МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИ
УМАНСЬКИЙ ДЕРЖАВНИЙ ПЕДАГОГІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ПАВЛА ТИЧИНИ

Математичний апарат педагогічної науки

Навчальний посібник
(конспект лекцій)

Укладач: Т. В. Поліщук

УДК 519.2 (075.8)
М 34

Укладач:
Т. В. Поліщук – кандидат фізико-математичних наук, доцент кафедри вищої математики та методики навчання математики Уманського державного педагогічного університету імені Павла Тичини.

Рецензенти:
М. В. Дудик – кандидат фізико-математичних наук, доцент, професор кафедри фізики та астрономії та методики їх викладання Уманського державного педагогічного університету імені Павла Тичини;
М. О. Медвєдева – кандидат педагогічних, доцент, завідувач кафедри інформатики та інформаційно-комунікаційних технологій Уманського державного педагогічного університетц імені Павла Тичини.

Рекомендовано до друку вченою радою
факультету фізики, математики та інформатики
Уманського державного педагогічного університету імені Павла Тичини
(протокол № 4 від 28 листопада 2019 року)

М 34 Математичний апарат педагогічної науки: навчально-методичний посібник / МОН України, Уманський держ. пед. ун-т імені Павла Тичини; уклад. Т. В. Поліщук. – Умань : Візаві, 2019. – 109 с.

У навчальному посібнику описані основні математичні поняття, що використовуються при математичній обробці результатів педагогічних досліджень, розглянуто основні критерії, наведено приклади обробки результату експерименту. Посібник буде корисним магістрантам, аспірантам та педагогам, які займаються дослідницькою та науково-дослідною роботою.

УДК 519.2 (075.8)

Зміст

ВСТУП

Більшість явищ та процесів, які відбуваються навколо нас можна точно описати за допомогою математичних формул. Сучасна наука здатна наперед передбачити розташування небесних світил, провести розрахунки орбіт штучних супутників, період розпаду радіоактивних речовин, передбачити руйнування матеріалів та багато іншого.

Усім добре відомий вислів *«математика – цариця наук»*. Інша наука стає дійсною наукою тільки тоді, коли вона починає використовувати математичний апарат. Проте багато вчених гуманітаріїв переконані, що не одна математика може претендувати на цей титул. Адже, наприклад, математику для доведення своїх положень зовсім не потрібно звертатись до психології, а в психології можна здійснювати відкриття, не використовуючи математику.

Прикладом можуть служити теорія психоаналізу, аналітична психологія К. Юнга, індивідуальна психологія А. Адлера, об'єктивна психологія В. М. Бехтерева, культурно-історична теорія Л. С. Виготського, концепція відношень особистості В. Н. Мясищева та багато інших теорій.

На даний час більшість психологічних та педагогічних концепцій піддаються сумніву тільки тому, що вони не були підтверджені статистично. Психологія та педагогіка – це науки, в яких немає одиниць вимірювання та уявлень про те, як запозичені ними одиниці вимірювання – міліметри, секунди і градуси – співвідносяться з психічними феноменами та педагогічними явищами.

Між тим, як зазначає Ірина Грекова *« ...явища, які є предметом вивчення гуманітарних наук, безумово складніші ніж ті, якими займаються точні наук. Вони набагато «важчі» (якщо взагалі) піддаються формалізації. Вербальний спосіб побудови дослідження тут, як не парадоксально, виявляється точніший ніж формально-логічний»*.

Педагогічні явища є значно складнішими і непередбачуваними. Зокрема, наперед, важко сказати, як опанують учні чи студенти програмний матеріал з тієї чи іншої теми, який навчальний предмет засвоюється краще, а який гірше, як змінеться рівень знань після вивчення певного блоку дисциплін.

Однак, і в сукупності цих, іноді випадкових явищ існують певні закономірності. Як відомо, вивченням *«випадкових»* закономірностей займається розділ математики, що називається теорія ймовірностей. Тому застосування математичних методів доцільне не лише під час проведення досліджень у природничо-математичних науках, а й у суспільних також.

Зазвичай, у ході педагогічного дослідження, накопичують велику кількість інформації, однак обробляють і аналізують її не завжди раціонально, а часом і незадовільно. Здебільшого, це пояснюється складністю педагогічних досліджень, а також відсутністю знань у дослідника з відповідних розділів математики, а саме математичної статистики і теорії ймовірностей, оскільки переважна більшість дослідників даного профілю мають гуманітарну освіту.

Ця проблема привернула увагу багатьох науковців. Так, у 1969 році вийшла праця П. М. Воловика «Теорія ймовірностей і математична статистика в педагогіці», яка мала за мету допомогти опанувати складні процеси математичних обчислень та ліквідувати своєрідну прогалину у знаннях дослідників.

Для розуміння матеріалу, викладеного у цій монографії, достатньо знань з математики за курс середньої школи. Крім того, в ній наведено чимало наочних прикладів обробки результатів педагогічних експериментів. Автор навів методи розв’язування задач, які постають у процесі обробки результатів педагогічних досліджень.

Для практичного проведення розрахунків недостатньо знати лише якісний характер закономірностей. Також необхідно вміти кількісно оцінити ці закономірності за даними спостережень і вимірювань, знайти середнє значення характеристик, що вивчаються, і визначити досить надійні межі можливих відхилень характеристик від їх середніх значень. Методику розв’язування цих задач дає нам математична статистика.

За виявленими з допомогою теорії ймовірностей закономірностями можна передбачити, як випадкові явища відбуватимуться надалі.

Сьогодні існує велика кількість комп’ютерних програм та додатків (STATISTICA, MS Excel, GEOGEBRA та інш.) для статистичної обробки даних. Але комп’ютер не пояснює як інтерпретувати отримані чисельні результати та які способи для цього потрібно використати.

Тема 1.

Загальні питання досліджень у педагогічних науках

Вимірювання результатів педагогічних досліджень.

У наукових дослідженнях досить часто досліди (спостереження) повторюються декілька разів, при цьому головні комплекси умов залишаються незмінними. Однак, незважаючи на це, їхні результати завжди зазнаватимуть випадкового розсіювання. Проте той факт, що результат кожного окремого досліду, який зазнає випадкового розсіювання, неможливо заздалегідь передбачити, ще не означає, що повторні проведення дослідів (спостережень) не виявлять певної закономірності. Детальне вивчення цієї закономірності показує, що математично її можна описати нормальною кривою розподілу (тим краще, чим більше дослідів (спостережень) проводиться).

Як правило, при цьому, основна маса результатів групується навколо певного середнього значення α, якому відповідає «реальна величина» результату досліду (спостереження). Відхилення від цього значення траплятиметься тим рідше, чим більша абсолютна величина таких відхилень. Амплітуда випадкових коливань результатів дослідів характеризується середнім квадратичним відхиленням σ. Сталі величини α *і* σ — параметри кривої розподілу.

Якщо мати велику кількість результатів дослідів (спостережень), то можна було б відносно легко знайти «реальне значення» величини цих результатів, а також основну зону їх розсіювання при застосуванні даного вимірювання (тобто σ), побудувавши емпіричну криву розподілу результатів дослідів (спостережень).

Слід зазначити, що для проведення великої кількості дослідів (спостережень) потрібно багато часу і коштів, а тому їх кількість доводиться обмежувати. За цих умов досліднikові необхідно знайти надійні середні показники, які були б найкращими оцінками параметрів розподілу (α *і* σ). Виконати це завдання можна за допомогою методів прикладної математики, а саме математичної статистики.

Статистика зародилась як учення про об’єднане у певну політичну одиницю народонаселення. Проте статистичні методи безпосередньо не мають справи із спільнотами людей чи із соціальними групами, оскільки майже ніколи не буває даних, які б повно і всебічно характеризували людину. Внаслідок цього сукупність, що вивчається, до певної міри завжди є абстрактною. Сукупності, що є предметом статистичного вивчення, завжди характеризуються мінливістю. Причини мінливості будь-якої змінної величини слід вивчати на підставі вимірювань та аналізу варіації.

Вчення про варіації призводить безпосередньо до концепції розподілу чисельностей. Ці розподіли можуть бути різними, а кількість класів, на які розподілена певна сукупність, – скінченою або нескінченною. У найпростішому з можливих випадків, коли є лише два класи (наприклад, встигаючі і невстигаючі студенти), розподіл визначається тією пропорцією, в якій зустрічаються члени, що належать до того чи іншого класу (відсоток встигаючих та невстигаючих студентів). В інших випадках мінливість може бути перервною з невизначеною наперед кількістю класів поділу. *Наприклад*, якщо розглядати сукупність студентів, що вчаться на «відмінно» у різних групах вузу, то тут розподіл визначається кількістю навчальних груп, в яких є відмінники $0, 1, 2, \ldots, n$. Класів поділу в цьому випадку має бути стільки, щоб включали всі навчальні групи. Варіююча величина – є дискретною випадковою величиною. Розподіл чисельностей показує частоту набуття випадковою величиною кожного з можливих своїх значень. У ряді випадків випадкова величина може набувати будь-яких проміжних значень всередині інтервалу її варіювання. Випадкова величина при цьому є неперервною, відповідний розподіл можна виразити у формі математичної функції випадкової величини.

У багатьох випадках знаходимо цілком закономірний зв’язок між числовими значеннями варіюючих ознак і ймовірністю реалізації цих значень у множині дослідів (спостережень), що проводяться. Саме ця обставина дає змогу побудувати загальну теорію, яка розробляє раціональні прийоми обробки дослідних даних, що стосуються масових явищ і відображають вплив розсіюючих випадкових факторів. Ці прийоми істотно пов’язані з припущенням сталості відносної частоти і наявністю імовірнісних законів розсіювання та називаються математико-статистичними методами. Вони уможливлюють вивчення варіації випадкових величин, дозволяють знайти середні показники, оцінити їхню точність і надійність, зробити припущення щодо істотності (невипадковості) відмінностей у двох серіях дослідів (спостережень) одного типу, провести порівняння ефективності того чи іншого методу (перевірка гіпотез), а також вивчити спільну варіацію двох або більше випадкових величин (кореляція).

Одним із найважливіших завдань будь-якого дослідження, зокрема і педагогічного, є встановлення зв'язку між тими числовими показниками, зміна яких визначає сутність процесу, що вивчається. Зв'язки у навколишньому світі є дуже складними й багатогранними. Щоб пізнати будь-яке явище необхідно вивчити не лише його зв'язки з навколишніми факторами, а й взаємозв'язки всіх його сторін, тобто слід встановити закономірності змін взаємопов'язаних явищ і показників, що їх характеризують.

У математичній статистиці взаємозв'язки явищ вивчаються методом кореляції. Потреба у застосуванні цього методу зумовлена тим, що не завжди можна врахувати вплив сторонніх факторів, або через те, що ці фактори невідомі, або їх не можна ізолювати. Названий метод кореляції уможливлює з'ясувати за складних взаємодій сторонніх впливів, якою була б залежність між результатом і фактором, за умови, що сторонні фактори не змінюються і своєю зміною не спотворюють основної залежності. Для виявлення закономірностей зв'язку необхідно мати достатньо велику кількість спостережень.

За допомогою кореляції розв'язують такі *дві задачі*:

1) на основі спостережень над значною кількістю фактів визначають, як в середньому змінюється результуюча ознака при зміні певного фактора, припускаючи при цьому, що інші фактори не змінюються, хоч насправді має місце їхній спотворюючий вплив;

2) визначають міру впливу спотворюючих факторів.

Для математичної обробки даних спостережень або експериментів застосовують методи теорії ймовірностей і математичної статистики, які дозволяють на підставі обмеженої кількості дослідів (спостережень) над випадковою величиною дійти обґрунтованого висновку про закони її розподілу та про її кількісні характеристики.

Експериментатор, маючи обмежений дослідний матеріал, не може знайти ні точних значень для математичного сподівання і дисперсії, ні отримати істинний закон розподілу випадкової величини, що вивчається. Можна, дати лише наближену оцінку для математичного сподівання і дисперсії та наближений вираз для закону розподілу, якому, на думку дослідника, підпорядковується дана випадкова величина. Отже головним завданням математичної обробки результатів спостережень є встановлення способів, за якими можна дати певну оцінку для числових параметрів, що характеризують випадкову величину, і наближений вираз її закону розподілу.

Слід зазначити, що поряд з використанням класичної математичної статистики необхідні пошуки і розробка нових методів дослідження та опису кількісних закономірностей педагогічних явищ і процесів з використанням теорії інформації, теорії ігор, теорії дослідження операцій і деяких інших, нових галузей математики та інформатики.

Підвищення ефективності навчального процесу вимагає подальшого піднесення якості науково-педагогічних досліджень, впровадження найновіших досягнень науки у практичну діяльність закладів вищої осіти. Відомо, що якість і надійність досліджень, а також ймовірність одержаних результатів залежать від методів. Метод науково-педагогічного дослідження – це шлях вивчення і опанування складних психолого-педагогічних процесів формування особистості, встановлення об’єктивної закономірності виховання і навчання. Методи дослідження є інструментом пізнання, тому їх вибору надають особливої уваги.

Традиційно педагогічна наука використовує такі добре відомі методи дослідження, як спостереження за ходом освітнього процесу, вивчення та узагальнення досвіду, анкетування, експеримент та моделювання. Однак дослідження освітнього процесу відзначаються надзвичайною складністю і різноманітністю, як і сам процес.

У наукових дослідженнях з гуманітарних наук важливою проблемою є знаходження об’єктивних методів та критеріїв-показників для виміру результатів досліджень.

Одиницю, яка досліджується називають *об’єктом,* який описують за допомогою *ознак* (наприклад результат контрольної роботи учня чи студента). Кожен об’єкт характеризується певним *значенням ознаки* (наприклад, 10 правильних відповідей). *Вимірюванням називається будь-який процес визначення значення*. Прикладом вимірювання є оцінювання роботи учня в школі, чи студента в університеті бал. При вимірювані результати дослідження порівнюють з значеннями відповідної вимірювальної шкали, де кожне реальне значення має формальне найменування, яким, як правило буває число.

Педагогічні вимірювання – це процес отримання кількісної інформації про знання тих хто навчається. Вимірювання в педагогіці, психології, соціології досить важкі за структурою. Вони не досягають настільки абсолютної і відносної точності, як наприклад, у фізичних вимірюваннях, де використовуються метричні шкали. Тому для педагогічних вимірювань вводяться спеціальні критерії оцінювання і менш точні оціночні шкали.

Рівні вимірювань за класифікацією Стівенса і шкали, що використовуються в педагогіці, показані на рис. 1.1 і у табл. 1.1.

Рис. 1.1. Типи вимірювань у педагогіці.

Таблиця 1.1

Рівні вимірювань у педагогіці

Рівень вимірювання (шкала)	Основна операція, що визначає рівень	Математичні і статистичні величини, що характеризують шкалу	Приклади з педагогіки, психології, соціології
Номінальний (номінальна)	Приписування однакових чисел – найменування об'єктам, що мають спільну ознаку	Мода, абсолютна, відносна, відсоткові частоти	Приписування числового коду з визначеними соціально-демографічними характеристиками. Список осіб в групі. Поділ студентів за ознакою «зараховано» – «не зараховано». Поділ студентів за видами навчання – «очне», «заочне», «денне».
Ординальний (порядкова)	Ранжування об'єктів за вираженням певної ознаки	Мода, медіана, квантилі (процентилі, децилі, квартилі та ін.). Рангові критерії	Ранжування спеціалістів за ступенем професійної придатності. Ранжування студентів за оцінками «відмінно», «добре», «задовільно», «незадовільно».
Інтервальний (інтервальна)	Визначення величини відмінностей між об'єктами.	Мода, медіана, вибірка, квантилі, рангові критерії, вибіркова дисперсія, стандартне квадратичне відхилення, коефіцієнт, кореляція	Вимірювання якості правильних відповідей на тест, завдання, анкету.
Вимірювання відношень (відношення)	Визначення рівності відношень величин	Всі арифметичні операції і всі поняття і методи математичної статистики	Вимірювання відношень тимчасових інтервалів, що характеризують навчальний процес.

Шкали вимірювання.

Номінальна шкала – найбільш проста форма вимірювань, у якій кожному об'єкту, що оцінюється приписується назва або число. У педагогічній практиці номінальна шкала застосовується для класифікації тих що навчаються по групах у відповідності з якоюсь ознакою (шкала статті: дівчина – хлопчик; нації: українець, росіянин, білорус; інтересів: інтерес до музики, літератури, спорту). В освітньому процесі цими ознаками можуть бути, наприклад, оцінки «зараховано» – «не зараховано». Арифметичні дії з числами в номінальній шкалі не мають змісту. Між виміряними величинами немає і не може бути ієрархічного зв'язку. Значення шкали не порівняльні (жодне її значення не є більшим, меншим, кращим, гіршим). При використані цієї шкали вимірювання часто називають визначенням чи класифікацією. За допомогою номінальної шкали можна виміряти тільки якісні ознаки. Кількісна оцінка обмежується декількома станами (≥2). Обробку кількісних даних при вимірюванні в номінальній шкалі слід проводити не з результатами вимірювань, а з питомою вагою кількості об'єктів даного класу. У цій шкалі можна використати наступні статистичні операції:

- розрахунок частот, частотних (питомої ваги) об'єктів даного класу;
- визначення моди досліджуваної ознаки.

Зміст понять *«частота» і «мода»* стосовно педагогічних вимірювань буде показаний далі на конкретних прикладах.

Порядкова шкала (ординальна). Дана шкала дозволяє також проводити тільки якісні вимірювання. Її значення мають певну послідовність, їх можна поставити в одному порядку в ряд від найменшого до найбільшого. Важливо те, що кроки, інтервали шкали не порівняльні. В порядковій шкалі об'єкти оцінюються з точки зору відношення *рівності* між ними чи відношення «більше-менше», тому не можна врахувати відстань між об'єктами, що є основним недоліком цієї шкали.

Приклад: шкала шкільних балів 1, 2, 3, 4, 5, … 12 не можна сказати, що різниця між 1 і 2 така ж сама як між 11 і 12. Шкільний бал показує якість, а не кількість вимірювального явища. Вони нерідко бувають суб'єктивними. При вимірюванні ознаки в порядковій шкалі можливі тільки монотонні перетворення. Арифметичні дії тут не мають сенсу, тобто так званий *«середній бал»* не відображає об'єктивні закономірності. Тому над ними неможна робити арифметичні дії такі як середнє арифметичне, дисперсія.

У шкалі порядку допустимі наступні статистичні операції:

- в якості «середньої» оцінки можна використовувати медіану;
- в якості міри розсіювання можна використовувати квантилі;
- в якості міри зв'язку двох ознак можна використовувати ранговий коефіцієнт кореляції.

Недоліки номінальної і порядкової шкали в якійсь мірі дозволяє усунути *інтервальна шкала*.

1. *Інтервальна шкала (кількісна).* За допомогою цієї шкали можна виміряти кількість правильних відповідей, даних студентом в результаті виконання тесту чи програмованого завдання, врахувати затрати часу на самопідготовку і т. д. Будучи якісною, інтервальна шкала, однак, не дозволяє судити, в скільки разів одна кількісна характеристика більша за іншу. Якщо, наприклад, студент відповів на 30 питань тесту, це не означає, що він знає в два рази більше ніж студент, що відповів на 15 питань.

В інтервальній шкалі допустимі майже всі статистичні операції, крім тих, які передбачають знання *«істинно»* нульової точки шкали. в інтервальній шкалі не можна використовувати такі характеристики, як середнє геометричне і коефіцієнт варіації ознаки, що спостерігається. Недоліком цієї шкали є невизначеність точки відліку і невизначеність абсолютного нуля. При оцінці, скажемо, виконання програмового завдання, нуль правильних відповідей не означає повної відсутності знань в студента (дійсно, температура 0°С не є абсолютним нулем, а температура +12°С не означає вдвічі більшу температуру в порівнянні з +6°С). При цій шкалі точно відомо наскільки один інтервал шкали більше чи менше від іншого. При рівномірній інтервальній шкалі (шкалі температури, маси, довжини) інтервали однакові. При «+», «-», показників інтервальні шкали результатам можна надавати конкретного змісту.

Особливим видам рівномірної інтервальної шкали є *шкала відношень.*
а) шкала відношень. В такій шкалі можна встановити рівень відношень чисел, що приписуються об'єктам (2/4=4/8). Приклад: шкали фізичних величин. В цій шкалі можна здійснювати всі арифметичні дії, використовувати поняття та методи математичної статистики. Будь-яка інтервальна шкала перетворюється у шкалу відношень якщо чітко фіксувати початок відліку. За допомогою критеріїв визначається рівень знань учня по відношенню до вимог

програми. Отримана при цьому бальна шкала є інтервальної. Критеріями можуть бути: кількість розв'язаних задач, кількість зроблених помилок, точність виконання роботи, швидкість її виконання.

б) альтернативна шкала - містить два значення: «Так – Ні». Тут є один інтервал ділення (0-1), (1-2). Деякі вчені стверджували, що в педагогічних науках не завжди можна використовувати шкали інтервального типу (крім шкали часу). Але це не так, дійсно коли один учень не розв'язав за урок жодної задачі, то нуль - означає відсутність розв'язаних ним задач. Якщо один учень розв'язав 5, а другий 10 задач, то можна сказати, що другий зробив у два рази більше.

Якщо знання учнів, студентів оцінювати за одними і тими ж критеріями, що вимірюють їх знання і вміння у відношенні до знань, що вимагає програма, то така шкала оцінок наближається до інтервальної. Адже вища оцінка показує відношення об'єктів до вищого інтервалу. У звичайній педагогічній діяльності можна користуватися, а в педагогічних дослідженнях оцінки потрібно використовувати обережно. Обовязково має бути і якісний аналіз педагогічних явищ.

Шкала відношень дозволяє отримати найвищий рівень вимірювання. За нею, на відміну від інтервальної шкали, можна визначити відношення чисел, що приписуються об'єкту. У даній шкалі в якості відліку вибраний абсолютний нуль. У педагогічних вимірюваннях її можна використовувати, наприклад, для знаходження тимчасових інтервалів, що характеризують навчальний процес і його елементи.

Аналіз табл. 1 показує, що кількість числових характеристик і клас допустимих математико-статистичних операцій розширюються при переході від одного рівня до другого, внаслідок чого використання шкали відношень відкриває широкі можливості для кількісних вимірювань. Однак вивчення вітчизняного і зарубіжного досвіду педагогічних вимірювань показує, що для більшості характеристик освітнього процесу цілком достатні номінальна, порядкова та інтервальні шкали.

Контроль рівня знань. Якісні педагогічні вимірювання за номінальною і порядковою шкалам можна отримати з допомогою всіх традиційних видів контролю, що застосовуються у закладах середньої та вищої освіти. Види, форми організації і методи контролю навчальної роботи студентів представлені в табл. 1.2.

Дані спостережень контролю до аналізу та інтерпретації необхідно узагальнити. Результати міжсесійного, квартального і підсумкового контролю можна представити у вигляді не згрупованого ряду довільної форми.

Таблиця 1.2

Види і форми організації контролю навчальної роботи студентів

Педагогічні вимоги до контролю	Функції контролю	Форми контролю	Види контролю
1. Цілеспрямованість (пізнавальна, навчально-методична, психологічна, організаційна) 2. Об'єктивність 3. Всебічність 4. Регулярність 5. Індивідуальність 6. Різноманіття форм 7. Диференційований підхід 8. Єдність вимог	1. Навчальна 2. Діагностична 3. Контролююча 4. Виховна 5. Розвиваюча 6. Прогностична	1. Індивідуальна 2. Фронтальна 3. Групова 4. Самоконтроль	1. Міжсесійний - попередній; - поточний; - тематичний 2. Підсумковий - заліковий; - екзаменаційний

Таблиця 1.2

Методи контролю навчальної роботи студентів

Спостереження за роботою	Усний контроль	Письмовий контроль	Лабораторно-практичний контроль	Програмований контроль
Систематичне спостереження за роботою студентів на заняттях і поза аудиторією	Індивідуальне опитування, фронтальне опитування, контрольне читання креслень, контрольне читання технічної документації, усний залік, усний екзамен, усний самоконтроль	Розв'язування задач, виконання вправ, контрольна письмова робота, письмовий екзамен, письмовий самоконтроль, домашня письмова робота (з навчальною метою), тестування	Контрольна, лабораторна робота, лабораторно-практичний самоконтроль, розв'язання задач, виконання креслень, графіків, схем	1. Безмашинний контроль: - контрольні карти безмашинного контролю; - програмована письмова робота; - програмований контрольний залік 2. Машинний контроль

Критерії оцінки результатів навчальної роботи.

Останнім часом для об'єктивної роботи результатів навчальної роботи в психолого–педагогічних дослідженнях використовують різні критерії.

Так, наприклад, вважаються показниками ефективності навчального процесу *індивідуальний темп (Т) роботи і успішність (а):*

$$T = \frac{t}{n}, \; a = \frac{k}{K}, \qquad (1.1)$$

де t – час, який витрачений на засвоєння матеріалу; n – об'єм засвоєного матеріалу; k – кількість правильних операцій (дій); K – кількість необхідна для отримання правильної відповіді операції (дії). Якщо задача розв'язана правильно, то $a=1$, при невирішеній задачі $a=0$, при незакінченому (половинному) розв'язку $0<a<1$).

Пропонується ще одна формула:

$$K_y = \frac{N_1}{N} \cdot 100, \qquad (1.2)$$

де K_y – успішність; N_1 – число правильних відповідей; N – загальне число заданих запитань.

Деякі дослідники вважають, що для визначення ефективності навчальної роботи достатньо порівняти успішність за оцінкам. До речі час, який був витрачений на засвоєння навчальної інформації, пропонується брати середнім для всього класу або групи. Так, І. Т. Огородніков для оцінки прийомів і методів навчання поділяє критерії на дві групи: критерії якості засвоєння (об'єм знань, системність знань, міцность, реальність знань) і критерії розвитку самостійності та творчої активності (допитливість, здатність аналізувати і узагальнювати вивчений матеріал, самостійність міркувань, умовиводів, вмінь користуватись вивченим матеріалом в подальшій діяльності) і т. д. Дослідниця Т. І. Андреева при оцінці ефективності навчання враховує три рівні засвоєння навчальної інформації: розуміння, вміння та запам'ятовування. Учений А. С. Батищев зазначав, що основним критерієм для визначення рівня знань з теоретичної дисципліни можуть бути:

- кількість правильно розв'язаних задач;
- співвідношення між розв'язаними та заданими задачами;
- самостійність;

– перенесення навичок;
– успішність учнів;
– рівень засвоєння знань;
– формування навичок і вмінь;
– пов’язування кінцевого результату із поставленою задачею;
– застосування набутих знань і т. д.

