У випадку, коли обидві змінні виміряні в порядкових шкалах, або проранжовані, використовують ранговий коефіцієнт кореляції Спірмена, який обчислюється за формулою:

$$r_s = 1 - \frac{6\sum_{i=1}^{n}\left(r_{x_i} - r_{y_i}\right)^2}{n\left(n^2 - 1\right)},$$

де r_{x_i} – ранг змінної x_i, r_{y_i} – ранг змінної y_i.

У разі наявності зв'язаних рангів (однакових) однієї чи обох змінних загальна формула для коефіцієнта рангової кореляції r_s Спірмена обчислює дещо неточне значення. У такому разі необхідно здійснити поправку на однакові ранги. Скоригована формула має такий вигляд:

$$r_s = 1 - \frac{6\sum_{i=1}^{n}\left(r_{x_i} - r_{y_i}\right)^2 + T_x + T_y}{n\left(n^2 - 1\right)},$$

де T_x, T_y – поправки, пов'язані з однаковими рангами;

$$T_x = \frac{\sum_{i=1}^{L_x}\left(T_{x_i}^3 - T_{x_i}\right)}{12},\ T_y = \frac{\sum_{i=1}^{L_y}\left(T_{y_i}^3 - T_{y_i}\right)}{12},$$

де L_x, L_y – кількість зв'язок (груп однакових рангів); T_{x_i}, T_{y_i} – розмір i-ої зв'язки (кількість елементів в них).

Якщо емпіричне значення $r_s \geq r_{крит}$ для заданого α-рівня, то роблять висновок, що коефіцієнт кореляції статистично значуще відрізняється від нуля.

Приклад 4.4. В групі учасників психологічного експерименту було виміряно дві ознаки *X* та *Y*. Чи можна стверджувати, що ознаки *X* та *Y* статистично значуще взаємопов'язані між собою? Результати дослідження подано в таблиці:

№	1	2	3	4	5	6	7	8	9	10	11	12	13
X	43	41	42	35	38	36	44	45	40	37	33	39	34
Y	32	31	35	26	25	27	29	34	28	33	22	37	21

Розв'язування. Проранжуємо дані обох рядів та складемо таблицю проміжних обрахунків:

№	X	r_{x_i}	Y	r_{y_i}	$d_i = r_{x_i} - r_{y_i}$	d_i^2
1	43	11	32	9	2	4
2	41	9	31	8	1	1
3	42	10	35	12	-2	4
4	35	3	26	4	-1	1
5	38	6	25	3	3	9
6	36	4	27	5	-1	1
7	44	12	29	7	5	25
8	45	13	34	11	2	4
9	40	8	28	6	2	4
10	37	5	33	10	-5	25
11	33	1	22	2	-1	1
12	39	7	37	13	-6	36
13	34	2	21	1	1	1

Обчислимо суму: $\sum_{i=1}^{n}\left(r_{x_i} - r_{y_i}\right)^2 = 116$.

Знайдемо ранговий коефіцієнт кореляції Спірмена:

$$r_s = 1 - \frac{6\sum_{i=1}^{n}\left(r_{x_i} - r_{y_i}\right)^2}{n\left(n^2 - 1\right)} = 1 - \frac{6 \cdot 116}{13(13^2 - 1)} = 0,68.$$

Отримане значення свідчить про пряму кореляцію середньої сили між змінними. Перевіримо її статистичну значущість при $\alpha = 0,05$: $r_s = 0,68 > r_{крит} = r_s = 0,553$, а, отже можна зробити висновок, що коефіцієнт кореляції статистично значуще відрізняється від нуля для рівня значущості $\alpha = 0,05$.

4.4. КОЕФІЦІЄНТ τ-КЕНДАЛЛА

Коефіцієнт τ-Кендалла є альтернативним коефіцієнтом для обчислення зв'язку в тих же випадках, що і коефіцієнт рангової кореляції Спірмена, але ґрунтується на дещо інших припущеннях.

Для підрахунку коефіцієнта τ-Кендалла використовують такий алгоритм:

1. Впорядковують ранги по одній із змінних, наприклад по X.
2. Підраховується кількість «співпадінь» (P): для кожного об'єкта підраховується, скільки разів його ранг по Y виявляється менше від рангів об'єктів, які знаходяться нижче нього.
3. Підраховується кількість «інверсій» (Q): для кожного об'єкта підраховується, скільки разів його ранг по Y виявляється більше, ніж ранги об'єктів, які знаходяться нижче.
4. Обчислюється коефіцієнт τ за формулою:

$$\tau = \frac{P - Q}{\frac{n(n-1)}{2}}.$$

У випадку наявності зв'язних рангів коефіцієнт τ обчислюється за формулою:

$$\tau = \frac{P - Q}{\sqrt{\frac{n(n-1)}{2} - K_x}\sqrt{\frac{n(n-1)}{2} - K_y}},$$

де $K_x = \frac{1}{2}\sum_{i=1}^{m_x} f_{x_i}\left(f_{x_i} - 1\right)$, $K_y = \frac{1}{2}\sum_{i=1}^{m_y} f_{y_i}\left(f_{y_i} - 1\right)$, m_x, m_y – кількість груп зв'язних рангів по X та по Y відповідно; f_{x_i}, f_{y_i} – кількість елементів i-ої групи зв'язних рангів по X та по Y відповідно.

Коефіцієнт кореляції τ-Кендалла немає своєї таблиці для знаходження критичних значень. Пошук критичних значень здійснюється за таблицею критичних значень критерію t-Стьюдента. Для обчислення емпіричного значення використовується формула:

$$t_{емп} = |\tau| \sqrt{\frac{n-2}{1-\tau^2}} .$$

Якщо $t_{емп} \geq t_{крит}$ для заданого α-рівня, то формується висновок, що коефіцієнт кореляції статистично значуще відрізняється від нуля.

<u>Приклад 4.5.</u> В групі учасників психологічного експерименту було виміряно дві ознаки X та Y. Після ранжування, впорядкували ранги по X. Чи можна стверджувати, що ознаки X та Y статистично значуще взаємопов'язані між собою?

№	r_x	r_y	P	Q
1	1	13	0	12
2	2	11	1	10
3	3	10	1	9
4	4	8	2	7
5	5	4	5	3
6	6	12	0	7
7	7	9	0	6
8	8	7	0	5
9	9	3	2	2
10	10	5	1	2
11	11	6	0	2
12	12	2	0	1
13	13	1	0	0
			12	**66**

Обчислюємо суму «співпадінь» та «інверсій»: $P = 12$; $Q = 66$.

Підставимо одержані значення у формулу для підрахунку коефіцієнта τ:

$$\tau = \frac{P-Q}{\frac{n(n-1)}{2}} = \frac{12-66}{\frac{13 \cdot 12}{2}} \approx -0{,}69 .$$

Отримане значення свідчить про обернену кореляцію середньої сили між змінними. Перевіримо її статистичну значущість при $\alpha = 0{,}05$. Для цього знайдемо емпіричне значення:

$$t_{емп} = 0{,}69 \cdot \sqrt{\frac{11}{1-(0{,}69)^2}} \approx 3{,}16\,.$$

Для ступенів свободи $\nu = n - 2 = 13 - 2 = 11$ при рівні значимості $\alpha = 0{,}05$ $t_{крит} = 2{,}20$, тому $t_{емп} > t_{крит}$, а, отже можна зробити висновок, що коефіцієнт кореляції статистично значуще відрізняється від нуля.

4.5. БІСЕРІАЛЬНИЙ КОЕФІЦІЄНТ РАНГОВОЇ КОРЕЛЯЦІЇ

У випадку, коли одна змінна виміряна в дихотомічній шкалі найменувань, а друга – в порядковій шкалі, використовують бісеріальний коефіцієнт рангової кореляції Кертена та Гласса.

Для підрахунку бісеріального коефіцієнту рангової кореляції використовують такий алгоритм:

1. Впорядкувати ранги та розписати ранги по принципу належності до 1 чи до 0.
2. Підрахунок співпадінь (P). У цьому випадку дані організовані таким чином, що кількість співпадінь рахується для рангів в стовпчику 1 і рівна кількості об'єктів, які знаходяться в стовпчику 0 нижче взятого рангу в стовпчику 1.
3. Підрахунок інверсій (Q). Кількість інверсій рахується для рангів у стовпчику 1 і рівна кількості об'єктів, які знаходяться в стовпчику 1 нижче взятого рангу стовпчику 0.
4. Підраховується сума співпадінь та інверсій і підставляються у формулу:

$$r_{rb} = \frac{P-Q}{n_0 n_1}\,,$$

де n_0 – кількість 0, n_1 – кількість 1.

Ця формула була виведена Кертеном, а дослідник Гласс довів, що вона алгебраїчно еквівалентна такій формулі:

$$r_{rb} = \frac{\overline{y}_1 - \overline{y}_0}{\frac{n_0 n_1}{2}},$$

де $\overline{y_1}$ – середній ранг об'єктів, що мають по X одиницю, $\overline{y_0}$ – середній ранг об'єктів, що мають по X нуль, n_1 – кількість об'єктів по X, які мають одиницю, n_0 – кількість об'єктів по X, які мають нуль.

Бісеріальний коефіцієнт рангової кореляції немає своєї таблиці для знаходження критичних значень. Пошук критичних значень здійснюється за таблицею критичних значень критерію t-Стьюдента. Для обчислення емпіричного значення використовується формула:

$$t_{емп} = |r_{rb}|\sqrt{\frac{n-2}{1-r_{rb}^{\ 2}}}.$$

Якщо $t_{емп} \geq t_{крит}$ для заданого α-рівня значущості, то формується висновок, що коефіцієнт кореляції статистично значуще відрізняється від нуля.

Приклад 4.6. Проведено дослідження зв'язку між статтю та рівнем тривожності. В результаті дослідження отримали такі дані:

стать	ж	ж	ч	ж	ч	ч	ж	ч	ж	ж	ж	ч	ч	ж	ч
тривожність	15	8	5	9	7	3	6	1	2	14	12	10	11	13	4

Розв'язування. Складемо таблицю для проміжних обрахунків:

№	*Стать*	*Тривож.*	*Тривож.*	*Стать*	1	0	*P*	*Q*
1	0	15	1	1	1		8	
2	0	8	2	0		0		6
3	1	5	3	1	1		7	
4	0	9	4	1	1		7	
5	1	7	5	1	1		7	

6	1	3	6	0		0		3
7	0	6	7	1	1		6	
8	1	1	8	0		0		2
9	0	2	9	0		0		2
10	0	14	10	1	1		4	
11	0	12	11	1	1		4	
12	1	10	12	0		0		0
13	1	11	13	0		0		0
14	0	13	14	0		0		0
15	1	4	15	0		0		0
							43	13

Підраховавши суму співпадінь та інверсій, знайдемо значення бісеріального коефіцієнту рангової кореляції:

$$r_{rb} = \frac{P - Q}{n_0 n_1} = \frac{43 - 13}{8 \cdot 7} \approx 0{,}54 .$$

Отримане значення свідчить про пряму кореляцію середньої сили між змінними. Перевіримо її статистичну значущість при $\alpha = 0{,}05$. Для цього обрахуємо емпіричне значення:

$$t_{емп} = 0{,}54 \cdot \sqrt{\frac{13}{1 - (0{,}54)^2}} \approx 2{,}75 .$$

Для ступенів свободи $\nu = n - 2 = 15 - 2 = 13$ при рівні значимості $\alpha = 0{,}05$ $t_{крит} = 2{,}16$, тому $t_{емп} > t_{крит}$, а, отже можна зробити висновок, що коефіцієнт кореляції статистично значуще відрізняється від нуля при рівні значимості $\alpha = 0{,}05$.

5. ПАРАМЕТРИЧНІ МЕТОДИ ПОРІВНЯННЯ ДВОХ ВИБІРОК

Для порівняння вибіркових середніх величин, які належать до двох вибірок, і для вирішення питання про те, чи відрізняються середні значення статистично достовірно один від одного, використовують *t*-критерій Стьюдента або його непараметричні аналоги.

t-критерій Стьюдента був розроблений Вільямом Госсетом для оцінки якості пива в компанії Гіннес. Щоб запобігти розкриттю конфіденційної інформації, Госсет публікував свої роботи під псевдонімом Стьюдент. Його найважливіше відкриття отримало назву розподіл Стьюдента.

Для коректного застосування t-критерію Стьюдента для двох груп необхідно обов’язкове виконання двох умов.

Першою умовою для застосування критерію є вимога нормальності розподілу змінної, що досліджується, у кожній з груп, що порівнюються.

Другою умовою, яка має виконуватися, є рівність дисперсій у групах, що порівнюються.

t-критерій Стьюдента використовується у трьох випадках:

1) порівняння середнього показника однієї вибірки із певною заданою величною (*t*-критерій для однієї вибірки);
2) порівняння середніх показників двох незалежних вибірок (*t*-критерій для незалежних вибірок);
3) порівняння середніх показників двох залежних вибірок (*t*-критерій для залежних вибірок).

В загальному алгоритм застосування t-критерію Стьюдента полягає в наступному:

1. Перевірка нормальності розподілу даних у вибірках, що порівнюються.
2. Перевірка рівності (гомогенності) дисперсій незалежних вибірок або перевірка наявності прямого кореляційного зв’язку між залежними вибірками.

3. Власне обчислення t-критерію (схема обчислення критерію відрізняється в різних випадках: порівнюються залежні чи незалежні вибірки; однаковий чи різний обсяг вибірок).
4. Порівнюється отримане емпіричне значення t-критерію із критичним табличним значенням і робиться висновок про підтвердження чи спростування нульової гіпотези щодо відсутності відмінностей.

5.1. *F*-КРИТЕРІЙ ФІШЕРА

Для перевірки гіпотези про рівність дисперсій генеральних сукупностей, розподілених за нормальним законом, використовується F-критерій Фішера. F-критерій Фішера називають дисперсійним відношенням, так як він знаходиться, як відношення двох незміщених оцінок дисперсій.

Алгоритм F-критерію Фішера:

1. Висуваємо гіпотези:

H_0: Дисперсії двох незалежних нормально розподілених вибірок рівні: $\sigma_1^2 = \sigma_2^2$.

H_1: Дисперсії двох незалежних нормально розподілених вибірок суттєво відрізняються: 1. $\sigma_1^2 > \sigma_2^2$, 2. $\sigma_1^2 \neq \sigma_2^2$.

2. Обчислюють виправлені дисперсії: σ_1^2, σ_2^2, де дисперсія обраховується за формулою: $\sigma^2 = \dfrac{\sum_{i=1}^{n}\left(x-\overline{x}\right)^2}{n-1}$.
3. Знаходять емпіричне значення критерію $F_{екс}$ за формулою:

$$F_{екс} = \frac{\sigma_1^2}{\sigma_2^2},$$

причому в чисельнику ставиться більша з двох дисперсій.

Визначають критичне значення критерію $F_{табл} = F\left(\alpha, f_1, f_2\right)$ за таблицею критичних точок розподілу Фішера-Снедекора за числами ступенів свободи $f_1 = n_1 - 1$ і $f_2 = n_2 - 1$, де n_1 та n_2 – кількість елементів першої та другої вибірки відповідно (n_1 – кількість елементів з більшим значенням σ^2).

Порівнюємо $F_{екс}$ з $F_{табл}$: якщо $F_{екс} < F_{табл}$, то приймається нульова гіпотеза про рівність генеральних дисперсій; якщо $F_{екс} \geq F_{табл}$, то приймається альтернативна гіпотеза.

Слід зауважити, що у випадку, коли у гіпотезі H_1 передбачається, що $\sigma_1^2 \neq \sigma_2^2$, то нульова гіпотеза приймається при виконанні умов:

$$F\left(1-\frac{\alpha}{2}, f_1, f_2\right) < F_{екс} < F\left(\frac{\alpha}{2}, f_1, f_2\right).$$

5.2. СТАТИСТИЧНИЙ КРИТЕРІЙ *t*-СТЬЮДЕНТА ДЛЯ ОДНІЄЇ ВИБІРКИ

Одновибірковий критерій t-Стьюдента – критерій, що дозволяє перевірити гіпотезу про те, що середнє значення ознаки, що вивчається, відрізняється від деякого відомого значення.

Алгоритм обчислення t-критерію Стьюдента для однієї вибірки:

1. Висувають гіпотези:

   H_0: Відмінності між $\bar{x}$ та величиною a випадкові і незначні.

   H_1: Відмінності між $\bar{x}$ та величиною a достовірні і значимі.

2. Знаходять значення t-критерію Стьюдента за формулою:

$$t_{емп} = \frac{|\bar{x} - a|}{S_x}\sqrt{n},$$

де $\bar{x}$ – середнє арифметичне ряду, S_x – стандартне відхилення, n – обсяг вибірки.

3. Знаходять число ступенів свободи: $\nu = n - 1$.
4. Для знайденого емпіричного значення визначаємо рівень значущості та використовуючи таблицю критичних точок розподілу Стьюдента знаходимо $t_{табл}$.

5. Робимо висновок: якщо $t_{емп} \geq t_{табл}$ – приймають гіпотезу H_1; якщо $t_{емп} < t_{табл}$ – приймають гіпотезу H_0.

<u>Приклад 5.1.</u> Визначено *IQ* випадково відібраних 30 студентів, що навчалися за спеціальною програмою. З'ясувати, чи відрізняється інтелект цих студентів від нормативного показника *а*=100?

Результати дослідження подані у таблиці:

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x	100	111	112	105	105	104	94	89	113	125	96	100	98	124	121
№	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
x	116	95	92	118	96	94	117	130	90	114	119	120	100	96	102

Розв'язування.

1. З метою перевірки ряду змінних на відповідність закону нормального розподілу використаємо критерій М.А. Плохинського.

