Житомирський державний університет імені Івана Франка
СОЦІАЛЬНО-ПСИХОЛОГІЧНИЙ ФАКУЛЬТЕТ
Кафедра соціальної та практичної психології

# МАТЕМАТИЧНІ МЕТОДИ У ПСИХОЛОГІЇ

НАВЧАЛЬНИЙ ПОСІБНИК

Климчук В.О.

ЖИТОМИРСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІМЕНІ ІВАНА ФРАНКА
СОЦІАЛЬНО-ПСИХОЛОГІЧНИЙ ФАКУЛЬТЕТ
КАФЕДРА СОЦІАЛЬНОЇ ТА ПРАКТИЧНОЇ ПСИХОЛОГІЇ

КЛИМЧУК В.О.

# МАТЕМАТИЧНІ МЕТОДИ У ПСИХОЛОГІЇ

*НАВЧАЛЬНИЙ ПОСІБНИК*

Київ
«Освіта України»
2009

**УДК 159.9.072.5**
**ББК ??.??**
**К 49**

*Рекомендовано Міністерством освіти і науки України*
*як навчальний посібник для студентів вищих навчальних закладів*

**Гриф надано Міністерством освіти і науки України**
**Лист № 1.4/18-Г-1773 від 11.07.08.**

**Рецензенти:**

**Н.В. Хазратова,** доктор психологічних наук, професор, завідувачка кафедри психології Київського славістичного університету (м. Київ);

**М.М. Заброцький,** професор, завідувач кафедри педагогіки та психології Обласного інституту післядипломної педагогічної освіти (м. Житомир);

**О.Є. Івашко,** кандидат психологічних наук, старший викладач кафедри педагогіки та психології Петрозаводського державного університету (м. Петрозаводськ, Росія).

**Климчук В.О.**

**К49** **Математичні методи у психології.** Навчальний посібник для студентів психологічних спеціальностей. — К.: Освіта України. — 2009. — 288 с.

**ISBN 978–966–188–032–9**

Навчальний посібник розроблений для використання в умовах кредитно-модульної системи організації навчального процесу, містить практичні основи застосування методів математичної статистики у психології, рекомендації до обробки даних експериментальних досліджень. У посібнику представлено правила графічного представлення статистичної інформації, обчислення середнього арифметичного, моди, медіани, дисперсії. Описано основи кореляційного аналізу, факторного аналізу, кластерного аналізу. Представлено роботу з методами статистичного висновку. Книга містить практикум зі статистичної обробки даних на базі програми Statistica.

Для студентів вищих навчальних закладів, психологів-дослідників, педагогів та осіб, що самостійно вивчають психологію та педагогіку.

**УДК 159.9.072.5**
**ББК ????????**

**ISBN 978–966–188–032–9**

© Климчук В.О., 2009
© «Освіта України», 2009

**Навчальний посібник**

**Климчук В.О.**

# МАТЕМАТИЧНІ МЕТОДИ У ПСИХОЛОГІЇ

Головний редактор **Задворний С.В.**
Комп'ютерна верстка **Климчук В.О.**
Дизайн обкладинки **Тимофеїва Н.В**

Підписано до друку 22.10.2008 року. Формат 60х84/16
Друк офсетний. Папір офсетний.
Ум. друк. арк. 18. Вид. обл. арк. 8.52

Наклад 300 прим.

Свідоцтво про внесення до Державного реєстру видавців
ДК № 1957 від 27.09.2004 р.
Тел./факс: (044) 411–43–97, 228–81–29, 237–59–92
E-mail; osvita2005@ukr.net; www.rambook.ru

Видавництво «Освіта України» запрошує авторів до співпраці з випуску видань,
що стосуються питань управління, модернізації, інноваційних процесів, технологій,
методичних і методологічних аспектів освіти та навчального процесу
у вищих навчальних закладах.
Надаємо усі види видавничих та поліграфічних послуг

# ПЕРЕДМОВА

Дослідження та експерименти, психодіагностика, підготовка звітів на педради та виступів на конференціях, розробка систем атестації персоналу – з цим далеко не вичерпним переліком робіт стикаються майже всі практичні психологи освітніх закладів, організаційні психологи, психотерапевти, психологи-дослідники. У всіх цих випадках психолог має аргументовано і об'єктивно довести опонентам свою точку зору, переконати у її істинності. Аналізуючи власний досвід наукових дискусій, досвід спілкування з практичними психологами у процесі навчання і по його завершенні, дозволю собі зробити висновок: у більшості випадків психолог може довести власну правоту лише посилаючись на результати конкретних досліджень. А зробити це він може лише у випадку уміння адекватно використовувати для опису, обробки та аналізу даних **математичний апарат**. Дуже влучно пише А.Д. Наслєдов: "Так, практична діяльність психолога – це передусім мистецтво застосування практичних методів. Але здорового глузду недостатньо для професійної роботи. Професіонал вирізняється тим, що може обґрунтувати свою точку зору, скажімо, перевірити ефективність того чи іншого практичного методу або спроможність організаційного рішення. При цьому він буде опиратися на науково обґрунтовані аргументи, а не лише на власну суб'єктивну думку".

Декілька слів про структуру і зміст книги.

Частина прикладів, наведених у курсі лекцій, є реальними психологічними дослідженнями, частина – гіпотетичні розробки. Велика частка прикладів побудована на основі ідей, взятих із книг О.В. Сидоренко, А.Д. Наслєдова, Д. Мартіна. Сама ідея переважання в курсі лекцій гіпотетичних прикладів взята з роботи американського дослідника Д. Мартіна "Психологічні експерименти", який обґрунтовує такий підхід необхідністю звернення до самої суті питання, не змушуючи студентів осмислювати величезну кількість другорядних деталей, які містяться в реальних експериментах. Такий підхід, як зазначає дослідник, має і свій недолік – студент, що вчиться лише за цим підручником, може закінчити курс, так і не побачивши справжніх експериментів і справжніх проблем, які виникають при спробі використати в них методи математичної статистики. Однак, вийти з цієї ситуації можна такими шляхами: спонукати студентів виконувати завдання, пов'язані з аналізом наукових статей, доповідей, наукових звітів, монографій; давати завдання проводити власні психологічні дослідження та

експерименти за заданими наперед темою та структурою; інтегрувати курс “Математичні методи у психології” з курсом “Експериментальна психологія”.

Допомогти реалізувати ці шляхи може практикум з методів математичної статистики у психології, який повинен використовуватися в комплексі з цим курсом лекцій. Крім того, всі гіпотетичні приклади, наведені у курсі лекцій, можуть слугувати гіпотезами наукових досліджень і я буду вдячний за їх перевірку (посилання на автора гіпотез у наукових роботах є обов’язковим ☺). Завдання інтеграції цього курсу з курсом “Експериментальна психологія”, я сподіваюсь, кожен студент зможе виконати самостійно, якщо має навички переносу знань з одного предмету на інший.

Ідея створення алгоритмів вибору статистичних критеріїв запозичена у О.В. Сидоренко, однак самі алгоритми дещо спрощені.

За браком місця у курсі лекцій не всі критерії описано детально – ґрунтовнішу інформацію про них можна знайти у рекомендованій літературі.

Нумерація формул, таблиць та рисунків йде в межах теми лекції. І ще одна особливість курсу лекцій – представлення більшості таблиць критичних значень не у додатках, а у тексті лекції. Це обумовлено зручністю викладу матеріалу – не потрібно переривати плин думки на пошук тих чи інших показників в кінці книги.

На завершення хочу висловити подяку людям, без яких цей курс не було б написано. Дякую *Музиці Олександру Леонідовичу* – за створення сприятливих умов роботи та ряд цінних ідей, *Музиці Олені Оксентівні* – за внесені корективи у зміст книги, *Оснадчуку Юрію Олеговичу* – за перевірку математичного апарату книги, *Горбуновій Вікторії Валеріївні* – за психологічну підтримку та натхнення.

## МЕТОДИЧНІ ПРОБЛЕМИ ВИКЛАДАННЯ “МАТЕМАТИЧНИХ МЕТОДІВ У ПСИХОЛОГІЇ”

Дуже часто при вивченні будь-якого курсу студенти передусім ставлять викладачу питання: “А навіщо нам вивчати цю дисципліну? Чи стане вона в нагоді по завершенню вузу?” І, на жаль, дуже часто після успішної здачі заліків та екзаменів студенти на це гносеологічне питання дають самі собі онтологічну відповідь: “Я марно витратив частинку свого життя”. Власне такі студентські питання і відповіді спонукали до вивчення проблем, які виникають при вивченні курсу “Математичні методи у психології”.

**Проблеми, з якими стикаються викладачі та студенти.**

До проблематики викладання математичних методів у психології зверталися ряд дослідників: О.Ю. Артем’єва, Д. Кемпбелл, Д. Мартін, А.Д. Наслєдов, С.С. Паповян. О.В. Сидоренко.

Так, О.Ю. Артем’єва та Є.М. Мартинов [1] зазначають, що рецептурна статистика вже давно відійшла на другий план, а задачею викладача є розвинути вміння відображати на психологічну реальність результати, отримані всередині абстрактних моделей (навчити принципам побудови ймовірнісно-статистичної мови). Проблемою ж дослідними вважають механічне копіювання викладачами стандартних математичних курсів без врахування специфіки формалізації саме у психології.

Аналізуючи роботи Д. Кемпбелла [5] з експериментальної психології, можна виокремити кілька проблем, які виникають у царині математико-статистичного аналізу експериментів: 1) студенти впевнені, що саме лише використання математичних методів забезпечує валідність результатів дослідження; 2) виникають проблеми при зіставленні даних, отриманих з допомогою кількісних методів, із нормативами здорового глузду; 3) часто спостерігається невміння поєднувати інтерпретації на основі кількісних та якісних даних.

В роботі з експериментальної психології Д. Мартін [6] теж описує ряд проблем, що виникають при використанні статистичних критеріїв. Перша проблема – студенти роблять помилку, коли при використанні статистичних критеріїв отримують низьку значимість відмінностей між двома незалежними вибірками, і на основі цього роблять висновок про достовірну тотожність цих вибірок. Друга проблема – абсолютизація рівнів 0,05 і 0,01. Д. Мартін пропонує ставитися до цих рівнів більш гнучко, враховуючи не лише математичні ймовірності, але й практичні наслідки своєї правоти чи помилки.

Третя проблема полягає у тому, що студенти змішують поняття “статистична значимість” і “практична значимість”, вважаючи, що статистично значимий результат є важливим, а статистично не значимий – неважливим, а іноді навіть помилковим і таким, що не вимагає оприлюднення. Четверта проблема – страх перед необхідність проводити обчислення, або “калькулятофобія”.

Аналізуючи роботу А.Д. Наслєдова [9] можна виділити такі проблеми, з якими стикаються студенти при засвоєнні математичних методів психологічного дослідження: 1) сумніви студентів у необхідності вивчення математичних методів і їх використання у психології; джерелом сумнівів є нерозрізнення психології як наукового способу пізнання дійсності та психології як мистецтва використання практичних методів (консультування, психотерапія тощо); 2) страх перед використанням математики; 3) невміння операціоналізовувати, і як наслідок – проблеми із підбором адекватних методів вимірювання психологічних явищ, шкал.

С.С. Паповян [10], вивчаючи проблему використання математичних методів у соціальній психології, доходить висновку, що у більшості випадків математико-статистичний аналіз результатів емпіричних досліджень обмежується виявленням простих кореляційних залежностей або проводиться з допомогою неадекватних статистичних методів

О.В. Сидоренко [12] описує проблему “чорного ящика”, яка виникла у зв’язку з широким розповсюдженням комп’ютерних статистичних програм. Студенти втрачають розуміння суті процесу обчислень; їх діяльність зводиться до завантаження у комп’ютер набору даних і отримання потому набору статистичних показників. Те ж, що відбувається у самому процесі обчислень, для них лишається таємницею, “чорним ящиком”. Більшою мірою це стосується складних математико-статистичних процедур, таких як факторний аналіз, багатомірне шкалювання, моделювання структурними рівняннями тощо.

Майже з усіма названими проблемами нам вдалося стикнутися у процесі викладання курсу “Математичні методи у психології”. Водночас, можна виділити ще кілька важливих моментів.

**Перша проблема** – це відсутність навичок переносу знань, отриманих у межах цього навчального курсу на інші дисципліни, зокрема, на експериментальну психологію, психосемантику, диференційну психологію, психологічну практику. Навіть студенти, які під час навчання демонстрували високий рівень знань, виявляються нездатними до використання того ж

факторного аналізу в іншому контексті, коли його проведення вже не є самостійним навчальним завданням, а має бути засобом для вирішення завдань вищого рівня.

**Друга проблема** – жорстка прив'язка деяких студентів до певних методів, які вони намагаються застосувати до досліджень найрізноманітнішої структури. Якщо ж задача дослідження виходить за межі застосування обраного методу – вона лишається нерозв'язаною або спрощується і підганяється під цей метод. Так, досить поширеними є випадки прив'язки до кореляційного аналізу, і студенти ніби забувають, що крім нього є ще ряд методів, які можуть дати багато цінної інформації – коваріаційний аналіз, канонічний аналіз, регресійний аналіз. Ця проблема має й інший бік – занадто велике захоплення статистичними показниками, які часто дублюють один одного. Були випадки використання одночасно критерію знаків, t-критерію та критерію кутового перетворення Фішера; поєднання факторного та кластерного аналізу із наступним багатомірним шкалюванням.

**Третя проблема** – неадекватність використання статистичних процедур. Так, студенти використовують кореляційний аналіз для порівняння середніх величин; кластерний аналіз для вирішення задачі зменшення розмірності набору особистісних конструктів.

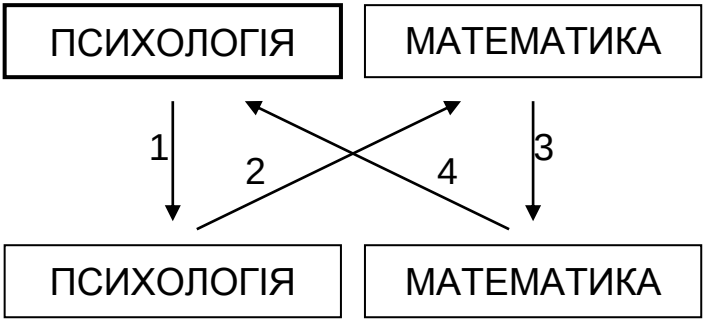
**Четверта проблема** – страх складних статистичних показників. Не всі студенти готові до обчислення вручну навіть t-критерія Стьюдента. З іншого боку іноді зустрічається нелюбов до простих методів. Тоді студенти намагаються знайти найскладніший метод для аналізу найпростішого експерименту.

**П'ята проблема** – неправильна інтерпретація статистичних показників, отриманих у результаті правильних обчислень. Найчастіше проблеми з інтерпретацією виникають при пошуку рівня статистичної значимості. Доволі складно студентам зрозуміти, як величина коефіцієнта кореляції може бути незначною, але значимою; або значною, але не значимою.

**Шоста проблема** – складнощі із розумінням математичних відмінностей між вимірювальними шкалами, і як наслідок – плутанина із межами застосування методів. Так, спостерігаються спроби використати середнє арифметичне до даних, отриманих в межах номінативної шкали.

Узагальнюючи результати аналізу відомих фахівців та спираючись на власний досвід викладання можна створити цілісну класифікацію проблем, з якими стикаються викладачі та студенти. В основі класифікації покладено виділення Г.В. Суходольським чотирьох видів інтерпретацій, якими повинен

володіти психолог-дослідник: психолого-психологічні, психолого-математичні, математико-математичні та математико-психологічні [13].




Психолого-психологічна інтерпретація (1) – операціоналізація понять, встановлення співвідношення між предметом дослідження та методами збору даних або як влучно пише А.Д. Наслєдов, між тим, *що* вивчається і *як* вивчається. Психолого-математична інтерпретація (2) – математична ідентифікація дослідницької ситуації, вибір методів аналізу даних. Математико-математична інтерпретація (3) – обчислення, перехід від набору даних до конкретних чисел-результатів обчислень. Математико-психологічна інтерпретація (4) – зворотній перехід від числової, математичної, реальності до реальності психологічної; формулювання висновків на основі обчислених показників.

***Класифікація проблем, які вникають у студентів при вивченні математичних методів у психології***

| Тип інтерпретації | Проблеми |
|---|---|
| Психолого-психологічна | А.Д. Наслєдов (№1, №3) |
| Психолого-математична | О.Ю. Артем'єва, Є.М. Мартинов, автор (№1, №2, №3, №6) |
| Математико-математична | Д. Мартін, А.Д. Наслєдов (№2), С.С. Паповян, О.В. Сидоренко, автор (№4) |
| Математико-психологічна | О.Ю. Артем'єва та Є.М. Мартинов, Д. Кемпбелл, автор (№5) |

Розроблена класифікація проблем дає змогу системно підійти до їх вирішення, адже вирішуючи кожну з проблем окремо викладач змушений би для кожного випадку шукати окремий методичний прийом, розробляти шляхи його впровадження тощо. Натомість нам вдалося знизити "*розмірність*" проблематики, перейшовши від масиву з 15 проблем до 4-х мірного простору.

**Окремі методи вирішення інтерпретаційних проблем.**

Для вирішення окремих проблем можна використати кілька методичних прийомів, в основу яких покладено *принципи синхронізації*. Робота з *психолого-психологічним* вектором проблем можлива в разі поєднання і *синхронізації* зусиль викладачів математичних методів та експериментальної психології. Оптимальний варіант – коли ці два курси читаються паралельно. В цьому випадку студенти одночасно вчаться операціоналізовувати поняття, розбираються в експериментальних планах. Координація зусиль двох викладачів полягає, по-перше, у спільному погодинному плануванні курсів: досить важливо, щоб студенти спочатку дізналися про вимірювальні шкали, а потім відразу навчилися операціоналізовувати поняття; дізналися про експериментальні плани, і лише потім перейшли до вивчення методів статистичного висновку, розібралися з кореляційною стратегією, а потім почали вивчати кореляційний аналіз. По-друге, досить важливою є проста обізнаність викладачів у навчальних планах один одного, щоб була можливість замість занурення у чужу проблематику, просто відіслати студенів до безпосереднього фахівця. Як показує досвід соціально-психологічного факультету, завдяки такій координації полегшується роботи із векторами *психолого-математичним* та *математико-психологічним*.

Коли говорити про *математико-математичний* вектор, то оптимізувати його можна, поєднавши і синхронізувавши зусилля викладачів теорії ймовірності і математичної статистики та математичних методів у психології. Як правило, першу дисципліну читають фахівці-математики, а другу – психологи, і математична дисципліна передує психологічній. В результаті, через велику математичне перевантаження та ігнорування викладачами специфіки студентів-гуманітаріїв, по завершенні курсу теорій ймовірності більшість студентів по-перше, виробляють стійкий імунітет проти математики, по-друге, не засвоюють на практичному рівні базових статистичних понять (середнє, дисперсія, статистична значимість тощо), по-третє, не набувають навичок елементарних ручних обчислень та роботи зі статистичними програмами. В результаті викладач-психолог повинен давати

потому масу фактично дублюючої, “непсихологічної” інформації, втрачаючи при цьому час на вирішення математичних задач.

Досвід викладання цих дисциплін на соціально-психологічному факультеті показу, що у разі тісної взаємодії викладачів-математиків та психологів (обговорення стратегії курсів, спільної розробки навчальних і робочих програм, підбір прикладів для аналізу) відбувається значна оптимізація навчального процесу, про що свідчать навіть опитування студентської думки.

В разі націленості викладача на перспективу, достатньо хороші результати може принести знайомство інших викладачів-психологів з навчальної і робочою програмами з математичних методів у психології (викладачів загальної психології, психосемантики, диференційної психології, інженерної психології, керівників курсових і дипломних робіт). В цьому разі студенти не випадають з математико-психологічного континууму, і не лише не втрачають набутих навичок, а й здобувають нові – навички переносу знань.

Роботу майже з усіма векторами проблем оптимізує активізація самостійної роботи студентів. Як робочий ми обрали принцип “реальної дії”, виходячи з того, що повноцінно засвоїти будь-які можна лише пропустивши їх через власний практичний досвід, зробивши їх “особистісними”. Для дотримання цього принципу було вирішено в структуру лабораторних занять ввести блок самостійної дослідницької роботи. До кожної лабораторної роботи, яка присвячена використанню певного статистичного методу, студенти на протязі тижня проводять власні психологічні дослідження, збираючи дані для обробки, а вже на самому занятті вони здійснюють ручну та комп’ютерну обробку цих даних. При цьому, оскільки кожен статистичний метод має специфічні умови використання, то й кожна лабораторна робота – це нове дослідження за розробленою викладачем структурою. Таким чином, студенти на власному досвіді, причому багато разів, проходять усі етапи психологічного дослідження, а значить, і всі види психологічних інтерпретацій – від психолого-психологічних до математико-психологічних. Наприклад, завдяки вимозі написання висновків до проведених обчислень забезпечується вирішення проблем з *вектором математико-психологічним*. Водночас, по завершенню курсу, студенти отримують великий досвід самостійного проведення психологічних досліджень, а тому готові до їх самостійної розробки і планування, що важливо для їх майбутніх курсових та дипломних робіт.

Однак, при реалізації вищеописаних методів, викладач, який працює в межах традиційної системи навчання, стикнеться з рядом перепон, як то: неможливість узгодити розклад занять з іншим викладачем, збої у розкладі через велике завантаження студентів на початку і в середині семестру і мале – в кінці, неможливість дати студентам достатньо часу для проведення дослідження, слабкі можливості контролю та корекції засвоєння усіма студентами окремих розділів і тематичних блоків. Ще однією проблемою, як зазначає В.В. Горбунова [2], є рівномірний розподіл занять практичного курсу, тобто кожен тиждень читається, наприклад, три заняття: одне лекційне, одне семінарське, одне лабораторне. В той же час є теми, які вимагають одного семінарського і жодної лабораторної, або навпаки, кілька лабораторних робіт і жодного семінару. Фактично, в умовах традиційної системи відбувається *десинхронізація* лабораторного, семінарського і лекційного циклів.

Ці, ніби то й незначні, перешкоди здатні повністю дезорганізувати навчальний процес і зруйнувати використання запропонованих методів. Вийти з критичного положення вдається, якщо навчання здійснюється в межах кредитно-модульної системи організації навчального процесу.

**Викладання "Математичних методів у психології" в умовах кредитно-модульної системи[1].**

Завдяки специфіці кредитно-модульній системі організації навчального процесу [2; 3; 7; 8; 11] вдається досить легко запровадити запропоновані вище методи вирішення інтерпретаційних проблем. Так, синхронізація зусиль викладачів математичних методів у психології, теорії ймовірності та експериментальної психології значно спрощується завдяки модульній організації змісту матеріалу. Кожен модуль виступає окремими самостійним блоком знань, а тому корекції змісту одного модуля робити достатньо легко і відбуваються вони "безболісно" для інших модулів.

Завдяки чіткому розподілу годин всередині модуля (один модуль дорівнює одному кредиту) з'являється можливість абсолютно чітко довести до студентів обсяг годин їх самостійної роботи. Деканат же має змогу обчислити загальний обсяг самостійної роботи студентів з усіх дисциплін, і відповідно вирішити, чи реальний він для виконання. Такий підхід дозволяє із годин, які відводяться для самостійної роботи, зробити реальний механізм

---

*[1] Тут і далі описано досвід впровадження кредитно-модульної системи на соціально-психологічному факультеті Житомирського державного університету імені Івана Франка (див. [7; 8])*

оптимізації навчального процесу. В контексті математичних методів у психології, – у студентів з'являється дійсно вільний час для проведення самостійних досліджень.

Достатньо важливим є те, що кредитно-модульна система дозволяє уникнути десинхронізації лекційного та лабораторно-семінарського циклів. Викладач сам вирішує, скільки необхідно лекцій, семінарів та лабораторних робіт в межах модуля, і крім того, може впливати на їх послідовність у розкладі. Таким чином, начальний процес прив'язується не до константного розкладу, а до варіативного змісту навчання.

Суттєвою перевагою модульного підходу є підвищення можливостей викладача в плані контролю процесу засвоєння знань. Кожен модуль завершується підсумковою модульної роботою, яка є письмовою, і результати якої здаються на збереження в деканат. Крім того, викладачу за перевірку модульних робіт нараховуються години, а значить, він не просто може контролювати знання студентів, але й повинен це обов'язково робити. Для студентів ці роботи також є достатньо корисними, оскільки змушують їх систематизувати свої знання не один раз на рік перед сесією, а кілька разів на семестр, перед кожною підсумковою модульною роботою. Так усувається традиційна проблема руху вперед по курсу і випускання по ходу окремих тем, а то й розділів. В ідеалі, в умовах кредитно-модульної системи студент просто не зможе вивчати курс далі до тих пір, поки не засвоїть попередні модулі, хоча, звісно, це поки що перспектива. Водночас, як зазначає Т.М. Майстренко, кредитно-модульна система не передбачає обов'язкового складання заліків чи екзаменів, що дозволяє студентам вивільнити ще більше часу для самостійної роботи [7].

Ще два резерви кредитно-модульної системи описує О.Л. Музика [8]. Перший резерв – завдяки можливості вибирати у студентів активізується їх внутрішня мотивація. Адже в межах цієї системи вони можуть обирати темп оволодіння навчальним матеріалом; здавати заліки та екзамени чи заробити автоматичні оцінки; на екзамені вони можуть вибрати одне з двох теоретичних питань, а також можуть погодитися на поточну оцінку чи спробувати підвищити її. Другий резерв – зростання самооцінки студентів у результаті виконання професійно-орієнтованих завдань, які розробляються з орієнтацією на майбутню професійну діяльність.

## Література


1. *Артемьева Е.Ю., Мартынов Е.М. Вероятностные методы в психологии. – М., 1975. – 206 с.*
2. *Горбунова В.В. До проблеми викладання курсу "Експериментальна психологія" // Соціальна психологія. – № 5 (13). – 2005. – С.169-176.*
3. *Горбунова В.В. Экспериментальная психология в схемах и таблицах: Учебное пособие. – Ростов-на-Дону: Феникс, 2005. – 184 с.*
4. *Корнилова Т.В. Экспериментальная психология: Теория и методы. – М.: Аспект Пресс, 2002. – 381 с.*
5. *Кэмпбелл Д. Модели экспериментов в социальной психологии и прикладных исследованиях. – СПб.: Социально-психологический центр, 1996. – 392 с.*
6. *Мартин Д. Психологические эксперименты – СПб.: прайм-ЕВРОЗНАК, 2002. – 480 с.*
7. *Майстренко Т.М. Кредитно-модульна система організації навчання фахівців психологічних спеціальностей / Навчально-методичне забезпечення кредитно-модульної системи організації навчального процесу в галузевих університетах: Матеріали VII Всеукраїнської науково-методичної конференції 9-10 листопада 2005 року. – Рівне: НУВГП, 2005. – С. 207-213.*
8. *Музика О.Л. Особливості підготовки психологів в умовах кредитно-модульної системи навчання / Тези XXXI наук.-практ. міжвуз. конф., присвяченої Дню університету 14-16 березня 2006 р. – Житомир, 2006. – С.280.*
9. *Наследов А.Д. Математические методы психологического исследования. Анализ и интерпретация данных. – СПб.: Речь, 2004. – 392 с.*
10. *Паповян С.С. Математические методы в социальной психологии. – М: Наука, 1983. – 344 с.*
11. *Положення про кредитно-модульну систему організації навчального процесу у Житомирському державному університеті імені Івана Франка. – Житомир, 2004. – 11 с.*
12. *Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: Речь, 2000. – 350 с.*
13. *Суходольский Г.В. Основы математической статистики для психологов. – Л.: ЛГУ, 1972. – 428 с.*

# Розділ І. ОСНОВНІ ПОНЯТТЯ МАТЕМАТИЧНОЇ СТАТИСТИКИ

## § 1.1. ОСНОВНІ ПОНЯТТЯ МАТЕМАТИЧНОЇ СТАТИСТИКИ

1. Змінні та їх вимірювання. Вимірювальні шкали.
2. Поняття про залежність між змінними. Залежні та незалежні змінні.
3. Основні символи змінних та операцій. Поняття масиву даних.

Література:


1. Артемьева Е.Ю., Мартынов Е.М. Вероятностные методы в психологии. – М., 1975.
2. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976.
3. Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО “Речь”, 2000.
4. Теорія статистики: Навчальний посібник /Вашків П.Г., Пастер П.І., Сторожук В.П., Ткач Є.І. – К.: Либідь, 2001.
5. Турчин В.М. Математична статистика в прикладах і задачах: Навч. посібник. – К.: НМК ВО, 1993.
6. Циба В.Т. Основи теорії кваліметрії: навчальний посібник. – К.: ІЗМН, 1997.
7. http://www.statsoft.ru/home/textbook/default.htm


### *1. ЗМІННІ ТА ЇХ ВИМІРЮВАННЯ*

*Нема жодної наукової галузі і жодної царини життєдіяльності,*
*де б не виникала потреба вимірювання властивостей будь-якої природи.*

*В.Т. Циба, “Основи теорії кваліметрії”, 1997*

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

**Що ж таке змінні?** Фактично, змінні – це все, що можна вимірювати, контролювати або змінювати у дослідженнях. Змінною може бути рівень особистісної тривожності, інтелект, рівень агресії, мотивація досягнення, тип темпераменту тощо. Що буде змінною у Вашому дослідженні і скільки буде цих змінних, залежить від його теми, мети, гіпотези, методів, завдань та структури. Навіть у дослідженнях, які виконані на одну тему, змінні можуть

бути різними, оскільки відрізнятимуться гіпотези чи використані для дослідження методики.

*У дослідженнях мотивації досягнення у дітей за* ***змінну*** *може бути обрано* ***тип поведінки*** *дитини в експериментальній ситуації, або ж* ***результати діагностики*** *з допомогою дитячого варіанту ТАТ.*

Крім того, змінні поділяються за своєю значимістю для дослідження на основні та додаткові.

*При дослідженні відмінностей у рівні мотивації досягнення дітей з повних сімей та дітей з неповних сімей* ***основні змінні*** *будуть такі: тип сім'ї дитини (повна чи неповна), рівень мотивації досягнення успіху та рівень мотивації уникання невдач. Також у процесі дослідження фіксуватимуться* ***додаткові змінні****, які можуть вплинути на його результати: стать дитини, вік, успішність тощо.*

**Що ж таке "вимірювання"?** Вимірювання – це процес приписування чисел певним явищам (змінним) відповідно до певних правил. Таким чином, у процесі вимірювання ми маємо справу з трьома елементами: 1) числа; 2) явища (змінні); 3) правила приписування чисел явищам. Графічно зв'язки між цими елементами можна зобразити так (рис. 1.1.1):

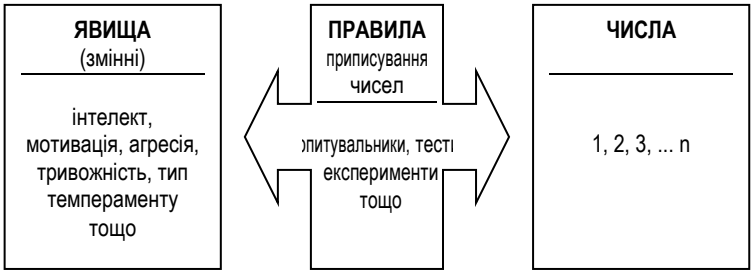



*Рис. 1.1.1. Зв'язки між елементами процесу вимірювання*

Таким чином, якщо мова йде про вимірювання рівня інтелекту, то змінною буде *інтелект*, числами – *показники від 0 до 200* (залежно від

використаного тесту), а правилом приписування певних чисел певному рівню інтелекту – *тест інтелекту* (тест Айзенка, тест Равена, тест Векслера тощо).

Процес вимірювання може бути різного рівня точності, що залежить як від типу досліджуваного явища, так і від інструменту його вимірювання. Якщо рівень інтелекту може бути виміряний з точністю $\pm 1$, то тип темпераменту з допомогою стандартизованих опитувальників може бути виміряний лише приблизно (в сторону переважання того чи іншого типу). Саме тому ще одним важливим для нас є поняття про **вимірювальні шкали**.

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

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

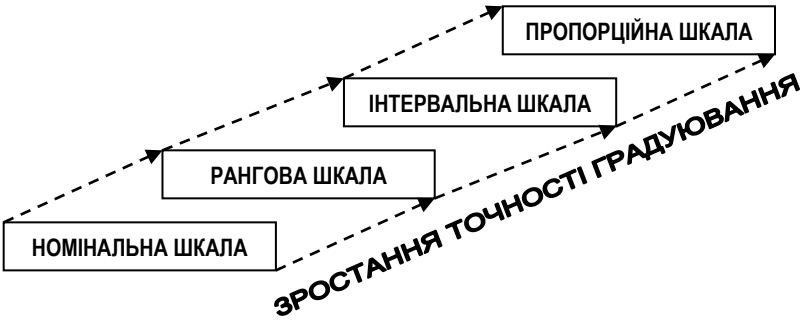



*Рис. 1.1.2. Класифікація вимірювальних шкал С. Стівенса (на основі точності градуювання шкали)*

Дослідники Дж. Гласс та Дж. Стенлі *(Гласс Дж., Стенли Дж.,1976, с.16)* наводять таблицю характеристик вимірювальних шкал С. Стівенса, яка має такий вигляд (табл. 1.1.1.):

*Таблиця 1.1.1*

*Характеристика вимірювальних шкал С. Стівенса (доповнена автором)*

| ШКАЛА | ХАРАКТЕРИСТИКА | МАТЕМАТИЧНІ ОПЕРАЦІЇ | ПРИКЛАДИ |
|---|---|---|---|
| Номінальна | У межах цієї шкали об'єкти класифікуються, а класи позначаються номерами. Таким чином, число тут слугує лише назвою певного класу, а тому нічого не говорить про властивості об'єкту, крім того, що він належить до певного класу. Частковим випадком шкали найменувань є дихотомічна шкала. | встановлення рівності | Колір очей, номери гравців спортивної команди, автомобільні чи інвентарні номери, типи темпераменту тощо. |
| Рангова | У межах цієї шкали об'єкти розташовуються в порядку спадання чи зростання у них певної якості. При цьому кожній градації якості приписується свій порядковий номер (ранг). Фактично, об'єкти лише впорядковуються. Особливість шкали – однакові різниці між сусідніми рангами не означають однакової різниці між ступенями прояву виміряної якості. | встановлення відношень "більше-менше" | Нагороди за заслуги, військові ранги, місця на олімпіаді, ранжування досліджуваних за проявами індивідуальних рис. |
| Інтервальна | У межах цієї шкали існує одиниця вимірювання, з допомогою якої можна не лише впорядкувати об'єкти, але й приписати їм числа так, щоб однакові різниці між числами виражали однакові відмінності у проявах вимірюваної якості. Особливість шкали – довільний нуль, який не говорить про відсутність якості у об'єкта. | встановлення рівності інтервалів (різниць) | Календарний час, шкала температур за Цельсієм. |

*Таблиця 1.1.1 (продовження)*

| ШКАЛА | ХАРАКТЕРИСТИКА | МАТЕМАТИЧНІ ОПЕРАЦІЇ | ПРИКЛАДИ |
|---|---|---|---|
| Пропорційна | У межах цієї шкали числа мають такі ж властивості, як і в шкалі інтервалів, але, крім того, відношення чисел виражають кількісні відношення ступенів прояву явища.<br>Особливість шкали – наявність абсолютного нуля, який означає відсутність явища | встановлення рівності відношень | Зріст, вага, температура за Кельвіном, рівень інтелекту, мотивація тощо. |

## *2. ПОНЯТТЯ ПРО ЗАЛЕЖНІСТЬ МІЖ ЗМІННИМИ. ЗАЛЕЖНІ ТА НЕЗАЛЕЖНІ ЗМІННІ*

Ми з'ясували, що "змінні" та "вимірювання" є одними із основних понять математичної статистики. Однак, оскільки ми говоримо про використання методів математичної статистики у психології, то варто сказати про деякі з **основних задач** психології.

*Першою задачею* є **пошук зв'язку** між явищами (змінними). Дійсно, чи пов'язані між собою вік та інтелект? Чи є зв'язок між мотивацією досягнення дітей та їх агресивною поведінкою в іграх? Чи зв'язані активність підлітків та їх статус у групі? Ці та багато інших питань пов'язані із задачею пошуку зв'язку.

*Другою базовою задачею* є виявлення впливу. Що на що впливає – мотивація досягнення на рівень агресії чи агресія на мотивацію досягнення? Чи призводить підвищена активність підлітків до зростання їх статусу у групі, чи навпаки, підвищення статусу впливає на зростання рівня активності? Це і є типові задачі, пов'язані з виявлення напрямку та форми впливу явищ одне на одне.

Розглянемо першу задачу (**пошук зв'язку**). Вважається, що дві чи більше змінних пов'язані між собою, коли їх значення розподілені узгоджено – коли систематично зміна однієї змінної в певному напрямку супроводжується зміною іншої змінної в іншому напрямку. Так, коли зростання рівня втоми супроводжується зростанням кількості помилок, говорять про існування

*прямого зв'язку* між цими явищами. З іншого боку, між явищами буде *обернений зв'язок*, коли, наприклад, збільшення часу роботи за комп'ютером супроводжується зменшенням кількості правильних операцій.

На рис. 1.1.3. наведено приклад графічного зображення зв'язку між явищами.

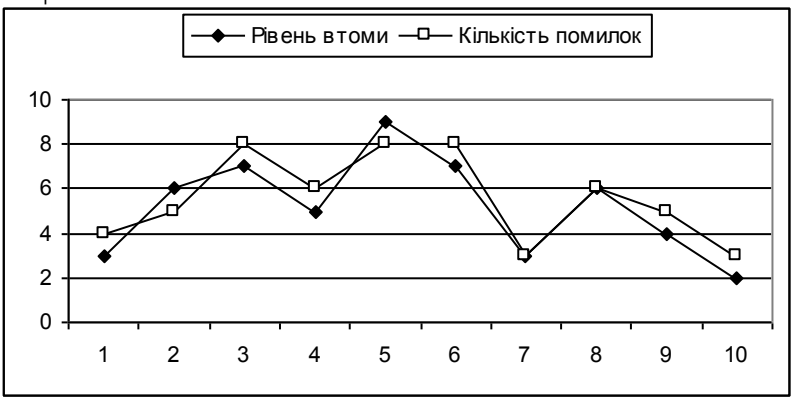



*Рис. 1.1.3. Зображення прямого зв'язку між рівнем втоми та кількістю помилок*

Як видно з рис. 1.1.3, показники втоми та кількості помилок розміщені узгодженим чином – високому рівню втоми відповідає велика кількість помилок, а низькому рівню втоми – мала кількість помилок. Такий зв'язок ми називали **прямим**.

Існують методи пошуку числового значення зв'язку (кореляційний аналіз) та інші способи графічного пошуку зв'язку (діаграми розсіювання), але про них ми поговоримо детальніше в розділі II.

Друга задача (**аналіз впливу**) випливає із проблеми, які ставить перша задача – якщо між явищами існує зв'язок, то чи означає це, що одне з них впливає на інше? А якщо так, то що на що впливає? Виявляється, методи пошуку зв'язку на ці питання не дають відповіді. Дійсно, справедливим може бути як те, що втома впливає на зростання кількості помилок, так і те, що збільшення кількості помилок призводить до нервового виснаження, і відповідно, до зростання кількості помилок. Водночас, обидва досліджувані явища можуть бути обумовлені сторонніми, не врахованими у дослідженні

чинниками – емоційним чи фізичним станом, сімейними проблемами досліджуваних, впливом самого дослідника тощо.

Виявляється, що для аналізу впливу одного явища на інше можна використати лише один дійсно надійний та валідний підхід – проведення експерименту. Суть експерименту полягає у тому, що дослідник систематично змінює одне з явищ, яке знаходиться у нього під контролем, і спостерігає, як буде змінюватися інше явище. У цьому випадку необхідно розрізнення двох типів змінних: **незалежних** та **залежних**.

***Незалежна змінна*** – явище, яке знаходиться під контролем експериментатора, і яке він може змінювати відповідно до наперед визначеної експериментальної процедури.

***Залежна змінна*** – явище, яке змінюється під впливом незалежної змінної.

*Так, при необхідності перевірити вплив втоми на кількість помилок, експериментатор повинен змінювати рівень втоми досліджуваних і спостерігати, скільки помилок вони зроблять. При перевірці впливу кількості помилок на рівень втоми експериментатор повинен непомітно для досліджуваних втручатися у процес їх роботи, штучно варіюючи кількість помилок, і спостерігаючи за їх рівнем втоми.*

Слід пам'ятати, що коли ми говоримо про **залежні** змінні – маємо на увазі задачу пошуку зв'язку, а коли говоримо про **залежну та незалежну** змінні – маємо на увазі задачу аналізу впливу.

## *3. ОСНОВНІ СИМВОЛИ ЗМІННИХ ТА ОПЕРАЦІЙ. ПОНЯТТЯ МАСИВУ ДАНИХ.*

Ми весь час не зможемо оперувати словом "змінні", тим більше, що досить часто змінних у дослідженні буває дуже багато. Тому розглянемо деякі символічні позначення, які використовуватимемо у подальшій роботі.

**Символи змінних.**

Для позначення певної змінної будемо використовувати латинські літери X, Y, Z і т.д.

Для позначення значень певної змінної будемо використовувати позначки типу $X_1$, $X_2$, $X_3$, $X_4$, $X_5$ і т.д. або $X_1$, $X_2$, ... $X_n$. У загальному окреме значення змінної позначатимемо $X_i$.

**Символи операцій.**

Статистика має справу з великими масивами даних, з якими слід проводити багато різних операцій. Тому для скороченого запису створили ряд символів, серед яких найчастіше нам буде зустрічатися СУМА ($\Sigma$).

Для запису $X_1 + X_2 + X_3 + X_4 + X_5$ пишуть так:

$$\sum_{i=1}^{5} X_i$$

**Масив даних.**

Під масивом даних ми будемо розуміти набір даних, отриманий у певному дослідженні. Це можуть бути показники інтелекту та віку досліджуваних, типи темпераменту, соціометричні статуси тощо.

# § 1.2. СТАТИСТИЧНЕ СПОСТЕРЕЖЕННЯ. ПОНЯТТЯ ПРО РЕПРЕЗЕНТАТИВНІСТЬ. ГРАФІЧНЕ ПРЕДСТАВЛЕННЯ ДАНИХ

1. Статистичне спостереження: види, способи, принципи та помилки.
2. Генеральна сукупність та вибірка. Репрезентативність вибірки.
3. Статистична значимість.
4. Статистичні таблиці та графіки.

Література:


1. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976.
2. Довідник здобувача наукового ступеня: Зб. нормат. докум. та ін форм. матеріалів з питань атестації наук. кадрів вищої кваліфікації /Упоряд. Ю.І. Церков. – К.: Ред. "Бюлетеня Вищої атестаційної комісії України": Вид-во "Толока", 2003.
3. ДСТУ 3008-95. Документація. Звіти у сфері науки і техніки. Структура і правила оформлення. – К.: Державний комітет України по стандартизації, метрології та сертифікації, 1996.
4. Теорія статистики: Навчальний посібник /Вашків П.Г., Пастер П.І., Сторожук В.П., Ткач Є.І. – К.: Либідь, 2001.
5. Хазанов Ю.С. Статистика. – М.: Статистика, 1974.
6. http://www.statsoft.ru/home/textbook/default.htm


## *1. СТАТИСТИЧНЕ СПОСТЕРЕЖЕННЯ: ВИДИ, СПОСОБИ, ПРИНЦИПИ ТА ПОМИЛКИ*

**Статистичне спостереження** – це процес науково організованого планомірного збору даних.

Можна виділити види статистичного спостереження:

1. За часом реєстрації фактів:
   - поточне (систематичний запис фактів по мірі їх виникнення – запис народження дітей, реєстрація шлюбів і т.д.),
   - періодичне (реєстрація фактів через строго визначені проміжки часу),
   - одночасне (реєстрація фактів по мірі їх необхідності в певний

момент часу)

2. За кількістю досліджуваних:
   - суцільне (вивчаються всі досліджувані певної сукупності),
   - несуцільне:
     - вибіркове (дає характеристику всієї сукупності на основі обстеження її частини),
     - спостереження основного масиву (спостерігаються об'єкти, які займають найбільшу питому вагу в досліджуваній сукупності),
     - монографічне (детальне і глибоке вивчення окремих одиниць досліджуваної сукупності).
3. За способами статистичне спостереження ділять на:
   - безпосереднє (отримання відомостей шляхом особистого огляду, підрахунку і т.д.),
   - документальне (отримання інформації з певних документів),
   - опитування (отримання інформації зі слів опитуваного).

До принципів статистичного спостереження відносять:

1. **Формулювання мети дослідження** – слід визначити мету дослідження, інакше буде зібрано багато непотрібної інформації, і мало – потрібної.
2. **Визначення об'єкта дослідження** – слід визначити, яке коло явищ досліджується і в якому аспекті.
3. **Розробка програми дослідження** – в якій послідовності і які факти вивчатимуться.

Отримані з допомогою статичного спостереження матеріали можуть бути відповідним чином оброблені і на основі обробки можна зробити певні висновки про досліджуваний процес чи явище. Однак, на висновок можуть вплинути помилки і похибки, які виникають при дослідженні.

1. Помилки реєстрації (виникають внаслідок неправильного запису фактів):
   - випадкові (описки, недостатнє знання і т.д.),
   - систематичні:
     - навмисні (коли опитувані чи реєстратор навмисно повідомляють чи записують неправильні дані),

- ненавмисні (зумовлені систематичними випадковими причинами – несправність приладів, втома досліджуваних після робочого дня тощо).

2. Помилки репрезентативності (характерні лише для несуцільного спостереження – виникають внаслідок того, що склад відібраної для дослідження вибірки не відтворює усієї сукупності досліджуваних).

Виявити помилки допомагає статистичний контроль (арифметичний та логічний).

## *2. ГЕНЕРАЛЬНА СУКУПНІСТЬ ТА ВИБІРКА. РЕПРЕЗЕНТАТИВНІСТЬ ВИБІРКИ.*

Можливості психологічних досліджень обмежені одним суттєвим фактором – вибіркою. **Вибірка** – це сукупність досліджуваних людей. Вибірка є частиною **генеральної сукупності** – усієї множини людей певної категорії. Поняття "генеральна сукупність" та "вибірка" співвідносяться як поняття "ціле" та "частина", або як "множина" і "підмножина" (рис. 1.2.1.):

*Рис. 1.2.1. Співвідношення понять "генеральна сукупність" та "вибірка"*

Така ситуація призводить до серйозної проблеми – дослідження проводять для виявлення всезагальних закономірностей, але на обмеженій вибірці. Де ж тоді впевненість, що закономірність, отримана на одній вибірці, буде справедливою для генеральної сукупності?

Загальним поняттям, що описує цю проблему, є поняття

“**репрезентативність вибірки**”. Репрезентативність вибірки – це ступінь представленості у ній генеральної сукупності. Справа у тому, що генеральна сукупність неоднорідна – вона має свою структуру, оскільки включає людей різних класів, професій, вікових параметрів, статі тощо. Якщо вибірка репрезентативна – то виявлені на ній закономірності можна перенести на генеральну сукупність. Загальний вигляд описаної ситуації представлено на рис. 1.2.2.

*Рис. 1.2.2. Різні ступені репрезентативності вибірки*

Таким чином, видно, що існують різні ступені репрезентативності, і зрозуміло чому – жоден дослідник не в змозі сформувати вибірку, яка б абсолютно точно відтворювала структуру генеральної сукупності. Це призвело б до того, що вибірка стала б рівною по обсягу самій генеральній сукупності. Тому дослідник мусить обирати – великі обсяги вибірки і дуже високу репрезентативність, або малі за обсягом вибірки і не дуже високу репрезентативність.

Існують загальні **методи підвищення репрезентативності** вибірки:

1. Планомірний підбір досліджуваних із генеральної сукупності. При використанні цього методу вивчають структуру генеральної сукупності та обирають для дослідження представників усіх виділених категорій. Цей метод неекономний і доволі громіздкий. Крім того, яким би точним не був аналіз генеральної сукупності, є ймовірність зробити помилку і не врахувати якісь категорії.
2. Рандомізований підбір досліджуваних з генеральної сукупності. Цей метод ґрунтується на уявленні про ймовірнісну характеристику розподілу ознак. Суть його полягає у підборі досліджуваних з

допомогою генератора випадкових чисел (або таблиць). У цьому випадку економиться час, матеріальні ресурси і є висока ймовірність попадання у вибірку більшості категорій генеральної сукупності.

## *3. СТАТИСТИЧНА ЗНАЧИМІСТЬ*

Тісно пов'язаним із рівнем репрезентативності вибірки є поняття "**статистична значимість**". Статистична значимість є числовим вираженням можливості узагальнення результатів, отриманих у вибірці, на генеральну сукупність. Чим вищий рівень репрезентативності вибірки, тим вища статистична значимість отриманих результатів.

З іншого боку, під статистичною значимістю можна розуміти ймовірність помилки при спробі узагальнення результатів і поширення їх на генеральну сукупність. Тоді, чим меншою є величина статистичної значимості, тим нижчою є ймовірність помилки.

*Традиційно* рівні статистичної значимості позначають буквою **p** та виділяють такі значення: 0,05, 0,01 та 0,001. Рівень статистичної значимості 0,05 означає, що існує лише 5% імовірності помилки, а рівень 0,001 – означає 0,1% імовірності помилки. Зазвичай результат, значимий на рівні 0,05 вважають прийнятним, результат на рівні значимості 0,01 – статистично значимим, а результат на рівні 0,001 – високо значимим.

*Таблиця 1.2.1.*

*Рівні статистичної значимості*

| Рівень статистичної значимості | Ступінь значимості результату | Імовірність помилки при узагальненні результату |
|---|---|---|
| p=0,05 | Прийнятний результат | 5% |
| p=0,01 | Значимий результат | 1% |
| p=0,001 | Високо значимий результат | 0,1% |

Як же визначити рівень статистичної значимості отриманих результатів? Для цього створені спеціальні статистичні таблиці, якими ми будемо користуватися далі. Для кожного із статистичних показників існують свої таблиці статистичної значимості, в яких поставлено в залежність рівень статистичної значимості від величини самого показника та від величини досліджуваної вибірки.