Ми бачимо, що кількість критеріїв оцінки навчальної роботи достатньо велика і початківцю в них розібратись важно.
Але із усієї сукупності можна виділити два види критеріїв:

– кількісні (отримують на основі інтервальної шкали та піддаються статистичним методам обробки);
– якісні (їх не можна виміряти безпосередньо в інтервальній шкалі, але за допомогою них можна аналізувати зміст педагогічного явища).

Кількісні критерії:

1. *Об’єм засвоєних знань* (I_0) (вимірюється кількістю вірно розв’язаних задач у контрольній роботі чи тесті). Якщо результати навчальної роботи виражені в одній і тій самій вимірювальній шкалі але максимально можливий результат при різних вимірах однаковий, то даним показником користуватися можна. Якщо максимально можливий результат різний – цим показником не користуються. Тоді застосовують довільне відносне число.

2. *Коефіцієнт засвоєння навчального матеріалу* (k).

$$k = \frac{I_0}{I_a}, \tag{1.3}$$

в чисельнику – об’єм навчального матеріалу *засвоєного* учнями за певний відрізок часу (вранці, 1-2 урок), у знаменнику – об’єм матеріалу, що *повідомляється* учням за той самий проміжок часу.

За одиницю часу беруть 1 урок або 2-3, що слідують один за одним. Брати більше не потрібно, тому що на засвоєння починають впливати фактори такі як забування та домашня підготовка.

Умовні одиниці виміру навчального матеріалу. Ними можуть бути як одиниці, що застосовуються в програмованому навчанні так і вбрані самим вчителем – формули, умовні знаки, правила.

У навчальному процесі може бути *два випадки:*

1. Певний навчальний математичний засвоюється учнями в різні одиниці часу (урок і 1 година).

2. Учні вивчають один і той самий матеріал самостійно всіма доступними способами (книга, зошит, Інтернет). В цьому випадку не виключено, що крім навчального матеріалу учень засвоює якусь нову тему чи ознайомиться з додатковим матеріалом.

3. *Швидкість засвоєння навчального матеріалу* (V) або співвідношення коефіцієнта засвоєння матеріалу і часу витраченого на його засвоєння k_1:

$$V = \frac{I_0}{t}, \ k_t = \frac{k}{t}, \qquad (1.4)$$

де t – час затрачений учнем на засвоєння матеріалу.

4. *Коефіцієнт надійності засвоєння навчального матеріалу (А)* дорівнює співвідношенню між навчальним матеріалом, що учні запам’ятали і навчальним матеріалом, що їм повідомлявся в процесі навчання:

$$A = \frac{I_m}{I_a}. \qquad (1.5)$$

Різниця між коефіцієнтами засвоєння навального і міцністю засвоєння навчального матеріалу показує *забування* вивченого матеріалу.

Якісні критерії:

1) рівень знання навчального матеріалу (учень, студент впізнає об’єкт на основі його суттєвих ознак);
2) рівень розуміння залежності навчального матеріалу (розуміння функціональної залежності між вивченими явищами та уміння описувати об’єкти);
3) рівень оволодіння навчальним матеріалом (учень, студент практично може застосовувати теоретичний матеріал для вирішення задач);
4) рівень оволодіння інтелектуальними навичками (учень, студент вільно оперує вивченим матеріалом, вміє трансформувати засвоєний матеріал в нових умовах свідомо і оперативно).

Якщо ви хочете визначити рівень засвоєння матеріалу учнями, студентами на протязі заняття – це варто робити через 2-3 години після заняття. Використовувати тести і короткі контрольні роботи, питання різного рівня складності, і їх результати потрібно врахувати окремо, або при сумування результатів використовувати різні коефіцієнти.

Методи педагогічних досліджень.

Як зазначалося вище педагогічна наука використовує такі добре відомі методи дослідження, як спостереження за ходом навчально-виховного процесу, вивчення та узагальнення досвіду, анкетування, експеримент та моделювання.

Методи математичного моделювання. Останнім часом у науці почали широко використовувати метод моделювання — абстрагованого відображення процесів явищ в іншій системі інформації, вивчення якої дає нові відомості для дослідника. Метод моделювання також є досить перспективним для дослідження освітнього процесу.

Термін *«модель»* походить від латинського слова *modulus* — міра. У найбільш поширеному вжитку він означає подібність певних предметів і явищ.

Під науковою моделлю розуміють уявну або матеріально створену систему, яка, відображаючи або відтворюючи об’єкт дослідження, здатна замінити його так, що при її вивченні дослідник має можливість отримати нову інформацію про даний об’єкт.

Всі існуючі моделі поділяють на такі три головні групи: речові (фізичні), речово-математичні (в тому числі й алгоритмізовані) та знакові (логіко-математичні).

Фізичне моделювання — це моделювання з допомогою спеціальних моделюючих установок, які переважно встановлюються в обладнаних лабораторіях. Вони дозволяють вивчати складні фізичні, хімічні й технологічні процеси.

Майже всі галузі наук надають фундаментального значення інформаційному моделюванню. При такому моделюванні властивості об’єкта чи процесу зображуються у певній знаковій системі. Найбільшого поширення набули знакові системи, які базуються на використанні алгоритмічних мов та мов програмування, а також логіко-математичних обчислень. Це робить згадані моделі зручними для опису кількісних характеристик і доступними для «розуміння» їх ЕОМ, що особливо актуально на сучасному етапі інформатизації більшості сфер діяльності людини, в тому числі й освітнього процесу. Названа перевага дозволяє на ЕОМ, у межах певної математичної моделі, автоматизувати розв'язання різних конкретних завдань.

Моделями освітнього процесу можна назвати абстраговане відображення педагогічних закономірностей з допомогою штучно створених матеріальних об'єктів чи опис цих закономірностей мовою математичної символіки.

Найкраще для моделювання освітнього процесу підходять фізичне та інформаційне моделювання.

Педагогічними інформаційними моделями є математичні та логіко-математичні програми. У цих групах моделей навчання і виховання виділяють моделі різної складності. Зокрема, слід виділити науково-практичні, призначені безпосередньо для практичного розв'язання наукових проблем у межах частинних гіпотез та вищі абстраговані моделі – для абстрагованого опису складних теоретичних положень.

За допомогою моделі можна з гіпотез вивести опорні точки для думки про достовірність цих же гіпотез.

При так званому дедуктивному моделюванні дослідник відштовхується від якоїсь гіпотези про заздалегідь постульований набір загальних визначень (які формулюють кількісні відносини, можливі між даними змінними), і від деякого набору умов, що визначають правила вибору тих або інших з можливих відносин в різних можливих випадках. Потім підбирається з наявних в математиці або спеціально конструюється формула, яка представляє структуру такої системи. Інтерпретована у відповідних змістовних термінах, вона розглядається як модель досліджуваного виду навчання їх поведінки.

За допомогою таких моделей в професійній педагогіці можна давати абстрактну схему деяких дій в певних ситуаціях вони допомагають розставляти знайдені емпіричним шляхом факти у пропонованій теоретичній системі. На підставі аналізу самих моделей можна робити нові теоретичні і практичні висновки про досліджування явищ.

Приведемо приклад використання моделі для формулювання висновків.

Якщо час, витрачений на виконання всіх одиниць продукції (вироби), відзначити T, $t_p -$ корисний час, витрачений безпосередньо на обробку виробу, t_x - час, витрачений працівником на виконання допоміжних операцій в процесі роботи устаткування, то:

$$T = t_p + t_x.$$

Максимальна продуктивність знарядь праці $Q_{макс}$ визначиться з наступного співвідношення:

$$Q_{\text{макс}} = \frac{k}{k \cdot t_x + 1}\ , \tag{1.6}$$

де $k = \frac{1}{t_p}$ величина, що характеризує технологічну продуктивність робочої машини.

Аналізуючи вище наведені співвідношення, можна зробити наступні висновки:

якщо витрати часу на виконання допоміжних операцій залишаються незмінними $t_x = const$, то знаряддя праці (верстат, автомат, калькулятор і т. д.) мають межу підвищення продуктивності;

якщо разом із збільшенням технологічної продуктивності знаряддя праці $(k \to \infty)$ скорочується одночасно час на допоміжні операції $(t_x \to 0)$, то продуктивність знарядь праці практично не має межі, тобто

$$Q_{макс} = lim \frac{k}{k \cdot t_x + 1} = 1. \tag{1.7}$$

Характер витрат праці на виконання виробничих робіт залежить від двох чинників: рівня автоматизації основних і другорядних операцій; ступені безпосередньої участі людини у виробничому процесі як наслідок дії першого чинника.

На основі вище наведеної математичної моделі і логічного міркування можна зробити висновок: чим вище рівень автоматизації виробництва, отже менші витрати ручної праці працівника на виконання допоміжних робіт, тим вище продуктивність праці, тим економічніше виробництво, вище його ефективність і краще якість продукції (при автоматизації виробничих процесів досягається вища точність і чистота обробки виробів, що виготовляються).

Метод моделювання є незамінним засобом і при прогнозуванні різних явищ, де неможливо точно аналізувати всі обставини або проводити спостереження і вимірювання. Неповним початковим даним тоді даються оцінки досвідченими експертами. У результаті колективних обговорень і доповнення професійних знань експертів можна сформувати досить точні оцінки досліджуваного явища, на базі яких можна сконструювати досить добре відповідну дійсності модель явища.

Наприклад, перш ніж формувати зміст навчання фахівців заданої кваліфікації і профілю, потрібно на основі експертних оцінок розробити прогностичну професійну модель фахівця. У результаті повинні отримати віддзеркалення науково обґрунтовані дані про найбільш ймовірні тенденції розвитку відповідної галузі науки, техніка, виробництва, а також достатньою мірою деталізовані вимоги до суспільно-політичних і професійних якостей, якими повинен володіти фахівець, що ідеалізується, оптимально вписується в умови не тільки сучасного виробництва, а й виробництва майбутнього.

За допомогою моделей можна встановлювати і описувати компоненти об'єкту, що вивчається, взаємозв'язок між ними, робити висновки про управління об'єкту і прогнозувати його розвиток. Тим самим моделі можуть характеризувати різні процеси, структури і зв'язки. Ці моделі завжди спеціалізовані і складаються з ряду атрибутів, взаємозв'язок між якими для досліджуваного явища визнається постійним. Модель в чомусь схематизувала явища дійсності, від деяких конкретних властивостей відмовляються. Тому вона завжди застосовується для опису тільки окремих сторін конкретних явищ за певних умов.

Наприклад, за допомогою моделі, сконструйованої на основі експериментів, можна описати і наочно представити структуру компонентів технічного мислення студентів коледжів та учнів загальноосвітніх шкіл, взаємозв'язок між ними, показати основні шляхи і засоби для утворення технічного мислення окремих закладів освіти, що вчаться прогнозувати основні напрями розвитку технічного мислення цих учнів у зв'язку з технічним прогресом в народномугосподарстві і розвитком систем придатної освіти.

У педагогіці моделі можуть виступати: як засіб наукового дослідження, як предмети дослідження та як засоби діяльності. Все більше застосовується математичне і кібернетичне моделювання для пояснення суті процесу навчання. Види цих моделювань розширюються у зв'язку із застосуванням в науці дослідницької і обчислювальної техніки.

Якщо моделі в педагогіці використовуються як засоби навчання, вони є для учнів предметом дослідження. Такими моделями є моделі різних машин, механізмів, макети і т. д.

Моделі, навчання, що є засобом, використовуються часто і як засоби діяльності учнів (якщо у них формуються, наприклад, способи вирішення технічних завдань, ми маємо справу з особливим видом моделювання).

Але в педагогічних дослідженнях не можна перебільшувати роль моделювання, а особливо роль математичного моделювання. Хоча в емпіричних дослідженнях зараз дуже модно використовувати термін *«модель»* і кожне систематичне уявлення вважають моделлю, не скрізь і не завжди для пояснення педагогічних явищ доцільно конструювати різного виду моделі. Сконструйовану модель слід вважати вдалою, якщо з її допомогою можна пояснити розкривані експериментальним шляхом факти або якщо вона дозволяє передбачити існування ще не відкритих фактів і дати привід для цілеспрямованих пошуків. Моделі, що використовуються в педагогіці для опису процесу навчання відображають тільки просту структуру навчання, що гранично схематизована. Невелике збільшення числа чинників, що враховуються, в цьому процесі або введення нелінійних залежностей робить модель надзвичайно складною, такою, що важко піддається математичному дослідженню і є практично непридатною.

Метод математичного моделювання в педагогіці буде успішним лише в тому випадку, якщо дослідник ретельно перевірить:

чи зберігаються в моделі ті сторони і особливості процесу, які є предметом досліджень;

чи дотримані при зборі фактичні додаткові умови, які забезпечують використання даного методу до аналізу цих даних;

чи не відбулося при конструюванні моделі розповсюдження отриманих висновків за її цінність і контрольованість.

Конструювання моделей – це творча і часто дуже важка робота, виконання якої відбувається у декілька етапів. У дослідженнях з педагогіки складання моделей не повинне стати кінцевою ціллю, а моделі повинні сприяти ходу розкриття важливої наукової проблеми. Їх складання і використання повинно бути комплексне зв'язане з іншими методами дослідження педагогічних явищ.

У 1860 році Г. Т. Фехнер зробив першу практичну спробу математичного моделювання у психології. Для опису зв'язку між силою подразника та інтенсивністю відчуттів він використав модель логарифмічної функції. У 1885 році Г. Еббінгауз побудував математичну модель – «криву забування», яка описує зв'язок між часом збереження у пам'яті інформації та її кількістю.

У 1907 році О. Щукарєв, а у 1908 році Т. Робертсон побудували свої педагогічні моделі, так звані *«криві навчання»*. Згодом для опису певних аспектів навчання були запропоновані найрізноманітніші математичні моделі: експоненціальної функції, гіперболічної, параболічної та низку інших.

Суть цього моделювання полягає в тому, що з різних видів відомих у математиці функціональних залежностей обирають таку, яка дозволяє достатньо близько описати експериментальні дані. Завдяки цьому є можливість математично описати певні прості педагогічні закономірності. Однак водночас він обмежений у своїх можливостях. До того ж при його використанні припускається, що структура процесу, який описується, відповідає певній математичній структурі. А оскільки доведення істинності такого припущення немає, то відповідно немає гарантій, що модель буде придатною і в умовах, які відрізнятимуться від даного конкретного досліду, чи для явищ іншого типу.

У 30-40 роках минулого століття, як самостійна дисципліна сформувалась теорія автоматичного управління, а згодом – кібернетика, що зумовило принципово новий підхід до моделювання психолого-педагогічних явищ. Поштовхом для подальшої розробки моделей навчання стала робота Мак-Калока і Піттса (1943 рік), присвячена аналізу абстрактних нервових розгалужень. Слідом за нею з’явилося ще кілька статей аналогічного змісту, а також праці з галузі абстрактних автоматів. Абстрактні нервові розгалуження й абстрактні автомати можна розглядати як своєрідні найпростіші моделі нервових зв’язків у мозку людини. Спираючись на ці праці, перші спроби моделювання процесу вироблення умовних рефлексів у тварин, а також моделювання певних найпростіших елементів навчання зробили такі вітчизняні вчені, як А. А. Ляпунов, Г. А. Шестопал, Ю. А. Шрейдер, М. Л. Цетлін, а також зарубіжні – Р. Буш і Ф. Мостеллер, У. К. Естес, П. Суппс, Р. Аткінсон та інші. Провідна ідея цих спроб – ймовірнісний характер процесів навчання та виховання.

Виходячи з наявних даних про природу явищ, що вивчаються, і на підставі логічних міркувань дослідник будує, так би мовити, гіпотетичну модель, а згодом застосовує її до фактичних результатів досліду.

Однак освітній процес та різні його закономірності ще не вивчено настільки, щоб можна побудувати таку модель, яка б цілком і правильно могла відображати перебіг цього процесу, враховуючи вплив різноманітних зовнішніх факторів.

У 50-х роках 20 століття. вчені Р. Буш та Ф. Мостеллер здійснили кількісний опис експериментальних досліджень навчання і пристосування живих організмів, придатний для практичного використання. Варто зазначити, що розробку своїх моделей вони здійснили з позицій біхевіоризму стосовно найпростіших випадків навчання.

Учені Р. Буш і Ф. Мостеллер детально розробили математичний апарат для опису і дослідження простих проявів навчання. Проілюструємо метод на простому прикладі.

Наприклад, студент повинен самостійно в позааудиторний час виконати певне завдання. Якщо йдеться про старанного студента, то можна сподіватися, що ймовірно він це завдання виконає. Щодо нестаранного студента, то цілком ймовірне протилежне. Однак може статися протилежна ситуація. Старанний студент з якихось поважних причин не виконає завдання, а лінивий у пориві ентузіазму візьметься за нього. Отже, кожний з двох можливих варіантів (виконав – не виконав) має для кожного студента певну ймовірність. Зрозуміло, що у наведеному прикладі багато залежить від викладача. Якщо він вимогливо ставиться до студентів і систематично перевіряє виконання поставлених завдань, то природно припустити, що ймовірність виконання завдання другим студентом зростатиме завдяки постійній вимогливості. І, навпаки, не виключено, що поведінка викладача може бути такою, що навіть старанний студент припинить виконувати завдання. Як бачимо, під дією зовнішніх впливів розподіл ймовірностей змінюється.

Принципова перевага такого роду моделей полягає в тому, що вони дають математичну схему, пристосовану до опису й аналізу багатьох різновидностей і властивостей навчання в різних ситуаціях.

Досить вдалою виявилась спроба створення ймовірнісної моделі програмованого навчання, здійснена в 60-х роках минулого століття працівниками інституту кібернетики і психології О. М. Довгялло, Г. О. Баллом, Ю. І. Машбицом.

Вважають, що позитивним у програмованому навчанні є: виділення головного, істотного у навчальному матеріалі, забезпечення оперативного контролю за ходом засвоєння знань, логічна послідовність у засвоєнні знань, можливість працювати в оптимальному темпі і здійснювати самоконтроль, можливості індивідуалізації навчання, а негативним – зведення ролі вчителя до інструктора, відсутність можливості для розвитку творчості учнів.

Як бачимо, у програмованому навчанні кожна наступна порція матеріалу однозначно визначається попередньою порцією і відповіддю учня. Це дозволяє описати роботу певної навчаючої машини (популярні на початку 70-х р., так звані *«паперові навчаючі машини»* - програмовані підручники, програмовані збірники задач тощо) як функціонування автомата, вхідними сигналами якого є відповіді учня, а вихідними – порції навчального матеріалу.

Вивчаючи певний матеріал, учень вводить у машину різні відповіді – і правильні, і помилкові. Це дозволяє кожну відповідь учня розглядати як випадкову подію, а послідовність таких подій – як ймовірнісний процес. Оскільки перехід до наступної частини навчального матеріалу залежить від відповідей учня на попередню, то послідовність цих частин також може бути представлена як ймовірнісний процес.

Отже, формальну сторону програмованого навчання описує ймовірнісний процес, стани якого - частини навчального матеріалу, а ймовірності переходу з одного стану в інший – це ймовірності відповідей учня. Моделлю цього процесу може бути добре відомий у математиці Маяковський ланцюг з поглинаючим станом. Стани цього ланцюга – це частини навчального матеріалу, ймовірності переходу з одного стану в інший – ймовірності відповідей учня. На основі такого підходу обґрунтовують ймовірнісну модель програмованого навчання.

Крім цього, у моделі вводиться поняття ймовірності засвоєння порції навчальної програми, вважаючи, що у випадку конструювання відповідей ймовірність засвоєння дорівнює ймовірності правильної відповіді, а у випадку вибіркових відповідей ймовірність правильної відповіді вища, ніж ймовірність засвоєння. При цьому використовується поняття ймовірності вгадування.

Запропонована модель дозволила математично описати і встановити кількісні співвідношення між такими важливими характеристиками навчальних програм, як схема програми, тип відповіді, розмір кроку, міра допомоги. Модель уможливлює розв’язання, наприклад, такого завдання: нехай за допомогою розгалуженої програми, складеної за певною схемою, належить навчити за певний час певний контингент учнів. Цей контингент задається середньою для всіх учнів ймовірністю засвоєння. Необхідно знайти таку міру допомоги, за якої середній час навчання не перевищуватиме заданий. Аналогічні завдання можна розв'язати і для заданої кількості помилок тощо.

Процес взаємодії учня з навчальною програмою може бути представлений за такою схемою. Приміром, учень – автомат А. Навчальна програма, або точніше, її окремі дози – автомат Б. Взаємодія між ними – це гра, в результаті якої кожен з автоматів А і Б прагне до виграшу. Виграшем автомата Б буде заборона переходу суперникові – автоматові А у новий стан (+1), тобто заборона учневі переходу до вивчення нової дози матеріалу, і, отже, одержання незадовільної оцінки. Виграшем автомата А вважається успішний перехід до нової дози матеріалу, що свідчить про успішне опанування попередньої, і одержання нової частини знань.

Модель уможливлює попередньо оцінити середній виграш автомата А при його заданих станах, визначити максимально можливий виграш цього ж автомата за даних умов, оцінити параметри (стани) автомата Б та розв’язати деякі інші питання.

У наш час знайдені методи, що дозволяють пов'язувати між собою системи понять математики і інших наук та пояснювати на їх основі достатньо складні явища.

Так, наприклад, теорія матриць і графів є засобом, за допомогою якого можна дуже ефективно аналізувати структуру навчального матеріалу, отримати хороший огляд про взаємні зв'язки соціальних явищ. Якщо теорія матриць розглядає комплексні явища в чистому вигляді і статичному, то теорію рівнянь можна успішно застосовувати при дослідженні динаміки явищ. Якщо дані процеси неперервні, тоді рівняння, особливо диференціальні, найкращий засіб формулювання цих зв'язків. Диференціальні рівняння дозволяють добре описувати також кібернетичний зворотний зв'язок освітнього процесу. Комбінуючи теорію матриць і теорію диференціальних рівнянь, можна конструювати моделі таких систем, які одночасно є комплексними та динамічними.

Наприклад, процес підвищення кваліфікації можна представити у вигляді дискретної функції, оскільки практично він йде як би стрибками (привласнення розряду). Розряд працівника на початку року позначається $i+1$ або $i+2$. При цьому побудуємо модель, де у вертикальному стовпці будуть розряди на початку року (початковий стан**),** а в горизонтальному рядку – розряди наприкінці року або п'ятирічки (нова кваліфікація).

Швидкість підвищення кваліфікації (v) для даної сукупності працівників може бути визначена за формулою:

$$v_k = \frac{\sum(r_{i+1} - r_1)}{\sum r_i}.$$

Поки що використання методів математичного моделювання в педагогічних дослідженнях впроваджується з неабиякими труднощами, пов’язаними із складністю педагогічного процесу, недостатнім вивченням психічної діяльності людини, а то й недосконалістю сучасного математичного апарату.

І все ж коректне використання математичних методів у дослідженнях вже нині відкриває великі перспективи перед педагогічною теорією та практикою.

Адже математичним та статистичним методам, моделюванню надається істотне значення у розвитку досліджень в галузі суспільних наук, зокрема в педагогіці та психології. А тому широке застосування цих методів для дослідження проблем теорії навчання і виховання сприятиме підвищенню рівня цих наук.

Тема 2.

Основні поняття, що використовуються в математичній обробці даних педагогічних досліджень

Ознаки та змінні.

Ознаки та змінні – вимірюванні психолого-педагогічні явища. Такими явищами можуть бути: час розв'язку задач, кількість допущенних помилок, показник інтелектуальної лабільності, кут повороту під час бесіди, інтенсивність агресивних реакцій, показник соціометричного статусу і т. д.

Поняття ознаки та змінної використовуються як такі, що замінюють одне одного. Педагогічні та психологічні змінні є випадковими величинами, тому що заздалегідь невідомо, яке значення вони приймуть.

Математична обробка – це робота з значеннями ознаки, що отриманні в піддослідних під час педагогічного дослідження. Такі індивідуальні показники називають «варіантами», «датами», «спостереженні значення». Значення ознаки визначають за допомогою шкал вимірювання. Про них ми згадували раніше.

Розподіл ознаки. Параметри розподілу.

Розподілом ознаки називається закономірність появи різних її значень. В педагогічних дослідженнях частіше посилаються на нормальний розподіл. Нормальний розподіл характеризується тим, що крайні значення ознаки зустрічаються достатньо рідко, а значення ближчі до середини достатньо часто. *Нормальним розподіл* називають тому що досить часто зустрічається в природничих науках. Графіком нормального розподілу є дзвінковоподібна крива (крива Гаусса).

Параметри розподілу – це його числові характеристики, що вказують де в середньому розміщені значення ознаки, наскільки ці значення стійкі і як часто з'являються певні значення ознаки. Найбільш практично важливими параметрами є математичне сподівання, дисперсія та показники асиметрії і ексцеса (мода та медіана). У реальних педагогічних явищах ми оперуємо не параметрами, а їх оцінкою (наближеними значеннями). Тому чим більша вибірка тим ближче оцінка параметра до його істинного значення.

Формулювання гіпотез систематизує припущення дослідника та представляє їх в чіткому та лаконічному виді. Статистичні характеристики (середні) дають дослідникові уявлення про сукупність, що вивчається, але не дозволяють робити узагальнення. Для формування висновків на основі знайдених характеристик потрібно їх порівнювати з такими ж, що були знайдені раніше, та зустрічаються в інших сукупностях чи теоретичних характеристиках. При цьому висувається статистична гіпотеза про суттєві або несуттєві відмінності, які потім за певним критерієм підтверджується або спростовується.

Статистичні гіпотези та критерії.

Під *статистичною гіпотезою* розуміється будь-яке припущення про властивості випадкових величин або подій. Статистична гіпотеза є – статистичним виразом загальнотеоретичної гіпотези результатом формалізації висунутою у дослідницькій роботі змістовної гіпотези. Вона виражається мовою теорії ймовірностей і математичної статистики.

Статистичні гіпотези, що виникають в педагогічних дослідженнях, ділять *на чотири основні групи:*

1) гіпотези про типи ймовірних законів розподілу випадкових величі;

2)гіпотези про властивості тіл, або інших числових параметрів;

3)гіпотези про стохастичну (ймовірною) залежність двох або більш ознак (чинників):

4)гіпотези про рівність або відмінність законів розподілу випадкових величин.