Складемо таблицю первинних даних та проміжних розрахунків:

№	x	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^4$
1	100	-6,53	42,6409	-278,45	1818,246
2	111	4,47	19,9809	89,315	399,236
3	112	5,47	29,9209	163,667	895,26
4	105	-1,53	2,3409	-3,582	5,480
5	105	-1,53	2,3409	-3,582	5,480
6	104	-2,53	6,4009	-16,194	40,972
7	94	-12,53	157,0009	-1967,221	24649,283
8	89	-17,53	307,3009	-5386,985	94433,843
9	113	6,47	41,8609	270,84	1752,335
10	125	18,47	341,1409	6300,872	116377,114
11	96	-10,53	110,8809	-1167,576	12294,574
12	100	-6,53	42,6409	-278,445	1818,246

Продовження таблиці

13	98	-8,53	72,7609	-620,65	5294,149
14	124	17,47	305,2009	5331,86	93147,589
15	121	14,47	209,3809	3029,742	43840,361
16	116	9,47	89,6809	849,278	8042,664
17	95	-11,53	132,9409	-1532,809	17673,283
18	92	-14,53	211,1209	-3067,587	44572,034
19	118	11,47	131,5609	1509,004	17308,270
20	96	-10,53	110,8809	-1167,576	12294,574
21	94	-12,53	157,0009	-1967,221	24649,283
22	117	10,47	109,6209	1147,731	12016,742
23	130	23,47	550,8409	12928,236	303425,697
24	90	-16,53	273,2409	-4516,672	74660,589
25	114	7,47	55,8009	416,833	3113,74
26	119	12,47	155,5009	1939,096	24180,53
27	120	13,47	181,4409	2444,009	32920,8
28	100	-6,53	42,6409	-278,445	1818,246
29	96	-10,53	110,8809	-1167,576	12294,574
30	102	-4,53	20,5209	-92,96	421,107
сума	3196		4025,467	12906,95	986164,301
Сер. Знач.	106,53		134,18	430,2317	32872,143
			$S_x = 11,78$		

Обрахуємо значення стандартного відхилення, показники асиметрії та ексцесу і їх помилки репрезентативності для ряду:

$$S_x = \sqrt{\sigma^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \sqrt{\frac{4025,467}{29}} \approx 11,78,$$

$$A_{sx}=\frac{n}{(n-1)(n-2)}\frac{\sum_{i=1}^{n}(x_i-\bar{x})^3}{S_x{}^3}=\frac{30}{29\cdot 28}\frac{12906{,}95}{(11{,}78)^3}\approx 0{,}292,$$

$$m_A=\sqrt{\frac{6}{n+3}}=\sqrt{0{,}18}=0{,}42,\ t_{A_x}=\frac{|A_{sx}|}{m_{A_x}}=\frac{0{,}292}{0{,}42}=0{,}69\leq 3.$$

$$E_x=\frac{n(n+1)}{(n-1)(n-2)(n-3)}\frac{\sum_{i=1}^{n}(x_i-\bar{x})^4}{S_x{}^4}-\frac{3(n-1)^2}{(n-2)(n-3)}=$$

$$=\frac{30\cdot 31}{29\times 28\times 27}\frac{986164{,}301}{(11{,}78)^4}-\frac{3\cdot 29^2}{28\times 27}\approx 2{,}172-3{,}337=-1{,}165,$$

$$m_E=2\sqrt{\frac{6}{n+3}}=2\sqrt{0{,}18}=0{,}84,\ t_{E_x}=\frac{|E_x|}{m_{E_x}}=\frac{1{,}165}{0{,}84}=1{,}38\leq 3.$$

Оскільки за результатами розрахунків показники асиметрії та ексцесу не перевищують утричі свої помилки репрезентативності, то можемо зробити висновок, що дані розподілені за нормальним розподілом.

1. Висунемо гіпотези:

   H_0: Відмінності між $\bar{x}$ та 100 випадкові і незначні.

   H_1: Відмінності між $\bar{x}$ та 100 достовірні і значимі.

2. Знаходимо значення t-критерію Стьюдента:

$$t_{емп}=\frac{|\bar{x}-a|}{S_x}\sqrt{n}=\frac{106{,}53-100}{11{,}78}\sqrt{30}\approx 3{,}036.$$

3. Знаходимо число ступенів свободи: $\nu=n-1=30-1=29$.
4. Встановлюємо прийнятний рівень значущості $\alpha=0{,}05$.

За таблицею граничних значень t–критерію Стьюдента знаходимо $t_{табл}$ =2,05.

Оскільки $t_{емп}=3{,}036\geq t_{табл}=2{,}05$, то приймають гіпотезу H_1.

Отже, інтелект студентів, що займаються за спеціальною програмою, статистично достовірно перевищує нормативний показник інтелекту a = 100.

5.3. СТАТИСТИЧНИЙ КРИТЕРІЙ *t*-СТЬЮДЕНТА ДЛЯ НЕЗАЛЕЖНИХ ВИБІРОК

Розрізняють два типи вибірок значень випадкових величин, що беруть в експерименті участь: незалежні та залежні.

Незалежні (незв'язні) вибірки характеризуються тим, що ймовірність вибору будь-якого досліджуваного однієї вибірки не залежить від вибору будь-якого досліджуваного іншої вибірки. При такому підході формування груп, досліджувані не можуть перебувати в обох групах одночасно. Прикладом незалежних вибірок можуть бути рандомізовані (сформовані випадковим чином) контрольна та експериментальна групи.

Залежні (зв'язні) вибірки характеризуються тим, що кожному респонденту однієї вибірки ставиться у відповідність за певним критерієм респондент з іншої вибірки. Найбільш типовим прикладом залежних вибірок є повторне вимірювання значення ознаки (властивості) на одній і тій же вибірці після впливу (ситуація «до – після»). При цьому вибірки максимально залежні, оскільки включають в себе одних і тих же досліджуваних.

Можуть бути і більш «слабші» варіанти залежності. Наприклад, чоловіки – це одна вибірка, їх дружини – інша вибірка (при дослідженні, наприклад, їх переваг). Або діти 5-7 років – одна вибірка, а їх брати або сестри – інша вибірка.

Розглянемо алгоритм обчислення t-критерію Стьюдента для великих незалежних вибірок:

1. Висувають гіпотези:

H_0: Відмінності між першою та другою групами випадкові.

H_1: Відмінності між першою та другою групами є достовірними, значимими.

2. Для кожної групи обчислюють стандартні відхилення :

$$S_{x_1} = \sqrt{\sigma_1^{\ 2}}\,,\ S_{x_2} = \sqrt{\sigma_2^{\ 2}}\,,$$

де, як і раніше, $\sigma^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ – вибіркова дисперсія.

3. Для кожної групи обчислюють величину середніх помилок за формулою: $m = \frac{S}{\sqrt{n}}$ (окремо шукаємо m_1, m_2).
4. Знаходять значення t-критерію Стьюдента за формулою:

$$t_{емп} = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{(m_1)^2 + (m_2)^2}}.$$

5. Знаходять число ступенів свободи, яке залежить від кількості досліджуваних в обох вибірках: $\nu = n_1 + n_2 - 2$, де n_1, n_2 – кількість досліджуваних в кожній вибірці.
6. Для знайденого емпіричного значення визначають рівень значущості та використовуючи таблицю критичних точок розподілу Стьюдента знаходять $t_{табл}$.
7. Роблять висновок: якщо $t_{емп} \geq t_{табл}$ – приймають гіпотезу H_1, якщо $t_{емп} < t_{табл}$ – приймають гіпотезу H_0.

У випадку малих незалежних вибірок користуються наступною формулою:

$$t_{емп} = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{\frac{\left(\sigma_1^2(n_1 - 1) + \sigma_2^2(n_2 - 1)\right)(n_1 + n_2)}{(n_1 + n_2 - 2)\cdot n_1 \cdot n_2}}}.$$

Приклад 5.2. Для проведення формувального експерименту було обрано дві групи: експериментальна та контрольна. Перевірити чи відмінності між обраними групами суттєві.

Результати дослідження подані у таблиці:

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	15	8	10	6	20	9	14	7	15	8	16	13	15		
2	10	15	8	8	4	19	6	9	7	15	8	17	14	16	20

Розв'язування. Розподіл в обох вибірках дослідження відповідає закону нормального розподілу. Переконаємося у цьому. Скористаємося критерієм Є.І. Пустильника.

Складемо таблицю первинних даних та проміжних розрахунків:

№	1	2	$(x_{i1}-\overline{x_1})^2$	$(x_{i2}-\overline{x_2})^2$	$(x_{i1}-\overline{x_1})^3$	$(x_{i2}-\overline{x_2})^3$	$(x_{i1}-\overline{x_1})^4$	$(x_{i2}-\overline{x_2})^4$
1	15	10	17,22	2,99	71,47	-5,18	296,53	8,94
2	8	15	8,12	10,69	8,12	34,97	65,93	114,28
3	10	8	0,72	13,91	0,72	-51,9	0,52	193,49
4	6	8	23,52	13,91	23,52	-51,9	553,19	193,49
5	20	4	83,72	59,75	83,72	-461,89	7009,04	3570,06
6	9	19	3,42	52,85	3,42	384,24	11,70	2793,12
7	14	6	9,92	32,83	9,92	-188,13	98,41	1077,81
8	7	9	14,82	7,45	14,82	-20,35	219,63	55,50
9	15	7	17,22	22,37	17,22	-105,82	296,53	500,42
10	8	15	8,12	10,69	8,12	34,97	65,93	114,28
11	16	8	26,52	13,91	26,52	-51,9	703,31	193,49
12	9	17	3,42	27,77	3,42	146,36	11,70	771,17
13	4	14	46,92	5,15	46,92	11,70	2201,49	26,52
14		16		18,23		77,85		332,33
15		20		68,39		565,61		4677,19
сума	**141**	**176**	**263,66**	**360,89**	**317,91**	**318,63**	**11533,9**	**14622,09**
Сер. ариф	**10,85**	**11,73**						

Обрахуємо показники асиметрії та ексцесу для обох рядів даних:

$$A_1=\frac{n_1\sum\limits_{i=1}^{n}\left(x_{i1}-\overline{x_1}\right)^3}{\left(n_1-1\right)\left(n_1-2\right)s_1^{\,3}}=\frac{13\cdot 317{,}91}{12\cdot 11\cdot\left(4{,}69\right)^3}=0{,}3\,,$$

$$A_2=\frac{n_2\sum\limits_{i=1}^{n}\left(x_{i2}-\overline{x_2}\right)^3}{\left(n_2-1\right)\left(n_2-2\right)s_2^{\,3}}=\frac{15\cdot 318{,}63}{14\cdot 13\cdot\left(5{,}08\right)^3}=0{,}2\,,$$

$$E_1=\frac{n_1(n_1+1)\sum_{i=1}^{n}\left(x_{i1}-\overline{x_1}\right)^4}{(n_1-1)(n_1-2)(n_1-3)s_1^{\,4}}-\frac{3(n_1-1)^2}{(n_1-2)(n_1-3)}=$$

$$=\frac{13\cdot 14\cdot 11533{,}9}{12\cdot 11\cdot 10\cdot(4{,}69)^4}-\frac{3\cdot 12}{11\cdot 10}=2{,}96\,,$$

$$E_2=\frac{n_2(n_2+1)\sum_{i=1}^{n}\left(x_{i2}-\overline{x_2}\right)^4}{(n_2-1)(n_2-2)(n_2-3)s_2^{\,4}}-\frac{3(n_2-1)^2}{(n_2-2)(n_2-3)}=$$

$$=\frac{15\cdot 16\cdot 14622{,}09}{14\cdot 13\cdot 12\cdot(5{,}08)^4}-\frac{3\cdot 14}{13\cdot 12}=2{,}14\,.$$

Обрахуємо критичні значення:

$$A^1_{крит}=3\sqrt{\frac{6(n_1-1)}{(n_1+1)(n_1+3)}}=3\sqrt{\frac{6\cdot 12}{14\cdot 16}}=1{,}7\,,$$

$$A^2_{крит}=3\sqrt{\frac{6(n_2-1)}{(n_2+1)(n_2+3)}}=3\sqrt{\frac{6\cdot 15}{16\cdot 18}}=1{,}68\,,$$

$$E^1_{крит}=5\sqrt{\frac{24n_1(n_1-2)(n_1-3)}{(n_1+1)^2(n_1+3)(n_1+5)}}=5\sqrt{\frac{24\cdot 13\cdot 11\cdot 10}{14^2\cdot 16\cdot 18}}=3{,}9\,,$$

$$E^2_{крит}=5\sqrt{\frac{24n_2(n_2-2)(n_2-3)}{(n_2+1)^2(n_2+3)(n_2+5)}}=5\sqrt{\frac{24\cdot 15\cdot 13\cdot 12}{16^2\cdot 18\cdot 20}}=3{,}9\,.$$

Оскільки показники асиметрії (A) й ексцесу (E) для обох рядів даних менші своїх критичних значень ($A_{крит}$ і $E_{крит}$), то згідно з критерієм Є. І. Пустильника розподіл вважається нормальним: $A_1=0{,}3<A^1_{крит}=1{,}7$, $A_2=0{,}2<A^2_{крит}=1{,}68$, $E_1=2{,}96<E^1_{крит}=3{,}9$, $E_2=2{,}14<E^2_{крит}=3{,}9$.

Перевіримо гомогенність дисперсій скориставшись F-критерієм Фішера.

Для кожної групи обчислюють вибіркові дисперсії:

$$\sigma_1{}^2 = \frac{\sum_{i=1}^{13}\left(x_i{}^1 - \overline{x_i{}^1}\right)^2}{13-1} = \frac{263,66}{12} \approx 21,97,$$

$$\sigma_2{}^2 = \frac{\sum_{i=1}^{15}\left(x_i{}^2 - \overline{x_i{}^2}\right)^2}{15-1} \approx 25,78.$$

Знайдемо емпіричне значення F-критерію – $F_{екс}$:

$$F_{екс} = \frac{\sigma_2^2}{\sigma_1^2} = \frac{25,75}{21,97} = 1,172.$$

Визначаємо критичне значення критерію :

$$F_{табл} = F\left(\alpha = 0,05, f_2 = 14, f_1 = 12\right) \approx 2,6 \ .$$

Отже, можна зробити висновок про рівність дисперсій обох вибірок.

Висунемо гіпотези:

H_0: Відмінності між першою та другою групами випадкові.

H_1: Відмінності між першою та другою групами достовірні, значимі.

Знайдемо значення t-критерію Стьюдента:

$$t_{емп} = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{\frac{\left(\sigma_1^2\left(n_1-1\right)+\sigma_2^2\left(n_2-1\right)\right)\left(n_1+n_2\right)}{\left(n_1+n_2-2\right)\cdot n_1\cdot n_2}}} =$$

$$= \frac{11,73-10,85}{\sqrt{\frac{\left(21,97\left(15-1\right)+25,78\left(13-1\right)\right)\left(15+13\right)}{\left(15+13-2\right)\cdot 15\cdot 13}}} = \frac{0,88}{1,84} \approx 0,47.$$

Для ступенів свободи: $\nu = 15+13-2 = 26$ та рівня значущості $\alpha = 0,05$ знаходимо $t_{табл} = 2,056$.

Оскільки $t_{емп} = 0,47 < t_{табл} = 2,056$, то приймають гіпотезу H_0.

Отже, відмінності між групами випадкові. Тому, групи подібні і ми можемо взяти обрану групу за контрольну.

5.4. СТАТИСТИЧНИЙ t-КРИТЕРІЙ СТЬЮДЕНТА ДЛЯ ЗАЛЕЖНИХ ВИБІРОК

Розглянемо алгоритм обчислення t-критерію Стьюдента для залежних вибірок:

1. Висувають гіпотези:

   H_0: Відмінності між першою та другою групами випадкові.

   H_1: Відмінності між першою та другою групами достовірні, значимі.

2. Обчислюють середнє арифметичне різниці:

$$\Delta\overline{x} = \frac{\sum_{i=1}^{n}\left(x_i^1 - x_i^2\right)}{n}.$$

3. Обчислюють величину середньої помилки середньої арифметичної різниці:

$$m_{\Delta\overline{x}} = \frac{S_{\Delta\overline{x}}}{\sqrt{n}}, \text{ де } S_{\Delta\overline{x}} = \sqrt{\frac{\sum_{i=1}^{n}\left(\Delta x_i - \Delta\overline{x}\right)^2}{n-1}}.$$

4. Знаходять значення t-критерію Стьюдента за формулою:

$$t_{емп} = \frac{\left|\Delta\overline{x}\right|}{m_{\Delta\overline{x}}}.$$

5. Знаходять число ступенів свободи, яке залежить від кількості досліджуваних в обох вибірках: $\nu = n - 1$.
6. Для знайденого емпіричного значення визначають рівень значущості та використовуючи таблицю критичних точок розподілу Стьюдента знаходимо $t_{табл}$.
7. Роблять висновок: якщо $t_{емп} \geq t_{табл}$ – приймають гіпотезу H_1, якщо $t_{емп} < t_{табл}$ – приймають гіпотезу H_0.

Приклад 5.3. Було виміряно час виконання завдання для групи респондентів в умовах А та В.

Результати дослідження подані у таблиці:

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A	35	37	37	43	45	43	65	50	60	55	55	47	65	35
B	45	46	45	45	40	45	50	50	45	50	50	40	60	40

Встановити рівень статистичної значущості відмінностей часу, затраченого респондентами в умовах А та В.

Розв'язування. У таблиці представлені результати однієї й тієї ж групи досліджуваних, з якими двічі проводилося психологічне тестування. Змістовно такі вибірки називають залежними (зв'язаними, парними).

Розподіл у обох вибірках дослідження відповідає закону нормального розподілу (перевірити самостійно).

Висунемо дві гіпотези:

H_0: Відмінності між результатами випадкові.

H_1: Відмінності між результатами достовірні та значимі.

Складемо таблицю первинних даних та проміжних розрахунків

№	Умови А	Умови В	$x_1 - x_2$	$(\Delta x - \Delta\bar{x})^2$
1	35	45	-10	132,25
2	37	46	-9	110,25
3	37	45	-8	90,25
4	43	45	-2	12,25
5	45	40	5	12,25
6	43	45	-2	12,25
7	65	50	15	182,25
8	50	50	0	2,25
9	60	45	15	182,25
10	55	50	5	12,25
11	55	50	5	12,25

12	47	40	7	30,25
13	65	60	5	12,25
14	35	40	-5	42,25
Сума			21	845,5

Обчислюємо середнє арифметичне різниці та величину середньої помилки середньої арифметичної різниці:

$$\Delta\bar{x} = 1{,}5,\ S_{\Delta\bar{x}} = \sqrt{\frac{\sum_{i=1}^{n}\left(\Delta x_i - \Delta\bar{x}\right)^2}{n-1}} = \sqrt{\frac{845{,}5}{13}} \approx 8{,}065,$$

$$m_{\Delta\bar{x}} = \frac{S_{\Delta\bar{x}}}{\sqrt{n}} = \frac{8{,}065}{\sqrt{14}} \approx 2{,}155.$$

Знаходимо значення t-критерію Стьюдента:

$$t_{емп} = \frac{\left|\Delta\bar{x}\right|}{m_{\Delta\bar{x}}} = \frac{1{,}5}{2{,}155} \approx 0{,}696.$$

Для числа ступенів свободи $\nu = 13$ та рівня значущості $\alpha = 0{,}05$, використовуючи таблицю критичних точок розподілу Стьюдента, знаходимо $t_{табл} = 2{,}160$.