**Статистичні таблиці** – це спосіб раціонального, наочного і систематичного викладу та аналізу цифрових характеристик досліджуваного процесу чи явища.

Складену, але не заповнену таблицю називають *макетом таблиці*.

Структура таблиці:

Номер таблиці

Назва таблиці

*Таблиця 1.2.2.*

*Середні показники внутрішньої мотивації у виділених поведінкових групах*

| Показники внутрішньої мотивації | Перша група | Друга група | Третя група |
|---|---|---|---|
| Задоволення | 13,6 | 15,5 | 18,2 |
| Зусилля | 8,7 | 7,9 | 7,6 |
| Компетентність | 13,9 | 13,7 | 13,4 |
| Сприймання вибору | 13,3 | 12,9 | 13,4 |

Верхні заголовки

Таблиці бувають числові та текстові. Однак, хоч текстові таблиці і не є статистичними, правила їх оформлення залишаються такими ж.

Сучасна комп'ютерна техніка дозволяє робити з таблицями все, що завгодно, однак, не слід забувати про те, що крім вас вони повинні бути ЗРОЗУМІЛИМИ для інших людей.

**Правила складання статистичних таблиць**

1. Таблиці мають бути наочними і виразними, їх не слід робити громіздкими та перевантажувати деталями. Іноді замість однієї загальної таблиці слід зробити декілька часткових.
2. Назва таблиці має розкривати її зміст у стислій і виразній формі. Назви рядків і стовпчиків пишуться коротко і без скорочень.
3. Якщо таблиця переноситься на наступний лист, її стовпчики слід або нумерувати, або переносити верхній заголовок на цей лист.
4. Якщо по певній ознаці змінна не має значень, це позначають "–", а якщо не має по ній відомостей, то "…"

Без сумніву, графічне представлення психологічних даних є дуже цінним доповненням до статистичного аналізу. Графіки чи діаграми притягують увагу читача, адже навряд чи він зупиниться на листах суцільного тексту чи на нагромадженні різних таблиць.

Часто графік слугує ефективним засобом вираження думки. Один маленький графік скаже більше, ніж десятки сторінок чи параграфів. Зазвичай кількісні дані дуже абстрактні, а графік – наочний. Наприклад, деякі елементи таблиці можна зобразити на діаграмі (рис. 1.2.3):

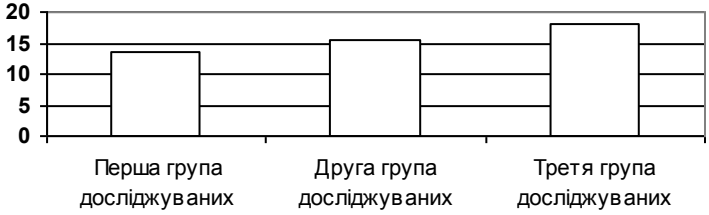



*Рис. 1.2.3. Рівень задоволення від роботи в різних групах досліджуваних*

Загалом виділяють три **цілі**, які переслідує графік:

1) краще розуміння даних;
2) глибший аналіз;
3) контроль точності.

**Загальні правила побудови графіків** *(за Гласс Дж., Стенли Дж.,1976)***:**

1. Графіки читають зліва направо.
2. Не можна зображати лінійні величини з допомогою об'ємів чи площ.
3. Бажано, щоб на графіку була зображена нульова відмітка.
4. Для графіків з відсотками бажано певним чином виділяти лінію 100% чи інші лінії, важливі для аналізу.
5. Рекомендується показувати не більше координатних ліній, ніж це необхідно для розуміння.
6. Криві лінії повинні різко відрізнятися від прямих.
7. Горизонтальна шкала читається зліва на право, а вертикальна – знизу вверх.
8. Цифри відповідних шкал розміщують зліва та знизу.
9. У графік можна включати цифрові дані.
10. Якщо цифрові дані не потрапили до графіка, їх бажано розмістити в таблиці.

Дослідники Дж. Гласс та Дж. Стенлі наводять приклади графіків, **виконаних без врахування вищеназваних правил**.

1. Внаслідок того, що **шкала починається не з нуля**, створюється враження, що відсоток виключених зріс в 4 рази, хоча насправді – лише на 5% (рис. 1.2.4).

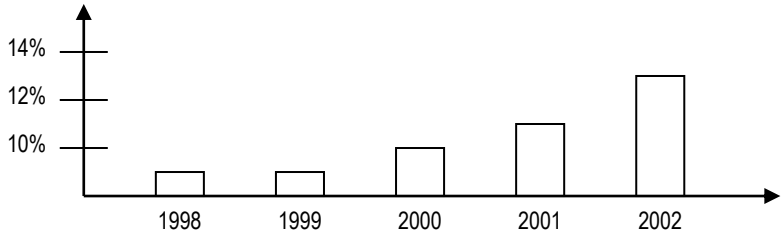



*Рис. 1.2.4. Відсоток виключених з вузу студентів після 1 курсу за 1998-2002 роки*

2. Суми, отримані від лотерей, **зображені як діаметри кіл**, а око сприймає **ПЛОЩІ**. Тому, хоч і різниця між районом А та В складає 4 тис., здається, що вона рівна принаймні 16 тис. грн. (рис. 1.2.5).

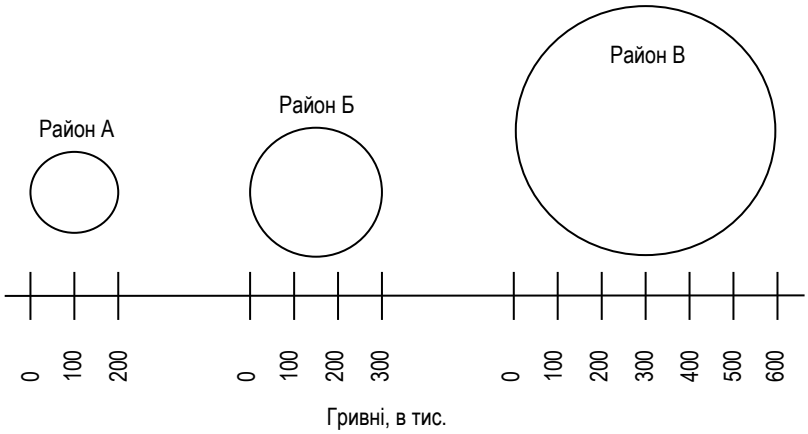



*Рис. 1.2.5. Прибуток від лотерей у районах А, Б та В*

# § 1.3. МІРИ ЦЕНТРАЛЬНОЇ ТЕНДЕНЦІЇ

1. Мода
2. Медіана та її обчислення
3. Середнє арифметичне: обчислення та властивості
4. Інтерпретація мір центральної тенденції
5. Вибір міри центральної тенденції

Література:


1. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976.
2. Теорія статистики: Навчальний посібник /Вашків П.Г., Пастер П.І., Сторожук В.П., Ткач Є.І. – К.: Либідь, 2001.
3. Хазанов Ю.С. Статистика. – М.: Статистика, 1974.
4. http://www.statsoft.ru/home/textbook/default.htm


## *1. МОДА*

Для опису даних, отриманих у результаті досліджень, часто використовують різноманітні статистичні показники, серед яких є міри центральної тенденції та міри мінливості.

Завдання мір центральної тенденції – виявити найтиповіший показник для даної вибірки.

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

Правила обрахування:

1. Якщо в даних всі значення зустрічаються однаково часто, кажуть, що в них немає моди:

$$(1, 2, 3, 4)$$

2. Якщо два сусідні значення мають однакову частоту, то модою називають їх середнє:

$$Mo(1, 2, 2, 3, 3, 4) = 2,5$$

3. Якщо два несусідні значення мають однакову частоту, то кажуть, що в даних є дві моди, а ряд даних називається бімодальним:

Mo(1, 1, 1, 2, 3, 4, 5, 5, 5)= 1 та 5

Крім того, в ряду даних можуть бути найбільша і менші моди, при цьому найбільша мода – єдине значення, яке задовольняє визначення моди.

## *2. МЕДІАНА*

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

Правила обчислення:

1. Якщо ряд містить непарну кількість значень, то медіана є центральним значенням Md(11,13,**25**,48,49)=25
2. Якщо ряд містить парну кількість значень, то медіана обраховується як середнє двох центральних значень: Md(11,**13,25**,48)=19
3. Якщо ряд достатньо великий, то щоб знайти місце медіани варто використати таку формулу:

$$N_{Md} = \frac{n+1}{2}$$

Де $N_{Md}$ – місце медіани в ряду даних,
$n$ – кількість елементів в ряду даних.

## *3. СЕРЕДНЄ АРИФМЕТИЧНЕ: ОБЧИСЛЕННЯ ТА ВЛАСТИВОСТІ*

Важливою мірою центральної тенденції є середнє арифметичне.

СЕРЕДНЄ АРИФМЕТИЧНЕ – $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Середнє арифметичне обчислюється досить просто навіть вручну. Однак, цей процес можна ще **спростити**, якщо окремі значення в ряду даних повторюються (табл. 1.3.1):

*Таблиця 1.3.1.*

*Спрощене обчислення середнього арифметичного*

| Вихідні дані | Дані ($X_i$) | Частоти ($f_i$) | $X_i f_i$ | Обчислення |
|---|---|---|---|---|
| 2 6 10 | 2 | 1 | 2 | $\overline{X} = \frac{\sum_i X_i f_i}{n} = \frac{166}{21} = 7{,}9$ |
| 3 6 10 | 3 | 2 | 6 | |
| 3 6 11 | 5 | 4 | 20 | |
| 5 8 11 | 6 | 3 | 18 | |
| 5 8 11 | 8 | 2 | 16 | |
| 5 9 15 | 9 | 2 | 18 | |
| 5 9 18 | 10 | 2 | 20 | |
| | 11 | 3 | 33 | |
| | 15 | 1 | 15 | |
| | 18 | 1 | 18 | |
| | $n = \sum_i f_i = 21$ | | $\sum_i X_i f_i = 166$ | |

Можлива ситуація, коли, знаючи середні арифметичні декількох різних груп, вам треба знайти загальне середнє арифметичне для всіх груп разом. Наприклад, є 3 класи з різною кількістю учнів і середніми показниками успішності (табл. 1.3.2.):

*Таблиця 1.3.2*

*Середні показники успішності груп А, Б та В*

| Група | $\overline{X}$ | *n* |
|---|---|---|
| А | 11,9 | 24 |
| Б | 8,2 | 30 |
| В | 10,8 | 28 |

Якби ми мали вихідні дані, за якими були обраховані середні, ми б додали 82 значення (24+30+28=82) і поділили їх на 82. Однак, цього ми зробити не можемо, а тому звернемося до формули середнього арифметичного:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

*n* нам відомо, треба лише знайти для кожної групи суму її показників. Виходячи з формули середнього арифметичного: $\sum_i X_i = n \cdot \overline{X}$.

Тому **середнє арифметичне для об'єднаних груп матиме вигляд**:

$$\overline{X} = \frac{n_а X_а + n_б X_б + n_в X_в}{n_а + n_б + n_в}$$

Середнє арифметичне має декілька дуже цікавих **властивостей**, які слід знати, щоб мати змогу їх використовувати. Вони ще стануть нам у нагоді в подальшій роботі.

1. *Сума відхилень всіх значень ряду від його середнього арифметичного рівна нулю.*

Відхилення – різниця між середнім арифметичним ($\overline{X}$) і конкретним значенням ряду ($X_i$).

Сума відхилень записується як $\sum_i \left(X_i - \overline{X}\right) = 0$

Перевіримо цю властивість на прикладі:

*Нехай маємо ряд (1, 2, 3, 4, 5): n=5;* $\overline{X}$ *=3*

*Таблиця 1.3.3.*

*Перевірка властивості №1*

| **Значення** | **–** | **Середнє** | **=** | **Відхилення** |
|---|---|---|---|---|
| 1 | – | 3 | = | 2 |
| 2 | – | 3 | | -1 |
| 3 | – | 3 | = | 0 |
| 4 | – | 3 | = | 1 |
| 5 | – | 3 | = | 2 |
| | | | СУМА | **0** |

Це можна довести для всіх випадків, якщо розписати формулу відхилення:

$$\sum_i \left(X_i - \overline{X}\right) = \sum_i X_i - \sum_i \overline{X} = \sum_i X_i - n\overline{X} = \sum_i X_i - n\frac{\sum_i X_i}{n} = \sum_i X_i - \sum_i X_i = 0$$

2. *Якщо до кожного значення ряду додати (або відняти) константу, то середнє арифметичне цього ряду зросте або зменшиться на цю ж константу.*

$$\frac{1}{n}\sum_i \left(X_i \pm c\right) = \overline{X} \pm c$$

Перевіримо (табл. 1.3.4.):

*Таблиця 1.3.4.*

*Перевірка властивості №2*

| **Значення** | **+** | **Константа** | **=** | **Нове значення** |
|---|---|---|---|---|
| 1 | + | | = | 2 |
| 2 | + | | = | 3 |
| 3 | + | 1 | = | 4 |
| 4 | + | | = | 5 |
| 5 | + | | = | 6 |
| **3** | СЕРЕДНЄ АРИФМЕТИЧНЕ | | | **4** |

Доведемо:

$$\frac{1}{n}\sum_i (X_i \pm c) = \frac{1}{n}\sum_i X_i \pm \frac{1}{n}\sum_i c = \overline{X} \pm \left(\frac{1}{n}\right) nc = \overline{X} \pm c$$

3. *Якщо кожен член ряду помножити (чи поділити) на константу, то середнє арифметичне цього ряду також слід помножити (поділити) на цю константу.*

$$\frac{1}{n}\sum_i (X_i \cdot c) = \overline{X} \cdot c$$

Доведіть це твердження самостійно і перевірте на прикладі.

4. *Сума квадратів відхилень значень від їх середнього арифметичного менше суми квадратів відхилень цих значень від будь-якого іншого числа.*

$$\sum_i \left(X_i - \overline{X}\right)^2 \prec \sum_i \left(X_i - b\right)^2$$

Перевірте це твердження самостійно на прикладі.

## *4. ІНТЕРПРЕТАЦІЯ МІР ЦЕНТРАЛЬНОЇ ТЕНДЕНЦІЇ*

Кожна з мір центральної тенденції має цікаву інтерпретацію в термінах похибок, які виникають тоді, коли одна ця статистична характеристика замінює всі значення в групі.

### МОДА

Якщо для заміни всіх чисел ряду обрати моду, то в рядах буде максимально велика кількість співпадань:

(1, 2, 2, **3, 3, 3, 3, 3**)

(3, 3, 3, **3, 3, 3, 3, 3**)

МЕДІАНА

Нехай ми маємо ряд (1, 3, 6, 7, 8): Md=6.

Знайдемо абсолютну різницю між Md (6), та всіма елементами ряду (табл. 1.3.5.):

*Таблиця 1.3.5.*

*Сума абсолютних різниць елементів ряду та медіани*

| **Ряд** | **-** | **Медіана** | **=** | **Абсолютна різниця** |
|---|---|---|---|---|
| 1 | - | | = | 5 |
| 3 | - | | = | 3 |
| 6 | - | 6 | = | 0 |
| 7 | - | | = | 1 |
| 8 | - | | = | 2 |
| | | | СУМА | **11** |

Знайдена нами сума абсолютних різниць завжди менше суми різниць відносно будь-якої іншої точки (перевірте це самостійно).

Якщо замість кожного значення ряду обрати медіану, то буде досягнута *мінімальна похибка* – якщо похибку визначити як суму абсолютних різниць кожного елементу ряду та міри центральної тенденції

СЕРЕДНЄ АРИФМЕТИЧНЕ

Якщо замість кожного значення ряду обрати середнє арифметичне, то буде досягнута *мінімальна похибка* – якщо похибку визначити як суму квадратів різниць кожного елементу ряду та міри центральної тенденції **(для доведення див. властивість середнього арифметичного №4)**

## *5. ВИБІР МІРИ ЦЕНТРАЛЬНОЇ ТЕНДЕНЦІЇ*

Кожна міра має певні характеристики, які роблять її цінною в певних умовах. Вибір тієї чи іншої міри іноді вимагає певних роздумів. Наведемо декілька корисних порад та характеристик мір центральної тенденції.

1. Моду та медіану обчислити найпростіше.

2. В малих групах мода нестабільна:

   Mo(1, 2, **3**, 3, 3, 4, 4, **5**)=3; Mo(1, 2, **2**, 3, 3, 4, 4, **4**)=4

3. На медіану не впливають величини крайніх значень ряду даних.
4. На величину середнього арифметичного впливають значення кожного елементу ряду (табл. 1.3.6.).

*Таблиця 1.3.6.*

*Порівняння мір центральної тенденції в рядах, що відрізняються одним значенням*

| | Мода | Медіана | Середнє |
|---|---|---|---|
| **1:** 1, 3, 3, 5, 6, 7, 8 | 3 | 5 | 4,7 |
| **2:** 1, 3, 3, 5, 6, 7, 16 | 3 | 5 | 5,9 |

5. Деякі множини даних можуть не мати реальної міри центральної тенденції:

Наводиться такий приклад *(Дж. Гласс, Дж. Стенлі, 1976)*:

*На лавці сидить 5 чоловіків. Два жебраки з майном 25 копійок. Третій – робітник, його збереження нараховували 5000 грн. Четвертий чоловік мав 100 000 грн. в різних фірмах. П'ятий – мультимільйонер з прибутком 5 000 000 грн.*

*Таким чином, ряд: (0,25, 0,25, 5000, 100 000, 5 000 000).*

| | |
|---|---|
| *Mo=0,25*<br>*Md=5000* | Отже, жодна з мір центральної тенденції для такого ряду не є адекватною |
| $\overline{X}$ *= 1021000* | Другий висновок – для групи з 5 значеннями навряд чи потрібна взагалі якась статистика |

6. Вибір міри центральної тенденції може бути обумовлений типом змінних, які аналізуються (*Мартин Д., 2002)*:

| Номінативна шкала (найменування) | Порядкова шкала | Інтервальна шкала або шкала відношень |
|---|---|---|
| **МОДА** | | |
| | **МЕДІАНА** | |
| | | **СЕРЕДНЄ** |

# § 1.4. МІРИ МІНЛИВОСТІ

1. Призначення мір мінливості.
2. Розмах та коефіцієнт осциляції.
3. Дисперсія: обчислення та властивості.
4. Стандартне відхилення та коефіцієнт варіації.

Література:


1. Артемьева Е.Ю., Мартынов Е.М. Вероятностные методы в психологии. – М., 1975.
2. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976.
3. Горбатов Д.С. Практикум по психологическому исследованию: Учеб. пособие. – Самара: Издательский дом "БАХРАХ-М", 2000.
4. Кимбл Г. Как правильно пользоваться статистикой: Пер с англ. – М., 1972.
5. Мартин Д. Психологические эксперименты: Секреты механизмов психики. – СПб.: прайм-ЕВРОЗНАК, 2002.
6. Молостов А.В. Элементы вариационной статистики. – К.: Урожай, 1965.
7. Свалов Н.Н. Вариационная статистика. Учебное пособие для вузов. – М.: Лесная промышленность, 1977.


## *1. ПРИЗНАЧЕННЯ МІР МІНЛИВОСТІ*

Міри центральної тенденції говорять про концентрацію групи значень навколо певного показника. Такі міри дають показник, який до певної міри представляє всю вибірку. Однак, у цьому випадку ігноруються відмінності, що існують між окремими значеннями. Для визначення цих відмінностей і використовують **міри мінливості**.

## *2. РОЗМАХ ТА КОЕФІЦІЄНТ ОСЦИЛЯЦІЇ*

РОЗМАХ – різниця максимального і мінімального значення в групі

Обчислення: $R = X_{max} - X_{min}$

$$R(1, 3, 4, 4, 5) = 4$$

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

**R(10 20 30 40 50)=40** та **R(10 10 10 10 50)=40**

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

КОЕФІЦІЄНТ ОСЦИЛЯЦІЇ відображає відносні коливання крайніх значень ряду відносно середнього показника.

$$K_o = \frac{R}{\overline{X}} \cdot 100$$

Знайдемо коефіцієнти осциляції для двох вище наведених рядів:

$K_{o1}$=(40/30)*100=133%; $K_{o2}$=(40/18)*100=222%

Таким чином, видно, що в другому ряду крайні значення коливаються відносно середнього більше, ніж в першому – **виникає питання про можливість порівнювати ці ряди між собою.**

## *3. ДИСПЕРСІЯ: ОБЧИСЛЕННЯ ТА ВЛАСТИВОСТІ*

Як ми відмітили, на розмах не впливають всі дані, що лежать між крайніми значеннями. Було б досить корисним знайти міру, яка б дозволяла при обрахуванні варіативності ряду враховувати всі його значення.

Ми знаємо, як порахувати відхилення кожного значення ряду від його середнього арифметичного. Сума цих відхилень була б непоганою мірою, однак вона завжди точно рівна нулю – ми це з'ясовували. Щоб обійти цю властивість сумують не різниці, а квадрати цих різниць.

Однак, сама сума квадратів як міра використовуватися не може, оскільки аналізовані вибірки можуть мати різний обсяг. Це обмеження знімається, коли сума ділиться на *n-1.* Отримана міра називається дисперсією.

ДИСПЕРСІЯ – $\sigma^2 = \dfrac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$

Обчислення дисперсії (табл. 1.4.1.):

*Таблиця 1.4.1.*

*Обчислення дисперсії*

| Дані | Різниця ($X_i - \overline{X}$) | Квадрат різниці $\left(X_i - \overline{X}\right)^2$ | Дисперсія ($\sigma^2$) |
|---|---|---|---|
| 1 | 1-2=-1 | 1 | $\sigma^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1} = \frac{12}{6-1} = 2{,}4$ |
| 3 | 3-2=1 | 1 | |
| 3 | 3-2=1 | 1 | |
| 0 | 0-2=-2 | 4 | |
| 4 | 4-2=2 | 4 | |
| 1 | 1-2=-1 | 1 | |
| $\overline{X}$ =2 | | $\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ =12 | |

Дисперсія, подібно до середнього арифметичного, має декілька властивостей (доведення – самостійно).

1. Додавання константи (С) до кожного значення ряду не змінює дисперсію.
2. Множення кожного значення ряду на константу (С) збільшує дисперсію в $C^2$ разів.

## *4. СТАНДАРТНЕ ВІДХИЛЕННЯ ТА КОЕФІЦІЄНТ ВАРІАЦІЇ*

Дуже тісно пов'язана з дисперсією наступна міра – стандартне відхилення (середнє квадратичне відхилення).

$$s_x = \sqrt{\sigma^2}$$

Щоб знайти стандартне відхилення, слід обчислити дисперсію, а потім знайти з неї корінь квадратний.

Для того, щоб мати змогу порівнювати якісь ряди, слід щоб виконувалася така вимога до отриманих значень:

$$\overline{X} - 3s_x \le X_i \le \overline{X} + 3s_x$$

Крім того, стандартне відхилення застосовують для обчислення коефіцієнта варіації.

КОЕФІЦІЄНТ ВАРІАЦІЇ відображає відносні коливання всіх значень ряду відносно його середнього арифметичного.

$$V = \frac{s_x}{\overline{X}} \cdot 100$$

Знайдемо коефіцієнти варіації для рядів:

(10 20 30 40 50) (10 10 10 10 50)

**Перший ряд** (табл. 1.4.2.):

*Таблиця 1.4.2.*

*Обчислення коефіцієнта варіації для першого ряду*

| Дані | Різниця $X_i - \overline{X}$ | Квадрат різниці $\left(X_i - \overline{X}\right)^2$ | Дисперсія $\sigma^2$ | Стандартне відхилення *S* | Коефіцієнт варіації ***V*** |
|---|---|---|---|---|---|
| 10 | -20 | 400 | 250 | 15,8 | 52,7% |
| 20 | -10 | 100 | | | |
| 30 | 0 | 0 | | | |
| 40 | 10 | 100 | | | |
| 50 | 20 | 400 | | | |
| $\overline{X}$ =30 | | $\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = 1000$ | | | |

**Другий ряд** (табл. 1.4.3.)**:**

*Таблиця 1.4.3.*

*Обчислення коефіцієнта варіації для другого ряду*

| Дані | Різниця $X_i - \overline{X}$ | Квадрат різниці $\left(X_i - \overline{X}\right)^2$ | Дисперсія $\sigma^2$ | Стандартне відхилення *S* | Коефіцієнт варіації ***V*** |
|---|---|---|---|---|---|
| 10 | -8 | 64 | 320 | 17,9 | 99,4% |
| 10 | -8 | 64 | | | |
| 10 | -8 | 64 | | | |
| 10 | -8 | 64 | | | |
| 50 | 32 | 1024 | | | |
| $\overline{X}$ =18 | | $\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = 1200$ | | | |

Зрозуміло, що ряди з такими коефіцієнтами варіації порівнювати буде зовсім некоректно.

*Інша справа, що взагалі неможливо порівнювати між собою ряди, які складаються з 5 елементів, шукати для них показники центральної тенденції та показники варіації (див. с. 22).*

# Розділ ІІ. АНАЛІЗ ЗВ'ЯЗКУ МІЖ ЗМІННИМИ

## § 2.1. МІРИ ЗВ'ЯЗКУ В ПАРАМЕТРИЧНІЙ СТАТИСТИЦІ

1. Загальне поняття про зв'язок. Статистичний та функціональний зв'язок.
2. Пошук зв'язку з допомогою діаграм розсіювання.
3. Коефіцієнт кореляції Пірсона
4. Обчислення коефіцієнта кореляції Пірсона
5. Область значень коефіцієнта кореляції та рівень його статистичної значимості
6. Особливості та помилки коефіцієнта кореляції

Література:


1. Артемьева Е.Ю., Мартынов Е.М. Вероятностные методы в психологии. – М., 1975.
2. Власов К.П., Власов П.К., Киселева А.А. Методы исследований и организация экспериментов. – Х.: Издательство "Гуманитарный центр", 2002.
3. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976.
4. Горбатов Д.С. Практикум по психологическому исследованию: Учеб. пособие. – Самара: Издательский дом "БАХРАХ-М", 2000.
5. Готтсданкер Р. Основы психологического эксперимента / Пер. с англ. – М., 1982.
6. Молостов А.В. Элементы вариационной статистики. – К.: Урожай, 1965.
7. Музика О.Л. Курсові роботи з психології: навчально-методичний посібник. – Житомир, 1999.
8. Солсо Р.Л., Джонсон Х.Х., Бил М.К. Экспериментальная психология: практический курс. – СПб.: прайм-ЕВРОЗНАК, 2001.


### *1. ЗАГАЛЬНЕ ПОНЯТТЯ ПРО ЗВ'ЯЗОК. СТАТИСТИЧНИЙ ТА ФУНКЦІОНАЛЬНИЙ ЗВ'ЯЗОК.*

Дослідників часто цікавить питання, чи пов'язані між собою два явища. Наприклад, чи навчаються краще діти, які раніше навчилися читати? Чи залежить успішність учнів від чисельності класу? Чи пов'язаний рівень агресивності з певними кліматичними та погодними умовами?

Однак, перш ніж перейти до конкретного розгляду цього питання, слід з'ясувати деякі фундаментальні положення, які стосуються поняття зв'язку.

Зв'язок між змінними може бути двох типів – **функціональний** та **статистичний**.

**Функціональний зв'язок** – одному значенню однієї змінної відповідає лише одне значення іншої змінної. Прикладом цього зв'язку може бути будь-яка математична функція: $y = a + bx$, $y = x^2$ тощо. Графік функціонального зв'язку наведено на рисунку 2.1.1. зліва.

**Статистичний зв'язок** – одному значенню однієї змінної може відповідати декілька значень другої змінної. Прикладом цього зв'язку можуть бути результати гіпотетичного дослідження рівня успішності та інтелекту, наведені на рисунку справа (рис. 2.1.1.).

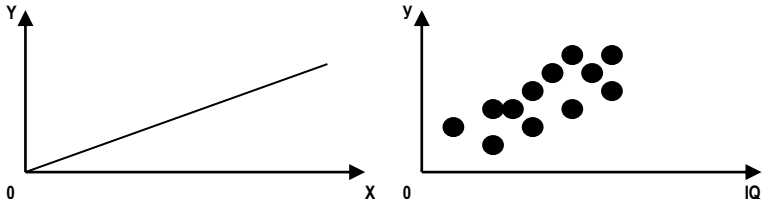



*Рис. 2.1.1. Функціональний та статистичний зв'язки*

Функціональний зв'язок є математично наближеним способом описання зв'язку між реальними процесами. Статистичний зв'язок відображає реальну картину цього зв'язку, однак, не дає можливості її узагальнення.

## *2. ПОШУК ЗВ'ЯЗКУ З ДОПОМОГОЮ ДІАГРАМ РОЗСІЮВАННЯ.*

Зв'язок між явищами можна зобразити графічно з допомогою **діаграм розсіювання**. Для побудови діаграми розсіювання на осі абсцис відкладають одну змінну, а на осі ординат – другу. Після цього результати кожного досліджуваного відкладають у цій системі координат. Утвориться картина, подібна до рис. 2.1.2.

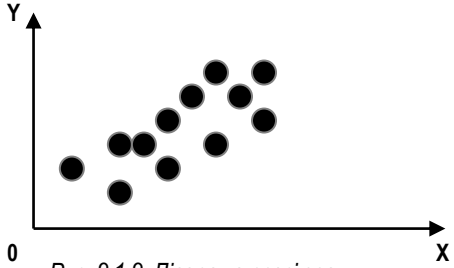



*Рис. 2.1.2. Діаграма розсіювання*

Як видно, між змінними Х та У існує позитивний зв'язок, оскільки зростання першої змінної призводить до зростання другої.

Однак, ми поки що не маємо узагальненої числової міри зв'язку між змінними, яка б дозволяла порівнювати між собою зв'язки різного рівня (прямі та обернені, слабкі та сильні).

## *3. КОЕФІЦІЄНТ КОРЕЛЯЦІЇ ПІРСОНА*

Для пошуку міри зв'язку між двома змінними розглянемо відхилення кожної змінної від середнього арифметичного: $X_i - \overline{X}$ та $Y_i - \overline{Y}$

Візьмемо для прикладу такі ряди даних (табл. 2.1.1.).

*Таблиця 2.1.1.*

| № | Інтелект | Успішність |
|---|---|---|
| | X | Y |
| 1 | 90 | 5 |
| 2 | 120 | 6 |
| 3 | 120 | 8 |
| 4 | 135 | 9 |
| 5 | 135 | 9 |
| 6 | 120 | 10 |
| 7 | 100 | 7 |
| 8 | 90 | 6 |
| 9 | 80 | 5 |
| 10 | 140 | 6 |
| 11 | 190 | 11 |
| $\sum$ | 1320 | 82 |
| $\overline{X}$ | 120 | 7,5 |

Тепер подивимося, що буде відбуватися з добутками $(X_i - \overline{X})(Y_i - \overline{Y})$ при їх підрахунку для різних досліджуваних. Візьмемо таких досліджуваних, у яких високий рівень інтелекту (І) та успішності (У), низький рівень І та У, високий рівень І та низький рівень У.

1. Високий рівень І та У (досліджуваний11) – добуток буде додатнім.
2. Низький рівень І та У (досліджуваний 8) – добуток теж буде додатнім.
3. Високий рівень І та низький рівень У (досліджуваний 10) – добуток буде від'ємним.

Таким чином, **якщо** змінні X та Y в основному пов'язані прямо, то більшість добутків ( $X_i - \overline{X}$ )( $Y_i - \overline{Y}$ ) будуть додатні, а значить і сума $\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)$ теж буде додатною.

**Якщо** ж більшість змінних пов'язані обернено, то більшість добутків ( $X_i - \overline{X}$ )( $Y_i - \overline{Y}$ ) будуть від'ємними, а значить і сума $\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)$ теж буде від'ємною.

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

У результаті отримують величину, яка називається **коваріацією**, і позначається $s_{xy}$ .

$$s_{xy} = \frac{\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{n-1}$$

Відмітьте, що коваріація змінної сама з собою є дисперсією.

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

У результаті ділення коваріації на стандартні відхилення цих груп отримується **коефіцієнт кореляції Пірсона**. Позначається він $r_{xy}$ :

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

З допомогою цієї формули коефіцієнт кореляції Пірсона означається, однак важко обчислюється. Тому для зручності обчислень з допомогою математичних перетворень виводять таку формулу:

$$r_{xy} = \frac{n \cdot \sum_i (x_i \cdot y_i) - \left(\sum_i x_i\right)\left(\sum_i y_i\right)}{\sqrt{\left(n \cdot \sum_i x_i^2 - \left(\sum_i x_i\right)^2\right) \cdot \left(n \cdot \sum_i y_i^2 - \left(\sum_i y_i\right)^2\right)}}$$

## 4. ОБЧИСЛЕННЯ КОЕФІЦІЄНТА КОРЕЛЯЦІЇ ПІРСОНА

*Нехай нам потрібно визначити, чи існує зв'язок між рівнем успішності учінв та їх особистісною тривожністю. Після проведення дослідження ми отримуємо два ряди даних. Обчислимо коефіцієнт лінійної кореляції Пірсона для того, щоб встановити характер зв'язку між ними.*

**ПРОЦЕДУРА ОБЧИСЛЕННЯ**

Передусім для спрощення обчислень дані варто згрупувати у таблицю (табл. 2.1.2.)

*Таблиця 2.1.2.*

*Обчислення коефіцієнта кореляції Пірсона*

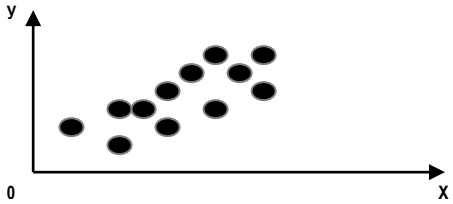
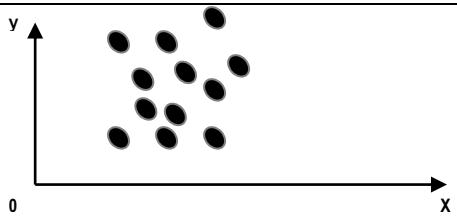
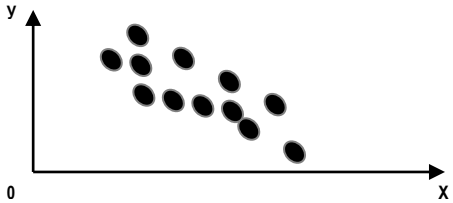
| № | **Учні** | Успішність учнів (**x**) | Особистісна тривожність (**y**) | $x_i^2$ | $y_i^2$ | $(x_i \cdot y_i)$ |
|---|---|---|---|---|---|---|
| 1. | А.О. | 10 | 48 | 100 | 2304 | 480 |
| 2. | Л.А. | 12 | 49 | 144 | 2401 | 588 |
| 3. | М.П. | 10 | 45 | 100 | 2025 | 450 |
| 4. | С.А. | 9 | 38 | 81 | 1444 | 342 |
| 5. | Р.К. | 8 | 34 | 64 | 1156 | 272 |
| 6. | В.О. | 7 | 25 | 49 | 625 | 175 |
| 7. | Д.К. | 10 | 42 | 100 | 1764 | 420 |
| 8. | В.К. | 9 | 41 | 81 | 1681 | 369 |
| 9. | М.А. | 11 | 43 | 121 | 1849 | 473 |
| 10. | Д.К. | 10 | 46 | 100 | 2116 | 460 |
| 11. | З.Ц. | 12 | 48 | 144 | 2304 | 576 |
| 12. | В.Ж. | 8 | 23 | 64 | 529 | 184 |
| | | $\sum_i x_i = 116$ | $\sum_i y_i = 482$ | $\sum_i x_i^2 = 1148$ | $\sum_i y_i^2 = 20198$ | $\sum_i (x_i \cdot y_i) = 4789$ |

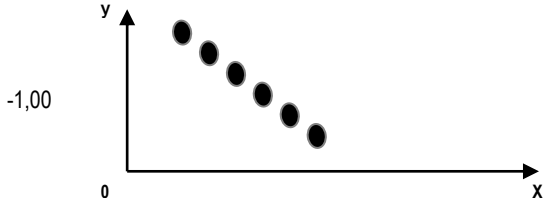
Після проведення всіх обчислень значення з нижніх комірок таблиці підставляються у формулу коефіцієнта лінійної кореляції Пірсона, і отримується значення $r_{xy}$ **=0,87**

## *5. ОБЛАСТЬ ЗНАЧЕНЬ КОЕФІЦІЄНТА КОРЕЛЯЦІЇ ТА РІВЕНЬ ЙОГО СТАТИСТИЧНОЇ ЗНАЧИМОСТІ*

*Таблиця 2.1.3.*

*Область значень коефіцієнта кореляції*

| Величина $r_{xy}$ | Тип діаграми розсіювання | Опис зв'язку |
|---|---|---|
| +1,00 | | Строгий прямий зв'язок |
| Близько +0,50 |  | Помірний прямий зв'язок |
| 0,00 |  | Немає зв'язку |
| Близько - 0,50 |  | Помірний обернений зв'язок |




Строгий обернений зв'язок

Для оцінки значимості $r_{xy}$ користуються таблицею критичних значень $r_{xy}$. Якщо отриманий показник $r_{xy}>r_{табл}$ на одному з рівнів значимості, то він є статистично достовірним (табл. 2.1.4.).

*Таблиця 2.1.4.*

*Критичні значення $r_{xy}$*
*(для рівнів значимості 0,05 та 0,01)*[2]

| N | 0,05 | 0,01 | N | 0,05 | 0,01 | N | 0,05 | 0,01 |
|---|---|---|---|---|---|---|---|---|
| 10 | 0,63 | 0,77 | 21 | 0,43 | 0,55 | 32 | 0,35 | 0,45 |
| 11 | 0,60 | 0,74 | 22 | 0,42 | 0,54 | 33 | 0,34 | 0,44 |
| 12 | 0,58 | 0,71 | 23 | 0,41 | 0,53 | 34 | 0,34 | 0,44 |
| 13 | 0,55 | 0,68 | 24 | 0,40 | 0,52 | 35 | 0,33 | 0,43 |
| 14 | 0,53 | 0,66 | 25 | 0,40 | 0,51 | 36 | 0,33 | 0,42 |
| 15 | 0,51 | 0,64 | 26 | 0,39 | 0,50 | 37 | 0,32 | 0,42 |
| 16 | 0,50 | 0,62 | 27 | 0,38 | 0,49 | 38 | 0,32 | 0,41 |
| 17 | 0,48 | 0,61 | 28 | 0,37 | 0,48 | 39 | 0,31 | 0,41 |
| 18 | 0,47 | 0,59 | 29 | 0,37 | 0,47 | 40 | 0,31 | 0,40 |
| 19 | 0,46 | 0,58 | 30 | 0,36 | 0,46 | 41 | 0,31 | 0,40 |
| 20 | 0,44 | 0,56 | 31 | 0,36 | 0,46 | 42 | 0,30 | 0,39 |

## *6. ОСОБЛИВОСТІ ТА ПОМИЛКИ ІНТЕРПРЕТАЦІЇ КОЕФІЦІЄНТА КОРЕЛЯЦІЇ*

1. **Причинність та кореляція**

   Те, що між величинами існує кореляційний зв'язок, ще не означає, що між ними існує причинний зв'язок:
   - явище 1 може бути причиною явища 2;
   - явище 2 може бути причиною явища 1;
   - причиною явищ 1 та 2 може бути подія 3.
   - причиною явищ 1 та 2 можуть бути явища 3, 4, 5, 6 в різних пропорціях.

[2] *Музика О.Л. Курсові роботи з психології: навчально-методичний посібник. – Житомир, 1999. – С. 26.*

Однак, хоча і кореляція не вказує на причинний зв'язок, вона може бути ключем до його знаходження, адже на її основі можна сформулювати певні гіпотези і потім їх детально перевірити в експерименті.

2. **Часто виникають помилкові кореляції:** наприклад, об'єднуються дві групи з різними середніми, в кожній з яких кореляція відсутня.
   Якщо, наприклад, в одну групу об'єднати хлопчиків та дівчаток, а потім знайти кореляцію між рівнем їх успішності та фізичною силою, то можна виявити значимий обернений зв'язок. Причина полягає у тому, що дівчатка в школі вчаться краще і фізично слабші, а хлопці вчаться дещо гірше, однак, фізично сильніші. Якщо спробувати знайти зв'язок між фізичною силою та успішністю окремо серед хлопців та окремо серед дівчат, то виявимо, що зв'язок відсутній.
3. **Коефіцієнт кореляції оцінює лише лінійний зв'язок**, а залежність між змінними може бути нелінійною: U-подібна, експоненційна, синусоїдна і т.д. Для виявлення цього типу залежностей використовують складніший метод – регресійний аналіз.
4. **Формула Пірсона основана на гіпотезі нормальності**, тому не може застосовуватися до малих вибірок і до даних, які розподілені ненормально.
5. Інколи виникають спроби використати коефіцієнт кореляції для порівняння середніх показників. Так, в одній із статей у поважному психологічному журналі описувався психологічний тренінг. Автор, взявши показники до та після тренінгу, порахував для них коефіцієнт кореляції і на основі отриманої величини зробив висновок про його ефективність.
   Це в принципі **неможлива задача**, адже з допомогою обчислення коефіцієнта кореляції можна знайти лише зв'язок між змінними. В нашому випадку автор знаходить зв'язок між показниками до і після тренінгу, – але ж це нічого не говорить про статистичну значимість відмінностей між цими ж показниками.
   Дані можуть мати високий рівень кореляції між собою, але не мати суттєвої різниці в середніх величинах. Наведемо гіпотетичний приклад, коли середні показники абсолютно однакові, а кореляція – висока:

| Учні | Показник успішності **до** інтелектуального тренінгу | Показник успішності **після** інтелектуального тренінгу |
|---|---|---|
| Антон | 6 | 6 |
| Василь | 7 | 8 |
| Петро | 8 | 9 |
| Віталій | 9 | 9 |

| | | |
|---|---|---|
| Іван | 10 | 9 |
| Марія | 11 | 10 |
| Ганна | 12 | 12 |
| **Середній показник** | **9** | **9** |
| **Коефіцієнт кореляції Пірсона** | r=0,93; p<0,002 | |

Як бачимо, в середня величина успішності учнів після тренінгу така ж сама, як і до нього. Тобто, його ефективність – нульова, а зміни в успішності окремих учнів швидше за все обумовлені сторонніми факторами. В той же час дані досить сильно між собою корелюють.

Можуть бути й такі варіанти: середні показники однакові, а кореляція низька; середні показники різні, а кореляція висока; середні показники різні, а кореляція низька. Тобто, **задача порівняння середніх** не має ніякого відношення до **задачі пошуку зв'язку**!

Для правильного вирішення задачі порівняння середніх користуються методами статичного висновку. Остільки і нашому випадку йдеться мова про залежні вибірки (до і після тренінгу діагностувалися одні і ті ж досліджувані), то можна було б застосувати критерій знаків або t-критерій Стьюдента для залежних вибірок (див. Розділ III).

## § 2.2. МІРИ ЗВ'ЯЗКУ В НЕПАРАМЕТРИЧНІЙ СТАТИСТИЦІ

1. Типи вимірювання змінних
2. Коефіцієнт кореляції φ
3. Точковий бісеріальний коефіцієнт кореляції
4. Тетрахоричний коефіцієнт кореляції
5. Бісеріальний коефіцієнт кореляції
6. Коефіцієнт рангової кореляції Спірмена
7. Коефіцієнт τ-Кендалла
8. Бісеріальний коефіцієнт рангової кореляції
9. Множинна кореляція

Література

1. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976.
2. Горбатов Д.С. Практикум по психологическому исследованию: Учеб. пособие. – Самара: Издательский дом "БАХРАХ-М", 2000.

3. Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО “Речь”, 2000.
4. Музика О.Л. Курсові роботи з психології: навчально-методичний посібник. – Житомир, 1999.

## *1. ТИПИ ВИМІРЮВАННЯ ЗМІННИХ*

Ми з’ясували, що для застосування коефіцієнта кореляції Пірсона необхідно, щоб дані були розподілені нормально, а вибірка була достатньо великою (параметрична статистика). Тепер ми розглянемо інші коефіцієнти кореляції, які використовуються у випадках, коли дані розподілені не нормально, або коли вони виміряні в інших шкалах, до яких традиційний коефіцієнт кореляції незастосовний (непараметрична статистика).

Застосування певного коефіцієнту кореляції залежить від типу шкали, в якій було проведено вимірювання. По відношенню до потреб пошуку зв’язку дані можуть бути виміряні 4-ма способами (табл. 2.2.1.):

1. Вимірювання в дихотомічній шкалі найменувань.
2. Вимірювання в дихотомічній шкалі найменувань з припущенням про нормальний розподіл даних.
3. Вимірювання в порядковій шкалі.
4. Вимірювання в шкалах інтервалів чи відношень.

*Таблиця 2.2.1.*

| | |
|---|---|
| Вимірювання в дихотомічній шкалі найменувань | З допомогою 0 та 1 фіксується наявність чи відсутність прояву певної ознаки (0-чоловік, 1 – жінка; 0-учень, 1-студент ...) |
| Вимірювання в дихотомічній шкалі найменувань з припущенням про нормальний розподіл даних | Робиться припущення, що точніші дані могли б забезпечити нормальний розподіл, але ті дані, що є, можуть бути закодовані лише як такі, що перевищують чи ні певний рівень ознаки. |
| Вимірювання в порядковій шкалі | Дані є послідовними рангами |
| Вимірювання в шкалах інтервалів чи відношень | Дані є інтервальними величинами чи відношеннями |

Для дослідження зв’язку використовують два ряди даних, а кожен з рядів може бути виміряний різним способом. Нижче наведені залежності використаних коефіцієнтів кореляції від способу виміру кожної із змінних (табл. 2.2.2).

*Таблиця 2.2.2.*

*Залежність типу коефіцієнта кореляції від способу вимірювання*

<table>
<tr><th>X<br>Y</th><th>Вимірювання в дихотомічній шкалі найменувань</th><th>Вимірювання в дихотомічній шкалі найменувань з припущенням про нормальний розподіл даних</th><th>Вимірювання в порядковій шкалі</th><th>Вимірювання в шкалах інтервалів чи відношень</th></tr>
<tr><td>Вимірювання в дихотомічній шкалі найменувань</td><td>Коефіцієнт φ</td><td>Коефіцієнт φ (в більшості випадків)</td><td rowspan="2">Бісеріальний ранговий коефіцієнт кореляції Кертена і Гласна</td><td>Точковий бісеріальний коефіцієнт кореляції</td></tr>
<tr><td>Вимірювання в дихотомічній шкалі найменувань з припущенням про нормальний розподіл даних</td><td>Коефіцієнт φ (в більшості випадків)</td><td>Тетрахоричний коефіцієнт кореляції</td><td>Бісеріальний коефіцієнт кореляції</td></tr>
<tr><td>Вимірювання в порядковій шкалі</td><td colspan="2">Бісеріальний ранговий коефіцієнт кореляції Кертена і Гласна</td><td>Коефіцієнт рангової кореляції Спірмена або τ Кендалла</td><td>*Міра не існує.* Слід шкалу інтервалів перетворити в рангову</td></tr>
<tr><td>Вимірювання в шкалах інтервалів чи відношень</td><td>Точковий бісеріальний коефіцієнт кореляції</td><td>Бісеріальний коефіцієнт кореляції</td><td>*Міра не існує.* Слід шкалу інтервалів перетворити в рангову</td><td>Коефіцієнт кореляції Пірсона</td></tr>
</table>

## *2. КОЕФІЦІЄНТ КОРЕЛЯЦІЇ φ*

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

*Нехай маємо ряди даних (табл. 2.2.3): Х – сімейний стан (0-неодружений, 1 – одружений), а У – навчання у вузі (0 – продовжує навчання, 1 - виключений).*

*Таблиця 2.2.3*

| № | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Х** | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 |
| **У** | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |

Очевидно, що до таких даних прямо застосувати коефіцієнт кореляції Пірсона неможливо. Однак, в результаті певних математичних міркувань з нього можна отримати формулу для нашого випадку.

Вона матиме такий вигляд:

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

де:

$p_{xy}$ – відсоток людей, що мають одиницю по Х та по У одночасно,

$p_x$ – відсоток людей, що мають одиницю по Х,

$p_y$ – відсоток людей, що мають одиницю по У,

$q_x$ – відсоток людей, що мають нуль по Х,

$q_y$ – відсоток людей, що мають нуль по У.

При цьому $q_x = 1 - p_x$, а $q_y = 1 - p_y$

Для отримання цієї формули достатньо формулу коефіцієнта кореляції Пірсона поділити на n та з'ясувати, що середнє арифметичне для дихотомічних шкал є вираженням відсотку кількості одиниць.

Якщо нас не цікавлять відсотки одиниць та нулів, ми користуємося **таблицею спряженості ознак**, яка має в загальному такий вигляд (табл. 2.2.4):

*Таблиця 2.2.4.*

| | | Змінна Х | | Всього |
|---|---|---|---|---|
| | | 0 | 1 | |
| Змінна У | 1 | a | b | **a+b** |
| | 0 | c | d | **c+d** |
| | Всього | **a+c** | **b+d** | n |

Тоді обчислення проводять за такою формулою:

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

Для нашого випадку таблиця спряженості ознак матиме такий вигляд (табл. 2.2.5):

*Таблиця 2.2.5*

| | | Змінна Х | | Всього |
|---|---|---|---|---|
| | | 0 | 1 | |
| Змінна У | 1 | 2 | 4 | **6** |
| | 0 | 5 | 1 | **6** |
| | Всього | **7** | **5** | 12 |

Отже, $\varphi = 0{,}507$

## *3. ТОЧКОВИЙ БІСЕРІАЛЬНИЙ КОЕФІЦІЄНТ КОРЕЛЯЦІЇ*

Цей коефіцієнт кореляції використовується у випадку, коли одна змінна виміряна у дихотомічній шкалі найменувань, а друга – в шкалі інтервалів чи відношень (бісеріальний – тому що мова йдеться про дві серії об'єктів – ті, що мають 0 по Х, і ті, що мають 1 по Х).