Для перевірки гіпотези обчислюють емпіричний критерій і порівнюють його з теоретичним, числові значення якого дані у відповідних таблицях. Статистичні методи ніколи не дозволяють доводити помилковість гіпотези з абсолютною точністю, завжди є деяка можливість помилки, яку характеризують рівні значущості (рівні достовірності). Межу достовірності підтвердження або спростування гіпотези встановлюють відповідно до характеру об'єкта дослідження.

Гіпотезу, що підлягає контролю називають ***гіпотезою частот*** і ***нульовою гіпотезою та*** позначають її H_0, *порівнювану ж з нею гіпотезу називають* ***альтернативною*** і позначають H_1.

Відповідним обчисленнями визначають, чи відкинути H_0 і тим самим прийняти H_1, або прийняти H_0 тим самим відкинути H_1.

Наприклад, нульовою гіпотезою (може бути твердження, що різниця в результатах роботи учнів, що навчаються в двох різних групах залежить не від методів навчання, а від інших випадкових причин; альтернативною гіпотезою може бути твердження, що відмінність в результатах роботи цих учнівських груп обумовлена не випадковими чинниками, а цілеспрямованим застосуванням різних методів навчання.

Види статистичних гіпотез:

1. *Нульові* – це гіпотеза про відсутність відмінностей. Це те, що ми хочемо спростувати, коли перед нами стоїть завдання довести значущість відмінностей H_0.
2. *Альтернативні* – це гіпотеза значущість відмінностей H_1. Це те, що ми хочемо довести (експериментальна гіпотеза). $X_1 - X_2 = 0$.(значення ознаки, що співставляються).
3. *Напрямлені* – H_0: X_1 не більше X_2; H_1: X_1 більше X_2.
4. *Не напрямлені* – H_0: X_1 не відрізняється від X_2; H_1: X_1 відрізняється від X_2.

Якщо ми хочемо довести, що в експериментальній групі (Е) пройшли якісні зміни на відміну від контрольної групи (К), то ми маємо сформулювати напрямлену гіпотезу. Перевірка гіпотези здійснюється за допомогою критеріїв статистичної оцінки відмінностей.

Статистичним критерієм називається вирішальне правило, що забезпечує надійну поведінку, тобто прийняття істинної та відхилення хибної гіпотези з високою ймовірністю. Статистичні критерії означають метод визначення певного числа і саме це число. Коли ми говоримо, що достовірність відмінностей визначалась за критерієм χ^2, то маємо на увазі, що використовували метод χ^2 для визначення цього числа. Коли ми кажемо, що $\chi^2 = 12{,}676$, маємо на увазі конкретне число визначене за методом χ^2. Це число називається емпіричним значенням критерію. Порівнявши емпіричне та критичне значення критерію ми можемо підтвердити чи спростувати нульову гіпотезу. Наприклад, якщо $\chi^2{}_e > \chi^2{}_{кр}$, H_0 відкидається.

Як правило, щоб визнати достовірність відмінностей (перевагу експериментальній методиці) необхідно, щоб емпіричне (експериментальне) значення перевищувало критичне (теоретичне). Хоча є критерії де потрібно дотримуватися протилежного правила (критерій Манна-Уітні чи критерій знаків).

У більшості випадків одне й те саме емпіричне значення критерію може виявитись достовірним чи недостовірним у залежності від кількості спостережень у досліджуваній вибірці (кількості учнів) *n* або так званої кількості ступенів вільності *N*. Ступінь вільності *N* дорівнює різниці між числом класів варіаційного ряду і числом умов при яких він був сформований. *Наприклад*, у нас є 50 учнів 3-х класів. Вміють працювати на комп'ютері 20 учнів. Недосконало працюють на комп'ютері – 20 учнів. Виходить, що решта – 10 учнів, не вміють працювати на комп'ютері. Якщо ми втратимо дані про те скільки учнів не вміють працювати на комп'ютері, то це легко можна визначити. Отже, ми вільні тільки в 1-му та 2-му класах (*N=3-1=2)*.

Критерії діляться на *параметричні та непараметричні.*

Параметричні – це критерії, до складу розрахункової формули яких входять параметри розподілу тобто середні та дисперсії (t-критерій Стьюдента, критерій F).

Непараметричні – це критерії, до складу розрахункової формули яких не входять параметри розподілу, іде оперування частотами чи рангами (критерій Розенбаума, критерій Вілкоксона). Кожен вид має свої недоліки.

Рівень значущості – це ймовірність того, що ми вважаємо відмінності суттєвими, а вони насправді є випадковими. Коли ми кажемо, що відмінності достовірні на 5% -му рівні достовірності (чи $p \leq 0{,}05$), то маємо на увазі, що ймовірність того, що вони недостовірні складає 0,05.

Коли ми кажемо, що відмінності достовірні на 1%-му рівня достовірності (чи $p \leq 0{,}01$), то маємо на увазі, що ймовірність того що вони недостовірні складає 0,01.

Рівень достовірності – це ймовірність відхилення нульової гіпотези, в той час коли вона вірна.

Правило відхилення H_0 і прийняття H_1.

Якщо емпіричне значення критерію дорівнює критичному значенню, що відповідає $p \leq 0{,}05$ або перевищує його, то H_0 відхиляється, але ми ще не можемо прийняти H_1.

Якщо емпіричне значення критерію дорівнює критичному значенню, що відповідає $p \leq 0{,}01$ або перевищує його, то H_0 відхиляється, і H_1.

Виключення: критерій Вілкоксона і критерій Манна-Уітні.

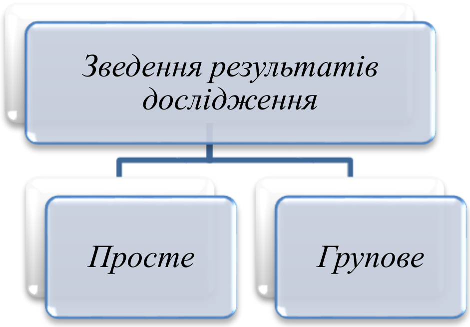
Потужність критерію – це його здатність виявляти відмінності, якщо вони є. Або це здатність відхилити нульову гіпотезу про відсутність відмінностей, якщо вона невірна. *М=1-м.*

Етапи роботи при статистичному досліджені:

- ✓ побудова моделі для того щоб за допомогою цифр описати суттєві властивості об'єкта що спостерігається – будуємо математичну модель (таблиця, матриця)яку на другому етапі заповнюємо цифрами;
- ✓ збір даних за допомогою спостереження з ціллю отримання кількісної інформації поро явище;
- ✓ зведення отриманих даних, знаходження узагальнених цифрових даних та їх обробка – працюємо тільки з формальною моделлю, визначають її межі;
- ✓ аналіз та інтерпретація даних – інтерпретація приводить результати дослідження у відповідність з реальним об'єктом дослідження.

Математичні методи застосовуємо на перших трьох етапах.

Зведення результатів дослідження – це етап на якому від даних зібраних про окремі елементи сукупності до даних, що характеризують сукупність як єдине ціле.



Просте – сукупність розглядається як аморфна маса.

Групове – досліджувану сукупність розділяють на групи за певної ознакою (вік, стать, середня успішність і т. д.). При такому зведенні групи можуть мати не рівну кількість елементів.

Групувати можна за *кількісними та якісними ознаками*:

Кількісні – результати тестів, вік, швидкість роботи, точність. Для них можна ввести одиницю виміру. А результат вибудувати по зростанню чи спаданню.

Якісні – не складають послідовність. Стать, національність, тип нервової діяльності, соціальне положення.

Групування буває:

Просте – розбиття на групи відбувається за однією ознакою. *Наприклад*, викладачі групи групуються за стажем роботи.

Комбіноване – розбиття на групи відбувається більше ніж за однією ознакою. *Наприклад,* викладачі групи групуються за стажем роботи і статтю.

У педагогічних дослідженнях групують не більше ніж за трьома ознаками.

Представлення результатів дослідження.

Результати (кількісні дані), що отримані вході дослідження можна представляти у вигляді:

1) *перерахувати у тексті роботи, якщо даних мало;*
2) *у вигляді таблиць – вони дозволяють представити дані у стислому вигляді, якщо таблиці грамотно складені то про кожне число можна отримати пояснення в таблиці і не потрібно повторюватися в тексті роботи;*
3) *у вигляді графічних зображень (графіків, діаграм, полігонів частот, кривими розподілу).*

Таблиця представляє дані і виступає як засіб їх інтерпретації. Правильно вибраний вид таблиці допоможе краще розкрити суть цих даних і зв'язок між ними.

До графічних методів зведення та представлення результатів дослідження відносять графік, гістограму, діаграму і т.д.

Тема 3.

Середні показники результатів дослідження та їх застосування

Середні об'єму і середні положення.

Для найбільш точного опису статистичних сукупностей використовують різні середні показники. Найчастіше поняття середнього не пов'язане з яким-небудь визначеним числом, а представляє собою загальну категорію мислення. *Наприклад*, середній учень, середній вчитель, середня успішність. Намагаючись точніше означити вище названі поняття, вимірювати і оцінювати їх можливо об'єктивними методами, можна присвоїти їм точне числове значення.

При багатьох педагогічних дослідженнях шляхом перерахунку можна визначити частоту подій. Це і є однією з форм квантифікацій педагогічних явищ. Наступною сходинкою квантифікації є визначення різних середніх (середніх значень) сукупностей. Оскільки середні являють собою квантитативні, в тому числі отримані в результаті вимірювання і обчислення величини, то у сукупності може бути багато середніх.

Середні можна розділити на дві групи:

1. *Середні об'єму* – середні, числове значення яких змінюється тоді, коли змінюється значення довільного елемента сукупності. В якості об'ємного середнього в педагогічних дослідженнях приймається звичайне середнє арифметичне.
2. *Середні положення* (структурні середні). Змінюються тоді, коли проходять зрушення в структурі сукупності (змінюється їх кількість, послідовність). В якості середнього положення використовують **медіану** (середній член впорядкованого розподілу частоти, по обидві сторони якого залишається рівна кількість членів) і **моду** (найбільш часто повторюване значення в статистичному розподілі частоти). Рідше використовують **квартилі** (розподіл частоти на чотири частини, в кожній з яких є рівна кількість членів ряду), **децили** (ділять статистичний ряд на десять рівних частин).

Значення середніх при вивченні будь-яких квантитативних явищ засноване на законі великих чисел і середні будуть достовірним джерелом інформації тільки тоді, коли при їх обчисленні враховували цей закон.

Обчислення середнього арифметичного. Середнє арифметичне обчислюється за формулою:

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{\sum fx}{N}, \tag{3.1}$$

де x – величина окремих елементів сукупності; f – частота появи окремої величини; N – кількість членів сукупності (об’єм сукупності).

Чим більша частота величини кожного елемента сукупності, тим більше вона приймає участь в утворенні середньої величини середнього сукупності. Тому середня, обчислена таким чином, називається **зваженим середнім арифметичним** чи **середньозваженою величиною**.

Для обчислення середньозваженої величини бажано внести в таблицю величини всіх членів і відповідних їм частот.

Якщо кожен варіант розподілу частоти з’являється тільки один раз, то маємо формулу:

$$\bar{x} = \frac{\sum x}{N}, \tag{3.2}$$

яку називають **звичайною арифметичною середньою**.

Часто зустрічаються помилки в дослідженнях із середніми арифметичними (при обчисленні середньої успішності, при обчисленні результатів експерименту в тому випадку, коли на основі вже знайдених середніх знаходять будь-яке нове середнє).

Замість того, щоб користуватися середньозваженою величиною, обмежуються простою середньою арифметичною, що призводить до спотворених результатів. При обчисленні середньої оцінки школи, училищ складають звичайну середню оцінку всіх груп і ділять отриману суму на число груп. Однак, це хибний алгоритм. У такому випадку потрібно користуватися середньозваженою величиною, причому потрібно враховувати різну кількість учнів (студентів) у кожній групі. За допомогою наступної формули можна обчислити середньозважену величину:

$$\bar{x} = \frac{n1x1 + n2x2 + \cdots + n\text{к}x\text{к}}{n1 + n2 + \cdots + n\text{к}}. \tag{3.3}$$

Обчислення середнього інтервального ряду розподілу. Середня арифметична обчислюється в тому випадку, коли елементи сукупності розподілу частоти згруповуються в інтервали. У такому випадку обчислюються середні величини всіх інтервалів. Зазвичай ми не маємо уявлення про те, як розподілити елементи в самих інтервалах. Тому за середню величину інтервалу беруть звичайне середнє арифметичне інтервалів. На основі першого інтервалу визнається середня величина інтервального ряду (табл. 3.1): $\frac{1+5}{2} = 3$.

Таблиця 3.1

Результати контрольних робіт

Оцінка x_1	*f*	*Середня величина інтервалу x_1*	*$f(x_1)$*
1-5	2	3	6
6-10	6	8	48
11-15	13	13	169
16-20	10	18	180
21-25	4	23	92
	$\sum f=35$	*$\sum fx_1=495$*	

*Наприклад, р*езультати контрольних робіт, оцінені в 25-бальній системі. Для обчислення середньої оцінки величина кожного інтервалу множиться на частоту інтервалу. Потім знаходять суму похідних і ділять на загальне число членів сукупності.

$$\bar{x} = \frac{\sum f\, x1}{\sum f} = \frac{495}{35} = 14{,}1.$$

Знаходження медіани і моди. Медіану і моду легко знайти на графіках, що зображають розподіл частоти. Як уже говорилося, медіану, квартили і децили легко знайти на відсотковій кривій кумулятивної частоти.

Графічно медіана являє собою абсцису вертикальної прямої, яка розділяє площину полігону частоти пополам (рис. 3.1 крива Б). При симетричному розподілі частоти величина середнього арифметичного і величина медіани рівні (рис. 3.1 крива А).

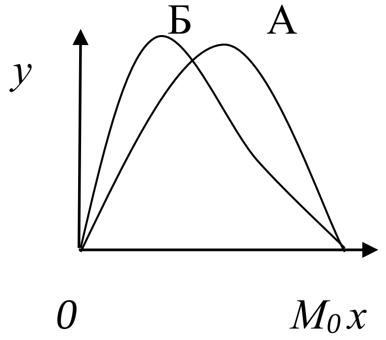


Рис. 3.1.

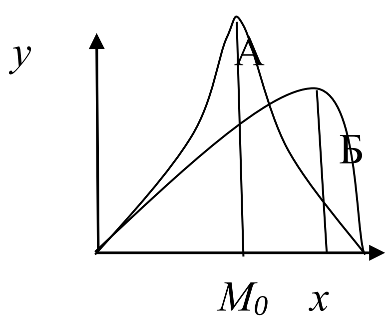


Рис. 3.2.

Моду можна знайти за допомогою полігона частот (якщо розподіл частот не має інтервалів) чи гістограми (якщо розподіл частоти має інтервали). Як уже зазначалося, **модою** (домінантою) називається величина, яка найбільш часто зустрічається (домінуює).

Отож, мода (M_0) являє собою абсцису максимуму полігона частоти чи абсцису максимуму диференціальної кривої частоти (рис. 3.2). Доволі легко знайти моду при розподілі частоти на інтервали. У такому випадку розподіл представляється у вигляді гістограми (рис. 3.3).

Рис. 3.3.

Перш за все, в ній знаходиться так званий модальний інтервал (інтервал з найбільшою частотою). Далі, вершини стосовно відношенню до інтервалу. Величиною моди є абсциса точки перетину діагоналей.

Обчислення медіани. Якщо ранговий ряд розподілу частоти має непарну кількість членів, то легко знайти медіану. Медіана чи центральна величина ряду – це величина члена, що знаходиться в середині ранжируваного ряду.

Наприклад, 3, 6, 7, 9, $\underbrace{11,}_{\text{медіана}}$ 14, 16, 17, 19.

У середині даного ряду знаходиться число 11, а отже і воно є медіаною. Порядковий номер медіани в ряду вираховуємо наступним чином *(N+1)/2*, де *N* – число членів ряду. Якщо в ранговому ряді парне число членів, то при використанні вищенаведеної формули порядковий номер медіани є числом, що знаходиться між двома середніми членами.

Наприклад, маємо ранговий ряд: 4, 6, 7, 9, 11, 13, 15, 16. Порядковий номер медіани (8+1)/2=4,5. Практично це означає, що в середині ряду знаходиться два члени. Для обчислення однієї величини медіани вираховується звичайне середнє арифметичне двох середніх членів. Отримане нами число і буде величиною медіани.

$$M_e = \frac{9+11}{2} = 10.$$

Медіана в інтервальному ряді визначається за формулою:

$$M_e = x_{M_e} + \frac{k\left(\frac{N}{2} - F\right)}{f_{M_e}}, \tag{3.4}$$

де x_{M_e} – значення нижньої границі медіанного інтервалу; $к$ – довжина медіанного інтервалу; N – число членів сукупності; F – сума частот, попередніх медіанному інтервалу; f_{M_e} – частота медіанного інтервалу.

Обчислення моди. Якщо ми маємо справу з простим, доволі коротким статистичним рядом, то можемо знайти моду «на око» із впорядкованого ряду.

Наприклад, 8, 4, 5, 8, 7, 6, 8, 9, 10, 3, 11, 8. Частіше всього зустрічається число 8, отож, воно і буде модою.

Якщо ми маємо справу з розподілом частот, що не має інтервалів, то модою вважається числова величина, що зустрічається найчастіше.

На практиці розрізняють так звану істинну моду і наближену моду. Істинна мода являє собою точку найбільшої концентрації розподілу. Ця точка може і не співпадати з серединою інтервалу, так як в рамках інтервалу величини елементів сукупності не розподіляються рівномірно. Істинна мода наближається до наближеної моди в тому випадку, якщо сукупність розподілена на доволі короткі інтервали, полігон частот вирівняний і число членів сукупності достатньо велике. У такому випадку обчислюємо моду:

$$M_0 = 3 \cdot M_e - 2 \cdot \bar{x} \tag{3.5}$$

$$M_0 = \bar{x} - 3 \cdot (\bar{x} - M_e)$$

При обчисленні моди інтервального ряду користуються формулою, що базується на графічному методі обчислення моди.

Формула має вигляд:

$$M_0 = x_{M_0} + \frac{k(f_{M_0} - f_{M_{0-1}})}{(f_{M_0} - f_{M_{0-1}}) + (f_{M_0} - f_{M_{0+1}})}, \tag{3.6}$$

де x_{M_0} – початкове значення модального інтервалу; k – величина модального інтервалу; f_{M_0} – частота модального інтервалу; $f_{M_{0-1}}$ – частота інтервалу, попереднього модальному; $f_{M_{0+1}}$ – частота інтервалу, наступного за модальним.

Практичне застосування середніх. Середні широко застосовуються для характеристики даних, отриманих у результаті дослідження. Слід мати на увазі, що можливості середніх відносно обмежені.

Середнє арифметичне дає можливість:

- охарактеризувати досліджувану сукупність одним числом;
- порівняти окремі величин з середнім арифметичним;
- визначити тенденцію розвитку будь-якого явища;
- порівняти різні сукупності;
- обчислити інші статистичні показники, так як багато статистичних обчислень спираються на середнє арифметичне.

Слід відзначити, що середнє арифметичне не характеризує різносторонні сукупності. Воно характеризує тільки загальний рівень довільної сукупності, навколо якого коливаються величини елементів.

Необхідно звернути увагу на те, що середнє арифметичне залежить не тільки від індивідуальних величин елементів, але і від відносної частоти явища. Так, *наприклад*, середня оцінка класу відразу змінюється, якщо збільшиться або зменшиться число незадовільних оцінок.

Середнє арифметичне зручно використовувати для опису сукупності, якщо розподіл параметрів здійснено симетрично по відношенню до середини. У випадку асиметричного розподілу чи багатовершинного полігону частот середнє арифметичне не можна використовувати для опису сукупності. У таких випадках для характеристики сукупності краще використовувати моду.

Медіану використовують тоді, коли хочуть визначити точну середину ряду. Деякі інтервали особливо великої частоти в значній мірі можуть вплинути на середнє арифметичне. Перевагою медіани є те, що на неї такі екстрені інтервали не впливають.

Середнім арифметичним може бути така величина значення, яка в ряду може бути відсутньою або взагалі не існувати. Медіана завжди є одним із членів ряду розподілу чи дуже близькою для нього величиною. Графічно медіана являє собою абсцису вертикальної лінії, що розподіляє навпіл площу під кривою розподілу частот. При симетричному розподілі частот величина середнього арифметичного дорівнює величині медіани.

На основі медіани, а точніше на основі квартилі, іноді можна дати доволі хорошу характеристику структури ряду. Можна стверджувати, чи розподіляються члени ряду навколо середнього рівномірно, чи накопичення величин відбувається по зростаючим чи спадним інтервалам.

Модою користуються в тому випадку, коли хочуть швидко охарактеризувати сукупність на основі явища, що зустрічається найчастіше. Причому треба мати на увазі найзагальнішу величину: оцінка, що найчастіше зустрічається в звичайних умовах, розмір одягу, що зустрічається найчастіше у дітей певного віку.

Наприклад, при виготовленні дитячого взуття за основу береться мода (розмір дитячого взуття, що найчастіше зустрічається), а не середнє арифметичне розмірів дитячої ноги.

Отже, слід відмітити, що, досліджуючи при розподілі частот взаємні величини середнього арифметичного, моди і медіани, можна зробити доволі цікаві висновки про структуру розподілу частот. Порівняння середніх одного розподілу частот з середнім іншого розподілу частот дає можливість порівняти особливості таких розподілів.

Тема 4.

Варіація та допустимі межі результатів дослідження

Варіація значень елементів сукупності.

Варіація ознаки є властивістю статистичної сукупності і зумовлена дією безлічі взаємопов’язаних причин, серед яких є основні і другорядні. Основні формують центр розподілу, другорядні – варіацію ознаки, сукупна їх дія – форму розподілу.

Конкретні умови, в яких знаходиться кожний з об’єктів, що вивчається, а також особливості їхнього власного розвитку (соціальні, економічні тощо) виражаються відповідними числовими рівнями статистичних показників. Таким чином, варіація, тобто незбіг (неспівпадання) рівнів одного й того ж показника у різних об’єктів, має об’єктивний характер і допомагає пізнати сутність явища, що вивчається. Дослідження варіації в статистиці та педагогічних дослідженнях має велике значення, тому що величина варіації ознаки в статистичній сукупності характеризує її однорідність. У статистичній практиці для вивчення та вимірювання варіації використовують різноманітні показники варіації в залежності від поставленого завдання. До них відносяться розмах варіації, середнє лінійне відхилення, середній квадрат відхилень (дисперсія), середнє квадратичне відхилення і коефіцієнт варіації.

Розмах варіації (R) є найпростішим вимірювачем варіації ознаки. Це різниця між найбільшим і найменшим значенням ознаки:

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

Однак розмах варіації показує лише крайні значення ознаки. Повторюваність проміжних значень не враховується.

Середнє лінійне відхилення ($\bar{d}$) представляє собою середню величину із відхилень варіантів ознаки від їхньої середньої. Його можна розрахувати за формулою середньої арифметичної простої або зваженої в залежності від відсутності або наявності часток в ряду розподілу.

$$\bar{d} = \frac{\sum |x_i - \bar{x}|}{n}$$ - проссте середнє лінійне відхилення,

$$\bar{d} = \frac{\sum |x_i - \bar{x}| \cdot f_i}{\sum f_i}$$ - зважене середнє лінійне відхилення,

де x_s – i-тий варіант ознаки; f_s – питома вага i-того варіанту; n – кількість варіантів; $\bar{x}$ - середня арифметична величина.

Показник середнього лінійного відхилення знайшов широке застосування на практиці. За його допомогою аналізуються, наприклад, склад працюючих, ритмічність виробництва, рівномірність постачань матеріалів, розробляються системи матеріального стимулювання.

Треба пам'ятати, що алгебраїчна сума відхилень від середнього рівня дорівнює нулю, тобто середнє значення відхилення для будь-якої випадкової величини прямує до нуля. Тому в статистичних наукових дослідженнях для виміру варіації частіше використовують показник дисперсії.

Наприклад, розрахуємо середнє лінійне відхилення для дискретного ряду розподілу на основі даних таблиці 4.1.

Таблиця 4.1.

Розподіл вчителів середніх шкіл району за стажем роботи

Стаж роботи (x_i, роки)	К-ть вчителів до загальної кількості (f_i %)	$x_i f_i$	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$	$\lvert x_i - \bar{x} \rvert \cdot f_i$
8	14	112	-2	2	-28
9	20	180	-1	1	-20
10	30	300	0	0	0
11	24	264	1	1	24
12	12	144	2	2	24
Разом	100	1000	-	-	96

Розмах варіації за стажем складає: *R=12–8=4* (роки). Допоміжні розрахунки наведені в графах 3 – 6 таблиці 4.1.

Середній стаж роботи визначимо за формулою середньозваженої арифметичної: $\bar{x} = \frac{1000}{100} = 10$ (років).

Середнє лінійне відхилення стажу роботи вчителів середніх шкіл району складає: $\bar{d} = \frac{96}{100} = 0{,}96$ (роки).

Дисперсія(σ^2*)* - представляє собою середній квадрат відхилень індивідуальних значень ознаки від її середньої величини. Дисперсія розраховується за формулами простої середньої незваженої і зваженої відповідно:

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n}; \quad \sigma^2 = \frac{\sum(x_i - \bar{x})^2 \cdot f_i}{\sum f_i}.$$

На практиці застосовують більш просту формулу для розрахунку дисперсії: $\sigma^2 = \overline{x^2} - \bar{x}^2$.

Середнє квадратичне відхилення (σ).

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$$ (для незгрупованих даних);

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2 \cdot f_i}{\sum f_i}}$$ (для згрупованих даних).

Наприклад, розрахуємо дисперсію і середнє квадратичне відхилення для наступного ряду розподілу (табл. 4.2).

Таблиця 4.2

Розподіл магазинів міста за товарообігом в II кварталі 1998 року

Групи магазинів за величиною товарообігу (тис. грн.)	Число магазинів(f_i)	Середина інтервалу (x_i ,тис. грн.)	$x_i f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
40-50	2	45	90	-49,2	2420,64	4841,28
50-60	4	55	220	-39,2	1536,64	6146,56
60-70	7	65	455	-29,2	852,64	5968,48
70-80	10	75	750	-19,2	368,64	3686,40
80-90	15	85	1285	-9,2	84,64	1269,60
90-100	20	95	1900	0,8	0,64	12,80
100-110	22	105	2310	10,8	116,64	2566,08
110-120	11	115	1265	20,8	432,64	4759,04
120-130	6	125	750	30,8	948,64	5691,84
130-140	3	135	405	40,8	1664,64	4993,92
Разом	100	-	9420	-	-	39936,00

При розрахунку показників варіації за інтервальним рядом розподілу необхідно спочатку визначити середини інтервалів, а потім вже вести подальший розрахунок, розглядаючи ряд середин інтервалів як дискретний ряд розподілу.