Оскільки $t_{емп} = 0{,}696 < t_{табл} = 2{,}160$, то приймають гіпотезу H_0, тобто відмінності між результатами випадкові.

6. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ НА ОСНОВІ НЕПАРАМЕТРИЧНИХ КРИТЕРІЇВ. ВИЯВЛЕННЯ ВІДМІННОСТЕЙ У РІВНІ ОЗНАКИ

6.1. Q -КРИТЕРІЙ РОЗЕНБАУМА

Q -критерій Розенбаума відносять до простих непараметричних статистичних критеріїв. Особливостями застосування критерію є:

1. Використовується для оцінки розбіжностей між двома вибірками за рівнем деякої досліджуваної ознаки, яка кількісно виміряна.
2. У кожній з вибірок повинно бути не менше 11 досліджуваних.
3. Вибірки повинні бути незалежними.
4. Потужність критерію невисока, проте він дозволяє швидко оцінити розбіжності між двома вибірками за якою-небудь ознакою.

Однак, якщо критерій не виявляє достовірних розбіжностей, то це ще не означає, що їх дійсно немає. В цьому випадку слід застосовувати критерій Фішера.

Якщо ж Q-критерій виявляє достовірні розбіжності між вибірками з рівнем значущості $\alpha = 0{,}01$, то можна обмежитись лише ним і уникнути труднощів застосування інших критеріїв.

5. Даний метод вимагає достатньо точно виміряних ознак.
6. Робота з Q -критерієм передбачає підрахунок так званих «хвостів» (критерій ще називають критерієм «хвостів»).

Обмеження на вибірки критерію:

1. Обсяги вибірок не повинні занадто відрізнятися:
   a) якщо обсяги вибірок менші 50, то $|n_1 - n_2| \le 10$, тобто абсолютне значення різниці між n_1 і n_2 не повинно бути більше 10 спостережень;

б) якщо обсяги вибірок між 50 і 100, то $|n_1 - n_2| \le 20$, тобто абсолютне значення різниці між n_1 і n_2 не повинно бути більше 20 спостережень;

в) якщо обсяги вибірок перевищують 100, то одна з вибірок не повинна перевищувати іншу більш, ніж у 1,5-2 рази.

3. Діапазони значень ознаки в двох вибірках не повинні співпадати між собою.
4. Якщо найбільше і найменше значення припадають на одну вибірку, то Q -критерій застосувати не можна.

Застосування критерію розпочинаємо з того, що впорядковуємо значення ознаки в обох вибірках за зростанням (або спаданням) ознаки.

Алгоритм підрахунку Q -критерію Розенбаума:

1. Перевіряють, чи дотримані обмеження.
2. Висувають гіпотези:

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

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

3. Впорядковують значення ознаки в обох вибірках за зростанням (або спаданням) ознаки. Розглядають як першу вибірку ту, в якій значення ймовірно вищі, і як другу – де значення ймовірно нижчі.
4. Визначають найвище (максимальне) значення у другій вибірці.
5. Підраховують кількість значень у першій вибірці, які перевищують максимальне значення у другій. Позначимо отримане значення як S_1.
6. Визначають найнижче (мінімальне) значення у вибірці 1.
7. Підраховують кількість значень у вибірці 2, які нижче мінімального значення вибірки 1. Позначимо отримане значення як S_2.
8. Обчислюють емпіричне значення Q за формулою: $Q_{емп} = S_1 + S_2$.
9. Визначають критичні значення Q для n_1 і n_2:

Якщо $n_1, n_2 < 26$, то по таблиці критичних значень критерію Розенбаума в залежності від n_1 і n_2 знаходять критичне значення критерію.

При $n_1, n_2 > 26$ порівнюють отриману емпіричну величину з $Q_{крит} = 8$ при $\alpha = 0{,}05$ і $Q_{крит} = 10$ при $\alpha = 0{,}01$.

Якщо $Q_{емп} < Q_{крит}$, то відмінності між вибірками статистично незначимі, тобто приймається гіпотеза H_0.

<u>Приклад 6.1.</u> Було проведено дослідження пам'яті учнів п'ятих класів, один з яких з поглибленим вивченням математики (А), а інший навчається за стандартною програмою (В). Встановити рівень статистичної значущості відмінностей рівня пам'яті учнів п'ятих класів, що навчаються у класі з поглибленим вивченням математики з рівнем пам'яті учнів п'ятих класів, що навчаються за стандартною програмою?

Результати дослідження наведено у таблиці:

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14
А	31	30	27	30	29	28	27	26	30	31	29	30	25	
В	32	33	29	32	31	28	28	29	27	30	30	32	26	27

Розв'язування. Оскільки обсяги вибірок менші 50, то перевіримо умову $|n_1 - n_2| \le 10$: $|13 - 14| \le 10$.

Впорядковуємо значення ознаки в обох вибірках за спаданням ознаки.

В	А
33	S_1
32	
32	
32	
31	31
	31
30	30
30	30
	30

Продовження таблиці

	30
29	29
29	29
28	28
28	
27	27
27	27
26	26
S_2	25

Висуваємо гіпотези:

H_0: Рівень пам'яті учнів п'ятих класів, що навчаються у класі з поглибленим вивченням математики не перевищує рівня пам'яті учнів п'ятих класів, що навчаються за стандартною програмою.

H_1: Рівень пам'яті учнів п'ятих класів, що навчаються у класі з поглибленим вивченням математики перевищує рівень пам'яті учнів п'ятих класів, що навчаються за стандартною програмою.

Обчислюємо емпіричне значення Q : $Q_{емп} = 4 + 1 = 5$.

Оскільки $n_1, n_2 < 26$, то по таблиці критичних значень критерію Розенбаума знаходимо критичні значення критерію: $Q_{крит} = 6$ при $\alpha = 0,05$.

Оскільки $Q_{емп} < Q_{крит}$, то приймається H_0, а, отже, рівень пам'яті учнів п'ятих класів, що навчаються у класі з поглибленим вивченням математики не перевищує рівня пам'яті учнів п'ятих класів, що навчаються за стандартною програмою.

6.2. *U*-КРИТЕРІЙ МАННА-УІТНІ

U-критерій Манна-Уітні призначений для оцінки відмінностей між двома вибірками на рівні деякої ознаки. Він дозволяє виявити відмінності між малими вибірками, коли $n_1, n_2 \geq 3$ або $n_1 = 2$, $n_2 \geq 5$, і є більш потужним, ніж критерій Розенбаума.

Обмеження критерію:

1. Кожна вибірка повинна мати не менше 3 спостережень $n_1, n_2 \geq 3$.
2. Допускається, що в одній вибірці може бути 2 спостереження, але в другій їх повинно бути не менше 5: $n_1 = 2$, $n_2 \geq 5$.
3. Кожна вибірка повинна мати не більше 60 спостережень: $n_1, n_2 \leq 60$.

Приклад формулювання гіпотез:

H_0: Відмінності між X_1 та X_2 випадкові (або їх взагалі не існує). Отже, наші групи подібні і ми можемо взяти вибрану групу як контрольну.

H_1: Відмінності між X_1 та X_2 достовірні, значимі. Вони можуть бути викликані великим розмахом значень індивідуальних показників, і тоді необхідно змінити або контрольну, або експериментальну групу.

Процедура перевірки висунутих гіпотез така:

1. Дані обох груп об'єднують у нову таблицю, розташувавши їх в порядку збільшення показників. Показники експериментальної групи кодують літерою X_1, а контрольної групи – X_2.
2. Кожному значенню отриманого ряду присвоюють його ранг. Якщо в ряді є декілька однакових числових значень, то використовують правило зв'язаних рангів.
3. Знаходять суму рангів окремо для експериментальної і контрольної груп.
4. Обчислюємо значення *U*-критерію за формулою:

$$U_{емп} = n_1 n_2 + \frac{n_{R_{\max}}\left(n_{R_{\max}} + 1\right)}{2} - R_{\max},$$

де n_1 – кількість досліджуваних у 1 групі; n_2 – кількість досліджуваних у 2 групі; $n_{R_{\max}}$ – кількість досліджуваних у групі із більшою ранговою сумою; $R_{\max}$ – найбільша рангова сума.

5. За таблицею критичних значень U-критерію знаходимо критичне значення $U_{крит}$ для n_1 і n_2:

   Якщо $U_{емп} > U_{крит}$, то приймається гіпотеза H_0.

   Якщо $U_{емп} \leq U_{крит}$, то приймається гіпотеза H_1.

   Чим нижче значення U, тим вище надійність відмінностей.

<u>Приклад 6.2.</u> Для двох груп підлітків А та В було досліджено рівень їх агресивності. Перевірити чи відмінності між А та В випадкові: чи можемо взяти групу В як контрольну.

Результати дослідження наведено у таблиці:

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
А	10	15	20	8	6	7	9	8	15	14					
В	8	10	19	4	15	20	8	6	7	9	8	15	14	16	17

Розв'язування. Кожна вибірка має не менше 3 спостережень.

Сформулюємо гіпотези:

H_0: Відмінності між А та В випадкові (або їх взагалі не існує). Отже, наші групи подібні і ми можемо взяти групу В як контрольну.

H_1: Відмінності між А та В достовірні, значимі.

Дані обох груп об'єднаємо у нову таблицю, розташувавши їх в порядку збільшення показників. Показники експериментальної групи позначимо літерою А, а контрольної групи – В:

№	група	показник	ранг	№	група	показник	ранг
1	В	4	1	14	В	10	13,5
2	В	6	2,5	15	А	14	15,5
3	А	6	2,5	16	В	14	15,5
4	В	7	4,5	17	А	15	18,5
5	А	7	4,5	18	А	15	18,5
6	А	8	8	19	В	15	18,5
7	А	8	8	20	В	15	18,5
8	В	8	8	21	В	16	21
9	В	8	8	22	В	17	22
10	В	8	8	23	В	19	23
11	А	9	11,5	24	А	20	24,5
12	В	9	11,5	25	В	20	24,5
13	А	10	13,5				

Кожному значенню отриманого ряду присвоюємо його ранг. Якщо в ряді є декілька однакових числових значень, то використовуємо правило зв'язаних рангів.

Знайдемо суму рангів окремо для групи А та В: $R_A = 125$, $R_B = 200$, отже $R_{\max} = 200$ для групи В, тому $n_{R_{\max}} = 15$

Обчислюємо значення U-критерію:

$$U_{емп} = n_1 n_2 + \frac{n_{R_{\max}}\left(n_{R_{\max}} + 1\right)}{2} - R_{\max} = 10 \cdot 15 + \frac{15(15+1)}{2} - 200 = 70.$$

За таблицею критичних значень U-критерію знаходимо критичне значення $U_{крит}$ для $n_1 = 10$ і $n_2 = 15$: $U_{крит} = 44$.

Оскільки $U_{емп} = 70 > U_{крит} = 44$, то приймається гіпотеза H_0, а, отже відмінності між групами незначимі, тому груму В можна обрати в якості контрольної групи.

6.3. H -КРИТЕРІЙ КРУСКАЛЯ-УОЛЛІСА

Критерій H використовується для оцінки відмінностей за ступенем вираженості аналізованої ознаки одночасно між трьома, чотирма і більше незв'язними вибірками (групами). Він дозволяє виявити ступінь зміни ознаки у вибірках, не вказуючи, однак, на напрямок цих змін. Цей критерій є продовженням критерію Манна-Уітні більш ніж на дві вибірки. Слід підкреслити, що у вибірках може бути різна, тобто не обов'язково однакова кількість респондентів.

Метод може бути застосований в наступних випадках:

- Якщо $\alpha = 0{,}05$, то при зіставленні трьох вибірок допускається, щоб в одній було 3 значення, а в другій та третій по 2 значення.
- Якщо $\alpha = 0{,}01$, то необхідно щоб в кожній вибірці було не менше 3 значень, або щоб в крайньому разі, в одній з вибірок було 4, а в двох інших – по 2 значення.
- Для оцінки отриманих результатів необхідно використовувати спеціальну таблицю критичних значень, яка розроблена тільки для трьох вибірок. Якщо будь-які відмінності стають «стертими», то їх можна виявити, попарно порівнявши між собою.
- При більшій кількості вибірок і об'єктів в кожній вибірці необхідно використовувати таблицю критичних значень критерію χ^2, оскільки критерій Крускаля-Уолліса асимптотично наближається до розподілу χ^2. Кількість ступенів свободи визначається за формулою: $\nu = c - 1$, де с – кількість вибірок, що розглядаються.
- Вибірки повинні бути незалежними.

Алгоритм H -критерію Крускаля-Уолліса:

1. Формулюємо гіпотези:

H_0: Між вибірками 1, 2, 3 і т. д. існують лише випадкові відмінності в рівні досліджуваної ознаки.

H_1: Між вибірками 1, 2, 3 і т. д. існують невипадкові відмінності в рівні досліджуваної ознаки.

2. Записуємо значення ознаки для кожної групи.
3. Дані обох груп об'єднують у нову таблицю, розташувавши їх в порядку збільшення показників.
4. Ранжують загальну вибірку у порядку зростання.
5. Знаходять суму рангів окремо для кожної групи.
6. Обчислюємо значення H-критерію за формулою:

$$H_{емп} = \frac{12}{n(n+1)} \cdot \sum_{j=1}^{k} \frac{R_j^{\,2}}{n_j} - 3(n+1),$$

де n – загальна кількість респондентів у об'єднаній вибірці; k – кількість груп; R_j, $j=\overline{1,k}$, – сума рангів для j-ої групи; n_j, $j=\overline{1,k}$, – кількість досліджуваних у кожній групі.

7. Знаходимо $H_{крит}$: якщо $H_{емп} < H_{крит}$, то приймається гіпотеза H_0, тобто між вибірками 1, 2, 3 і т. д. існують лише випадкові відмінності в рівні досліджуваної ознаки.

<u>Приклад 6.3.</u> Чотири групи досліджуваних виконували тест Бурдона в різних експериментальних умовах. Різниця в умовах полягала в наступному: шум вентилятора, відтворення аудіозапису шкільної перерви, відтворення аудіозапису магістралі, повна тиша. Питання в тому, чи залежить ефективність тесту від умов експерименту. Кожна група включала в себе чотири людини.

Результати дослідження наведено в таблиці:

№	*1 група*	*2 група*	*3 група*	*4 група*
1	23	45	34	21
2	20	12	24	22
3	34	34	25	26
4	35	11	40	27

Розв'язування. Формулюємо гіпотези:

H_0: Між вибірками 1, 2, 3, 4 існують лише випадкові відмінності в рівні досліджуваної ознаки.

H_1: Між вибірками 1, 2, 3, 4 існують невипадкові відмінності в рівні досліджуваної ознаки.

Дані обох груп об’єднуємо у нову таблицю, розташувавши їх в порядку збільшення показників та ранжуємо загальну вибірку у порядку зростання:

№	*Група*	*Показник*	*ранг*
1	2	11	1
2	2	12	2
3	1	20	3
4	4	21	4
5	4	22	5
6	1	23	6
7	3	24	7
8	3	25	8
9	4	26	9
10	4	27	10
11	1	34	12
12	2	34	12
13	3	34	12
14	1	35	14
15	3	40	15
16	2	45	16

Знаходимо суму рангів окремо для кожної групи: $T_1 = 35$, $T_2 = 31$, $T_3 = 42$, $T_4 = 28$. Обчислюємо значення H-критерію:

$$H_{емп} = \frac{12}{16(16+1)} \cdot \left(\frac{35^2}{4} + \frac{31^2}{4} + \frac{42^2}{4} + \frac{28^2}{4} \right) - 3(16+1) \approx 1{,}213 .$$

Оскільки в експерименті брало участь 4 вибірки, то необхідно використовувати таблицю критичних значень критерію χ^2 з $\nu = c - 1 = 4 - 1 = 3$ ступенів свободи.

При $\alpha = 0{,}05$ – $H_{крит} = 7{,}8$. Оскільки $H_{емп} = 1{,}213 < H_{крит} = 7{,}8$, то приймається гіпотеза H_0, а, отже між вибірками 1, 2, 3, 4 існують лише випадкові відмінності в рівні досліджуваної ознаки.

6.4. *S*-КРИТЕРІЙ ТЕНДЕНЦІЙ ДЖОНКІРА

S-критерій тенденцій Джонкіра призначений для виявлення тенденцій зміни ознаки при переході від вибірки до вибірки при порівнянні трьох і більше вибірок.

Умови застосування *S*-критерію тенденцій Джонкіра:

1. Вимірювання може здійснюватися за шкалою порядку, інтервалів або співвідношень.
2. Вибірки повинні бути незалежними.
3. Кількість елементів в кожній вибірці має бути однакова. Якщо кількість спостережень неоднакова, то необхідно урівняти їх.
4. Нижня межа застосування критерію: не менше трьох вибірок і не менше двох елементів в кожному спостереженні. Верхня межа – не більше 6 вибірок і не більше 10 об'єктів у кожній вибірці. У всіх інших випадках слід використовувати H -критерій Крускаля-Уолліса.

Приклад формулювання гіпотез:

H_0: Тенденція до збільшення значень ознак від вибірки до вибірки є випадковою.

H_1: Тенденція до збільшення значень ознак від вибірки до вибірки не є випадковою.

Алгоритм *S*-критерію тенденцій Джонкіра:

1. Перевірити чи виконуються обмеження критерію. Якщо кількість значень у вибірках не однакова, то зрівняти їх орієнтуючись на меншу, обираючи данні для урівнювання з інших вибірок випадковим чином.

2. У кожній вибірці впорядкувати значення за зростанням.
3. Сформулювати гіпотези.
3. Починаючи з крайнього лівого стовпця підрахувати для кожного індивідуального значення S_i – кількість показників, що перевищують його значення у всіх стовпчиках справа. Отримані суми приписати поруч з кожним індивідуальним значенням.
4. Знайти суму всіх S_i: $A = \sum S_i$.
5. Обрахувати максимально ймовірну кількість всіх перевищень для наших груп:

$$B = \frac{c(c-1)}{2} n^2,$$

де c – кількість груп, що досліджується; n – кількість досліджуваних в одній групі.