Прикладом цього випадку може бути спроба дослідження зв'язку між статтю та рівнем тривожності (табл. 2.2.6):

*Таблиця 2.2.6.*

| № | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Стать (1-ч, 0-ж) (**Х**) | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 |
| Тривожність (**У**) | 150 | 170 | 160 | 165 | 140 | 183 | 157 | 152 | 163 | 168 | 160 | 155 | 157 | 160 | 152 |

Формула для цього коефіцієнта кореляції теж отримана з формули Пірсона, і має вигляд:

$$r_{pb} = \frac{\overline{Y}_1 - \overline{Y}_0}{s_y}\sqrt{\frac{n_1 n_0}{n(n-1)}},$$

де:

$\overline{Y}_1$ – середнє арифметичне змінної У, пораховане для об'єктів, що мають по Х одиницю,

$\overline{Y}_0$ – середнє арифметичне змінної У, пораховане для об'єктів, що мають по Х нуль,

$s_y$ – стандартне відхилення У,

$n_1$ – кількість об'єктів по Х, які мають одиницю,

$n_0$ – кількість об'єктів по Х, які мають нуль,

$n$ – загальна кількість об'єктів.

Для наших даних: $n$ =15, $n_0$ =7, $n_1$=8, $s_y$ =8,94, $\overline{Y}_1$ =163,25, $\overline{Y}_0$ =156,57 і в результаті $r_{pb}$ =0,41.

## *4. ТЕТРАХОРИЧНИЙ КОЕФІЦІЄНТ КОРЕЛЯЦІЇ*

Якщо ми маємо справу із двома змінними, виміряними в дихотомічних шкалах, але вважаємо, що з використанням інших складніших методів, які нам поки що недоступні, можна було б отримати точніші дані, які розподілилися б нормально, то варто використати тетрахоричний коефіцієнт кореляції.

Як можна отримати ці виміри?

> *Уявімо, що в нас немає лінійки, а є лише планка висотою 1,6 м. Тоді, ми можемо виявити лише людей, зріст яких менше нашої планки (позначаємо їх зріст 0), або вище (позначаємо 1). Шкала дихотомічна, однак, ми знаємо, що висота зросту серед людей одного віку розподілена за нормальним законом.*

Уявімо, що було проведено дослідження зв'язку між мотивацією досягнення та рівнем тривожності, у якому взяли участь 100 осіб. Однак, внаслідок відсутності ліцензії на тестові методики, які вимірюють ці якості точно, ми змушені скористатися спрощеним варіантом, який може сказати лише, є чи немає мотивації досягнення (Х: 0 та 1) та є чи немає тривожності (У: 0 та 1).

Для обчислення тетрахоричного коефіцієнту кореляції використовують таблицю спряженості ознак (табл. 2.2.7.):

*Таблиця 2.2.7.*

| | | Змінна X | | |
|---|---|---|---|---|
| | | 0 | 1 | Всього |
| Змінна У | 1 | a=5 | b=25 | **30** |
| | 0 | c=64 | d=6 | **70** |
| Всього | | **69** | **31** | 100 |

Формула для обчислення тетрахоричного коефіцієнта кореляції має такий вигляд:

$$r_{tet} = \cos\frac{180^\circ}{1+\sqrt{bc/ad}}$$

У нашому випадку $r_{tet}$ =0,93.

Однак, цей показник має певні обмеження: його можна застосовувати лише у випадках, коли величина $\frac{a+b}{n}$ або $\frac{b+d}{n}$ незначно відхиляється від 0,50. У випадках, коли ці величини більше 0,7 чи менше 0,3 – використовувати цю формулу взагалі не можна, оскільки вона дає значну похибку (завищує прямий зв'язок).

## *5. БІСЕРІАЛЬНИЙ КОЕФІЦІЄНТ КОРЕЛЯЦІЇ*

Бісеріальний коефіцієнт кореляції обчислюється у випадках, коли одна змінна виміряна в дихотомічній шкалі з припущенням її нормальності, а друга – в шкалі інтервалів чи відношень.

Виведення бісеріального коефіцієнту кореляції ґрунтується на теорії регресії, яку ми поки що не розглядали, а тому приймемо його без доведення.

$$r_{bis} = \frac{\overline{X}_1 - \overline{X}_0}{s_x} \cdot \frac{n_1 n_0}{un\sqrt{n^2 - n}}$$

Якщо змінна X виміряна в шкалі інтервалів, а У – в дихотомічній шкалі з припущенням нормальності, то

$\overline{X}_1$ – середнє арифметичне значення змінної X, пораховане для об'єктів, які мають 1 по У

$\overline{X}_0$ – середнє арифметичне значення змінної Х, пораховане для об'єктів, які мають 0 по У

$s_x$ – стандартне відхилення змінної Х

$n_1$ – кількість об'єктів, які мають 1 по У

$n_0$ – кількість об'єктів, які мають 0 по У

$n$ – загальна кількість об'єктів

$u$ – висота нормованого нормального розподілу в точці, за якою лежить $n_1/n$ відсотків його площі (величина шукається по таблицях)

Особливістю цього коефіцієнту кореляції є те, що він може виходити за межі (-1; +1), однак, це означає лише, що гіпотеза нормальності для нашої вибірки виявилася некоректною.

## *6. КОЕФІЦІЄНТ РАНГОВОЇ КОРЕЛЯЦІЇ СПІРМЕНА*

У випадку, коли обидві змінні виміряні в порядкових шкалах, використовують ранговий коефіцієнт кореляції Спірмена, який легко можна вивести із коефіцієнта кореляції Пірсона, зробивши певні припущення стосовно рангових величин (Х та У приймають значення, рівні 1, 2, 3, ... ).

Уявімо, що ми провели дослідження, в якому експерт з допомогою рангування визначав, наскільки студенту подобається його група (Х), а також його куратор (У) – табл. 2.2.8.

Формула Спірмена ґрунтується на припущенні, що $\sum X_i = \sum Y_i$, і має такий вигляд:

$$r_s = 1 - \frac{6\sum\left(X_i - Y_i\right)^2}{n\left(n^2 - 1\right)}$$

В нашому випадку $r_s$ =0,43

*Таблиця 2.2.8.*

*Обчислення коефіцієнта рангової кореляції Спірмена*

| № | **X** | **У** | $(X_i - Y_i)$ | $(X_i - Y_i)^2$ |
|---|---|---|---|---|
| 1 | 1 | 5 | -4 | 16 |
| 2 | 8 | 1 | 7 | 49 |
| 3 | 3 | 2 | 1 | 1 |
| 4 | 2 | 3 | -1 | 1 |
| 5 | 4 | 4 | 0 | 0 |
| 6 | 5 | 6 | -1 | 1 |
| 7 | 6 | 9 | -3 | 9 |
| 8 | 7 | 10 | -3 | 9 |
| 9 | 10 | 8 | 2 | 4 |
| 10 | 9 | 7 | 2 | 4 |
| $\sum$ | 55 | 55 | | 94 |

Однак, при обчисленні коефіцієнта кореляції Спірмена може виникнути **проблема зв'язаних рангів**.

Уявімо, що наш експерт не зміг прорангувати 1, 3 та 4 досліджуваних. В такому випадку їм слід приписати ранги за таким правилом:

> *Уявно все ж таки проставити цим досліджуваним ранги (1, 2, 3), потім знайти їх середнє арифметичне і реально проставити їм знайдений середній ранг (2). Подальше рангування слід продовжувати вже не з 2, а з того рангу, на якому закінчилося уявне рангування (після 3).* ***Такі ранги називають зв'язаними.***

Однак, якщо ми тепер порахуємо суми рангів в обох змінних, то побачимо, що базове для формули Спірмена припущення $\sum X_i = \sum Y_i$ не справджується. У такому випадку див. табл. 2.2.9:

*Таблиця 2.2.9.*

| **Кількість зв'язаних рангів** | Зв'язаних рангів мало | Зв'язаних рангів багато |
|---|---|---|
| **Проблема** | Відмінності між сумами незначні | Відмінності між сумами значні |
| **Вирішення** | Ігнорувати відмінності, пам'ятаючи, що формула може дати похибку | Застосувати формулу Пірсона або τ-Кендалла (див. далі) із поправкою на зв'язані рангів |

## 7. КОЕФІЦІЄНТ $\tau$-КЕНДАЛЛА

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

*Таблиця 2.2.10.*

| № | **X** (подобається група) | **У** (подобається куратор) | Співпадання (P) | Інверсії (Q) |
|---|---|---|---|---|
| 1 | 1 | 5 | 5 | 4 |
| 4 | 2 | 3 | 6 | 2 |
| 3 | 3 | 2 | 6 | 1 |
| 5 | 4 | 4 | 5 | 1 |
| 6 | 5 | 6 | 4 | 1 |
| 7 | 6 | 9 | 1 | 3 |
| 8 | 7 | 10 | 0 | 3 |
| 2 | 8 | 1 | 2 | 0 |
| 10 | 9 | 7 | 1 | 0 |
| 9 | 10 | 8 | 0 | 0 |
| | | $\sum$ | 30 | 15 |

Для підрахунку коефіцієнта $\tau$-Кендалла використовують такий алгоритм (табл. 2.2.10):

1. Впорядковують ранги по одній із змінних.
2. Підраховується кількість "співпадань" (P): для кожного об'єкта підраховується, скільки разів його ранг по У виявляється менше від рангів об'єктів, які знаходяться нижче нього.
3. Підраховується кількість "інверсій" (Q): для кожного об'єкта підраховується, скільки разів його ранг по У виявляється більше, ніж ранги об'єктів, які знаходяться нижче.
4. Обчислюється коефіцієнт $\tau$ за формулою:

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

де:

$P$ - кількість співпадань,

$Q$ - кількість інверсій,

$n$ - загальна кількість об'єктів.

Цю формулу можна спростити, якщо врахувати, що $n(n-1)/2$ – це загальна кількість можливих пар $n$ об'єктів, і $n(n-1)/2 = P + Q$. Тоді можна отримати дві абсолютно еквівалентні формули:

$$\tau = \frac{4P}{n(n-1)} - 1 \text{ або } \tau = 1 - \frac{4Q}{n(n-1)}$$

В нашому випадку ми отримаємо τ=0,33.

Як можна помітити, τ-Кендалла та r-Спірмена для одних і тих же даних дають **різні числові результати**. Причина – в різних логіках побудови коефіцієнтів. Якщо r-Спірмена можна проінтерпретувати як коефіцієнт Пірсона для рангових величини, то τ-Кендалла відображає різницю ймовірностей між тим, що дані в обох рядах даних мають однаковий порядок і тим, що дані в обох рядах даних мають різний порядок.

Співвідношення між цими коефіцієнтами можна виразити так:

$$-1 \le 3 \cdot \tau - 2 \cdot r_s \le 1$$

Ще одна проблема – **зв'язані ранги**. Якщо вони є, то у формулу вводять певні поправки:

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

де:

$$K_x = \frac{1}{2}\sum f_{xi}(f_{xi} - 1) \text{ та } K_y = \frac{1}{2}\sum f_{yi}(f_{yi} - 1)$$

$f_{xi}$ - число зв'язаних рангів у кожній групі зв'язків по Х,

$f_{yi}$ - число зв'язаних рангів у кожній групі зв'язків по У.

А тепер давайте повернемося до випадку, коли одна змінна виміряна в дихотомічній шкалі найменувань, а друга – в порядковій шкалі. У цьому випадку послуговуються бісеріальним коефіцієнтом рангової кореляції Кертена та Гласса. Ми розглянули його в кінці тому, що він розроблений на основі тих же методів, що і коефіцієнт $\tau$ Кендалла.

Нехай ми знову провели дослідження зв'язку між статтю та рівнем тривожності, однак, оскільки в нас не було змоги дістати методики, необхідні для виявлення рівня тривожності, ми скористалися оцінкою психолога-експерта, який ранжував наших досліджуваних за їх рівнем тривожності (табл. 2.2.11.).

*Таблиця 2.2.11*

| № | **X** (стать, 1 – ч, 0 - ж) | **y** (тривожність) |
|---|---|---|
| 1 | 0 | 1 |
| 2 | 1 | 10 |
| 3 | 0 | 2 |
| 4 | 1 | 9 |
| 5 | 0 | 5 |
| 6 | 0 | 8 |
| 7 | 1 | 4 |
| 8 | 1 | 7 |
| 9 | 0 | 3 |
| 10 | 0 | 6 |

У такому випадку слід, по-перше, впорядкувати ранги, після чого розписати ранги по принципу належності до 1 чи до 0 (табл. 2.2.12).

Наступний етап – підрахунок співпадань та інверсій. У цьому випадку дані організовані таким чином, що кількість співпадань рахується для рангів в стовпчику 1 і рівна кількості об'єктів, які знаходяться в стовпчику 0 нижче взятого рангу в стовпчику 1.

Кількість інверсій рахується для рангів у стовпчику 1 і рівна кількості об'єктів, які знаходяться в стовпчику 1 нижче взятого рангу стовпчику 0.

*Таблиця 2.2.12*

| № | **X** (стать, 1 – ч, 0 - ж) | **У** (тривожність) | **Ранги по У** | | Співпадання (P) | Інверсії (Q) |
|---|---|---|---|---|---|---|
| | | | **X=1** | **X=0** | | |
| 2 | 1 | 10 | 10 | | 6 | |
| 4 | 1 | 9 | 9 | | 6 | |
| 6 | 0 | 8 | | 8 | | 2 |
| 8 | 1 | 7 | 7 | | 5 | |
| 10 | 0 | 6 | | 6 | | 1 |
| 5 | 0 | 5 | | 5 | | 1 |
| 7 | 1 | 4 | 4 | | 3 | |
| 9 | 0 | 3 | | 3 | | |
| 3 | 0 | 2 | | 2 | | |
| 1 | 0 | 1 | | 1 | | |
| | | | | $\sum$ | 20 | 4 |

Після цього підраховується сума співпадань та інверсій і підставляється у формулу:

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

де:

$n_0$ – кількість нулів по дихотомічній змінній,

$n_1$ – кількість одиниць по дихотомічній змінній.

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

$$r_{rb} = \frac{2}{n}\left(\overline{Y}_1 - \overline{Y}_0\right),$$

де:

$\overline{Y}_1$ – середній ранг об'єктів, що мають по X одиницю,

$\overline{Y}_0$ – середній ранг об'єктів, що мають по X нуль.

Таким чином, можна користуватися будь-яким з цих двох коефіцієнтів, залежно від того, який Вам зручніше. Однак, слід пам'ятати, що при розрахунках в межах рангових шкал з'являється проблема зв'язаних рангів. У таких випадках наведені вище формули застосовувати не можна, а тому слід уникати таких ситуацій.

## 9. МНОЖИННА КОРЕЛЯЦІЯ

Останній метод кореляційного аналізу, який ми розглянемо – це метод множинної кореляції. До цього ми розглядали лише випадки, які стосувалися зв'язку між двома змінними. Однак, достатньо часто нас цікавить більша кількість змінних. Тому для оцінки зв'язку в таких випадках було розроблено цей метод.

Ми розглянемо лише один випадок, для якого варто застосовувати цей метод. У інших випадках продуктивнішими є методи регресійного та факторного аналізу, які ми розглянемо пізніше.

Нехай нас цікавить зв'язок між змінною У та іншими двома змінними $X_1$ та $X_2$.

*Нехай $X_1$ – рівень інтелекту, $X_2$ – рівень мотивації досягнення, а У – бал, отриманий на державних іспитах.*

Коефіцієнт множинної кореляції матиме такий вигляд:

$$R_{y-x_1,x_2} = \sqrt{b_1 r_{yx_1} - b_2 r_{yx_2}} \ ,$$

де:

$$b_1 = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{1 - r^2_{x_1x_2}} \text{, а } \ b_2 = \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{1 - r^2_{x_1x_2}} \ .$$

У цих формулах:

$r_{yx_1}$ – кореляція між У та $X_1$

$r_{yx_2}$ – кореляція між У та $X_2$

$r_{x_1x_2}$ – кореляція між $X_1$ та $X_2$

*Приміром, ми отримали такі значення цих коефіцієнтів:*

$r_{yx_1}$ *=0,50;* $r_{yx_2}$ *=0,50;* $r_{x_1x_2}$ *=0.*

Для обчислень множинної кореляції та аналізу зв'язків між досліджуваними змінними дані зручно організовувати в **інтеркореляційну матрицю** (табл. 2.2.13):

*Таблиця 2.2.13.*

| **Інтеркореляційна матриця** | | ЗМІННІ | | |
|---|---|---|---|---|
| | | **$X_1$** | **$X_2$** | **У** |
| ЗМІННІ | **$X_1$** | 1,0 | 0 | 0,50 |
| | **$X_2$** | | 1,0 | 0,50 |
| | **У** | | | 1,0 |

Після цього обчислюють $b_1$ та $b_2$, $b_1=0,50$, $b_2=0,50$.

У результаті отримуємо $R_{y-x_1,x_2}$ =0,71.

Таким чином, видно, що хоча дві Х-ві змінні між собою не пов'язані, а кожна з них окремо не сильно пов'язана з У, існує тісний зв'язок між ними та змінною У.

# § 2.3. РЕГРЕСІЙНИЙ АНАЛІЗ

1. Задачі регресійного аналізу.
2. Визначення коефіцієнтів регресії.
3. Обчислення значень залежної змінної.
4. Обчислення похибки рівняння регресії.
5. Види рівнянь регресії.
6. Загальні принципи вибору рівняння регресії.

Література


1. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976.
2. Власов К.П., Власов П.К., Киселева А.А. Методы исследований и организация экспериментов. – Х.: Издательство “Гуманитарный центр”, 2002.
3. Молостов А.В. Элементы вариационной статистики. – К.: Урожай, 1965.
4. Свалов Н.Н. Вариационная статистика. Учебное пособие для вузов. – М.: Лесная промышленность, 1977.


## *1. ЗАДАЧІ РЕГРЕСІЙНОГО АНАЛІЗУ*

Раніше ми зазначали, що пошук зв’язку між явищами є однією з основних задач будь-якої науки. Однак, аналіз зв’язку має дві сторони:

1. Сила зв’язку і його напрям (прямий чи обернений).
2. Форма зв’язку (пряма, параболічна тощо).

Для аналізу сили зв’язку та його напрямку ми з Вами використовували кореляційний аналіз, і наголошували на тому, що досить часто зв’язок може носити досить складний нелінійний характер, а тому одного лише кореляційного аналізу недостатньо. Необхідно аналізувати діаграму розсіювання, на якій наочно видно, якої форми може бути зв’язок.

Але одного лише наочного зображення для переконливих висновків недостатньо. Статистика дозволяє нам з великою точністю визначити форму зв’язку та записати її у вигляді математичного рівняння з залежною (результативною) та незалежною (факторною) змінною. Методом, з допомогою якого можна це зробити є регресійний аналіз.

*Таблиця 2.3.1.*

| ЗАДАЧІ | МЕТОДИ ВИРІШЕННЯ |
|---|---|
| Пошук сили зв'язку та його напряму | Кореляційний аналіз |
| Пошук форми зв'язку, передбачення змін | Регресійний аналіз |

Сам термін "регресія" запропоновано англійським вченим Френсісом Гальтоном.

***Френсіс Гальтон (16.02.1822, Бірмінгем - 17.01.1911)*** *вніс величезну кількість ідей у найрізноманітніші галузі науки – від метеорології до антропології. Тривалий період життя він провів у подорожах, результатом яких стала книга "Мистецтво подорожувати", яка мала величезний успіх як посібник по організації подорожей. Йому належить поняття "антициклон". Він розробив машину для отримування енергії з морських хвиль. Саме йому належить створення науки дактилоскопії.*

*Серед його важливих внесків у науку є розробка та впровадження у психологію статистичного методу. Гальтона цікавила проблема, абсолютно нова на той час, – спадкова основа індивідуальних відмінностей. Саме для пошуку закономірностей індивідуальних варіацій він і застосовував статистичний апарат.*

Але повернемося до регресійного аналізу. Ми сказали, що з його допомогою можна визначити форму зв'язку та розробити формулу, яка досить точно описуватиме цей зв'язок, і відповідно, дозволить здійснювати прогноз поведінки певних явищ.

Найпростішим випадком є лінійна форма зв'язку, а тому розгляд регресійного аналізу ми почнемо саме з нього.

## *2. ВИЗНАЧЕННЯ КОЕФІЦІЄНТІВ РЕГРЕСІЇ*

В одному з попередніх питань ми розглядали завдання – визначити зв'язок між кількістю учбовою успішністю учнів та їх особистісною тривожністю. Для

вирішення цього завдання ми користувалися коефіцієнтом лінійної кореляції Пірсона, і отримали такі результати – $r_{xy}$ **= 0,87** (див. табл. 2.1.2 і табл. 2.3.2.).

*Таблиця 2.3.2*

| № | **Учні** | Успішність учнів (**x**) | Особистісна тривожність (**y**) | $x_i^2$ | $y_i^2$ | $(x_i \cdot y_i)$ |
|---|---|---|---|---|---|---|
| 1. | А.О. | 10 | 48 | 100 | 2304 | 480 |
| 2. | Л.А. | 12 | 49 | 144 | 2401 | 588 |
| 3. | М.П. | 10 | 45 | 100 | 2025 | 450 |
| 4. | С.А. | 9 | 38 | 81 | 1444 | 342 |
| 5. | Р.К. | 8 | 34 | 64 | 1156 | 272 |
| 6. | В.О. | 7 | 25 | 49 | 625 | 175 |
| 7. | Д.К. | 10 | 42 | 100 | 1764 | 420 |
| 8. | В.К. | 9 | 41 | 81 | 1681 | 369 |
| 9. | М.А. | 11 | 43 | 121 | 1849 | 473 |
| 10. | Д.К. | 10 | 46 | 100 | 2116 | 460 |
| 11. | З.Ц. | 12 | 48 | 144 | 2304 | 576 |
| 12. | В.Ж. | 8 | 23 | 64 | 529 | 184 |
| | | $\sum_i x_i = 116$ | $\sum_i y_i = 482$ | $\sum_i x_i^2 = 1148$ | $\sum_i y_i^2 = 20198$ | $\sum_i (x_i \cdot y_i) = 4789$ |

Давайте спробуємо застосувати до цих же даних регресійний аналіз і встановимо форму та рівняння зв'язку між тривожністю та успішністю. Дійсно, результати кореляційного аналізу дають нам можливість сказати лише, що чим вища успішність, тим вища тривожність (або навпаки, нижча тривожність, тим нижча успішність), але ми не знаємо, на скільки зростає одна змінна при зменшенні іншої – на скільки зростає тривожність при зростанні успішності на 1 бал? Або: на скільки зростає успішність при зростанні тривожності на 1 бал?

Перш за все, необхідно побудувати діаграму розсіювання (ми вже знаємо, як це робити, див. рис. 2.3.1а). Нагадаємо, що у регресійному аналізові, як і в кореляційному, позначання однією змінної "У", а другої "Х" – це умовність. Що на що впливатиме ми не будемо знати до тих пір, поки не проведемо експеримент!!!

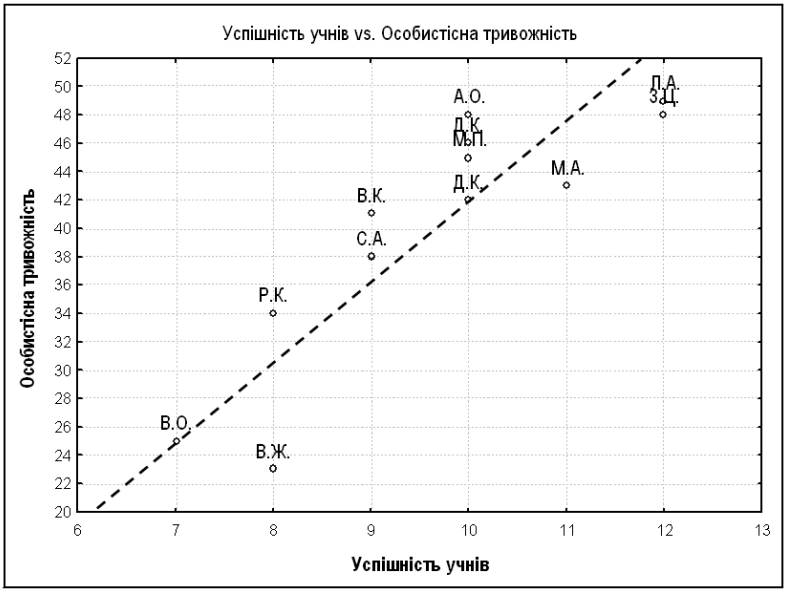



*Рис. 2.3.1а. Діаграма розсіювання, побудована для регресійного аналізу*

Ми домовилися, що будемо говорити про лінійний зв'язок між змінними. Очевидно, що нашою задачею буде знайти лінію, яка максимально наближатиметься до всіх точок діаграми розсіювання, і відповідно, наближено описуватиме всіх їх.

Загальне рівняння прямої лінії має вигляд:

$$Y = a + bX ,$$

де X – умовно незалежна змінна (факторна змінна), У – умовно залежна змінна (результативна змінна) (умовно – тому що ми на основі кореляційного та регресійного аналізу не можемо визначити, що є причиною, а що – наслідком).

Коефіцієнти **a** та **b називають коефіцієнтами регресії**, а саму лінію – лінією регресії.

У нашому випадку ***X – успішність учнів, У – особистісна тривожність.***

Найпростішим, але найнедосконалішим методом регресійного аналізу є **МЕТОД ВИБРАНИХ ТОЧОК**. Однак, його розгляд дасть нам можливість згодом зрозуміти точніші та складніші методи.

Суть **методу вибраних точок** полягає у тому, що на графіку обираються дві точки, так, щоб проведена через них лінія ділила все поле точок на приблизно однакові частини. У нашому випадку – це точки (В.О.) та (Д.К.) – рис. 2.3.1а. Підставляючи значення цих точок у рівняння прямої отримують систему з двох рівнянь, з яких знаходять коефіцієнти загального рівняння **a** та **b**.

У нашому випадку:

В.О.: X=7, Y=25;

Д.К.: X=10, Y=46.

Відповідно, утворюється система рівнянь:

$$\begin{cases} 25 = a + b \cdot 7 \\ 46 = a + b \cdot 10 \end{cases}$$

Розв'язавши її, отримуємо: b=7; a-24 .

А тому, загальне рівняння регресії для наших даних виглядатиме так:

$$Y = -24 + 7 \cdot X$$

Маючи це рівняння, ми можемо вийти за межі отриманих нами конкретних даних, зробивши прогноз і з'ясувавши, як буде змінюватися тривожність учнів із зростанням чи спаданням їх успішності.

Перевагою методу вибраних точок є його швидкість, а недоліком – неточність та приблизність. Він не враховує всіх значень, отриманих у дослідженні, а тому його використання обмежене первинним аналізом.

Тепер, розібравшись у найпростішому методі, можемо перейти до складнішого, але точнішого – **МЕТОДУ НАЙМЕНШИХ КВАДРАТІВ.**

Проведемо по нашій діаграмі розсіювання нову уявну лінію регресії (уявну, бо саме рівняння цієї лінії ми будемо шукати, це ніби ескіз) – рис. 2.3.1б. Для кожної реальної експериментальної точки можна на лінії побудувати теоретичну точку. Відповідно, рівняння цієї теоретичної кривої матиме вигляд:

$$Y_{теор} = a + bX \ ,$$

Де $Y_{теор}$ – теоретична ордината.

Відповідно, можна знайти різницю між теоретичними та експериментальними ординатами кожної точки:

$$\begin{cases} Y_{теор\,1} - Y_1 = a + bX_1 - Y_1 \\ \dots \\ Y_{теор\,10} - Y_{10} = a + bX_{10} - Y_{10} \end{cases}$$

Розв'язати таку систему рівнянь можна, однак, у процесі її розв'язку можна стикнутися з такою трудністю – внаслідок того, що різниця $\overline{Y_1} - Y_1$ може бути як додатною, так і від'ємною їх сума може бути рівною нулю. Тому вирішили знаходити не суми різниць експериментальних і теоретичних ординат, а суми квадратів різниць цих ординат – так зникає вплив знаків.

Отже, вирішення поставленої задачі зводиться до отримання такої лінії, для якої сума квадратів відхилень експериментальних значень від обчислених теоретично є найменшою.

У результаті розв'язку цих рівнянь (методами диференціального числення) отримується така загальна система:

$$\begin{cases} aN + b\sum X = \sum Y \\ a\sum X + b\sum X^2 = \sum XY \end{cases}$$

Отже, для пошуку лінії регресії методом найменших квадратів слід знайти всі необхідні суми, підставити їх у систему рівнянь і розв'язати цю систему – коефіцієнти будуть знайдені.

Для нашого випадку це буде таке **рівняння**:

$$\begin{cases} 12a + 116b = 482 \\ 116a + 1148b = 4789 \end{cases}$$

Розв'язавши його, ми отримаємо: a=-39,49; b=8,24.

А тому, загальне рівняння регресії для наших даних виглядатиме так:

$$Y = -39{,}49 + 8{,}24 \cdot X$$

Як бачимо, між методом вибраних точок та методом найменших квадратів існують певні відмінності.

## *3. ОБЧИСЛЕННЯ ЗНАЧЕНЬ ЗАЛЕЖНОЇ ЗМІННОЇ*

Рівняння регресії дає можливість знайти теоретичне значення Y, яке іноді називають обчисленим значенням, вирівняним значенням або найбільш імовірним значенням. Зокрема, якщо ми знаємо рівняння регресії для швидкості читання/кількості помилок, то можемо робити досить таки достовірний прогноз щодо нового учня, який прийшов до класу – вимірявши його швидкість читання, ми можемо знайти найбільш імовірну кількість помилок. Якщо виявиться, що його кількість помилок реальна значно відрізняється від теоретичної, то це може свідчити про певні особливості психічного розвитку учня, і значить, варто звернути на нього особливу увагу.

Те саме стосується нашого дослідження зв'язку між успішністю та особистісною тривожністю.

## *4. ОБЧИСЛЕННЯ ПОХИБКИ РІВНЯННЯ РЕГРЕСІЇ*

Лінія регресії не точно описує всю сукупність даних – за означенням вона максимально наближається до всіх точок діаграми розсіювання. Тому існує можливість відхилення даних від лінії регресії. Чим більше це відхилення, тим менш ефективною є регресія, і навпаки, чим менше це відхилення, тим регресія більш очевидна і ефективна.

Величину, яка характеризує похибку рівняння регресії (середню квадратичну похибку рівняння, вибіркове стандартне відхилення від лінії регресії) шукають так:

1. Обчислити всі теоретичні значення "залежної" змінної за формулою регресії.
2. Знайти різниці між теоретичними та експериментальними значеннями "залежної" змінної.
3. Піднести різниці до квадрату та знайти їх суму.
4. Поділити суму квадратів на N-2.
5. Знайти квадратний корінь з отриманого числа.

Формула для обчислення похибки рівняння регресії виглядає так:

$$s_{yx} = \sqrt{\frac{\sum\left(Y_{теор\,i} - Y_i\right)^2}{N-2}}$$

Для нашого випадку (табл. 2.3.3.):

*Таблиця 2.3.3.*

| № | **Учні** | Успішність учнів | Особистісна тривожність | Теоретичне значення | Різниця | Квадрат різниці |
|---|---|---|---|---|---|---|
| | | **X** | **Y** | $Y_{теор}$ | $\left(Y_{теор}-Y\right)$ | $\left(Y_{теор}-Y\right)^2$ |
| 1. | А.О. | 10 | 48 | 42,91 | -5,09 | 25,91 |
| 2. | Л.А. | 12 | 49 | 59,39 | 10,39 | 107,95 |
| 3. | М.П. | 10 | 45 | 42,91 | -2,09 | 4,37 |
| 4. | С.А. | 9 | 38 | 34,67 | -3,33 | 11,09 |
| 5. | Р.К. | 8 | 34 | 26,43 | -7,57 | 57,30 |
| 6. | В.О. | 7 | 25 | 18,19 | -6,81 | 46,38 |
| 7. | Д.К. | 10 | 42 | 42,91 | 0,91 | 0,83 |
| 8. | В.К. | 9 | 41 | 34,67 | -6,33 | 40,07 |
| 9. | М.А. | 11 | 43 | 51,15 | 8,15 | 66,42 |
| 10. | Д.К. | 10 | 46 | 42,91 | -3,09 | 9,55 |
| 11. | З.Ц. | 12 | 48 | 59,39 | 11,39 | 129,73 |
| 12. | В.Ж. | 8 | 23 | 26,43 | 3,43 | 11,76 |
| $s_{yx}=\sqrt{\frac{\sum\left(Y_{теор\,i}-Y_i\right)^2}{N-2}}=\sqrt{\frac{511{,}36}{10}}=7{,}15$ | | | | | | $\sum\left(Y_{теор}-Y\right)^2=$ 511,36 |

Можна знайти показник ефективності регресії, поділивши стандартне відхилення по У ($s_y$) на похибку регресії ($s_{yx}$): 8,73/7,15=1,22 або 122%.

Як і у випадку з дисперсією та стандартним відхиленням, значення похибки регресії саме по собі мало що каже. Воно проявляється лише у порівнянні. Зокрема, можна підрахувати похибку регресії для методу вибраних точок і порівняти з похибкою, отриманою при аналізі методом найменших квадратів. Те рівняння, яке буде мати меншу похибку, і буде більш точно відображати дійсність.

Також можна порівнювати похибки для різних рівнянь регресії, знайдених для одних і тих же даних. Так, можна спробувати побудувати логарифмічну регресію, параболічну, гіперболічну, лінійну, для кожного рівняння знайти похибки і порівняти їх між собою – те рівняння, яке дасть найменшу похибку, і буде найближчим до наших даних, і буде найбільш прогностичним.

## *5. ВИДИ РІВНЯННЯ РЕГРЕСІЇ*

Ми зазначили, що лінія регресії може бути не лише прямою, але і кривою. Використовують ряд типових кривих регресії, для яких розроблені свої рівняння.

1. Рівняння гіперболічного виду: $Y = a + \frac{b}{X}$

Його коефіцієнти можна визначити з такої системи:

$$\begin{cases} \sum YX = a\sum X + bN \\ \sum YX^2 = a\sum X^2 + b\sum X \end{cases}$$

2. Рівняння логарифмічної параболи: $Y = a + b \cdot \lg X$

Його коефіцієнти можна визначити з такої системи:

$$\begin{cases} \sum Y = aN + b\sum \lg X \\ \sum Y \cdot \lg X = a\sum \lg X + b\sum \lg^2 X \end{cases}$$

3. Рівняння степеневої функції: $Y = aX^b$

Його коефіцієнти можна визначити з такої системи:

$$\begin{cases} \sum \lg Y = N \cdot \lg a + b\sum \lg X \\ \sum \lg Y \cdot \lg X = \lg a\sum \lg X + b\sum \lg^2 X \end{cases}$$

## *6. ЗАГАЛЬНІ ПРИНЦИПИ ВИБОРУ РІВНЯННЯ РЕГРЕСІЇ*

Вибір певної форми регресії на різних етапах дослідження детермінується різними чинниками (табл. 2.3.4).

*Таблиця 2.3.4.*

| **ЕТАП ДОСЛІДЖЕННЯ** | **СПОСІБ ВИБОРУ РІВНЯННЯ РЕГРЕСІЇ** |
|---|---|
| Попередній аналіз | На основі досвіду та інтуїції |
| Поглиблений аналіз | 1. Будуються декілька ліній регресії.<br>2. Обчислюються похибки регресій та перевіряються значимості коефіцієнтів регресії<br>3. Обирається лінія регресії з мінімальною похибкою та максимальною значимістю коефіцієнтів. |

Хоча на перший погляд регресійний аналіз виглядає досить громіздким і довгим, сьогодні, з використанням комп'ютерної техніки, він проводиться буквально за секунди. Однак, ручний варіант обробки даних є просто необхідним для розуміння суті регресійного аналізу, знання його особливостей, переваг та недоліків.

# Розділ III. МЕТОДИ СТАТИСТИЧНОГО ВИСНОВКУ

## § 3.1. СТАТИСТИЧНІ ГІПОТЕЗИ

1. Психологічна і статистична гіпотези.
2. Направлені і ненаправлені статистичні гіпотези.
3. Особливості перевірка статистичної гіпотези.
4. Помилка I роду та рівень значимості статистичного критерію.
5. Помилка II роду та потужність статистичного критерію.

Література:


1. Власов К.П., Власов П.К., Киселева А.А. Методы исследований и организация экспериментов. – Х.: Издательство "Гуманитарный центр", 2002.
2. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976.
3. Готтсданкер Р. Основы психологического эксперимента / Пер. с англ. – М., 1982.
4. Дружинин В.Н. Экспериментальная психология. – СПб.: Питер, 2000.
5. Корнилова Т.В. Экспериментальная психология: Теория и методы: Учебник для вузов. – М.: Аспект Пресс, 2002.
6. Мартин Д. Психологические эксперименты: Секреты механизмов психики. – СПб.: прайм-ЕВРОЗНАК, 2002.
7. Мельников В.М., Ямпольский Л.Т. Введение в экспериментальную психологию личности. – М., 1985.
8. Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО "Речь", 2000.
9. Солсо Р.Л., Джонсон Х.Х., Бил М.К. Экспериментальная психология: практический курс. – СПб.: прайм-ЕВРОЗНАК, 2001.


### *ВСТУП*

До сьогоднішнього дня ми розглядали методи, які дозволяють визначити силу зв'язку між явищами, його значимість та форму. Сьогодні ми починаємо розділ, вивчення якого дозволить нам з'ясовувати також напрям цього зв'язку (аналізувати причинно-наслідкові зв'язки). Щоб відразу знати, що для чого використовувати, давайте побудуємо таблицю (табл. 3.1.1.):

*Таблиця 3.1.1.*

| ЗАДАЧІ | МЕТОДИ ВИРІШЕННЯ |
|---|---|
| Аналіз сили зв'язку та його напряму | Кореляційний аналіз |
| Аналіз форми зв'язку, передбачення змін | Регресійний аналіз |
| Причинно-наслідковий аналіз | Методи статистичного висновку у поєднанні з психологічним експериментом |

Ми ознайомимося з такими методами статистичного висновку:

- ✓ методи виявлення відмінностей в рівні прояву ознаки
- ✓ методи оцінки достовірності зсуву в значення
- ✓ методи виявлення відмінностей у розподілі ознаки
- ✓ багатофункціональні статистичні критерії

Однак, будь-який висновок неможливий без попереднього висунення гіпотези. Саме від того, наскільки правильно та грамотно ми сформулюємо гіпотезу, і залежатиме правильність наших висновків.

## *1. ПСИХОЛОГІЧНА І СТАТИСТИЧНА ГІПОТЕЗА*

Перш за все необхідно з'ясувати, що ж ми розуміємо під гіпотезою.

> *Гіпотеза – це висловлювання, істинність чи хибність якого не відома, але може бути встановлена експериментально;*
>
> *каузальна гіпотеза – висловлювання стосовно причин та пояснень того чи іншого явища.*

Виявляється, що гіпотези можуть бути двох рівнів. Перший рівень – експериментальна психологічна гіпотеза, другий рівень – статистична гіпотеза. Власне, над методами перевірки статистичних гіпотез ми і будемо працювати на наших заняттях. Однак, щоб зрозуміти їх особливість, варто також розглянути психологічні гіпотези, які є предметом експериментальної психології.

**Психологічна гіпотеза** – це дійсно певне висловлювання, яке можна перевірити дослідним шляхом. Це розумне, обґрунтоване та розвинуте припущення. Загалом, вона формулюється як теорема. Формулювання психологічної гіпотези – справжній творчий акт, який вимагає високого рівня знань та досвіду від науковця.

**Статистична гіпотеза** – це просте твердження відносно невідомого параметра. Зазвичай, статистичну гіпотезу позначають буквою Н, і

записують так – (Н: r=0). Твердження в дужках означає, що висунута статистична гіпотеза відносно рівності нулю певного параметра.

Відношення між статистичною гіпотезою та психологічною гіпотезою можна зобразити так (рис. 3.1.1.):

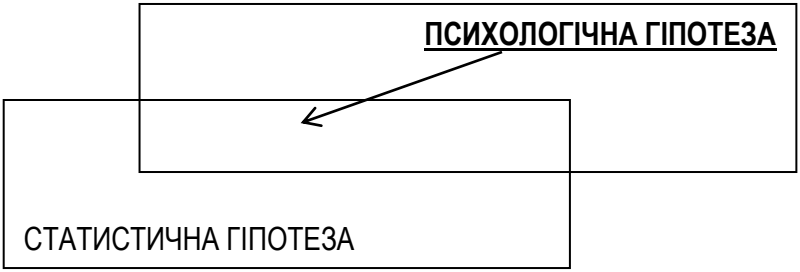



*Рис. 3.1.1. Психологічна та статистична гіпотези*

На рисунку зображено, що психологічна та статистична гіпотези знаходяться у відношенні підпорядкування – статистична гіпотеза формулюється на основі психологічної гіпотези. Водночас, зображені два поля не співпадають – не всі психологічні гіпотези можна перевірити статистичними методами, і не всі статистичні гіпотези мають науковий інтерес для психології.

Які ж бувають статистичні гіпотези?

Статистичні гіпотези поділяють на два типи: нульову та альтернативну. Що це означає?

**Нульова гіпотеза** – це гіпотеза про відсутність відмінностей. Вона позначається $\mathbf{H_0}$ і називається нульовою тому, що містить 0: $X_1$-$X_2$=0, де $X_1$ та $X_2$ – показники, між якими з'ясовують відмінність.

**Альтернативна гіпотеза** – це гіпотеза про існування та значимість відмінностей. Позначається $\mathbf{H_1}$. Альтернативна гіпотеза – це те, що ми в більшості випадків хочемо довести.

Наприклад, для психологічної гіпотези ($П_1$) про зв'язок між творчістю та тривожністю можна сформулювати дві протилежні гіпотези:

$\mathbf{H_0}$ – у досліджуваних відсутній зв'язок між рівнями творчості та тривожності,

**$H_1$** – у досліджуваних існує значимий зв'язок між рівнями творчості та тривожності.

Яку з гіпотез прийняти, залежатиме від результатів дослідження. Приміром, ми підтвердили гіпотезу **$H_1$.**

Наступним кроком треба з'ясувати, що на що впливає – творчість на тривожність, чи тривожність на творчість?

Для цього необхідно вже не просто кореляційне дослідження, а реальний експеримент. Для його проведення можна сформулювати ще дві ПСИХОЛОГІЧНІ гіпотези:

$П_2$ – рівень тривожності впливає на рівень творчості,

$П_3$ – рівень творчості впливає на рівень тривожності.

СТАТИСТИЧНІ ГІПОТЕЗИ для $П_2$:

**$H_0$** – після штучного підвищення рівня тривожності рівень творчості у досліджуваних не зміниться,

**$H_1$** – після штучного підвищення рівня тривожності рівень творчості у досліджуваних значимо зміниться.

СТАТИСТИЧНІ ГІПОТЕЗИ для $П_3$:

**$H_0$** – після навчання досліджуваних творчим стратегіям діяльності рівень тривожності у досліджуваних не зміниться,

**$H_1$** – навчання досліджуваних творчим стратегіям діяльності рівень тривожності у досліджуваних значимо зміниться.

У результаті статистичного аналізу можна зробити висновок, яка з гіпотез підтверджується, а яка відкидається. Загалом структуру такого дослідження можна зобразити так (рис. 3.1.2.):

*Рис. 3.1.2. Процес побудови гіпотез в експериментальному дослідженні*

Отже, з'ясовано, що рівень творчості впливає на рівень тривожності. Однак, у справжніх дослідників з'явиться питання – який напрямок цього впливу (рівень творчості призводить до зниження чи до підвищення рівня тривожності)?

Для з'ясування цього питання розглянемо ще одну класифікацію гіпотез.

Отже, ми стикнулися з проблемою напрямку змін. Вирішити її дозволить ще одна класифікація статистичних гіпотез. Справа у тому, що гіпотези можуть бути направлені і ненаправлені.

Приклад **ненаправленої гіпотези** ми розглянули в попередньому питанні. Це гіпотези типу:

| |
|---|
| **$H_0$**: $X_1$ не відрізняється від $X_2$ |
| **$H_1$**: $X_1$ відрізняється від $X_2$ |

Або (рис. 3.1.3.):

| | | |
|---|---|---|
| $П_3$ | Творчість впливає на тривожність | |
| С | $H_0$: немає змін у тривожності після впливу | $H_1$: є зміни у тривожності після впливу |

*Рис. 3.1.3. Приклад ненаправленої статистичної гіпотези*

Однак, існують і направлені гіпотези, які виглядають так:

| |
|---|
| **$H_0$**: $X_1$ не перевищує $X_2$ |
| **$H_1$**: $X_1$ перевищує $X_2$ |

У такому контексті статистичні гіпотези відносно $П_3$ **виглядали б так** (рис. 3.1.4.)**:**

| | | |
|---|---|---|
| $П_3$ | Творчість впливає на тривожність | |
| С | $H_0$: тривожність після впливу не перевищує тривожність до впливу | $H_1$: тривожність після впливу значимо перевищує тривожність до впливу |

*Рис. 3.1.4. Приклад направленої статистичної гіпотези*

У загальному класифікацію статистичних гіпотез можна зобразити так (рис. 3.1.5.):

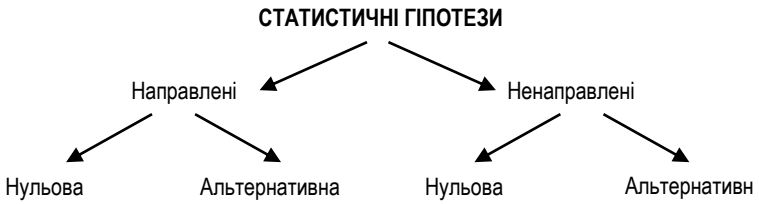



*Рис. 3.1.5. Класифікація статистичних гіпотез*

## *3. ОСОБЛИВОСТІ ПЕРЕВІРКИ СТАТИСТИЧНОЇ ГІПОТЕЗИ*

Перевірка статистичних гіпотез здійснюється з допомогою СТАТИСТИЧНИХ КРИТЕРІЇВ. **Статистичний критерій** – це метод обчислення, який дозволяє прийняти істинну та відхилити хибну статистичну гіпотезу.

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

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

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

Особливістю перевірки статичних гіпотез є те, що рішення дослідника ніколи не приймається з повною впевненістю – він завжди припускає ризик прийняття неправильного висновку.

Цей ризик іманентно властивий будь-якому психологічному дослідженню, і причину його можна знайти в таких поняттях як вибірка та генеральна сукупність (рис. 3.1.6.):

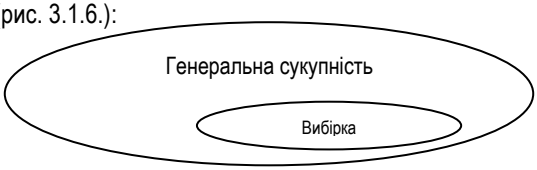



*Рис. 3.1.6. Вибірка та генеральна сукупність*

Хоча дослідник робить все від нього залежне, щоб структура вибірки досліджуваних відображала структуру генеральної сукупності, на яку результати дослідження будуть поширені, повністю цього ніколи досягти не вдається. Це відображається у такому понятті, як рівень статистичної значимості (або ймовірність помилки при перенесенні результатів дослідження вибірки на генеральну сукупність).

Так, якщо ми хочемо з'ясувати вплив творчості на тривожність, ми не можемо дослідити всіх людей – мають бути представлені всі вікові категорії, всі класові прошарки, всі професії, всі культури. Тому дослідник обмежується дослідженням якоїсь однієї категорії, з якої він обирає вибірку – це можуть бути, наприклад, учні молодших класів. Але й тоді виникає ряд проблем – всього в Україні може бути 200 000 учнів, а дослідити ми можемо всього 200!!!

Саме цим і обумовлений той факт, що рішення про існування чи відсутність певної закономірності завжди приймається з урахуванням можливості помилки.

Фактично, визначити цю можливість і покликані методи статистичного висновку, а максимально її знизити – експериментальні методи планування і проведення експерименту.

## *4. ПОМИЛКА І РОДУ ТА РІВЕНЬ ЗНАЧИМОСТІ СТАТИСТИЧНОГО КРИТЕРІЮ*

Ми з'ясували, що в процесі статистичного висновку завжди існує ймовірність помилки. Помилки можна класифікувати. Виділяють помилки двох типів:

- ✓ Помилка І роду.
- ✓ Помилка ІІ роду.

Розглянемо **помилку І роду**.

**Помилка І роду** полягає у відхиленні нульової гіпотези, тоді як вона виявляється правильною.

Ймовірність помилки І роду позначається $\alpha$. Відповідно, ймовірність правильного рішення: 1– $\alpha$. Відповідно, чим менше $\alpha$, тим більша ймовірність правильного рішення.

Історично склалося так, що виділяють три рівні статистичної значимості критерію, який визначається співвідношенням його експериментального та критичного значення (табл. 3.1.2.).

*Таблиця 3.1.2.*

| РІВНІ СТАТИСТИЧНОЇ ЗНАЧИМОСТІ | СПОСІБ ВИЗНАЧЕННЯ |
|---|---|
| 0,05 (мінімальний) | Аналіз співвідношення експериментального (отриманого в результаті обчислень експериментальних результатів) та критичного (отриманого з статистичних таблиць) значень того чи іншого статистичного критерію. |
| 0,01 (середній) | |
| 0,001 (максимальний) | |

## *5. ПОМИЛКА II РОДУ ТА ПОТУЖНІСТЬ СТАТИСТИЧНОГО КРИТЕРІЮ*

Інший варіант помилки – помилка II роду.

**Помилка II роду** полягає у прийнятті нульової гіпотези, тоді як вона виявляється неправильною.

Ймовірність помилки II роду позначають β. А величину 1 – β називають потужністю критерію.

Потужність критерію визначають емпіричним шляхом. Одна й та ж задача може бути вирішена різними критеріями, при цьому виявляється, що деякі критерії можуть виявити відмінності там, де інші їх не бачать. Звісно, можна поставити питання – навіщо використовувати менш потужні критерії?

Проблема в тому, що **на вибір критерію впливають** не лише його потужність, а й ряд інших факторів:

1) простота обчислень,
2) діапазон використання (обсяг вибірки, тип шкали),
3) застосовність до вибірок з неоднаковими обсягами,
4) інформативність результатів (односторонні та двосторонні критерії).