Результати допоміжних розрахунків для визначення дисперсії та середнього квадратичного відхилення знаходяться в графах 2 – 6 таблиці 4.2. Середній розмір товарообігу визначається за середньозваженою арифметичною і складає: $\bar{x} = \dfrac{9420}{100} = 94{,}2$ тис. грн.

Дисперсія товарообігу $\sigma^2 = \dfrac{39936{,}00}{100} = 399{,}36$.

Середнє квадратичне відхилення товарообігу визначається як квадратний корінь з дисперсії: $\sigma = \sqrt{399{,}36} = 19{,}98 \approx 20$ тис. грн.

Для порівняння варіації різних ознак або однієї ознаки в різних сукупностях використовують *відносні показники варіації*.

Коефіцієнти варіації розраховуються як відношення абсолютних характеристик варіації $(\sigma, \bar{d}, R)$ до центру розподілу і часто виражаються відсотками. Отже:

1. *Коефіцієнт осциляції:* $V_R = \dfrac{R}{\bar{x}} \cdot 100\%$;
2. *Лінійний коефіцієнт варіації:* $V_{\bar{d}} = \dfrac{\bar{d}}{\bar{x}} \cdot 100\%$;
3. *Квадратичний коефіцієнт варіації:* $V_\sigma = \dfrac{\sigma}{\bar{x}} \cdot 100$.

У статистиці сукупності, що мають коефіцієнт варіації (V_σ) більший 30-35%, вважаються неоднорідними.

Приклад. За даними вибіркових обстежень домогосподарств, середньодушові витрати на харчування становили 80 гр. од.; на придбання промислових товарів – 35; дисперсії відповідно – 256 та 196. Порівняти ступінь варіації витрат домогосподарств на харчування та придбання промислових товарів можна за допомогою квадратичного коефіцієнта варіації:

витрати на харчування:

$$V_\sigma = \frac{\sqrt{256}}{80} \cdot 100 = 20\%;$$

витрати на придбання промислових товарів:

$$V_\sigma = \frac{\sqrt{196}}{35} \cdot 100 = 40.$$

Отже, ступінь варіації витрат на придбання промислових товарів значно вищий.

Тема 5.

Параметричні та непараметричні методи порівняння результатів дослідження

Параметричні методи порівняння результатів дослідження.

Критерії діляться на параметричні та непараметричні. Параметричні критерії у розрахунковій формулі містять середні та дисперсії (*t*- критерій Стьюдента, критерій *F* та інші). Непараметричні критерії у розрахунковій формулі не містять середні та дисперсії, а основані на опереванні частотами чи рангами (критерій *Q*- Розенбаума, критерій *Т*-Вілкоксона та інші).

Описані нижче методи порівняння середніх називаються **параметричними.** За їх допомогою порівнюються параметри *генеральних сукупностей* (середніх значень, диспепсій і **т.** д.). Параметричні методи вимагають кількісних вимірювань **за** шкалою інтервалів або шкалою відносин і нормального розподілу сукупності вибірки.

***F*-критерій Фішера (критерій порівняння дисперсій).** Вище розглядалося описання елементів однієї сукупності (ряд виміряних чисел), причому з'ясувалася ймвірність того, що яке-небудь окреме виміряне число належить до вибірки і визначалися межі достовірності якого-небудь параметра сукупності (середнього або стандартного відхилення).

У педагогічних дослідженнях досить часто доводиться порівнювати дві вибірки, що відносяться до однієї і тієї ж загальної сукупності. Тільки на підставі порівняння середніх, їх меж достовірності і варіювання не можна отримати достатньо точних даних для порівняння відмінності середніх вибірок.

На практиці при порівнянні двох середніх виходять із умови, що відмінність арифметичних середніх обох вибірок $\left(\overline{x_1}-\overline{x_2}=D\right)$, якщо вибірки відносяться до однієї і тієї ж спільної вибірки, дорівнює нулю (або в межах випадковості близька до нуля). Це і буде нульовою гіпотезою про різницю двох сукупностей вибірки однієї і тієї ж спільної сукупності.

Перш ніж приступити до дослідження достовірності відмінностей арифметичних середніх сукупностей, необхідно з'ясувати, чи відмінності дисперсій істотні чи ні.

Нехай дисперсія однієї сукупності вибірки з n_1 елементів дорівнює σ_{12}, а іншої сукупності вибірки з n_2 елементів дорівнює σ_{22}. Порівняємо дисперсії склавши їх відношення:

$$F_{emp} = \frac{\sigma_1^2}{\sigma_2^2} \qquad (5.1)$$

(дисперсії вибирають так, щоб відношення їх квадратів завжди було більше або рівне одиниці).

Потім знаходимо ймовірність нульової гіпотези зі ступенем вільності $n_1 -1$ *і* $n_2 -1$ в F – таблиці (табл. 5.1)**.** У заголовку таблиці (в першому рядку) знаходять ступінь вільності чисельника (більшої дисперсії) і у першому стовпці таблиці ступінь вільності знаменника. У таблиці 5.1 представлені значення F для p=5%, але використовуються і інші значення F, в основному, для p=1%. Якщо з'ясується, що дисперсії значно відрізняються $\sigma_1^2 >> \sigma_2^2$ $[F_{emp} > F_{krt}]$ і отримане при діленні відношення F_{emp} істотно відрізняється від отриманого в таблиці значення (F_{krt}), тобто якщо ймовірність нульової гіпотези нижче 5%, тоді зрозуміло, що сукупності відрізняються одна від одної. Необхідність порівнювати середні відпадає.

У педагогіці відмінності дисперсій мають велике значення. Припустимо, що під час проведення педагогічного експерименту в експериментальних і контрольних групах отримали однакові середні при оцінці рівня знань учнів (студентів), але дисперсії в цих класах (групах) є суттєво різними. Це означає, що в одному класі (групі) оцінки були тісно зосереджені навколо середнього арифметичного (розподіл оцінок був однорідним), в іншій групі було багато як високих (позитивних) так і низьких (негативних) оцінок. Тобто результати розсіялися (розподіл оцінок був гетерогенним). Якщо дисперсії істотно не відрізняються (ймовірність нульової гіпотези $p>5\%$), то для оцінки значущості відмінності середніх застосовують ***F***- критерій.

Таблиця 5.1

Таблиця – критерій (*F – значення p=5%*)

Число ступеня вільності знаменника	Число ступеня вільності чисельника								
	1	2	3	4	5	6	12	24	∞
1	161,0	199,5	215,7	224,6	230,2	234,0	244,0	249,0	254,3
2	18,5	19,2	19,2	19,3	13,3	19,3	19,4	19,5	19,5
3	10,1	9,6	9,3	9,1	9,0	8,9	8,7	8,6	8,5
4	7,7	6,9	6,6	6,4	6,3	6,2	5,9	5,8	5,6
5	6,6	5,8	5,4	5,2	5,1	5,0	4,7	4,5	4,4
6	6,0	5,1	4,8	4,5	4,4	4,3	4,0	3,8	3,7
7	5,6	4,7	4,4	4,4	4,0	3,9	3,6	3,4	3,2
8	5,3	4,5	4,1	3,8	3,7	3,6	3,3	3,1	2,9
9	5,1	4,3	3,9	3,6	3,5	3,4	3,1	2,9	2,7
10	5,0	4,1	3,7	3,5	3,3	3,2	2,9	2,7	2,5
11	4,8	4,0	3,6	3,4	3,2	3,1	2,8	2,6	2,4
12	4,8	3,9	3,5	3,3	3,1	3,0	2,7	2,5	2,3
13	4,7	3,8	3,4	3,2	3,0	2,9	2,6	2,4	2,2
14	4,6	3,7	3,3	3,1	3,0	2,9	2,5	2,3	2,1
15	4,5	3,7	3,3	3,1	2,9	2,8	2,5	2,3	2,1
16	4,5	3,6	3,2	3,0	2,9	2,7	2,4	2,2	2,0
17	4,5	3,6	3,2	3,0	2,8	2,7	2,4	2,2	2,0
18	4,4	3,6	3,2	2,9	2,8	2,7	2,3	2,1	1,9
19	4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,9
20	4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,8
22	4,3	3,4	3,1	2,8	2,7	2,6	2,2	2,0	1,8
24	4,3	3,4	3,0	2,8	2,6	2,5	2,2	2,0	1,7
26	4,2	3,4	3,0	2,7	2,6	2,5	2,2	2,0	1,7
28	4,2	3,3	3,0	2,7	2,6	2,4	2,1	1,9	1,7
30	4,2	3,3	2,9	2,7	2,5	2,4	2,1	1,9	1,6
40	4,1	3,2	2,9	2,6	2,5	2,3	2,0	1,8	1,5
60	4,0	3,2	2,8	2,5	2,4	2,3	1,9	1,7	1,4
120	3,9	3,1	2,7	2,5	2,3	2,2	1,8	1,6	1,3
∞	3,8	3,0	2,6	2,4	2,2	2,1	1,8	1,5	1,0

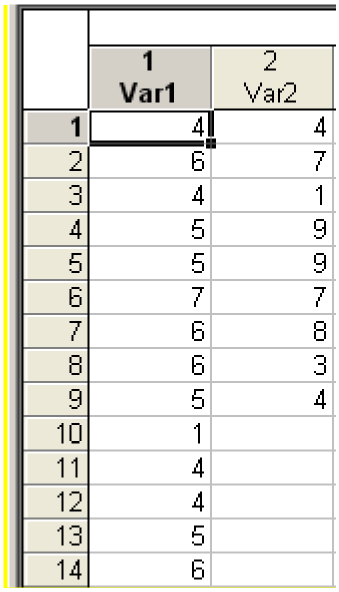
У системі STATISTICA реалізовано односторонній критерій Фішера (в якості σ_1^2 завжди беруть максимальну дисперсію). У такому випадку нульова гіпотеза H_0 відхиляється на користь альтернативної H_1. Розглянемо приклад.

Приклад. Нехай необхідно порівняти ефективність навчання двох груп студентів. Рівень успішності – характеризує рівень управління процесом навчання, а дисперсія якість управління та ступінь організованості процесу навчання. Обидва показники є незалежними і в загальному випадку мають розглядатися одночасно. Рівень успішності (математичне сподівання) кожної групи студентів характеризується середніми арифметичними $\overline{X}_1$ і $\overline{X}_2$, а якість характеризуєтся відповідними дисперсіями оценок (вибірок): S_1^2 і S_2^2. У ході оцінки рівня поточної успішності виявилося, що він однаковий в обох групах: $\overline{x}_1 = \overline{x}_2$ =4,0. Дисперсії вибірок: $S_1^2 = 0.50$ і $S_2^2 = 0.08$. Числа степенів вільності, які відповідають цим оцінкам: $f_1 = 24$ і $f_2 = 30$. Звідси, для встановлення відмінностей в ефективності навчання ми можемо скористатися стабільністю успішності, тобто перевірити гіпотезу $H_0 : S_1^2 = S_2^2$.

Вирахуємо $F = S_1^2 / S_2^2$ (в чисельнику має бути більша дисперсія), $F = 0.5/0.08 = 6.25$. По таблицям (**STATISTICA – Probability Distribution Calculator)** знаходимо $F(0.975, 24, 30) = 2.135879$, яке меньше вирахуваного, тому нульова гіпотеза має бути відхилена на користь альтернативної $H_1 : S_1^2 \neq S_2^2$. Даний висновок не може задовольнити дослідника, томущо його цікавить точна величина відношення S_1^2 / S_2^2 (у нас в чисельнику завжди більша дисперсія). При перевірці одностороннього критерія отримаємо $F(0.95, 24, 30) = 1.8874$, що менше вирахуваного вище значення. Отже, нульова гіпотеза має бути відхилена на користь альтернативної $H_1 : S_1^2 > S_2^2$.

Критерій Фішера в програмі STATISTICA в середовищі Windows

Для прикладу перевірки гіпотези (критерій Фішера) використаємо (створюємо) файл з двома змінними (fisher.sta) (рис. 5.1):



	1 Var1	2 Var2
1	4	4
2	6	7
3	4	1
4	5	9
5	5	9
6	7	7
7	6	8
8	6	3
9	5	4
10	1	
11	4	
12	4	
13	5	
14	6	

Рис. 5.1. Таблиця з двома незалежними змінними.

Щоб перевірити гіпотезу необхідно в базовій статистиці базовой статистике (**Basic Statistics and Tables**) (рис. 5.2) вибрати перевірку за Стьюдентом для незалежних змінних. (**t-test, independent, by variables**).

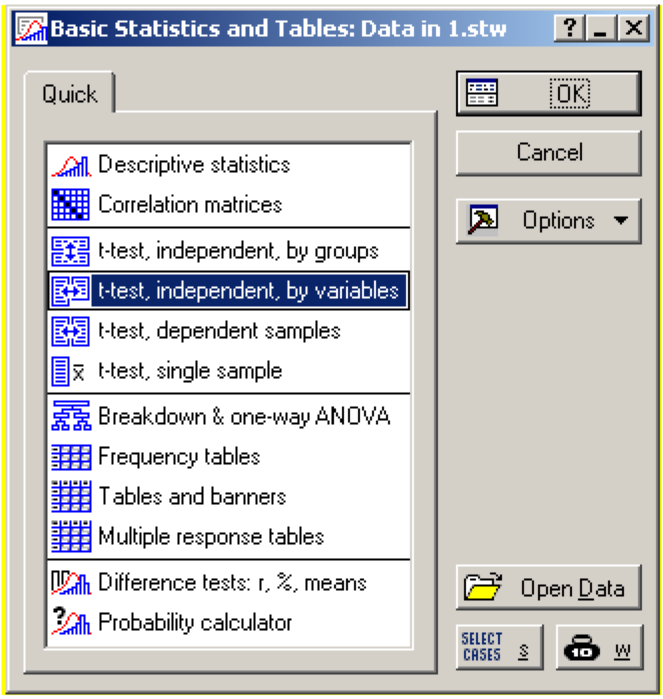


Рис. 5.2. Перевірка параметричних гіпотез.

Після вибору змінних та натискування на клавішу **Summary** проводиться підрахунок значень середньоквадратичних відхилень та критерія Фішера. Крім цього визначається рівень достовірності ***p***, при якому відмінності не є суттєвими (рис. 5.3).

Group 1 vs. Group 2	T-test for Independent Samples (Data in 1.stw) Note: Variables were treated as independent samples Valid N Group 1	Valid N Group 2	Std.Dev. Group 1	Std.Dev. Group 2	F-ratio Variances	p Variances
Var1 vs. Var2	14	9	1.460092	2.862594	3.843786	0.031209

Рис. 5.3. Результати перевірки гіпотези (F- критерій).

Використовуючи **Probability Calculator** та задавши значення параметрів можна побудувати графік розподілу Фішера з позначкою вирахуваного значення (рис. 5.4).

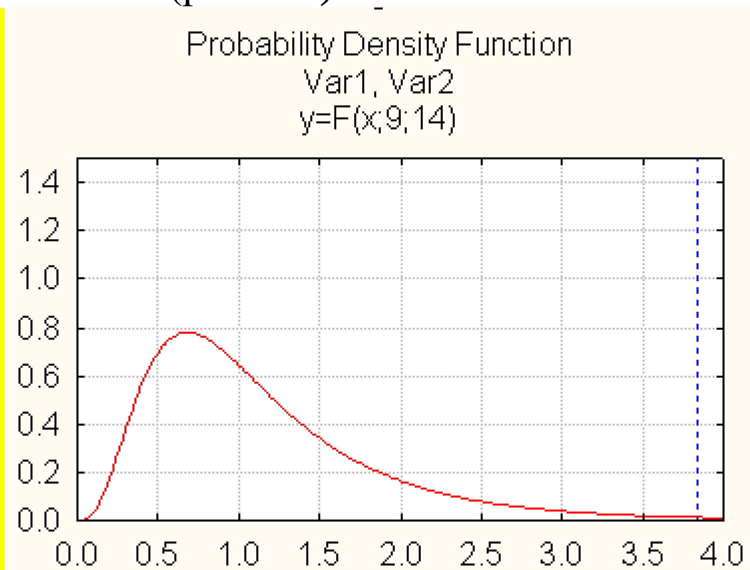


Рис. 5.4. Область прийняття (відхилення) гіпотези (F- критерій).

Стандартна помилка різниці арифметичних середніх. Після того, як з допомогою F-тесту перевірили, чи зв'язані дисперсії множин вибірки між собою гомогенно або гетерогенно, визначають і стандартну помилку різниць арифметичних середніх. Це потрібно буде пізніше при оцінці значущості відмінності середніх.

Для того, щоб визначити, яка варіативна різниця всіх середніх сукупностей (попарно) навколо різниці середніх

$$\overline{x}_1 - \overline{x}_2 = D$$

на практиці застосовують формулу:

$$m_D = \sqrt{m_{\overline{x}_1}^2 + m_{\overline{x}_2}^2} \qquad (5.2)$$

Оскільки стандартну помилку арифметичних середніх загальної сукупності $m_{\overline{x}_1}$ і $m_{\overline{x}_2}$ можна виразити формулами:

$$m_{\overline{x}_1} = \frac{\sigma N_1}{\sqrt{N_1}} \quad i \quad m_{\overline{x}_2} = \frac{\sigma N_2}{\sqrt{N_2}}, \tag{5.3}$$

то отримуємо формулу

$$m_D = \sqrt{\frac{\sigma^2 N_1}{\sqrt{N_1}} + \frac{\sigma^2 N_2}{\sqrt{N_2}}} \tag{5.4}$$

де m_D - стандартна помилка різниць середніх арифметичних сукупностей; σ^2_{N1} і σ^2_{N2} - дисперсія першої і другої спільної сукупностей; N_1 *і* N_2 - кількість елементів (спостережень) першою і другою спільних сукупностей.

Якщо відома стандартна помилка арифметичних середніх спільних сукупностей або дисперсії сукупностей і кількість членів сукупності, то при ***однорідних дисперсіях*** можна легко знайти стандартну помилку різниць середніх арифметичних за формулою 5.4.

На практиці зазвичай не відомі ні стандартна помилка арифметичних середніх спільної сукупності $m_{\overline{x}}$, ні її дисперсія σ_x^2, то часто виходять з дисперсій окремих розподілів (сукупностей вибірки) σ^2 і обчислюють $m^2_{\overline{x}}$ за формулою:

$$m^2_{\overline{x}} = \frac{\sigma^2}{n-1} = \frac{\frac{\sum\left(x_1 - \overline{x}_2\right)^2}{n}}{n-1} \tag{5.5}$$

Підставляючи вищенаведені значення в 5.1 отримуємо:

$$m_D = \sqrt{\frac{\sum\left(x_{1i} - \overline{x}_1\right)^2 + \sum\left(x_{2i} - \overline{x}_2\right)^2}{n_1 + n_2 - 2} \cdot \frac{n_1 + n_2}{n_1 \cdot n_2}}, \tag{5.6}$$

де m_D - помилка стандартного відхилення різниць всіх арифметичних середніх, якщо дисперсії вибірок однорідні; $(x_{1i} - \overline{x_1})$ - різниця кожного члена (першої сукупності і середньої сукупності; $(x_{2i} - \overline{x_2})$ - різниця кожного члена другої сукупності і середньої сукупностей; n_1 і n_2 - об'єми вибірки першою і другий сукупностей; $n_1 + n_2 - 2$ - ступені вільності сукупностей.

Якщо значення $\sum(x_{1i} - \overline{x_1})$ *і* $\sum(x_{2i} - \overline{x_2})$ окремо не обчислюють, виходять з квадрата формули помилки арифметичного середнього:

$$\sigma^2 = \frac{\sum\left(x_1 - \bar{x}\right)^2}{n}$$

і звідти $\sum(x_i - \bar{x})^2 = n \cdot \sigma^2$.
Тим самим формула може отримати вигляд:

$$m_D = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2 - 2} \cdot \frac{n_1 + n_2}{n_1 \cdot n_2}}, \tag{5.7}$$

де σ_1^2 і σ_2^2 — дисперсії першої і другої вибірок.

При неоднорідних (гетерогенних) дисперсіях ($p<5\%$) для обчислення стандартної помилки різниці арифметичних середніх використовують формулу:

$$m_D = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}, \tag{5.8}$$

при невеликих вибірках:

$$m_D = \sqrt{\frac{\sigma_1^2}{n_1 - 1} + \frac{\sigma_2^2}{n_2 - 1}}, \tag{5.9}$$

де σ_1^2 і σ_2^2 – дисперсія вибірок, n_1 і n_2– об'єми вибірок.

Достовірність різниці арифметичних середніх. Для того, щоб визначити, чи відрізняються вибірки одна від одної випадково, так, що можна припустити, що обидві є вибірками з однієї і тієї ж загальної сукупності (дисперсії однорідні і ймовірність нульової гіпотези *$p>5\%$),* або ж відмінність є достовірною (дисперсії гетерогенні і ймовірність нульової гіпотези *$p<5\%$)*), використовують t - і *F*- критерії.

Спочатку визначають $F_{emp} = \frac{\sigma_1^2}{\sigma_2^2}$. Якщо дисперсії однорідні, m_D обчислюють відповідно до формул для однорідних дисперсій. Потім знаходять різницю арифметичних, середніх вибірок $(D = \left|\bar{x}_1 - \bar{x}_2\right|)$.

Використовуючи формулу: $t = \frac{\bar{x} - x_1}{\sigma}$ знаходять t_{emp}

$$t_{emp} = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{m_D}, \tag{5.10}$$

або

$$t_{emp} = \frac{D}{m_D}, \tag{5.11}$$

Підставляючи 5.9 у 5.11 формулу:

$$t_{emp} = \frac{\left|\overline{x}_1 - \overline{x}_2\right|}{\sqrt{\frac{\sum\left(x_{11} - \overline{x}_2\right)^2 + \sum\left(x_{21} - \overline{x}_2\right)^2}{n_1 + n_2 - 2}}} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}, \tag{5.12}$$

Після знаходження t_{emp} порівнюють його з t_{kpm}, узятим з таблиці.

При отриманні t_{kpm} з таблиці враховують, що число ступенів вільності n_1+n_2-2.

Якщо $t_{emp} \geq t_{kpm}$, то відмінності (різниці) середніх достовірні, якщо ж $t_{emp} < t_{kpm}$, то відмінності середніх недостовірні.

Якщо відмінність дисперсій значуща (ймовірність нульової гіпотези нижче 5%), але все одно хочуть порівняти відмінності середніх, то роблять це аналогічно як і в попередньому випадку, тільки у формулі 5.10 m_D обчислюють за формулою:

$$m_D = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \tag{5.13}$$

тоді

$$t_{emp} = \frac{|\overline{x_1} x_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \tag{5.14}$$

і при знаходженні t_{krt} з таблиці число ступенів вільності обчислюють наближеним середнім.

Кількість ступенів вільності обчислюється за формулою $\frac{n_1+n_2-2}{2}$.

Непараметричні методи порівняння результатів.

Непараметричними *називаються такі статистичні методи, при яких на параметри увага не звертається і припущень про закон розподілу не робиться* (нормальний розподіл не обов'язковий). Більшість непараметрчених методів можна застосовувати і тоді, коли результати дослідження виміряні за допомогою шкал найменування і порядку.

Варто звернути увагу на той факт, що застосовувати ***F-II*** критерії необхідно обережно, оскільки їх можна використовувати тільки при нормальному розподілі, тобто коли ознаки аналізованих і порівнюваних рядків нормально розподіляються навколо арифметичного середнього. У такому разі диференціальна крива розподілу ряду повинна сфоєю формою наближатися до кривої Гауса. Особливо важлива при цьому симетричність.

Необхідно враховувати, що нормальний розподіл найбільш поширений у біологічних та соціальних явищах і тому при великих вибірках помилки дуже малі.

У психологічних дослідженнях часто буває важливо довести, що в результаті дії довільних факторів відбулися достовірні зміни («зсуви») у вимірюваних показниках. До числа таких факторів повинен бути віднесений насамперед фактор часу. Зіставлення показників, отриманих у тих самих випробуваних по тим самим методиках, але в різний час, дає нам часовий зсув.

Зіставлення показників, отриманих по тим самим методиках, але в різних умовах виміру (наприклад, «спокою» і «стресу»), дає нам ситуаційний зсув. Умови виміру можуть змінюватися як реально та і умовно. *Наприклад*, ми можемо попросити піддослідного *«уявити»,* що він виявився в інших умовах: у майбутньому, у позиції інших людей, які оцінюють його як би з боку, у стані розгніваного батька й т.п. Зіставляючи показники, вимірювані у звичайних і уявлюваних умовах, ми одержуємо умовний зсув.

Ми можемо створити спеціальні експериментальні умови, що приблизно впливають на ті або інші показники, і зіставити виміри, зроблені до й після експериментального впливу. Якщо зсув виявиться статистично достовірними, це дозволить нам стверджувати, що експериментальні впливи були істотними (ефективними).

Наприклад, ми можемо зробити висновок про те, що дана програма тренінгу дійсно сприяє розвитку цифрової комптентності, або що даний спосіб навіюваного впливу впливає на зміну відносин піддослідних до тієї або іншої проблеми, або що психодраматична заміна ролей підтверджує постулат Дж. Л. Морено про зближення позиції учасників суперечки після того, як їм довелося відігравати роль свого опонента й т.д.

У всіх цих випадках ми говоримо про зсув під впливом контрольованих або не контрольованих впливів. І тут ми зустрічаємося з методичними труднощами, які можна перебороти тільки шляхом введення контрольної групи, яка б не відчувала на собі впливу даного експериментального фактору. У разі відсутності контрольної групи, зсув в експериментальній групі може бути пояснений дією різних причин: часом доби, у який робилися заміри, важливою для випробовуваних подією, що відбулася між 1-м і 2-м вимірами і по потужності впливу значно перевищує експериментальний фактор і т.д. Ми ніколи не зможемо виключити тієї можливості, що зміни, досягнуті, як нам здається, в результаті наших впливів, насправді виявляються неврахованими причинами. От якщо в експериментальній групі зсуви виявляться достовірними, а в контрольній групі – недостовірними, те це, дійсно, може свідчити про ефективність впливів. При відсутності контрольної групи ми констатуємо, що зсув відбувся, але не маємо права вважати причиною зсуву наші фактори впливу.