6. Знайти емпіричне значення критерію S за формулою:

$$S_{емп} = 2A - B.$$

7. За таблицею критичних значень S-критерію тенденцій Джонкіра для рівня значимості α, заданої кількості груп (c) і кількості елементів в кожній групі (n) знаходимо $S_{крит}$: якщо $S_{емп} < S_{крит}$, то приймається гіпотеза H_0; якщо $S_{емп} \ge S_{крит}$, то приймається гіпотеза H_1.

Приклад 6.4. Студенти першого, другого та третього курсу були протестовані. Тест містив 50 запитань. Чи можна вважати, що збільшення числа правильних відповідей при переході від курсу до курсу є випадковим?

Результати дослідження наведено у таблиці:

№	*1 курс*	*2 курс*	*3 курс*
1	31	35	41
2	33	35	40
3	32	39	44
4	34	37	44
5	35	40	42

Розв'язування. Кількість значень у вибірках є однаковою. Знайдемо середнє арифметичне для кожної групи: $\bar{x}_{1курсу}=33$, $\bar{x}_{2курсу}=37{,}2$, $\bar{x}_{3курсу}=42{,}2$.

У кожній вибірці впорядковуємо значення за зростанням:

№	*1 курс*	S_1	*2 курс*	S_2	*3 курс*
1	31	10	35	5	40
2	32	10	35	5	41
3	33	10	37	5	42
4	34	10	39	5	44
5	35	8	40	4	44

Починаючи з крайнього лівого стовпця підраховуємо для кожного індивідуального значення S_i – кількість показників, що перевищують його значення у всіх стовпчиках справа. Отримані суми дописуємо біля кожного значення у стовпці S_1 та S_2. Знаходимо $A=\sum S_i=72$.

Обраховуємо $B=\frac{c(c-1)}{2}n^2=\frac{3(3-1)}{2}5^2=75$.

Знаходимо емпіричне значення критерію *S*:

$$S_{емп}=2A-B=2\cdot 72-75=65.$$

За таблицею критичних значень *S*-критерію тенденцій Джонкіра для рівня значимості $\alpha=0{,}01$ знаходимо $S_{крит}=S_{крит}(0{,}01;3;5)=45$.

Оскільки $S_{емп}=65\geq S_{крит}=45$, то гіпотеза H_0 відхиляється на рівні значимості $\alpha=0{,}01$, а, отже, збільшення числа правильних відповідей при переході від курсу до курсу не є випадковим.

7. ОЦІНЮВАННЯ ВІРОГІДНОСТІ ЗСУВУ У ЗНАЧЕННЯХ ОЗНАКИ

7.1. *G*-КРИТЕРІЙ ЗНАКІВ

G-критерій знаків призначений для виявлення загального напрямку зсуву досліджуваної ознаки. Він дозволяє встановити, в якому напрямку у вибірці вцілому змінюються значення ознаки при переході від першого вимірювання до другого (від кращого до гіршого, від слабкого до сильного тощо). Він дозволяє виявити існування зсуву та його напрямок, однак, не дозволяє встановити інтенсивність зсуву.

Порівнюючи дві вибірки, можна помітити наявність переважаючих змін (зсувів) у бік збільшення або зменшення відносно первинних значень. Цей критерій оцінює наскільки ці зрушення можуть розглядатися як невипадкові. Як правило, дослідник вже в процесі експерименту може помітити, що для більшості досліджуваних показники в другому замірі мають тенденцію, наприклад, до збільшення. Однак все ж необхідно довести, що позитивний зсув є переважаючим.

Зсуви, які здаються нам переважаючими, називаються «типовими», тобто, якщо ми, наприклад, знижуємо рівень агресивності, то зниження агресивності, підвищення емоційної стійкості це є типові реакції (зсуви). Тоді нетиповими реакціями є підвищення агресивності, зниження емоційної стійкості. І при розрахунку цього критерію можна встановити, чи не занадто багато нетипових зсувів, щоб вважати зсув в типовому напрямку суттєвим. Отже, чим менше «нетипових» значень, тим достовірніший типовий зсув.

Ще є можливість «нульових» зсувів, коли реакція не змінюється або показники не збільшуються або не зменшуються, а залишаються на колишньому рівні. Проте такі «нульові» зсуви в критерії знаків виключаються з розгляду. При цьому кількість пар, що порівнюються зменшується на кількість таких «нульових» зсувів.

Суть критерію знаків полягає в тому, що він визначає, чи не занадто багато «нетипових зсувів», щоб вважати переважаючим зсув в «типовому» напрямку.

$G_{емп}$ – це число «нетипових» зсувів. Чим менше $G_{емп}$ тим більша ймовірність того, що зсуви в «типовому» напрямку є статистично достовірними.

Обмеженнями критерію знаків є: кількість спостережень в обох замірах – не менше 5 і не більше 300.

Алгоритм G-критерію знаків:

1. Підрахувати кількість нульових реакцій і вилучити їх із розгляду.
2. Визначити переважаючий напрямок змін. Вважати зсуви в переважаючому напрямку типовими.
3. Висунути гіпотези:

   H_0: Переважання типового напрямку зсуву є випадковим.

   H_1: Переважання типового напрямку зсуву не є випадковим.

4. Визначити кількість нетипових зсувів, яке і дорівнює $G_{емп}$.
5. За таблицею критичних значень критерію знаків G визначити $G_{крит}$ для n, де n – кількість порівнюваних пар даних.

Якщо $G_{емп} \leq G_{крит}$, то приймається гіпотеза H_1 і зсув в типову сторону можна вважати достовірним.

Приклад 7.1. Після проведення корекційної роботи по зниженню тривожності психолог одержав результати:

№	1	2	3	4	5	6	7	8	9	10
До проведення корекційної роботи	30	39	35	34	40	35	22	22	32	23
Після проведення корекційної роботи	25	39	26	25	34	32	20	23	29	20

Продовження таблиці

№	11	12	13	14	15	16	17	18	19	20
До проведення корекційної роботи	28	26	32	40	38	35	34	29	30	28
Після проведення корекційної роботи	30	26	27	36	37	30	30	24	29	25

З'ясувати, чи є ефективною проведена корекційна робота.

Розв'язування. Пари порівнюваних результатів під номером 2 та 12 мають «нульовий» зсув, тому їх вилучимо з розгляду:

№	1	2	3	4	5	6	7	8	9
До проведення корекційної роботи	30	35	34	40	35	22	22	32	23
Після проведення корекційної роботи	25	26	25	34	32	20	23	29	20
Зсув	+	+	+	+	+	+	–	+	+
№	10	11	12	13	14	15	16	17	18
До проведення корекційної роботи	28	32	40	38	35	34	29	30	28
Після проведення корекційної роботи	30	27	36	37	30	30	24	29	25
Зсув	–	+	+	+	+	+	+	+	+

Визначаємо переважаючий напрямок змін.

Типовий зсув: зменшення тривожності після проведення корекційної роботи.

Нетиповий зсув: збільшення тривожності після проведення корекційної роботи.

Висуваємо гіпотези:

H_0: Переважання типового напрямку зсуву є випадковим.

H_1: Переважання типового напрямку зсуву не є випадковим.

Визначаємо кількість нетипових зсувів: $G_{емп} = 2$.

За таблицею критичних значень критерію знаків G визначаємо $G_{крит} = 3$ для $n = 18$ при $\alpha = 0{,}01$. Оскільки $G_{емп} = 2 \leq G_{крит} = 3$, то приймається гіпотеза H_1 і зсув в типову сторону можна вважати достовірним. Отже, можна вважати, що проведена корекційна робота є вдалою.

7.2. Т-КРИТЕРІЙ ВІЛКОКСОНА

Т-критерій Вілкоксона застосовується для зіставлення показників, що виміряні в різних умовах на одній і тій самій вибірці. Він дозволяє встановити не тільки напрямок змін, але і їх вираженість. За допомогою нього визначають чи є зсув показників в одному напрямку більш інтенсивним ніж в іншому.

Обмеженнями *Т*-критерію Вілкоксона є:

- Вибірки повинні бути залежними і мати однаковий об’єм.
- Мінімальна кількість досліджуваних, що проходять вимірювання в двох умовах – 5 людей, максимальна – 50.
- «Нульові» зсуви виключаються з кількості спостережень.

Суть методу полягає в тому, що зіставляють вираженість зсувів в тому чи іншому напрямку по модулю. Для цього ранжують всі величини зсувів по модулю, а потім сумують ранги. Якщо зсуви в додатну та від’ємну сторони відбуваються випадково, то суми рангів будуть приблизно рівними. Якщо зсув в деякому напрямку буде переважати зсув в протилежному напрямку, то і сума рангів для переважаючого випадку буде значно вищою.

Типовим будемо вважати зсув в напрямку, що зустрічається частіше. Іноді, при деяких значеннях n, а саме $n > 18$, можна взагалі відмовитись від поняття типовості. Якщо зсувів порівну, то типових немає, але по інтенсивності певний напрямок зсуву може мати значиму тенденцію.

Алгоритм *T*-критерію Вілкоксона:

1. Скласти список досліджуваних у довільному порядку.
2. Обрахувати різницю між індивідуальними значеннями в першому та другому вимірюванні. Визначити, що буде вважатися типовим зсувом і сформулювати відповідні гіпотези.

Наприклад:

H_0: Інтенсивність зсувів у типовому напрямку не перевищує інтенсивність зсувів у нетиповому напрямку.

H_1: Інтенсивність зсувів у типовому напрямку переважає інтенсивність зсувів у нетиповому напрямку.

3. Знайти модуль знайдених різниць.
4. Проранжувати знайденні абсолютні значення.
5. Підрахувати суму рангів, що відповідають нетиповому зсуву:

$$T_{емп} = \sum R_r ,$$

де R_r – рангові значення зсувів, що відповідають знаку, який зустрічається рідше.

6. За таблицею критичних значень критерію *T*-Вілкоксона визначити $T_{крит}$ для n.

Якщо $T_{емп} \leq T_{крит}$, то приймається гіпотеза H_1 і зсув в типову сторону по інтенсивності переважає.

<u>Приклад 7.2.</u> Психолог проводить з молодшими школярами корекційну роботу по формуванню розподілу уваги. Для цього визначають кількість помилок до та після корекційних вправ. Чи буде зменшуватися кількість помилок у молодших школярів після проведення корекційних вправ?

Результати дослідження подані у таблиці:

№	1	2	3	4	5	6	7	8	9	10
До	24	12	42	30	40	55	50	52	50	22
Після	22	12	41	31	32	44	50	32	32	21
№	11	12	13	14	15	16	17	18	19	
До	33	78	79	25	28	16	17	12	25	
Після	34	56	78	23	22	12	16	18	25	

Розв'язування. Очевидно, що пари під номером 2, 7, 19 мають нульовий зсув. Вилучаємо їх з розгляду:

№	X	Y	$X-Y$	$\|X-Y\|$	*Ранги*
1	24	22	2	2	7,5
2	42	41	1	1	3,5
3	30	31	-1	1	3,5
4	40	32	8	8	12
5	55	44	11	11	13
6	52	32	20	20	15
7	50	32	18	18	14
8	22	21	1	1	3,5
9	33	34	-1	1	3,5
10	78	56	22	22	16
11	79	78	1	1	3,5
12	25	23	2	2	7,5
13	28	22	6	6	10,5
14	16	12	4	4	9
15	17	16	1	1	3,5
16	12	18	-6	6	10,5

Обраховуємо різницю між індивідуальними значеннями в першому та другому вимірюванні. Типовим зсувом буде зменшення кількості помилок.

Формулюємо гіпотези:

H_0: Інтенсивність зсувів у типовому напрямку не перевищує інтенсивність зсувів у нетиповому напрямку.

H_1: Інтенсивність зсувів у типовому напрямку переважає інтенсивність зсувів у нетиповому напрямку.

Знаходимо модуль знайдених різниць та ранжуємо знайденні абсолютні значення.

Рахуємо суму рангів, що відповідають нетиповому зсуву:

$$T_{емп} = \sum R_r = 3{,}5 + 3{,}5 + 10{,}5 = 17{,}5\,.$$

За таблицею критичних значень критерію *T*-Вілкоксона визначаємо $T_{крит} = 32$ для $n = 16$ і $\alpha = 0{,}01$.

Оскільки $T_{емп} = 17{,}5 \le T_{крит} = 32$, то на рівні значимості $\alpha = 0{,}01$ приймається гіпотеза H_1 і зсув в типову сторону по інтенсивності переважає. Отже, можна зробити висновок про позитивний вплив корекційних вправ на розподіл уваги у молодших школярів.

7.3. χ_r^2-КРИТЕРІЙ ФРІДМАНА

χ_r^2-критерій Фрідмана використовують для порівняння показників, виміряних у трьох і більше умовах на одній і тій же вибірці досліджуваних. Він дозволяє фіксувати тільки наявність змін, але не їх напрямок. Використовується для зіставлення показників, що вимірянні в c умовах ($c \ge 3$) на тій самій вибірці з n досліджуваних ($n \ge 2$).

Алгоритм розрахунку χ_r^2-критерію Фрідмана:

1. Сформулювати гіпотези:

H_0: Між показниками, отриманими в різних умовах, існують лише випадкові відмінності.

H_1: Між показниками, отриманими в різних умовах, існують невипадкові відмінності.

2. Для кожного досліджуваного проранжувати його значення, одержані в різних умовах, за зростанням, тобто меншому значенню приписуємо менший ранг.
3. Підраховати суму рангів T_i, $i=\overline{1,c}$, для кожної умови експерименту.
4. Обраховати емпіричне значення χ_r^2-критерію Фрідмана за формулою:

$$\chi^2_{r\,емп} = \frac{12}{n \cdot c \cdot (c+1)} \sum_{i=1}^{c} T_i^2 - 3n(c+1),$$

де c – кількість умов експерименту, n – обсяг вибірки.

5. Знайти критичне значення $\chi^2_{r\,крит}$ для заданого рівня значимості α:

- при $c=3$ і $n \le 9$ або при $c=4$ і $n \le 5$ значення $\chi^2_{r\,крит}$ визначається за таблицею критичних значень χ_r^2- критерію Фрідмана;
- в інших випадках значення $\chi^2_{r\,крит}$ визначається за таблицею критичних значень критерію χ^2 при $\nu = c-1$.

Якщо $\chi^2_{r\,емп} \ge \chi^2_{r\,крит}$, то гіпотеза H_0 відхиляється на рівні значимості α, в протилежному випадку гіпотеза H_0 приймається на рівні значимості α.

Приклад 7.3. Визначити, чи залежить студентська активність на заняттях від способу поведінки викладача. В якості різних моделей поведінки викладача скористались класифікацією лідерів К. Левіна – демократична, ліберальна, авторитарна. Активність студентів спостерігали незалежні експерти, які за 20-бальною шкалою визначали їх загальний емоційний стан, частоту відповідей, зацікавленість у занятті. У результаті проведеного дослідження було отримано такі дані:

Номер досліджуваного	***Авторитарний стиль***	***Демократичний стиль***	***Ліберальний стиль***
1	15	19	6
2	8	15	3
3	18	12	3
4	14	18	12
5	13	14	7
6	3	10	7

Розв'язування. Формулюємо гіпотези:

H_0: Між показниками, отриманими в різних умовах, існують лише випадкові відмінності.

H_1: Між показниками, отриманими в різних умовах, існують невипадкові відмінності.

Для кожного досліджуваного проранжуємо його значення, одержані в різних умовах:

Номер досліджуваного	**Авторитарний стиль**	**Ранг**	**Демократичний стиль**	**Ранг**	**Ліберальний стиль**	**Ранг**
1	15	2	19	3	6	1
2	8	2	15	3	3	1
3	18	3	12	2	3	1
4	14	2	18	3	12	1
5	13	2	14	3	7	1
6	3	1	10	3	7	2
		$T_1 = 12$		$T_2 = 17$		$T_3 = 7$

Підраховуємо суму рангів T_i, $i = \overline{1,3}$, для кожної умови експерименту:

$$T_1 = 2 + 2 + 3 + 2 + 2 + 1 = 12;$$

$$T_2 = 3 + 3 + 2 + 3 + 3 + 3 = 17;$$

$$T_3 = 1+1+1+1+1+2 = 7.$$

Обраховуємо емпіричне значення χ_r^2-критерію Фрідмана:

$$\chi^2_{r\,емп} = \frac{12}{n \cdot c \cdot (c+1)} \sum_{i=1}^{c} T_i^2 - 3n(c+1) =$$

$$= \frac{12}{6 \cdot 3 \cdot (3+1)} \left(12^2 + 17^2 + 7^2\right) - 3 \cdot 6 \cdot (3+1) = 8{,}333.$$

Знаходимо критичне значення $\chi^2_{r\,крит}$. Оскільки $c=3$ і $n \le 9$, то значення $\chi^2_{r\,крит}$ визначимо за таблицею критичних значень χ_r^2-критерію Фрідмана: $\chi^2_{r\,крит} = 6{,}33$ для $p < 0{,}052$ і $\chi^2_{r\,крит} = 8{,}33$ для $p < 0{,}012$.

Оскільки $\chi^2_{r\,крит} = 8{,}33 \le \chi^2_{r\,емп} = 8{,}333$, то приймається гіпотеза H_1, а, отже, між показниками, отриманими в різних умовах існують невипадкові відмінності.

7.4. *L*-КРИТЕРІЙ ПЕЙДЖА

L-критерій Пейджа застосовується у випадках, коли потрібно обґрунтувати достовірність існування певної тенденції у зміні ознаки в одній і тій же вибірці, якщо кількість вимірювань становить три і більше. Наприклад, ми здійснюємо експериментальний вплив та вимірюємо значення критеріального параметру кожні півроку впродовж двох років спостережень; якщо при цьому параметр змінювався монотонно, то постає питання: чи дійсно є об'єктивно існуюча тенденція зміни параметра чи зміни випадкові? За допомогою *L*-критерію Пейджа можна дати відповідь на дане питання.

Обмеження застосування *L*-критерію за кількістю умов c та обсягом вибірки n: нижній поріг – $c=3$, $n=2$; верхній поріг – $c=6$, $n=12$.

Алгоритм *L*-критерію Пейджа:

1. Проранжувати індивідуальні значення кожного респондента, одержані в різних умовах, за зростанням, тобто меншому значенню приписати менший ранг.
2. Підрахувати суму рангів T_i, $i=\overline{1,c}$, по кожній серії вимірювання ознаки.

3. Стовбці, що відповідають окремим дослідженням розташувати у порядку зростання рангових сум.

4. Сформулювати гіпотези:

H_0: Зростання індивідуальних показників при переході від однієї умови до іншої є випадковим.