Узагальнюючи вищесказане, можна об'єднати типи помилок у таку таблицю (табл. 3.1.3., за *Гласс Дж., Стенли Дж., 1979*):

*Таблиця 3.1.3.*

| | | РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ | |
|---|---|---|---|
| | | **$H_0$ правильна** | **$H_0$ неправильна** |
| РІШЕННЯ | **Відхилити $H_0$** | Помилка I роду (ймовірність α) | Правильне рішення (ймовірність 1 – β) |
| | **Прийняти $H_0$** | Правильне рішення (ймовірність 1 – α) | Помилка II роду (ймовірність β) |

# § 3.2. ВИЯВЛЕННЯ ВІДМІННОСТЕЙ У РІВНІ ПРОЯВУ ОЗНАКИ

1. Задача зіставлення і порівняння.
2. U-критерій Манна-Уітні.
3. t-критерій Стьюдента для незалежних вибірок.
4. S-критерій Джонкіра.
5. Алгоритм вибору критерію порівняння.

Література


1. Музика О.Л. Курсові роботи з психології: навчально-методичний посібник. – Житомир, 1999.
2. Готтсданкер Р. Основы психологического эксперимента / Пер. с англ. – М., 1982.
3. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976.
4. Мартин Д. Психологические эксперименты: Секреты механизмов психики. – СПб.: прайм-ЕВРОЗНАК, 2002.
5. Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО “Речь”, 2000.
6. Солсо Р.Л., Джонсон Х.Х., Бил М.К. Экспериментальная психология: практический курс. – СПб.: прайм-ЕВРОЗНАК, 2001.


## *1. ЗАДАЧА ЗІСТАВЛЕННЯ І ПОРІВНЯННЯ*

Досить часто перед дослідником у психології стоїть задача виявлення відмінностей між двома, трьома і більше вибірками досліджуваних. Наприклад, задача визначення психологічних особливостей хронічно хворих юнаків у порівнянні із здоровими, правопорушників – із законослухняними громадянами, працівників державних установ – із працівниками приватних фірм тощо.

Крім того, досить часто буває необхідно не виявити відмінності, а навпаки, довести, що ці відмінності відсутні. Наприклад, довести, що студенти з різним рівнем успішності не відрізняються за рівнем інтелекту. Проблема доведення відсутності відмінностей характерна для формуючих експериментів, коли необхідно довести еквівалентність експериментальної та контрольної груп до експерименту:

$$O_1 \to X \to O_2$$
$$O_1^* \to\to\to O_2^*$$

Для того, щоб результати експерименту були валідними, необхідно щоб була виконана умова:

$$O_1 = O_1^*$$

Це і є задача виявлення відмінностей в рівні прояву ознаки. Вирішити цю задачу дозволяють такі статистичні критерії: **U-критерій Манна-Уітні; S-критерій Джонкіра; t-критерій Стьюдента.**

Почнемо наш розгляд із критерію Манна-Уітні.

> *З метою з'ясування необхідності психологічної роботи в лікарні адміністрацією було запрошено на тимчасову роботу психолога. Психолог, щоб довести важливість і необхідність своєї роботи, вирішив побудувати її за експериментальним планом. Він обрав для дослідження дві палати, причому в одній з них він проводив регулярні психологічні консультації та бесіди, а в іншій – нічого не робив. Для діагностики змін у самопочутті пацієнтів він використав опитування лікарів – до початку своєї роботи він опитав їх відносно емоційного стану пацієнтів в обох палатах, і таке ж опитування провів після завершення терміну своєї роботи (через 2 тижні).*

Отже, ми маємо експеримент, в якому:

$$O_1 \to X \to O_2$$
$$O_1^* \to\to\to O_2^*$$

$O_1$ та $O_2$ – опитування лікарів відносно самопочуття експериментальної групи (групи, з якою проводилися психологічні консультації) до та після експерименту;

$O_1^*$ та $O_2^*$ – опитування лікарів відносно самопочуття контрольної групи (групи, з якою не проводилися психологічні консультації) до та після експерименту;

$X$ – експериментальний вплив (психологічні консультації).

Перш за все, необхідною умовою вдалості експерименту є відсутність відмінностей між експериментальною та контрольною групами до експерименту. Адже, навіть, якби лікарі помітили різницю в емоційних станах пацієнтів обох палат, то це могло б бути обумовлено лише тим, що в палаті, в якій працював психолог, були більш оптимістичніші пацієнти. Саме тому перевірка рівності $O_1 = O_1^*$ є обов'язковою! А здійснити її можна з допомогою **U-критерій Манна-Уітні.**

## *2. U-КРИТЕРІЙ МАННА-УІТНІ*

Подивимося, які бали за 20-бальною шкалою поставили пацієнтам лікарі (табл. 3.2.1.)

*Таблиця 3.2.1.*

| $n_1$ та $n_2$ | **ПОЗИТИВНИЙ ЕМОЦІЙНИЙ СТАН** (до експерименту) | |
|---|---|---|
| | ЕКСПЕРИМЕНТАЛЬНА ГРУПА (палата 1) | КОНТРОЛЬНА ГРУПА (палата 2) |
| 1. | 10 | 8 |
| 2. | 15 | 10 |
| 3. | 20 | 19 |
| 4. | 8 | 4 |
| 5. | 6 | 15 |
| 6. | 7 | 20 |
| 7. | 9 | 8 |
| 8. | 8 | 6 |
| 9. | 15 | 7 |
| 10. | 14 | 9 |
| 11. | | 8 |
| 12. | | 15 |
| 13. | | 14 |
| 14. | | 16 |
| 15. | | 17 |
| $\bar{x}$ | 11,2 | 11,7 |

Виходячи з середніх показників можемо припустити, що в обох групах

емоційні стани приблизно однакові. Для того, щоб перевірити це припущення, висунемо дві гіпотези:

1. $H_0$ – відмінності між $\bar{x}_1$ та $\bar{x}_2$ випадкові (або їх взагалі не існує). Отже, наші групи подібні і ми можемо взяти вибрану групу як контрольну.
2. $H_1$ – відмінності між $\bar{x}_1$ та $\bar{x}_2$ достовірні, значимі. Вони можуть бути викликані великим розмахом значень індивідуальних показників, і тоді необхідно змінити або контрольну, або експериментальну групу.

Процедура перевірки висунутих гіпотез така:

1. Дані обох груп об'єднують у нову таблицю, розташувавши їх в порядку зменшення показників. Показники експериментальної групи кодують літерою Е, а контрольної групи – К.
2. Кожному значенню отриманого ряду присвоють його ранг. Якщо в ряду є декілька однакових числових значень, то використовують **правило зв'язаних рангів** (див. питання про коефіцієнт кореляції Спірмена)
3. Знаходять суму рангів (Σ**R**) окремо для експериментальної (Σ**R1**) і контрольної (Σ**R2**) груп (табл. 3.2.2.).

*Таблиця 3.2.2.*

| № | Група | Емоційний стан | Ранг | № | Група | Емоційний стан | Ранг |
|---|---|---|---|---|---|---|---|
| 1. | К | 4 | 1 | 14. | К | 10 | 13,5 |
| 2. | Е | 6 | 2,5 | 15. | Е | 14 | 15,5 |
| 3. | К | 6 | 2,5 | 16. | К | 14 | 15,5 |
| 4. | Е | 7 | 4,5 | 17. | Е | 15 | 18,5 |
| 5. | К | 7 | 4,5 | 18. | Е | 15 | 18,5 |
| 6. | Е | 8 | 8 | 19. | К | 15 | 18,5 |
| 7. | Е | 8 | 8 | 20. | К | 15 | 18,5 |
| 8. | К | 8 | 8 | 21. | К | 16 | 21 |
| 9. | К | 8 | 8 | 22. | К | 17 | 22 |
| 10. | К | 8 | 8 | 23. | К | 19 | 23 |
| 11. | Е | 9 | 11,5 | 24. | Е | 20 | 24,5 |
| 12. | К | 9 | 11,5 | 25. | К | 20 | 24,5 |
| 13. | Е | 10 | 13,5 | | | | |

**ΣR₁** = 125

**ΣR₂** = 200

4. Обчислюємо значення **U**-критерію за формулою:

$$U = n_1 \cdot n_2 + \frac{n_{R\max} \cdot (n_{R\max} + 1)}{2} - R_{\max};$$

де $n_1$ – кількість досліджуваних у 1 групі;

$n_2$ – кількість досліджуваних у 2 групі;

$n_{R\max}$ – кількість досліджуваних у групі із більшою ранговою сумою;

$R_{\max}$ – найбільша рангова сума.

$$U = 10 \cdot 15 + \frac{15 \cdot (15 + 1)}{2} - 200 = 70$$

5. За таблицею 3.2.3. знайдемо критичне значення **U**-критерію для $n_1$ = 10 і $n_2$ = 15. Їх два: $U_{крит}$ = 39

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

*Якщо* $U < U_{крит}$, *то приймається гіпотеза* $H_1$.

У нашому випадку 70 > 39, а тому приймається гіпотеза $H_0$ (табл. 3.2.3.). Отже, відмінності між нашими групами незначні, тому вибрану групу можна взяти як контрольну.

*Таблиця 3.2.3.*

*Критичні значення U-критерію (p = 0,05)*[3]

| $n_1$ | $n_2$ | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** |
| **3** | 1 | 2 | 2 | 3 | 3 | 4 | 4 | 5 | 5 | 6 | 6 | 7 | 7 | 8 |
| **4** | 3 | 4 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 11 | 12 | 13 | 13 |
| **5** | 5 | 6 | 7 | 8 | 9 | 11 | 12 | 13 | 14 | 15 | 17 | 18 | 19 | 20 |
| **6** | 6 | 8 | 10 | 11 | 13 | 14 | 18 | 17 | 19 | 21 | 22 | 24 | 25 | 27 |
| **7** | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 24 | 26 | 28 | 30 | 32 | 34 |
| **8** | 10 | 13 | 15 | 17 | 19 | 22 | 24 | 26 | 23 | 31 | 34 | 36 | 38 | 41 |
| **9** | 12 | 15 | 17 | 20 | 23 | 28 | 28 | 31 | 34 | 37 | 39 | 42 | 45 | 48 |
| **10** | 14 | 17 | 20 | 23 | 26 | 29 | 33 | 36 | 39 | 42 | 45 | 48 | 52 | 55 |
| **11** | 16 | 19 | 23 | 26 | 30 | 33 | 37 | 40 | 44 | 47 | 51 | 55 | 53 | 62 |
| **12** | 18 | 22 | 26 | 29 | 33 | 37 | 41 | 45 | 49 | 53 | 57 | 61 | 65 | 69 |
| **13** | 20 | 24 | 28 | 33 | 37 | 41 | 45 | 50 | 54 | 59 | 63 | 67 | 72 | 76 |
| **14** | 22 | 26 | 31 | 86 | 40 | 45 | 50 | 55 | 59 | 64 | 67 | 74 | 78 | 83 |
| **15** | 24 | 29 | 34 | 39 | 44 | 49 | 54 | 59 | 64 | 70 | 75 | 80 | 85 | 90 |
| **16** | 26 | 31 | 37 | 42 | 47 | 53 | 59 | 64 | 70 | 75 | 81 | 86 | 32 | 98 |
| **17** | 28 | 34 | 39 | 45 | 51 | 57 | 63 | 67 | 75 | 81 | 87 | 93 | 99 | 105 |
| **18** | 30 | 36 | 42 | 48 | 55 | 61 | 67 | 74 | 80 | 86 | 93 | 99 | 106 | 112 |
| **19** | 32 | 38 | 45 | 52 | 58 | 65 | 72 | 78 | 85 | 92 | 99 | 106 | 113 | !19 |
| **20** | 34 | 41 | 48 | 55 | 82 | 69 | 78 | 83 | 90 | 96 | 105 | 112 | 118 | 127 |

## *3. t-КРИТЕРІЙ СТЬЮДЕНТА ДЛЯ НЕЗАЛЕЖНИХ ВИБІРОК*

Після проведення формуючого експерименту також виникає потреба порівняти середні показники емоційного стану пацієнтів у експериментальній та контрольній групах після експерименту:

$$\begin{matrix} O_1 \to X \to O_2 \\ O_1^* \to\to\to O_2^* \end{matrix} \Rightarrow \begin{matrix} O_1 = O_1^* \quad (1) \\ O_2 \neq O_2^* \quad (2) \end{matrix}$$

Відношення (1) ми перевірили з допомогою U-критерію Манна-Уітні. Для перевірки (2) можна знову скористатися U-критерієм Манна-Уітні. Ви вже знаєте як це робити. Але, можливо, чисельність вибірки у вас буде значно

[3] *Мартин Д. Психологические эксперименты: Секреты механизмов психики. – СПб.: прайм-ЕВРОЗНАК, 2002. – С. 453.*

більшою. Тоді краще скористатися t-критерієм Стьюдента як для обгрунтування вибору контрольної групи, так і для перевірки відмінностей між середніми показниками після формуючого експерименту.

Після експерименту позитивний емоційний стан у експериментальній групі зріс до 14,7 балів ( $\bar{X}_{\text{експ.}}$ = 14,7), а в контрольній став рівний 10,6 ( $\bar{X}_{\text{контр.}}$= 10,6).

Якщо виходити з середніх показників, то можна припустити, що позитивний емоційний стан у експериментальній групі зріс не випадково, а завдяки формуючому експерименту.

Для того, щоб перевірити це припущення, висунемо дві гіпотези:

1. $H_0$ – відмінності між $\bar{X}_{\text{експ.}}$ та $\bar{X}_{\text{контр.}}$ випадкові. Експеримент не вдався.
2. $H_1$ – відмінності між $\bar{X}_{\text{експ.}}$ та $\bar{X}_{\text{контр.}}$ достовірні, значимі. Отже, емоційний стан в експериментальній групі зріс завдяки роботі психолога.

Процедура перевірки висунутих гіпотез така[5]:

1. Дані щодо емоційного стану кожного пацієнта в експериментальній і контрольній групах заносять до таблиці (табл. 3.2.4.).

*Таблиця 3.2.4.*

| **Експериментальна група** | | **Контрольна група** | |
|---|---|---|---|
| Пацієнти | Емоційний стан | Пацієнти | Емоційний стан |
| 1. | 15 | 1. | 7 |
| 2. | 18 | 2. | 11 |
| 3. | 20 | 3. | 18 |
| 4. | 13 | 4. | 4 |
| 5. | 9 | 5. | 14 |
| 6. | 11 | 6. | 21 |
| 7. | 13 | 7. | 5 |
| 8. | 10 | 8. | 5 |
| 9. | 18 | 9. | 7 |
| 10. | 20 | 10. | 9 |
| | | 11. | 8 |
| | | 12. | 12 |
| | | 13. | 12 |
| | | 14. | 16 |
| | | 15. | 10 |
| $\bar{X}$ = 14,7 | | $\bar{X}$ = 10,6 | |

2. Для кожної групи обчислюють стантадртні відхилення (див. відповідне питання у темі “Міри мінливості”). У нашому випадку для експериментальної групи **σ** = 4,111, для контрольної - **σ** = 4,997.
3. Обчислюють для кожної групи величину середніх помилок за формулою**:**

$$m = \frac{\sigma}{\sqrt{n}}$$

Для експериментальної групи **m** = 1,300, для контрольної – **m** = 1,290.

4. Знаходять значення t-критерію Стьюдента за формулою[5]:

$$t_{емп} = \frac{\bar{x}_{експ.} - \bar{x}_{контр.}}{\sqrt{m^2_{експ.} + m^2_{контр.}}}$$

У нашому випадку **t** = 2,24
5. Знаходять число ступенів свободи, яке залежить від кількості досліджуваних в обох вибірках: **ν** =$n_1$+$n_2$ – 2. У нашому випадку **ν** = 10+15 – 2 = 23
6. За таблицею 3.2.5. визначають рівень статистичної значимості.

*Якщо $t_{емп} \geq t_{табл}$ – примають гіпотезу $H_1$*

*Якщо $t_{емп} \prec t_{табл}$ – примають гіпотезу $H_0$*

Отже, прийнята гіпотеза $H_1$ на рівні статистичної значимості 5% (p<0,05). Отже, експеримент є вдалим (табл. 3.2.5.).

*Таблиця 3.2.5.*

*Граничні значення t –критерію Стьюдента[4]*

| ν | Рівні значимості | | | ν | Рівні значимості | | |
|---|---|---|---|---|---|---|---|
| | **0,05** | **0,01** | **0,001** | | **0,05** | **0,01** | **0,001** |
| 1 | 12,71 | 63,60 | | 21 | 2,08 | 2,83 | 3,82 |
| 2 | 4,30 | 9,93 | 31,60 | 22 | 2,07 | 2,82 | 3,79 |
| 3 | 3,18 | 5,84 | 12,94 | 23 | 2,07 | 2,81 | 3,77 |
| 4 | 2,78 | 4,60 | 8,61 | 24 | 2,06 | 2,80 | 3,75 |
| 5 | 2,57 | 4,03 | 6,86 | 25 | 2,06 | 2,79 | 3,73 |
| 6 | 2,45 | 3,71 | 5,96 | 26 | 2,06 | 2,78 | 3,71 |
| 7 | 2,37 | 3,50 | 5,41 | 27 | 2,05 | 2,77 | 3,69 |
| 8 | 2,31 | 3,36 | 5,04 | 28 | 2,05 | 2,76 | 3,67 |
| 9 | 2,26 | 3,25 | 4,78 | 29 | 2,04 | 2,76 | 3,66 |

[4] *Музика О.Л. Курсові роботи з психології: навчально-методичний посібник. – Житомир, 1999.*

| $\nu$ | Рівні значимості | | | $\nu$ | Рівні значимості | | |
|---|---|---|---|---|---|---|---|
| | **0,05** | **0,01** | **0,001** | | **0,05** | **0,01** | **0,001** |
| 10 | 2,23 | 3,17 | 4,59 | 30 | 2,04 | 2,75 | 3,65 |
| 11 | 2,20 | 3,11 | 4,44 | 40 | 2,02 | 2,70 | 3,55 |
| 12 | 2,18 | 3,06 | 4,32 | 50 | 2,01 | 2,68 | 3,50 |
| 13 | 2,16 | 3,01 | 4,22 | 60 | 2,00 | 2,66 | 3,46 |
| 14 | 2,15 | 2,98 | 4,14 | 80 | 1,99 | 2,64 | 3,42 |
| 15 | 2,13 | 2,95 | 4,07 | 100 | 1,98 | 2,63 | 3,39 |
| 16 | 2,12 | 2,92 | 4,02 | 120 | 1,98 | 2,62 | 3,37 |
| 17 | 2,11 | 2,90 | 3,97 | 200 | 1,97 | 2,60 | 3,34 |
| 18 | 2,10 | 2,88 | 3,92 | 500 | 1,96 | 2,59 | 3,31 |
| 19 | 2,09 | 2,86 | 3,88 | ∞ | 1,96 | 2,58 | 3,29 |
| 20 | 2,09 | 2,85 | 3,85 | | | | |

## *4. S-КРИТЕРІЙ ДЖОНКІРА*

*Наш психолог-дослідник зміг переконати адміністрацію лікарні в ефективності своєї роботи, однак, постав перед новою проблемою – як краще ефективно організувати свій робочий час? І тоді він вирішує застосувати в своїй роботі групові методи – тоді він зможе одночасно працювати з більшою кількістю людей, і відповідно, принесе більшу користь лікарні. Але як же з'ясувати, чи буде новий груповий метод роботи ефективнішим за індивідуальний?*

*Найпростіший спосіб – в одній з палат проводити індивідуальні консультації, в іншій – групові, і потім порівняти результати опитування лікарів обох палат. Але під час роботи він розробив ще один метод роботи – індивідуально-груповий, і хотів би перевірити одночасно його ефективність також.*

*Тому він для свого дослідження обрав **три палати**: в першій палаті він проводив роботу індивідуальними методом, в другій – груповим, в третій – індивідуально-груповим. І, відповідно, висунув **ПСИХОЛОГІЧНУ** гіпотезу – ефективність роботи зростатиме при переході від індивідуального методу до індивідуально-групового.*

Відповідно, можна побудувати статистичні гіпотези:

1. $H_0$ – тенденція до зростання позитивного емоційного стану при переході від вибірки до вибірки є випадковим.
2. $H_1$ – тенденція до зростання позитивного емоційного стану при переході від першої вибірки до другої є невипадковою та значимою.

Для перевірки цих статистичних гіпотез використовують S-критерій Джонкіра.

Одна з основних умов критерію – однакові обсяги вибірок, що порівнюються.

Давайте подивимося на дані, отримані дослідником через два тижні після роботи з групами (перед цим було з'ясовано, що відмінності між групами досліджуваних до експерименту відсутні – у всіх приблизно однаковий емоційний стан) (табл. 3.2.6):

*Таблиця 3.2.6.*

| | Палата 1: Індивідуальний метод | Плата 2: Груповий метод | Палата 3: Індивідуально-груповий метод |
|---|---|---|---|
| ЕМОЦІЙНИЙ СТАН | 2 | 3 | 5 |
| | 3 | 4 | 6 |
| | 4 | 5 | 5 |
| | 4 | 4 | 6 |
| | 5 | 5 | 4 |
| | 4 | 5 | 7 |
| | 22 | 26 | 33 |

1. Для обчислення критерію необхідно організувати дані у таблицю так, щоб в межах кожної групи показники емоційного стану знаходилися в зростаючому порядку (табл. 3.2.7.).

*Таблиця 3.2.7.*

| | Палата 1: Індивідуальний метод | $S_i$ | Плата 2: Груповий метод | $S_i$ | Палата 3: Індивідуально-груповий метод |
|---|---|---|---|---|---|
| ЕМОЦІЙНИЙ СТАН ДОСЛІДЖУВАНИХ | 2 | 12 | 3 | 6 | 4 |
| | 3 | 11 | 4 | 5 | 5 |
| | 4 | 8 | 4 | 5 | 5 |
| | 4 | 8 | 5 | 3 | 6 |
| | 4 | 8 | 5 | 3 | 6 |
| | 5 | 3 | 5 | 3 | 7 |
| Сума $S_i$ | | 50 | | 25 | |

2. Після того, як дані розташовані в порядку зростання (в межах кожного стовпчика), необхідно підрахувати для кожного індивідуального значення його $S_i$ – кількість показників, які перевищують його в правих стовпчиках.
3. Після цього необхідно знайти суму всіх $S_i$.

$$A = \sum S_i$$

У нашому випадку А=50+25=**75**

4. Необхідно підрахувати максимально можливу кількість всіх перевищень для наших груп:

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

де $c$ — кількість досліджуваних груп,

$n$ - кількість досліджуваних у одній з груп.

У нашому випадку:

$$B = \frac{3(3-1)}{2} \cdot 6^2 = 108$$

5. Тепер можна обчислити емпіричне значення критерію S:

$$S_{емп} = 2 \cdot A - B$$

У нашому випадку:

$$S_{емп} = 2 \cdot 75 - 108 = 42$$

6. На основі порівняння $S_{емп}$ з $S_{крит}$ і роблять висновок про прийняття однієї з статистичних гіпотез:

$$S_{емп} \geq S_{крит} \Rightarrow H_1$$

$$S_{емп} \prec S_{крит} \Rightarrow H_0$$

Давайте з'ясуємо чому рівне $S_{крит}$ (табл. 3.2.8.):

*Таблиця 3.2.8.*

*Критичні значення S-критерію Джонкіра*[5]

| *c* | *n* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* |
| *p=0,05* | | | | | | | | | |
| *3* | 10 | 17 | 24 | 33 | 42 | 53 | 64 | 76 | 88 |
| *4* | 14 | 26 | 38 | 51 | 66 | 82 | 100 | 118 | 138 |
| *5* | 20 | 34 | 51 | 71 | 92 | 115 | 140 | 166 | 194 |
| *6* | 26 | 44 | 67 | 93 | 121 | 151 | 184 | 219 | 256 |

[5] *Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО "Речь", 2000. – С. 323.*

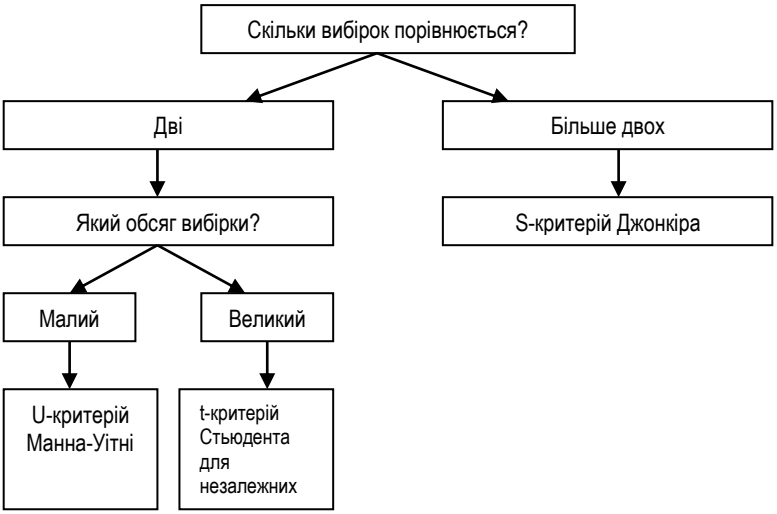
| *p=0,01* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *3* | - | 23 | 32 | 45 | 59 | 74 | 90 | 106 | 124 |
| *4* | 20 | 34 | 50 | 71 | 92 | 115 | 140 | 167 | 195 |
| *5* | 26 | 48 | 72 | 99 | 129 | 162 | 197 | 234 | 274 |
| *6* | 34 | 62 | 94 | 130 | 170 | 213 | 260 | 309 | 361 |

$S_{крит}$ = 42 (*p=0,05*) – робимо висновок про прийняття гіпотези $H_1$.

## *5. АЛГОРИТМ ВИБОРУ КРИТЕРІЮ ПОРІВНЯННЯ*

Ми ознайомилися з декількома статистичними критеріями, призначеними для оцінки відмінностей між різними вибірками. Давайте зобразимо на схемі алгоритм вибору того чи іншого критерію.

**АЛГОРИТМ ВИБОРУ КРИТЕРІЮ №1 (виявлення відмінностей)**



# § 3.3. ОЦІНКА ДОСТОВІРНОСТІ ЗСУВУ ЗНАЧЕННЯ

1. Задача дослідження змін.
2. Критерій знаків (G).
3. t-критерій Стьюдента для залежних вибірок.
4. Критерій $\chi_r^2$ Фрідмана.
5. L-критерій тенденцій Пейджа.
6. Алгоритм вибору критерію оцінки змін.

Література


1. Готтсданкер Р. Основы психологического эксперимента / Пер. с англ. – М., 1982.
2. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976.
3. Мартин Д. Психологические эксперименты: Секреты механизмов психики. – СПб.: прайм-ЕВРОЗНАК, 2002.
4. Музика О.Л. Курсові роботи з психології: навчально-методичний посібник. – Житомир, 1999.
5. Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО "Речь", 2000.
6. Солсо Р.Л., Джонсон Х.Х., Бил М.К. Экспериментальная психология: практический курс. – СПб.: прайм-ЕВРОЗНАК, 2001.


## *1. ЗАДАЧА ДОСЛІДЖЕННЯ ЗМІН*

Перевірку ефективності роботи психолога в лікарні можна було б організувати і з допомогою дещо іншого способу порівняння середніх. У попередньому випадку ми перевіряли дві рівності:

$$[1]\quad \begin{matrix} O_1 \to X \to O_2 \\ O_1^* \to\to\to O_2^* \end{matrix} \Rightarrow \begin{matrix} O_1 = O_1^* & (1) \\ O_2 \neq O_2^* & (2) \end{matrix}$$

Однак, можна організувати роботу іншим чином:

$$[2]\quad \begin{matrix} O_1 \to X \to O_2 \\ O_1^* \to\to\to O_2^* \end{matrix} \Rightarrow \begin{matrix} O_1 \neq O_2 & (1^*) \\ O_1^* = O_2^* & (2^*) \end{matrix}$$

Тобто, з іншого боку, потрібно довести, що показники експериментальної групи до експерименту та після нього відрізняються, а контрольної групи – не відрізняються (змін не сталося).

Таким чином, перед нами задача дослідження змін – у випадку [1] порівнювалися незалежні вибірки, тоді як у випадку [2] відбувається

порівняння вибірок залежних, і для цього можна використати t-критерій Стьюдента для залежних вибірок або критерій знаків.

## *2. КРИТЕРІЙ ЗНАКІВ (G)*

Критерій знаків дає можливість виявити загальний напрямок зсуву певних показників в одній групі після експериментального впливу. Він дозволяє виявити існування зсуву та його напрямок, однак, не дозволяє встановити інтенсивність зсуву. Критерій знаків дуже простий у використанні і належить до методів непараметричної статистики, тому його використання до великих вибірок обмежене. Давайте з допомогою критерію знаків з'ясуємо, чи сталися зміни в контрольній групі після експерименту. Фактично ми перевірятимемо рівність (2*): $O_1^* = O_2^*$

1. Висунемо статистичні гіпотези:

   $H_0$ – відмінності між $\overline{x}_{\text{контр. до}}$ та $\overline{x}_{\text{контр. після}}$ випадкові і не значимі.

   $H_1$ – відмінності між $\overline{x}_{\text{контр. до}}$ та $\overline{x}_{\text{контр. після.}}$ достовірні, значимі.

2. Запишемо дані контрольної групи до та після експерименту до таблиці і з допомогою знаків "+", "–" та "=" позначимо кожну індивідуальну зміну у показниках емоційного стану (табл. 3.3.1):

*Таблиця 3.3.1.*

| Контрольна група (палата 2) | | ЗСУВ |
|---|---|---|
| До експерименту | Після експерименту | |
| 8 | 7 | - |
| 10 | 11 | + |
| 19 | 18 | - |
| 4 | 4 | = |
| 15 | 14 | - |
| 20 | 21 | + |
| 8 | 5 | - |
| 6 | 5 | - |
| 7 | 7 | = |
| 9 | 9 | = |
| 8 | 8 | = |
| 15 | 12 | - |
| 14 | 12 | - |
| 16 | 16 | = |
| 17 | 10 | - |

3. З'ясуємо, що будемо вважати "типовим зсувом" і для цього зсуву визначимо коефіцієнт $G_{емп}$ (як кількість типових зсувів).

   Очевидно, що типовим для нас буде зсув "+" – підвищення позитивного емоційного стану. Тоді:

$$G_{емп} = 2 \qquad N = 15$$

4. За таблицею 3.3.2 знайдемо $G_{кр}$ для N = 15

*Таблиця 3.3.2.*

*Критичні значення критерію знаків*

| N | $G_{кр}$ (p=0,05) | N | $G_{кр}$(p=0,05) | N | $G_{кр}$(p=0,05) |
|---|---|---|---|---|---|
| 6 | 6 | 16 | 13 | 26 | 19 |
| 7 | 7 | 17 | 13 | 27 | 20 |
| 8 | 8 | 18 | 14 | 28 | 20 |
| 9 | 8 | 19 | 15 | 29 | 21 |
| 10 | 9 | 20 | 15 | 30 | 21 |
| 11 | 10 | 21 | 16 | 31 | 22 |
| 12 | 10 | 22 | 17 | 32 | 23 |
| 13 | 11 | 23 | 17 | 33 | 23 |
| 14 | 12 | 24 | 18 | 34 | 24 |
| 15 | 12 | 25 | 18 | 35 | 24 |

$$G_{кр} = 12$$

5. Якщо $G_{емп} >= G_{кр}$, то приймають гіпотезу $H_1$, якщо ж $G_{емп} < G_{кр}$, то приймають гіпотезу $H_0$.

   Таким чином, ми з'ясували, що в контрольній групі не сталося змін після експерименту, і тепер можна приступати до аналізу експериментальної групи.

## *3. t-КРИТЕРІЙ СТЬЮДЕНТА ДЛЯ ЗАЛЕЖНИХ ВИБІРОК*

Отже, нам необхідно з'ясувати, чи сталися в експериментальній групі пацієнтів зміни в емоційному стані після роботи з ними психолога ($O_1 \neq O_2$). Для цього ми скористаємося t-критерієм Стьюдента для залежних вибірок (хоча можна було і використати критерій знаків).

Середні показники говорять, що зміни сталися ( $\overline{x}_{\text{експ. до}}$ = 11,2; $\overline{x}_{\text{експ. після}}$ = 14,7). Однак, чи дійсно це так, ми зможемо стверджувати лише перевіривши статистичні гіпотези:

1. $H_0$ – відмінності між $\overline{x}_{\text{експ. до}}$ та $\overline{x}_{\text{експер. після}}$ випадкові. Отже, експеримент не вдався.
2. $H_1$ – відмінності між $\overline{x}_{\text{експ. до}}$ та . $\overline{x}_{\text{експ. після.}}$ достовірні, значимі.

Для обчислення t-критерію Стьюдента для залежних вибірок використовується формула:

$$t = \frac{\overline{X_d}}{\frac{\sigma_d}{\sqrt{N-1}}},$$

де:

$\overline{X_d}$ – середнє арифметичне різниць значень до та після експерименту,

$\sigma_d$ – стандартне відхилення різниць значень до та після експерименту,

$$\sigma_d = \sqrt{\frac{\sum(X_d - \overline{X}_d)^2}{N-1}}$$

$N$ – обсяг вибірки.

Згрупуємо дані до нової таблиці (табл. 3.3.3.):

*Таблиця 3.3.3.*

| № | До експерименту | Після експерименту | $X_d$ ($X_{після}$–$X_{до}$) | $(X_d - \overline{X}_d)$ | $(X_d - \overline{X}_d)^2$ |
|---|---|---|---|---|---|
| 1 | 10 | 15 | 5 | 1,5 | 2,25 |
| 2 | 15 | 18 | 3 | -0,5 | 0,25 |
| 3 | 20 | 20 | 0 | -3,5 | 12,25 |
| 4 | 8 | 13 | 5 | 1,5 | 2,25 |
| 5 | 6 | 9 | 3 | -0,5 | 0,25 |
| 6 | 7 | 11 | 4 | 0,5 | 0,25 |
| 7 | 9 | 13 | 4 | 0,5 | 0,25 |
| 8 | 8 | 10 | 2 | -1,5 | 2,25 |
| 9 | 15 | 18 | 3 | -0,5 | 0,25 |
| 10 | 14 | 20 | 6 | 2,5 | 6,25 |
| | | | $\overline{X}_d$ =3,5 | | $\sum(X_d - X_{d\text{сер}})^2 = 26,5$ |

$$\sigma_d = \sqrt{\frac{26,5}{10-1}} = 1,72$$

$$t = \frac{\overline{X_d}}{\frac{\sigma_d}{\sqrt{N-1}}} = \frac{3,5}{\frac{1,72}{\sqrt{9}}} = 6,11$$

Знаходимо число ступенів свободи, яке залежить від кількості досліджуваних в обох вибірках:

$\nu$ =N-1=10-1=9

Перевіримо за таблицею граничних значень (див. попереднє питання): $t_{крит}$=4,78, $t_{експер}$=6,11.

Оскільки $t_{експер} > t_{крит}$, то приймаємо гіпотезу $H_1$ ($p \leq 0,001$).

Отже, ми з'ясували, що існують значимі відмінності між показниками експериментальної групи до та після експерименту.

## *4. КРИТЕРІЙ $\chi_r^2$ ФРІДМАНА*

Критерій $\chi_r^2$ Фрідмана використовують для порівняння показників, виміряних в трьох і більше умовах на одній і тій же вибірці. Критерій дозволяє встановити, що величини показників змінюються від умови до умови, але при цьому не вказує на напрямок змін.

Проведемо такий психологічний експеримент:

> *Перевіримо, чи залежить студентська активність на заняттях від способу поведінки викладача. Для цього оберемо одну досліджувану групу з 5 студентів заочного відділення. В якості різних моделей поведінки викладача скористаємося класифікацією лідерів К. Левіна – демократичного, ліберального, авторитарного.*
>
> *Маючи три різні способи поведінки, пропонуємо викладачеві на трьох заняттях із студентами вести себе по-різному: 1 заняття – авторитарно, 2 заняття – демократично, третє заняття – ліберально[6].*

---

[6] *Коректніше було б провести не одну серію, а декілька з різною черговістю стилів управління, і не на одній групі, а на декількох (див. Дружинин В.Н.,2000; Кемпбелл Д.,1996 )*

*Активність студентів будемо визначати з допомогою незалежних експертів, які за 10-бальною шкалою визначатимуть їх загальний емоційний стан, частоту відповідей, зацікавленість у занятті.*

У результаті проведеного дослідження було отримано такі дані (табл. 3.3.4.):

*Таблиця 3.3.4.*

| **Студент** | **Активність** | | |
|---|---|---|---|
| | Авторитарний стиль | Демократичний стиль | Ліберальний стиль |
| А.П. | 4 | 6 | 5 |
| П.Р. | 5 | 10 | 5 |
| В.А. | 4 | 8 | 6 |
| К.А. | 3 | 9 | 4 |
| М.С. | 2 | 6 | 5 |

Для статистичного підтвердження висунутої гіпотези про вплив стилю поведінки викладача на активність студентів необхідно застосувати до отриманих даних $\chi_r^2$ Фрідмана.

1. Сформулюємо статистичні гіпотези:

$H_0$ – відмінності між трьома показниками активності відсутні або випадкові.

$H_1$ – відмінності між трьома показниками активності існують і є значимими.

2. Прорангуємо дані, отримані в різних умовах для кожного досліджуваного студента (перше місце припишемо найнижчій активності) – табл. 3.3.5:

*Таблиця 3.3.5.*

| **Студент** | **Активність** | | |
|---|---|---|---|
| | Авторитарний стиль (ранг) | Демократичний стиль (ранг) | Ліберальний стиль (ранг) |
| А.П. | 1 | 3 | 2 |
| П.Р. | 1,5 | 3 | 1,5 |
| В.А. | 1 | 3 | 2 |
| К.А. | 1 | 3 | 2 |
| М.С. | 1 | 3 | 2 |
| Сума ($T_i$) | 5,5 | 15 | 9,5 |

3. Знайдемо суми рангів по кожній з умов експерименту і загальну суму рангів.

5,5+15+9,5=30

4. Знайдемо розрахункову суму рангів (це крок необхідний лише для перевірки правильності нашого рангування).

$$\sum R_i = n \cdot \frac{c(c+1)}{2},$$

де *c* – кількість умов,

*n* – кількість досліджуваних.

Якщо розрахункова сума рангів не збігається з обчисленою, то ми неправильно прорангували і слід виправити помилку.

5. Знайдемо емпіричне значення $\chi_r^2$ за формулою:

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

6. Знайдемо з допомогою статистичних таблиць критичне значення $\chi_r^2$ і порівняємо його з емпіричним (табл. 3.3.6.). Якщо емпіричне значення $\chi_r^2$ перевищує критичне – приймається гіпотеза $H_1$.

*Таблиця 3.3.6.* [7]

*Критичні значення $\chi_r^2$, c=3, 2<n<9*

| n=2 | | n=3 | | n=4 | | n=5 | |
|---|---|---|---|---|---|---|---|
| $\chi_r^2$ | *p* | $\chi_r^2$ | *p* | $\chi_r^2$ | *p* | $\chi_r^2$ | *p* |
| 0 | 1,000 | 0,000 | 1.000 | 0,0 | 1,000 | 0.0 | 1,000 |
| 1 | 0,833 | 0,667 | 0,944 | 0,5 | 0,931 | 0,4 | 0,954 |
| 3 | 0,500 | 2,000 | 0,528 | 1,5 | 0,653 | 1,2 | 0,691 |
| 4 | 0,167 | 2,667 | 0,361 | 2,0 | 0,431 | 1.6 | 0.522 |
| | | 4.667 | 0,194 | 3.5 | 0,273 | 2,8 | 0,367 |
| | | 6,000 | 0,028 | 4,5 | 0,125 | 3,6 | 0,182 |
| | | | | 6,0 | 0.069 | 4,8 | 0,124 |
| | | | | 6,5 | 0,042 | 5,2 | 0,093 |
| | | | | 8,0 | 0,0046 | 6,4 | 0,039 |
| | | | | | | 7,6 | 0,024 |
| | | | | | | 8,4 | 0,0085 |
| | | | | | | 10,0 | 0,00077 |

[7] *Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО "Речь", 2000. – С. 325.*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *n=6* | | n=7 | | n=8 | | n=9 | |
| $\chi_r^2$ | *p* | $\chi_r^2$ | *p* | $\chi_r^2$ | *p* | $\chi_r^2$ | *p* |
| 0,00 | 1,000 | 0,000 | 1,000 | 0,00 | 1,000 - | 0,000 | 1,000 |
| 0.33 | 0,956 | 0,286 | 0,964 | 0,25 | 0,967 | 0,222 | 0,971 |
| 1.00 | 0.740 | 0,857 | 0,768 | 0,75 | 0.794 | 0,667 | 0,814 |
| 1,33 | 0,570 | 1,143 | 0,620 | 1,00 | 0,654 | 0,889 | 0,865 |
| 2,33 | 0,430 | 2,000 | 0.486 | 1,75 | 0,531 | 1,556 | 0.569 |
| 3,00 | 0,252 | 2,571 | 0,305 | 2,25 | 0,355 | 2,000 | 0,398 |
| 4,00 | 0,184 | 3.429 | 0,237 | 3,00 | 0,285 | 2,667 | 0,328 |
| 4,33 | 0,142 | 3,714 | 0,192 | 3,25 | 0,236 | 2,889 | 0,278 |
| 5,33 | 0,072 | 4.571 | 0,112 | 4,00 | 0,149 | 3,556 | 0,187 |
| 6,33 | 0,052 | 5,429 | 0,085 | 4,75 | 0,120 | 4,222 | 0,154 |
| 7.00 | 0,029 | 6,000 | 0,052 | 5,25 | 0,079 | 4,667 | 0,107 |
| 8,33 | 0,012 | 7,143 | 0,027 | 6,25 | 0,047 | 5,556 | 0,069 |
| 9,00 | 0,0081 | 7,714 | 0,021 | 6,75 | 0,038 | 6,000 | 0,057 |
| 9,33 | 0,0055 | 8,000 | 0,016 | 7,00 | 0,030 | 6,222 | 0,048 |
| 10,33 | 0,0017 | 8.857 | 0,0084 | 7,75 | 0,018 | 6,889 | 0,031 |
| 12,00 | 0,00013 | 10,286 | 0,0036 | 9,00 | 0,0099 | 8,000 | 0,019 |
| | | 10,571 | 0,0027 | 9,25 | 0,0080 | 8,222 | 0,016 |
| | | 11,143 | 0,0012 | 9,75 | 0,0048 | 8,667 | 0,010 |
| | | 12,286 | 0,00032 | 10,75 | 0,0024 | 9,556 | 0,0060 |
| | | 14,000 | 0,000021 | 12,00 | 0,0011 | 10,667 | 0.0035 |
| | | | | 12,25 | 0,00086 | 10,889 | 0,0029 |
| | | | | 13,00 | 0,00026 | 11,556 | 0,0013 |
| | | | | 14,25 | 0,000061 | 12,667 | 0.00066 |
| | | | | 16,00 | 0,0000036 | 13,556 | 0,00035 |
| | | | | | | 14,000 | 0,00020 |
| | | | | | | 14.222 | 0,000097 |
| | | | | | | 14,889 | 0,000054 |
| | | | | | | 16,222 | 0,000011 |
| | | | | | | 18,000 | 0,0000006 |

## *5. L-КРИТЕРІЙ ТЕНДЕНЦІЙ ПЕЙДЖА*

Цей критерій дуже схожий на критерій Фрідмана, однак, він дає змогу встановити не лише існування відмінностей в трьох різних умовах для однієї і тієї ж вибірки, але і з'ясувати напрямок змін – тенденцію.

Обчислимо цей критерій для нашого дослідження стилів поведінки викладачів. Ми з'ясували, що вони по-різному впливають на студентську активність, але ми хочемо піти далі – з'ясувати, як саме. Для цього ми висуваємо гіпотезу: *найсприятливішим для розвитку активності*

*студентів є демократичний стиль управління, найнесприятливіший – авторитарний, а ліберальний – знаходиться посередині.*

Перевіримо цю гіпотезу з допомогою критерію тенденцій Пейджа.

1. Сформулюємо статистичні гіпотези:

   $H_0$ – тенденція до зростання показників активності випадкова,

   $H_1$ – тенденція до зростання показників активності невипадкова і значима.

2. Розмістимо показники активності студентів відповідно до нашої психологічної гіпотези (табл. 3.3.7). Зверніть увагу – дані можуть розташовуватися лише у порядку зростання або спадання.

*Таблиця 3.3.7.*

| **Студент** | **Активність** | | |
|---|---|---|---|
| | Авторитарний стиль | Ліберальний стиль | Демократичний стиль |
| А.П. | 4 | 5 | 6 |
| П.Р. | 5 | 5 | 10 |
| В.А. | 4 | 6 | 8 |
| К.А. | 3 | 4 | 9 |
| М.С. | 2 | 5 | 6 |
| Сума | 18 | 25 | 39 |

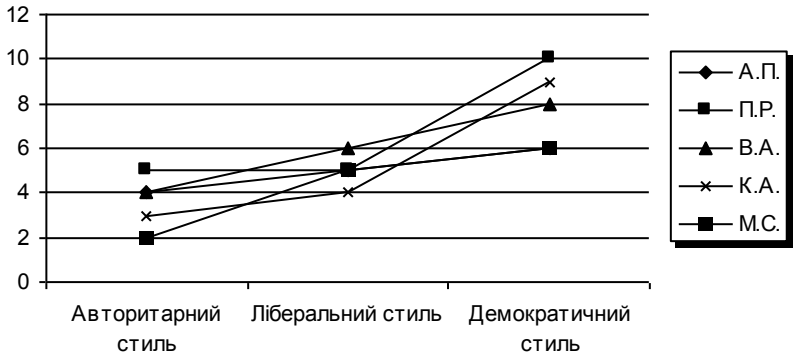



*Рис. 3.3.1. Показники активності студентів при різних стилях поведінки з ними*

3. Прорангуємо показники в межах кожного стовпчика (табл. 3.3.8.):

*Таблиця 3.3.8.*

| **Студент** | **Активність** | | |
|---|---|---|---|
| | Авторитарний стиль (ранг) | Ліберальний стиль (ранг) | Демократичний стиль (ранг) |
| А.П. | 1 | 2 | 3 |
| П.Р. | 1,5 | 1,5 | 3 |
| В.А. | 1 | 2 | 3 |
| К.А. | 1 | 2 | 3 |
| М.С. | 1 | 2 | 3 |
| Сума ($T_i$) | 5,5 | 9,5 | 15 |

4. Знайдемо суми рангів по кожній з умов експерименту і загальну суму рангів.

$$5{,}5+15+9{,}5=30$$

5. Знайдемо розрахункову суму рангів (це крок необхідний лише для перевірки правильності нашого рангування).

$$\sum R_i = n \cdot \frac{c(c+1)}{2},$$

де *c* – кількість умов,

*n* – кількість досліджуваних.

Якщо розрахункова сума рангів не збігається з обчисленою, то ми неправильно прорангували і слід виправити помилку.

6. Знайдемо емпіричне значення $L_{емп}$ за формулою:

$$L_{емп} = \sum (T_i \cdot i)$$

$$L_{емп} = (5{,}5 \cdot 1) + (9{,}5 \cdot 2) + (15 \cdot 3) = 69{,}5$$

7. Знайдемо з допомогою статистичних таблиць критичне значення $L_{крит}$ і порівняємо його з емпіричним (табл. 3.3.9.). Якщо емпіричне значення перевищує критичне, то приймається гіпотеза $H_1$. Інакше – $H_0$.

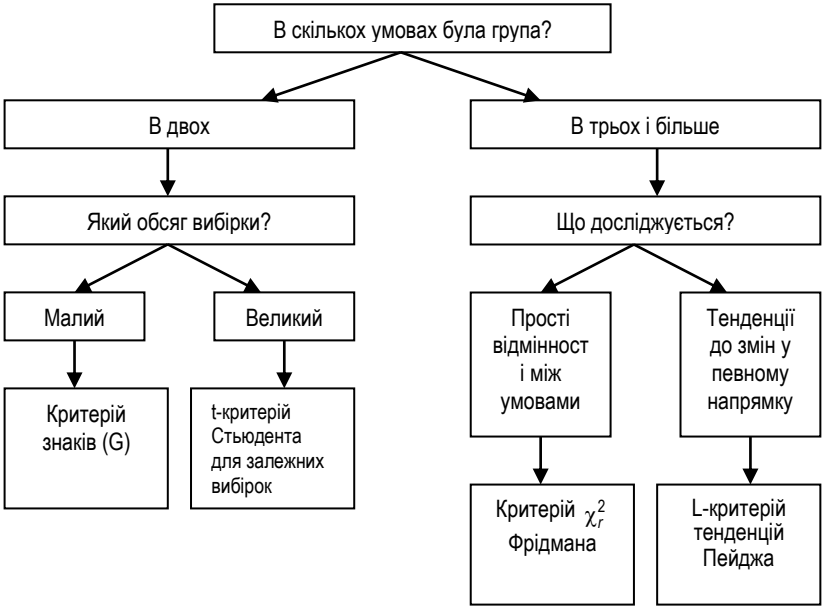
*Таблиця 3.3.9.*

*Критичні значення L*[8]

| 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 |

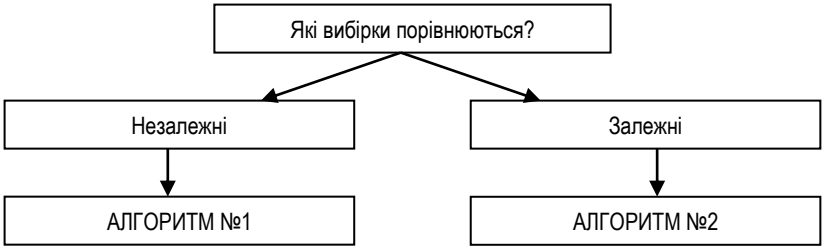
[8] *Там же, с. 327.*

**АЛГОРИТМ ВИБОРУ КРИТЕРІЮ №2 (оцінка змін – зсуву значень)**




Ми з вами ознайомилися з великою кількістю критеріїв. Давайте розглянемо алгоритм, який нам дасть змогу систематизувати уявлення про вже відомі методи статистичного висновку.