Нехай, ми встановили, що після того, як двом конфліктуючим підгрупам довелося грати роль своїх опонентів у суперечці, підсилилося відчуття розуміння цих опонентів «з середини». Але ми не можемо виключити можливості, що якщо б ми не проводили психодраматичної заміни ролей, взаєморозуміння все-таки б покращилося просто в силу того, що обидві підгрупи якийсь час вчилися й працювали разом.

Трапляються випадки, коли ми не маємо контрольної групи, проте ми маємо 2 або більше експериментальних групи, які відрізняються умовами і способами впливу на них. Крім експериментальних умов, можуть бути, і різні природні умови життя, навчання, робота, спілкування й навіть харчування, водопостачання, географічне розташування й т.д. Порівняння груп, які відрізняються за цими ознаками, дозволить нам уточнити специфічну дію експериментальних або природно діючих факторів, хоча при цьому нам варто пам'ятати, що вплив неврахованих факторів може виявитися ще більше потужнішим.

У висновках ми все-таки будемо обмежені, якщо не перевірили свої результати на контрольній групі, у якій заміри проводилися паралельно.

Крім тимчасових, ситуаційних, умовних зсувів і зсувів під впливом, можна розглянути ще особливу категорію структурних зсувів.

Ми можемо зіставляти між собою різні показники тих самих піддослідних, якщо вони вимірювані в одних і тих самих одиницях, по одній і тій самій шкалі.

Наприклад, ми можемо досліджувати час розв'язування двох задач, що вимірюється в хвилинах, або успішність осіб, що навчаються по різних дисциплінах і т.д.

У принципі, ми могли б для такого роду «перепадів» використати критерії оцінки ймовірності середніх тенденцій для незалежних вибірок: *U*- критерій, *Q*- критерій і кутове перетворення Фішера. Однак, перед нами - залежні ряди значень, оскільки вони вимірювані на тих самих піддослідних, тому більш обґрунтованим буде використання критеріїв оцінки ймовірності зсувів для зв'язаних вибірок. Виключення можуть становити випадки, коли ми зіставляємо величини зсувів у двох незалежних групах піддослідних, наприклад експериментальної і контрольної. Допустимо, якщо ми встановили, що позитивний зсув убік поліпшення взаєморозуміння спостерігається як в експериментальній так і в контрольній групах. Ми можемо спробувати довести, що в експериментальній групі цей зсув ймовірно більший, ніж у контрольної. Як висновок, експериментальний вплив все-таки істотний.

Останнє важливе запитання стосується того, чи повинні ми завжди робити обидва заміри в одній і тій самій вибірці, або «зсув» можна вивчати на подібних, так званих «урівноважених» вибірках, що збігаються одна з одною по статі, віку, професії й іншими значимими для дослідника характеристикам.

По суті, допускається зіставлення показників різних вибірок, урівноважених по всім значимим для дослідження ознакам. Іншими словами, можна рівень тривоги або об'єм уваги до іспиту вимірювати в одній підгрупі, а після іспиту – в іншій підгрупі, за умови, що вони *«урівноважені»*. Однак, як показує досвід, створити *«урівноважені»* підгрупи практично неможливо. Ми завжди впираємося у факт існування відмінностей між виділеними підгрупами, які можуть у значній мірі вплинути на результат. У підсумку виявиться, що ми досліджували не вплив екзаменаційного стресу на рівень тривоги або об'єм уваги, а відмінності за цим показником між двома виділеними підгрупами. На жаль, у значній мірі це відноситься й до проблеми порівняння експериментальної й контрольної груп: ми майже ніколи не можемо бути впевнені, що виявлені відмінності пояснюються дією досліджуваних факторів, а не відмінностями між двома вибірками.

Багато дослідників обходять цю проблему найпростішим способом: зсув є – значить, вплив ефективний! І дійсно, при відсутності контрольної вибірки можна поміркувати про те, які ж причини, крім передбачуваних, можуть пояснити отримані зсуви.

Інший варіант «зрівноважування» ─ ведення паралельних форм тесту. У тих випадках, коли на результатах повторних вимірів можуть позначитися ефекти навченості, доводиться вимірювати *«до»* реакції піддослідного за допомогою одного інструмента, а *«після»* ─ за допомогою іншого. У результаті на замірах може відбитися й дія фактору часу, і розходження в паралельних формах тесту, і незрозуміло що ще. Створити паралельну форму методики не менш важко, ніж підібрати *«урівноважену»* групу піддослідних. І все-таки, у тих випадках, коли в нас немає іншого виходу, доводиться звертатися до цього способу.

У тих випадках, коли ми хочемо оцінити розходження в інтенсивності зсувів у двох групах піддослідних (контрольної й експериментальної або двох експериментальних), ми можемо використати різні варіанти зіставлень:

1) робити зіставлення окремо у двох групах, використовуючи критерії L і $\chi^2_г$;

2) зіставляти показники зрушення у двох групах. Оскільки групи незалежні, значення зсувів також незалежні, і ми можемо застосовувати стосовно них критерії Q-Розенбаума, U-Манна-Уітні й φ* - кутове перетворення Фішера.

У педагогічних явищах, коли працюють з малими вибірками і мають в своєму розпорядженні обмежений набір кількісних показників, доцільно застосовувати непараметричні методи при порівнянні результатів дослідження.

Знаковий, або сигнатурний тест (*G*- критерій знаків) – є **о**дним з найбільш простих непараметричних методів, які можна використовувати для порівняння двох випадкових рядів з різною кількісною ознакою. *G*-критерій призначений для встановлення загального напрямку зсуву досліджуваної ознаки. Він дозволяє встановити, у яку сторону у вибірці вцілому змінюються значення ознаки при переході від першого виміру до другого: чи змінюються показники у бік покращення, підвищення або посилення або, навпаки, убік погіршення, зниження або послаблення. Критерій знаків застосуємо й до тих зсувів, які можна визначити лише *якісно* (*наприклад*, зміна негативного відношення до чого-небудь на позитивне) та до тих зсувів, які можуть бути вимірювані кількісно (наприклад, скорочення часу роботи над завданням після експериментального впливу).

У другому випадку, якщо зсув варіюють у досить широкому діапазоні, краще застосовувати критерій *T*-Вілкоксона. Він ураховує не тільки напрямок, але й інтенсивність зсуву і може виявитися більш «сильнішим» у визначенні ймовірності зсувів, ніж критерій знаків.

Як правило, дослідник уже в процесі експерименту може помітити, що в більшості піддослідних показники у другому вимірі мають тенденцію, скажемо, підвищуватися. Однак досліднику потрібно довести, що переважним є позитивний зсув.

Для початку ми назвемо зсуви, які на нашу думку є переважними (типовими зсувами), а зсуви більше рідшого, протилежного напрямку, нетиповими. Якщо значення показника підвищують в більшої кількості піддослідних, то цей зсув ми будемо вважати типовим. Якщо ми досліджуємо відношення випробовуваних до якої-небудь події або пропозиції, і після експериментальних впливів у більшості піддослідних негативне відношень змінилося на позитивне, то цей зсуву ми назвемо типовим.

Можуть бути «нульові» зсуви, коли реакція не змінюється або показники не підвищуються й не знижуються, а залишаються на колишньому рівні. Однак такі «нульові» зсуви в критерії знаків виключаються за бажанням. При цьому кількість пар зменшується на число таких «нульових» зсувів.

Суть критерію знаків полягає в тому, що він визначає, чи не занадто багато спостерігається «нетипових зсувів», щоб зсув в «типовому» напрямку вважався переважним? Зрозуміло, що чим менше «нетипових зсувів», тим більша ймовірність, що перевага «типового» зсуву є переважною. $G_{емп}$ – кількість нетипових зсувів. Чи менше $G_{емп}$, тим більша ймовірність, що зсув у «типовому» напрямку статистично достовірний.

Гіпотези:

H_0: Перевага типового напрямку зсуву є випадковою.

H_1: Перевага типового напрямку зсуву є невипадковою.

На рис. 5.5 «типові» зсуви зображені у вигляді світлої хмари, а нетипові зсуви - темної хмари. Ми бачимо, що темна хмара значно менше. Допустимо, після виступу оратора більшість слухачів змінили своє негативне відношення до якоїсь пропозиції на позитивне. Разом з тим, частина слухачів змінила своє позитивне відношення на негативне, проявивши «нетипову» реакцію. Критерій знаків, який дозволяє визначити, чи не занадто значна частина слухачів «нетипово» прореагувала на виступ оратора? Чи поглинає світла хмара невелику темну хмару?

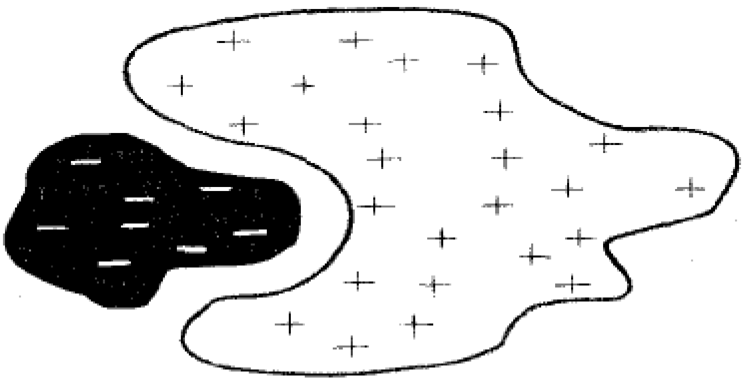


Рис. 5.5. Графічне подання позитивних і негативних зсувів у формі хмари: світла хмара – позитивний зсув, темна хмара – негативний зсув.

У табл. 5.3 дані критичні значення критерію знаків. Оскільки критерій знаків являє собою одне із трьох виключень із загального правила, представимо узагальнену *«вісь достовірності»* для цього критерію графічно (рис. 5.5).

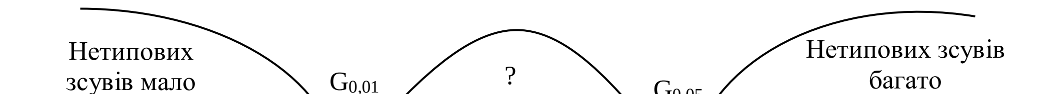


Рис. 5.5. Узагальнена *«вісь значимості»* для критерію знаків.

Зона достовірності простягається вліво, убік більш менших значень, оскільки чим менше *«нетипових»* знаків, тим достовірніший «типовий» зсув. Зона недостовірності простягається вправо, убік більш більших значень *G*. Поступово «нетипових» зрушень стає так багато, що губиться саме відчуття якоїсь переваги у напрямках зсувів. Зона недостовірності характеризує ситуацію, коли зсуви обох напрямків змішуються.

Обмеження критерію знаків. Кількість спостережень в обох замірах – не менш 5 і не більше 300.

Алгоритми розрахунку критерію знаків.

Перший.

1. Підрахувати кількість нульових реакцій і виключити їх з розгляду. У результаті *n* зменшиться на кількість нульових реакцій.
2. Визначити переважний напрям змін. Вважати зсуви в переважних напрямах *«типовими»*.
3. Визначити кількість *«нетипових»* зсувів. Вважати це число емпіричним значенням *G*.
4. За табл. 5.3 визначити критичні значення G для даного n.
5. Зіставити $G_{емп}$ *з* $G_{кр}$. Якщо $G_{емп}$ менше $G_{кр}$ або принаймні дорівнює йому, зсув у типову сторону може вважатися достовірним.

Приклад. Нехай, дослідника цікавить питання, чи є результати, проведеної в експериментальному класі контрольної роботи достовірно кращими за результати контрольної роботи в контрольному класі (при оцінці застосовувалися однакові об'єктивні критерії, наприклад, кількість правильних відповідей на 10 питань). Результати заносяться попарно до таблиці у вигляді випадкового ряду. Основою його може бути, наприклад, порядковий номер учнів в класі (табл. 5.2).

Як видно з таблиці, між кожною парою чисел експериментального і контрольного класу x1 і y1 можуть бути відношення

$$(x_1 > y_1(+);\ x_1 < y_1(-) \text{ і } x_1 = y_1(0).$$

У четвертому рядку таблиці відзначається не різниця числових значень результатів контрольних робіт, а результат порівняння величин результатів у вигляді знаків «+», «-», «0». Це значно спрощує обчислення.

При підведенні підсумків знак «0» не розглядається і при цьому також відповідно зменшують кількість порівнюваних пара чисел. Далі виписують кількість частот із знаками «+» і «-» і перевіряють за допомогою нульової гіпотези, є відмінність знаків «+» і «-» випадковим чи ні.

Таблиця 5.2

Кількість правильних відповідей

Порядковий номер учня у списку класу	1	2	3	4	5	6	…	*n*
Правильні відповіді в експериментальному класі x_1	6	7	5	8	5	6	…	6
Правильні відповіді у контрольному класі y_1	4	8	3	7	5	7	…	6
Знак (x_1 - y_1)	+	-	+	+	0	-	…	0

Краще знайти відповідь на нульову гіпотезу за допомогою z-критерію. При цьому справедлива формула:

$$z = \frac{|f \ll +\gg \ -n \cdot P\{+\}| - 0{,}5}{\sqrt{n(P\{+\} \cdot (P\{-\})}} \tag{5.15}$$

де f «+» – кількість найбільш часто (чи f «-» – найбільш рідко) зустрічається; n – кількість пар чисел, які порівнюємо; $P\{+\}$ – передбачувана ймовірність частоти знаку «+»; $P\{-\}$ – передбачувана ймовірність частоти знаку «-»; 0,5 – коефіцієнт.

Примітка. Частота зустрічі знаків «+» і «-» рівноймовірна і рівна половині, тобто P=0,5.

Припустимо, що розглядалися 40 пар чисел, причому знак «+» зустрічався 25 разів, знак «-» 10 разів, знак «0» – 5 раз.

Отже, $z = \frac{|25-35\cdot 0{,}5|}{\sqrt{35\cdot 0{,}5\cdot 0{,}5}} = \frac{7}{\sqrt{8{,}75}} \approx 2{,}4.$

Значення $z=2{,}4$ показує, що з одної сторони арифметичного середнього між значеннями і арифметичним середнім знаходиться 48,9% випадків, з іншого боку також 48,9% випадків, всього 97,8%. Тим самим *1%<p<5%,* на підставі чого можна стверджувати, що відмінність між випадковими рядами, що знаходиться на рівні значущості 5%, – достовірно, а на рівні *1%* – недостовірно.

Таблиця 5.3

Достовірність критерію знаків

n	*p=1%*	*p=5%*	*n*	*p=1%*	*p=5%*
10	0	1	55	17	19
15	2	3	60	19	21
20	3	5	65	21	24
25	5	7	70	23	26
30	7	9	75	25	28
35	9	11	80	28	30
40	11	13	85	30	32
45	13	15	90	32	35
50	15	17			

Достовірність результатів знакового тесту легко перевірити і за допомогою відповідної таблиці (табл. 5.3).

За допомогою табл. 5.3 можна знайти критичне значення *mn,* яке показує, яким може бути найбільша кількість мінусів (або плюсів), що забезпечує задану достовірність.

У нашому прикладі коли *n*=35, при 1% рівні значущості мінусів може бути 9, при 5% рівні значущості 11. Оскільки у нас було 10 мінусів, то результат достовірний при 5% рівні значущості і недостовірний при 1% рівні значущості.

У табл. 5.3 наведено критичні значення для 90 пар спостережень. Якщо пар спостережень більше 90, то можна значення $\overline{m_n}$ на 5% рівні значущості досить точно обчислити за формулою:

$$\overline{m_n} = \frac{n-1}{2} - 0{,}98\sqrt{n+1}, \qquad (5.16)$$

де *n* – кількість пар спостереження.

Знаковий тест можна використати і тоді, коли $x_1 - y_1$ представляють собою не різницею чисел, а правильні «+» або неправильні «-» відповіді на будь яке питання контрольної роботи.

Наприклад, провели контрольну роботу в експериментальному і контрольному класах (групах), причому відповідь кожного питання оцінювали як правильну *(о)* і неправильну *(v)*. Неправильними вважали і питання, що залишилися без відповіді. Результати занесли до таблиці.

Якщо в експериментальному класі (групі) учень (студент) відповів правильно, а в контрольному класі неправильно, то в графу (рядок «знак $(x_1 - y_1)$» заноситься знак «+», якщо навпаки, то знак

«-». Якщо відповіді в обох класах однакові, правильні або неправильні в графу заноситься знак «0» і цей результат надалі не розглядається.

Таблиця 5.4

Відповіді на питання

Порядковий номер учня у списку класу	1	2	3	4	5	…	*n*
Правильні відповіді в експериментальному класі x_1	$\bar{0}$	$\bar{0}$	υ	υ	$\bar{0}$	…	$\bar{0}$
Правильні відповіді у контрольному класі y_1	υ	$\bar{0}$	υ	$\bar{0}$	υ	…	υ
Знак (x_1 - y_1)	+	0	0	-	+	…	+

Припустимо, що при 30 парах учнів отримали 20 знаків «+», 5 знаків «-» і 5 знаків «0». Використовуючи приведену вище формулу, отримаємо:

$$\frac{|20-25\cdot0{,}5|-0{,}5}{\sqrt{25\cdot0{,}5\cdot0{,}5}} = \frac{7}{\sqrt{6{,}25}} = \frac{7}{2{,}5} = 2{,}8 \approx 3.$$

З z-таблиці видно, що в даному випадку $(z=3)$ 49,865% значень знаходиться по одну сторону від арифметичного середнього, а по іншу сторону знаходиться стільки ж частот. Отже, ступінь ймовірності 99,73%. З таблиці з'ясовується, що критичне значення $\overline{m_n}$ при 1 % рівні значущості 5 знаків «-» і при 5% рівні значущості 7 знаків «-». Отже, наш результат достовірний як при 1 %, так і при 5% рівні значущості.

Визначення достовірності відмінностей сукупностей методом χ^2 (*hii* квадрата). Якщо дослідника цікавить не достовірність відмінності двох сукупностей, а питання, чи існує відмінність між рядами яких-небудь показників двох сукупностей, то зручно користуватися методом χ^2 (критерій Пірсона). χ^2 ще називають критерієм згоди. Це непараметричний критерій оцінки, оскільки при його обчисленні користуються замість характеристик сукупності (середні дисперсії і ін.) частотами, що характеризують розподіл варіантів.

Величину χ^2 знаходимо за формулою:

$$\chi^2 = \sum\left[\frac{(\acute{f}_E-\acute{f}_k)^2}{\acute{f}_k}\right], \tag{5.17}$$

де $\acute{f}_E$ – відносна частота інтервалу одного ряду, $\acute{f}_k$– іншого ряду.

Якщо два ряди явищ схожі, то частоти одних і тих же інтервалів в обох рядах не повинні значно відрізнятися. Кількість спостережень в обох рядах повинна бути рівною.

Якщо кількість спостережень у рядах відрізняється, необхідно використовувати відносний розподіл частоти, сума якого завжди 100%. Потрібно також, щоб частота жодного інтервалу не була дуже маленькою (частота повинна бути не менше 4 - 5). Оскільки інтервали не повинні бути рівними, можна інтервали з малою частотою підсумовувати. Необхідно підкреслювати, що χ^2 критерій можна використовувати тільки при згрупованих даних.

Кількість ступенів вільності при χ^2 критерій *n-1*, де n – кількість інтервалів.

Приклад. Припустимо, що провели контрольну роботу в експериментальних і контрольних класах, причому бали оцінок розподілилися так, як показано в таблиці 5.6.

Таблиця 5.6

Розподіл оцінок контрольних робіт

Бали (оцінки)	Частота в експериментальних класах (f'_E)	Частота в контрольних класах (f'_k)
0 – 10	2	3
11 – 20	5	3
21 – 30	4	4
31 – 40	11	13
41 – 50	12	14
51 – 60	13	16
61 – 70	10	14
71 – 80	8	6
81 – 90	3	5
92 - 100	2	2
	$\sum f'_E = 70$	$\sum f'_k = 80$

Нас цікавить питання, чи є суттєвими відмінності у згаданих рядах.

З аналізу таблиці з'ясовується, що хоча дані і згруповані, але частоти в деяких інтервалах невеликі і суми частот в експериментальних і контрольних класах різні. Для того, щоб достовірно з'ясувати відмінності рядів можна було використовувати χ^2 критерій, при цьому перш за все підсумовують інтервали з маленькою частотою, а потім знаходять відносні частоти $\acute{f}_E$ **і** $\acute{f}_k$**.** Дані знову заносяться до таблиці 5.7.

Потім знаходять: $(\acute{f}_E - \acute{f}_k)$ i $\sum\left[\frac{(\acute{f}_E - \acute{f}_k)^2}{\acute{f}_k}\right]$

Кількість ступенів вільності 5, оскільки інтервалів 6. Із таблиці 5.8 з'ясовується, що 5-му ступеню вільності та 95% рівні ймовірності відповідає значення χ^2=11,1. Оскільки знайдене нами при обчисленні χ^2 ≈1,88 (x_{emp}<x_{krt}), то в даному випадку не можна відкидати нульову гіпотезу.

Таблиця 5.7

Робоча таблиця для χ^2 критерія

Номер інтер-валу	Частота (f_E)	Частота (f_k)	Відносна частота (f'_E%)	Відносна частота (f'_k%)	$(f'_E - f'_k)$	$(f'_E - f'_k)^2$	$\frac{(f'_E - f'_k)^2}{f'_k}$
0 – 30	11	10	15,7	12,5	3,2	10,24	0,82
31 – 40	11	13	15,7	16,25	-0,55	0,30	0,01
41 – 50	12	14	17,1	17,5	-0,4	0,16	0,02
51 – 60	13	16	18,6	20,0	-1,4	1,96	0,10
61 – 70	10	14	14,3	17,5	-3,2	10,24	0,59
71 - 100	13	13	18,6	16,25	2,35	5,52	0,34
	$\sum f_E = 70$	$\sum f_k = 80$	100%	100%	0	χ^2≈1,88	

Таблиця 5.8

Таблиця χ^2 критерія

n-1	**95 %**	**99 %**	***n-1***	**95 %**	**99 %**
1	3,84	6,63	16	26,3	32,0
2	5,99	9,21	17	27,6	33,4
3	7,81	11,3	18	28,9	34,8
4	9,49	13,3	19	30,1	36,2
5	11,1	15,1	20	31,4	37,6
6	12,6	16,8	21	32,7	38,9
7	14,1	18,5	22	33,9	40,3
8	15,5	20,1	23	35,2	41,6
9	16,9	21,7	24	36,4	43,0
10	18,3	23,2	25	37,7	44,3
11	19,7	24,7	26	38,9	45,6
12	21,0	26,2	27	40,1	47,0
13	22,4	27,7	28	41,3	48,3
14	23,7	29,1	29	42,6	49,6
15	25,0	30,6	30	43,8	50,9

Отже, проведені в експериментальному і контрольному класах контрольні роботи достатньо схожі, і застосована в експериментальному класі незалежна змінна істотно не вплинула на результати експерименту у порівнянні з результатами контрольного класу.

У загальному вигляді формулу для обчислення χ^2 можна виразити таким чином:

$$\chi^2 = \sum \left[\frac{(x-y)^2}{y}\right], \tag{5.18}$$

де x – результати експерименту; *y* – раніше отримані результати або теоретично отримані результати.

***T*-критерій Вілкоксона.** Даний критерій застосовується для зіставлення показників, виміряних у двох різних умовах в одній і тій же вибірці піддослідних.

Він дозволяє встановити не тільки напрям змін, але і їх вираження. За його допомогою ми визначаємо, чи є зсув показників у якомусь одному з напрямків більше інтенсивним, ніж в іншому.

Цей критерії застосуємо в тих випадках, коли ознаки виміряні принаймні по шкалі порядку, і зсуви між другим та першим вимірами теж можуть бути упорядковані. Для цього вони повинні варіювати у досить широкому діапазоні. У принципі, можна застосовувати критерії *T* і в тих випадках, коли зсуви приймають тільки три значення: -1, 0 і +1, але тоді критерії *T* навряд чи додасть що-небудь нове до тих висновків, які можна було б одержати за допомогою критерію знаків. От якщо зсуви змінюються, скажемо, від -30 до +45, тоді має смисл їх рангувати і потім сумувати ранги.

Суть методу полягає в тому, що ми зіставляємо значення (вирази), зсувів у тому чи іншому напрямках по абсолютній величині. Для цього ми спочатку рангуємо всі абсолютні величини зсувів, а потім підсумуємо ранги. Якщо зсуви в позитивну й в негативну сторону відбуваються випадково, то сума рангів абсолютних значень їх будуть приблизно рівні. Якщо ж інтенсивність зсуву в одному з напрямків переважає, то сума рангів абсолютних значень зсувів у протилежну сторону буде значно нижче, ніж це могло бути при випадкових змінах.

Спочатку ми виходимо із припущення про те, що типовим зсувом буде зсув в напрямку, що частіше зустрічається, а нетиповим зсувом - зсув напрямку, що рідко зустрічається.

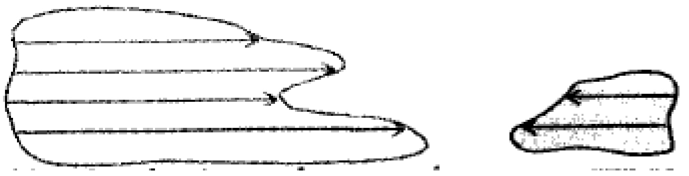
Гіпотези:

H_0: Інтенсивність зсувів у типовому напрямку не переважає інтенсивності зсувів у нетиповому напрямку.

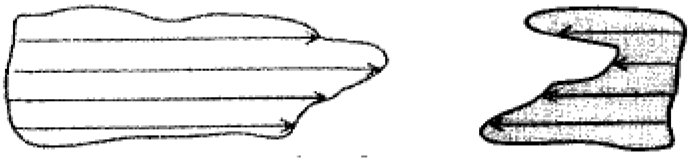
H_1: Інтенсивність зсувів у типовому напрямку переважає інтенсивності зсувів у нетиповому напрямку.