H_1: Зростання індивідуальних показників при переході від однієї умови до іншої не є випадковим.

5. Обраховуємо емпіричне значення $L_{емп}$ за формулою:

$$L_{емп} = \sum_{i=1}^{c} (T_i \cdot i).$$

6. Знайти критичне значення $L_{крит}$ для заданого рівня значимості α.

7. Зробити висновок: якщо $L_{емп} \geq L_{крит}$, то гіпотеза H_0 відхиляється на рівні значимості α, в протилежному випадку гіпотеза H_0 приймається на рівні значимості α.

Приклад 7.4. Розглянемо приклад залежності студентської активність на заняттях від способу поведінки викладача, що розглядався у попередньому пункті. Перевіримо чи зміна індивідуальних показників при переході від умови до умови невипадкова.

Скористаємося даними попереднього прикладу, заздалегідь розташувавши за зростанням порядкових сум:

Номер досліджуваного	Ліберальний стиль	Ранг	Авторитарний стиль	Ранг	Демократичний стиль	Ранг
1	6	1	15	2	19	3
2	3	1	8	2	15	3
3	3	1	18	3	12	2
4	12	1	14	2	18	3
5	7	1	13	2	14	3
6	7	2	3	1	10	3
		$T_3 = 7$		$T_1 = 12$		$T_2 = 17$

Сформулюємо гіпотези:

H_0: Зростання індивідуальних показників при переході від однієї умови до іншої є випадковим.

H_1: Зростання індивідуальних показників при переході від однієї умови до іншої не є випадковим.

Обраховуємо емпіричне значення $L_{емп}$:

$$L_{емп} = \sum_{i=1}^{c} \left(T_i \cdot i \right) = 7 \cdot 1 + 12 \cdot 2 + 17 \cdot 3 = 82.$$

Знаходимо критичне значення $L_{крит}$:

$$L_{крит} = \begin{cases} 83, \alpha \le 0{,}001; \\ 81, \alpha \le 0{,}01; \\ 79, \alpha \le 0{,}05. \end{cases}$$

Оскільки $L_{емп} = 82 \ge L_{крит} = 81$ для рівня значимості $\alpha \le 0{,}01$, тому гіпотеза H_0 відхиляється на рівні значимості α, а, отже тенденція до збільшення показника при переході від умови до умови не є випадковою.

8. ВИЯВЛЕННЯ ВІДМІННОСТЕЙ У РОЗПОДІЛІ ОЗНАКИ

8.1. χ^2-КРИТЕРІЙ ПІРСОНА

χ^2-критерій Пірсона (критерій узгодженості, критерій χ^2) є найпотужнішим непараметричним критерієм. Його широко застосовують у дисперсійному аналізі та інших методах аналізу даних. χ^2-критерій Пірсона можна застосовувати як до числових, рангових, так і до номінальних даних. До того ж кількість порівнюваних розподілів не обмежується. Цей критерій оперує не первинними даними, а їх розподілом за класами. Критерій χ^2 відповідає на запитання, чи з однаковою частотою трапляються різні значення ознаки у двох та більше розподілах?

Критерій χ^2 в основному використовується у двох випадках:

- для співставлення емпіричного розподілу ознаки с теоретичним (рівномірним, нормальним або деяким іншим);
- для співставлення двох або більше емпіричних розподілів однієї і тієї ж ознаки.

χ^2-критерій Пірсона побудований таким чином, що при повному співпаданні емпіричного і теоретичного розподілів (або емпіричних розподілів) величина $\chi^2_{емп}$ дорівнює нулеві, і чим більші відмінності між порівнюваними розподілами, тим більша і величина $\chi^2_{емп}$.

Обмеження застосування χ^2-критерію Пірсона:

- обсяг вибірки повинен бути достатньо великим $n \geq 30$. При $n < 30$ критерій χ^2 дає дуже наближені значення. Точність критерію зростає при збільшенні n;
- значення частот для кожної комірки таблиці не повинно бути менше 5, тобто якщо число розрядів задано наперед і не може змінюватися, то ми не зможемо застосувати критерій χ^2 не накопичивши деякої мінімальної кількості спостережень;

- вибрані розряди повинні «вичерпувати» розподіли, тобто охоплювати весь діапазон варіативності ознак. При цьому групування на розряди має бути однакове для всіх порівнюваних розподілів;
- розряди не повинні перетинатися: якщо спостереження віднести до одного розряду, то його вже не можна віднести ні до якого іншого розряду. Сума спостережень по розрядах має дорівнювати загальній кількості спостережень;
- при співставленні в розподілах ознак, які набувають всього два значення, необхідно вносити «поправку на неперервність».

Формулюються основна та альтернативна гіпотези:

H_0: Порівнювані розподіли не відрізняються.

H_1: Порівнювані розподіли відрізняються.

Обчислення емпіричного значення $\chi^2_{емп}$.

Для співставлення емпіричного і теоретичного розподілів:

$$\chi^2_{емп} = \sum_{i=1}^{k} \frac{\left(f_i^{емп} - f_i^{теор}\right)^2}{f_i^{теор}},$$

де k – кількість класових інтервалів; $f_i^{емп}$ – частота емпіричного розподілу, що відповідає i-му класовому інтервалу; $f_i^{теор}$ – частота теоретичного розподілу, що відповідає i-му класовому інтервалу.

Якщо аналізовані дані вимірювали в кількісних або порядкових шкалах, то при порівнянні вибірок однакового обсягу, значення критерію обчислюють за формулою:

$$\chi^2_{емп} = \sum_{i=1}^{k} \frac{\left(f_i - g_i\right)^2}{f_i + g_i},$$

де k – кількість класових інтервалів; f_i, g_i – частоти порівнюваних емпіричних розподілів, що відповідають i-му класовому інтервалу.

У випадку, якщо аналізуються вибірки різного обсягу значення критерію обчислюють за формулою:

$$\chi^2_{емп} = \frac{1}{n_1 n_2} \sum_{i=1}^{k} \frac{\left(n_2 f_i - n_1 g_i\right)^2}{f_i + g_i},$$

де n_1 та n_2 – обсяги першої та другої вибірок відповідно.

Критерій χ^2 можна застосовувати також і для порівняння вибірок значень номінальних ознак. У цьому випадку аналізують дані, подані у вигляді таблиці спряженості ознак. Елементами таблиці є числа, рівні кількостям елементів досліджуваних вибірок, для яких досліджувана ознака набуває значень, котрі відповідають певному класу. Кожний рядок таблиці характеризує розподіл елементів відповідної вибірки за класами, а кожний стовпець – наповненість певного класу в різних вибірках. Значення критерію розраховують за формулою:

$$\chi^2_{емп} = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{\left(a_{ij} - e_{ij}\right)^2}{e_{ij}},$$

де k – кількість класових інтервалів (рядків); r – кількість вибірок (стовпців), a_{ij}, $i = \overline{1,r}$, $j = \overline{1,k}$, – елементи таблиці спряженості ознак; e_{ij}, $i = \overline{1,r}$, $j = \overline{1,k}$, – очікуванні величини, що відповідають значенням a_{ij}:

$$e_{ij} = \frac{a_i a_j}{\sum_{i=1}^{r} \sum_{j=1}^{k} a_{ij}}, \quad a_i = \sum_{j=1}^{k} a_{ij}, \quad a_j = \sum_{i=1}^{r} a_{ij}.$$

Для таблиці зв'язаності ознак розміром 2×2 застосовується критерій χ^2 з поправкою на неперервність:

$$\chi^2_{емп} = \frac{n\left(\left|ad - bc\right| - \frac{n}{2}\right)^2}{(a+b)(c+d)(a+c)(b+d)},$$

де a, b, c, d – дані, що визначаються згідно з таблицею зв'язності розміром 2×2, n – обсяг досліджуваних.

Зауважимо, що формула обчислення значення $\chi^2_{емп}$ містить квадрати різниць частот, а, отже, цей критерій фіксує тільки існування відмінностей в розподілах, але не напрямок змін ознаки.

За таблицею критичних точок розподілу χ^2 знаходимо критичне значення для заданого α-рівня та кількості ступенів свободи: $\nu = k-1$, для таблиць – $\nu = (k-1)(r-1)$, де k – кількість рядків; r – кількість стовпців.

Якщо $\chi^2_{емп}$ менше $\chi^2_{крит}$, то робимо висновок про відсутні статистично значущих відмінностей між розподілами (приймаємо гіпотезу H_0). Якщо $\chi^2_{емп}$ більше або дорівнює $\chi^2_{крит}$, то нульова гіпотеза відхиляється та приймається альтернативна.

Приклад 8.1. При дослідженні рівня успішності навчання математики учнів шостих класів було проведено контрольну роботу. Перевірялась гіпотеза про рівномірний розподіл результатів контрольної роботи між рівнями навчальних досягнень учнів.

Результати контрольної роботи наведено в таблиці:

Рівні навчальних досягнень	***Високий***	***Середній***	***Достатній***	***Низький***	***Всього***
Кількість	18	14	11	11	54

Розв’язування. Оскільки згідно з розглядуваною гіпотезою результати контрольної роботи відповідають рівномірному розподілу, тому теоретичні частоти для вибірки повинні бути однакові і дорівнювати: $\frac{54}{4} = 13{,}5$.

Тоді остаточний варіант розрахункової таблиці даних буде:

Рівні навчальних досягнень	$f_i^{емп}$	$f_i^{теор}$	$f_i^{емп} - f_i^{теор}$	$\left(f_i^{емп} - f_i^{теор}\right)^2$	$\frac{\left(f_i^{емп} - f_i^{теор}\right)^2}{f_i^{теор}}$
Високий	18	13,5	4,5	20,25	1,5
Середній	14	13,5	0,5	0,25	0,019
Достатній	11	13,5	-2,5	6,25	0,.463
Низький	11	13,5	-2,5	6,25	0,463
Сума	54	54	0	33	2,445

Сформулюємо гіпотези:

H_0: Результати контрольної роботи розподілені з однаковою частотою за рівнями навчальних досягнень учнів.

H_1: Результати контрольної роботи розподілені з різною частотою за рівнями навчальних досягнень учнів.

Зробивши проміжні розрахунки, обраховуємо $\chi^2_{емп}$:

$$\chi^2_{емп} = \sum_{i=1}^{k} \frac{\left(f_i^{емп} - f_i^{теор}\right)^2}{f_i^{теор}} = 2{,}445.$$

За таблицею критичних точок розподілу χ^2 знаходимо критичне значення для заданого α-рівня та кількості ступенів свободи: $\nu = k - 1 = 4 - 1 = 3$: $\chi^2_{крит}(\alpha = 0{,}05) = 7{,}8$.

Отже, $\chi^2_{емп} = 2{,}445 < \chi^2_{крит} = 7{,}8$, тому приймається гіпотеза H_0, а, отже, можна вважати, що результати контрольної роботи розподілені з однаковою частотою за рівнями навчальних досягнень учнів.

Приклад 8.2. В двох школах міста психолог досліджував питання про функціонування психологічної служби. Для цього було проведення опитування вчителів, які давали відповіді в номінальній шкалі – «да», «ні».

З'ясувати, чи однаковий рівень функціонування психологічної служби в школах.

Результати опитування наведено у таблиці:

	Школа 1	Школа 2	
Відповідь «да»	$a = 25$	$b = 17$	$a + b = 42$
Відповідь «ні»	$c = 15$	$d = 19$	$c + d = 34$
	$a + c = 40$	$b + d = 36$	76

Розв'язування. Сформулюємо гіпотези:

H_0: Рівень функціонування в школах психологічної служби однаковий.

H_1: Рівень функціонування в школах психологічної служби суттєво відрізняється.

Оскільки данні подано таблицею зв'язаності ознак розміром 2×2, то застосуємо формулу для обрахунку значення критерій χ^2 з поправкою на неперервність:

$$\chi^2_{емп} = \frac{n\left(\left|ad - bc\right| - \frac{n}{2}\right)^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{76\left(\left|25 \cdot 19 - 17 \cdot 15\right| - \frac{76}{2}\right)^2}{42 \cdot 34 \cdot 40 \cdot 36} = 1{,}22.$$

За таблицею критичних точок розподілу χ^2 знаходимо критичне значення для заданого α-рівня та кількості ступенів свободи: $\nu = (2-1)(2-1) = 1$: $\chi^2_{крит}(\alpha = 0{,}05) = 3{,}8$.

Отже, $\chi^2_{емп} = 1{,}22 < \chi^2_{крит} = 3{,}8$, тому приймається гіпотеза H_0, а, отже, можна вважати, що організація психологічних служб знаходиться на одному рівні в обох школах.

Приклад 8.3. Психолог порівнює два емпіричні розподіли, в кожному з яких респондентів було обстежено за допомогою тесту інтелекту. Чи відрізняються між собою ці два розподіли?

Результати дослідження наведено у таблиці:

Рівень інтелекту	60	70	80	90	100	110	120	130	140	150	сума
Перша група	1	8	23	30	38	12	7	4	1	0	124
Друга група	0	0	1	11	18	14	3	4	1	1	53

Розв'язування. Сформулюємо гіпотези:

H_0: Емпіричні розподіли в обох групах однакові.

H_1: Емпіричні розподіли в обох групах різні.

Рівень інте-лекту	f_i	g_i	$n_2 f_i$	$n_1 g_i$	$n_2 f_i - n_1 g_i$	$(n_2 f_i - n_1 g_i)^2$	$f_i + g_i$	$\frac{(n_2 f_i - n_1 g_i)^2}{f_i + g_i}$
60	1	0	53	0	53	2809	1	2809
70	8	0	265	0	424	179776	8	22472
80	23	1	1219	124	1095	1199025	24	49959,38
90	30	11	1590	1364	226	51076	41	1245,76
100	38	18	2014	2232	-218	47524	56	848,64
110	12	14	636	1736	-1100	1210000	26	46538,46
120	7	3	371	372	-1	1	10	0,1
130	4	4	212	496	-284	80656	8	10082
140	1	1	53	124	-71	5041	2	2520,5
150	0	1	0	124	-124	15376	1	15376
	124	53				2681733	177	151851,8

Оскільки аналізуються вибірки різного обсягу то застосуємо відповідну формулу для обрахунку значення критерій χ^2:

$$\chi^2_{емп} = \frac{1}{n_1 n_2} \sum_{i=1}^{k} \frac{\left(n_2 f_i \ - n_1 g_i\right)^2}{f_i \ + g_i} = \frac{151851{,}8}{124 \cdot 53} = 23{,}01 .$$

За таблицею критичних точок розподілу χ^2 знаходимо критичне значення для заданого α-рівня та кількості ступенів свободи: $\nu = k - 1 = 10 - 1 = 9$: $\chi^2_{крит}(\alpha = 0{,}01) = 21{,}7$.

Отже, $\chi^2_{емп} = 23{,}01 > \chi^2_{крит} = 21{,}7$, тому гіпотеза H_0 відхиляється, а, отже, розподіл рівнів інтелекту в двох вибірках значимо відрізняються.

8.2. λ-КРИТЕРІЙ КОЛМОГОРОВА-СМІРНОВА

λ-критерій Колмогорова-Смірнова призначений для знаходження точки, в якій накопичене відхилення одного розподілу від другого є максимальним (порівнюватись можуть емпіричний розподіл з теоретичним, або два емпіричні), і оцінки статистичної достовірності цього відхилення. Накопиченим відхиленням двох розподілів називається тому, що в цьому методі виробляється послідовне порівняння частот попадання в першу градацію, потім у першу і другу, далі в першу, другу і третю і т.д.

Якщо різниця між двома розподілами статистично достовірна, то на деякому кроці накопичене відхилення перевищить критичне значення.

Обмеження застосування:

- при зіставленні двох емпіричних розподілів об'єм обох вибірок – $n_1, n_2 \geq 50$; при зіставлені емпіричного розподілу з теоретичним – $n \geq 5$;
- класи інтервалів мають бути упорядковані за зростанням або спаданням певної ознаки.

Формулюється основна та альтернативна гіпотези:

H_0: Різниці між двома розподілами недостовірні (відносно максимального накопиченого відхилення між ними).

H_1: Різниці між двома розподілами достовірні (відносно максимального накопиченого відхилення між ними).

Алгоритм λ-критерію Колмогорова-Смірнова для двох емпіричних розподілів по накопиченому відхиленню полягає в наступному:

1. Обрахувати накопичення відхилень за допомогою рекурентної формули:

$$d_j = d_{j-1} + \left(\frac{n_{j1}}{n_1} - \frac{n_{j2}}{n_2} \right), \quad j = 1, \ldots, k,$$

де k – кількість класових інтервалів, n_1 та n_2 – обсяги першої та другої вибірок відповідно, n_{j1} та n_{j2}, $j = 1, \ldots, k$, – кількість досліджуваних, що володіють ознакою на j-му рівні (потрапляють в j-й інтервальний клас) у першій та другій вибірці відповідно; $d_0 = 0$.

2. Знайти $d_m = \max\limits_{1 \le j \le k} \left| d_j \right|$ – найбільше накопичене відхилення.

3. Обрахувати значення критерію λ:

$$\lambda_{емп} = d_m \sqrt{\frac{n_1 n_2}{n_1 + n_2}}.$$

4. Визначити, якому рівню статистичної значущості відповідає $\lambda_{емп}$. Для стандартних рівнів значимості $\alpha = 0{,}05$ критичне значення $\lambda_{крит} = 1{,}36$ та $\alpha = 0{,}01$ критичне значення $\lambda_{крит} = 1{,}63$. Якщо $\lambda_{емп} \le \lambda_{крит}$, то відмінності між розподілами мають випадковий характер.

Алгоритм λ-критерію Колмогорова-Смірнова для порівняння емпіричного розподілу з теоретичним полягає в наступному:

1. Обчислити відносну частоту емпіричного розподілу: $\omega_j = \dfrac{n_j}{n}$, $j = 1,...,k$, де n_j – кількість досліджуваних, що володіють ознакою на j-му рівні (потрапляють в j-й інтервальний клас); n – обсяг вибірки.
2. Обрахувати накопичення відхилень за допомогою рекурентної формули:

$$d_j = d_{j-1} + \left(\omega_j - g_j \right),\ j = 1,...,k,$$

де g_j, $j = 1,...,k$, – теоретична частота.

3. Знайти $d_m = \max\limits_{1 \le j \le k} \left| d_j \right|$ – найбільше накопичене відхилення.