**АЛГОРИТМ ВИБОРУ КРИТЕРІЮ №3 (вибір типу критерію)**



# § 3.4. ВИЯВЛЕННЯ ВІДМІННОСТЕЙ У РОЗПОДІЛІ ОЗНАК

1. Задача порівняння розподілу ознак
2. Критерій $\chi^2$-Пірсона
3. Критерій $\lambda$ Колмогорова-Смірнова
4. Алгоритм вибору критерію порівняння розподілів
5. Поняття про багатофункціональні критерії. Критерій $\varphi *$ – кутове перетворення Фішера

Література


1. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976.
2. Мартин Д. Психологические эксперименты: Секреты механизмов психики. – СПб.: прайм-ЕВРОЗНАК, 2002.
3. Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО "Речь", 2000.
4. Солсо Р.Л., Джонсон Х.Х., Бил М.К. Экспериментальная психология: практический курс. – СПб.: прайм-ЕВРОЗНАК, 2001.
5. Справочник по прикладной статистике. В 2-х т. /Под ред. Э. Ллойда, У. Ледермана, Ю.Н. Тюрина. – М.: Финансы и статистика, 1989.


## *1. ЗАДАЧА ПОРІВНЯННЯ РОЗПОДІЛУ ОЗНАК*

Досить часто для підтвердження певної гіпотези необхідно не лише знайти коефіцієнт кореляції, порівняти середні показники чи проаналізувати дисперсії рядів даних, але й порівняти один чи декілька розподілів між собою. За якими ж параметрами можна порівнювати розподіли?

1. Можна аналізувати відмінності дисперсій розподілів.

Наприклад, було висунуто таку гіпотезу: *дорослі, що виховувалися в середовищі, збагаченому стимулами, будуть мати варіативніші показники когнітивної складності, ніж дорослі, що виховувалися в середовищі, збідненому стимулами*.

У результаті проведеного дослідження середовища виховання дітей та їх когнітивної складності в дорослому віці було отримано такі розподіли (рис. 3.4.1.):

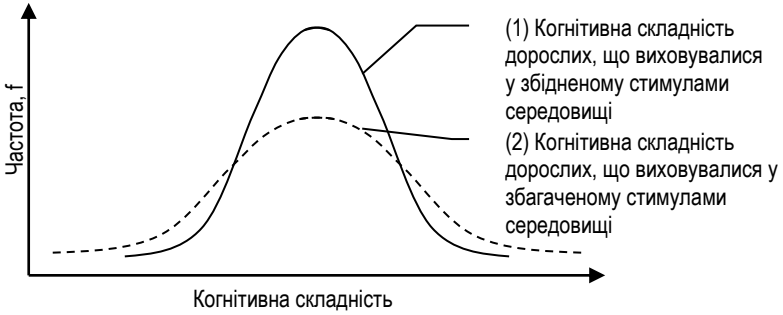



*Рис. 3.4.1. Розподіл когнітивної складності*

З частотного розподілу когнітивної складності видно, що крива 1 менш варіативна, ніж крива 2, що і свідчить про менший діапазон мінливості когнітивної складності у дорослих, що виховувалися в збагаченому стимулами середовищі.

2. Можна аналізувати відмінності в асиметрії розподілів.

Наприклад, при розробці тестових завдань нам необхідно оцінити кожне з них за рівнем складності. Візьмемо для цього два завдання, і дамо їх для вирішення нашим досліджуваним. Вимірявши час вирішення ними цих задач, отримаємо змогу побудувати такі розподіли (рис. 3.4.2):

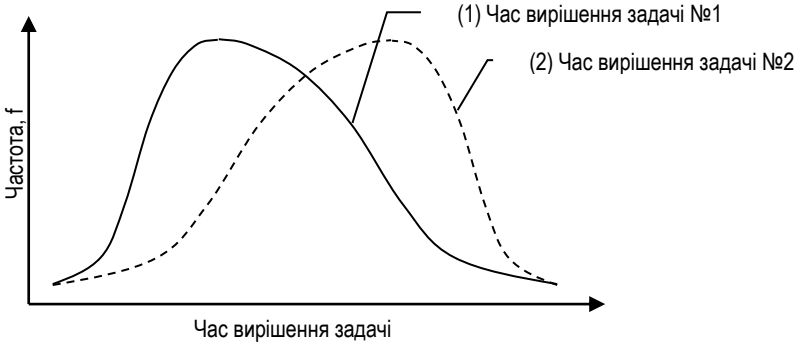



*Рис. 3.4.2. Розподіл часу вирішення задач*

З аналізу розподілів видно, що крива розподілу часу вирішення задачі №1 асиметрична і зсунута вліво, тоді як крива розподілу часу вирішення задачі №2 асиметрична і зсунута вправо. Це свідчить про те, що більшість досліджуваних швидше вирішували задачу №1, ніж задачу №2.

Такий аналіз може, наприклад, дати підстави до включення цих задач до розділів тесту з різними рівнями складності.

3. Можна аналізувати відмінності чи схожості між емпіричним та теоретичним розподілами (нормальним розподілом, рівномірним розподілом тощо).

Така необхідність виникає тоді, коли, наприклад, слід перевірити отримані дані на відповідність критеріям нормального розподілу, щоб до них можна було застосувати методи параметричної статистки. Також досить часто буває необхідно перевірити відповідність емпіричного розподілу, отриманого в результаті дослідження, теоретичному розподілу, передбаченому в межах певної теоретичної моделі досліджуваного явища.

Таким чином, маємо два типи задач, пов'язаних з порівнянням розподілів:

I тип – порівняння емпіричних розподілів (приклад 1 та 2);

II тип – порівняння емпіричного розподілу з теоретичним (приклад 3).

Ці задачі можна вирішити, застосувавши $\chi^2$-критерій Пірсона або $\lambda$-критерій Колмогорова-Смірнова. Далі ми детальніше розглянемо кожен з цих критеріїв з точки зору процедури обчислення, переваг та обмежень, а також алгоритму вибору.

## *2. КРИТЕРІЙ $\chi^2$ -ПІРСОНА*

Цілі застосування цього критерію такі:

1) порівняння емпіричного та теоретичного розподілів;

2) порівняння 2, 3 і більше емпіричних розподілів.

Критерій також має декілька обмежень:

1. Кількість спостережень (обсяг вибірки) повинна були більше 30, інакше буде велика імовірність неточності.
2. Теоретична частота повинна бути не менше 5.

Цей критерій має свою особливість – хоча його можна застосовувати для порівняння емпіричних розподілів між собою, в процесі обчислення

необхідно побудувати до кожного з емпіричних розподілів свій теоретичний розподіл і потім знову ж таки порівняти один емпіричний розподіл та один теоретичний. У зв'язку з цим використання критерію $\chi^2$-Пірсона для порівняння емпіричних розподілів дещо ускладнюється у порівнянні з використанням для цієї мети критерію $\lambda$-Колмогорова-Смірнова.

Наведемо приклад використання критерію $\chi^2$-Пірсона **для порівняння емпіричного та теоретичного розподілів.**

*Нехай, працюючи психологом в архітектурній проектній компанії, ми отримали завдання з'ясувати, яку з двох однакових симетричних стежинок, що ведуть до її центрального входу можна знищити, а яку залишити і модернізувати (ідея прикладу – з Сидоренко Е.В., 2000).*

Описану ситуацію можна графічно зобразити так, як показано на рис. 3.4.3.
Тут А – точка, де дві стежини розходяться, Б – центральний вхід компанії, Л – ліва стежина, П – права стежина.

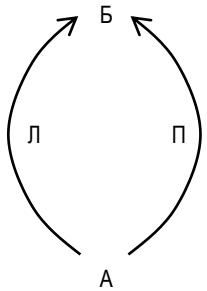


*Рис. 3.4.3. Симетричні стежини*

Для вирішення цієї задачі ми можемо на основі уявлень про білатеральне регулювання висунути гіпотезу – *оскільки люди з переважанням правої ноги в більшості випадків роблять коло проти годинникової стрілки, а люди з переважанням лівої – за годинниковою стрілкою, і більшість людей мають перевагу правої ноги, то більшість з відвідувачів компанії оберуть саме праву доріжку.*[9]

Спробуємо перевірити цю гіпотезу – проведемо спостереження протягом певного часу за 90 відвідувачами компанії, відмічаючи, по якій доріжці вони пішли (по лівій, чи по правій).
Нехай в результаті спостереження ми отримали такі дані: по лівій доріжці пішло 32 особи (Л=32), а по правій – 68 (П=58). На жаль, порівнюючи лише

[9] *Звісно, для того, щоб мати право висунути цю гіпотезу, ми маємо переконатися, що доріжки рівнозначні за своїми характеристикамиу. Інакше – вплив того факту, що на одній з доріжок знаходиться велика калюжа зведе нанівець всю нашу експериментальну роботу.*

кількості людей, ми нічого не можемо сказати однозначно. Ми мусимо перевірити значимість цих відмінностей між доріжками. Для цього ми використаємо критерій $\chi^2$-Пірсона.
Яким же чином? Дані, які ми отримали в результаті спостереження, – це і є наш емпіричний розподіл. Теоретичний розподіл, який ми візьмемо для порівняння, буде рівномірним – дійсно, якби не було різниці між правою та лівою ногами, і доріжки були рівнозначні, то люди розподілилися б між ними рівномірно (50 на 50).

Для перевірки гіпотези про нерівномірність розподілу відвідувачів по доріжках сформулюємо такі статистичні гіпотези:
$H_0$ – емпіричний розподіл відвідувачів значимо не відрізняється від теоретичного рівномірного розподілу,
$H_1$ – емпіричний розподіл відвідувачів достатньо значимо відрізняється від теоретичного рівномірного розподілу.
Перш за все, необхідно обчислити теоретичні частоти для рівномірного розподілу. Обчислення проводиться за такою формулою:

$$f_{теор} = \frac{n}{k},$$

де n – кількість спостережень,
k – кількість розрядів досліджуваної ознаки.
У нашому випадку n=90, k=2, а тому:

$$f_{теор} = \frac{90}{2} = 45$$

Наступний етап роботи – обчислення емпіричного значення критерію $\chi^2$-Пірсона. Для цього складемо таку таблицю (табл. 3.4.1.):

*Таблиця 3.4.1.*

| Розряд | $f_{емп}$ | $f_{теор}$ | $f_{емп} - f_{теор}$ | $(f_{емп} - f_{теор})^2$ | $\frac{(f_{емп} - f_{теор})^2}{f_{теор}}$ |
|---|---|---|---|---|---|
| Ліва доріжка | 32 | 45 | -13 | 169 | 3,756 |
| Права доріжка | 58 | 45 | 13 | 169 | 3,756 |
| **Сума** | **90** | **90** | **0** | | **7,512** |

Обчислюється емпіричне значення критерія $\chi^2$-Пірсона за такою формулою:

$$\chi^2_{емп} = \sum \frac{\left(f_{емп} - f_{теор}\right)^2}{f_{теор}}$$

Як бачимо з таблиці, $\chi^2_{емп}$ =7,512.

Тепер з таблиці критичних значень знайдемо $\chi^2_{крит}$ (табл. 3.4.2).

*Таблиця 3.4.2.*

***Критичні значення критерію $\chi^2$ Пірсона***[10]

| ν | p | | ν | p | | ν | p | |
|---|---|---|---|---|---|---|---|---|
| | 0,05 | 0,01 | | 0,05 | 0,01 | | 0,05 | 0.01 |
| 1 | 3,841 | 6,635 | 35 | 49,802 | 57,342 | 69 | 89,391 | 99,227 |
| 2 | 5,991 | 9,210 | 36 | 50,998 | 58,619 | 70 | 90,631 | 100,425 |
| 3 | 7,815 | 11,345 | 37 | 52,192 | 59,892 | 71 | 91,670 | 101,621 |
| 4 | 9,488 | 13,277 | 38 | 53,384 | 61,162 | 72 | 92,808 | 102,816 |
| 5 | 11,070 | 15,086 | 39 | 54,572 | 62,428 | 73 | 93,945 | 104,010 |
| 6 | 12,592 | 16,812 | 40 | 55,758 | 63,691 | 74 | 95,081 | 105,202 |
| 7 | 14,067 | 18,475 | 41 | 56,942 | 64,950 | 75 | 96,217 | 106,393 |
| 8 | 15,507 | 20,090 | 42 | 58,124 | 66,206 | 76 | 97,351 | 107,582 |
| 9 | 16,919 | 21,666 | 43 | 59,304 | 67,459 | 77 | 98,484 | 108,771 |
| 10 | 18,307 | 23,209 | 44 | 60,481 | 68,709 | 78 | 99,617 | 109,958 |
| 11 | 19,675 | 24,725 | 45 | 61,656 | 69,957 | 79 | 100,749 | 111,144 |
| 12 | 21,026 | 26,217 | 46 | 62,830 | 71,201 | 80 | 101,879 | 112,329 |
| 13 | 22,362 | 27,688 | 47 | 64,001 | 72,443 | 81 | 103,010 | 113,512 |
| 14 | 23,685 | 29,141 | 48 | 65,171 | 73,683 | 82 | 104,139 | 114,695 |
| 15 | 24,996 | 30,578 | 49 | 66,339 | 74,919 | 83 | 105,267 | 115,876 |
| 16 | 26,296 | 32,000 | 50 | 67,505 | 76,154 | 84 | 106,395 | 117,057 |
| 17 | 27,587 | 33,409 | 51 | 68,669 | 77,386 | 85 | 107,522 | 118,236 |
| 18 | 28,869 | 34,805 | 52 | 69,832 | 78,616 | 86 | 108,648 | 119,414 |
| 19 | 30,144 | 36,191 | 53 | 70,993 | 79,843 | 87 | 109,773 | 120,591 |
| 20 | 31,410 | 37,566 | 54 | 72,153 | 81,069 | 88 | 110,898 | 121,767 |
| 21 | 32,671 | 38,932 | 55 | 73,311 | 82,292 | 89 | 112,022 | 122,942 |
| 22 | 33,924 | 40,289 | 56 | 74,468 | 83,513 | 90 | 113,145 | 124,116 |
| 23 | 35,172 | 41,638 | 57 | 75,624 | 84,733 | 91 | 114,268 | 125,289 |
| 24 | 36,415 | 42,980 | 58 | 76,778 | 85,950 | 92 | 115,390 | 126,462 |
| 25 | 37,652 | 44,314 | 59 | 77,931 | 87,166 | 93 | 116,511 | 127,633 |
| 26 | 38,885 | 45,642 | 60 | 79,082 | 88,379 | 94 | 117,632 | 128,803 |
| 27 | 40,113 | 46,963 | 61 | 80,232 | 89,591 | 95 | 118,752 | 129,973 |

[10] *Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО "Речь", 2000. – С. 328.*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 28 | 41,337 | 48,278 | 62 | 81,381 | 90,802 | 96 | 119,871 | 131,141 |
| 29 | 42,557 | 49,588 | 63 | 82,529 | 92.010 | 97 | 120,990 | 132,309 |
| 30 | 43,773 | 50,892 | 64 | 83,675 | 93,217 | 98 | 122,108 | 133,476 |
| 31 | 44,985 | 52,191 | 65 | 84,821 | 94,422 | 99 | 123,225 | 134,642 |
| 32 | 46,194 | 53,486 | 66 | 85,965 | 95,626 | 100 | 124,342 | 135,807 |
| 33 | 47,400 | 54,776 | 67 | 87,108 | 96,828 | | | |
| 34 | 48,602 | 56,061 | 68 | 88,250 | 98,028 | | | |

Для його пошуку необхідно обчислити число ступенів свободи $\nu$ за формулою:

$$\nu = k - 1,$$

де k – кількість розрядів.

У нашому випадку $\nu = 2 - 1 = 1$, і $\chi^2_{крит}$ =6,635 (p=0,01).

*Якщо* $\chi^2_{емп} < \chi^2_{крит}$ *– приймається гіпотеза* $H_0$*,*

*якщо* $\chi^2_{емп} > \chi^2_{крит}$ *– приймається гіпотеза* $H_1$*.*

У нас випадок, коли $\chi^2_{емп} > \chi^2_{крит}$, а отже, ми приймаємо гіпотезу $H_1$ про значимість відмінностей між емпіричним та теоретичним рівномірним розподілами.

Таким чином, ми можемо стверджувати, що дійсно, більшість відвідувачів обирають праву доріжку, а тому ліву можна закрити і перебудувати для інших цілей.

## *3. КРИТЕРІЙ $\lambda$ КОЛМОГОРОВА-СМІРНОВА*

Критерій $\lambda$ Колмогорова-Смірнова має ту ж сферу застосувань, що й критерій $\chi^2$-Пірсона, але з певними відмінностями.

По-перше, якщо з допомогою критерію $\chi^2$-Пірсона можна було порівнювати два і більше емпіричних розподілів між собою, то з допомогою критерію $\lambda$ Колмогорова-Смірнова можна порівнювати лише два емпіричних розподіли між собою.

По-друге, критерій Колмогорова-Смірнова має одну дуже цінну властивість – він дозволяє знайти точку, в якій сума накопичених відмінностей між двома розподілами є найбільшою, та оцінити достовірність цієї відмінності.

Також критерій Колмогорова-Смірнова має свої обмеження. У випадку, коли

аналізується відмінність між двома емпіричними розподілами, вибірка повинна бути **більше 50** досліджуваних, а у випадку, коли аналізуються відмінності між емпіричним та теоретичним розподілами – **більше 5** досліджуваних.

Для ілюстрації алгоритму використання критерію Колмогорова-Смірнова у випадку **порівняння емпіричного та теоретичного розподілів** наведемо такий приклад.

> *Медичний психолог однієї з лікарень вирішив оцінити емоційний стан її пацієнтів. Для цього він використав тест Люшера. Аналізуючи його результати, він виявив цікаву закономірність – більшість пацієнтів на перші місця ставили чорний колір. Перед ним постала проблема – довести, що розміщення чорного кольору на перших позиціях є невипадковим (ідея прикладу – з Сидоренко Е.В., 2000).*

Щоб довести невипадковість такої ситуації, можна використати критерій Колмогорова-Смірнова. Нехай було отримано такі дані (табл. 3.4.3):

*Таблиця 3.4.3.*

| Позиція чорного кольору | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Кількість досліджуваних, що поставили чорний колір на цю позицію | 32 | 20 | 18 | 12 | 13 | 8 | 9 | 8 |

Сформулюємо дві альтернативні гіпотези:

$H_0$ – емпіричний розподіл чорного кольору значимо не відрізняється від теоретичного рівномірного розподілу,

$H_1$ – емпіричний розподіл чорного кольору достатньо значимо відрізняється від теоретичного рівномірного розподілу.

Для перевірки цих гіпотез проведемо ряд обчислень, які оформимо у таблицю 3.4.4:

*Таблиця 3.4.4.*

| Позиція чорного кольору | Емпірична частота $f_{емп}$ | Емпірична частість $f^*_{емп}$ | Накопичена емпірична частість $\sum f^*_{емп}$ | Накопичена теоретична частість $\sum f^*_{теор}$ | Різниця $d$ |
|---|---|---|---|---|---|
| 1 | 32 | 0,267 | 0,267 | 0,125 | 0,142 |
| 2 | 20 | 0,167 | 0,434 | 0,250 | 0,184 |
| 3 | 18 | 0,150 | 0,584 | 0,375 | **0,209** |
| 4 | 12 | 0,100 | 0,684 | 0,500 | 0,184 |
| 5 | 13 | 0,108 | 0,792 | 0,625 | 0,167 |
| 6 | 8 | 0,067 | 0,859 | 0,750 | 0,109 |
| 7 | 9 | 0,075 | 0,934 | 0,875 | 0,059 |
| 8 | 8 | 0,066 | 1,000 | 1,000 | 0,000 |
| Сума | 120 | 1,000 | | | |

Розглянемо елементи таблиці.

**Емпірична частість** ($f^*_{емп}$) обчислюється за такою формулою:

$$f^*_{емп} = \frac{f_{емп}}{n},$$

де $n$ – кількість досліджуваних у вибірці.

**Накопичена емпірична частість** ($\sum f^*_{емп}$) – обчислюється для кожного з розрядів ознаки, починаючи з першого (для кожної позиції чорного кольору) як сума відповідної емпіричної частості та попередньої емпіричної частості. Наприклад, для першого розряду вона залишиться рівна 0,267, для другого розряду (позиції) вона буде рівна (0,267+0,167)=0,434, для третього розряду – (0,434+0,150)=0,584 і так далі.

**Накопичена теоретична частість** ($\sum f^*_{теор}$) – обчислюється аналогічно до накопиченої емпіричної частості, але сумуються вже теоретичні частості. Для визначення теоретичної частості використовують таку формулу:

$$f^*_{теор} = \frac{1}{k},$$

де k – кількість розрядів.

У нашому випадку

$$f^*_{теор} = \frac{1}{8} = 0{,}125$$

**Різниця** ( $d$ ) – обчислюється як модуль різниці відповідних емпіричних та теоретичних частостей.

Після заповнення таблиці, слід в стовпчику "Різниця" знайти найбільшу величину різниці між емпіричними та теоретичними частостями. В нашому випадку це буде число **0,209** в третьому розряді. Це число позначимо як $d_{max}$.

$$d_{\max} = 0{,}209$$

Тепер необхідно знайти критичне значення $d_{кр}$ (табл. 3.4.5.). Для цього варто скористатися таблицею критичних значень для n<100, та наступною формулою, якщо n>100.

$$d_{кр} = \begin{cases} \dfrac{1{,}36}{\sqrt{n}} \Leftarrow (p \le 0{,}05) \\ \dfrac{1{,}63}{\sqrt{n}} \Leftarrow (p \le 0{,}01) \end{cases}$$

*Таблиця 3.4.5.*

*Критичні значення $d_{max}$* [11]

| *n* | Максимальний модуль різниці накопичених частостей $d_{max}$ | | *n* | Максимальний модуль різниці накопичених частостей $d_{max}$ | |
|---|---|---|---|---|---|
| | p=0,05 | p=0,01 | | p=0,05 | p=0,01 |
| 5 | 0,6074 | 0,7279 | 50 | 0,1921 | 0,2302 |
| 10 | 0,4295 | 0,5147 | 60 | 0,1753 | 0,2101 |
| 15 | 0,3507 | 0,4202 | 70 | 0,1623 | 0,1945 |
| 20 | 0,3037 | 0,3639 | 80 | 0,1518 | 0,1820 |
| 25 | 0,2716 | 0,3255 | 90 | 0,1432 | |
| 30 | 0,2480 | 0,2972 | 100 | 0.1358 | |
| 40 | 0,2147 | 0,2574 | >101 | $1{,}36/\sqrt{n}$ | $1{,}63/\sqrt{n}$ |

[11] *Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО "Речь", 2000. – С. 329.*

У нашому випадку:

$$d_{кр} = \begin{cases} \dfrac{1{,}36}{\sqrt{120}} = 0{,}124 \Leftarrow (p \le 0{,}05) \\ \dfrac{1{,}63}{\sqrt{120}} = 0{,}149 \Leftarrow (p \le 0{,}01) \end{cases}$$

*Якщо* $d_{\max} < d_{кр}$ *– приймається гіпотеза* $H_0$,

*якщо* $d_{\max} \ge d_{кр}$ *– приймається гіпотеза* $H_1$

У нашому випадку приймається гіпотеза $H_1$ про значимість відмінності емпіричного розподілу чорного кольору від теоретичного рівномірного розподілу на рівні $p \le 0{,}05$.

Таким чином, психолог-дослідник може зробити висновок, що така ситуація з емоційним станом пацієнтів є невипадковою і вимагає глибшого аналізу.

Для ілюстрації алгоритму використання критерію Колмогорова-Смірнова у випадку **порівняння двох емпіричних розподілів** продовжимо аналіз емоційного стану пацієнтів лікарні.

*Медичний психолог вже знайомої нам лікарні не зупинився на констатації факту негативного емоційного стану пацієнтів. Він вирішив порівняти емоційний стан пацієнтів своєї лікарні та сусідньої. Для цього він знову використав тест Люшера, провівши його в сусідній лікарні. Після отримання даних з'явилася необхідність порівняти два отриманих розподіли (в рідній лікарні та в сусідній).*

Нехай було отримано такі дані (табл. 3.4.6.):

*Таблиця 3.4.6.*

| Позиція чорного кольору | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| Кількість досліджуваних, що поставили чорний колір на цю позицію | Лікарня 1 | 32 | 20 | 18 | 12 | 13 | 8 | 9 | 8 |
| | Лікарня 2 | 18 | 18 | 20 | 16 | 12 | 16 | 10 | 9 |

Сформулюємо дві альтернативні гіпотези:

$H_0$ – емпіричний розподіл чорного кольору в лікарні №1 значимо не відрізняється від емпіричного розподілу в лікарні №2 розподілу,

$H_1$ – емпіричний розподіл чорного кольору в лікарні №1 значимо відрізняється від емпіричного розподілу в лікарні №2 розподілу.

Для перевірки цих гіпотез проведемо ряд обчислень, які оформимо у таблицю 3.4.7:

*Таблиця 3.4.7.*

| Позиція чорного кольору | Емпіричні частоти | | Емпіричні частості | | Накопичені емпіричні частості | | Різниця $d$ |
|---|---|---|---|---|---|---|---|
| | $f_1$ | $f_2$ | $f_1^*$ | $f_2^*$ | $\sum f_1^*$ | $\sum f_1^*$ | |
| 1 | 32 | 18 | 0,267 | 0,151 | 0,267 | 0,151 | 0,116 |
| 2 | 20 | 18 | 0,167 | 0,151 | 0,434 | 0,302 | **0,132** |
| 3 | 18 | 20 | 0,150 | 0,168 | 0,584 | 0,470 | 0,114 |
| 4 | 12 | 16 | 0,100 | 0,134 | 0,684 | 0,604 | 0,080 |
| 5 | 13 | 12 | 0,108 | 0,101 | 0,792 | 0,705 | 0,087 |
| 6 | 8 | 16 | 0,067 | 0,134 | 0,859 | 0,839 | 0,020 |
| 7 | 9 | 10 | 0,075 | 0,084 | 0,934 | 0,923 | 0,011 |
| 8 | 8 | 9 | 0,066 | 0,077 | 1,000 | 1,000 | 0,000 |
| Сума | *120* | *119* | *1,000* | *1,000* | | | |

Як бачимо, максимальна різниця припадає на другий розряд і

$$d_{\max} = 0{,}132$$

Після отримання значення $d_{max}$ можна приступити до обчислення значення $\lambda$.

$$\lambda = d_{\max} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

У нашому випадку:

$$\lambda = 0{,}132 \cdot \sqrt{\frac{120 \cdot 119}{120 + 119}} = 1{,}020$$

Далі з допомогою таблиці критерію $\lambda$ Колмогорова-Смірнова визначають рівень статистичної значимості отриманої величини (табл. 3.4.8.). Виявляється, що

$$p = 0{,}25$$

Очевидно, що такий рівень статистичної значимості недостатній для прийняття гіпотези $H_1$ про значимість відмінностей між двома розподілами чорного кольору. Отже, приймаємо гіпотезу $H_0$.

*Таблиця 3.4.8.*

***Критичні значення критерію λ Коломогорова-Смірнова*** [12]

| λ | λ, останній десятковий знак | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| | p – десяткові знаки (знак "0," опущено) | | | | | | | | | |
| 0,3 | 99999 | 99998 | 99995 | 99991 | 99983 | 99970 | 99949 | 99917 | 99872 | 99807 |
| 0,4 | 99719 | 99603 | 99452 | 99262 | 99027 | 98741 | 98400 | 97998 | 97532 | 96998 |
| 0,5 | 96394 | 95719 | 94969 | 94147 | 93250 | 92282 | 91242 | 90134 | 88960 | 87724 |
| 0,6 | 86428 | 85077 | 83678 | 82225 | 80732 | 79201 | 77636 | 76042 | 74422 | 72781 |
| 0,7 | 71124 | 69453 | 67774 | 66089 | 64402 | 62717 | 61036 | 59363 | 57700 | 56050 |
| 0.8 | 54414 | 52796 | 51197 | 49619 | 48063 | 46532 | 45026 | 43545 | 42093 | 40668 |
| 0,9 | 39273 | 37907 | 36571 | 35266 | 33992 | 32748 | 31536 | 30356 | 29206 | 28087 |
| 1.0 | 27000 | 25943 | 24917 | 23922 | 22957 | 22021 | 21114 | 20236 | 19387 | 18566 |
| 1.1 | 17772 | 17005 | 16264 | 15550 | 14861 | 14196 | 13556 | 12939 | 12345 | 11774 |
| 1,2 | 11225 | 10697 | 10190 | 09703 | 09235 | 08787 | 08357 | 07944 | 07550 | 07171 |
| 1,3 | 06809 | 06463 | 06132 | 05815 | 05513 | 05224 | 04949 | 04686 | 04435 | 04196 |
| 1,4 | 03968 | 03751 | 03545 | 03348 | 03162 | 02984 | 02815 | 02655 | 02503 | 02359 |
| 1.5 | 02222 | 02092 | 01969 | 01852 | 01742 | 01638 | 01539 | 01446 | 01357 | 01274 |
| 1,6 | 01195 | 01121 | 01051 | 00985 | 00922 | 00864 | 00808 | 00756 | 00707 | 00661 |
| 1.7 | 00618 | 00577 | 00539 | 00503 | 00469 | 00438 | 00408 | 00380 | 00354 | 00330 |
| 1,8 | 00307 | 00285 | 00265 | 00247 | 00229 | 00213 | 00198 | 00186 | 00170 | 00158 |
| 1,9 | 00146 | 00136 | 00126 | 00116 | 00108 | 00100 | 00092 | 00085 | 00079 | 00073 |
| 2,0 | 00067 | 00062 | 00057 | 00053 | 00048 | 00045 | 00041 | 00038 | 00035 | 00032 |
| 2,1 | 00030 | 00027 | 00025 | 00023 | 00021 | 00019 | 00018 | 00016 | 00015 | 00014 |
| 2,2 | 00013 | 00011 | 00010 | 00010 | 00009 | 00008 | 00007 | 00007 | 00006 | 00006 |
| 2,3 | 00005 | 00005 | 00004 | 00004 | 00004 | 00003 | 00003 | 00003 | 00002 | 00002 |
| 2,4 | 00002 | 00002 | 00002 | 00001 | 00001 | 00001 | 00001 | 00001 | 00001 | 00001 |

Таким чином, можна зробити висновок, що емоційні стани пацієнтів обох лікарень значимо не відрізняються один від одного, а значить, пошук причин такого стану може стати серйозним джерелом подальших гіпотез та досліджень (навіть ваших власних).

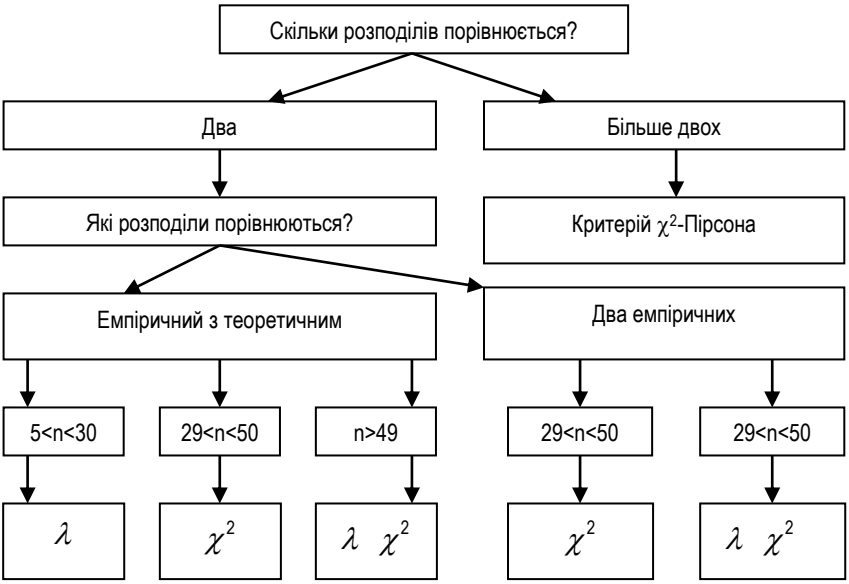
## *4. АЛГОРИТМ ВИБОРУ КРИТЕРІЮ ПОРІВНЯННЯ РОЗПОДІЛІВ*

Для правильного руху по алгоритму вибору критерію слід врахувати два фактори – кількість розподілів та які саме розподіли порівнюються (емпіричні чи емпіричний та теоретичний).

---

[12] *Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО "Речь", 2000. – С. 329.*

**АЛГОРИТМ ВИБОРУ КРИТЕРІЮ №4 (порівняння розподілів)**




## *5. ПОНЯТТЯ ПРО БАГАТОФУНКЦІОНАЛЬНІ КРИТЕРІЇ. КРИТЕРІЙ $\varphi*$ – КУТОВЕ ПЕРЕТВОРЕННЯ ФІШЕРА*

Більшість критеріїв, які ми розглядали, мали вузьке призначення – вони використовувалися лише в межах певного типу шкал (номінативних, порядкових чи інтервальних), або лише для підтвердження певного типу гіпотез (про зв'язок, про відмінності в рівні прояву ознаки, про зсув значення, про відмінності в розподілі ознак).

На відміну від цих уже розглянутих критеріїв, існує декілька інших типів критеріїв, які дають ряд нових можливостей. Ці критерії можуть застосовуватися як самостійно, так і в комплексі з іншими критеріями. Побудовані вони на зіставлені долей, виражених у частках одиниць чи відсотків.

Суть критеріїв – у визначенні того, яка доля спостережень характеризується ефектом, а яка – ні.

Ці критерії можуть бути застосовані:

- до будь-якого типу шкал;
- до залежних і незалежних вибірок;
- до задач про відмінності в рівні прояву ознаки, про зсув значення, про відмінності в розподілі ознак.

Далі ми розглянемо деякі з можливостей **критерію кутового перетворення Фішера** ( $\varphi *$ ) за О.В. Сидоренко .

Критерій $\varphi *$ дає можливість оцінити достовірність відмінностей між відсотковими долями **двох** вибірок, в яких зареєстровано певний ефект. Наприклад, чи достовірною є відмінність між кількістю відмінників на соціологічному (50) та на біологічному (90) факультетах, якщо на соціологічному факультеті навчається 250 студентів, а на біологічному – 327?

Такого роду проблеми допомагає вирішити кутове перетворення Фішера (критерій $\varphi *$ ).

Він має такі можливості:

1. Зіставлення двох вибірок по якісно визначеній ознаці.
2. Зіставлення двох вибірок по кількісно визначеній ознаці.
3. Використання критерію в комплексі з критерієм $\lambda$ Колмогорова-Смірнова для максимально точного результату у випадках, коли є сумніви щодо точки розподілу, яка може використовуватися як критична для поділу досліджуваних на тих, у кого є ефект, і тих, у кого його немає.

Розглянемо детально перший випадок – **зіставлення двох вибірок по якісній ознаці**.

*Для виявлення рівня здатності до прийняття рішень в екстремальних умовах в двох різних підрозділах промислового підприємства, психологом було поставлено експеримент. Керівник за домовленістю із психологом дав кожному з членів підрозділу завдання, яке вимагало роботи в швидкому темпі та нестандартного рішення. У результаті було з'ясовано, що в першому підрозділі з завданням справилося 12 осіб з 20, а в другому – 19 із 26. Перед психологом постала задача – перевірити, чи значимі ці відмінності.*

Для вирішення цієї задачі можна скористатися кутовим перетворенням Фішера. Згрупуємо тримані дані у таблицю 3.4.9:

*Таблиця 3.4.9.*

| Підрозділ | Справилися з задачею | Не справилися з задачею | Всього |
|---|---|---|---|
| №1 | 12 | 8 | 20 |
| №2 | 19 | 7 | 26 |

Сформулюємо дві альтернативні гіпотези:

$H_0$ – доля осіб, які справилися із задачею в першому підрозділі, не більша долі осіб, які справилися з задачею в другому підрозділі,

$H_1$ – доля осіб, які справилися із задачею в першому підрозділі, значимо більша долі осіб, які справилися з задачею в другому підрозділі.

Далі побудуємо таблицю відсоткових долей (табл. 3.4.10):

*Таблиця 3.4.10.*

| Підрозділ | Справилися з задачею | Не справилися з задачею | Всього |
|---|---|---|---|
| №1 | 60% | 40% | 100% |
| №2 | 73% | 27% | 100% |

По статистичній таблиці (див. додаток, табл. 1) визначимо величину кута $\varphi$ для підрозділу 1 та підрозділу 2 (60% та 73%):

$$\varphi_{60\%} = 1{,}772$$

$$\varphi_{73\%} = 2{,}049$$

Далі обчислюється емпіричне значення $\varphi *$.

$$\varphi^*_{емп} = \left|\varphi_1 - \varphi_2\right| \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

Де:

$n_1$ – загальна кількість досліджуваних у першій групі;

$n_2$ – загальна кількість досліджуваних у другій групі;

У нашому випадку:

$$\varphi^*_{емп} = \left|1{,}772 - 2{,}049\right| \cdot \sqrt{\frac{20 \cdot 26}{20 + 26}} = 0{,}277 \cdot 3{,}314 = 1{,}083$$

Звенемося до статистичної таблиці цього критерію. За нею можна точно визначити рівень статистичної значимості отриманого числа (табл. 3.4.11).

*Таблиця 3.4.11*

*Рівні статистичної значимості критерію φ* Фішера*[13]

| p | p (останній десятковий знак) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 0,00 | 2,91 | 2,81 | 2,70 | 2.62 | 2,55 | 2,49 | 2.44 | 2,39 | 2,35 | |
| **0,01** | **2,31** | 2,28 | 2,25 | 2,22 | 2,19 | 2,16 | 2,14 | 2,11 | 2,09 | 2,07 |
| 0,02 | 2,05 | 2,03 | 2,01 | 1,99 | 1,97 | 1,96 | 1,94 | 1.92 | 1,91 | 1,89 |
| 0,03 | 1,88 | 1,86 | 1,85 | 1,84 | 1,82 | 1,81 | 1,80 | 1,79 | 1,77 | 1,76 |
| 0,04 | 1,75 | 1,74 | 1,73 | 1,72 | 1,71 | 1,70 | 1,68 | 1,67 | 1,66 | 1,65 |
| **0,05** | **1,64** | 1,64 | 1,63 | 1,62 | 1,61 | 1,60 | 1,59 | 1,58 | 1,57 | 1,56 |
| 0,06 | 1,56 | 1,55 | 1,54 | 1,53 | 1,52 | 1,52 | 1,51 | 1,50 | 1,49 | 1,48 |
| 0,07 | 1,48 | 1,47 | 1,46 | 1,46 | 1,45 | 1,44 | 1,43 | 1,43 | 1,42 | 1,41 |
| 0,08 | 1,41 | 1,40 | 1,39 | 1,39 | 1,38 | 1,37 | 1,37 | 1,36 | 1,36 | 1,35 |
| 0,09 | 1,34 | 1,34 | 1,33 | 1,32 | 1,32 | 1,31 | 1,31 | 1,30 | 1,30 | 1,29 |
| 0,10 | 1,29 | | | | | | | | | |

Потрібно у цій таблиці знайти значення $\varphi^*_{емп}$ – у відповідній стрічці ви знайтеде значення “р” з точністю до сотих, а у відповідному стовпчику – третій знак після коми, тобто, уточните значення “р” до тисячних.

Виявляється, що для нашого значення $\varphi^*_{емп}$ p>0,10 – це свідчить про низьку статистичну значимість відмінностей між групами, а значить, відхиляємо гіпотезу $H_1$ і приймаємо гіпотезу $H_0$ про відсутність відмінностей між двома вибірками.

На початку ми згадували про можливість використання кутового перетворення Фішера для зіставлення вибірок по **кількісно визначеній ознаці**. Як це зробити? В принципі хід роботи той же – лише потрібно звести отримані дані до таблиці, розсортувавши результати дослідження по рівнях “є ефект” чи “немає ефекту”. Однак, досить часто сам розподіл по рівнях є проблемою. Якщо задано 20-бальну шкалу, то коли вважати, що ефект є – при 10, 11 чи 15 балах? А коли ефекту немає? У багатьох випадках для цього можна використати критерій λ Колмогорова-Смірнова. У попередній лекції ми згадували про одну з можливостей цього критерію – можливість знайти точку в розподілах, в якій накопичуються максимальні розходження. Цією точкою є $d_{max}$. Якщо тепер застосувати до отриманих даних критерій λ Колмогорова-Смірнова, то можна знайти цю точку, а потім використати її як критичну для розміщення результатів досліджуваних в розряди “є ефект” або “немає ефекту”.

---

[13] *Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО “Речь”, 2000. – С. 332.*

## ЗАВДАННЯ ДЛЯ САМОКОНТРОЛЮ

***Тестові завдання:***

1. Військові звання – це:
   а) номінативна шкала;
   б) шкала відношень;
   в) порядкова шкала.
2. Поняття, яке відображає рівень відповідності структури вибірки структурі генеральної сукупності – це:
   а) адекватність;
   б) значимість;
   в) репрезентативність.
3. Показник, який ділить упорядкований ряд даних на дві однакових частини – це:
   а) мода;
   б) медіана;
   в) середнє квадратичне.
4. Кореляційний аналіз дозволяє виявити:
   а) форму зв'язку між змінними;
   б) силу та напрям зв'язку між змінними;
   в) вплив змінних одна на одну.
5. Регресійний аналіз використовується для:
   а) пошуку сили зв'язку між змінними;
   б) пошуку залежних та незалежних змінних;
   в) пошуку форми зв'язку між змінними.
6. Методи статистичного висновку спрямовані на:
   а) виявлення зв'язку між явищами;
   б) пошук причинно-наслідкового зв'язку;
   в) пошук форми і напрямку зв'язку.
7. Для виявленні відмінностей у рівні прояву ознаки використовують:
   а) критерій Манна-Уітні;
   б) критерій Пейджа;
   в) критерій знаків.
8. Критерій Манна-Уітні використовують для:
   а) перевірки гіпотези про відмінність у рівні прояву ознаки;
   б) перевірки гіпотези про зсув значень;
   в) перевірки гіпотези про зв'язок між змінними.
9. Критерій Джонкіра використовують для:
   а) перевірки гіпотези про відмінність у рівні прояву ознаки;
   б) перевірки гіпотези про зсув значень;
   в) перевірки гіпотези про зв'язок між змінними.
10. Для аналізу тенденцій використовують:

а) критерій Пейджа;
б) критерій Пірсона;
в) критерій знаків.

11. Номери телефонів – це:
    а) номінативна шкала;
    б) порядкова шкала;
    в) інтервальна шкала.
12. Поняття, яке відображає ймовірність помилки при перенесенні результатів, отриманих на вибірці, на генеральну сукупність – це:
    а) репрезентативність вибірки;
    б) статистична значимість;
    в) похибка репрезентативності.
13. Міра центральної тенденції, яка враховує всі значення ряду – це:
    а) мода;
    б) медіана;
    в) середнє арифметичне.
14. Для графічного виявлення зв'язку між змінними використовують:
    а) діаграму розсіювання;
    б) діаграму кореляцій;
    в) гістограму кореляцій.
15. Для підбору адекватної лінії регресії використовують:
    а) критерій Антонова-Мірошнікова;
    б) обчислення похибок;
    в) критерій знаків.
16. Помилка I роду при перевірці статистичної гіпотези полягає у тому, що:
    а) відхиляють нульову гіпотезу, тоді як вона є істинна;
    б) приймають нульову гіпотезу, тоді як вона є хибною;
    в) приймають альтернативну гіпотезу, тоді як вона є хибною.
17. Для виявленні відмінностей у рівні прояву ознаки використовують:
    а) критерій Джонкіра;
    б) критерій Пейджа;
    в) критерій Фрідмана.
18. Критерій Фрідмана використовують для:
    а) перевірки гіпотези про відмінність у рівні прояву ознаки;
    б) перевірки гіпотези про зсув значень;
    в) перевірки гіпотези про зв'язок між змінними.
19. Критерій Пейджа використовують для:
    а) перевірки гіпотези про відмінність у рівні прояву ознаки;
    б) перевірки гіпотези про зсув значень;
    в) перевірки гіпотези про зв'язок між змінними.
20. Для виявлення статистичної значимості критерію Пейджа використовують:
    а) статистичні таблиці;

б) критерій Колмогорова-Смірнова;
в) коефіцієнт Пейджа.

21. Показники рівня інтелекту за тестом Айзенка – це:
    а) інтервальна шкала;
    б) шкала відношень;
    в) номінативна шкала.
22. Показник, який у ряді даних зустрічається найчастіше – це:
    а) мода;
    б) медіана;
    в) середнє арифметичне.
23. Міра центральної тенденції, яка використовується в межах усіх вимірювальних шкал – це:
    а) медіана;
    б) мода;
    в) середнє арифметичне.
24. Для пошуку зв'язку між змінними, виміряними у порядковій шкалі використовують:
    а) коефіцієнт кореляції Спірмена;
    б) коефіцієнт кореляції Пірсона;
    в) критерій Колмогорова-Смірнова.
25. Якщо в регресійному аналізі одна змінна називається залежною, а інша незалежною, то:
    а) це просто умовні позначення;
    б) це дійсно означає, що одна змінна впливає на іншу;
    в) це означає, що необхідно провести додатковий аналіз.
26. Помилка II роду при перевірці статистичної гіпотези полягає у тому, що:
    а) відхиляють нульову гіпотезу, тоді як вона є істинна;
    б) приймають нульову гіпотезу, тоді як вона є хибною;
    в) приймають альтернативну гіпотезу, тоді як вона є хибною.
27. Для оцінки достовірності зсуву значень використовують:
    а) критерій знаків;
    б) критерій Манна-Уітні;
    в) критерій Джонкіра.
28. Критерій Стьюдента для залежних змінних використовують для:
    а) перевірки гіпотези про відмінність у рівні прояву ознаки;
    б) перевірки гіпотези про зсув значень;
    в) перевірки гіпотези про зв'язок між змінними.
29. Критерій знаків використовують для:
    а) перевірки гіпотези про відмінність у рівні прояву ознаки;
    б) перевірки гіпотези про зсув значень;
    в) перевірки гіпотези про зв'язок між змінними.
30. Для виявлення статистичної значимості критерію знаків використовують:

а) статистичні таблиці;
б) критерій Колмогорова-Смірнова;
в) знаковий коефіцієнт Стьюдента.

***Теоретичні питання:***

1. Змінні та їх вимірювання. Вимірювальні шкали
2. Генеральна сукупність та вибірка. Репрезентативність вибірки
3. Статистична значимість
4. Залежність між змінними. Залежні та незалежні змінні
5. Статистичні таблиці та графіки
6. Параметрична та непараметрична статистика
7. Призначення мір центральної тенденції
8. Мода та медіана
9. Середнє арифметичне: обчислення та властивості
10. Вибір міри центральної тенденції
11. Призначення мір мінливості
12. Розмах та коефіцієнт осциляції
13. Дисперсія: обчислення та властивості
14. Стандартне відхилення та коефіцієнт варіації
15. Загальне поняття про зв'язок. Статистичний та функціональний зв'язок.
16. Пошук зв'язку з допомогою діаграм розсіювання.
17. Пошук зв'язку з допомогою коефіцієнта кореляції. Коефіцієнт кореляції Пірсона
18. Область значень коефіцієнта кореляції, рівень його статистичної значимості та обмеження
19. Коефіцієнт кореляції Спірмена, коефіцієнт кореляції φ
20. Множинна кореляція
21. Задачі регресійного аналізу
22. Визначення коефіцієнтів регресії
23. Обчислення значень залежної змінної на основі регресії
24. Обчислення похибки рівняння регресії
25. Види рівнянь регресії
26. Загальні принципи вибору рівняння регресії
27. Психологічна і статистична гіпотези
28. Направлені і ненаправлені статистичні гіпотези
29. Помилка I роду та рівень значимості статистичного критерію
30. Помилка II роду та потужність статистичного критерію
31. Задача зіставлення і порівняння
32. U-критерій Манна-Уітні
33. t-критерій Стьюдента для незалежних вибірок
34. S-критерій Джонкіра
35. Алгоритм вибору критерію порівняння

36. Задача дослідження змін
37. Критерій знаків (G)
38. t-критерій Стьюдента для залежних вибірок
39. Критерій $\chi_r^2$ Фрідмана
40. L-критерій тенденцій Пейджа
41. Алгоритм вибору критерію оцінки змін

***Творчі завдання:***

Розробіть структуру дослідження, результати якого можна було б обробити з допомогою запропонованих нижче методів. Змоделюйте його результати, обробіть їх та зробіть висновки.

Результати роботи:

- Структура дослідження (тема, гіпотеза, план дослідження)
- Числові результати дослідження
- Обробка результатів (вручну)
- Висновки про підтвердження гіпотези

| | |
|---|---|
| – Мода та медіана<br>– Середнє арифметичне<br>– Розмах та коефіцієнт осциляції<br>– Дисперсія<br>– Стандартне відхилення та коефіцієнт варіації<br>– Коефіцієнт кореляції Спірмена<br>– Коефіцієнт кореляції φ<br>– Множинна кореляція<br>Коефіцієнт кореляції Пірсона | – Регресійний аналіз<br>– U-критерій Манна-Уітні<br>– t-критерій Стьюдента для незалежних вибірок<br>– S-критерій Джонкіра<br>– Критерій знаків (G)<br>– Критерій $\chi_r^2$ Фрідмана<br>– L-критерій тенденцій Пейджа<br>t-критерій Стьюдента для залежних вибірок |

# Розділ IV. ДИСПЕРСІЙНИЙ АНАЛІЗ

## § 4.1. ФАКТОРНІ ЕКСПЕРИМЕНТИ

1. Поняття про факторні гіпотези.
2. Типи факторних планів.
3. Результати факторних експериментів.

Рекомендована література:


1. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976.
2. Готтсданкер Р. Основы психологического эксперимента / Пер. с англ. – М., 1982.
3. Дружинин В.Н. Экспериментальная психология. – СПб.: Питер, 2000.
4. Корнилова Т.В. Экспериментальная психология: Теория и методы: Учебник для вузов. – М.: Аспект Пресс, 2002.
5. Мартин Д. Психологические эксперименты: Секреты механизмов психики. – СПб.: прайм-ЕВРОЗНАК, 2002.
6. Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО "Речь", 2000.
7. Солсо Р.Л., Джонсон Х.Х., Бил М.К. Экспериментальная психология: практический курс. – СПб.: прайм-ЕВРОЗНАК, 2001.
8. Экспериментальная психология / Ред.-сост. П. Фресс, Ж. Пиаже. – М.: Прогресс, 1966.
9. Хок Р.Р. 40 исследований, которые потрясли психологию. Секреты выдающихся экспериментов. – СП.: прайм-ЕВРОЗНАК, 2003.