За аналогією з критерієм знаків, зсуви в протилежні сторони ми можемо уявити собі при вигляді двох хмар. Величина хмари залежить не тільки від кількості відповідних зсувів, але й від їхньої інтенсивності, вираженої у довжині стрілок (рис. 5.6). По суті, хмари протистоять одна одній, як два повітряні фронти: вони не просто змагаються по величині, вони міряються силами! При визначенні *n,* а саме при $n>18$, ми взагалі можемо відмовитися від поняття типового зсуву. Зсувів у ту й іншу сторону може виявитися порівну якщо 9 менших зсувів будуть відноситися до одного напрямку, а 9 більших зсувів - до протилежного, то ми можемо констатувати достовірну перевагу цього протилежного напрямку зсувів. Згадаємо, що критерії знаків в цьому випадку не виявив б жодних достовірних розходжень.

На рис. 5.6(а) *«світлий фронт»* переважає над *«темним фронтом»* і за кількістю зсувів, і за їх інтенсивністю. На рис. 5.6(б) *«світлий фронт»* переважає над *«темним фронтом»* тільки по інтенсивності, але по кількості зсувів вони рівні. На рис. 5.6(в) *«світлий фронт»* поступається *«темному»* за кількістю зсувів, але найінтенсивніші зсуви належать «світлому фронту». Тут критерій знаків може констатувати перевагу змін, які відповідають *«темному фронту»*. Проте, інтенсивність протилежних, хоч і рідких, зсувів, настільки велика, що робити якісь однозначні висновки було б необачно.



а) *«світлий фронт»* переважає над *«темним фронтом»* по кількості зсувів та по їх інтенсивності.



б) *«світлий фронт»* переважає над *«темним фронтом»* тільки по інтенсивності, але по кількості зсувів вони рівні.

в) *«світлий фронт»* уступає «*темному»* по кількості зсувів, але найінтенсивніші зсуви належать «світлому фронту».

Рис. 5.6. Варіанти співвідношень *«світлого» та «темного фронтів»* – зсуви двох різних напрямків.

Обмеження і застосування критерію Вілкоксона.

1. Мінімальна кількість учасників експерименту, які пройшли вимірювання в двох умовах – 5 осіб. Максимальна кількість – 50 осіб.
2. Нульові зсуви виключаються з розгляду, і кількість спостережень *n* зменшується на кількість нульових зсувів. Можна обійти це обмеження, сформулювавши гіпотези, що включають відсутність змін, наприклад: «Зсув убік збільшення значень перевищує зсув убік зменшення значень і тенденцію збереження їх на колишньому рівні».

Алгоритм підрахунок критерію *T*-Вілкоксона.

1. Скласти список учасників експерименту у довільному порядку, наприклад, алфавітному.
2. Розрахувати різницю між індивідуальними значеннями в другому

і першому замірах («після» - «до»). Визначити, що буде вважатися «типовим» зсувом і сформулювати відповідні гіпотези.

3. Перевести різницю в абсолютні величини й записати їх окремим стовпцем (інакше важко відволіктися від знака різниці).

4. Прорангувати абсолютні величини відмінностей, нараховуючи меншому значенню менший ранг. Перевірити збіг отриманої суми рангів з розрахунковою.

5. Відзначити довільними знаками ранги, що відповідають зсувам у «нетиповому» напрямку.

6. Підрахувати суму цих рангів за формулою: $T=\Sigma R_r$, де R_r рангові значення зсувів з більше рідким знаком.

7. Визначити критичні значення *T* для даного *n* за таблицею. Якщо *T* менше або дорівнює *T*, зсув в «типову» сторону по інтенсивності ймовірно переважає.

Критерій $\chi^2_г$ Фрідмана. Критерій $\chi^2_г$ застосовується для зіставлення показників, вимірюваних у трьох або більше умовах в одній і тій самій вибірці учасників експерименту. Критерій дозволяє встановити, що величини показників від умови до умови зміняються, але при цьому не вказує на напрямок змін.

Даний критерій є поширенням критерію Т-Вілкоксона на кількість умов виміру більше 2. Однак тут ми рангувати не абсолютні величини зрушень, а самі індивідуальні значення, отримані даним випробуваним в 1, 2, 3 і т.д. вимірах.

Наприклад, якщо в учасника експерименту в першому вимірі визначена швидкість проходження графічного лабіринту 54 сек, у другому вимірі - 42 сек, а в третьому вимірі - 63 сек, те ці показники одержать ранги, відповідно, 2, 1, 3, оскільки меншому значенню, отриманому в другому вимірі, ми нарахуємо ранг 1, середньому значенню, отриманому в першому вимірі - ранг 2, а найбільшому значенню, отриманому в третьому вимірі - ранг 3.

Після того, як всі значення будуть проранговані, підраховуються суми рангів по стовпцях для кожного зі зроблених замірів.

Якщо розходження між значеннями ознаки, отриманими в різних умовах, випадкові, то суми рангів по різних умовах будуть приблизно рівні. Але якщо значення ознаки змінюються в різних умовах якимось закономірним чином, то в одних умовах будуть переважати високі ранги, а в інші - низькі. Суми рангів будуть ймовірно відрізнятися між собою. Емпіричне значення критерію $\chi^2_г$ і вказує на те, наскільки розрізняються суми рангів. Чим більше емпіричне значення $\chi^2_г$ тим більш істотні розбіжності сум рангів воно відбиває.

Якщо $\chi^2_г$ дорівнює критичному значенню або перевищує його, розходження статистично достовірні.

Гіпотези:

H₀: Між показниками, отриманими (виміряними) у різних умовах, існують лише випадкові розходження.

H₁: Між показниками, отриманими в різних умовах, існують невипадкові розходження.

Графічно це буде виглядати як «пучок» ламаних ліній зі зламами в тих самих місцях. На рис. 5.7 представлено графіки зміни часу рішення анаграм у ході експерименту з дослідження інтелектуальної наполегливості. Ми бачимо, що «сирі» значення п'яти досліджуваних дають пучок, що розсипається, хоча й з помітною тенденцією до зламу в одній і тій самій точці - на анаграмі № 2. На рис. 5.8 представлено графіки, побудовані за рангованими даними того ж дослідження. Ми бачимо, що тут «пучок» зібраний практично в одну жирну лінію, з єдиної кривій, що вибивається з нього. По суті, критерій $\chi^2_г$ дозволяє нам оцінити, чи достатньо узгоджено вигинається пучок при переході від умови до умови. До тих пір більше $\chi^2_г$ поки більш вираженими є розходження.

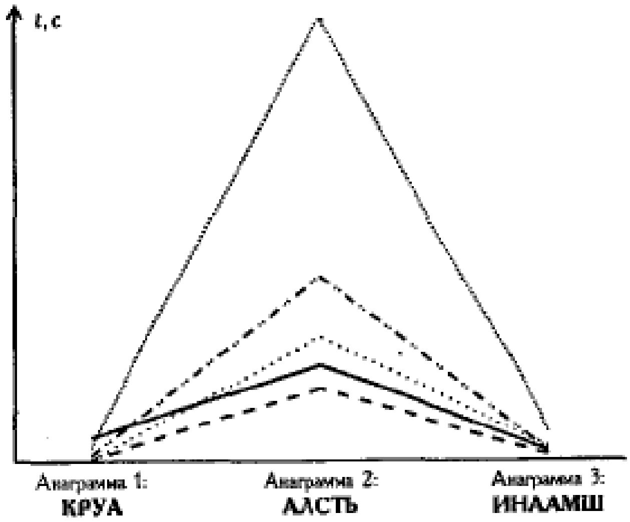


Рис. 5.7. Графіки зміни часу рішення трьох послідовно представлених анаграм (у сек) у п'яти піддослідних.

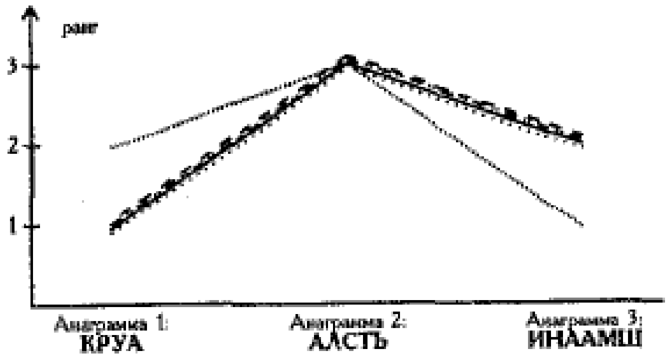


Рис. 5.8. Графіки зміни рангованих показників часу вирішення анаграм.

Обмеження критерію.

1. Нижній поріг: не менш 2-х учасників експерименту (n>2), кожний з яких пройшов не менш 3-х вимірів (n>3).
2. При c=3, n<9, при c=4, n<4, рівень достовірності отриманого емпіричного значення $\chi^2_{\text{г}}$ визначається за табл. 5.9; при більшій кількості учасників експерименту або умов, отримані емпіричні значення $\chi^2_{\text{г}}$ зіставляються із критичними значеннями $\chi^2_{\text{г}}$, характерними для критерія Пірсона. Це пояснюється тим, що χ^2 має розподіл, подібний з розподілом χ^2. Число ступенів вільності V визначається за формулою: $V=c-1$, де c - кількість умов виміру (вимірів).

Таблиця 5.9

Критичні значення критерію Фрідмана

Кількість умов (вибірок) c = 3							
n=2		**n=3**		**n=4**		**n=5**	
χ_r^2	*p*	χ_r^2	*p*	χ_r^2	*p*	χ_r^2	*p*
0	1.000 0,833	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
n=6		**n=7**		**n=8**		**n=9**	
χ_r^2	*p*	χ_r^2	*p*	χ_r^2	*p*	χ_r^2	*p*
0,00	1,00000	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
						14,222	0,0Q0097
						14,889	0,000054
						16,222	0,000011
						18,000	0,0000006

Продовження таблиці 5.9

Кількість умов (вибірок) c = 4							
n=2		n=3		n=4			
χ^2_r	p	χ^2_r	p	χ^2_r	p	χ^2_r	p
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

Алгоритм підрахунку критерію χ^2г Фрідмана.

1. Прорангувати індивідуальні значення першого учасника експерименту, отримані ним в 1-м, 2-м, 3-м і т.д. замірах.
2. Пробробити те ж саме стосовно всіх іншим учасниками експерименту.
3. Просумувати ранги за умовами, у яких здійснювалися виміри. Перевірити збіг загальної суми рангів з розрахунковою сумою.
4. Визначити емпіричне значення χ^2г, за формулою:

$$\chi_r^2 = \left[\frac{12}{n \cdot c \cdot (c+1)} \cdot \sum T_j^2\right] - 3 \cdot n \cdot (c+1)$$

де c - кількість умов; n - кількість випробуваних; T_j - суми рангів по кожному з умов.

5. Визначити рівні статистичної значимості для χ^2гемп:

а) при c=3, n<9 ;

б) при c=4, $n<4$.

6. При більшій кількості умов і/або учасників експерименту – визначити кількість ступенів вільності V за формулою: $v=c-1$, де с – кількість умов (замірів).

По таблицею визначити критичні значення критерію χ^2г, при даному числі ступенів вільності. Якщо χ^2гемп дорівнює критичному значенню або перевищує його, розходження достовірні.

***L*-критерій тенденції Пейджа.** Критерій *L*-Пейджа застосовується для зіставлення показників, вимірюваних у трьох і більше умовах в одній і ті самій вибірці піддослідних. Критерій дозволяє виявити тенденції в зміні величини ознаки при переході від умови до умови. Його можна розглядати як продовження тесту Фрідмана, оскільки він не тільки констатує розходження, але й вказує на напрямок змін.

Даний критерій дозволяє перевірити наші припущення про певну вікову або ситуативно обумовлену динаміку тих чи інших ознак. Він дозволяє об'єднати кілька зроблених замірів єдиною гіпотезою про тенденції зміни значень ознаки при переході від заміру до заміру. Якби не його обмеження, критерій був би незамінний в «поздовжніх» дослідженнях.

На жаль, наявні таблиці критичних значень розраховані тільки на невелику вибірку (n<12) обмежена кількість вимірів, що зіставляють (с<6).

У випадку, якщо ці обмеження не виконуються, потрібно використати критерій $\chi^2_г$ Фрідмана.

У критерії ***L*** застосовується таке ж рангування умов по кожному піддослідному, як і в критерії $\chi^2_г$. Якщо піддослідний у першому досдіді допустив 17 помилок, у другому - 12, а в третьому - 5, то 1-й ранг одержує третю умову, 2-й ранг - другу, а 3-й ранг - першу умова. Після того, як значення всіх випробуваних будуть проранговані, підраховуються суми рангів по кожній умові. Потім всі умови розташовуються в порядку зростання рангових сум: на першому місці ліворуч виявиться умова з меншою ранговою сумою, за нею - умова з наступної по величині ранговою сумою, і т.д., поки праворуч не виявиться умова з найбільшою ранговою сумою, Далі ми за допомогою спеціальної формули підрахунку ***L*** перевіряємо, чи дійсно значення зростають зліва на право. Емпіричне значення критерію ***L*** відбиває ступінь розходження між ранговими сумами, тому чим вище значення ***L***, тим більше істотні розходження.

Гіпотези:

H₀: Збільшення індивідуальних показників при переході від першої умови до другого, а потім до третього й далі, випадкове.

H₁: Збільшення індивідуальних показників при переході від першої умови до другого, а потім до третього й далі, невипадкове.

При формулюванні гіпотез ми маємо на увазі нову нумерацію умов, що відповідає передбачуваним тенденціям.

Графічне подання критерію. Використаємо для ілюстрації приклад із пред'явленням анаграм приблизно зростаючої складності. Задум експериментатора полягав у тому, щоб кожна наступна задача жадала від випробуваних усе більше тривалих роздумів.

Судячи із графіка на рис. 5.9, у більшості випробуваних анаграм 1 стоїть на першому ранговому місці, тобто вирішується швидше двох інших, анаграма 3 на 2-м ранговому місці, а анаграма 2 - на 3- м. Очевидно, їх варто було б пред'являти в іншій послідовності: 1, 3, 2. Графік, що виражає таку гіпотетичну послідовність задач, представлений на рис. 5.10.

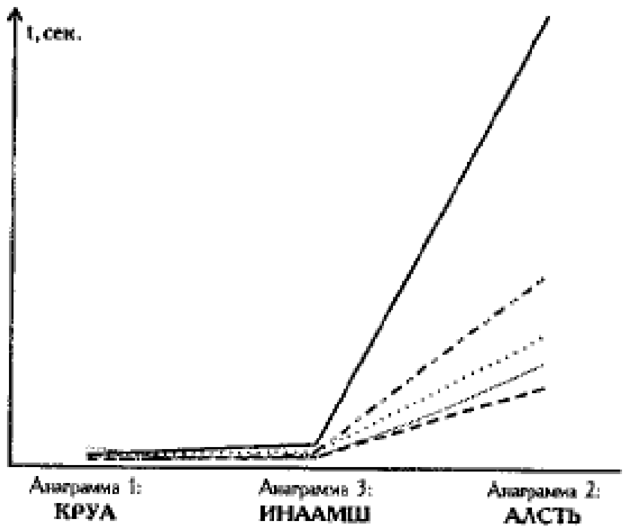


Рис. 5.9. Графіки зміни показників часу вирішення (сек.) анаграм п’яти випробовуваних в новій (гіпотетичній) послідовності їх представлення.

Символом достовірної, виразної тенденції в зміні показників при переході від умови до умови буде досить «зібрана» ламана крива, спрямована догори або, навпаки, до низу. Якщо на рис. 5.6 характерною рисою всіх індивідуальних кривих був крутий злам в одній і тій же точці графіка, то в цьому випадку на деяких відрізках підвищення кривій характеризується більшою крутістю, а на інші - меншою крутістю. Очевидно, ймовірність тенденцій буде забезпечуватися саме відрізками більше крутого сходження, але тест тенденцій поблажливо поширить цей ефект і на більше пологі відрізки.

На рис. 5.11 графіки представлені вже для рангування показників. Тут уже всі розходження в крутості згладжені. **L**-тест побудований на зіставленні сум рангів, а рангування неминуче трохи огрубляє отримані показники. Досвід показує, однак, що **L**-тест є досить потужним критерієм, хоча й обмеженим по сфері застосування через відсутність таблиць критичних значень для більших *n*.

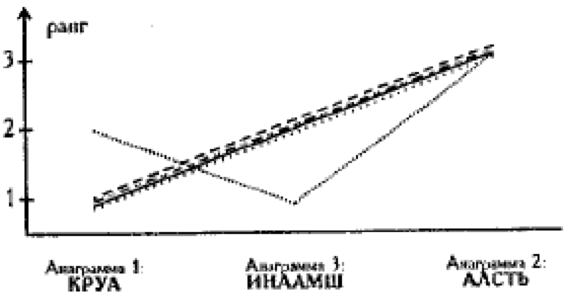


Рис. 5.11. Графіки зміни рангованих показників часу рішення анаграм п'ятьома випробуваними *в* новій (гіпотетичної) послідовності їх представлення.

Обмеження критерію Пейджа.

1. Нижній поріг - 2 випробуваних, кожний з яких пройшов не менш 3-х вимірів у різних умовах. Верхній поріг - 12 учасників експерименту і 6 умов (n<12, c<6). Критичні значення критерію **L** дані під керівництвом J.Greene, M. D'Olivera (1989р.), Вони передбачають три рівні статистичної достовірності: p<0,05; p<0,01; p<0,001.
2. Необхідною умовою застосування тесту є впорядкованість стовпців даних: ліворуч повинен розташовуватися стовпець із найменшою ранговою сумою показників, праворуч - з найбільшої. Можна просто пронумерувати заново всі стовпці, а потім вести розрахунки не ліворуч праворуч, а по номерах, але так легше заплутатися.

Алгоритм підрахунку критерію тенденцій L- Пенджа:

1. Прорангувати індивідуальні значення першого піддослідного, отримані ним в 1-м, 2-м, 3-м і т.д. вимірах. При цьому першим може бути будь-який піддослідний, наприклад перше за алфавітом ім’я.
2. Проробити те ж саме стосовно всіх інших піддослідних.
3. Просумувати ранги за умовами, у яких здійснювалися заміри. Перевірити збіг загальної суми рангів з розрахунковою сумою.
4. Розташувати всі умови а порядку зростання їхніх рангових сум у

$= \sum T_j \cdot j$

таблиці.

5. Визначити емпіричне значення L по формулі: L
по даній умові; де T_j - сума рангів,
j - порядковий номер, приписаний даній умові в упорядкованій послідовності умов.

6. За табл. 5.10 визначити критичні значення L для даної кількості випробуваних n і даної кількості умов c. Якщо $L_{емп}$ дорівнює критичному значенню або перевищує його, тенденція достовірна.

7.

Таблиця 5.10

Кількість респондентів n	c (кількість умов)				p
	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
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

***Q* - критерій Розенбаума.** Критерій використовується для оцінки відмінностей між двома вибірками за рівнем довільної кількісно вимірювальної ознаки. У кожній із вибірок має бути не менше 11 учасників експерименту. Це дуже простий не параметричний критерій, який дозволяє швидко оцінити відмінності між двома вибірками за довільною ознакою. Проте, якщо критерій Q не виявив достовірних відмінностей, то абсолютно не означає, що вони дійсно відсутні. У такому випадку краще застосувати критерій φ^* Фішера. Якщо ж Q-критерій виявляє достовірні відмінності між вибірками з рівнем достовірності $p<0{,}01$, то можна обмежитися тільки ним.

Критерій застосовують тільки у тих випадках, коли дані представлені принаймні у порядковій шкалі. Ознака має варіювати в довільному діапазоні значень, в протилежному випадку порівняння за допомогою Q– критерія просто неможливе. *Наприклад*, якщо у нас тільки 3 значення ознаки, 1, 2 і 3 – нам дуже важко встановити відмінності. Метод Розенбаума вимагає, достатньо «тонко» виміряних ознак.

Застосування критерія почнемо з того, що впорядкуємо значення ознаки в обох вибірках по зростанню (чи спаданню) ознаки. найкраще, коли дані кожного учасника експерименту представлені на окремій карточці. Тоді достатньо легко впорядкувати два ряда значень за ознакою яка нас цікавить – розкравши картки на столі. Так мы відразу побачимо чи співпадають діапазони значень, і якщо ні, то наскільки один ряд значень "вище" (S_1), а другий – "нище" (S_2). Для того, щоб не запутатися, в цьому та інших критеріях рекомендується першим рядом (вибіркою, групою) вважати тот ряд, де значення вище, а другим рядом – той, де значення нижщі.

Гипотези:

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

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

Графічне представлення критерія *Q*.

На рис. 5.12 представлено три варіанти співвідношення рядів значень у двох вибірках. У варіанті (а) всі значення першого ряда вищі усіх значень другого ряду. Відмінності, безумовно, достоверны, при соблюдении условия, что $n_1, n_2 \geq 11$.

У варіанті (б), навпаки, обидва ряди знаходяться на одному і томуж самому рівні: відмінності не достовірні. У варіанті (в) ряди частково перетинаються, але все таки перший ряд вищий ніж другий. Чи достатньо великі зони S_1 і S_2 (в сумі Q), можна визначити за таблицею критичних значень Q для разних n. Чим більша Q, тим більш достовірні відмінності ми можемо константувати.

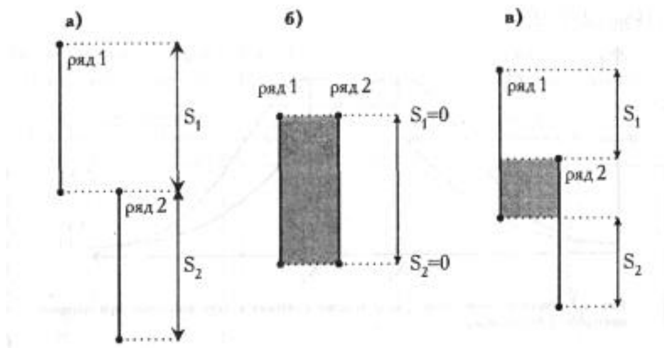


Рис. 5.12. Можливі співвідношення рядів у двох вибірках:
S_1 – зона значень 1-го ряду, які більші максимального значення 2-го ряду;
S_2 – зона значень 2-го ряду, які менші мінімального значення 1-го ряду;
пунктиром відмічені зони двох рядів, що перетинаються.

Обмеження критерію *Q*.

1. У кожній вибірці, що співставляється має бути не менше 11 спастережень. При цьому об'єми вибірок приблизно мають співпадати. Е.В. Гублером зазначаються наступні правила:

а) якщо в кожній вибірці спостережень менше 50, то абсолютна величина різниці між n_1 *і* n_2 не повина бути більше 10 спостережень;

б) якщо в кожній вибірці спостережень більше 51, але менше 100, то абсолютна величина різниці між між n_1 *і* n_2 не повина бути більше 20 спостережень;

в) якщо в кожній вибірці спостережень більше 100, то допускається, щоб одна із вибірок була більше іншої не більше ніж у 1,5-2 раза (Гублер Е.В., 1978, с. 75).

2. Діапазони розсіювання значень у двох вибірках мають не співпадати між собою, інакше не має сенсу застосовувати даний критерій. Але можливі випадки, коли, діапазони розсіювання значень співпадають, але, унаслідок різносторонньої асиметрії двох розподілів, відмінності в середніх величинах ознак суттєві (рис. 5.13, 5.14).

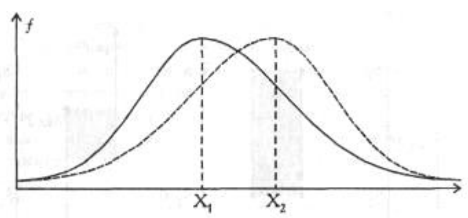


Рис. 5.14. Варіант співвідношення розподілів ознаки у двох вибірках, при якому критерій *Q* не діє.

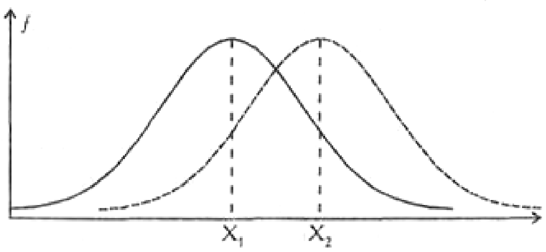


Рис. 5.15. Варіант співвідношення розподілів ознаки у двох вибірках, при якому критерій *Q* може бути суттєвим.

Приклад. У ймовірних учасників психологічного експерименту, що моделює діяльність повітряного диспетчера, було виміряно рівень вербального та невербального інтелекту за допомогою методики Д. Векслера. Було досліджено 26 юнаків у віці від 18 до 24 років (середній вік 20,5 років). 14 із них були студентами фізико-математичного факультету, а 12 - студентами факультету психології. Показники вербального інтелекта представлені в таблиці 5.11.

Чи можна стверджувати, що одна із груп має переваги над другою за рівнем вербального інтелекту?

Таблица 5.11

Індивідуальні значення вербального інтелекту у вибірках студентів фізико-математичного факультету (n_1=14) та факультету психології (n_2=12)

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

Упорядкуємо значення в обох вибірках, а потім сформулюємо гіпотези:

H_0: Студенти-фізики не мають переваг над студентами-психологами за рівнем вербального інтелекту.
H_1: Студенти-фізики мають переваги над студентами-психологами за рівнем вербального інтелекту.

Таблиця 5.12

Упорядковані ряди індивідуальних значень вербального інтелекту (за спаданням) у двох студентських вибірках

1 ряд - студенты-физики				2 ряд - студенты-психологи			
1	См.А.	136					
2	Б.Л.	136					
3	Ф.В.	136	S_1				
4	С.А.	135					
5	К.А.	134					
6	И.А.	132		1	Е.В.	132	
7	П.А.	132					
8	Ст.А.	132					
9	Ф.А.	132					
10	Т.А.	131					
11	К.Ан.	129					
12	Ц.А.	127		2	О.В.	127	
				3	Н.Т.	126	
				4	И.Ч.	126	
13	К.Е.	124					
				5	К.О.	123	
				6	О.К.	123	
14	Ч.И.	121					
				7	Ф.О.	120	
				8	И.В.	120	
				9	Р.Р.	120	S_2
				10	И.Н.	119	
				11	Р.И.	116	
				12	Н.К.	115	

Як видно із таблиці 5.11, ми правильно позначили ряди: перший, той, що «вище» - ряд фізиків, а другий, той, що «нижче» - ряд психологів.