4. Знайти $d_{крит}$ за формулою:

$$d_{крит} = \begin{cases} \dfrac{1{,}36}{\sqrt{n}}, \text{ для } p \le 0{,}05, \\ \dfrac{1{,}63}{\sqrt{n}}, \text{ для } p \le 0{,}01. \end{cases}$$

5. Якщо $d_m < d_{крит}$, то емпіричний розподіл не відрізняється від теоретичного.

<u>Приклад 8.4.</u> Психолог дитячого садочку вирішив порівняти емоційний стан дітей двох однакових вікових груп, застосувавши тест Люшера. Після отримання даних з'явилася необхідність порівняти два отриманих розподіли.

Результати дослідження наведено у таблиці:

Позиція чорного кольору		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Кількість дітей, що поставили чорний на відповідну позицію	***1 група***	2	4	3	3	5	3	5	8	9	8
	2 група	3	2	3	1	4	9	8	9	6	5

Розв'язування. Сформулюємо гіпотези:

H_0: Відмінності між розподілами обох груп недостовірні.

H_1: Відмінності між розподілами обох груп достовірні.

Обрахуємо накопичення відхилень:

Класові інтервали	1 група	2 група	$\frac{n_{j1}}{n_1}$	$\frac{n_{j2}}{n_2}$	$\frac{n_{j1}}{n_1}-\frac{n_{j2}}{n_2}$	d_j
1	2	3	0,04	0,06	-0,02	-0,02
2	4	2	0,08	0,04	0,04	0,02
3	3	3	0,06	0,06	0	0,02
4	3	1	0,06	0,02	0,04	0,06
5	5	4	0,1	0,08	0,02	0,08
6	3	9	0,06	0,18	-0,12	-0,04
7	5	8	0,1	0,16	-0,06	-0,1
8	8	9	0,16	0,18	-0,02	-0,12
9	9	6	0,18	0,12	0,06	-0,06
10	8	5	0,16	0,1	0,06	0
Сума	50	50	1	1		

$d_m = 0{,}12$ – найбільше накопичене відхилення.

Обрахуємо значення критерію λ:

$$\lambda_{емп} = d_m \sqrt{\frac{n_1 n_2}{n_1 + n_2}} = 0{,}12\sqrt{\frac{50^2}{100}} = 0{,}6 .$$

Оскільки $\lambda_{емп} = 0{,}6 < \lambda_{крит} = 1{,}36$ для рівня значимості $\alpha = 0{,}05$, то приймається гіпотеза H_0, а, отже, відмінності між розподілами обох груп носять випадковий характер.

<u>Приклад 8.5.</u> У вибірці учнів 11 класів проводилося тестування з математики. Чи можна стверджувати, що розподіл результатів тестування відрізняється від рівномірного розподілу.

Результати тестування наведено у таблиці:

Частка правильних відповідей, %	**Кількість учасників**
0-20	4
21-40	15
41-60	18
61-80	7
81-10	1
сума	45

Розв'язування. Сформулюємо гіпотези:

H_0: Емпіричний розподіл результатів тестування з математики не відрізняється від рівномірного.

H_1: Емпіричний розподіл результатів тестування з математики відрізняється від рівномірного.

Теоретичні частоти для рівномірного розподілу обчислюється за формулою: $g_j = \frac{1}{k}$, $j = 1,...,k$.

Обчислюємо відносну частоту емпіричного розподілу та теоретичну частоту:

Класові інтервали	n_j	ω_j	g_j	$\omega_j - g_j$	d_j
1	4	0,089	0,2	-0,111	-0,111
2	15	0,333	0,2	0,133	0,022
3	15	0,333	0,2	0,133	0,155
4	10	0,222	0,2	0,022	0,177
5	1	0,022	0,2	-0,178	-0,001
сума	45	1			

$d_m = 0{,}177$ – найбільше накопичене відхилення.

Знаходимо $d_{крит}$:

$$d_{крит} = \begin{cases} \dfrac{1{,}36}{\sqrt{45}} = 0{,}203, \text{ для } p \le 0{,}05, \\ \dfrac{1{,}63}{\sqrt{n}} = 0{,}243, \text{ для } p \le 0{,}01. \end{cases}$$

Оскільки $d_m = 0{,}177 < d_{крит} = 0{,}203$ при $p \le 0{,}05$, то приймається гіпотеза H_0, а, отже, емпіричний розподіл результатів тестування з математики не відрізняється від рівномірного.

9. ЕЛЕМЕНТИ РЕГРЕСІЙНОГО АНАЛІЗУ

Регресійний аналіз – статистичний метод встановлення залежності між незалежними і залежними змінними.

Регресійна модель – це функція незалежної величини та коефіцієнтів з включеними випадковими змінними.

Регресійний аналіз тісно пов'язаний з кореляційним аналізом. В кореляційному аналізі досліджується напрямок та міцність зв'язку між незалежними змінними. В регресійному аналізі досліджується форма залежності (модель зв'язку, вираженої у функції регресії) між незалежними змінними. Фактично два методи вивчають однаковий взаємозв'язок, але з різних сторін, доповнюючи один одного.

Регресійний аналіз використовується у випадку, якщо відношення між змінними можуть бути виражені кількісно у вигляді деякої комбінації цих змінних. Отримана комбінація використовується для передбачення значення, що може приймати залежна змінна, яка обчислюється на заданому наборі значень незалежних змінних. У найпростішому випадку для цього використовується лінійна регресія.

Необхідно зазначити, що серед регресійних моделей виділяють:

- однопараметричні моделі (залежність від однієї змінної);
- багатопараметричні моделі (залежність від декількох змінних);
- лінійні моделі відносно незалежних змінних;
- моделі нелінійні за змінними та нелінійні за параметрами.

Розглянемо основні етапи регресійного аналізу.

Перший етап. Припущення. На цьому етапі відбувається вибір форми зв'язку між змінними (моделі).

Другий етап. Параметризація – відбувається оцінка значень параметрів у вибраній формулі статистичного зв'язку. Форма зв'язку (функція) – лінійна, нелінійна.

Третій етап. Перевірка надійності отриманих оцінок. На цьому етапі виконуються такі тести: F-тест (перевірка статистичної значущості вибраної форми зв'язку), *t*-тест (перевірка статистичної значущості знайдених числових значень параметра).

Одне із важливих питань психології є встановлення зв'язку між змінними. Вибір типу регресійної моделі є нетривіальним завданням. Для встановлення форми зв'язку, перш за все, необхідно побудувати діаграму розсіювання, за якою можна зробити припущення про лінію, яка максимально наближається до всіх точок діаграми розсіювання.

Це дає можливість визначити наявність чи відсутність залежності між досліджуваними величинами, а також зробити певні припущення про тип залежності.

9.1. ЛІНІЙНІ ОДНОФАКТОРНІ МОДЕЛІ

Найпростішим для аналізу і найбільш дослідженим є випадок лінійної кореляційної залежності між двома змінними *X* та *Y*. Наявність лінійного зв'язку можна перевірити, розрахувавши коефіцієнт парної кореляції Пірсона.

Розглянемо випадок лінійної регресії: $y = a + bx$ (x – умовно незалежна змінна; y – умовно залежна змінна), тобто коли близькою до точок діаграми буде саме пряма. Коефіцієнти a і b називають коефіцієнтами рівняння регресії, а саму лінію – лінією регресії.

Для визначення лінії регресії використовують метод найменших квадратів. Для відображення того факту, що кожне індивідуальне значення y_i відхиляється від відповідного умовного y в модель вводять випадковий доданок u_i: $y_i = a + bx_i + u_i$.

Таким чином регресійне рівняння набуває вигляду: $y = ax + b + u$.

Мірою якості знайдених оцінок є $\sum_{i=1}^{n}(y_i - a - bx_i)^2 = \sum_{i=1}^{n}(u_i)^2$.

Метод найменших квадратів передбачає, що сума квадратів відхилень буде найменшою, тобто

$$\min\sum_{i=1}^{n}(u_i)^2 = \min\sum_{i=1}^{n}(y_i - a - bx_i)^2 .$$

Розглянувши функцію $z=\sum_{i=1}^{n}\left(y_i-a-bx_i\right)^2$ та дослідивши її за допомогою частинних похідних на мінімум, одержимо

$$\begin{cases} a=\overline{y}-b\overline{x}, \\ b=\dfrac{\overline{xy}-\overline{x}\,\overline{y}}{\overline{x^2}-\left(\overline{x}\right)^2}, \end{cases}$$

де $\overline{x}$ – середнє значення по x, $\overline{y}$ – середнє значення по y, $\overline{xy}$ – середнє значення по xy.

Параметр b називається коефіцієнтом регресії, його величина показує на скільки в середньому змінюється y зі змінною x на одиницю.

Рівняння регресії завжди доповнюється показником тісноти зв’язку. При використанні лінійної регресії, як такий показник виступає лінійний коефіцієнт кореляції:

$$r_{xy}=b\frac{S_x}{S_y}.$$

Після того, як знайдено рівняння лінійної регресії, проводиться оцінка значимості як рівняння загалом, так і окремих його параметрів.

Перевірити значимість рівняння регресії означає встановити, чи відповідає аналітична модель, що виражає залежність між змінними, експериментальним даним, і чи достатньо включених до рівняння пояснюючих змінних (однієї або декількох) для описання залежної змінної.

Оцінка значущості рівняння регресії загалом проводиться на основі F-критерію Фішера. Емпіричне значення критерію Фішера обраховують за такою формулою:

$$F_{емп}=\frac{r_{xy}{}^2}{1-r_{xy}{}^2}\left(n-2\right).$$

Знаходять за таблицями розподілу Фішера критичне значення $F_{крит}$ для вибраного рівня значимості α і ступенів свободи $v_1=1$, $v_2=n-2$.

Якщо $F_{емп}<F_{крит}$, тоді побудована регресійна модель вважається неадекватною дійсній, тобто не є значимою. В цьому випадку потрібно знаходити інший тип моделі, для описання зв’язку між змінними x та y.

Якщо $F_{емп} > F_{крит}$, тоді вважається, що побудована модель адекватна дійсній, тобто є значимою.

У парній лінійній регресії оцінюється значимість не тільки рівняння загалом, а й окремих його параметрів.

t-розподіл Стьюдента застосовується для перевірки суттєвості коефіцієнта регресії. Спочатку визначається стандартна помилка коефіцієнта регресії S_b за формулою:

$$S_b = \frac{\sqrt{S_{ост}^2}}{S_x \sqrt{n}},$$

де $S_{ост}^2 = \frac{\sum \left(y_{теор} - y_i \right)^2}{n-2}$ – вибіркова остаточна дисперсія.

Для оцінки суттєвості коефіцієнта регресії його величина порівнюється з його стандартною помилкою, тобто визначається емпіричне значення t-критерію Стьюдента:

$$t_b = \frac{b}{S_b}.$$

Довірчий інтервал для коефіцієнта регресії визначається за такою формулою:

$$b \pm t_{табл} S_b.$$

Знак коефіцієнта регресії вказує на зростання результативної ознаки y при збільшенні x $(b > 0)$, зменшення результативної ознаки y при збільшенні x $(b < 0)$ або її незалежність від незалежної змінної $(b = 0)$.

Стандартна помилка параметра a визначається за формулою:

$$S_a = \frac{\sqrt{S_{ост}^2 \sum x^2}}{S_x \cdot n}.$$

Процедура оцінювання суттєвості даного параметра не відрізняється від розглянутої вище для коефіцієнта регресії. Обчислюється t-критерій: $t_a = \frac{a}{S_a}$, його величина порівнюється з табличним значенням при $n-2$ ступенях свободи.

Значимість лінійного коефіцієнта кореляції перевіряється з урахуванням величини помилки коефіцієнта кореляції S_r: $S_r = \sqrt{\frac{1-r^2}{n-2}}$.

Фактичне значення t-критерію Стьюдента визначається як $t_r = \frac{r}{S_r}$, яке порівнюється з табличним значенням при $n-2$ ступенях свободи.

<u>Приклад 9.1.</u> За даними проведеного експерименту відомий зв’язок між двома змінними X та Y.

Одержані дані наведено в таблиці:

№	1	2	3	4	5	6	7	8	9	10	11	12
X	107	109	110	113	120	122	123	128	136	140	145	150
Y	102	105	108	110	115	117	119	125	132	130	141	144

Побудувати лінійне регресійне рівняння та зробити висновок про якість побудованої моделі.

Складемо таблицю первинних даних та проміжних розрахунків

№	x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$	$\tilde{y}_i$	$u_i^2 = (y_i - \tilde{y}_i)^2$
1	107	102	11449	10404	10914	103,70	2,89
2	109	105	11881	11025	11445	105,56	0,31
3	110	108	12100	11664	11880	106,49	2,28
4	113	110	12769	12100	12430	109,28	0,52
5	120	115	14400	13225	13800	115,79	0,62
6	122	117	14884	13689	14274	117,65	0,42
7	123	119	15129	14161	14637	118,58	0,18
8	128	125	16384	15625	16000	123,23	3,13
9	136	132	18496	17424	17952	130,67	1,77
10	140	130	19600	16900	18200	134,39	19,27
11	145	141	21025	19881	20445	139,04	3,84
12	150	144	22500	20736	21600	143,69	0,10
сума	1503	1448	190617	176834	183577	1448,07	35,34
середнє	125,25	120,67	15884,75	14736,17	15298,08		

Знайдемо коефіцієнти лінійної моделі регресі розв’язавши систему рівнянь:

$$\begin{cases} a = \overline{y} - b\overline{x}, \\ b = \dfrac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2}. \end{cases}$$

Обрахуємо коефіцієнт b:

$$b = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2} = \frac{15298{,}08 - 125{,}25 \cdot 120{,}67}{15884{,}75 - \left(125{,}25\right)^2} = 0{,}93,$$

та підставимо одержані значення у перше рівняння системи:

$$a = \overline{y} - b\overline{x} = 120{,}67 - 0{,}93 \cdot 125{,}25 = 4{,}19.$$

Тоді лінійне регресійне рівняння буде мати вигляд: $\tilde{y} = 4{,}19 + 0{,}93x$.

Знайдемо значення змінної $\tilde{y}$ для кожного заданого x:

$$\tilde{y}_1 = 4{,}19 + 0{,}93 \cdot 107 = 103{,}70,$$

$$\tilde{y}_2 = 4{,}19 + 0{,}93 \cdot 109 = 105{,}56,$$

$$\tilde{y}_3 = 4{,}19 + 0{,}93 \cdot 110 = 106{,}49,$$

$$\tilde{y}_4 = 4{,}19 + 0{,}93 \cdot 113 = 109{,}28,$$

$$\tilde{y}_5 = 4{,}19 + 0{,}93 \cdot 120 = 115{,}79,$$

$$\tilde{y}_6 = 4{,}19 + 0{,}93 \cdot 122 = 117{,}65,$$

$$\tilde{y}_7 = 4{,}19 + 0{,}93 \cdot 123 = 118{,}58,$$

$$\tilde{y}_8 = 4{,}19 + 0{,}93 \cdot 128 = 123{,}23,$$

$$\tilde{y}_9 = 4{,}19 + 0{,}93 \cdot 136 = 130{,}67,$$

$$\tilde{y}_{10} = 4{,}19 + 0{,}93 \cdot 140 = 134{,}39,$$

$$\tilde{y}_{11} = 4{,}19 + 0{,}93 \cdot 145 = 139{,}04,$$

$$\tilde{y}_{12} = 4{,}19 + 0{,}93 \cdot 150 = 143{,}69.$$

Для визначення тісноти зв’язку обрахуємо лінійний коефіцієнт кореляції:

$$S_x = 14,67,$$

$$S_y = 13,85$$

$$r_{xy} = b\frac{S_x}{S_y} = 0,93 \cdot \frac{14,67}{13,85} = 0,985.$$

Близькість коефіцієнта кореляції до 1 вказує на тісний лінійний зв'язок між ознаками.

Оцінимо якість рівняння за допомогою *F*-критерію Фішера. Обрахуємо емпіричне значення критерію Фішера:

$$F_{емп} = \frac{{r_{xy}}^2}{1 - {r_{xy}}^2}(n-2) = \frac{0,985^2}{1 - 0,985^2} \cdot 10 = 325,85.$$

Знаходимо за таблицями розподілу Фішера критичне значення $F_{крит}$ для вибраного рівня значимості $\alpha = 0,05$ і ступенів свободи $\nu_1 = 1$, $\nu_2 = 10$: $F_{крит} = 4,96$.

Оскільки $F_{емп} > F_{крит}$, то побудована модель адекватна дійсній, тобто є значимою.

За допомогою *t*-розподілу Стьюдента перевіряємо суттєвість коефіцієнта регресії.

Спочатку визначимо стандартну помилку коефіцієнта регресії S_b:

$$S^2_{ост} = \frac{\sum\left(y_{теор} - y_i\right)^2}{n-2} = \frac{35,34}{10} = 3,534,$$

$$S_b = \frac{\sqrt{S^2_{ост}}}{S_x\sqrt{n}} = \frac{\sqrt{3,525}}{14,67\sqrt{12}} = 0,037.$$

Визначимо емпіричне значення *t*-критерію Стьюдента:

$$t_b = \frac{b}{S_b} = \frac{0,93}{0,037} = 25,14.$$

Оскільки критичне значення *t*-критерію Стьюдента при $\alpha = 0,05$ і числу степенів свободи $\nu = n - 2 = 10$ рівне $t_{крит} = 2,23$,то

$$t_b = 25,14 > t_{крит} = 2,23,$$

а, отже, на рівні значимості $\alpha = 0{,}05$ визнаємо статистичну значимість коефіцієнта регресії.

Знайдемо довірчий інтервал для коефіцієнта регресії:

$$\left(b - t_{табл} S_b ; b + t_{табл} S_b\right) = \left(0{,}93 - 2{,}23 \cdot 0{,}037 ; 0{,}93 + 2{,}23 \cdot 0{,}037\right) = \left(0{,}847 ; 1{,}013\right).$$

Визначимо стандартну помилку параметра a:

$$S_a = \frac{\sqrt{S_{ост}^2 \sum x^2}}{S_x \cdot n} = \frac{\sqrt{3{,}534 \cdot 190617}}{14{,}67 \cdot 12} = 4{,}66.$$

Обчислюємо t-критерій: $t_a = \dfrac{a}{S_a} = \dfrac{4{,}19}{4{,}66} = 0{,}899$, та порівняємо його із табличним значенням при $\alpha = 0{,}05$ та $n-2$ ступенях свободи: $t_a = 0{,}899 < t_{крит} = 2{,}23$, а, отже, коефіцієнт a незначно відрізняється від нуля.