### *1. ПОНЯТТЯ ПРО ФАКТОРНІ ГІПОТЕЗИ*

Інколи достатньо складно, а то й неможливо сформулювати гіпотезу про вплив одного явища на інше. Причина цього – системна та комплексна взаємодія різних чинників, які накладаючись один на другого, утворюють складну і багатогранну картину оточуючої нас реальності. Тому простих гіпотез про вплив одного явища на інше буває мало для розуміння цієї реальності – необхідні складніші підходи.

Одним із таких підходів є **факторний експеримент.** Для статистичної обробки результатів факторного експерименту і використовують **дисперсійний аналіз**.

Отже, факторний експеримент використовують у випадках, коли необхідно перевірити складні гіпотези про зв'язок між змінними.
Загальний вигляд складної гіпотези такий:

**Якщо $A_1$, $A_2$, ..., $A_n$ – то B**

При цьому, **$A_1$, $A_2$, ..., $A_n$** – незалежні змінні, а **B** – залежна змінна. Якщо користуватися термінологією факторних експериментів, то незалежні змінні варто називати **факторними ознаками,** або просто **факторами**, а залежні – **результуючими ознаками**.
Існують декілька типів факторних гіпотез, які перевіряються з допомогою дисперсійного аналізу:

- ☑ перший тип – гіпотези про окремий вплив кожного із факторів,
- ☑ другий тип – гіпотези про взаємодію факторів (коли присутність одного з факторів змінює вплив іншого).

При цьому, відношення між факторами можуть бути різноманітні. В.М. Дружинін *(Дружинин В.Н., 2000, с. 118)* називає такі:

- відношення диз'юнкції;
- відношення кон'юнкції;
- лінійна залежність;
- адитивність;
- мультиплікативність тощо.

## *2. ТИПИ ФАКТОРНИХ ПЛАНІВ*

Суть факторного планування експерименту полягає в тому, що усі рівні незалежних змінних (факторів) потрібно комбінувати один з одним. При цьому необхідно, щоб кількість експериментальних груп була рівна кількості комбінацій рівнів незалежних змінних.
**Першою основою** для виділення типів факторних планів є кількість незалежних змінних (факторів). Відповідно, виділяють такі типи планів:

- план для **двох** факторів;
- план для **трьох** факторів.

Використання одночасно більше трьох факторів для планування експерименту є недоцільним, оскільки навіть обробка результатів трьохфакторного експерименту є достатньо громіздкою.
**Другою основою** для типології є кількість рівнів кожного з факторів:

- план для **двох** рівнів фактора;
- план для **трьох** рівнів фактора.

Узагальнюючи наведені типології, можна побудувати таблицю 4.1.1:

*Таблиця 4.1.1.*

*Типи факторних планів для двох факторів*

| | | Фактор 1 | |
|---|---|---|---|
| | | 2 рівні | 3 рівні |
| Фактор 2 | 2 рівні | **2x2** | 3x2 |
| | 3 рівні | 2x3 | **3x3** |

Найтиповішими є плани типу 2x2 та 3x3. Факторний план 2x2 дозволяє виявити ефект впливу двох факторів на результуючу ознаку, а план 3x3, крім того, дає можливість визначити форму зв'язку між змінними.

Очевидно, що два рівні фактора можна позначити, наприклад, як "низький" та "високий", або як "є ефект" та "немає ефекту". План 2x2 має такий вигляд (табл. 4.1.2):

*Таблиця 4.1.2.*

*Факторний план для двох факторів та двох рівнів кожного з факторів*

| | | Фактор 1 | |
|---|---|---|---|
| | | Є вплив | Немає впливу |
| Фактор 2 | Є вплив | **$A_1$** | **$A_2$** |
| | Немає впливу | **$A_3$** | **$A_4$** |

У таблиці $A_1$, $A_2$, $A_3$, $A_4$ – досліджувані групи, кожна з яких потрапляє в свої умови впливу фактора. Група $A_1$ перебуває в умовах впливу обох факторів, група $A_2$ – в умовах впливу лише другого фактора, група $A_3$ – в умовах впливу лише першого фактора, група $A_4$ – в умовах відсутності впливу обох факторів.

Факторний план 3x3 також достатньо часто використовують, особливо тоді, коли фактори мають складнішу градацію, ніж "є вплив – немає впливу". В цьому випадку факторний план має такий вигляд (табл. 4.1.3):

*Таблиця 4.1.3.*

| | | Фактор 1 | | |
|---|---|---|---|---|
| | | Є вплив | Середній рівень впливу | Немає впливу |
| Фактор 2 | Є вплив | **$A_1$** | **$A_2$** | **$A_3$** |
| | Середній рівень впливу | **$A_4$** | **$A_5$** | **$A_6$** |
| | Немає впливу | **$A_7$** | **$A_8$** | **$A_9$** |

Відповідно, тут також **$A_1$-$A_9$** – досліджувані групи, кожна з яких перебуває у своїх експериментальних умовах.

Для трьох факторів можливі різноманітні варіанти, типу 2х3х2, 2х3х3, 3х2х3 тощо. Однак, на думку В.М. Дружиніна, найтиповішим є план типу **2х2х2** *(Дружинин В.Н., 2000, с. 118)*.

Як бачимо, із зростанням кількості факторів або рівнів факторів хоча б на один, план експерименту значно ускладнюється. Дійсно, навіть план 3х3 вже достатньо складно реалізувати – необхідно 9 незалежних вибірок досліджуваних.

У зв'язку з цим можна виділити ще одну, **третю, основу для класифікації** типів факторних планів – спосіб підбору досліджуваних груп:

- **збалансований** план;
- **ротаційний** план.

У **збалансованих планах** кожній комбінації рівнів факторів піддається окрема досліджувана група. недолік цього плану – велика кількість груп, і відповідно, ускладнення процедури експерименту.

У **ротаційних планах** цей недолік знімається – тут всі досліджувані піддаються впливу всіх комбінацій рівнів факторів. Однак, на зміну одному недоліку приходить інший – з'являється ефект тренування. Дійсно, при такій побудові експерименту кожен попередній експериментальний вплив залишає свій слід у психіці досліджуваного, відповідно, впливаючи на його поведінку в нових умовах. Накопичуючись, ці впливи можуть повністю спотворити результати дослідження. Про способи контролю цих сторонніх чинників можна довідатися з рекомендованої літератури.

Розглянемо, у якій формі постають перед дослідником результати факторних експериментів, та як можна їх інтерпретувати.

Перше, що повинен зробити дослідник після проведення факторного експерименту, це представити його результати графічно. Це дасть змогу уявити тип взаємодії між факторами, а також спланувати напрямок подальшого дисперсійного аналізу та його доцільність. Дійсно, якщо графічне зображення говорить про відсутність будь-яких ефектів, навряд чи варто проводити подальший аналіз. З іншого боку, видимий графічно ефект насправді може бути лише видимістю, – тому для підтвердження дисперсійний аналіз просто необхідний.

Найцікавішою для аналізу Т.В. Корнилова називає взаємодію двох факторів (*Корнилова Т.В., 2002, с. 231*). Вона називає два типи взаємодії факторів, залежно від кількості факторів у експерименті: взаємодію першого порядку (між двома факторами) та взаємодію другого порядку (між трьома факторами).

Виділяють також такі види взаємодії: 1) **нульова**; 2) **перехресна**; 3) **розбіжна**.

Для ілюстрації видів взаємодії наведемо приклад дослідження.

> *Для дослідження впливу рівня інтелектуального розвитку та стану здоров'я на статус дитини у групі було проведено дослідження за факторним планом. Визначався рівень інтелекту, загальний стан здоров'я та соціометричний статус дітей*[14].

У результаті аналізу результатів такого дослідження можна отримати різні форми взаємодії факторів.

**Нульова взаємодія** має такий вигляд (рис. 4.1.1.):

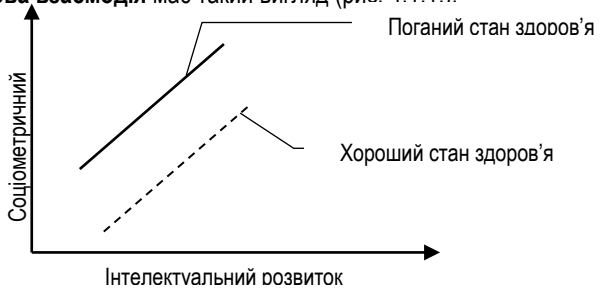



*Рис. 4.1.1. Приклад нульової взаємодії між факторами*

[14] *Ідея прикладу запозичена у В.М. Дружиніна (2000, с. 121).*

При нульовій взаємодії між факторами величина впливу другого фактора на результуючу ознаку однакова при всіх рівнях першого фактора – між факторами взаємодії немає.

**Перехресна взаємодія** має такий вигляд (рис. 4.1.2.):

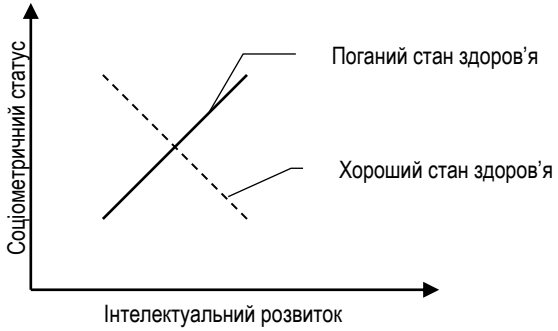



*Рис. 4.1.2. Приклад перехресної взаємодії між фактором інтелекту та фактором здоров'я*

При перехресній взаємодії є можливість наочно спостерігати, в яких взаємозв'язках знаходяться фактори між собою. У зображеному випадку виявляється, що статус дитини в групі підвищується при низькому інтелекті та хорошому здоров'ї, а також при високому рівні інтелекту та поганому здоров'ї.

**Розбіжна взаємодія** зображається так (рис. 4.1.3.):

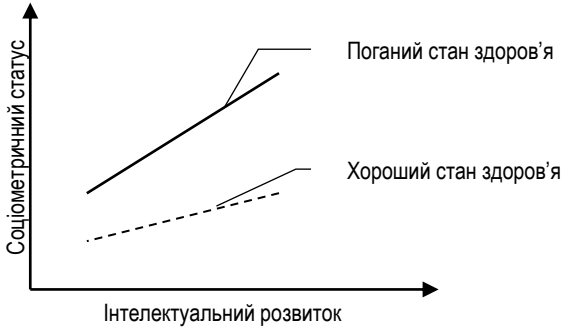



*Рис. 4.1.3. Приклад розбіжної взаємодії між фактором інтелекту та фактором здоров'я*

У випадку зображеної розбіжної взаємодії фактор “здоров’я” підсилює дію фактора “інтелект” – при наявності високого інтелекту погане здоров’я підвищує статус дитини.
Існують також інші варіанти взаємодії між факторами, однак, у всіх випадках слід пам’ятати одне – якою б наочною не була графічна картина, проведення дисперсійного аналізу є обов’язковим.

# § 4.2. ДВОФАКТОРНИЙ ДИСПЕРСІЙНИЙ АНАЛІЗ ДЛЯ НЕЗАЛЕЖНИХ ВИБІРОК

1. Задача двофакторного дисперсійного аналізу.
2. Обмеження двофакторного дисперійсного аналізу.
3. Підготовка даних до дисперсійного аналізу.
4. Проведення двофакторного дисперійсного аналізу. Критерій Фішера F.

Рекомендована література:


1. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976.
2. Готтсданкер Р. Основы психологического эксперимента / Пер. с англ. – М., 1982.
3. Мартин Д. Психологические эксперименты: Секреты механизмов психики. – СПб.: прайм-ЕВРОЗНАК, 2002.
4. Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО “Речь”, 2000.
5. Солсо Р.Л., Джонсон Х.Х., Бил М.К. Экспериментальная психология: практический курс. – СПб.: прайм-ЕВРОЗНАК, 2001.


## *1. ЗАДАЧА ДВОФАКТОРНОГО ДИСПЕРСІЙНОГО АНАЛІЗУ*

При аналізі результатів двофакторного експерименту варто послуговуватися методом двофакторного дисперсійного аналізу. Він дозволяє не лише оцінити вплив кожного із факторів на результуючу ознаку, але й оцінити їх взаємодію. Необхідність врахування взаємодії факторів можна побачити на таких прикладах.

**Приклад 1.**

*Дослідження впливу часового обмеження на швидкість вирішення задач може не вдатися без врахування фактору*

*інтелекту. Може, наприклад, виявитися, що при скороченні часу досліджувані з високим рівнем інтелекту будуть покращувати результати, а досліджувані з низьким – навпаки, будуть їх знижувати.*

**Приклад 2.**

*В умовах жорсткого контролю за результатами діяльності внутрішньо мотивовані працівники можуть знижувати ефективність своєї діяльності, а мотивовані зовнішньо – підвищувати її.*

**Приклад 3.**

*Залежність між мотивацією досягнення успіху та рівнем тривожності може бути прямою для жінок, та оберненою для чоловіків.*

## *2. ОБМЕЖЕННЯ ДВОФАКТОРНОГО ДИСПЕРІЙСНОГО АНАЛІЗУ*

Для того, щоб дисперсійний аналіз було проведено правильно, необхідно враховувати певні обмеження та рекомендації:

1. Кожен фактор повинен мати не менше двох градацій.
2. В кожній комірці дисперсійного комплексу повинно бути не менше двох спостережень (про дисперсійний комплекс див. далі).
3. Дисперсійний комплекс повинен бути врівноваженим.
4. Дисперсійний комплекс повинен бути симетричним (кожній градації першого фактора повинна відповідати однакова кількість градацій другого фактора).
5. Фактори повинні бути незалежними (між ними не повинно існувати значимого кореляційного зв'язку).
6. Результуюча ознака повинна бути розподілена нормально.

## *3. ПІДГОТОВКА ДАНИХ ДО ДИСПЕРСІЙНОГО АНАЛІЗУ*

Підготовка даних до дисперсійного аналізу полягає у перевірці всіх вище наведених обмежень. Більшість із них можна перевірити без складних математичних обчислень, крім 5 та 6 застереження.

Розглянемо **спосіб перевірки 5 застереження** стосовно незалежності

факторів. Дійсно, ми ж говоримо про фактори як про незалежні змінні – такими вони й повинні бути між собою. Тут варто відмітити, що взаємовплив факторів – це не те ж саме, що зв'язок чи залежність між ними. Фактори можуть впливати один на одного, але не бути залежними, але водночас, якщо фактори пов'язані між собою, то їх вплив настільки ускладнюється, що провести адекватний дисперсійний аналіз стає неможливим.

Нагадаємо приклад:

> *В умовах жорсткого контролю за результатами діяльності внутрішньо мотивовані працівники можуть знижувати ефективність своєї діяльності, а мотивовані зовнішньо – підвищувати її.*

Тут на робітників діє два фактори – жорсткість контролю та їх мотивація. Результуючою ознакою є ефективність праці. Дослідник, варіюючи ступінь жорсткості контролю та спостерігаючи за реакцією внутрішньо мотивованих та зовнішньо мотивованих робітників, отримує інформацію про ефективність їх праці. Однак, якщо виявиться, що фактори "жорсткість контролю" та "мотивація" не є незалежними? Якщо сама варіація жорсткості контролю впливає не лише на продуктивність праці, але й змінює мотивацію робітників?

Для того, щоб уникнути таких ситуацій **необхідно ще до початку експерименту спробувати з допомогою кореляційного дослідження знайти рівень зв'язку між незалежними факторами. Якщо виявиться, що зв'язок не значимий, можна приступати до проведення факторного експерименту**. Інакше – результати експерименту можуть бути абсолютно неочікуваними.

Стосовно **6 застереження,** то варто відмітити, що дисперсійний аналіз є методом параметричним, а тому перевірка нормальності розподілу результуючої ознаки є обов'язковою (хоча багато дослідників цим нехтують, коли мають справу з великими вибірками).

Для перевірки нормальності розподілу необхідно знайти його асиметрію та ексцес і порівняти із критичними значеннями *(Сидоренко Е.В., 2000, с. 232)*.

**а) обчислення емпіричних значень асиметрії ($A_{емп}$) та ексцесу ($E_{емп}$)**

$$A_{емп} = \frac{\sum \left(x_i - \bar{x}\right)^3}{n \cdot s_x^3}$$

$$E_{емп} = \frac{\sum\left(x_i - \bar{x}\right)^4}{n \cdot s_x^4} - 3,$$

де n – кількість спостережень,
$s_x$ – стандартне відхилення;

**б) обчислення критичних значень асиметрії ($A_{крит}$) та ексцесу ($E_{крит}$)**

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

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

де n – кількість спостережень;

**в) зіставлення емпіричних та критичних значень асиметрії та ексцесу**

Розподіл результуючої ознаки є нормальним, якщо виконуються такі умови:

$$\begin{cases} A_{емп} < A_{крит} \\ E_{емп} < E_{крит} \end{cases}$$

## *4. ПРОВЕДЕННЯ ДВОФАКТОРНОГО ДИСПЕРСІЙНОГО АНАЛІЗУ. КРИТЕРІЙ ФІШЕРА F*

Якщо виявиться, що результуюча ознака розподілена нормально, можна розпочинати аналіз. Прикладом факторного плану типу 2х2 може бути дослідження, яке описують Р. Солсо, Х. Джонсон, К. Біл (*Солсо Р., Джонсон Х., Бил М., 2001, с. 89* ).

*Тривалий час існувало дві гіпотези відносно поведінки у ситуації змушеного захисту поглядів, протилежних власним.*

*Перша – гіпотеза підкріплення. Відповідно до неї, чим більшою*

*буде винагорода за захист поглядів, тим з більшою готовністю досліджуваний змінить свої власні погляди.*

*Друга – гіпотеза дисонансу. Відповідно до теорії когнітивного дисонансу, навпаки, чим меншою буде винагорода, тим швидше досліджуваний змінить свої погляди.*

*Для перевірки обох цих гіпотез було побудовано модель експерименту 2х2. Студентам коледжу запропонували написати твір, спрямований на підтримку закону про обмеження студентського самоврядування (це суперечило їх власним поглядам). Було обрано 4 групи досліджуваних. Першій та другій групам просто наказували написати твір на цю тему, а третій та четвертій запропонували на вибір – писати чи не писати. Крім того, першій та третій групі заплатили по $0,50, а другій та четвертій – по $2,50. Таким чином, було два фактори – наявність ситуації вибору (є та немає) та рівень винагороди (низький та високий). Результуючою ознакою був ступінь позитивного ставлення до закону.*

Спробуємо змоделювати процедуру цього експерименту (табл. 4.2.1):

*Таблиця 4.2.1*

| | | **Фактор А:** ВИНАГОРОДА | |
|---|---|---|---|
| | | Мала | Велика |
| **Фактор В:** ВИБІР | Немає вибору | **Група 1** | **Група 2** |
| | Є вибір | **Група 3** | **Група 4** |

Нехай отримані такі первинні результати по кожному з досліджуваних (всього 16 досліджуваних, по 4 на кожну експериментальну групу) (табл. 4.2.2):

*Таблиця 4.2.2.*

| | | **Фактор А:** ВИНАГОРОДА | |
|---|---|---|---|
| | | Мала | Велика |
| **Фактор В:** ВИБІР | Немає вибору | **10, 9, 5,6** | **6, 4, 2, 3** |
| | Є вибір | **5, 4, 2, 4** | **8, 6, 5, 6** |

*Можна висунути такі **гіпотезу**: між факторами "ВИБІР" та "ВИНАГОРОДА" буде спостерігатися взаємодія: при великій винагороді ставлення стає позитивним тоді, коли відсутній вибір, а при малій винагороді ставлення стає позитивним тоді, коли є можливість вибору.*

Для подальшого аналізу важливо дотримуватися чіткого алгоритму.

**1 крок:** ПОБУДОВА ДИСПЕРСІЙНОГО КОМПЛЕКСУ
Дисперсійний комплекс має такий вигляд (табл. 4.2.3):

*Таблиця 4.2.3.*

| | | **Фактор А:** ВИНАГОРОДА | | $T_B$ |
|---|---|---|---|---|
| | | Мала | Велика | |
| **Фактор В:** ВИБІР | Немає вибору | **30** | **15** | $T_{B1}$**=45** |
| | Є вибір | **15** | **25** | $T_{B2}$**=40** |
| | $T_A$ | $T_{A1}$**=45** | $T_{A2}$**=40** | **T=85** |

Тут $T_A$ та $T_B$ – відповідні суми показників зміни поглядів.

**2 крок:** ПЕРЕВІРКА НАЗВАНИХ ВИЩЕ ОБМЕЖЕНЬ ДИСПЕРСІЙНОГО АНАЛІЗУ
Вимога симетричності дисперсійного комплексу полягає в тому, щоб у кожній його комірці була однакова кількість спостережень. У нашому випадку на кожну комірку припадає по 4 спостереження.

**3 крок:** ГРАФІЧНЕ ЗОБРАЖЕННЯ РЕЗУЛЬТАТІВ ЕКСПЕРИМЕНТУ
Можна зображати сумарні величини по кожній з комірок, а можна – середні арифметичні (рис. 4.2.1.).

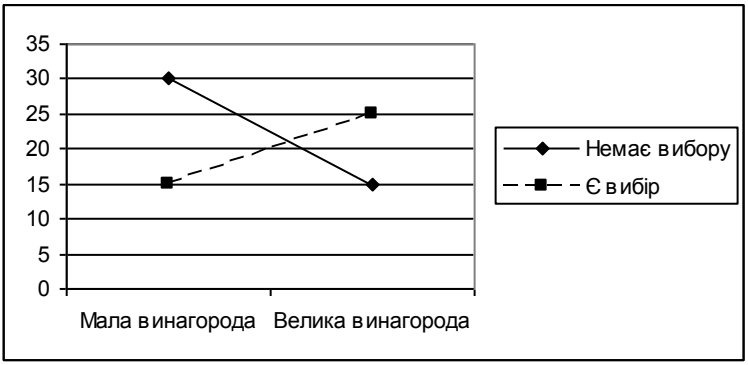



*Рис. 4.2.1. Графічне зображення взаємодії факторів*

**4 крок:** ПОБУДОВА КОМПЛЕКТУ СТАТИСТИЧНИХ ГІПОТЕЗ

КОМПЛЕКТ №1: гіпотези про вплив фактора А

| | |
|---|---|
| **$H_0$** | Відмінності у зміні ставлення до закону, обумовлені фактором А, є не більше вираженими, ніж випадкові відмінності. |
| **$H_1$** | Відмінності у зміні ставлення до закону, обумовлені фактором А, є більше вираженими, ніж випадкові відмінності. |

КОМПЛЕКТ №2: гіпотези про вплив фактора В

| | |
|---|---|
| **$H_0$** | Відмінності у зміні ставлення до закону, обумовлені фактором В, є не більше вираженими, ніж випадкові відмінності. |
| **$H_1$** | Відмінності у зміні ставлення до закону, обумовлені фактором В, є більше вираженими, ніж випадкові відмінності. |

КОМПЛЕКТ №3: гіпотези про вплив взаємодії факторів А та В

| | |
|---|---|
| **$H_0$** | Вплив фактора А на ставлення до закону однаковий при різних градаціях фактора В, і навпаки. |
| **$H_1$** | Вплив фактора А на ставлення до закону різний при різних градаціях фактора В, і навпаки. |

**5 крок:** ПІДГОТОВКА ДО ОБЧИСЛЕНЬ. ОБЧИСЛЕННЯ ОКРЕМИХ ВЕЛИЧИН

Цей крок та сам подальший алгоритм обчислення запропоновано О.В. Сидоренко *(Сидоренко Е.В., 2000, с. 250-251),* див. табл. 4.2.4.

*Таблиця 4.2.4.*

| ВЕЛИЧИНА | НАЗВА | ЗНАЧЕННЯ |
|---|---|---|
| $T_A$ | суми по градаціях фактора А | **$T_{A1}$=45; $T_{A2}$=40** |
| $\sum T_A^2$ | суми квадратів цих сум | $\sum T_A^2 = 45^2 + 40^2 = 3625$ |
| $T_B$ | суми по градаціях фактора В | **$T_{B1}$=45; $T_{B2}$=40** |
| $\sum T_B^2$ | суми квадратів цих сум | $\sum T_B^2 = 45^2 + 40^2 = 3625$ |
| $T_{AB}$ | суми по комірках | **30, 15, 15, 25** |
| $\sum T_{AB}^2$ | суми квадратів цих сум | $\sum T_{AB}^2 = 30^2 + 15^2 + 15^2 + 25^2 = 1975$ |
| n | кількість досліджуваних у кожній комірці | **n=4** |
| a | кількість градацій фактора А | **a=2** |
| b | кількість градацій фактора В | **b=2** |

| | | |
|---|---|---|
| N | загальна кількість індивідуальних значень | **N=16** |
| $\sum x_i$ | сума всіх індивідуальних значень | $\sum x_i = 85$ |
| $\frac{1}{N}\cdot\left(\sum x_i\right)^2$ | константа | $\frac{1}{N}\cdot\left(\sum x_i\right)^2 = 451{,}563$ |
| $\sum x_i^2$ | сума квадратів індивідуальних значень | $\sum x_i^2 = 529$ |

**6 крок:** ОБЧИСЛЕННЯ КРИТЕРІЮ F ФІШЕРА

Обчислюватися будуть такі величини (табл. 4.2.5):

$SS_A$ варіативність ознаки, обумовлена фактором А

$SS_B$ варіативність ознаки, обумовлена фактором В

$SS_{AB}$ варіативність ознаки, обумовлена взаємодією факторів А та В

$SS_{заг}$ загальна варіативність ознаки

$SS_{вип}$ варіативність ознаки, обумовлена випадковими факторами

df число ступенів свободи

MS середній квадрат

$F_{емп}$ емпіричне значення критерію F

$F_{крит}$ критичне значення критерію F

*Таблиця 4.2.5.*

| **Величина** | **Формула для обчислення** | **Результат обчислення** |
|---|---|---|
| $SS_A$ | $SS_A = \frac{1}{nb}\cdot\sum T_A^2 - \frac{1}{N}\cdot\left(\sum x_i\right)^2$ | $SS_A$=1,562 |
| $SS_B$ | $SS_A = \frac{1}{na}\cdot\sum T_B^2 - \frac{1}{N}\cdot\left(\sum x_i\right)^2$ | $SS_A$=1,562 |
| $SS_{AB}$ | $SS_{AB} = \frac{1}{n}\cdot\sum T_{AB}^2 - \frac{1}{N}\cdot\left(\sum x_i\right)^2 - SS_A - SS_B$ | $SS_{AB}$=39,063 |
| $SS_{заг}$ | $SS_{заг} = \sum x_i^2 - \frac{1}{N}\cdot\left(\sum x_i\right)^2$ | $SS_{заг}$=77,437 |
| $SS_{вип}$ | $SS_{вип} = SS_{заг} - SS_A - SS_B - SS_{AB}$ | $SS_{вип}$=35,250 |
| df | $df_A = a-1;\quad df_B = b-1;\quad df_{AB} = df_A \cdot df_B;$<br>$df_{заг} = N-1;\quad df_{вип} = df_{заг} - df_A - df_B - df_{AB}$ | $df_A$=1<br>$df_B$=1 |

| | | |
|---|---|---|
| | | $df_{AB}$=1<br>$df_{заг}$=15<br>$df_{вип}$=12 |
| MS | $MS_A = \frac{SS_A}{df_A}$; $MS_B = \frac{SS_B}{df_B}$;<br>$MS_{AB} = \frac{SS_{AB}}{df_{AB}}$; $MS_{вип} = \frac{SS_{вип}}{df_{вип}}$ | $MS_A$=1,562<br>$MS_B$=1,562<br>$MS_{AB}$=39,063<br>$MS_{вип}$=2,938 |
| $F_{емп}$ | $F_{емп(A)} = \frac{MS_A}{MS_{вип}}$; $F_{емп(B)} = \frac{MS_B}{MS_{вип}}$;<br>$F_{емп(AB)} = \frac{MS_{AB}}{MS_{вип}}$ | $F_{емп\,(A)}$=0,532<br>$F_{емп\,(B)}$=0,532<br>$F_{емп\,(AB)}$=13,296 |
| $df_1$, $df_2$ | $df_1$=$df_A$ (для фактора А);<br>$df_1$=$df_B$ (для фактора В);<br>$df_1$=$df_{AB}$ (для взаємодії факторів А та В);<br>$df_2$=$df_{вип}$ | $df_1$=1<br>$df_2$=12 |
| $F_{крит}$ | Шукається за таблицями критичних значень (див. додаток, табл. 2) | $F_{крит} = 4{,}75(p \le 0{,}05)$<br>$F_{крит} = 9{,}33(p \le 0{,}01)$ |
| Перевірка гіпотез | $\left[\begin{array}{l} F_{еми} < F_{крит} \Rightarrow H_0 \\ F_{еми} \ge F_{крит} \Rightarrow H_1 \end{array}\right.$ | $F_{емп\,(A)} < F_{крит}$ – $H_0$<br>$F_{емп\,(B)} < F_{крит}$ – $H_0$<br>$F_{емп\,(AB)} > F_{крит}$ – $H_1$ |

**7 крок:** ВИСНОВОК

У комплектах гіпотез 1 та 2 приймається гіпотеза $H_0$

У комплекті гіпотез 3 приймається гіпотеза $H_1$ про значимість впливу взаємодії двох факторів (вибору та винагороди) на зміну ставлення до закону про обмеження самоврядування студентів.

# Розділ V. МЕТОДИ БАГАТОМІРНОГО СТАТИСТИЧНОГО АНАЛІЗУ[15]

## § 5.1. ФАКТОРНИЙ АНАЛІЗ

1. Мета факторного аналізу.
2. Підготовка даних до факторного аналізу.
3. Інтерпретація результатів факторного аналізу.
4. Задача про кількість факторів.
5. Особливості факторного аналізу.
6. Місце факторного аналізу в структурі експерименту.
7. Проведення факторного аналізу вручну (за Ф. Франселлою та Д.Банністером).

Рекомендована література:


1. Исследование природы конфликтных смыслов //Столин В. В. Самосознание личности.- М.: Изд-во Моск. ун-та, 1983.
2. Левандовский Н.Г. О корректности применения факторного анализа и о критериях факторизации // Вопр. психол. - 1980. - №5.
3. Петренко В.Ф. Психосемантика сознания. – М.: Изд-во Моск. ун-та, 1988.
4. Похилько В.И., Федотова Е.О. Техника репертуарных решеток в экспериментальной психологии личности //Вопр. психол. – № 3, 1984. – С.151-157.
5. Теплов Б.М. Простейшие способы факторного анализа /Типологические особенности высшей нервной деятельности человека. – Т.5. – М., 1967.
6. Франселла Ф. Баннистер Д. Новый метод исследования личности – М., 1987.
7. Факторный анализ // Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. – СПб.: Питер, 2000.
8. Факторный анализ интеллекта // Анастази А., Урбина С. Психологическое тестирование. – СПб.: Питер, 2003.
9. *http://www.statsoft.ru/home/textbook/default.htm*
10. *http://redyar.samara.ru/stat/wfuck.html*


### *1. МЕТА ФАКТОРНОГО АНАЛІЗУ*

Факторно-аналітичний підхід ґрунтується на уявленні про комплексний характер досліджуваного явища, що виявляється, зокрема, у взаємозв'язках між окремими його ознаками.

**Мета факторного аналізу** – сконцентрувати вихідну інформацію,

---

[15] *Методи статистичного аналізу, представлені у цій лекції, розраховані на комп'ютерну обробку даних з допомогою спеціалізованих програм: STATISTICA, SPSS, NCSS, S+ тощо.*

представлену у вигляді масиву даних і виразити якомога більшу кількість ознак через якомога меншу кількість характеристик. Вважається, що наймісткіші характеристики і будуть найсуттєвішими. Саме ці узагальнені місткі характеристики і називаються **факторами**.

Можна виділити чотири напрямки, в яких можна використовувати факторний аналіз:

1. Зменшення кількості змінних (редукція даних).
2. Групування, класифікація та компактна візуалізація даних.
3. Пошук прихованих змінних.
4. Генерація нових ідей.

## *2. ПІДГОТОВКА ДАНИХ ДО ФАКТОРНОГО АНАЛІЗУ*

Факторний аналіз спирається на обчислення коефіцієнтів кореляції між змінними, а тому масив даних, які готуються для факторного аналізу, повинен відповідати певним вимогам.

1. Масив повинен бути представлений у вигляді двовимірної матриці.
2. У стовпчиках матриці повинні бути занесені аналізовані змінні, а в рядках – значення цих змінних.
3. У матриці не повинно бути пропусків.
4. Бажано, щоб кількість рядків була більшою за кількість стовпчиків.
5. Кількість змінних (стовпчиків) повинна бути достатньо великою (більше 10).

Нижче наведено приклад масиву даних, підготовленого для факторного аналізу. Нехай ми провели опитування, в якому просили людей оцінити політичних лідерів за якістю “розум” (за 10-бальною шкалою). Після цього дані представили у таблиці 5.1.1, де стовпчики – закодовані імені лідерів (змінні), а рядки – опитані особи (значення змінних). Дійсно, факторний аналіз може дати відповідь на питання, як сприймаються людьми політичні лідери, оскільки це і є задача на з’ясування структури зв’язків між змінними.

*Таблиця 5.1.1.*

*Дані, підготовлені для факторного аналізу*

| | | Політичні лідери (змінні) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | _1 | _2 | _3 | _4 | _5 | _6 | _7 | _8 | _9 | _10 |
| Опитані особи (значення змінних – об'єкти) | 1 | 2 | 1 | 3 | 7 | 8 | 6 | 9 | 9 | 4 | 5 |
| | 2 | 1 | 5 | 3 | 6 | 4 | 8 | 9 | 9 | 2 | 7 |
| | 3 | 3 | 1 | 5 | 4 | 7 | 7 | 7 | 7 | 6 | 2 |
| | 4 | 2 | 3 | 5 | 6 | 6 | 1 | 6 | 6 | 6 | 4 |
| | 5 | 1 | 2 | 3 | 6 | 7 | 8 | 9 | 9 | 5 | 4 |
| | 6 | 6 | 1 | 4 | 9 | 7 | 8 | 3 | 2 | 10 | 5 |
| | 7 | 4 | 1 | 2 | 10 | 5 | 8 | 3 | 7 | 9 | 6 |
| | 8 | 5 | 1 | 4 | 7 | 6 | 10 | 8 | 2 | 9 | 3 |
| | 9 | 10 | 5 | 6 | 3 | 9 | 2 | 8 | 1 | 4 | 7 |
| | 10 | 9 | 4 | 5 | 6 | 1 | 3 | 10 | 7 | 8 | 2 |
| | 11 | 4 | 3 | 6 | 7 | 1 | 5 | 2 | 9 | 8 | 10 |
| | 12 | 8 | 1 | 2 | 7 | 6 | 9 | 3 | 5 | 10 | 4 |
| | 13 | 10 | 3 | 4 | 9 | 5 | 8 | 2 | 6 | 7 | 1 |
| | 14 | 2 | 3 | 6 | 7 | 4 | 5 | 1 | 8 | 9 | 10 |
| | 15 | 5 | 1 | 2 | 8 | 10 | 9 | 7 | 3 | 4 | 6 |
| | 16 | 7 | 2 | 3 | 1 | 10 | 6 | 5 | 9 | 8 | 4 |
| | 17 | 6 | 1 | 3 | 9 | 8 | 10 | 5 | 2 | 4 | 7 |
| | 18 | 4 | 1 | 2 | 9 | 5 | 10 | 6 | 3 | 7 | 8 |
| | 19 | 2 | 1 | 4 | 6 | 5 | 8 | 9 | 7 | 10 | 3 |
| | 20 | 2 | 1 | 4 | 6 | 5 | 8 | 9 | 7 | 10 | 3 |

## *3. ІНТЕРПРЕТАЦІЯ РЕЗУЛЬТАТІВ ФАКТОРНОГО АНАЛІЗУ*

Основні результати факторного аналізу виражаються у *факторних навантаженнях та факторних полях*, *факторних вагах* та *власних значеннях факторів*.

**Факторні навантаження** (factor loadings) – це коефіцієнти кореляції кожної із аналізованих змінних із кожним з виділених факторів.

Чим тісніший зв'язок змінної із фактором, тим більшим є її факторне навантаження. Позитивний знак факторного навантаження вказує на прямий зв'язок змінної з фактором, негативний – на обернений. Таблиця факторних

навантажень, отримана в результаті факторного аналізу, містить кількість стрічок, рівну кількості змінних (стовпчиків у вихідній матриці), та кількість стовпчиків, рівну кількості виділених факторів. Факторні навантаження, отримані в результаті факторного аналізу даних наведені в табл. 5.1.2.

*Таблиця 5.1.2.*

*Факторні навантаження*

| | Factor | |
|---|---|---|
| | 1 | 2 |
| _1 | -0,03995 | -0,05209 |
| _2 | ***-0,83896*** | -0,02758 |
| _3 | ***-0,78238*** | -0,29903 |
| _4 | ***0,612751*** | ***-0,52075*** |
| _5 | 0,321634 | ***0,69117*** |
| _6 | ***0,877572*** | 0,02443 |
| _7 | -0,17285 | ***0,745851*** |
| _8 | -0,38113 | -0,06763 |
| _9 | 0,215095 | ***-0,57704*** |
| _10 | -0,10559 | -0,41135 |
| Expl.Var | 2,799167 | 1,905432 |
| Prp.Totl | 0,279917 | 0,190543 |

Як бачимо, виділилося 2 фактори. Найтісніше пов'язаними з першим фактором виявилися лідери 2, 3, 4, та 6, причому 2 та 3 протистоять 4 та 6. Якщо подивитися на характер оцінок цих лідерів, то стане зрозуміло, що як розумних оцінюють лідерів 4 та 6 (у них вищі бали), а як не зовсім розумних – 2 та 3. Інші лідери не потрапили до цього фактору, а тому нічого про них сказати ми не можемо.

Аналогічно аналіз проводиться по другому фактору. Тут ми бачимо протиставлення лідерів 5 та 7 лідерам 4 та 9. Така ситуація може стати джерелом подальших досліджень, оскільки наявність другого фактору не передбачалася (згадайте, що оцінювання проводилося лише за параметром "розум").

Цікаво, що факторні навантаження можна представити графічно (рис. 5.1.1.).

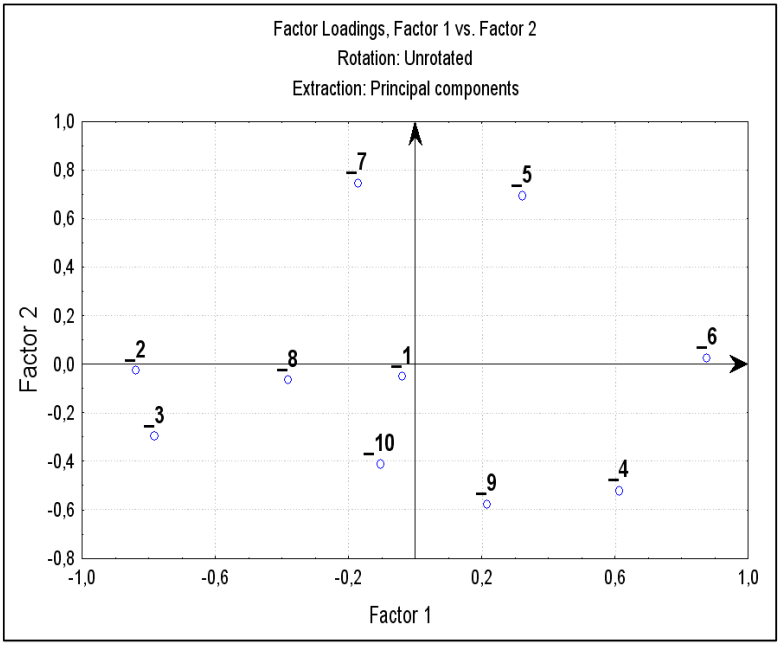



*Рис. 5.1.1. Графічне представлення факторних навантажень (факторне поле)*

Можна висунути таку гіпотезу стосовно причин виділення другого непередбачуваного фактора:

> *причиною виділення другого фактора може бути різне розуміння опитаними людьми поняття "розум": для одного – це **інтелект**, а для другого – життєва **практичність**.*

З'ясувати це дозволить наступний результат факторного аналізу – факторні ваги.

---

**Факторні ваги** (factor scores) – кількісні значення зв'язку виділених факторів з об'єктами[16]

---

[16] *Говорити про зв'язок тут не зовсім коректно з математичної точки зору, оскільки ці значення обраховуються способом, відмінним від кореляційного. Однак, ми використовуємо тут це поняття для кращого розуміння самого поняття "факторна вага"*

Об'єкту з більшою факторною вагою властивий більший рівень прояву властивостей виділеного фактора (більший рівень зв'язку з фактором). Позитивні факторні ваги відповідають об'єктам, що мають рівень прояву властивостей фактора більше середнього, а негативні – нижче середнього. Таблиця факторних ваг складається з кількості стрічок, рівної кількості об'єктів, та кількості стовпчиків, рівної кількості виділених факторів.

Нижче наведена таблиця факторних ваг, отримана в результаті дослідження сприймання політичних лідерів (табл. 5.1.3.).

*Таблиця 5.1.3.*

*Факторні ваги*

| | Factor | |
|---|---|---|
| | 1 | 2 |
| 1 | 0,111905 | ***1,112585*** |
| 2 | -0,88068 | 0,698867 |
| 3 | -0,28344 | ***0,870449*** |
| 4 | -1,23565 | 0,138548 |
| 5 | 0,041222 | ***1,062457*** |
| 6 | ***0,93953*** | -0,8847 |
| 7 | ***1,075413*** | -1,08295 |
| 8 | ***0,829232*** | 0,209612 |
| 9 | -1,85359 | ***1,07553*** |
| 10 | -1,52139 | -0,20612 |
| 11 | -1,19151 | -2,17375 |
| 12 | ***1,080925*** | -0,51853 |
| 13 | 0,188427 | -0,72121 |
| 14 | -0,94014 | -1,95754 |
| 15 | ***1,180397*** | ***1,103994*** |
| 16 | -0,41152 | ***1,155574*** |
| 17 | ***1,174646*** | 0,219696 |
| 18 | ***1,256649*** | -0,41478 |
| 19 | 0,21979 | 0,156133 |
| 20 | 0,21979 | 0,156133 |

Як бачимо з таблиці, найбільші факторні ваги по першому фактору мають опитані 6, 7, 8, 12, 15, 17, 18. По другому фактору – 1, 3, 5, 9, 15, 16. Таким чином, виявилося, що опитувані по різному пов'язані з виділеними факторами, а значить, і за основу для оцінки брали різні уявлення про "розум". Провівши додаткове дослідження, можна буде визначити, з яким

фактором (першим чи другим) який тип “розуму” пов’язується, а значить, можна буде встановити фактори уявлень про розум політичних лідерів у масовій свідомості.

Однак, це ще не всі можливості факторного аналізу – з його допомогою можна встановити, який із факторів найбільш значимий. Для цього використовується аналіз власних значень факторів.

**Власні значення** (eigenvalues) – це дисперсії, які пояснюються факторами.

В табл. 5.1.4. наведені власні значення виділених факторів.

*Таблиця 5.1.4.*

*Власні значення факторів*

| Eigenvalues (factor1.sta)<br>Extraction: Principal components | | | | |
|---|---|---|---|---|
| Factors | Eigenval | % total Variance | Cumul. Eigenval | Cumul. % |
| 1 | 2,799167 | ***27,99167*** | 2,799167 | 27,99167 |
| 2 | 1,905432 | ***19,05432*** | 4,704598 | ***47,04598*** |

У стовпчику (*Eigenval*) наведено дисперсії кожного фактора, а в стовпчику (*% total Variance*) – відсоток від загальної дисперсії для кожного фактора. У нашому випадку перший фактор пояснює приблизно 28% дисперсії, а другий фактор – 19% дисперсії. Стовпчик (*Cumul.Eigenval*) містить накопичену або кумулятивну дисперсію виділених факторів, а стовпчик (*Cumul%*) – накопичений відсоток від загальної дисперсії.

Для аналізу важливими є такі показники:

- накопичений відсоток дисперсії обох факторів (47%), який визначає, наскільки повно вдалося описати нашу сукупність даних з допомогою виділених факторів. Чим вищий цей показник, тим більшу частину масиву даних вдалося факторизувати;
- відсоток загальної дисперсії для кожного фактора, який вказує на значимість цього фактора. Чим більший відсоток дисперсії пояснює фактор, тим він значиміший і тим більше змінних він в себе включає.

У нашому випадку перший фактор пояснює 28% дисперсії, а другий – 19%. Разом вони описують 47% дисперсії, тобто, майже половину масиву даних. Це означає, що факторизація не повна, і що існують ще й інші фактори, менш значимі, але теж достатньо важливі.

Постає питання: скільки ж факторів варто виділити, щоб, з одного боку, вони максимально повно описували весь масив даних, а з іншого – були достатньо значимими? Про це ми поговоримо далі.

Зробимо підсумок. Для інтерпретації результатів факторного аналізу необхідні такі величини:

- факторні навантаження;
- факторне поле;
- факторні ваги;
- власні значення факторів.

В текстах наукових робіт представляють лише факторні поля, а при описі цих факторних полів поруч з кожною змінною в дужках ставлять її факторне навантаження. Самі факторні навантаження, ваги та власні значення факторів представляють у додатках.

## *4. ЗАДАЧА ПРО КІЛЬКІСТЬ ФАКТОРІВ*

Взагалі, рішення про кількість виділених факторів достатньо довільне. Зазвичай, спочатку виділяють максимально можливу кількість факторів, а потім на основі аналізу таблиці власних значень приймають рішення про найбільш збалансовану кількість виділених факторів і проводять знову аналіз вже для обраної їх кількості.

Однак, існують загальні рекомендації, слідування яким дасть непоганий результат.

### Критерій Кайзера

Цей критерій, розроблений Х.Ф. Кайзером (1960), найширше використовується і полягає у тому, що відбираються лише фактори, власні значення яких більше 1. Цей критерій автоматично включається в модулі факторного аналізу програми STATISTICA.

### Критерій кам'янистого насипу

Це є графічний критерій, вперше описаний Р.Б. Кеттелом (1966). На графіку в порядку спадання зображують власні значення кожного виділеного фактора (рис. 5.1.2.).

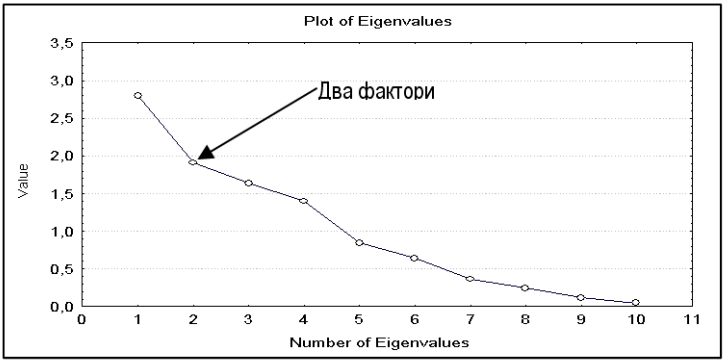



*Рис. 5.1.2. Критерій кам'янистого насипу*

Кеттел запропонував знайти таке місце на графіку, де зменшення власних значень зліва направо максимально сповільнюється. Відповідно до цього критерію варто залишити 2 фактори. Назва цього критерію походить від геологічного терміну, яким позначають уламки гірських порід, які накопичуються в нижній частині скелястого схилу – фактори справа і є тими уламками.

**Який же критерій використовувати?**

Обидва критерії достатньо глибоко вивчені та проаналізовані і це дозволяє зробити певні висновки. Критерій Кайзера іноді зберігає занадто багато факторів, а критерій кам'янистого насипу – навпаки, іноді зберігає занадто мало факторів. Однак, обидва критерії дають непоганий результат, якщо їх застосовувати в нормальних умовах (невелика кількість факторів та велика кількість змінних і об'єктів).

## *5. ОСОБЛИВОСТІ ФАКТОРНОГО АНАЛІЗУ*

Результати факторного аналізу будуть вдалими, якщо вдасться їх проінтерпретувати, виходячи із смислу показників, які характеризують виділені фактори. Цей етап роботи є надзвичайно відповідальним, оскільки вимагає від дослідника досвіду аналізу та чіткого уявлення про аналізовані змінні. Саме тому включення в набір змінних якомога більшої кількості різноманітних показників ні до чого не призведе. Факторний аналіз не дає

якогось нового знання, він лише допомагає виявити структуру заданих дослідником змінних. У зв'язку з цим при підборі змінних для факторного аналізу варто керуватися не бажанням "объять необъятное", а смисловим наповненням цих змінних.

З приводу питання про те, скільки ж факторів аналізувати, якщо їх вибір є досить довільними, то зазвичай досліджують декілька рішень з різною кількістю факторів, а потім обирається одне, найбільш осмислене.

Однак, розглянемо декілька методичних питань, пов'язаних із особливостями факторного аналізу.

1. Існує багато різних методів факторного аналізу (метод головних компонент, центроїдний метод, метод головних осей і т.д.). Більшість із них не статистичні в строгому смислі цього слова, оскільки для них не розроблені методи поширення результатів на генеральну сукупність. Вихідну кореляційну матрицю вважають заданою і обчислення проводяться без врахування статистичної значимості окремих кореляцій. Саме тому питання про значимість факторних навантажень вирішується з допомогою емпіричних порогів значимості (0,4 або 0,7), а змістове наповнення фактора визначається на основі змінних, які мають високі (значимі) факторні навантаження.

   Виключенням є метод максимальної правдоподібності та канонічний факторний аналіз, для яких способи перевірки статистичної значимості розроблені.