За таблицею 5.12 визначаємо кількість значень першого ряду, які більші за максимальне значення другого ряду: $S_1=5$. Потім визначаємо кількість значень другого ряду, які менші мінімального значення першого ряду: $S_2=6$. Вираховуємо $Q_{емп}$ за формулою:

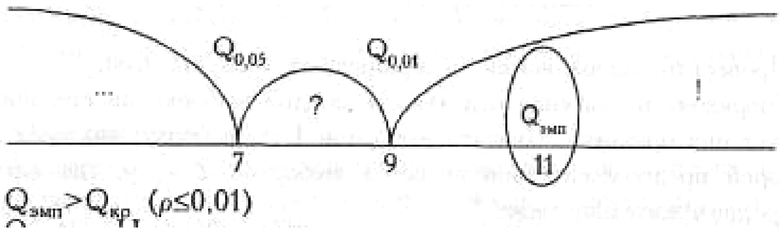
$$Q_{емп}=S1+S_2=5+6=ll.$$

За таблицею теоретичних значень визначаємо критичне значення Q для $n_1=14$, $n_2=12$:

$Q_{кр}=7$ $(p\leq 0,05)$ *та* $Q_{кр}=9$ $(p\leq 0,01)$.

Як бачимо, чим більша різниця між вибірками, тим більша величина Q. H_o відхиляється (див. рис.) при $Q_{емп}>Q_{кр}$, а при $Q_{емп}<Q_{кр}$ ми будемо вимушені прийняти H_o.

Побудуємо «вісь достовірності» (рис. 5.15).



Приймається H_1. Студенти-фізики мають переваги над студентами-психологами за рівнем вербального інтелекту (p<0,01). Зазначимо, що у тих випадках, коли емпірична величина критерію виявляється на межі зони не достовірності, ми маємо право стверджувати, що відмінності достовірні при p<0,05, якщо ж вона виявляється між двома критичними значеннями, то ми можемо стверджувати, що p< 0,05.

Якщо емпіричне значення критерію виявляється на межі зони достовірності, p<0,01, в зоні достовірності - p<0,01. Оскільки рівень достовірності виявлених відмінностей достатньо високий (p<0,01), мы могли б на цьому зупинитися. Але можна спробувати зіставити вибірки за рівнем не вербального інтелекту, оскільки саме невербальний інтелект визначає рівень інтелекту вцілому та ступінь його організованості.

Повернемося до цього прикладу під час розгляду критерія Манна-Уітні та спробуємо відповісти на питання про співвідношення рівнів невербального інтелекту у дфох вибірках.

Алгоритм підрахунку критерія Розембаума.

1. Перевірити чи виконуються обмеження $n_1 n_2 \geq 11$, n_1 $n_2 \approx n_2$.
2. Упорядкувати значення окремо у кожній вибірці по ступуню зростання ознаки. Вважати вибіркою першою ту вибірку, значення в якій ймовірно більше, а другою ту, де значення ймовірно менше.
3. Визначити найбільше (максимальне) значення у другій вибірці.
4. Підрахувати кількість значень у вибірці 1, які більші максимального значення у другій вибірці. Позначити отриману величину S_1.
5. Визначити найменше (мінімальне) значення у другій вибірці.
6. Підрахувати кількість значень у другій вибірці, які менші мінімального значення першої вибірки. Позначити отриману величину як S_2.
7. Порахувати емпіричне значення Q за формулою: $Q = S_1 + S_2$

8. За таблицею визначити критичне значення Q для даних n_1 *і* n_2. Якщо $Q_{емп}$ дорівнює $Q_{0,05}$ або перевищує його, H_0 відкидається.
9. При $n_1 * n_2 > 26$ співставити отримане емпіричне значення з $Q_{кр}=8$ ($p \leq 0{,}05$) і $Q_{кр}=10$ ($p \leq 0{,}01$). Якщо $Q_{емп}$ більше або дорівнює $Q_{кр}=8$, H_0 відкидається.

Використання χ^2 - методу для визначення достовірності різних якісних ознак. У педагогічній дослідницькій роботі дуже часто доводиться оцінювати достовірність таких явищ, які неможливо безпосередньо вимірювати і які не відповідають нормальному розподілу.

Приклад. У 40 учнів запитали: Що подобається їм більше: працювати з робочим зошитом чи без нього. Учні повинні були відповісти на перше питання так або ні. Якщо учень відповів на перше питання «так», то відповідь на друге питання була «ні». В даному випадку ми маємо справу з альтернативними ознаками явища, тобто з такими ознаками, при яких необхідно вибрати одну з двох можливостей.

Результати оцінок заносяться в таблицю по 4 - польній схемі таким чином (табл. 5.11).

Таблиця 5.11

Результати оцінок

Метод	Оцінка		
	Так	Ні	Всього
1. Робота з робочим зошитом	30(A)	10(B)	40(A+B)
2. Робота без робочого зошита	10(C)	30(D)	40(C+D)
Всього:	40(A+C)	40(B+D)	80(A+B+C+D)

З таблиці з'ясовується, що прихильників використання робочого зошита в тричі більше.

Для перевірки, чи є така оцінка статистично достовірною чи ні, використовують χ^2- тест, χ^2 обчислюється за формулою:

$$\chi^2 = \frac{(A-B-1)^2}{A+B}, \qquad (5.19)$$

де A - найбільше число 4 - польної схеми; B - найменше число схеми.

У нашому прикладі: $\chi^2 = \frac{(30-10-1)^2}{30+10} = \frac{361}{40} = 9{,}025.$

За ступінь вільності тут береться кількість стовпців мінус одиниця. У даному випадку число ступенів вільності *2-1=1*.

Порівнюючи отримане значення χ^2 з табличним значенням χ^2 з'ясовуємо, що $\chi^2_{emp} > \chi^2_{krt99\%} > \chi^2_{krt95\%}$. Отже, таких значень χ^2 можна набути при випадковому відборі учнів з ймовірністю більше 99%.

Для порівняння результатів двох вибірок можна ще, окрім названих критеріїв, застосовувати і інші непараметричні критерії.

У таблиці 5.12 наведено можливості та обмеження критеріїв.

Таблиця 5.12

Можливості та обмеження критеріїв

ПАРАМЕТРИЧНІ КРИТЕРІЇ	НЕПАРАМЕТРИЧНІ КРИТЕРІЇ
1. Дозволяють прямо оцінити відмінності в середніх, що отримані в двох вибірках (t - критерій Стьюдента).	Дозволяють оцінити лише середні тенденції, наприклад, дати відповідь на питання: чи частіше у вибірці А зустрічаються більш високі, а у вибірці Б – більш низькі значення ознаки (критерій Q, U, φ* та інш.).
2. Дозволяють прямо оцінити відмінності в дисперсіях (критерій Фішера).	Дозволяють оцінити лише відмінності в діапазонах варіативності ознаки (критерій φ*).
3. Дозволяють виявити тенденції зміни ознаки при переході від умови до умови (дисперсійний однофакторний аналіз), але лише за умови нормального розподілу ознаки.	Дозволяють виявити тенденції зміни ознаки при переході від умови до умови при будьякому розподілі ознаки (критерії тенденцій L та S).
4. Дозволяє оцінити взаємодію двох і більше факторів в їх впливі на зміну ознаки (двухфакторний дисперсійний аналіз).	Дана можливість відсутня.
5. Експериментальні дані мають задовольняти дві або три умови: а) значення ознаки вимірюються за интервальною шкалою; б) розподіл ознаки є нормальним; в) у дисперсійному аналізі має виконуватися умова рівності дисперсій в комірках комплекса.	Експериментальні дані можуть не задовольнити жодну з цих умов: а) значення ознаки може бути представленне у бідь якій шкалі, починаючи зі шкали найменувань; б) розподіл ознаки може бути будь-яким і його спів падання з довільним теоретичним законом розподілу не є обов'язковим та не потребує перевірки; в) вимога рівності дисперсій відсутня.
6. Математичні розрахунки досить складні.	Математичні розрахунки відносно прості (за виключенням критеріїв χ^2 *та* λ).
7. Якщо умови, що зазначені в п.5 виконуються то параметричні критерії виявляються більш потужними ніж не параметричні.	Якщо умови, що зазначені в п.5 не виконуються то непараметричні критерії виявляються більш потужними ніж параметричні, томущо вони менш чутливі до «засмічування».

Як бачимо з таблиці параметричні критерії можуть виявитися дещо «сильнішими» ніж непараметричні. Це можливо лише за умови, що якщо ознака виміряна за інтервальною шкалою та нормально розподіленна.

Непараметричні критерії не мають таких обмежень та не потребують довгих та складних викладок. У порівнянні з параметричними вони обмежені лише в одному – за їх допомогою не можна оцінити взаємодію двох чи більше умов (факторів), які впливають на зміну ознаки.

Тема 6.

Методи визначення зв'язку між явищами

Кореляція та її види.

За характером залежності явищ розрізняють функціональні (повні) та кореляційні (неповні) зв'язки.

Функціональний зв'язок передбачає, що певному значенню факторної ознаки завжди відповідає одне або кілька значень результативної ознаки. Функціональні зв'язки характеризуються повною відповідністю між причиною і наслідком. Завдяки цьому функціональну залежність описують точні математичні формули. Функціональні залежності вивчають точні науки, такі як математика, фізика, хімія тощо. У суспільних процесах – це переважно зв'язок між елементами розрахункових формул, наприклад, залежність фондовіддачі від обсягу виробництва продукції та вартості основних виробничих фондів або прямо пропорційна залежність між продуктивністю праці та виробництвом продукції.

Кореляційний зв'язок проявляється в середньому, для масових спостережень, коли кожному значенню ознаки *x* відповідає певна множина ознаки *y*, які варіюють і утворюють ряд розподілу. Прикладом такого зв'язку можна навести залежність між кольором очей та кольором волосся.

За напрямком зв'язок буває *прямим*, коли залежна змінна збільшується із збільшенням факторної ознаки; та *зворотнім*, при якому зростання факторної ознаки спричиняє зменшення результату.

За своєю аналітичною формою зв'язки бувають *лінійними та нелінійними*. В першому випадку між ознаками проявляються в середньому лінійні співвідношення. Нелінійний взаємозв'язок виражається нелінійною функцією, а змінні пов'язані між собою в основному нелінійно.

За кількістю факторів, що розглядаються розрізняють *парний* (якщо характеризується зв'язок двох ознак) та *множинний, або багатофакторний* зв'язок (якщо вивчаються більш ніж дві змінні).

За силою розрізняють *слабкий та сильний* зв'язок. Ця характеристика виражається конкретними величинами.

У найбільш загальному вигляді завдання статистики в галузі вивчення взаємозалежностей складається в кількісній оцінці їх наявності та напрямку, а також в характеристиці сили та форми впливу одних факторів на інші. Для його вирішення застосовують методи кореляційного та регресійного аналізу.

Завдання кореляційного аналізу полягають в вимірі щільності зв’язку між ознаками, визначенні невідомих причинних зв’язків і в оцінці факторів, що мають найбільший вплив на результативну ознаку.

Регресійний аналіз має на меті встановлення форми залежності, визначення функції регресії, використання рівняння для оцінки невідомих значень залежної змінної.

У педагогічних дослідженнях нерідко стає дуже важливим виявити причинно-наслідкові зв'язки між двома різними змінними, наприклад, між засвоєнням матеріалу двох предметів, циклів предметів тощо. Діалектичне розуміння розвитку припускає, що кожний предмет, кожне явище формуються під впливом не тільки необхідних, істотних, але і випадкових, несуттєвих факторів. Між випадковими величинами може існувати тільки зв'язок особливого роду, при якому зміна одного параметра викликає зміну іншого (стохастичний або кореляційний зв'язок). На відміну від функціональної залежності, при якій одній із взаємопов'язаних величин відповідає одне конкретне значення іншої, кореляційний зв'язок означає, що деякому значенню однієї сукупності може відповідати декілька значень іншої сукупності, які зазвичай точно не визначені.

Серед досліджуваних у педагогіці характеристик випадковими можуть бути: кількість помилок, що допускаються учнями (студентами); час, витрачений на виконання якогось завдання; оцінки, отримані учнями (студентами) (у деякій мірі це умовний параметр) та багато іншого.

Між зазначеними параметрами існує стохастичний зв'язок. Крім нього є також випадкові зв'язки, наявність яких залежить від умов одержання даних.

Кореляція буває повною (завершеною) або частковою, але вона може бути і відсутньою.

Одним з основних показників стохастичного зв'язку є **коефіцієнт кореляції** $\boldsymbol{r}$, що служить кількісним вираженням цього зв'язку. Повну кореляцію прийнято позначати 1($r=1$), часткову - $0<r<\pm1$; відсутність кореляції - 0($r=0$). Знак коефіцієнта кореляції (±) показує, який зв'язок існує між досліджуваними процесами, явищами - прямий чи зворотній. Якщо $r=+1$, то це означає, що існує пряма повна кореляція, тобто збільшення одного з двох параметрів викликає збільшення й іншого. Значення коефіцієнта кореляції $r= -1$ вказує на повну зворотну кореляцію - в цьому разі при збільшенні одного з двох параметрів зменшується інший. Коефіцієнти кореляції в межах $0< r< \pm 1$ означають існування зв'язку, причому щільність зв'язку визначається абсолютним значенням коефіцієнта, а характер кореляції знаком коефіцієнта.

Коефіцієнти кореляції (розміри рангового порядку) можуть оцінюватися і порівнюватися. Тому окремо узятий коефіцієнт кореляції може бути відносно низьким, середнім або високим.

При інтерпретації величини коефіцієнта кореляції звичайно вважають, що коефіцієнт від 0 до ± 0,20 означає незначний ступінь кореляції (прирівнюється до 0); від ± 0,20 до ± 0,40 – низький ступінь кореляції; від ± 0,40 до ± 0,70 - яскраво виражену кореляцію; від ± 0,70 до ± 1,00 – високу або дуже високий ступінь кореляції.

Коефіцієнти кореляції $r=+1$ *і* $r=-1$ у педагогічних дослідженнях практично не зустрічаються.

Слід зазначити, що коефіцієнт кореляції не дає підстав робити висновки про причини чи умови зв'язків. Він лише показує і підтверджує, що між певними явищами існує взаємний зв'язок, але не з'ясовує, чи зумовлене друге явище першим, перше другим, або перше і друге зумовлене третім.

Кореляційним моментом називають математичне сподівання я добутків відхилень випадкових величин x та y від їх середніх:

$$\mu = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{n}$$

Мірою кореляційного зв’язку між ознаками x та y є коефіцієнт кореляції Пірсона (коефіцієнт лінійної парної кореляції), який розраховується на основі парної вибірки значень величин x та y за формулою

$$r_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2)(\overline{y^2} - \bar{y}^2)}}, \quad (6.1)$$

де N – об’єм вибірки; y_i, x_i – значення і-того елемента вибірок $\boldsymbol{x}$ та $\boldsymbol{y}$ відповідно; $\bar{x}$, $\bar{y}$ – вибіркові середні $\boldsymbol{x}$ та $\boldsymbol{y}$ відповідно; $\overline{xy}$ – середнє значення добутку x_i і y_i; $\overline{y^2}, \overline{x^2}$ – середнє значення квадратів ознак y та x.

Коефіцієнт кореляції Пірсона є мірою **тільки лінійного зв'язку**, що є його суттєвим обмеженням. Застосування цієї міри є коректним тільки тоді, коли вигляд кореляційне поля вказує на наявність лінійного зв'язку. Коли зв'язок нелінійний, навіть явно видимий з кореляційного поля, коефіцієнт кореляції Пірсона не дає правильних результатів – наприклад, показує відсутність зв'язку.

У випадку нелінійного зв'язку застосування застосування коефіцієнта кореляції Пірсона можливе після лінеаризації експериментальної залежності (наприклад, логарифмування даних, які підпорядковуються експоненціальній залежності). У випадку складної нелінійної залежності можлива її кусково-лінійна апроксимація, тобто заміна нелінійної залежності лінійною на окремих ділянках (на графіку залежності це відповідає заміні окремих ділянок графіка нелінійної функції прямолінійними відрізками так, щоб вони мали відхилялись від графіка нелінійної залежності). Після цього коефіцієнт кореляції Пірсона коректно застосовується для оцінки сили лінійного зв'язку на окремих ділянках. При цьому має місце значне збільшення обсягів обчислень і ускладнення математичної моделі залежності.

Коефіцієнт кореляції Пірсона дозволяє оцінити силу (глибину, тісноту) лінійного зв'язку між величини, а також визначити напрям зв'язку.

Графічне зображення кореляційного зв'язку між величинами *x і y* називають *лінією регресії*, а відображувана цією лінією функція *y=y(x) – функцією регресії y на x*. Коли обидві функції регресії *y=y(x) і x=x (y),* що характеризують кореляційну залежність між величинами *y і x*, мають лінійний характер, то така залежність носить назву лінійної кореляції. Лінії регресії, котрі відповідають лінійній кореляції, є прямими, і їх називають прямими регресії.

Рівняння прямої регресії *y на x* у загальному випадку має вигляд

$$y=Ax+B. \qquad (6.2)$$

Якщо за результатами експерименту отримані n пар дослідних значень, то найбільш ймовірнісні значення коефіцієнтів А і В відповідно до методу найменших квадратів будуть дорівнювати:

$$\mathrm{A} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}, \quad \mathrm{B} = \frac{n\sum x^2 \sum y - \sum x \sum xy}{n\sum x^2 - (\sum x)^2}, \qquad (6.3)$$

де x , y, xy і x^2 - середні арифметичні значення величин x, y, xy і x^2. При необмеженому збільшені числа дослідних точок n формули наберуть вигляду:

$$\mathrm{A} = \frac{M(xy) - M(x)M(y)}{M(x^2) - [M(x)]^2} = \frac{M(xy) - M(x)M(y)}{\sigma_x^2}, \qquad (6.3)$$

$$B = M(y) - AM(x).$$

де M – математичне очікування відповідних величин; σ_x = А – дисперсія величини x. Отриманий за формулою кутовий коефіцієнт А називають коефіцієнтом регресії *у на х* і позначають p.

Тіснота зв'язку між величинами y і x, що відповідає ступеню наближення дослідних точок до прямих регресії, визначається величиною коефіцієнта кореляції

$$r_{xy} = \frac{M(xy) - M(x)M(y)}{\sigma_x\ \sigma_y} = \frac{K(xy)}{\sigma_x\ \sigma_y}, \qquad (6.4)$$

$$B = M(y) - AM(x).$$

де σ_x , σ_y - середні квадратичні відхилення величин x і y.

Чисельник у (6.4) є кореляційним моментом між величинами x і y, залежить від одиниць, у яких виражають x і y, і тому сам по собі не може бути показником зв'язку.

Коефіцієнт кореляції r_{xy} лежить в межах від -1 до +1 і характеризує ступінь наближення залежності між величинами x і y до лінійної функціональної залежності, котрій відповідають граничні значення коефіцієнта кореляції ±1. Якщо $r_{xy}>0$, наявна позитивна кореляція, яка означає, що при зростанні однієї з величин друга так само дещо зростає. У випадку $r_{xy}<0$ має місце від'ємна кореляція, коли при зростанні однієї з величин друга має тенденцію дещо зменшуватися.

У разі відсутності статистичного зв'язку між величинами x і y коефіцієнт кореляції $r_{xy}=0$, тому що в цьому випадку $M(xy)=M(x)M(y)$. Але протилежне твердження слід вважати неправомірним. При $r_{xy}=0$ величини x і y все таки можуть знаходитися в тісному кореляційному зв'язку, якщо лінії регресії значно відрізняються від прямих. Сюди відносяться, зокрема, випадки, коли лінії регресії мають вісь симетрії, паралельну до однієї з координатних осей *х або у*.

Коефіцієнти регресії мають той самий знак, що і коефіцієнт кореляції. Коефіцієнти $\sigma_{x/y}$ і $\sigma_{y/x}$ утворюють пару спряжених коефіцієнтів регресії.

Кореляційно-регресійний аналіз є задачею, яка ефективно розв'язується за допомогою електронних таблиць. Типовим прикладом застосування електронних таблиць до кореляційно-регресійного аналізу є процесор електронних таблиць MS Excel (і його вільний аналог Open Office Calc). Розрахунок коефіцієнта кореляції Пірсона реалізується спеціальною вбудованою функцією КОРРЕЛ. Обчислення коефіцієнтів регресії ефективно реалізується за допомогою формул робочого аркуша, побудова яких полегшується наявністю вбудованих функцій, які реалізують проміжні розрахунки (такі як СУММПРОИЗВ – обчислює суму добутків відповідних елементів масивів, СУММКВ – обчислює суму квадратів тощо).

До складу ПЕТ MS Excel входить надбудова Пакет аналізу, яка являє собою набір спеціалізованих інструментів, які реалізують типові складні функції статистичного аналізу даних. До складу пакету входять зокрема інструменти кореляційно-регресійного аналізу. Це – інструменти **Корреляция** и **Регрессия**. Обидва інструменти забезпечують виконання багатофакторного кореляційного і регресійного аналізу, включаючи 1-факторну модель як частинний випадок.

Вибір інструменту здійснюється у вікні надбудови Пакет аналізу (рис. 6.1) зі списку інструментів.

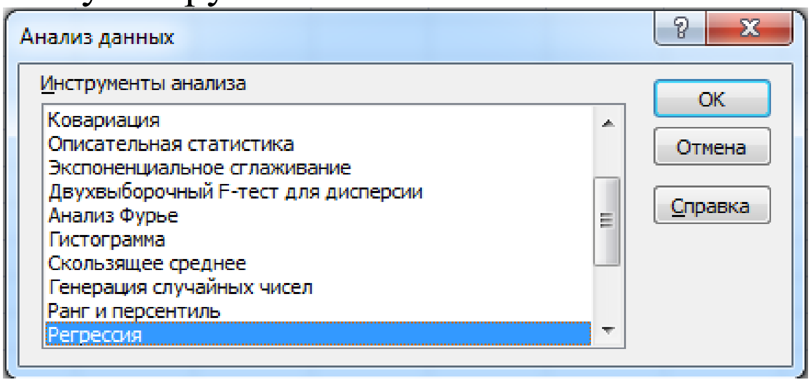


Рис. 6.1. Вікно меню надбудови **Пакет аналізу**.

Налаштування обраного інструменту виконується у його діалоговому вікні, де задаються діапазони даних та деякі допоміжні параметри. Діалогове вікно інструменту **Корреляция** показане на рис. 6.2, інструменту **Регрессия** – на рис. 6.3.

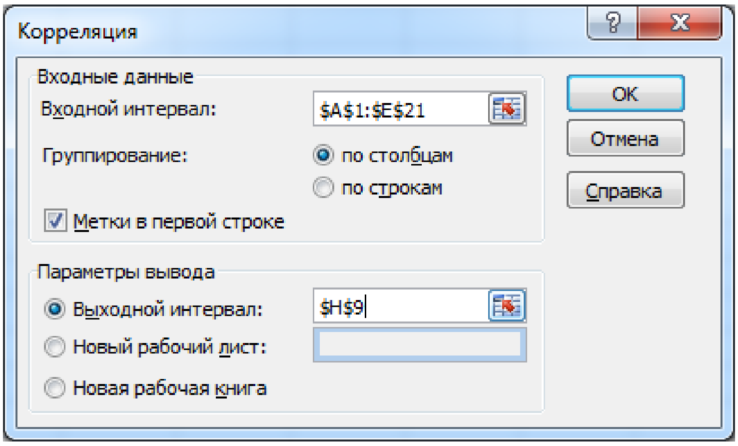


Рис.6.2. Діалогове вікно інструменту **Корреляция**.

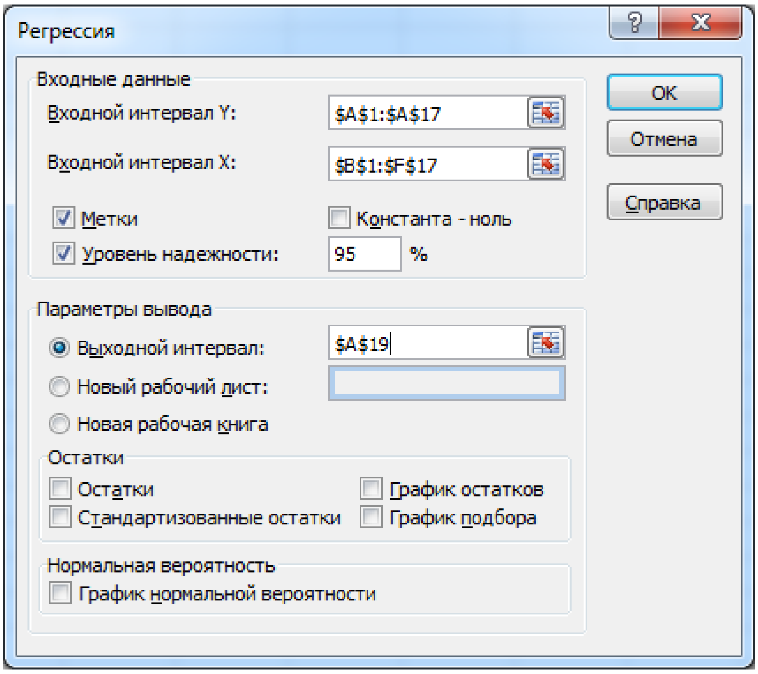


Рис. 6.3. Діалогове вікно інструменту **Регрессия**.

Результатом обробки даних інструментом **Корреляция** є кореляційна матриця, яка являє собою квадратну нижню трикутну матрицю, яка містить попарні коефіцієнти кореляції Пірсона усіх величин $Y, X_1, \ldots X_n$.

У результаті обробки даних інструментом **Регрессия** створюється звіт, який відображається у заданому місці і містить коефіцієнти регресії і ряд статистичних оцінок, які характеризують значущість отриманої регресійної моделі і потрібні для подальшого поліпшення моделі.

Порядок проведення кореляційного аналізу

1) побудувати точкову діаграму;
2) позначити кореляційне поле;

3) позначити емпіричну лінію регресії;
4) зробити попередній висновок;
5) обчислити середні вибіркові $\bar{y}$ та $\bar{x}$;
6) обчислити коефіцієнт кореляції за формулою:

$$r_{xy} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(x_i - \bar{x})^2}};$$

7) оцінити характер (прямий чи зворотній), та глибину кореляційного зв'язку за таблицею:

r	*Глибина зв'язку*	*r*	*Глибина зв'язку*
$r=0$	відсутня	$0,7<r\le 0,9$	сильна
$0<r\le 0,3$	слабка	$0,9<r<1$	дуже сильна
$0,3<r\le 0,5$	помірна	$r=1$	повна
$0,5<r\le 0,7$	значна		

8) оцінити ймовірність коефіцієнта кореляції $r_{x,y}$:

а) $m_r = \pm\frac{1-r^2}{\sqrt{n}}$, якщо $n\to\infty$;

$m_r = \pm\sqrt{\frac{1-r^2}{n-2}}$, якщо $n<100$;

б) $t_r = \frac{r}{m_r}$;

в) $\nu=n-2$, де n – кількість кореляційних пар;

г) знаючи ν, знайти за таблицею критерії Стьюдента t_{st} для трьох порогів ймовірності α_1, α_2, α_3: $t_{st} = \{t_{0,95}, \ldots t_{0,99}, \ldots t_{0.999}\}$;

д) порівняти t_r і t_{st} та зробити висновок щодо вірогідності коефіцієнта кореляції.