Перевіримо значимість лінійного коефіцієнта кореляції:

$$S_r = \sqrt{\frac{1 - r^2}{n - 2}} = \sqrt{\frac{1 - \left(0{,}985\right)^2}{10}} = 0{,}017,$$

$$t_r = \frac{0{,}985}{0{,}017} = 57{,}94,$$

та порівняємо його із табличним значенням при $\alpha = 0{,}05$ та $n-2$ ступенях свободи: $t_r = 57{,}94 > t_{крит} = 2{,}23$, а, отже визнаємо статистичну значимість коефіцієнта тісноти зв’язку.

9.2. НЕЛІНІЙНІ МОДЕЛІ ПАРНОЇ РЕГРЕСІЇ

1. Рівняння гіперболи: $y = a + \frac{b}{x}$.

   За допомогою заміни $\frac{1}{x} = t$ переходимо до лінійної моделі $y = a + bt$.

2. Рівняння параболи: $y = a_0 + a_1 x + a_2 x^2$.

   Коефіцієнти a_0, a_1, a_2 обраховуються за формулою:

$$\begin{cases} \overline{yx^2} - a_0 \overline{x^2} - a_1 \overline{x^3} - a_2 \overline{x^4} = 0, \\ \overline{xy} - a_0 \overline{x} - a_1 \overline{x^2} - a_2 \overline{x^3} = 0, \\ \overline{y} - a_0 - a_1 \overline{x} - a_2 \overline{x^2} = 0. \end{cases}$$

3. Рівняння логарифмічної функції: $y = a + b\lg x$.

   За допомогою заміни $\lg x = t$ переходимо до лінійної моделі $y = a + bt$.

4. Степенева функція: $y = ax^b$.

   За допомогою логарифмування переходимо до рівняння:

$$\lg y = \lg a + b\lg x,$$

$$\tilde{y} = \tilde{a} + b\tilde{x},$$

де $\tilde{y} = \lg y$, $\tilde{a} = \lg a$, $\tilde{x} = \lg x$.

Для оцінки тісноти зв'язку нелінійної регресії знаходять індекс кореляції:

$$\rho_{xy} = \sqrt{\frac{\sum (\tilde{y}_i - \overline{y})^2}{\sum (y_i - \overline{y})^2}}.$$

ТАБЛИЦІ КРИТИЧНИХ ЗНАЧЕНЬ СТАТИСТИЧНИХ КРИТЕРІЇВ

Таблиця 1

Критичні значення коефіцієнта кореляції r_{xy} Пірсона

(r_s Спірмена для $n > 10$) для перевірки неспрямованих альтернатив

Об'єм вибірки, n	α			
	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

Продовження таблиці 1

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	0,32	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	0,38	0,474
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

Продовження таблиці 1

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	0,33	0,414
61	0,213	0,252	0,327	0,411
62	0,211	0,25	0,325	0,408
63	0,209	0,248	0,322	0,405
64	0,207	0,246	0,32	0,402
65	0,206	0,244	0,317	0,399
66	0,204	0,242	0,315	0,396
67	0,203	0,24	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	0,22	0,286	0,361
90	0,174	0,207	0,27	0,341
100	0,165	0,197	0,256	0,324
110	0,158	0,187	0,245	0,31

Продовження таблиці 1

120	0,151	0,179	0,234	0,297
130	0,145	0,172	0,225	0,285
140	0,14	0,166	0,217	0,275
150	0,135	0,16	0,21	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
350	0,088	0,105	0,138	0,175
400	0,082	0,098	0,129	0,164
450	0,078	0,092	0,121	0,155
500	0,074	0,088	0,115	0,147
600	0,067	0,08	0,105	0,134

Таблиця 2

Критичні точки розподілу χ^2

Число ступе-нів свобо-ди, ν	α							
	0,01	**0,025**	**0,05**	**0,1**	**0,9**	**0,95**	**0,975**	**0,99**
1	6,635	5,024	3,841	2,706	0,016	0,004	0,001	0,0002
2	9,210	7,378	5,991	4,605	0,211	0,103	0,051	0,020
3	11,345	9,348	7,815	6,251	0,584	0,352	0,216	0,115
4	13,277	11,143	9,488	7,779	1,064	0,711	0,484	0,297
5	15,086	12,833	11,070	9,236	1,610	1,145	0,831	0,554
6	16,812	14,449	12,592	10,645	2,204	1,635	1,237	0,872
7	18,475	16,013	14,067	12,017	2,833	2,167	1,690	1,239
8	20,090	17,535	15,507	13,362	3,490	2,733	2,180	1,646
9	21,666	19,023	16,919	14,684	4,168	3,325	2,700	2,088
10	23,209	20,483	18,307	15,987	4,865	3,940	3,247	2,558
11	24,725	21,920	19,675	17,275	5,578	4,575	3,816	3,053
12	26,217	23,337	21,026	18,549	6,304	5,226	4,404	3,571
13	27,688	24,736	22,362	19,812	7,042	5,892	5,009	4,107
14	29,141	26,119	23,685	21,064	7,790	6,571	5,629	4,660
15	30,578	27,488	24,996	22,307	8,547	7,261	6,262	5,229
16	32,000	28,845	26,296	23,542	9,312	7,962	6,908	5,812
17	33,409	30,191	27,587	24,769	10,085	8,672	7,564	6,408
18	34,805	31,526	28,869	25,989	10,865	9,390	8,231	7,015
19	36,191	32,852	30,144	27,204	11,651	10,117	8,907	7,633
20	37,566	34,170	31,410	28,412	12,443	10,851	9,591	8,260

Продовження таблиці 2

21	38,932	35,479	32,671	29,615	13,240	11,591	10,283	8,897
22	40,289	36,781	33,924	30,813	14,041	12,338	10,982	9,542
23	41,638	38,076	35,172	32,007	14,848	13,091	11,689	10,196
24	42,980	39,364	36,415	33,196	15,659	13,848	12,401	10,856
25	44,314	40,646	37,652	34,382	16,473	14,611	13,120	11,524
26	45,642	41,923	38,885	35,563	17,292	15,379	13,844	12,198
27	46,963	43,195	40,113	36,741	18,114	16,151	14,573	12,879
28	48,278	44,461	41,337	37,916	18,939	16,928	15,308	13,565
29	49,588	45,722	42,557	39,087	19,768	17,708	16,047	14,256
30	50,892	46,979	43,773	40,256	20,599	18,493	16,791	14,953
31	52,191	48,232	44,985	41,422	21,434	19,281	17,539	15,655
32	53,486	49,480	46,194	42,585	22,271	20,072	18,291	16,362
33	54,776	50,725	47,400	43,745	23,110	20,867	19,047	17,074
34	56,061	51,966	48,602	44,903	23,952	21,664	19,806	17,789
35	57,342	53,203	49,802	46,059	24,797	22,465	20,569	18,509
36	58,619	54,437	50,998	47,212	25,643	23,269	21,336	19,233
37	59,893	55,668	52,192	48,363	26,492	24,075	22,106	19,960
38	61,162	56,896	53,384	49,513	27,343	24,884	22,878	20,691
39	62,428	58,12	54,572	50,660	28,196	25,695	23,654	21,426
40	63,691	59,342	55,758	51,805	29,051	26,509	24,433	22,164
41	64,950	60,561	56,942	52,949	29,907	27,326	25,215	22,906
42	66,206	61,777	58,124	54,090	30,765	28,144	25,999	23,650
43	67,459	62,99	59,304	55,230	31,625	28,965	26,785	24,398
44	68,710	64,201	60,481	56,369	32,487	29,787	27,575	25,148
45	69,957	65,410	61,656	57,505	33,350	30,612	28,366	25,901
46	71,201	66,617	62,830	58,641	34,215	31,439	29,160	26,657

Продовження таблиці 2

47	72,443	67,821	64,001	59,774	35,081	32,268	29,956	27,416
48	73,683	69,023	65,171	60,907	35,949	33,098	30,755	28,177
49	74,919	70,222	66,339	62,038	36,818	33,930	31,555	28,941
50	76,154	71,420	67,505	63,167	37,689	34,764	32,357	29,707
51	77,386	72,616	68,669	64,295	38,560	35,600	33,162	30,475
52	78,616	73,810	69,832	65,422	39,433	36,437	33,968	31,246
53	79,843	75,002	70,993	66,548	40,308	37,276	34,776	32,018
54	81,069	76,192	72,153	67,673	41,183	38,116	35,586	32,793
55	82,292	77,380	73,311	68,796	42,060	38,958	36,398	33,570
56	83,513	78,567	74,468	69,919	42,937	39,801	37,212	34,350
57	84,733	79,752	75,624	71,040	43,816	40,646	38,027	35,131
58	85,950	80,936	76,778	72,160	44,696	41,492	38,844	35,913
59	87,166	82,117	77,931	73,279	45,577	42,339	39,662	36,698
60	88,379	83,298	79,082	74,397	46,459	43,188	40,482	37,485
61	89,591	84,476	80,232	75,514	47,342	44,038	41,303	38,273
62	90,802	85,654	81,381	76,630	48,226	44,889	42,126	39,063
63	92,010	86,830	82,529	77,745	49,111	45,741	42,950	39,855
64	93,217	88,004	83,675	78,860	49,996	46,595	43,776	40,649
65	94,422	89,177	84,821	79,973	50,883	47,450	44,603	41,444
66	95,626	90,349	85,965	81,085	51,770	48,305	45,431	42,240
67	96,828	91,519	87,108	82,197	52,659	49,162	46,261	43,038
68	98,028	92,689	88,250	83,308	53,548	50,020	47,092	43,838
69	99,228	93,856	89,391	84,418	54,438	50,879	47,924	44,639
70	100,425	95,023	90,531	85,527	55,329	51,739	48,758	45,442
71	101,621	96,189	91,670	86,635	56,221	52,600	49,592	46,246
72	102,816	97,353	92,808	87,743	57,113	53,462	50,428	47,051

Продовження таблиці 2

73	104,010	98,516	93,945	88,850	58,006	54,325	51,265	47,858
74	105,202	99,678	95,081	89,956	58,900	55,189	52,103	48,666
75	106,393	100,839	96,217	91,061	59,795	56,054	52,942	49,475
76	107,583	101,999	97,351	92,166	60,690	56,920	53,782	50,286
77	108,771	103,158	98,484	93,270	61,586	57,786	54,623	51,097
78	109,958	104,316	99,617	94,374	62,483	58,654	55,466	51,910
79	111,144	105,473	100,749	95,476	63,380	59,522	56,309	52,725
80	112,329	106,629	101,879	96,578	64,278	60,391	57,153	53,540
90	124,116	118,136	113,145	107,565	73,291	69,126	65,647	61,754
100	135,807	129,561	124,342	118,498	82,358	77,929	74,222	70,065
110	147,414	140,917	135,480	129,385	91,471	86,792	82,867	78,458
120	158,950	152,211	146,567	140,233	100,624	95,705	91,573	86,923
130	170,423	163,453	157,610	151,045	109,811	104,662	100,331	95,451
140	181,840	174,648	168,613	161,827	119,029	113,659	109,137	104,034
150	193,208	185,800	179,581	172,581	128,275	122,692	117,985	112,668
200	249,445	241,058	233,994	226,021	174,835	168,279	162,728	156,432
250	304,940	295,689	287,882	279,050	221,806	214,392	208,098	200,939
300	359,906	349,874	341,395	331,789	269,068	260,878	253,912	245,972
350	414,474	403,723	394,626	384,306	316,550	307,648	300,064	291,406

Таблиця 3

Критичні точки розподілу t-Стьюдента

ν \ α	**0,10**	**0,05**	**0,01**	**0,001**
1	6,31	12,70	63,70	637,0
2	2,92	4,30	9,92	31,60
3	2,35	3,18	5,84	12,90
4	2,13	2,78	4,60	8,61
5	2,01	2,57	4,03	6,86
6	1,94	2,45	3,71	5,96
7	1,89	2,36	3,50	5,40
8	1,86	2,31	3,36	5,04
9	1,83	2,26	3,25	4,78
10	1,81	2,23	3,17	4,59
11	1,80	2,20	3,11	4,44
12	1,78	2,18	3,05	4,32
13	1,77	2,16	3,01	4,22
14	1,76	2,14	2,98	4,14
15	1,75	2,13	2,95	4,07
16	1,75	2,12	2,92	4,01
17	1,74	2,11	2,90	3,96
18	1,73	2,10	2,88	3,92
19	1,73	2,09	2,86	3,88
20	1,73	2,09	2,85	3,85
21	1,72	2,08	2,83	3,82
22	1,72	2,07	2,82	3,79
23	1,71	2,07	2,81	3,77
24	1,71	2,06	2,80	3,74

Продовження таблиці 3

25	1,71	2,06	2,79	3,72
26	1,71	2,06	2,78	3.71
27	1,71	2,05	2,77	3,69
28	1,70	2,05	2,76	3,66
29	1,70	2,05	2,76	3,66
30	1,70	2,04	2,75	3,65
40	1,68	2,02	2,70	3,55
60	1,67	2,00	2,66	3,46
120	1,66	1,98	2,62	3,37
∞	1,64	1,96	2,58	3,29

Таблиця 4

Критичні значення вибіркового коефіцієнта кореляціїрангів r_s Спірмена

n	α		**n**	α		**n**	α	
	0,05	**0,01**		**0,05**	**0,01**		**0,05**	**0,01**
5	0,94	-	**17**	0,48	0,62	**29**	0,37	0,48
6	0,85	-	**18**	0,47	0,60	**30**	0,36	0,47
7	0,78	0,94	**19**	0,46	0,58	**31**	0.36	0,46
8	0,72	0,88	**20**	0,45	0,57	**32**	0,36	0,45
9	0,68	0,83	**21**	0,44	0,56	**33**	0,34	0,45
10	0,64	0,79	**22**	0,43	0,54	**34**	0,34	0.44
11	0,61	0,76	**23**	0,42	0,53	**35**	0,33	0,43
12	0,58	0,73	**24**	0,41	0,52	**36**	0,33	0,43
13	0,56	0,70	**25**	0,49	0,51	**37**	0,33	0.43
14	0,54	0,68	**26**	0,39	0.50	**38**	0,32	0,41
15	0,52	0,66	**27**	0,38	0,49	**39**	0,32	0,41
16	0,50	0,64	**28**	0,38	0,48	**40**	0,31	0,40

Таблиця 5

Критичні точки розподілу F Фішера-Снедекора для
α = 0,05 та α = 0,01
(df_1 – число ступенів вільності більшої дисперсії «чисельник»,
df_2 – число ступенів вільності меншої дисперсії «знаменник»)

α = 0,05													
df_2 \ df_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
1	161	200	216	225	230	234	237	239	241	242	243	244	245
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41	19,42
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74	8,73
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91	5,89
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68	4,66
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57	3,55
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28	3,26
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07	3,05
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91	2,89
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79	2,76
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69	2,66
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60	2,58
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53	2,51
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48	2,45

α = 0,01													
df_2 / df_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
1	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056	6082	6106	6126
2	98,49	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,41	99,42	99,42
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05	26,98
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37	14,31
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,29	10,15	10,05	9,96	9,89	9,82
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72	7,66
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47	6,41
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67	5,61
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11	5,05
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71	4,65
11	9,65	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40	4,34
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16	4,10
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96	3,91
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80	3,75
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67	3,61

Таблиця 6

Критичні значення критерію Q Розенбаума

α = 0,05																
n	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**
11	6															
12	6	6														
13	6	6	6													
14	7	7	6	6												
15	7	7	6	6	6											
16	8	7	7	7	6	6										
17	7	7	7	7	7	7	7									
18	7	7	7	7	7	7	7	7								
19	7	7	7	7	7	7	7	7	7							
20	7	7	7	7	7	7	7	7	7	7						
21	8	7	7	7	7	7	7	7	7	7	7					
22	8	7	7	7	7	7	7	7	7	7	7	7				
23	8	8	7	7	7	7	7	7	7	7	7	7	7			
24	8	8	8	8	8	8	8	8	8	8	7	7	7	7		
25	8	8	8	8	8	8	8	8	8	7	7	7	7	7	7	
26	8	8	8	8	8	8	8	8	8	8	7	7	7	7	7	7

Продовження таблиці 6

α = 0,01																
n	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**
11	9															
12	9	9														
13	9	9	9													
14	9	9	9	9												
15	9	9	9	9	9											
16	9	9	9	9	9	9										
17	10	9	9	9	9	9	9									
18	10	10	9	9	9	9	9	9								
19	10	10	10	9	9	9	9	9	9							
20	10	10	10	10	9	9	9	9	9	9						
21	11	10	10	10	9	9	9	9	9	9	9					
22	11	11	10	10	10	9	9	9	9	9	9	9				
23	11	11	10	10	10	10	9	9	9	9	9	9	9			
24	12	11	11	10	10	10	10	9	9	9	9	9	9	9		
25	12	11	11	10	10	10	10	10	9	9	9	9	9	9	9	
26	12	12	11	11	10	10	10	10	10	9	9	9	9	9	9	9

Таблиця 7

Критичні значення критерію U Манна-Уітні для неспрямованих альтернатив (двобічна область)

α = 0,05																	
n	**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																		
n	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
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	24	27								
13	1	3	7	10	13	17	20	24	27	31	34							
14	1	4	7	11	15	18	22	26	30	34	38	42						
15	2	5	8	12	16	20	24	29	33	37	42	46	51					
16	2	5	9	13	18	22	27	31	36	41	45	50	55	60				
17	2	6	10	15	19	24	29	34	39	44	49	54	60	65	70			
18	2	6	11	16	21	26	31	37	42	47	53	58	64	70	75	81		
19	3	7	12	17	22	28	33	39	45	51	56	63	69	74	81	87	93	
20	3	8	13	18	24	30	36	42	48	54	60	67	73	79	86	92	99	105

Таблиця 8

Критичні значення критерію U Манна-Уітні для спрямованих альтернатив (однобічна область)

α = 0,05																			
n	**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

Продовження таблиці 8

α = 0,01																			
n	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
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

Таблиця 9

Критичні значення H-критерію Крускала-Уолліса

Об'єми вибірок			Значення *H*	*p*
n_1	n_2	n_3		
2	1	1	2,7000	0,500
2	2	1	3,6000	0,200
2	2	2	4,5714	0,067
3	1	1	3,2000	0,300
3	2	1	4,2857	0,100
			3,8571	0,133
3	2	2	5,3572	0,029
			4,7143	0,048
			4,5000	0,067
			4,4643	0,105
3	3	1	5,1429	0,043
			4,5714	0,100
			4,0000	0,129
3	3	2	6,2500	0,011
			5,3611	0,032
			5,1389	0,061
			4,5556	0,100
			4,2500	0,121
3	3	3	7,2000	0,004
			6,4889	0,011
			5,6889	0,029
			5,6000	0,050
			5,0667	0,086
			4,6222	0,100