2. Однією з проблем факторного аналізу є проблема повороту факторних структур. Будь-який ортогональний поворот факторів призводить до такої ж факторизації, але з перерозподілом навантажень. Необхідність повороту виникає часто тоді, коли виділеним факторам не вдається дати чітку смислову інтерпретацію. Наприклад, в одному факторі навантаження різних змінних можуть бути близькі за величиною і тоді важко знайти єдиний смисловий інваріант цього фактора. Поворот же дозволяє зробити факторні навантаження контрастнішими за рахунок певного зменшення навантажень по одних змінних і збільшення по інших. Це сприяє чіткішому прояву ознак, які формують той чи інший фактор.

## *6. МІСЦЕ ФАКТОРНОГО АНАЛІЗУ В СТРУКТУРІ ЕКСПЕРИМЕНТУ*

Узагальнюючи короткий огляд факторного аналізу, слід вказати на його місце в структурі експериментальної роботи. Виділяють два основні підходи до його використання:

1. Пошуковий підхід
2. Направлений підхід.

**Пошуковий підхід** до використання факторного аналізу полягає у його використанні на першій стадії дослідження складного явища з метою пошуку гіпотез про його структуру. Використання факторного аналізу на ранніх стадіях експерименту є дійсно надзвичайно продуктивним, однак, слід пам'ятати, що "він не чарівна паличка, яка може з гори сирих фактів витягнути приховані закономірності". Факторний аналіз – це, передусім, засіб перевірки та селекції гіпотез.

З іншого боку, розглядають **направлений факторний аналіз**, який має на меті проведення експерименту для підтвердження вже висунутої теоретичної гіпотези. Цей факторний аналіз використовують на пізніших етапах дослідження, коли вже необхідно, наприклад, визначити розмірність досліджуваного явища чи побудувати певний узагальнений індекс для цього явища.

## *7. ПРОВЕДЕННЯ ФАКТОРНОГО АНАЛІЗУ ВРУЧНУ (ЗА Ф. ФРАНСЕЛЛОЮ ТА Д.БАННІСТЕРОМ)*[17]

Для пояснення алгоритму проведення факторного аналізу вручну проведемо дослідження сприймання підлітком свого значимого оточення. Нас цікавитимуть фактори, які це сприймання визначають.

*Для цього можна з допомогою асоціативного експерименту виявити якості, які досліджуваний приписує людям зі свого оточення, та серед виявлених асоціацій з допомогою частотного аналізу виділити найзначиміші. Наступним кроком може бути створення матриці "якості-значимі особи", в якій досліджуваного просять проранжувати значимих осіб за ступенем вираження у них відповідних рис (або оцінити).*

Після ранжування матриця може мати такий вигляд (табл. 5.1.5.):

[17] *Питання лекції написано спільно з В.В. Горбуновою*

*Таблиця 5.1.5.*

| **Якість** | я | мати | батько | брат | сестра | кохана людина |
|---|---|---|---|---|---|---|
| Досягнення | 2 | 5 | 6 | 1 | 3 | 4 |
| Мудрість | 5 | 3 | 6 | 2 | 4 | 1 |
| Чіткість планів | 4 | 5 | 3 | 1 | 6 | 2 |
| Пошук | 2 | 4 | 5 | 1 | 3 | 6 |
| Рух | 3 | 4 | 5 | 2 | 1 | 6 |
| Досягнення вищого | 1 | 4 | 6 | 2 | 3 | 5 |
| Визначеність | 5 | 4 | 3 | 1 | 6 | 2 |
| Хитрість | 5 | 4 | 6 | 2 | 3 | 1 |

Тепер, маючи дані досліджуваного, можна приступати до їх факторного аналізу.

**1 крок:** КОРЕЛЯЦІЙНИЙ АНАЛІЗ

Якщо досліджуваний оцінював значимих осіб, то для обробки даних слід використати коефіцієнт кореляції Пірсона.

Якщо досліджуваний ранжував значимих осіб, то для обробки слід використати коефіцієнт кореляції Спірмена (див. Розділ І).

**Коефіцієнти кореляції обчислюють між рядками заповненої матриці.**

У нашому випадку досліджуваний ранжував значимих осіб, а тому ми провели обчислення з допомогою коефіцієнта рангової кореляції, і отримали такі результати (табл. 5.1.6.):

*Таблиця 5.1.6.*

| | Досягнення | Мудрість | Чіткість планів | Пошук | рух | Досягнення вищого | Визначеність | Хитрість |
|---|---|---|---|---|---|---|---|---|
| Досягнення | 1,00 | 0,31 | 0,26 | 0,83 | 0,66 | 0,89 | 0,09 | 0,43 |
| Мудрість | 0,31 | 1,00 | 0,43 | -0,09 | -0,14 | 0,03 | 0,54 | 0,94 |
| Чіткість планів | 0,26 | 0,43 | 1,00 | 0,03 | -0,37 | -0,09 | 0,94 | 0,37 |
| Пошук | 0,83 | -0,09 | 0,03 | 1,00 | 0,83 | 0,89 | -0,09 | -0,03 |
| Рух | 0,66 | -0,14 | -0,37 | 0,83 | 1,00 | 0,71 | -0,43 | 0,03 |
| Досягнення вищого | 0,89 | 0,03 | -0,09 | 0,89 | 0,71 | 1,00 | -0,26 | 0,09 |
| Визначеність | 0,09 | 0,54 | 0,94 | -0,09 | -0,43 | -0,26 | 1,00 | 0,43 |
| Хитрість | 0,43 | 0,94 | 0,37 | -0,03 | 0,03 | 0,09 | 0,43 | 1,00 |

**2 крок:** ОБЧИСЛЕННЯ ПОКАЗНИКІВ ВЗАЄМОЗВ'ЯЗКУ (табл. 5.1.7.)

Взаємозв'язок конструктів (R) обчислюється за такою формулою:

$$R_{ij} = r_{ij}^2 \cdot 100$$

де $R_{ij}$ – показник взаємозв'язку між і-ю та j-ю якостями,

$r_{ij}^2$ – квадрат коефіцієнта кореляції між і-ю та j-ю якостями.

*Таблиця 5.1.7.*

| | Досягнення | Мудрість | Чіткість планів | Пошук | Рух | Досягнення вищого | Визначеність | Хитрість |
|---|---|---|---|---|---|---|---|---|
| Досягнення | | 9,877 | 6,612 | 68,653 | 43,183 | 78,448 | 0,734 | 18,367 |
| Мудрість | 9,877 | | 18,367 | 0,734 | 2,040 | 0,081 | 29,469 | 88,89 |
| Чіткість планів | 6,612 | 18,367 | | 0,081 | 13,795 | 0,734 | 88,897 | 13,795 |
| Пошук | 68,653 | 0,734 | 0,081 | | 68,653 | 78,448 | 0,734 | 0,081 |
| Рух | 43,183 | 2,040 | 13,795 | 68,653 | | 51,020 | 18,367 | 0,081 |
| Досягнення вищого | 78,448 | 0,081 | 0,734 | 78,448 | 51,020 | | 6,612 | 0,734 |
| Визначеність | 0,734 | 29,469 | 88,897 | 0,734 | 18,367 | 6,612 | | 18,367 |
| Хитрість | 18,367 | 88,897 | 13,795 | 0,081 | 0,081 | 0,734 | 18,367 | |
| **Загальна дисперсія, що пояснюється якістю** | 225,88 | 149,47 | 142,29 | 217,39 | 197,14 | 216,08 | 163,18 | 140,33 |

Після обчислення всіх показників взаємозв'язку знаходять суму всіх балів по кожній з якостей. Це і буде загальна дисперсія, що пояснюється якістю (в принципі, є аналогом **власних значень факторів**).

Самі показники взаємозв'язку є аналогом **факторних навантажень**.

**3 крок:** ПОБУДОВА ФАКТОРНОГО ПРОСТОРУ

Необхідно розташувати всі якості у просторі двох факторів-осей. **Перший фактор** буде утворений якістю, яка має найбільшу суму показників взаємозв'язку, тобто найтісніше пов'язана з іншими. **Другий фактор** утворює друга за силою взаємозв'язку якість, яка, однак, не корелює з першим фактором. Всі інші якості слід розмістити у просторі двох виділених факторів відповідно до коефіцієнтів кореляції "якість-фактор" (рис. 5.1.3.).

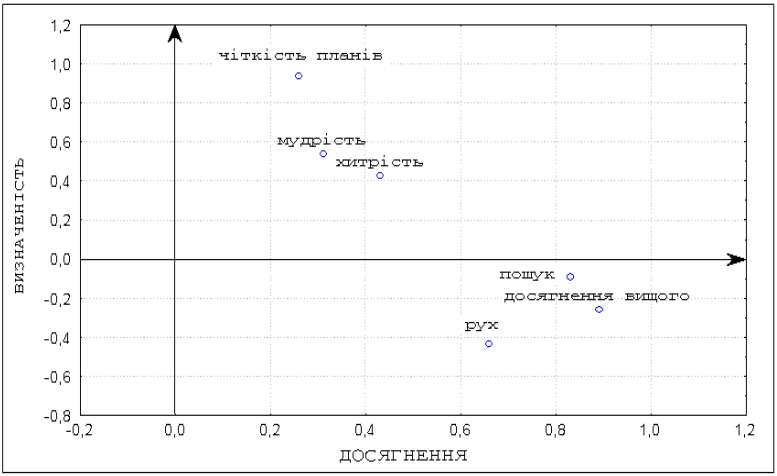



*Рис. 5.1.3. Факторний простір*

**4 крок:** ІНТЕРПРЕТАЦІЯ

Етап інтерпретації є найважливішим у факторному аналізі, адже саме для цього й проводять всі ці складні і громіздкі обчислення.

Загалом при інтерпретації результатів ручного факторного аналізу враховують такі показники:

1. Коефіцієнти кореляції між якостями.
2. Показники взаємозв'язку між якостями.
3. Загальні дисперсії, які пояснюються якостями.
4. Віднесеність якостей до факторів та їх полюсів.
5. Цілісна картина факторного простору.

Блискучі зразки інтерпретації результатів факторного аналізу можна знайти в роботах таких дослідників, як В.Ф. Петренко, Ф. Франселла та Д. Банністер, В.І. Похилько та Е.О. Федотова, О.Г. Шмельов.

## § 5.2. БАГАТОМІРНЕ ШКАЛЮВАННЯ

1. Загальна мета багатомірного шкалювання. Підгототовка даних до БШ
2. Логіка багатомірного шкалювання
3. Проблема розмірності при багатомірному шкалюванні
4. Інтерпретація результатів багатомірного шкалювання
5. Багатомірне шкалювання та факторний аналіз. Реплікуюче БШ.

Рекомендована література:


1. *Дейвисон М. Многомерное шкалирование. Методы наглядного представления данных: Пер. с англ. В.С. Каменского с предсл. С.А. Айвазяна и В.С. Каменского. – М.: Финансы и статистика, 1988. – 254 с.*
2. *Дронов С.В. Многомерный статистический анализ. – Барнаул: Изд-во Алтайского гос. ун-та, 2003. – 213 с.*
3. *Климчук В.О. Багатомірне шкалювання: використання у психологічних дослідженнях // Практична психологія та соціальна робота. – 2007. – №6. – С. 17-21.*
4. *Климчук В.О. Відображення у свідомості студентів-психологів термінологічного апарату психології здібностей // Науковий часопис НПУ імені М.П. Драгоманова, 2005. – №6 (30). – Ч.І. – С. 369-374.*
5. *Крускал Дж. Взаимосвязь между многомерным шкалированием и кластер-анализом. – http://lib.socio.msu.ru*
6. *Крылов В.Ю. О многомерном шкалировании в неметрических пространствах // Психологический журнал – 1987. – Т.8. – №5. – С. 140-142.*
7. *Лосик Г.В. Исследование восприятия гласных методом многомерного шкалирования // Психологический журнал – 1992. – Т.13. – №2.*
8. *Наследов А.Д. Математические методы психологического исследования. – СПб.: Речь, 2004. – 392 с.*
9. *Петров В.М. Опыт применения неметрического многомерного шкалирования при изучении предпочтений молодежи в области авторской песни // Социология 4 М. – 1991. – № 1. – С. 99-114.*
10. *Соколов Е.Н., Измайлов Ч.А., Завгородняя В.Л. Многомерное шкалирование знаковых конфигураций // Вопросы психологии. – 1985. – №1. – С. 133-139.*
11. *Терехина А. Ю. Многомерное шкалирование в психологи // Психологический журнал. – Том 4. – №1. – 1983. – С.76-88*
12. *Толстова Ю.Н. Основы многомерного шкалирования. – М.: Университет. – 157 с.*
13. *Шепард Р. Н. Многомерное шкалирование и безразмерное представление различий // Психологический журнал – 1980. – т. I. – № 4 – с. 72—83.*

## *1. МЕТА БАГАТОМІРНОГО ШКАЛЮВАННЯ. ПІДГОТОВКА ДАНИХ ДО БШ.*

Багатомірне шкалювання (БШ) є до певної міри альтернативою факторному аналізу. Перевага БШ – можливість у якості вихідних даних використовувати довільні матриці схожості (це можуть бути і кореляційні матриці, і прості матриці схожості (відмінності) об'єктів у шкалі відношень). Багатомірне шкалювання – це не стільки математична процедура, скільки **спосіб ефективного розміщення об'єктів у просторі певної кількості ознак**, який максимально зберігає відстані між цими об'єктами.

Отже, для проведення багатомірного шкалювання перш за все потрібно мати матрицю схожості (відмінності) об'єктів. Її можна отримати різними шляхами. *Оптимальними є дані, отримані шляхом експертної оцінки близькості об'єктів*. Якщо відомості про близькість об'єктів отримані не прямо, а опосередковано (на основі кореляційного чи коваріаційного аналізу), то вважається, що для таких даних БШ є не найкращим методом (варто подумати все ж таки про факторний аналіз).

Матриця схожості отримується дуже просто – учасникам дослідження чи експертам пропонують пари об'єктів і ставлять завдання оцінити їх внутрішню схожість чи відмінність.

Методи багатомірного шкалювання:

- метричні (близькість об'єктів оцінена в інтервальній чи пропорційній шкалах);
- неметричні (близькість об'єктів оцінена в ранговій шкалі).

Вихідна матриця повинна мати певні характеристики (для програми Statistica 5.5):

1. Кількість рядків = кількість стовпчиків + 4.
2. Назви рядків повинні точно відповідати назвам стовпчиків.
3. Остання 4-та стрічка повинна називатися **matrix** та містити число 2 (якщо аналізується матриця схожості) або число 3 (якщо аналізується матриця відмінностей).
4. Останні 1, 2, 3 стрічки можуть бути без назв і мають бути пустими.

Для подальшої роботи давайте візьмемо ілюстративний приклад.

> *Метою нашого дослідження буде сприймання студентами першого курсу психологічних теорій після вивчення курсу "Теорії особистості". Для цього ми створимо перелік цих теорій чи авторів цих теорій і попоросимо студентів попарно оцінити за 10-бальною шкалою, наскільки близькі ці теорії.*

Отриманий сирий результат може мати такий вигляд:

| | З. Фройд | К. Юнг | А. Адлер | Дж. Уотсон | Б. Скіннер | К. Халл | А. Маслоу | К. Роджерс | Р. Мей |
|---|---|---|---|---|---|---|---|---|---|
| З. Фройд | 10 | 7 | 8 | 1 | 3 | 3 | 2 | 1 | 2 |
| К. Юнг | | 10 | 6 | 3 | 1 | 4 | 1 | 1 | 2 |
| А. Адлер | | | 10 | 3 | 3 | 3 | 4 | 4 | 2 |
| Дж. Уотсон | | | | 10 | 8 | 7 | 3 | 4 | 3 |
| Б. Скіннер | | | | | 10 | 7 | 3 | 4 | 3 |
| К. Халл | | | | | | 10 | 3 | 3 | 3 |
| А. Маслоу | | | | | | | 10 | 8 | 9 |
| К. Роджерс | | | | | | | | 10 | 7 |
| Р. Мей | | | | | | | | | 10 |

Дані, підготовлені для БШ у програмі Statistica 5.5 матимуть такий вигляд:

| | ФРОЙД | ЮНГ | АДЛЕР | УОТСОН | СКІННЕР | ХАЛЛ | МАСЛОУ | РОДЖЕРС | МЕЙ |
|---|---|---|---|---|---|---|---|---|---|
| Фройд | 10 | 7 | 8 | 1 | 3 | 3 | 2 | 1 | 2 |
| Юнг | 7 | 10 | 6 | 3 | 1 | 4 | 1 | 1 | 2 |
| Адлер | 8 | 6 | 10 | 3 | 3 | 3 | 4 | 4 | 2 |
| Уотсон | 1 | 3 | 3 | 10 | 8 | 7 | 3 | 4 | 3 |
| Скіннер | 3 | 1 | 3 | 8 | 10 | 7 | 3 | 4 | 3 |
| Халл | 3 | 4 | 3 | 7 | 7 | 10 | 3 | 3 | 3 |
| Маслоу | 2 | 1 | 4 | 3 | 3 | 3 | 10 | 8 | 9 |
| Роджерс | 1 | 1 | 4 | 4 | 4 | 3 | 8 | 10 | 7 |
| Мей | 2 | 2 | 2 | 3 | 3 | 3 | 9 | 7 | 10 |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| matrix | 2 | | | | | | | | |

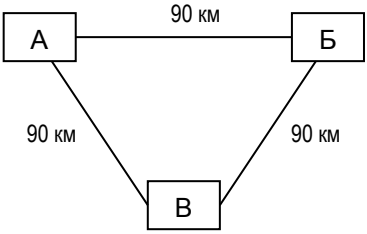
Ми з'ясували, що метою БШ є розміщення об'єктів у просторі певної розмірності, максимально зберігши відстані між ними. Логіку цього процесу можна проілюструвати на деяких прикладах.

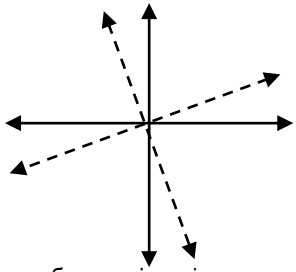
Нехай маємо матрицю попарних відстаней міст України. Обравши розмірність нового простору рівну 2, метод БШ розташує ці міста у 2-х мірному просторі, максимально зберігши відстані між ними. Отримане розташування точок можна буде використовувати як приблизну карту України. Самі ці 2 виміри можна тепер проінтерпретувати як координати "Пн\Пд" та "Зх\Сх"[18].

Для деталізації прикладу уявимо, що маємо 3 міста, а також відстані між цими містами:

| | Місто А | Місто Б | Місто В |
|---|---|---|---|
| Місто А | 0 | | |
| Місто Б | 90 км | 0 | |
| Місто В | 90 км | 90 км | 0 |

Тепер ці міста можна розташувати у певному n-мірному просторі. Нехай це буде двохмірний простір (площина). Тоді всі три міста дійсно можна розташувати на цій площині так, щоб зберегти відстані між ними:



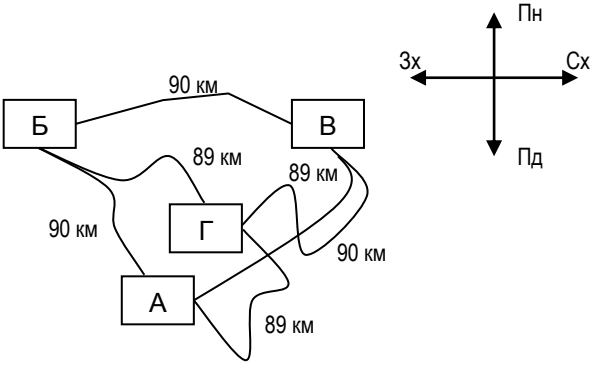



При цьому, розмістити ці міста можна так, щоб вони відповідали реальному розміщенню на карті, повертаючи осі координат.

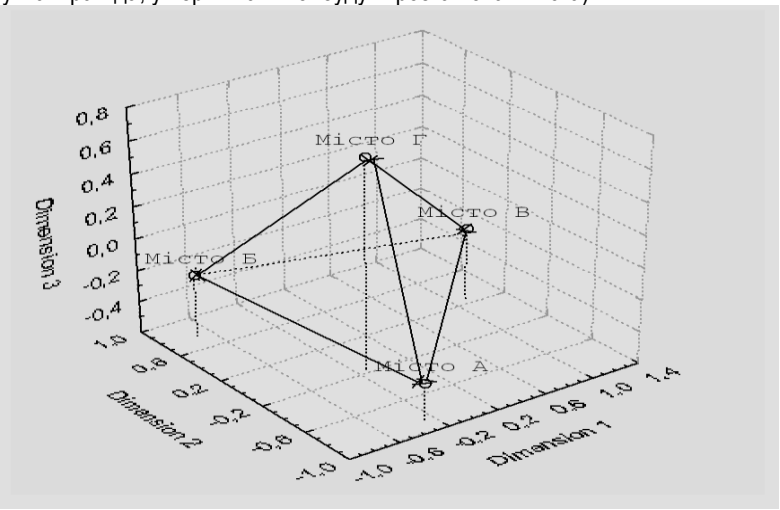
Ситуація ускладнюється, якщо маємо не 3, а 4 міста, причому 4-те місто Г знаходиться на відстані 89 км від усіх інших міст. Це може бути тоді, коли відстань між містами виміряна не прямо, а по довжині доріг:

[18] *Приклад з підручника StatSoft*

| | Місто А | Місто Б | Місто В | Місто Г |
|---|---|---|---|---|
| Місто А | 0 | | | |
| Місто Б | 90 км | 0 | | |
| Місто В | 90 км | 90 км | 0 | |
| Місто Г | 89 км | 89 км | 89 км | 0 |




Тоді процедура БШ буде не такою прозорою. Зберегти відстані у просторі осей при БШ вдасться лише у випадку 3-мірного простору (побудується 3-кутна піраміда, у вершинах якої будуть розташовані міста).

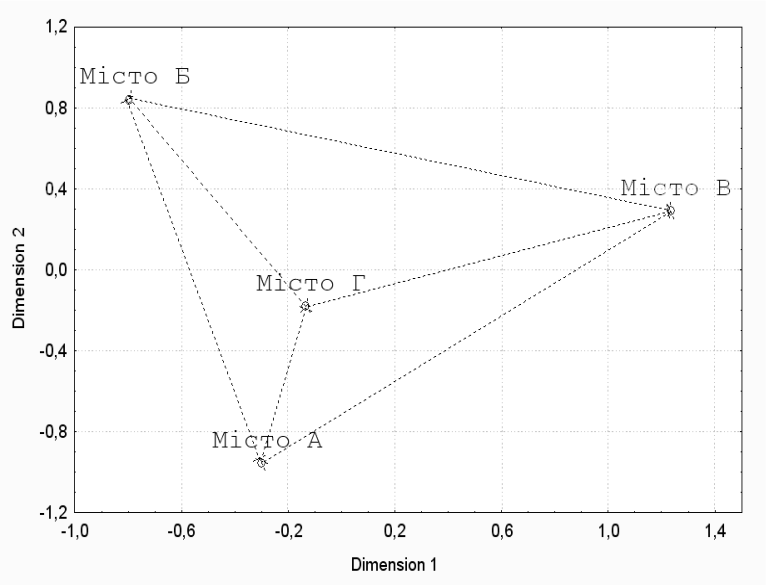


Але така картина навряд чи буде придатна до інтерпретації – вона не буде мати ніякого смислу. Тоді ми змушені спробувати розмістити ці міста у 2-мірній площині. Наслідком цього буде те, що процедура БШ на основі вихідної матриці відстаней обчислить оновлену матрицю відстаней, так, щоб оптимізувати розміщення міст – малі відстані можуть бути зменшені ще, і більші – збільшені тощо. .

В результаті **оновлена матриця** буде мати такий вигляд:

| | Місто А | Місто Б | Місто В | Місто Г |
|---|---|---|---|---|
| Місто А | 0 | | | |
| Місто Б | 1,86 | 0 | | |
| Місто В | 1,98 | 2,10 | 0 | |
| Місто Г | 0,79 | 1,22 | 1,45 | 0 |

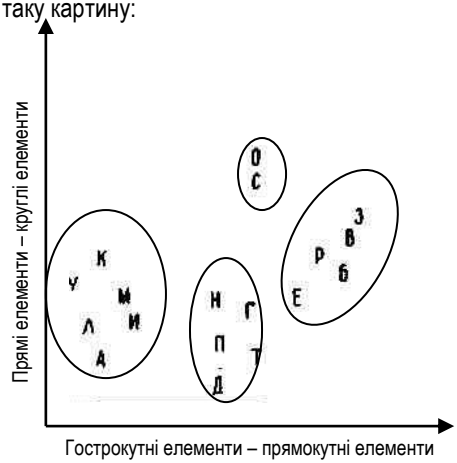
А розміщення міст на площині буде таке:




Порівняйте результат, отриманий з допомогою БШ та вихідний рисунок – вони дуже схожі. Дійсно, вийшло, що БШ перетворило відстані, виміряні по довжині доріг, у реальні геометричні відстані між містами. Таку картину тепер

можна проінтерпретувати як реальне розміщення міст на карті, а осі утвореного простору можна вважати осями сторін світу. Єдино, що можливо необхідно трішки повернути (див. ротацію факторів при ФА).

А.Ю. Терьохіна наводить приклад дослідженні візуального сприймання букв російського алфавіту. Учасникам дослідження пропонувалося оцінити пари букв за їх схожістю. Після обробки групової матриці схожості було проведено БШ та отримано таку картину:




Букви згрупувалися за схожістю в просторі двох осей, які на основі змістового аналізу вдалося проінтерпретувати як “прямі-круглі” та “гострокутні-прямокутні”.

Якщо продовжити наш приклад про сприймання психологічних теорій, то за логікою БШ повинні утворитися три групи теорій, найбільш близькі між собою (психоаналітичні – Фройд, Юнг, Адлер; поведінкові – Скіннер, Толмен, Халл; гуманістичні – Маслоу, Роджерс, Мей).

## *3. ПРОБЛЕМА РОЗМІРНОСТІ ПРИ БАГАТОМІРНОМУ ШКАЛЮВАННІ*

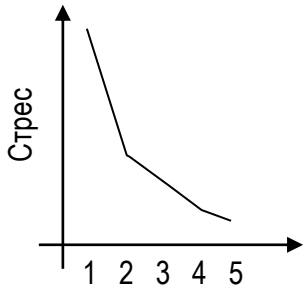
На попередньому прикладі з містами ми побачили, що чим більша розмірність простору, у якому ми проводимо БШ, тим більше буде відповідності між вихідною матрицею даних та **оновленою матрицею**. Якщо

кількість вимірів простору буде рівна кількості змінних, ми тримаємо повну відповідність між цими матрицями. Однак, це ж не є нашою задачею. Навіть навпаки, нам цікаво щоб кількість вимірів простору БШ була якомога меншою. Але тоді буде більше розбіжностей між вихідною та оновленою матрицею відстаней. З'являється дилема, яку необхідно вирішити.

Виявляється, існує два суттєвих фактори, які впливають на вибір мірності простору БШ.

**Перший фактор** – інтерпретованість. Необхідно підбирати таку мірність простору, щоб отримана конфігурація точок мала смисл і була якомога простішою та зрозумілою. Якщо конфігурація має вигляд хаотичної хмари, то навіть якщо усі математичні показники ідеальні – вони мало чим допоможуть. Ми не зможемо зробити жодних висновків. У такому випадку слід спробувати змінити мірність простору (збільшити чи зменшити) – можливо, тоді з'явиться очікувана структура.

**Другий фактор** – критерій кам'янистого насипу. Цей критерій був запропонований Р. Кеттелом для оптимального вибору кількості факторів у факторному аналізі (там йшлося про графік залежності власних значень факторів від їх кількості). У випадку ж багатомірного шкалювання можна говорити про графік залежності величини стресу (*про стрес більш детально поговоримо далі*) від мірності простору. Необхідно на цьому графіку знайти точку, у якій спадання стресу максимально сповільнюється. Ця кількість вимірів і буде оптимальною.




*Мірність простору*

Однак, проблема у тому, що Statistica автоматично не будує кам'янистий насип у БШ. Його необхідно побудувати вручну, провівши БШ з різними показниками мірності простору і відмітивши відповідно стрес для кожного БШ.

Для інтерпретації результатів БШ використовують певний набір показників.

**Вихідна матриця відстаней** – матриця, яку ми створили для проведення БШ.

У нашому випадку – це матриця оцінок студентами близькості різних психологічних теорій.

**Оновлена матриця відстаней (distance matrix)** – матриця відстаней, створена процедурою БШ для розміщення об'єктів у n-мірному просторі.

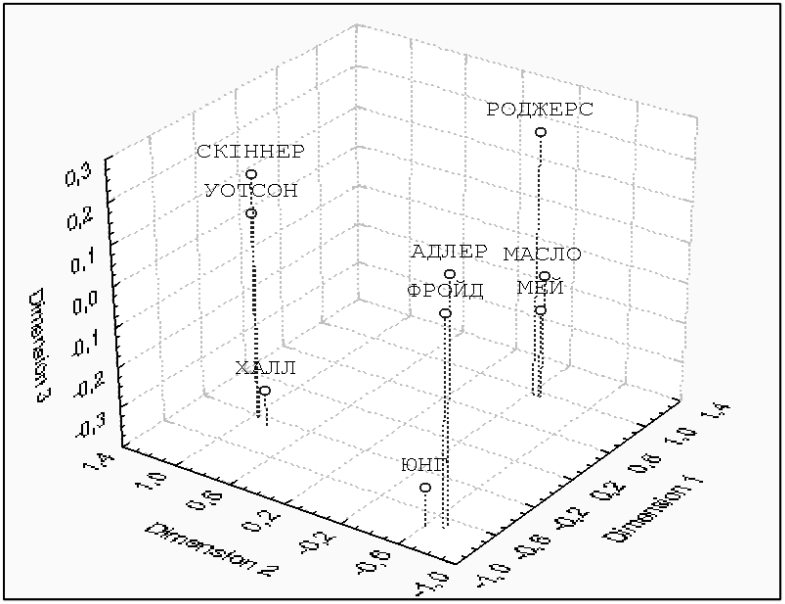
Оновлена матриця відстаней може не співпадати з вихідною матрицею відстаней, оскільки є її певним спрощенням. Для наших даних оновлена матриця відстаней має такий вигляд:

| | ФРОЙД | ЮНГ | АДЛЕР | УОТСОН | СКІННЕР | ХАЛЛ | МАСЛОУ | РОДЖЕРС | МЕЙ |
|---|---|---|---|---|---|---|---|---|---|
| Фройд | ***0,00*** | ***0,44*** | ***0,09*** | 1,74 | 1,75 | 1,68 | 1,73 | 1,74 | 1,75 |
| Юнг | ***0,44*** | ***0,00*** | ***0,53*** | 1,70 | 1,73 | 1,54 | 1,74 | 1,83 | 1,74 |
| Адлер | ***0,09*** | ***0,53*** | ***0,00*** | 1,73 | 1,73 | 1,70 | 1,70 | 1,69 | 1,73 |
| Уотсон | 1,74 | 1,70 | 1,73 | ***0,00*** | ***0,09*** | ***0,44*** | 1,73 | 1,69 | 1,74 |
| Скіннер | 1,75 | 1,73 | 1,73 | ***0,09*** | ***0,00*** | ***0,53*** | 1,74 | 1,68 | 1,75 |
| Халл | 1,68 | 1,54 | 1,70 | ***0,44*** | ***0,53*** | ***0,00*** | 1,70 | 1,74 | 1,68 |
| Маслоу | 1,73 | 1,74 | 1,70 | 1,73 | 1,74 | 1,70 | ***0,00*** | ***0,35*** | ***0,09*** |
| Роджерс | 1,74 | 1,83 | 1,69 | 1,69 | 1,68 | 1,74 | ***0,35*** | ***0,00*** | ***0,43*** |
| Мей | 1,75 | 1,74 | 1,73 | 1,74 | 1,75 | 1,68 | ***0,09*** | ***0,43*** | ***0,00*** |

Первинний аналіз можна провести уже за цією матрицею. Так, бачимо, що у нас виділяться 3 групи теорій, найбільш близькі між собою.

**N-мірна діаграма розсіювання (Graph final configuration)** – графічне розміщення об'єктів в n-мірному просторі

Ця діаграма є основним результатом БШ, і саме по ній як правило здійснюють інтерпретацію результатів БШ. Наша 3-мірна діаграма розсіювання матиме такий вигляд:




При аналізі діаграми розсіювання необхідно:

1. Проаналізувати виявлені групи точок, а також конфігурацію цих груп, їх кількість тощо.
2. Проінтерпретувати осі координат (виміри).

У нашому випадку виділилося 3 групи теорій (психоаналітична, гуманістична та поведінкова). Осі координат можна описати як: 1 вісь – психоаналіз-гуманістичний підхід, 2 вісь – поведінковий підхід – гуманістичний підхід, 3 вісь – *слабо піддається інтерпретації*, і очевидно, є лишньою. Можна провести повторне БШ в 2-мірному просторі.

---

**Таблиця кінцевої конфігурації (final configuration)** – координати кожного об'єкта по виділених осях. Це числове представлення діаграми розсіювання, і схоже на таблицю факторних навантажень.

---

Дійсно, з таблиці кінцевої конфігурації видно, що 3 вимір (3-тя вісь) є зайвою, оскільки жоден з об'єктів сильно його не навантажує.

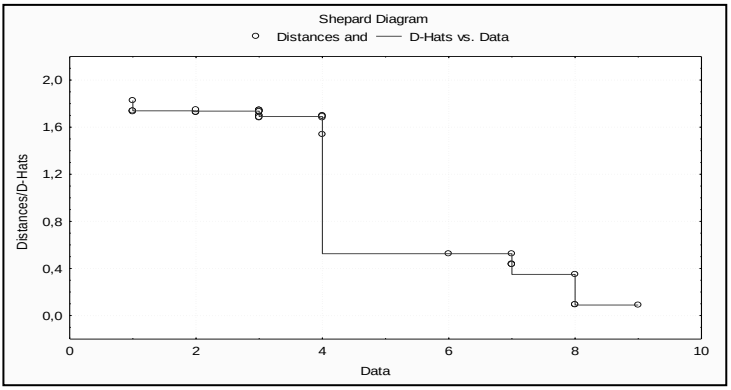
| D-star: Raw stress = ,0027762; Alienation = ,0058544 | | | |
|---|---|---|---|
| D-hat: Raw stress = ,0012195; **Stress = ,0038801** | | | |
| | **DIM. 1** | **DIM. 2** | **DIM. 3** |
| ФРОЙД | ***-0,70577*** | ***-0,71113*** | *0,121468* |
| ЮНГ | ***-0,74058*** | ***-0,60869*** | *-0,30383* |
| АДЛЕР | ***-0,66259*** | ***-0,71264*** | *0,204988* |
| УОТСОН | *-0,24198* | ***0,963722*** | *0,109729* |
| СКІННЕР | *-0,22722* | ***0,96582*** | *0,200653* |
| ХАЛЛ | *-0,2872* | ***0,86177*** | *-0,31107* |
| МАСЛОУ | ***0,952193*** | *-0,27553* | *-0,09334* |
| РОДЖЕРС | ***0,956395*** | *-0,22415* | *0,252479* |
| МЕЙ | ***0,956752*** | *-0,25918* | *-0,18108* |

**Стрес** – показник, який характеризує ступінь відмінності між вихідною матрицею даних та оновленою матрицею. Чим більше це число, тим менш надійними є результати БШ.

У нашому випадку стрес рівний 0,0038801, що є достатньо непоганим показником.

**Діаграма Шепарда (Shepard diagram)** – діаграма, за якою візуально можна побачити ступінь відмінності між вихідною та оновленою матрицею даних.

Діаграма Шепарда для нашого випадку має такий вигляд:



Для аналізу діаграми Шепарда дивляться, наскільки добре точки на ній лягають на ступінчату лінію. Якщо точки і лінія майже співпадають, то аналіз можна вважати надійним – оновлена матриця не спотворила вихідні дані.

## *5. БАГАТОМІРНЕ ШКАЛЮВАННЯ ТА ФАКТОРНИЙ АНАЛІЗ. РЕПЛІКУЮЧЕ БШ.*

На початку нашої роботи ми відмітили, що БШ може слугувати певною альтернативою ФА. Для кращого орієнтування у проблемі вибору аналізу слід описати деякі відмінності цих методів.

1. ФА накладає на вихідні дані вимогу відповідності критеріям нормального розподілу та статистичної значимості коефіцієнтів кореляції. В БШ такого обмеження немає.
2. ФА вимагає, щоб змінні були оцінені за певним списком характеристик, і лише тоді обчислює кореляційну матрицю. У БШ можна прямо оцінювати попарну схожість між об'єктами.
3. Для роботи з ФА необхідно мати коефіцієнти кореляцій між змінними. Для БШ крім коефіцієнтів кореляцій можна використовувати будь-які міри схожості чи відмінності між об'єктами – бали чи ранги.
4. ФА ґрунтується на уявленні про лінійний зв'язок між аналізованими змінними. БШ такого обмеження теж не має.
5. ФА намагається виділити більшу кількість факторів, ніж БШ. З цієї точки зору БШ дає більш прості для інтерпретації рішення.

Часто дослідник стикається з проблемою БШ результатів багатьох індивідуальних матриць. В такому випадку він змушений або на основі усереднення цих матриць будувати одну групову, і з нею проводити БШ, або проводити БШ окремо для кожної матриці, і потім узагальнювати результати. Однак, не так давно було розроблено метод ***реплікуючого БШ***. Він полягає у тому, що програма одночасно обробляє набір індивідуальних матриць, і будує комплексний груповий простір об'єктів.

## § 5.3. КЛАСТЕРНИЙ АНАЛІЗ

1. Мета та методи кластерного аналізу.
2. Підготовка даних до кластерного аналізу.
3. Деревоподібна кластеризація: пошук відстаней між об'єктами, стратегії кластеризації та інтерпретація.
4. Представлення результатів кластерного аналізу.
5. Кластерний аналіз за методом В.Ю. Крилова та Т.В. Острякової.

Рекомендована література:


1. Крылов В.Ю., Острякова Т.В. Новые методы кластерного анализа на основе психологической теории развития понятий Л.С. Выготского // Психологический журнал. – 1995. – Т16. – №1. – С. 130-137.
2. Общая психодиагностика /Под ред. А. А. Бодалева, В. В. Столина. – М.: Изд-во МГУ, 1987.
3. Петренко В.Ф. Психосемантика сознания. – М.: Изд-во Моск. ун-та, 1988.
4. http://www.statsoft.ru/home/textbook/default.htm
5. http://www.codenet.ru/progr/alg/ai/htm/gl3_10.php


### *1. МЕТА ТА МЕТОДИ КЛАСТЕРНОГО АНАЛІЗУ*

Достатньо багато досліджень ставлять за мету організацію отриманих даних у наглядні структури. Так, в біології часто метою є розбиття сукупності тварин на види і підвиди, у психології – класифікація видів поведінки, у педагогіці – таксономія виховних цілей тощо. Допомогти це зробити може *кластерний аналіз*.

Фактично, кластерний аналіз є набором різноманітних алгоритмів класифікації. Техніка кластеризації застосовується в дуже різноманітних сферах діяльності. Так, автори підручника Statsoft наводять приклади застосування кластерного аналізу в медицині – кластеризації піддаються симптоми захворювання чи види лікування, і отримуються достатньо цікаві класифікації. Теж саме стосується психіатрії та психотерапії. Відомі також ряд прикладів застосування кластерного аналізу в сфері маркетингу. Загалом, коли необхідно класифікувати великі масиви інформації на групи, які придатні для подальшого аналізу – кластерний аналіз є незамінним інструментом. Цікаві можливості кластерного аналізу у сфері психологічних досліджень групових процесів та явищ.

Кластерний аналіз має одну суттєву особливість – він не є звичайним статистичним методом, оскільки до нього у більшості випадків незастосовні процеси перевірки статистичної значимості. Кластерний аналіз дає найбільш можливо-значиме рішення. Саме тому досить часто його використовують тоді, коли дослідник має набір даних, але не має жодної апріорної гіпотези про класи цих даних.

Перш ніж перейти до безпосередньо алгоритмів кластеризації, виділимо декілька **зауважень**, які слід враховувати, використовуючи кластерний аналіз:

a. Більшість методів кластерного аналізу є доволі таки простими евристичними процедурами, які, як правило, не мають статистичного обґрунтування.
b. Різні методи кластеризації можуть породжувати різні кластерні рішення для одних і тих же даних. Це звичне явище у більшості прикладних дослідженнях, у тому слід по-перше обирати найбільш осмислене рішення, по-друге – завжди вказувати, який саме метод кластеризації було використано.
c. Використовуючи кластерний аналіз, дослідник має на меті виявлення структури даних. В той же час дія кластерного аналізу полягає у привнесені структури у аналізовані дані, тобто, кластеризація може призвести до появи артефактів (виявлення структури в даних, які її не мають тощо).
d. Осмислене рішення при кластерному аналізі можна обрати лише тоді, коли є базис для його осмислення – теорія. Без теоретичної моделі, без гіпотези стосовно структури даних з'являється небезпека наївного емпіризму, коли результати кластеризації приймаються на істину у кінцевій інстанції.

Яка ж загальна структура проведення кластерного аналізу? Можна виділити такі етапи роботи:

1. Проведення дослідження.
2. Підготовка даних до кластерного аналізу.
3. Вибір методу кластерного аналізу.
4. Вибір міри відстані між об'єктами та її обчислення.
5. Вибір стратегії кластеризації.
6. Застосування обраної стратегії для утворення кластерів.
7. Перевірка результатів кластерного аналізу на осмисленість і їх інтерпретація.

Виділяють три основних методи кластерного аналізу: деревоподібна кластеризація, метод К-середніх та двовходове об'єднання.

**Метод деревоподібної кластеризації (ієрархічна кластеризація, tree clustering)** дозволяє побудувати ієрархічне кластерне дерево, що має такий вигляд (рис. 5.3.1.):

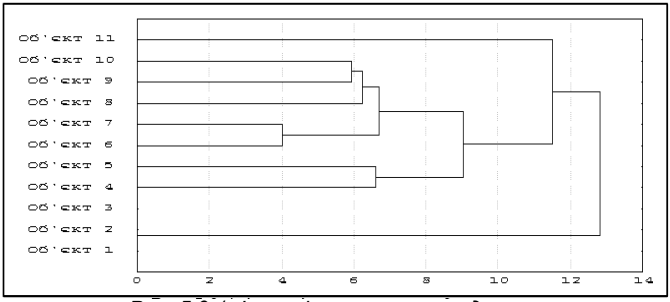



*Рис. 5.3.1. Ієрархічне кластерне дерево*

**Метод К-середніх (K-means clustering)** використовується тоді, коли у дослідника вже є певні апріорні гіпотези стосовно кількості кластерів. В межах цього методу дослідник має наперед задати кількість кластерів, і алгоритм кластеризації дозволить знати ці кластери так, щоб вони максимально різнилися один від одного. Перевагою цього методу є можливість перевірки статистичної значимості відмінностей між виділеними кластерами.

На рис. 5.3.2 наведено графічне зображення результатів кластерного аналізу методом К-середніх.

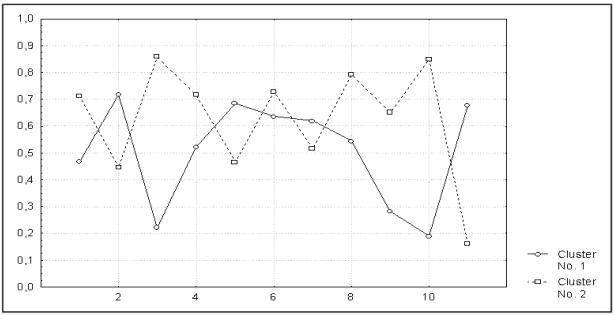



*Рис. 5.3.2. Кластерний аналіз методом К-середніх.*

**Метод двовходового об'єднання (two-way joining)** використовують у випадках, коли хочуть провести одночасну кластеризацію об'єктів (змінних) та спостережень. На рис. 5.3.3. наведено графічне зображення результатів використання цього методу.

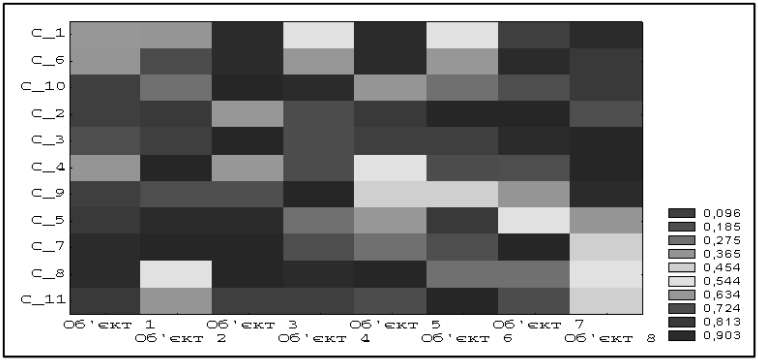



*Рис. 5.3.3. Результати кластерного аналізу методом двовходового об'єднання*

## *2. ПІДГОТОВКА ДАНИХ ДО КЛАСТЕРНОГО АНАЛІЗУ*

Підготовка даних до кластерного аналізу відбувається так же, як і при факторному аналізі, із збереженням усіх поставлених вимог (див. пункт "Основи факторного аналізу" цього практикуму).

*Нехай ми провели дослідження позитивного ставлення студентів психологічного факультету до студентів інших факультетів та до майбутніх професійних ролей, і нас цікавить об'єднання студентів у групи на основі схожого ставлення. Для цього ми створили рольовий перелік факультетів та можливих спеціальностей випускників психологічного факультету. Потім ми запропонували студентам оцінити своє ставлення до всіх ролей за 10-бальною шкалою. В результаті було отримано масив даних (табл. 5.3.1)[19].*

[19] *Висловлюю подяку В.В. Горбуновій за надані результати експериментального дослідження.*

*Таблиця 5.3.1.*

*Масив даних, підготовлений до кластерного аналізу*

| | | **Об'єкти (досліджувані студенти)** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | А.О. | О.В. | М.П. | К.П. | Р.А. | К.В. | В.Д. | П.Р. | Д.К. | Е.О. | З.А. |
| **Спостереження (результати оцінювання)** | Студент психологічного факультету | 5 | 9 | 9 | 5 | 5 | 10 | 3 | 5 | 1,5 | 1,5 | 1,5 |
| | Студент педагогічного факультету | 2 | 0,5 | 3 | 1 | 2 | 3 | 4 | 2 | 9 | 9 | 9 |
| | Студент фізико-математичного факультету | 3 | 1 | 2 | 4 | 3 | 4 | 6 | 3,5 | 10 | 10 | 10 |
| | Студент природничого факультету | 9 | 2 | 1,5 | 3 | 4 | 2 | 7 | 7 | 9,5 | 9,5 | 9,5 |
| | Студент історичного факультету | 4 | 3 | 2,5 | 3,5 | 6 | 5 | 9 | 9 | 7 | 7 | 7 |
| | Студент філологічного факультету | 10 | 4 | 1 | 2 | 1 | 6 | 1,5 | 1 | 8 | 8 | 8 |
| | Психолог | 4,5 | 9,5 | 10 | 10 | 10 | 9 | 2,5 | 5,5 | 2 | 2 | 2 |
| | Соціальний педагог | 5,5 | 10 | 8 | 9,5 | 9,5 | 9,5 | 3,5 | 4 | 1 | 1 | 1 |
| | Вчитель | 8 | 5 | 5 | 7 | 5 | 7 | 5 | 3 | 6 | 6 | 6 |
| | Викладач | 6 | 6 | 7 | 9 | 8 | 5,5 | 8 | 8 | 5 | 5 | 5 |
| | Керівник | 3,5 | 7 | 6 | 8 | 7 | 4,5 | 10 | 10 | 3 | 3 | 3 |
| | Консультант | 7 | 8 | 4 | 6 | 6 | 8 | 2 | 6 | 4 | 4 | 4 |

При проведенні кластерного аналізу слід враховувати, що традиційно кластеризація проводиться по об'єктах (стовпчиках масиву даних).

## *3. ДЕРЕВОПОДІБНА КЛАСТЕРИЗАЦІЯ: ПОШУК ВІДСТАНЕЙ МІЖ ОБ'ЄКТАМИ, СТРАТЕГІЇ КЛАСТЕРИЗАЦІЇ ТА ІНТЕРПРЕТАЦІЯ*

Ми з'ясували, що для проведення кластерного аналізу перш за все необхідно визначитися із методом. Оберемо метод деревоподібної кластеризації як найбільш показовий і зрозумілий (більшість початківців вперше знайомляться саме з ним).

Після визначення з методом необхідно обрати міру відстані між об'єктами та обчислити ці відстані. Розглянемо, які бувають **міри відстаней** між об'єктами.

**Евклідова відстань (Euclidian distances).** Це найуживаніша міра відстані між об'єктами, яка являє собою геометричну відстань між об'єктами у багатомірному просторі. Формула для обчислення Евклідової відстані має такий вигляд:

$$L = \sqrt{\sum_i (x_i - y_i)^2}$$

Евклідова відстань обчислюється по вихідним не стандартизованим даним, а тому всі змінні повинні бути виміряні в одному масштабі (якщо це сантиметри, то всі змінні повинні бути виміряні в сантиметрах тощо).

**Квадрат Евклідової відстані (Squared Euclidian distances)**. Ця міра використовується тоді, коли хочуть на порядок збільшити значення відстаней між дуже віддаленими між собою об'єктами. Формула має такий вигляд:

$$L = \sum_i (x_i - y_i)^2$$

**Манхеттенівська відстань (відстань міських кварталів – City-block (Manhattan) distances)**. Ця міра у більшості випадків призводить до таких же результатів, як і Евклідова відстань, але зменшується вплив окремих великих різниць (викидів) через те, що відстань обчислюється по простим різницям координат:

$$L = \sum_i |x_i - y_i|$$

**Відстань Чебишева (Chebychev distance metric).** Використовується тоді, коли хочуть позначити два об'єкти як різні, якщо вони відрізняються якимсь одним виміром (однією координатою). Формула має такий вигляд:

$$L = \textit{Максимум} |x_i - y_i|$$

**Відсоток невідповідності (Percent disagreement).** Міра використовується у випадках, коли дані належать номінативній шкалі, і обчислюється за такою формулою:

$$L = \frac{\textit{Кількість}\, x_i \neq y_i}{i}$$

**1-коефіцієнт кореляції Пірсона (1-Pearson r).** Використовується у випадку, абсолютні значення та різниці між об'єктами несуттєві, а більш важливим є наявність зв'язку між ними – $L = 1 - r$. Ця міра чутлива лише до схожості профілів об'єктів. Призводить до результатів, близьких до факторного аналізу – кластери можуть наближатися до факторів. Формулу коефіцієнта кореляції Пірсона див. у пункті "Міри зв'язку" цього практикуму.