Порядок проведення регресійного аналізу (для лінійного зв'язку) методом найменших квадратів.

1) $a = \frac{\sum_{i=1}^{n}(y_i - \bar{y})\cdot(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$;

2) $b = \bar{y} - a\bar{x}$;

3) записати рівняння $y=ax+b$ та побудувати графік.

Тема 7.

Визначення критеріїв тестів успішності та оформлення результатів дослідження

Складання та методи перевірки тесту.

Тест успішності може бути придатним засобом вимірювання знань та умінь учнів, лише в тому випадку, коли він складений дуже ретельно і перевірений. Складання і перевірка предметного тесту повинна проходити паралельно. Складання тесу повинно закінчуватись його попереднім контролем.

Далі розглянемо деякі математичні методи обрахування показників тесту успішності, які необхідно враховувати,як при складанні, так і при контролі тексту.

Об'єктивність. Тест успішності вважається об'єктивним тоді, коли його результати не залежать від особистості людини, яка проводить роботу, оцінює, підводить підсумки та інтерпретує отримані результати. Об'єктивність характеризується кореляцією між результатами, отриманими двома особами, що оцінюють. В цьому випадку необхідно, щоб коефіцієнти були близькими до одиниці ($r=1$).

Для забезпечення об'єктивності проведення оцінки і обробки результатів роботи повинні бути строго нормованими. Для цього потрібно скласти точні інструкції щодо проведення роботи, виправлення помилок і оцінки результатів роботи. Суворо потрібно слідкувати за тим, щоб учні не списували роботу один в одного. Щоб цього уникнути тест потрібно проводити з зовсім окремими один від одного учнями в двох чи навіть в чотирьох паралельних варіантах.

Валідність. Валідність тесту успішності характеризується кореляцією між результатами тесту і будь-яким вибраним критерієм валідності. Такими критеріями можуть бути оцінка знань учнів учителем, результати, отримані за допомогою інших методів оцінювання (наприклад, опитування чи контрольної роботи), які проаналізовані досвідченими методистами чи вчителями.

Найлегше для оцінки валідності тесту і контрольної роботи використовувати порядкову чи рангову кореляцію за Спірменом. (Q).

При визначенні валідності на основі цієї кореляції порівнюють дві групи учнів досліджуваної групи – група за знаннями на основі тесту успішності та оцінок вчителів, чи групи, що складена на основі оцінок в журналі протягом якогось періоду.

Валідність вважається достатньо високою, якщо $Q>0,6$, низькою, коли $Q>0,4$. Коефіцієнт кореляції 0,7–0,9 свідчить про високу валідність можна тесту якості знань. Якщо коефіцієнт досягає лише 0,45–0,65, то валідність можна вважати задовільною.

Діагностична цінність. Для визначення діагностичної цінності тесту проводять попередній експеримент з так званим нейтральним класом, результати якого в подальшій дослідницькій роботі не використовуються. Результати попередньої контрольної роботи розташовують у зростаючому порядку і визначають медіану результатів, тобто величину члена, що знаходиться всередині ряду. Учнів, які отримали оцінку вище медіани, вважають «слабкими», тих, хто отримав оцінку вище медіани, вважають «сильними».

Далі при кожному завданні вираховується кількість правильних і неправильних відповідей «сильними» і «слабкими». Результати заносяться до схеми чотирьох полів.

Припустимо, що на перше запитання з 15 «сильних» учнів 10 відповіло правильно, 5 – неправильно; зі «слабких» - 3 правильно, 12 неправильно.

Складемо схему чотирьох полів (табл. 6.1.).

Таблиця 6.1

Результати тесту успішності

	К-ть правильних відповідей	К-ть неправильних відповідей
«Сильні» учні	10	5
«Слабкі» учні	3	12

Діагностична цінність задачі являє собою відношення перехресних сум діагоналей чотирьох полів, тобто $\frac{10+12}{5+3} = \frac{22}{8} = 2.75.$

Критичною величиною діагностичної цінності задачі вважають 1,5. Якщо діагностична цінність отримана шляхом обрахунків; більша критичної величини 1,5, тоді задача відповідає контрольній роботі. Тобто вона має потрібну діагностичну цінність.

Потрібно звернути увагу на те, що при одній задачі поданій в попередній контрольній роботі, не доцільно ділити експериментальну групу на «сильних» і «слабких», це потрібно робити на основі оцінок учителя. Для оцінки діагностичної цінності (D) кожного питання правильні і неправильні відповіді учнів на кожне питання заносяться до таблиці 6.2.

Таблиця 6.2

Питання	Не правильні відповіді в «слабкій» групі V_N	Неправильні відповіді в «сильній» групі V_T	Залишок відповідей «сильної» і «слабкої» V_N - V_T	Сума відповідей «сильних» і «слабких» $V_N + V_T$
1	2	0	2	2
2	12	2	10	14
3	14	10	4	24
і т.д.				
$\sum$			80	60

Діагностична цінність D кожного питання обраховується за формулою:

$$D=\frac{K(V_N+V_T)}{2n\ (K-1)}100\%$$

де K – загальна кількість питання (відповідно і можливих відповідей);

N – кількість учнів в «сильній» («слабкій») групи;

$n = \frac{N*27}{100}$, де N загальна кількість учнів, які написали роботи;

V_N – помилки в «слабкій» групі;

V_T – помилки в «сильній групі».

Пропустимо, що для визначення діагностичної цінності питань контрольної роботи була проведена контрольна робота з 50 учнями, при цьому правильні і неправильні відповіді трьох перших питань розташувалися так, як наведено в таблиці (в обох групах було 14 учнів):

$n = \frac{50*27}{100} = 14$, кількість питань 15.

Діагностична цінність першого питання:
$\frac{15\,(2+10)}{2*14\,(15-1)} * 100\% \approx 7{,}7\%$
Діагностична цінність другого питання:
$\frac{15\,(12+2)}{2*14\,(15-1)} * 100\% \approx 54\%$
Діагностична цінність третього питання
$\frac{15\,(14+10)}{2*14\,(15-1)} * 100\% \approx 92\%$

Практично діагностичними вважають ті задачі, діагностична цінність яких становить від 16% до 84%. Якщо задача занадто легка і більшість учнів можуть розв'язати її, тоді діагностична цінність задачі нижча 16%; якщо задача занадто важка, і її не може розв'язати більшість учнів, тоді діагностична цінність цієї задачі більше 84%. З цього випливає,що перша задача, наведена в нашому прикладі, занадто легка, третя – занадто важка, а друга має задовільну діагностичну цінність.

На основі таблиці також можна зробити висновки про діагностичні цінність задач. На це вказує число *(V_N-V_T)*, наведене в четвертому стовпчику таблиці. Чим більше це число, тим вища діагностична цінність запитання. П'ятий стовпчик показує ступінь важкості задачі: чим більше це число, тим важча ця задача.

Інколи трапляється, що число, яке знаходиться в четвертому стовпчику від'ємне, тобто на деякі питання «сильні» відповідають неправильно, «слабкі» - правильно. Зазвичай такі питання неясно сформульовані, тому відповіді випадкові. Такі питання потрібно переформулювати по-іншому.

Діагностичну цінність тесту чи контрольної роботи можна легко визначити за допомогою знакового тесту. На основі попереднього досвіду вибирають «сильних» і «слабких» учнів (27%) і проводять з ними попередню контрольну роботу. Для кожної задачі відмічають в «слабкій » і «сильній» групах правильні і неправильні відповіді учня буквою П або Н. далі визначають відмінності знаків. Якщо загальна кількість однорідних знаків (або «+» або «-»), отримані шляхом обчислень, більше критичного числа, наведеного в таблиці, тоді задача має достовірну діагностичну цінність.

Хорошим способом перевірки валідності тесту чи контрольної роботи є і кореляція результатів тесту з оцінками учителя або зрівнювання з результатами отримані шляхом проведення тесту сформульованого по-іншому. У такому випадку потрібно також перевірити достовірність коефіцієнта кореляції.

Слід звернути увагу і на те, що безпосереднє вимірювання будь-якого явища завжди валідніше ніж опосередковане вимірювання того ж явища.

Після встановлення діагностичної цінності питань неможна виключати питання з контрольної роботи механічно, тому що низька діагностична цінність питання не завжди залежить від змісту, вона може залежати і від формулювання. Рекомендується перефразувати запитання, що мають малу діагностичну цінність і потім вдруге дати їм учням в попередньому експерименті.

Практично питання основного експерименту можуть бути ті, які в попередньому досліді мали приблизно 50% правильних і 50% неправильних відповідей.

Якщо тест чи контрольна робота складається з окремих запитань, які не обов'язково систематизувати за вмістом, тоді запитання основного експерименту повинні бути подані в тесті в порядку їх складності – спочатку більш легші, в кінці більш складніші. Тест чи контрольна робота має хорошу діагностичну цінність в тому випадку, коли за встановлений час на всі питання зможуть відповісти 90% учнів (дехто вважає 75%).

Якщо всі учні за виділений час зможуть відповісти на всі запитання, то тест занадто легкий, якщо за цей час на запитання зможуть відповісти менше 90% учнів, тест занадто важкий.

Темп розв'язування контрольної роботи чи тесту повинен бути достатньо напруженим. Якщо учні повинні навести в відповідях факти, то їм можна поставити за хвилину 3-4 питання, якщо для відповіді потрібне обмірковування, то 1-2 питання в хвилину, а при вибірковій відповіді – 2-3 питання в хвилину.

Надійність тесту успішності. Характеризується кореляцією між результатами двох паралельних (суворо однакової складності) варіантів одної роботи чи між результатами двохразового виконання цієї роботи в одному і тому ж учнівському колективі. При тесті успішності коефіцієнт надійності повинен бути $r > 0{,}7$ (коефіцієнт кореляції).

Методи перевірки тестів.

Існує кілька методів перевірки надійності тесту. При методі паралельного тесту для вимірювання одних і тих же знань конструюється два варіанти тесту. В обох варіантах повинно бути однакове число еквівалентних однотипних завдань.

Надійність тесту (контрольної роботи) буде перевірка з допомогою еквівалентного варіанту контрольної роботи (тесту). Складання еквівалентних тестів для природничо – математичних дисциплін достатньо легке, для гуманітарних більш складніше.

Метод паралельного тесту. Еквівалентність кожного питання можна перевірити методом χ^2 (на основі прикладу). Спочатку учні виконують контрольну роботу за першим варіантом, потім за другим. Далі обраховується кореляція між результатами контрольних робіт. Тест можна вважати надійним в тому випадку якщо коефіцієнт кореляції результатів тесту $r>0{,}7$.

Для визначення кореляції використовують формулу :

$$r = \frac{n\sum x_i\, y_i - \sum x_i\, y_i}{\sqrt{(n\sum x^2 - (\sum x_1)^2) * (n\sum y^2 - (\sum y_1)^2)}}$$

$$r = \frac{n\sum x_i\, y_i - \sum x_i\, y_i}{\sqrt{(n\sum x^2 - (\sum x_1)^2) * (n\sum y^2 - (\sum y_1)^2)}}$$

де, x_i – результати отримані в першому досліді;

y_i – результаті отримані в повторному досліді;

n – число досліджуваних.

Метод ділення навпіл. По суті не відрізняється від паралельного тесту. Різниця лише в тому, що один і той же тест розділяють на дві половини. Для рівноваги навчаючого і засвоєного ефекту рекомендується ділення за парними і непарними числами. Тест проводиться один раз у звичайному порядку. Результати тесту визначаються по пів тестах. До першого пів тесту входять завдання з непарним порядковим номером, до другого з парним. Визначаються результати кожного досліджуваного а двома пів тестами. Надійність тесту в цілому визначається за формулою

$$R = \frac{2r_{1,2}}{1 + r_{1,2}}$$

де $R -$ надійність тесту успішності;

$r_{1,2} -$ коефіцієнт кореляції між половинами тесту успішності.

Метод повторного тесту. Визначається стабільність тесту. Тест успішності проводиться два рази і обраховується коефіцієнт кореляції між отриманими результатами. Цей коефіцієнт називається коефіцієнтом самокореляції тесту або коефіцієнтом регресії і зазвичай позначаєься r_j. Тест має бути складений і його проведення організоване таким чином, щоб самокореляцію тесту $r_j>0,7$.

Метод повторного тесту дозволяє встановити стабільність тесту, тобто на скільки є стальними результати тесту проведеного в різний час. Метод ділення навпіл і метод паралельних тестів вимірюють також внутрішню сталість тесту. Остання показує в якій мірі завдання тесту вимірюють одну і ту саму властивість.

Метод порівняння груп. Одним з найпростіших методів визначення надійності контрольних робіт є порівняння «сильних» і «слабких» груп, отриманих на основі попередньої контрольної роботи. Результати цих груп підставляються в наступну формулу, обраховується коефіцієнт кореляції:

$$r \approx \frac{k}{k-1}\left\{1-\frac{1-2nk\sum(V_n+V_T)-[\sum(V_N+V_T)]^2}{0.666k[\sum(V_N-V_T)]^2}\right\},$$

де k – кількість питань;

n - *27%* від загальної кількості учнів N.

$\sum(V_n+V_T)$ - сума помилок в «слабкій» та «сильній» групах.

Порівняння ступенів складності задач. Зазвичай порівнюють задачі двох варіантів за допомогою x^2 тесту. При цьому застосовується система чотирьох полів:

Облік результатів тестів успішності.

$x^2=\frac{(Б-В)^2}{Б+В}$, де Б – кількість учнів, які правильно вирішили 1-у задачу і неправильно 2-у; В – кількість учнів, які вирішили правильно 2-у задачу, неправильно 1-у.

Таблиця 6.3.

Порівняння ступеня складності задач

Перша задача	Друга задача		Всього
	Правильно відповіли	Неправильно відповіли	
Правильно відповіли	40(А)	10(Б)	50
Неправильно відповіли	20(В)	5(Д)	25
Всього	60	15	75

$\chi^2=\frac{(10-20)^2}{10+20}=\frac{(-10)^2}{30}=\frac{100}{30}\approx 3.333$

Число ступенів вільності I, тому можна вважати різницю між ступенями складності 1-ї і 2-ї задачі незначною і ці задачі можна застосовувати в різних варіантах контрольної роботи.

При тестах і контрольних роботах, що застосовуються у дослідницькій роботі доцільно за кожну правильну відповідь давати один бал. При тестах розташування по-порядку і злиттям однієї задачі можна знайти ряд правильних відповідей. І в такому випадку кожна правильна відповідь рахується за один бал. Діагностичну цінність тесту не збільшує прийом, за якого більш легші питання оцінюються 0,5 і більш складніші – 2-а балами.

Результати тесту успішності рекомендується занести до таблиці, позначивши їх літерами П (правильна відповідь), Н (неправильна відповідь) і О (питання без відповіді).

Номер з п. учня	Номер задачі					Всього			
	1	2	3	…	25	Правильна відповідь	Неправильна відповідь	Без відповідей	Загаль на кількість
1	П	Н	П	…	П	20	2	3	25
2	Н	О	П	…	П	4	10	11	25
3	О	П	П	…	П	10	13	2	25
4	П	П	П	…	П	10	10	5	25
…	…	…	…	…	…	…	...	…	…
Кількість учнів, що дали правильну відповідь	15	12	6	…	20	246	210	44	500
Кількість учнів, що дали неправильну відповідь	2	7	12	…	0	210			
Кількість учнів, що не дали відповідь зовсім	3	1	2	…	0	44			
Всього	20	20	20	…	20	500			

На основі цієї таблиці легко визначити ступінь складності кожної задачі «сильних» і «слабких» учнів, оцінити весь тест і діагностичну цінність окремих питань.

Якщо ми внесемо результати контрольних і експериментальних класів в окремій таблиці, то можна за допомогою знакового тесту чи інших критеріїв визначити достовірність відмінностей результатів усього тесту, і на основі окремих відповідей достовірно виявити, чи вирішили в експериментальному класі в порівнянні з контрольним класом яку-небудь задачу краще чи ні. Це показує чи впливав експериментальний фактор в загальній мірі на засвоєння окремих питань чи ні.

Щоб дати загальну оцінку результату одного і того ж самого тесту проведеного в експериментальному та контрольному класі, потрібно порівняти між собою кількість правильних відповідей в експериментальному і контрольному класах, це робиться за допомогою знакового тесту.

Якщо в експериментальному класі було більше правильних відповідей, то в таблиці порівняння ставиться «+», якщо менше правильних відповідей, то «-». Достовірність отриманих результатів буде перевірено за допомогою відповідної таблиці.

Опираючись на приклад застосування знакового тесту можна легко перевірити достовірність відмінності відповідей на окремі питання. З цією метою порівнюють правильні і неправильні відповіді на питання в експериментальному і контрольному класах. Якщо учень в експериментальному класі відповів правильно, а в контрольному класі неправильно, то в рядок «знак» буде занесений «+». Якщо в експериментальному класі відповідь була неправильною, а в контрольному правильно, тоді буде внесений «-». Достовірність результатів перевіряється за допомогою знакового тесту.

П’ятибальну шкалу оцінювання, при якому контрольна робота чи тест буде оцінюватись від 1-5, не слід застосовувати в дидактичній дослідницькій роботі.

Якщо при оцінці результатів експериментальної дослідницької роботи застосовують точно і об’єктивно виміряні квантитативні показники і результати відповідають нормальному розподілу, то для перевірки різної статистичної достовірності доведеться застосовувати *z-*, *t-*, *чи F-* критерії. Їх застосування дає можливість точно визначити життєвість результатів, але зазвичай це пов’язано з достатньо об’ємними обчисленнями.

При застосуванні тесту успішності, а особливо при вибіркових і правильно-неправильних тестів, легко перевірити достовірність різниці результатів за допомогою x^2 критеріях чи знакового тесту.

Легко застосовувати і знаковий тест для повірки достовірності різниці відповідей на окреме питання.

При застосуванні знакового тесту, при порівнянні *t*-, *і* *F*-критеріями втрачається в деякій мірі велика інформація отримана у результаті дослідження. У порівнянні зі знаковим тестом ці критерії чутливі, тому що при їх зменшенні можна зробити достовірні висновки при значно меншій вибірці і диференціації між явищами, що порівнюються, ці висновки можуть бути незначними.

Далі для оцінки дидактичного дослідження часто розподілені не так як слід і не завжди підлягають перевірці, тому нема необхідності для більш об'єктивних результатів за допомогою інших критеріїв.

Знаковий тест, як і інші критерії заснований на теорії імовірності і з наочної точки зору і науковості його застосування неможна сумніватися. Перевагою знакового тесту є його легкість застосування, внаслідок чого він при ненормальному розподілі повинен знайти визначене місце з урахуванням суті результатів дидактичних досліджень.

Список використаних джерел

1. Бабанский Ю.К. Проблемы повышения эффективности педагогических исследований. Москва, 1989. С. 436-546.
2. Бережнова Е.В. Фундаментальное и прикладное в педагогических исследованиях. Педагогика. 2001. № 4. С. 3-7.
3. Берка К. Измерения: понятия, теории, проблемы. Москва, 1987.
4. Бернштейн М.С. К методике составления и проверки тестов. Вопр. психологии. 1988. № 1.
5. Бешелев С.Д. Экспертные оценки. М., 1983.
6. Біркентале В.В., Борова А.О. Про застосування кореляційно-регресійного аналізу до визначення рівня безробіття в Україні // Стратегічний розвиток національної економіки, регіонів і підприємств: т. 4. Моделі і методи ефективного стратегічного планування соціально-економічного розвитку / Матеріали всеукраїнської науково-практичної Інтернет-конференції, 15-17 листопада 2012 р., Донецьк, 2012. С. 27-31.
7. Боровков А.А. Теория вероятностей. Москва, 1976.
8. Брызгалова С.И. Научно-педагогическое исследование: учеб. Калининград, 1996.
9. Гершунский Б.С. В поисках практико-ориентированных образовательных концепций). М., 1998. 608 с.
10. Валеев Г.Х. Постановка проблемы педагогического исследования. Педагогика. 2001. № 4. С. 19-23.
11. Василенко П.М. Элементы методики метематической обработки результатов эксперементальных исследований. Київ, 1959.
12. Введение в научное исследование по педагогике: Учеб. пособие для студентов пед. ин-тов / Под ред. В.И. Журавлева. М.: Просвещение, 1988. 239 с.
13. Воловик П.М. Теорія ймовірностей та математична статистика в педагогіці. Київ, 1969.
14. Гласс Д., Стенли Д. Статистические методы в педагогике и психологии. М., 1976.

15. Готтсданкер Р. Основы психологического эксперимента. Москва, 1982.
16. Грабарь Н.И., Краснянская К.А. Применение математической статистики в педагогических исследованиях. М., 1977.
17. Гудвин Дж. Исследование в психологии: методы и планирование. Питер, 2004.
18. Давыдов В.П. Основы методологии, методики и технологии педагогического исследования: Научно-методическое пособие. М., 1997.
19. Дружинин В.Н. Экспериментальная психология: Учеб. пособие. М, 1997.
20. Загвязинский В.И., Атаханов Р. Методология и методы психолого-педагогического исследования: Учебное пособие. М., 2001.
21. Зельдович Я.Б., Мішкис А.Д. Элементи прикладной математики. М., 1972.
22. Кловак Г.Т. Основи педагогічних досліджень. Чернігів, 2003.
23. Королев Ф.Ф. Системный подход и возможности его применения в педагогических исследованиях // Сов. педагогика. 1990. № 9.
24. Краевский В.В. Методология педагогического исследования: Пособие для педагога-исследователя. Самара, 1994. 165 с.
25. Краевский В.В. Соотношение педагогической науки и педагогической практики. М., 1977.
26. Майоров А.Н. Тесты школьных достижений: конструирование, проведение, использование. СПб., 1997.
27. Маслов С.И. Выполнение студентами научно-исследовательской работы по педагогике: методические рекомендации для студентов. Тула: Изд-во Тул. гос. пед. ун-та им. Л. Н. Толстого, 2001.
28. Методология педагогики: сб. ст. Ч. III / Ред.-сост. В.О. Кутьев. М., 1991.
29. Методы системного педагогического исследования / Под ред. Н.В. Кузьминой. Л., 1980.

30. Михеев В.И. Методика получения и обработки экспериментальных данных в психолого-педагогических исследованиях: Монография. М., 1986.
31. Мкртычян Г.А. Параметры педагогической экспериментальной деятельности. Педагогика. 2001. № 5. С. 45-50.
32. Новиков А.М. Научно-экспериментальная работа в образовательном учреждении: Деловые советы. М., 1996.
33. Ньюман Л. Анализ качественных даннях. *Социологические исследования*. 1998. № 12.
34. Педагогика: Учеб. пособие для студентов пед. вузов и пед. колледжей / Под ред. П.И. Пидкасистого. М., 1998. Гл. 2: Методология и методы педагогических исследований. С. 32-61.
35. Педагогика: Учеб. пособие для студентов пед. учеб. заведений / В.А. Сластенин, И.Ф. Исаев, А.И. Мищенко, Е.Н. Шиянов. М., 1998. Гл. 6. С. 93-110.
36. Пискунов А.И., Воробьев Г.В. Методы педагогических исследований. М., 1987.
37. Полонский В.М. Научно-педагогическая информация: Слов.-справ. М., 1995. 256 с.
38. Полонский, В. М. Исследование в педагогике. Рос. пед. энц.: В 2 т. Т. 1. М., 1993.
39. Проблемы методологии педагогики и методики исследований / Под ред. М.А.Данилова и Н.И. Болдырева. М., 1971.
40. Психологическая диагностика детей и подростков / Под ред. К.М. Гуревича, Е.М. Борисовой. М., 1995.
41. Розенберг Н.М. Проблемы измерений в дидактике. Киев, 1979.
42. Руденко В.М., Руденко Н.М. Математичні методи в психології. Київ, 2009.
43. Рузавин Г.И. Методология научного исследования: Учеб. пособие для вузов. М.,1999. 317 с.
44. Сидоренко Е.В. Методы математической обработки в психологии. СПб., 2002.
45. Скалкова Я. и коллектив. Методология и методы педагогического исследования: Пер. с чешск. М., 1989.

46. Скаткин М.Н. Методология и методика педагогических исследований (в помощь начинающему исследователю). М., 1986.
47. Солсо Р.Л., Джонсон Х.Х., Бил М.К. Экспериментальная психология: практический курс. СПб.: прайм-ЕВРОЗНАК, 2001.
48. Сорокин Н.А. Дипломные работы в педагогических вузах: Учеб. пособие для студентов пед. ин-тов. М., 1986.Т. 2: Приложение Б. Статистика и обработка данных. М., 1992. С. 277-317.
49. Татарова Г.Г. Методология анализа данных в социологии (введение): Учеб. пособие для вузов. М., 1998.
50. Теория и практика педагогического эксперимента / Под ред. А.И. Пискунова и др. М., 1979.
51. Тутуболін В.Н. Статистическая обработка рядов наблюдений. М., 1973.
52. Черепанов В.С. Экспертные методы в педагогике: Учеб. пособие. Пермь, 1988.
53. Шевандрин Н.И. Социальная психология в образовании: Учеб. пособие. Ч. 1: Концептуальные и прикладные основы социальной психологии. - М., 1995. - Раздел 6. Основные методы социальной психологии (п.п. 6.1 Основные методы исследования и диагностики, 6.2. Статистические методы обработки и интерпретации данных). С. 278-313.
54. Эко У. Как написать дипломную работу. Гуманитарные науки: Учебно-метод. пособие / Пер. с итал. М., 2003.
55. Эксперимент в школе: организация и управление / Под ред. М.М. Поташника. М., 1991.
56. Ядов В.А. Социологическое исследование (методология, программа, методы). М., 1972.
57. Сисоєва С.О., Кристопчук Т.Є. Методологія науково-педагогічних досліджень: Підручник. Рівне, 2013. 360 с.