Продовження таблиці 9

4	1	1	3,5714	0,200
4	2	1	4,8214	0,057
			4,5000	0,076
			4,0179	0,114
4	2	2	6,0000	0,014
			5,3333	0,033
			5,1250	0,052
			4,4583	0,100
			4,1667	0,105
4	3	1	5,8333'	0,021
			5,2083	0,050
			5,0000	0,057
			4,0556	0,093
			3,8889	0,129
4	3	2	6,4444	0,008
			6,3000	0,011
			5,4444	0,046
			5,4000	0,051
			4,5111	0,098
			4,4444	0,102
4	3	3	6,7455	0,010
			6,7091	0,013
			5,7909	0,046
			5,7273	0,050
			4,7091	0,092
			4,7000	0,101

Продовження таблиці 9

4	4	1	6,6667	0,010
			6,1667	0,022
			4,9667	0,048
			4,8667	0,054
			4,1667	0,082
			4,0667	0,102
4	4	2	7,0364	0,006
			6,8727	0,011
			5,4545	0,046
			5,2364	0,052
			4,5545	0,098
			4,4455	0,103
4	4	3	7,1439	0,010
			7,1364	0,011
			5,5985	0,049
			5,5758	0,051
			4,5455	0,099
			4,4773	0,102
4	4	4	7,6538	0,008
			7,5385	0,011
			5,6923	0,049
			5,6538	0,054
			4,6539	0,097
			4,5001	0,104
5	1	1	3,8571	0,143

Продовження таблиці 9

5	2	1	5,2500	0,036
			5,0000	0,048
			4,4500	0,071
			4,2000	0,095
			4,0500	0,119
5	2	2	6,5333	0,008
			6,1333	0,013
			5,1600	0,034
			5,0400	0,056
			4,3733	0,090
			4,2933	0,122
5	3	1	6,4000	0,012
			4,9600	0,048
			4,8711	0,052
			4,0178	0,095
			3,8400	0,123
5	3	2	6,9091	0,009
			6,8218	0,010
			5,2509	0,049
			5,1055	0,052
			4,6509	0,091
			4,4945	0,101
5	3	3	7,0788	0,009
			6,9818	0,011
			5,6485	0,049
			5,5152	0,051
			4,5333	0,097
			4,4121	0,109

Продовження таблиці 9

5	4	1	6,9545	0,008
			6,8400	0,011
			4,9855	0,044
			4,8600	0,056
			3,9873	0,098
			3,9600	0,102
5	4	2	7,2045	0,009
			7,1182	0,010
			5,2727	0,049
			5,2682	0,050
			4,5409	0,098
			4,5182	0,101
5	4	3	7,4449	0,010
			7,3949	0,011
			5,6564	0,049
			5,6308	0,050
			4,5487	0,099
			4,5231	0,103
5	4	4	7,7604	0,009
			7,7440	0,011
			5,6571	0,049
			5,6176	0,050
			4,6187	0,100
			4,5527	0,102

Продовження таблиці 9

5	5	1	7,3091	0,009
			6,8364	0,011
			5,1273	0,046
			4,9091	0,053
			4,1091	0,086
			4,0364	0,105
5	5	2	7,3385	0,010
			7,2692	0,010
			5,3385	0,047
			5,2462	0,051
			4,6231	0,097
			4,5077	0,100
5	5	3	7,5780	0,010
			7,5429	0,010
			5,7055	0,046
			5,6264	0,051
			4,5451	0,100
			4,5363	0,102
5	5	4	7,8229	0,010
			7,7914	0,010
			5,6657	0,049
			5,6429	0,050
			4,5229	0,099
			4,5200	0,101

Продовження таблиці 9

5	5	5	8,0000	0,009
			7,9800	0,010
			5,7800	0,049
			5,6600	0,051
			4,5600	0,100
			4,5000	0,102

Таблиця 10

Критичні значення критерію тенденцій S Джонкіра

α = 0,05									
кіль-кість груп, C	**Кількість респондентів у кожній групі N**								
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
3	10	17	24	33	42	53	64	76	88
4	14	26	38	51	66	82	100	118	138
5	20	34	51	71	92	115	140	166	194
6	26	44	67	93	121	151	184	219	256
α = 0,01									
кіль-кість груп, C	**Кількість респондентів у кожній групі N**								
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
3	–	23	32	45	59	74	90	106	124
4	20	34	50	71	92	115	140	167	193
5	26	48	72	99	129	162	197	234	274
6	34	62	94	130	170	213	260	309	361

Таблиця 11

Критичні значення критерію знаків G

n	**Рівень значущості**		**n**	**Рівень значущості**		**n**	**Рівень значущості**	
	α = 0,05	α = 0,01		α = 0,05	α = 0,01		α = 0,05	α = 0,01
5	0	-	**36**	12	10	**84**	33	30
6	0	-	**37**	13	10	**86**	34	31
7	0	0	**38**	13	11	**88**	35	32
8	1	0	**39**	13	11	**90**	36	33
9	1	0	**40**	14	12	**92**	37	34
10	1	0	**41**	14	12	**94**	38	35
11	2	1	**42**	15	13	**96**	39	36
12	2	1	**43**	15	13	**98**	40	37
13	3	1	**44**	16	13	**100**	41	37
14	3	2	**45**	16	14	**110**	45	42
15	3	2	**46**	16	14	**120**	50	46
16	4	2	**47**	17	15	**130**	55	51
17	4	3	**48**	17	15	**140**	59	55
18	5	3	**49**	18	15	**150**	64	60
19	5	4	**50**	18	16	**160**	69	64
20	5	4	**52**	19	17	**170**	73	69
21	6	4	**54**	20	18	**180**	78	73
22	6	5	**56**	21	18	**190**	83	78
23	7	5	**58**	22	19	**200**	87	83
24	7	5	**60**	23	20	**220**	97	92
25	7	6	**62**	24	21	**240**	106	101
26	8	6	**64**	24	22	**260**	116	110

Продовження таблиці 11

27	8	7	**66**	25	23	**280**	125	120
28	8	7	**68**	26	23	**300**	135	129
29	9	7	**70**	27	24			
30	10	8	**72**	28	25			
31	10	8	**74**	29	26			
32	10	8	**76**	30	27			
33	11	9	**78**	31	28			
34	11	9	**80**	32	29			
35	12	10	**82**	33	30			

Таблиця 12

Критичні значення критерію Т Вілкоксона для двох залежних вибірок (однобічна область)

n	Рівень значущості для однобічної області		n	Рівень значущості для однобічної області	
	α = 0,05	α = 0,01		α = 0,05	α = 0,01
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

Таблиця 13

Критичні значення критерію χ_r^2 Фрідмана

Кількість умов (вибірок) *c = 3*							
n = 2		***n = 3***		***n = 4***		***n = 5***	
χ_r^2	α	χ_r^2	α	χ_r^2	α	χ_r^2	α
0	1.000 0,833	0,000	1.000 0,944	0,0	1,000 0,931	0,0	1,000
1	0,500 0,167	0.667	0,528 0,361	0,5	0,653 0,431	0,4	0,954
3		2,000	0,194 0,028	1,5	0,273 0,125	1.2	0,691
4		2,667		2,0	0,069 0,042	1.6	0,522
		4,667		3,5	0,0046	2,8	0,367
		6,000		4,5		3,6	0,182
				6,0		4,8	0,124
				6,5		5,2	0,093
				8.0		6.4	0,039
						7.6	0,024
						8.4	0,0085
						10,0	0,00077

Продовження таблиці 13

n = 6		**n = 7**		**n = 8**		**n = 9**	
χ_r^2	α	χ_r^2	α	χ_r^2	α	χ_r^2	α
0,00	1,00000	0,000	1,000000	0,00	1,000	0,000	1,000
0,33	0,95600	0,286	0,964000	0,25	0,967	0,222	0,971
1,00	0,74000	0,857	0,768000	0,75	0,794	0,667	0,814
1.33	0,57000	1,143	0,620000	1,00	0,654	0,889	0,865
2,33	0,43000	2,000	0,486000	1,75	0,531	1,556	0,569
3.00	0,25200	2,571	0,305000	2,25	0,355	2,000	0,398
4.00	0,18400	3,429	0,237000	3,00	0,285	2,667	0,328
4,33	0,14200	3,714	0,192000	3.25	0,236	2,889	0,278
5,33	0,07200	4,571	0,112000	4,00	0,149	3,556	0,187
6.33	0,05200	5,429	0,085000	4,75	0,120	4,222	0,154
7.00	0,02900	6,000	0,052000	5,25	0,079	4,667	0,107
8.33	0,01200	7,143	0,027000	6,25	0,047	5,556	0,069
9,00	0,00810	7,714	0,021000	6,75	0,038	6,000	0,057
9,33	0,00550	8,000	0,016000	7,00	0,030	6,222	0,048
10,33	0,00170	8,857	0,008400	7,75	0,018	6,889	0,031
12,00	0,00013	10,286	0,003600	9,00	0,0099	8,000	0,019
		10,571	0,002700	9,25	0,0080	8,222	0,016
		11,143	0,001200	9,75	0,0048	8,667	0,010
		12,286	0,000320	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

Продовження таблиці 13

						14,222 14,889 16,222 18,000	0,000097 0,000054 0,000011 0,0000006
Кількість умов (вибірок) *c = 4*							
n = 2		***n = 3***		***n = 4***			
χ_r^2	α	χ_r^2	α	χ_r^2	α	χ_r^2	α
0,0	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	8,1	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

Таблиця 14

Критичні значення критерію тенденцій L Пейджа

Кількість респондентів n	c (кількість умов)				α
	3	**4**	**5**	**6**	
2	–	–	109	178	0,001
	–	60	106	173	0,01
	28	58	103	166	0,05
3	–	89	160	260	0,001
	42	87	155	252	0,01
	41	84	150	244	0,05
4	56	117	210	341	0,001
	55	114	204	331	0,01
	54	111	197	321	0,05
5	70	145	259	420	0,001
	68	141	251	409	0,01
	66	137	244	397	0,05
6	83	172	307	499	0,001
	81	167	299	486	0,01
	79	163	291	474	0,05
7	96	198	355	577	0,001
	93	193	346	563	0,01
	91	189	338	550	0,05
8	109	225	403	655	0,001
	106	220	393	640	0,01
	104	214	384	625	0,05

Продовження таблиці 14

9	121	252	451	733	0,001
	119	246	441	717	0,01
	116	240	431	701	0,05
10	134	278	499	811	0,001
	131	272	487	793	0,01
	128	266	477	777	0,05
11	147	305	546	888	0,001
	144	298	534	869	0,01
	141	292	523	852	0,05
12	160	331	593	965	0,001
	156	324	581	946	0,01
	153	317	570	928	0,05

Таблиця 15

Критичні значення критерію λ Колмогорова-Смірнова (зіставлення емпіричного розподілу з теоретичним)

n	Найбільше накопичене відхилення d_m	
	Рівень статистичної значущості α = 0,05	*Рівень статистичної значущості* α = 0,01
5	0,6074	0,7279
10	0,4295	0,5147
15	0.3507	0,4202
20	0,3037	0,3639
25	0,2716	0,3255
30	0,2480	0,2972
40	0,2147	0,2574
50	0,1921	0,2302
60	0,1753	0,2101
70	0,1623	0,1945
80	0,1518	0,1820
90	0,1432	–
100	0,1358	–
>100	$\frac{1,36}{\sqrt{n}}$	$\frac{1,63}{\sqrt{n}}$

Таблиця 16

Критичні значення критерію λ Колмогорова-Смірнова (зіставлення двох емпіричних розподілів)

λ	λ, останній десятковий знак				
	0	**1**	**2**	**3**	**4**
	Рівень статистичної значимості				
0,3	0,99999	0,99998	0,99995	0,99991	0,99983
0,4	99719	99603	99452	99262	99027
0.5	96394	95719	94969	94147	93250
0.6	86428	85077	83678	82225	80732
0,7	71124	69453	67774	66089	64402
0,8	54414	52796	51197	49619	48063
0,9	39273	37907	36571	35266	33992
1,0	27000	25943	24917	23922	22957
1,1	17772	17005	16264	15550	14861
1,2	11225	10697	10190	09703	09235
1,3	06809	06463	06132	05815	05513
1,4	03968	03751	03545	03348	03162
1,5	02222	02092	01969	01852	01742
1,6	01195	01121	01051	00985	00922
1,7	00618	00577	00539	00503	00469
1,8	00307	00285	00265	00247	00229
1,9	00146	00136	00126	00116	00108
2,0	00067	00062	00057	00053	00048
2,1	00030	00027	00025	00023	00021
2,2	00013	00011	00010	00010	00009
2,3	00005	00005	00004	00004	00004
2,4	00002	00002	00002	00001	00001

Продовження таблиці 16

λ	5	6	7	8	9
0,3	0,99970	0,99949	0,99917	0,99872	0,99807
0,4	98741	98400	97998	97532	96998
0.5	92282	91242	90134	88960	87724
0.6	79201	77636	76042	74422	72781
0,7	62717	61036	59363	57700	56050
0,8	46532	45026	43545	42093	40668
0,9	32748	31536	30356	29206	28087
1,0	22021	21114	20236	19387	18566
1,1	14196	13556	12939	12345	11774
1,2	08787	08357	07944	07550	07171
1,3	05224	04949	04686	04435	04196
1,4	02984	02815	02655	02503	02359
1,5	01638	01539	01446	01357	01274
1,6	00864	00808	00756	00707	00661
1,7	00438	00408	00380	00354	00330
1,8	00213	00198	00186	00170	00158
1,9	00100	00092	00085	00079	00073
2,0	00045	00041	00038	00035	00032
2,1	00019	00018	00016	00015	00014
2,2	00008	00007	00007	00006	00006
2,3	00003	00003	00003	00002	00002
2,4	00001	00001	00001	00001	00001

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Авраменко В. І., Карімов І. К. Теорія ймовірностей і математична статистика: навч. посіб. 2-е вид. Дніпродзержинськ: ДДТУ, 2013. 245 с.
2. Бахрушин В. Є. Методи аналізу даних: навч. посіб. для студентів. Запоріжжя: КПУ, 2011. 268 с.
3. Боснюк В. Ф. Математичні методи в психології: курс лекцій. Харків: НУЦЗУ, 2020. 141 с.
4. Вашків П. Г., Пастер П. І., Сторожук В. П., Ткач Є. І. Теорія статистики: навч. посіб. Київ: Либідь, 2001. 201 с.
5. Вдовенко В. В. Математичні методи в психології: навчально-методичний посібник. Кіровоград: ПП «Центр оперативної поліграфії» Авангард», 2017. 112 с.
6. Горонескуль М. М. Таблиці функцій та критичних точок розподілів. Розділи: Теорія ймовірностей. Математична статистика. Математичні методи в психології. Харків: УЦЗУ, 2009. 90 с.
7. Донченко В. С., Сидоров М. В.-С. Теорія ймовірностей та математична статистика для соціальних наук: навч. посіб. Київ: ВПЦ «Київський університет», 2015. 400 с.
8. Дубровін В. І. Статистичні критерії: навч. посіб. Запоріжжя, 1998. 40 с.
9. Климчук В. О. Математичні методи в психології: навч. посіб. для студентів психологічних спеціальностей. Київ: Освіта України, 2009. 288 с.
10. Олефір В. О. Математичні методи в психології: методичні вказівки з організації та планування самостійної роботи студентів для здобувачів освітньо-кваліфікаційного рівня бакалавр за спеціальністю 053 – психологія. Харків, 2016. 59 с.
11. Питьовка О. Ю. Математичні методи в психології: методичні рекомендації для виконання самостійної роботи студентів денної та заочної форми навчання спеціальності 053 «Психологія». Мукачево: МДУ, 2017. 61 с.
12. Руденко В. М., Руденко Н. М. Математичні методи в психології: підручник для студентів вищих навчальних закладів. Рівне: Видавець Олег Зень, 2008. 496 с.
13. Руська Р. В. Математичні методи у психології. Курс лекцій. Тернопіль, 2018. 203 с.
14. Руська Р. В. Теорія імовірності та математична статистика в психології: навч. посіб. Тернопіль, 2020. 112 с.
15. Татьянчиков А. О. Математичні методи в психології: навчально-методичні рекомендації (в допомогу до самостійної роботи для здобувачів вищої освіти ступеня бакалавра факультету психології, політології та соціології); кафедра психології НУ «Одеська юридична академія». Одеса: Фенікс, 2021. 48 с.
16. Телейко А. Б., Чорней Р. К. Математико-статистичні методи в соціології та психології: навчальний посібник. Київ: МАУП, 2007. 424 с.
17. Турчин В. М. Теорія ймовірностей і математична статистика: Основні поняття, приклади, задачі. Дніпропетровськ: ДНУ, 2006. 476 с.

ЗМІСТ

Міністерство освіти і науки України
Кам’янець-Подільський національний університет імені Івана Огієнка

НАВЧАЛЬНЕ ЕЛЕКТРОННЕ ВИДАННЯ

ГУДИМА Уляна Василівна,
кандидат фізико-математичних наук, доцент,
доцент кафедри математики Кам’янець-Подільського
національного університет імені Івана Огієнка
ГУДИМА Олександр Васильович,
кандидат психологічних наук, доцент,
доцент кафедри загальної та практичної психології
Кам’янець-Подільського національного університет імені Івана Огієнка

МАТЕМАТИЧНІ МЕТОДИ В ПСИХОЛОГІЇ: ОСНОВНІ ПОНЯТТЯ ТА ПРИКЛАДИ

НАВЧАЛЬНИЙ ПОСІБНИК

ЕЛЕКТРОННЕ ВИДАННЯ

Підписано 30.01.2023. Формат 60x84/16. Гарнітура «Cambria».
Об’єм даних 5,3 Мб. Обл.-вид. арк. 5,4. Зам. № 1017.

Кам’янець-Подільський національний університет
імені Івана Огієнка,
вул. Огієнка, 61, м. Кам’янець-Подільський, 32300.
Свідоцтво серії ДК № 3382 від 05.02.2009 р.

Виготовлено в Кам’янець-Подільському національному
університеті імені Івана Огієнка,
вул. Огієнка, 61, м. Кам’янець-Подільський, 32300.