Оберемо Евклідову відстань, і проведемо обчислення. Для прикладу (табл. 5.3.2) обчислимо відстань між першим та другим об'єктами (досліджувані А.О. та О.В. з вихідного масиву даних табл. 20).

*Таблиця 5.3.2.*

*Обчислення відстані між об'єктами А.О. і О.В.*

| | Об'єкти | | $(x_i - y_i)$ | $(x_i - y_i)^2$ | $L = \sqrt{\sum_i (x_i - y_i)^2}$ |
|---|---|---|---|---|---|
| | А.О. | О.В. | | | |
| Студент психологічного факультету | 5 | 9 | -4 | 16 | |
| Студент педагогічного факультету | 2 | 0,5 | 1,5 | 2,25 | |
| Студент фізико-математичного факультету | 3 | 1 | 2 | 4 | |
| Студент природничого факультету | 9 | 2 | 7 | 49 | |
| Студент історичного факультету | 4 | 3 | 1 | 1 | |
| Студент філологічного факультету | 10 | 4 | 6 | 36 | |
| Психолог | 4,5 | 9,5 | -5 | 25 | $L = \sqrt{175{,}75} = 13{,}26$ |
| Соціальний педагог | 5,5 | 10 | -4,5 | 20,25 | |
| Вчитель | 8 | 5 | 3 | 9 | |
| Викладач | 6 | 6 | 0 | 0 | |
| Керівник | 3,5 | 7 | -3,5 | 12,25 | |
| Консультант | 7 | 8 | -1 | 1 | |
| | | | $\sum_i (x_i - y_i)^2$ | 175,75 | |

Провівши послідовні обчислення відстаней між усіма об'єктами будують **таблицю відстаней (distances matrix)** (табл. 5.3.3). Процес обчислень відстаней вручну є достатньо громіздким і тривалим, а тому доцільніше для цього використати комп'ютерні програми.

*Таблиця 5.3.3.*

*Таблиця відстаней між усіма об'єктами*

| | А.О. | О.В. | **М.П.** | К.П. | Р.А. | К.В. | В.Д. | П.Р. | **Д.К.** | **Е.О.** | **З.А.** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| А.О. | 0,00 | 13,26 | 14,81 | 13,41 | 13,51 | 11,51 | 14,37 | 13,63 | 12,82 | 12,82 | 12,82 |
| **О.В.** | 13,26 | 0,00 | **6,24** | 7,00 | 7,26 | 5,94 | 16,76 | 12,94 | **21,66** | **21,66** | **21,66** |
| **М.П.** | 14,81 | 6,24 | 0,00 | 6,82 | 6,69 | 8,02 | 14,94 | 12,37 | **20,41** | **20,41** | **20,41** |
| К.П. | 13,41 | 7,00 | 6,82 | 0,00 | 4,03 | 8,82 | 13,44 | 10,95 | 19,62 | 19,62 | 19,62 |
| Р.А. | 13,51 | 7,26 | 6,69 | 4,03 | 0,00 | 8,86 | 12,35 | 9,04 | 18,89 | 18,89 | 18,89 |
| К.В. | 11,51 | 5,94 | 8,02 | 8,82 | 8,86 | 0,00 | 16,43 | 13,83 | 18,69 | 18,69 | 18,69 |
| В.Д. | 14,37 | 16,76 | 14,94 | 13,44 | 12,35 | 16,43 | 0,00 | 6,61 | 12,85 | 12,85 | 12,85 |
| П.Р. | 13,63 | 12,94 | 12,37 | 10,95 | 9,04 | 13,83 | 6,61 | 0,00 | 15,97 | 15,97 | 15,97 |
| Д.К. | 12,82 | 21,66 | 20,41 | 19,62 | 18,89 | 18,69 | 12,85 | 15,97 | 0,00 | 0,00 | 0,00 |
| Е.О. | 12,82 | 21,66 | 20,41 | 19,62 | 18,89 | 18,69 | 12,85 | 15,97 | 0,00 | 0,00 | 0,00 |
| З.А. | 12,82 | 21,66 | 20,41 | 19,62 | 18,89 | 18,69 | 12,85 | 15,97 | 0,00 | 0,00 | 0,00 |

Чим менше значення у комірці таблиці, тим ближче знаходяться між собою відповідні об'єкти. Так, з табл. 5.3.3 видно, що найбільша відстань – між студентами О.В. та Д.К., Е.О., З.А. (L=21,66), а також між М.П. та Д.К., Е.О., З.А. (L=20,41). Очевидно, що вже на цьому етапі аналіз можна зробити висновок про наявність принаймні двох груп студентів, які різняться за свої ставленням до навчання та майбутньої професії – група О.В., М.П., та група Д.К., Е.О., З.А. Подивимося на відстані всередині кожної з груп. Дійсно, виявляється, що дуже близькі між собою студенти О.В. та М.П. (L=6,24), а також ідентичні між собою (знаходяться на нульовій відстані) студенти Д.К., Е.О. та З.А. (L=0).

Можна продовжити цей аналіз, шукаючи близькі та віддалені групи студентів, орієнтуючись виключно на таблицю відстаней, а можна продовжити деревоподібний кластерний аналіз, використавши одну із **стратегій кластеризації**. Стратегії кластеризації являють собою правила об'єднання об'єктів (змінних) у кластери. Вони переглядають таблицю схожостей об'єктів, і на кожному кроці послідовно об'єднують пару найбільш схожих

об'єктів (змінних чи кластерів). Завершується процес утворенням одного кінцевого великого кластера, який включає в себе всі об'єкти. Основна різниця між стратегіями – це спосіб вимірювання відстаней. Однак, тут уже мова йдеться не про безпосередні відстані між об'єктами – на першому кроці кластеризації кожен об'єкт являє собою окремий кластер, і відстані між ними визначаються обраною мірою (Евклідова відстань, відстань Чебишева тощо). Мова вже йде про той випадок, коли декілька об'єктів зв'язуються разом – як тепер визначити відстані між утвореними кластерами? Розглянемо для цього найпоширеніші **стратегії кластеризації**.

**Стратегія найближчого сусіда (Nearest neighbor) або стратегія одиночного зв'язок (Single lincage)**. Тут відстань між двома кластерами визначається як відстань між двома найближчими об'єктами (найближчими сусідами). Стратегія ніби нанизує об'єкти один на один, і в результаті кластери представляються у вигляді довгих "ланцюжків". Стратегія пов'язує два кластери разом, коли будь-які два об'єкти в цих кластерах ближче один до одного, ніж усі інші.

**Стратегія найвіддаленішого сусіда (Furthest neighbor) або стратегія повного зв'язку (Complete lincage)**. При використанні цієї стратегії відстань між кластерами визначається найбільшою відстанню між двома об'єктами з різних кластерів (між найвіддаленішими сусідами). Стратегія добре працює, коли об'єкти реально належать різним класам. Якщо є природним типом кластерів в отриманих даних є ланцюжки, то ця стратегія є непридатною. Стратегія утворює в основному "кущі" об'єктів.

**Стратегія незваженого попарного середнього (Unweighted pair-group average).** Відстань між двома кластерами визначається як середня відстань між всіма парами об'єктів у них. Метод ефективний випадку реального об'єднання об'єктів як у "кущі", так і в "ланцюжки".

**Стратегія зваженого попарного середнього (Weighted pair-group average)**. Стратегія відрізняється від попередньої тим, що при обчисленнях розмір відповідного кластера (кількість об'єктів, які він містить) використовується в якості вагового коефіцієнта. Тому цю стратегію використовують тоді, коли передбачають появу кластерів нерівного розміру.

**Стратегія незваженого попарного центроїдного усереднення (Unweighted pair-group centroid)**. В цій стратегії відстань між кластерами визначається як відстань між їх центрами ваги.

**Стратегія зваженого попарного центроїдного усереднення (Weighted pair-group centroid) або стратегія медіанної кластеризації (Median clustering)**. Стратегія аналогічна до попередньої, з тією різницею, що при обчисленнях враховуються самі ваги для врахування розмірів кластерів.

**Стратегія Варда (Ward's method)**. Ця стратегія суттєво відрізняється від попередніх, оскільки використовує методи дисперсійного аналізу для оцінки відстаней між кластерами. Ця стратегія мінімізує суму квадратів (SS) для двох гіпотетичних кластерів, які можуть бути сформовані на кожному кроці процесу кластеризації. Метод вважається ефективним, але намагається створювати кластери малого розміру.

Які ж загальні властивості описаних стратегій? Якщо уявити таблицю схожостей як багатомірний простір, а об'єкти – як точки цього простору, то можна описати, що роблять з точками цього простору розглянуті нами стратегії.

Перший тип стратегій – **стискаючі**. Стратегії цього типу ніби стискають простір об'єктів, зменшуючи відстані між усіма групами даних. Коли черговий об'єкт піддається такій стратегії, він, швидше за все, буде віднесений до вже існуючого кластера, ніж стане джерелом утворення нового. Сюди можна віднести *стратегію найближчого сусіда*.

Другий тип стратегій – **розширюючі**. Ці стратегії ніби розширюють простір об'єктів, збільшуючи відстані між ними. Точки ніби розступаються, утворюючи дрібніші, але чіткіші групи (кластери). Ці стратегії створюють кластери гіперсферичної форми, приблизно однакові за об'ємом. До цього типу належать *стратегія найвіддаленішого сусіда* та *стратегія Варда*.

Третій тип стратегій – **зберігаючі**. Методи, які належать до цього типу, залишають вихідний простір об'єктів без змін. Це стратегії *зваженого* та *незваженого попарного середнього*, а також *зваженого* та *незваженого попарного центроїдного усереднення*.

Давайте до даних таблиці відстаней застосуємо **стратегію найближчого сусіда**.

Найменша відстань у таблиці – між об'єктами Д.К., Е.О., З.А. (L=0). Очевидно, вони утворюють **перший** кластер. Друга по величині відстань – між об'єктами К.П. і Р.А. (L=4,03). Їх варто об'єднати у **другий** кластер.

**Третій** кластер утворять об'єкти О.В. і К.В. (L=5,94). До третього кластера також слід приєднати об'єкт М.П., оскільки його відстань від О.В. рівна L=6,25. Таким чином, **четвертий кластер** буде мати свою структуру – первинний кластер з об'єктів О.В. і К.В., і вторинний кластер, який включає в себе ще об'єкт М.П. **П'ятий кластер** буде утворений об'єктами В.Д. і П.Р. (L=6,61).

На цьому етапі кластеризації майже кожен об'єкт увійшов до якогось кластера, і тепер слід зв'язати між собою вже утворені кластери. Наступна за величиною є відстань L=6,69 – це відстань між об'єктами М.П. (четвертий кластер) і Р.А. (другий кластер). Очевидно, що другий та третій кластери слід об'єднати у кластер вищого порядку – **шостий кластер**. На наступному кроці варто вже шукати відстані між п'ятим і шостим кластерами. Нею буде L=9,04 між об'єктами П.Р. і Р.А. і утвориться **сьомий кластер**. Далі аналізуючи відстані можна побачити, що до сьомого кластеру приєднується об'єкт А.О. з мінімальною відстанню L=11,51 від об'єкта К.В. з сьомого кластера. Так утворюється **восьмий кластер**. І нарешті **дев'ятий кластер** утворений об'єднанням першого та восьмого з відстанню L=12,81 (об'єкти Д.К., Е.О., З.А та А.О.)

Можна узагальнено представити описаний процес кластеризації у **таблиці об'єднань** (amalgamation schedule) – табл. 5.3.4.

*Таблиця 5.3.4.*

*Таблиця об'єднань*

| Крок | Відстані між найближчими сусідами | ОБ'ЄКТИ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 1 | 0,00 | Д.К | Е.О. | | | | | | | | | |
| 2 | 0,00 | Д.К. | Е.О. | З.А. | | | | | | | | |
| 3 | 4,03 | К.П. | Р.А. | | | | | | | | | |
| 4 | 5,94 | О.В. | К.В. | | | | | | | | | |
| 5 | 6,25 | О.В. | К.В. | М.П. | | | | | | | | |
| 6 | 6,61 | В.Д. | П.Р. | | | | | | | | | |
| 7 | 6,69 | О.В. | К.В. | М.П. | К.П. | Р.А. | | | | | | |
| 8 | 9,04 | О.В. | К.В. | М.П. | К.П. | Р.А. | В.Д. | П.Р. | | | | |
| 9 | 11,51 | А.О. | О.В. | К.В. | М.П. | К.П. | Р.А. | В.Д. | П.Р. | | | |
| 10 | 12,82 | А.О. | О.В. | К.В. | М.П. | К.П. | Р.А. | В.Д. | П.Р. | Д.К. | Е.О. | З.А. |

В наведеній таблиці представлено кожен описаний вище крок, біля кожного кроку стоїть відстань між найближчими сусідами, а справа – виділені на

кожному з кроків кластери. На цьому процедура кластеризації завершена, однак, можна побудувати графічне зображення отриманих кластерів – **кластерне дерево (hierarchical tree plot)** – рис. 5.3.4.

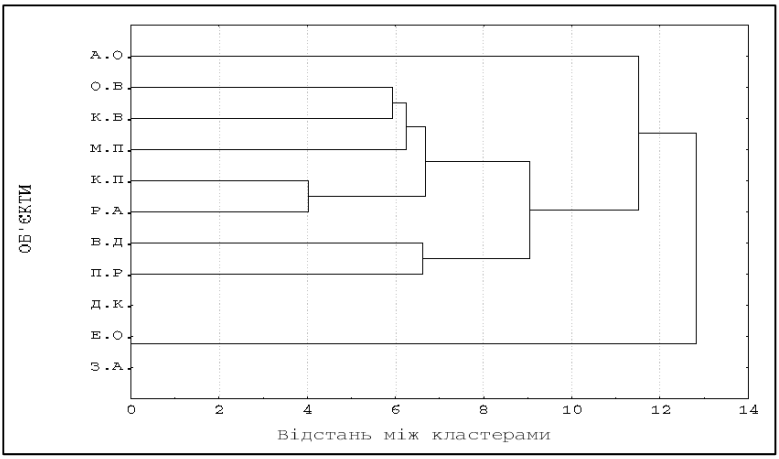



*Рис. 5.3.4. Кластерне дерево*

Дійсно, тепер наочно і чітко видно утворену кластерну структуру – кластери студентів, що мають однакове ставлення до своїх рольових позицій і до студентів інших факультетів. Перша група студентів – Д.К., В.О., З.А., друга група – К.П. та Р.А., третя група – О.В. та К.В., до якої приєднується М.П., четверта група – В.Д. та П.Р. Далі від усіх знаходиться студент А.О., який не входить у жоден з первинних кластерів, а значить, найбільше відрізняється від усіх інших.

Останній етап кластерного аналізу – інтерпретація. Інтерпретуючи кластерні дерева, слід намагатися в межах кожного із виділених кластерів знайти певний смисловий інваріант, який би пояснив причину об'єднання об'єктів у цей кластер. Для цього можна використати: 1) описові статистики об'єктів, 2) описові статистики спостережень (значень об'єктів), 3) вихідний масив даних. В результаті такого аналізу необхідно знайти певну психологічну характеристику, яка одночасно зможе слугувати назвою кластера.

Так, до першого кластера потрапили студенти Д.К., Е.О., З.А. Звернувшись до таблиці вихідних даних (табл. 5.3.1), можна побачити що це студенти, які

найпозитивніше ставляться до студентів інших факультетів, і одночасно – не дуже позитивно до майбутніх рольових позицій. Студенти другого кластеру (К.П. та Р.А.) навпаки, позитивно ставляться до майбутніх професійних рольових позицій, і не дуже позитивно – до студентських колективів. Студенти третього кластеру (О.В., К.В., М.П.) позитивно ставляться до професійних позицій психолога та до студентів своєї спеціальності. Студенти четвертого кластеру (В.Д. та П.Р.) найпозитивніше ставляться до професійних позицій педагога та до студентів такої ж орієнтації. Ставлення студента А.О. доволі таки важко визначити – очевидно, його професійна ідентифікація ще не відбулася.

Узагальнюючи результати проведеного аналізу, можна припустити наявність такої основи для класифікації студентів, як професійна ідентифікація. Студенти, у яких **несформована професійна ідентифікація** поділяються на два класи – ті, які більш позитивно ставляться до студентського життя (Д.К., Е.О., З.А.), і ті, які позитивно ставляться до майбутньої діяльності, не диференціюючи її на види (К.П., Р.А.). Студенти, у яких **професійна ідентифікація сформована,** також діляться на дві класи – орієнтовані на психологічну діяльність (О.В., К.В., М.П.) та орієнтовані на педагогічну діяльність (В.Д., П.Р.)

| | | |
|---|---|---|
| **Несформована** професійна ідентифікація | позитивно ставляться до студентського життя | Д.К., Е.О., З.А. |
| | позитивно ставляться до майбутньої діяльності, не диференціюючи її на види | К.П., Р.А. |
| **Сформована** професійна ідентифікація | орієнтовані на психолого-педагогічну діяльність | О.В., К.В., М.П. |
| | орієнтовані на керівну діяльність | В.Д., П.Р. |

Очевидно, що не всі аспекти емпірично виявленої класифікації повністю накладаються на структуру кластерного дерева, але, власне, у психології такої абсолютно точності навряд чи коли вдасться досягти. Зроблені узагальнення можуть стати джерелом подальших більш ґрунтовних досліджень і висновків з них. Кластерний же аналіз на цьому етапі дослідницької роботи свою задач виконав – виявив структуру студентської групи на основі рівня професійної ідентифікації.

## *4. ПРЕДСТАВЛЕННЯ РЕЗУЛЬТАТІВ КЛАСТЕРНОГО АНАЛІЗУ.*

Більшість результатів досліджень науковці представляють у формі наукових звітів, наукових робіт (дисертаційних, конкурсних), наукових статей та тез доповідей. Це робиться з метою донесення до широкого загалу результатів власних досліджень та висновків. Звичайно, ж, що основною вимогою до робіт такого типу є доказовість, переконливість та зрозумілість.

Якщо йдеться мова про представлення результатів кластерного аналізу, то слід визначитися із наборами показників, які варто включати в текст та у додатки до роботи. Розглянемо це питання з точки зору способу представлення.

**Науковий звіт.** Готується по завершенню виконання певного етапу наукової роботи, і подається замовнику цієї роботи. Науковий звіт є технічною документацією, і його оформлення регулюється вимогами ДСТУ 3008-95. Відповідно до вимог, у науковому звіті повині бути представлені ВСІ числові результати дослідження. Основна суть цих вимог – щоб будь-хто міг взявши ваші дані провести повторні обчислення і отримати ті ж самі результати. Якщо мова йдеться про кластерний аналіз, то необхідно навести *таблиці вихідних даних, таблиці відстаней, таблиці об'єднань і кластерні графіки*. При цьому *таблиці вихідних даних* та *таблиці об'єднань* можна навести у додатках, а *таблиці відстаней* та *кластерні графіки* – у тексті звіту. З метою економії місця в основному тексті (деякі замовники обмежують обсяг звіту) в додатки можна винести і *таблиці відстаней*.

**Наукова робота.** Готується по завершенню цілісного наукового дослідження. Наукові роботи можуть бути дисертаційними та конкурсними. Оформлення *дисертаційної роботи* регулюється вимогами ДСТУ 3008-95 та ВАК України. Відповідно до вимог, у дисертаційній роботі теж мають бути представлені ВСІ числові результати дослідження. Вимоги до представлення результатів кластерного аналізу – ті ж самі, що і у попередньому пункті. Вимоги до *конкурсної роботи* диктуються організацією, яка оголошує конкурс наукових робіт, однак, у більшості випадків, вони залишаються стандартними, визначеними держстандартами.

**Наукова стаття.** Вимоги до оформлень наукових статей формулюються редакторати тих чи інших видань. Однак, якщо

говорити про результати кластерного аналізу, то для їх представлення у статті достатньо *кластерного графіка*, а для більшої доказовості – *таблиці відстаней*.

**Тези доповіді.** В тезах доповіді не прийнято представляти ні графічної, ні табличної інформації. Для представлення результатів кластерного аналізу зазвичай описують гіпотезу дослідження, метод та отримані результати.

У всіх випадках обов'язково слід вказати:

1) метод кластеризації;
2) міру відстані між об'єктами;
3) стратегію кластеризації.

*Наприклад: "Кластерний аналіз було проведено методом деревоподібної кластеризації, мірою відстані слугувала відстань Чебишева, а кластеризація проводилася з допомогою стратегії Варда"*

## *5. КЛАСТЕРНИЙ АНАЛІЗ ЗА МЕТОДОМ В.Ю. КРИЛОВА ТА Т.В. ОСТРЯКОВОЇ.*

В.Ю. Крилов та Т.В. Острякова запропонували декілька нових цікавих стратегій кластеризації, які досить просто можна реалізовувати вручну. Цікаво також те, що ґрунтуються ці стратегії на виявлених Л.С. Виготським етапах розвитку понять. Це такі етапи:

1. Асоціативний комплекс;
2. Ланцюговий комплекс;
3. Асоціативно-ланцюговий комплекс;
4. Комплекс-колекція тощо.

Виявляється, на основі уявлень Виготського можна легко будувати кластери відповідної структури. Розглянемо основі стратегії кластеризації.

1. **Простий асоціативний кластер**. Для побудови асоціативного кластера обирається один елемент, який буде ядром. Далі підбираються усі елементи, які мають найближчу відстань до ядра – рис. 5.3.5. Таким чином, ядер може бути стільки, скільки є елементів у матриці. Для вибору того чи іншого елементу ядром, слід враховувати зміст задачі, які вирішує дослідник, особливості матриці відстаней...

*Нехай нас у матриці відстаней (табл. 22) цікавить студент **О.В.** (можливо, він найбільш активний або успішний у навчанні).*

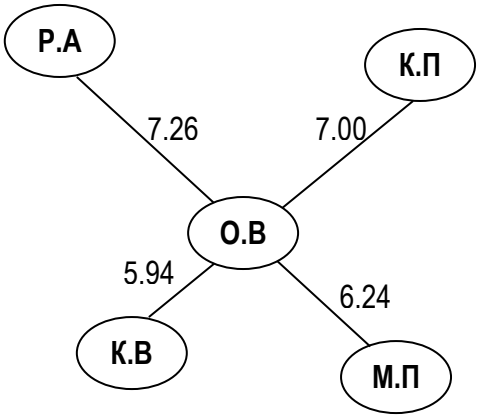



*Рис. 5.3.5. Простий асоціативний кластер. Ядро – О.В.*

2. **Складний асоціативний кластер**. Від простого відрізняється тим, що в якості ядра виступають декілька елементів. Для побудови такого кластеру обирають складне ядро (як правило, декілька близьких між собою елементів), потім для кожного елемента підбирають його найближчих сусідів. Після цього обирається певна мінімальна відстань і для залишаються у кластері лише ті елементи, які мають цю мінімальну відстань від ядра (рис. 5.3.6).

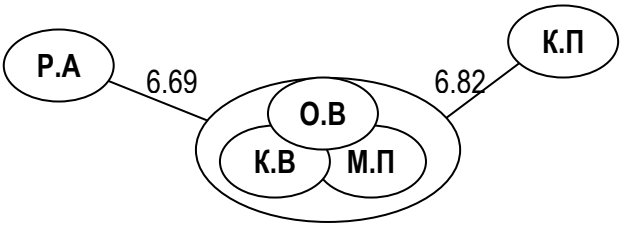



*Рис. 5.3.6. Складний асоціативний кластер. Ядро – О.В., К.В., М.П.*

3. **Ланцюговий кластер**. Для його побудови обирається з матриці один початковий елемент (№1), далі шукається найближчий до

нього елемент (№2) і т.д. Побудова ланцюгового кластеру завершується тоді, коли вичерпані усі близькі елементи – рис. 5.3.7.

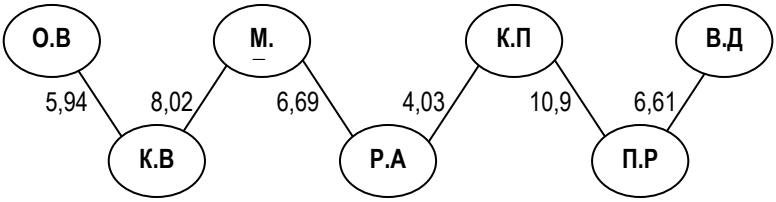



*Рис. 5.3.7. Ланцюговий кластер*

Побудова асоціативного та ланцюгового кластерів переслідує різні цілі. При побудові асоціативного кластера метою є пошук елементів, найближчих до ядра, а при побудова ланцюгового кластера – пошук послідовного зв'язку початкового елементу з усіма іншими елементами матриці.

*Прикладом використання ланцюгового кластера може бути дослідження циркуляції чуток на приватному підприємстві. Для цього слід створити матрицю, у якій мірою близькості між працівниками організації є кількість щоденних контактів між ними. Тоді побудувавши ланцюгові кластери для кожного працівника можна буде виявити найімовірніші шляхи розповсюдження чуток.*

Однак, було б цікаво розробити таку стратегію, яка б поєднувала у собі переваги як асоціативного, так і ланцюгового кластерів. Таким буде асоціативно-ланцюговий кластер.

4. **Асоціативно-ланцюговий кластер**. Для його побудови спочатку використовується стратегія побудови асоціативного кластера, а потім до кожного елемента цього кластера використовується стратегія побудови ланцюгового кластера.

## ЗАВДАННЯ ДЛЯ САМОКОНТРОЛЮ

***Тестові завдання:***

1. Факторний аналіз використовується для:
   а) виявленню факторів впливу;
   б) зменшення розмірності масиву даних;
   в) факторного диференціювання інтегральних ознак.
2. При підготовці даних до факторного аналізу бажаним є:
   а) щоб кількість рядків була більшою за кількість стовпчиків;
   б) щоб кількість стовпчиків була більшою за кількість рядків;
   в) щоб кількість стовпчиків дорівнювала кількості рядків.
3. Факторні ваги - це:
   а) коефіцієнти кореляції кожної із аналізованих змінних із кожним з виділених факторів;
   б) кількісні значення зв'язку виділених факторів з об'єктами;
   в) дисперсії, які пояснюються факторами.
4. Для оцінки значимості фактора використовують:
   а) факторні навантаження;
   б) факторні ваги;
   в) власні значення факторів.
5. Осмислена інтерпретація результатів кластерного аналізу можлива за умови:
   а) що вона ґрунтується на теоретичній моделі;
   б) що вона ґрунтується на багатьох повторних дослідженнях;
   в) що вона ґрунтується на використанні методів перевірки статистичної значимості.
6. Не існує такого методу кластерного аналізу:
   а) методу деревоподібної кластеризації;
   б) методу двовходового об'єднання;
   в) методу імовірнісних кластерів.
7. Для пошуку відстаней між об'єктами у кластерному аналізі використовують:
   а) відстань Спірмена;
   б) відстань Йокогами;
   в) відстань Чебишева.
8. При описі результатів кластерного аналізу у статті, в її текст необхідно включити:
   а) кластерні дерева;
   б) кластерні дерева та таблиці об'єднань;
   в) кластерні дерева, таблиці об'єднань та таблиці відстаней.
9. Для проведення багатомірного шкалювання необхідно:
   а) мати таблицю схожості об'єктів;
   б) провести перевірку нормальності розподілу;

в) перевірити статистичну значимість об’єктів.

10. Надійність результатів багатомірного шкалювання визначається таким показником:
    а) депресія;
    б) стрес;
    в) невроз.
11. Факторний аналіз використовується для:
    а) пошуку прихованих змінних;
    б) виявленню факторів впливу;
    в) факторного диференціювання інтегральних ознак.
12. Вимогою до підготовки даних до факторного аналізу є:
    а) представлення масиву даних у вигляді двохвимірної матриці;
    б) представлення масиву даних у вигляді N-вимірної матриці;
    в) представлення масиву даних будь-якому вигляді.
13. До основних результатів факторного аналізу відносять:
    а) факторні навантаження, факторні ваги, власні значення факторів;
    б) факторні навантаження, факторні ваги, власні значення факторів, таблицю об’єднань;
    в) факторні навантаження, факторні ваги, власні значення факторів, характеристичну матрицю.
14. Власні значення факторів - це:
    а) коефіцієнти кореляції кожної із аналізованих змінних із кожним з виділених факторів;
    б) кількісні значення зв’язку виділених факторів з об’єктами;
    в) дисперсії, які пояснюються факторами.
15. Метою кластерного аналізу є:
    а) організація отриманих у дослідженні даних у наглядні структури;
    б) пошук значимих кластерів – детермінант досліджуваних процесів;
    в) кластеризація.
16. Не існує такого методу кластерного аналізу:
    а) методу деревоподібної кластеризації;
    б) методу К-середніх;
    в) методу квазісиметричної кластеризації.
17. Для пошуку відстаней між об’єктами у кластерному аналізі використовують:
    а) Евклідову відстань;
    б) Піфагорову відстань;
    в) регресійну лінію.
18. Результати кластерного аналізу наочно відображаються у:
    а) кластерному дереві;
    б) кластерній рослині;
    в) кластерній піраміді.
19. Мета багатомірного шкалювання - це:

а) ефективне розміщення об’єктів у просторі певної кількості ознак;
б) ефективне розміщення об’єктів у факторних структурах;
в) зниження розмірності масиву даних.

20. При виборі мірності простору при багатомірному шкалювання слід врахувати:
а) інтерпретованість результату;
б) нічого не треба враховувати;
в) критерій Кайзера.

21. Факторний аналіз використовується для:
а) виявленню факторів впливу;
б) групування, класифікації та візуалізації даних;
в) факторного диференціювання інтегральних ознак.

22. Вимогою до підготовки даних до факторного аналізу є:
а) представлення масиву даних у вигляді N-вимірної матриці;
б) занесення до стовпчиків таблиці аналізованих змінних, а до рядків – значень цих змінних;
в) представлення масиву даних будь-якому вигляді.

23. Факторні навантаження - це:
а) коефіцієнти кореляції кожної із аналізованих змінних із кожним з виділених факторів;
б) кількісні значення зв’язку виділених факторів з об’єктами;
в) дисперсії, які пояснюються факторами.

24. Для візуалізації результатів факторного аналізу використовуються:
а) факторні поля;
б) факторні діаграми;
в) факторні кластери.

25. До кластерного аналізу незастосовні:
а) процеси перевірки статистичної значимості;
б) верифікація результатів у повторному дослідженні;
в) методи перевірки надійності та валідності.

26. Не існує такого методу кластерного аналізу:
а) методу деревоподібної кластеризації;
б) методу N-мірної кластерної ієрархізації;
в) методу двовходового об’єднання.

27. Для пошуку відстаней між об’єктами у кластерному аналізі використовують:
а) Манхеттенівську відстань;
б) Московську відстань;
в) алгебраїчну суму.

28. Для пошуку відстаней між кластерами використовують:
а) стратегії кластеризації;
б) факторний аналіз;
в) міри відстані.

29. Багатомірного шкалювання є альтернативою:
    а) кластерному аналізу;
    б) факторному аналізу;
    в) канонічному аналізу.
30. При виборі мірності простору при багатомірному шкалювання слід врахувати:
    а) критерій Цезаря;
    б) критерій кам'янистого насипу;
    в) критерій Кайзера.

***Теоретичні питання:***

1. Поняття про факторні гіпотези.
2. Типи факторних планів.
3. Результати факторних експериментів.
4. Задача двофакторного дисперсійного аналізу.
5. Обмеження двофакторного дисперійсного аналізу.
6. Підготовка даних до дисперсійного аналізу.
7. Проведення двофакторного дисперійсного аналізу. Критерій Фішера F.
8. Мета факторного аналізу.
9. Підготовка даних до факторного аналізу.
10. Інтерпретація результатів факторного аналізу.
11. Задача про кількість факторів у факторному аналізі.
12. Особливості факторного аналізу.
13. Місце факторного аналізу в структурі експерименту.
14. Загальна мета багатомірного шкалювання. Підготовка даних до БШ.
15. Логіка багатомірного шкалювання.
16. Проблема розмірності при багатомірному шкалюванні.
17. Інтерпретація результатів багатомірного шкалювання.
18. Багатомірне шкалювання та факторний аналіз: порівняльний аналіз.
19. Мета та методи кластерного аналізу.
20. Підготовка даних до кластерного аналізу.
21. Деревоподібна кластеризація: пошук відстаней між об'єктами.
22. Деревоподібна кластеризація: стратегії кластеризації та інтерпретація.
23. Представлення результатів кластерного аналізу.
24. Кластерний аналіз за методом В.Ю. Крилова та Т.В. Острякової.
25. Факторний аналіз та кластерний аналіз: порівняльний аналіз.

***Творчі завдання:***

1. Розробіть структуру дослідження, у якому використовується двофакторний дисперсійний аналіз. Змоделюйте його результати та зробіть інтерпретацію.

2. Розробіть структуру дослідження, у якому використовується факторний аналіз. Змоделюйте його результати на рівні факторного поля та зробіть інтерпретацію.
3. Розробіть структуру дослідження, у якому використовується кластерний аналіз. Змоделюйте його результати на рівні кластерного дерева та зробіть інтерпретацію.
4. Розробіть структуру дослідження, у якому використовується багатомірне шкалювання. Змоделюйте його результати на рівні двохмірного простору та зробіть інтерпретацію.

# ПЕРЕДМОВА

Лабораторний практикум з курсу “Математичні методи у психології” розроблений для студентів психологічних спеціальностей, які навчаються в умовах *кредитно-модульної системи навчання*.

Практикум побудований так, щоб максимально активізувати самостійну роботу студентів. У зв’язку з цим усі лабораторні роботи вимагають попереднього проведення самостійних досліджень за наперед заданою структурою. Однак, якщо студент може запропонувати свою тему та структуру дослідження, – він може побудувати роботу за власними результатами. Крім можливості використовувати результати власних досліджень у роботі на заняттях, можна самостійно обирати темп роботи.

Практикум розрахований на використання на заняттях пакету прикладних статистичних програм **STATISTICA 5.5**[20]. У зв’язку з певними особливостями цього пакету в роботі слід дотримуватися таких правил:

- кожен студент має працювати із закріпленим за ним комп’ютером;
- для роботи у папці ***Statlab*** слід створити папку, названу своїм прізвищем ***латинськими літерами***;
- у своїй папці на кожне лабораторне заняття слід створити папки, названі номерами цих занять (кожному заняттю відповідатиме папка зі своїм номером, наприклад ***Lab1, Lab2, Lab3, …***);
- всі файли, створені під час заняття, повинні зберігатися у папці з відповідним номером; а також обов’язково – на дискетах 3,5”

Для допуску до підсумкової модульної роботи необхідно здати усі лабораторні заняття відповідного модуля. Кожне лабораторне заняття здається викладачу в електронному вигляді (на диску 3,5” або CD-R/RW). Студенти можуть дотримуватися індивідуального темпу здачі лабораторних робіт.

---

[20] *© StatSoft, Inc. (1999). STATISTICA for Windows [Computer program manual]. Tulsa, OK: StatSoft, Inc., 2300 East 14th Street, Tulsa, OK 74104, phone: (918) 749-1119, fax: (918) 749-2217, email: info@statsoft.com, WEB: http://www.statsoft.com*

## ЛАБОРАТОРНЕ ЗАНЯТТЯ №1:
## ОПИСОВІ СТАТИСТИКИ

***Теоретична частина:***

1. **Способи представлення статистичних даних**
2. **Міри центральної тенденції**
   - Мода.
   - Медіана та її обчислення.
   - Середнє арифметичне: обчислення та властивості.
   - Інтерпретація мір центральної тенденції.
   - Вибір міри центральної тенденції.
3. **Міри мінливості**
   - Призначення мір мінливості.
   - Розмах та коефіцієнт осциляції.
   - Дисперсія: обчислення та властивості.
   - Стандартне відхилення та коефіцієнт варіації.

**Література для підготовки до заняття:**


1. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976. – 495 с.
2. Дружинин В.Н. Экспериментальная психология. – СПб.: Питер, 2000. – 320 с.
3. Електронний підручник з статистики StatSoft ***http://www.statsoft.ru/home/textbook/default.htm***
4. Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО “Речь”, 2000. – 350 с.
5. Солсо Р.Л., Джонсон Х.Х., Бил М.К. Экспериментальная психология: практический курс. – СПб.: прайм-ЕВРОЗНАК, 2001. – 528 с.
6. Теорія статистики: Навчальний посібник /Вашків П.Г., Пастер П.І., Сторожук В.П., Ткач Є.І. – К.: Либідь, 2001. – 320 с.


***Практична частина:***

**Завдання 1. Проведення психологічного дослідження *(до початку заняття)***

*Проведіть дослідження динаміки емоційного стану студентів протягом тижня.*

**Завдання 2. Побудова статистичних графіків (STATISTICA 5.5.)**

*За результатами проведеного дослідження побудуйте статистичній графіки. Результати роботи збережіть у своїй папці та на дискеті 3,5”.*

**Завдання 3. Обчислення мір центральної тенденції та мір мінливості**

*За допомогою програми Statistica обчисліть міри центральної тенденції та міри мінливості для отриманих у дослідженні даних.*

## ХІД РОБОТИ: ЗАВДАННЯ 1

## ДОСЛІДЖЕННЯ ДИНАМІКИ ЕМОЦІЙНОГО СТАНУ СТУДЕНТІВ ПРОТЯГОМ РОБОЧОГО ТИЖНЯ

**Об'єкт дослідження**: емоційно-вольова сфера студентів.

**Предмет дослідження**: динаміка емоційного стану студентів протягом робочого тижня.

**Гіпотеза**: зміни емоційного стану студентів протягом робочого тижня мають вигляд інвертованої U-подібної кривої.

**Методи дослідження**: методика „Емоційний градусник",

**Методи обробки даних**: описові статистики (середнє арифметичне, стандартне відхилення, асиметрія, ексцес), статистичні графіки та діаграми.

**Характеристики вибірки**: 10-15 студентів 1-5 курсів.

**Структура дослідження**:

***До лабораторного заняття:***

Знайдіть майбутніх учасників дослідження. Роздайте їм бланки методики "Емоційний градусник" та дайте інструкції з подальшої роботи (*див. Додаток 1*). По завершенню дослідження (через тиждень) зберіть заповнені бланки.

***На лабораторному занятті:***

1. Зведіть отримані сирі дані у статистичну таблицю (*див. Завдання 2*):

| | Понеділок | Вівторок | Середа | Четвер | П'ятниця |
|---|---|---|---|---|---|
| Учасник 1 | | | | | |
| Учасник 2 | | | | | |
| ... | | | | | |
| Учасник 9 | | | | | |

2. Зобразіть на одному полі графіки динаміки емоційного стану усіх учасників дослідження (*див. Завдання 2 – побудова графіків з основної таблиці даних*). Чи спостерігається там якась тенденція?
3. Обчисліть в межах кожного дня середні показники емоційного стану учасників дослідження, а також міри мінливості (дисперсію, стандартне відхилення, асиметрію та ексцес) (*див. Завдання 3*).

4. Побудуйте графік динаміки середнього показника емоційного стану учасників дослідження (*див. Завдання 2 – побудова графіків з таблиці результатів*)
5. Зробіть висновки про особливості зміни емоційного стану студентів протягом робочого тижня. Чи підтвердилася гіпотеза дослідження?

***Після лабораторного заняття:***

Підготуйте та здайте звітні матеріали. Поінформуйте учасників дослідження про їх результати.

## ХІД РОБОТИ: ЗАВДАННЯ 2

***Створення та збереження файлу даних***

1. Для запуску програми STATISTICA слід двічі клацнути мишею по її ярлику на Робочому столі або ввійти в меню ПУСК, як показано на рис. 1.1.

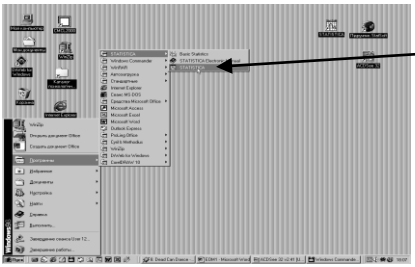



*Рис.1.1. Запуск програми STATISTICA*

2. Після запуску програми з'явиться *Вікно Модулів (STATISTICA Module Switcher)*, зображене на рис.1.2.

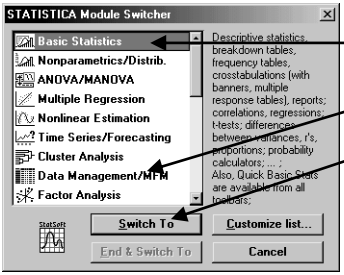



*Рис.1.2. Вікно модулів*

3. Для створення файлу даних запустіть *Модуль управління даними (Data Managemen/MFM).* Після запуску *Модуля* з'явиться вікно *Управління даними (Data Management),* зображене на рис. 1.3.

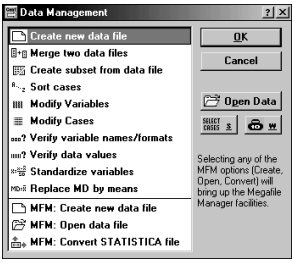



*Рис.1.3. Вікно Управління даними*

4. Оберіть пункт *Створити новий файл даних (Create new data file)* та натисніть кнопку *ОК.* У новому діалоговому вікні (рис. 1.4.)

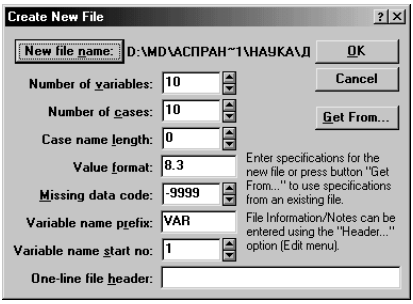



*Рис.1.4. Вікно Створення нового файлу*

у полі *Number of variables* введіть кількість стовпчиків (змінних), а у полі *Number of cases* введіть кількість рядків (випадків). Потім натисніть кнопку *New file name* та збережіть файл даних у папку ***D:\Statlab\(Прізвище)\Lab1*** під назвою ***data***. Після цього натисніть *ОК.*

5. **Файл даних створено (рис. 1.5.)!!!** Тепер можна перейти до занесення даних до створеного файлу. НЕ ЗАБУВАЙТЕ ПЕРІОДИЧНО НАТИСКУВАТИ КНОПКУ **"СОХРАНИТЬ"**.

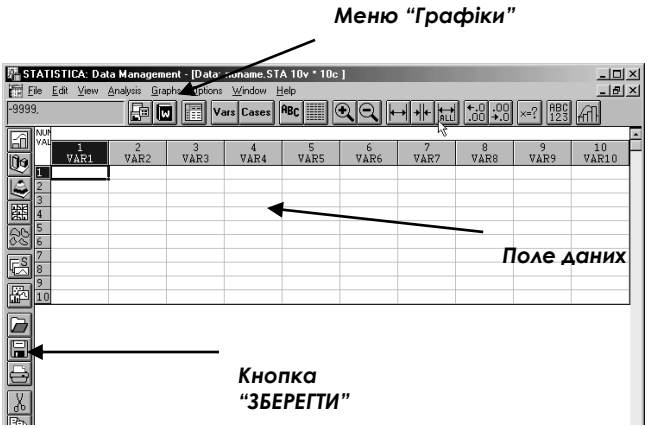



*Рис.1.5. Файл даних*

6. Введіть свої дані до поля даних (якщо Ви вже їх маєте). Збережіть введені дані. ***Краще лишній раз зберегти, ніж потім все знову набирати***.
7. Назвіть свої рядки та стовпчики. Для **перейменування стовпчиків** (змінних) слід двічі клацнути мишею по кнопці VAR1(перша змінна поля даних), та у полі *Ім’я* (*Name)* вікна, що з’явилося (рис. 1.6), зітріть старе та введіть нове ім’я змінної (до 8 символів латинськими літерами). Для переходу до іншої змінної (якщо є потреба) достатньо натиснути *Кнопку переходу* (**>>**). Коли перейменування завершено, натисніть *ОК.*

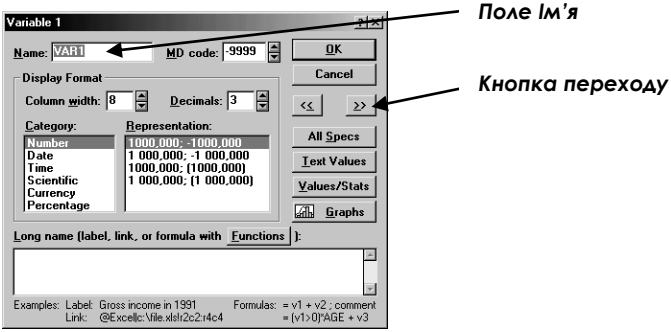



*Рис. 1.6. Вікно перейменування змінних*

Для перейменування рядків (випадків) слід двічі клацнути мишею по кнопці 1 (перший рядок поля даних). З'явиться 1 вікно *Менеджеру імен рядків.* В полі Width введіть максимальну кількість букв, які Ви будете використовувати в назвах рядків та натисніть *ОК.* З'явиться 2 вікна *Менеджеру імен рядків,* де необхідно вести послідовно імена рядків (можна вводити кирилицею) та натиснути *ОК*

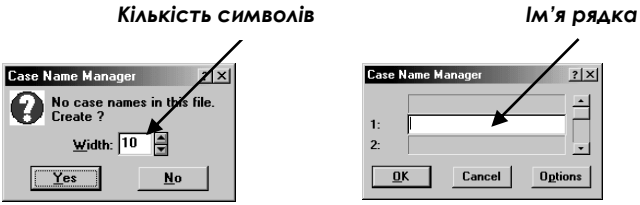



*Рис. 1.7. Вікна Менеджеру імен рядків (зліва – 1, справа – 2)*

***Побудова графіка з основної таблиці даних***

1. Виділіть мишею ті рядки, по яких Ви будуєте графік. Після цього ввійдіть у меню *ГРАФІКИ (Graphs),* підменю *Статистичні 2-мірні графіки (Stats 2D Graphs)* та натисніть пункт *Профілі рядків* (*Line Plots (Case Profiles)) – рис. 1.8.*

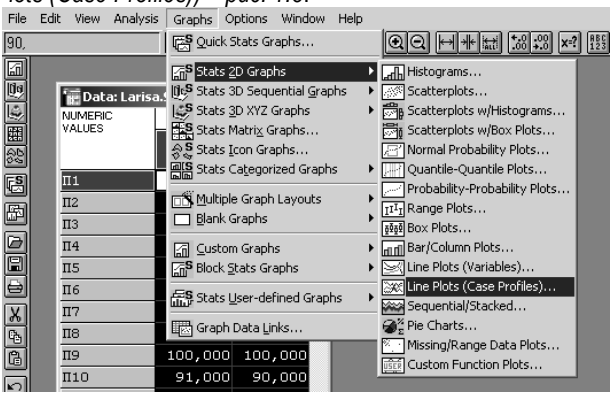



*Рис. 1.8. Побудова графіків з основної таблиці*

2. Після цього перед Вами з'явиться вікно – рис. 1.9. В ньому слід обрати тип графіка. Нас цікавить багатолінійний графік, тому оберемо пункт *Multiple* – віконце внизу зліва.

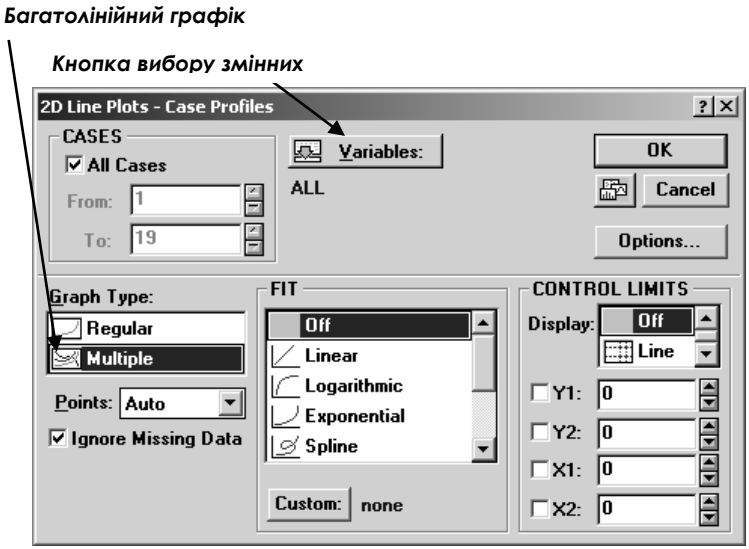



*Рис. 1.9. Вибір змінних та типу графіка*

3. Після цього слід вибрати змінні, які ввійдуть до профілю рядків. Для цього слід натиснути кнопку *Змінні (Variables)*. Перед Вами з'явиться вікно вибору змінних – рис. 1.10.

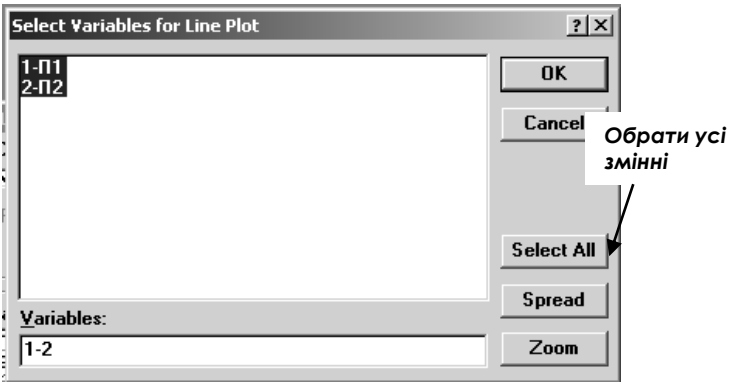



*Рис. 1.10. Вікно вибору змінних*

4. У цьому вікні натисніть кнопку *Обрати усі змінні (Select All)*. Після

цього Ви повернетеся до вікна, зображеного на рис. 1.9, у якому слід натиснути кнопку OK, і Ви отримаєте свій графік. Цього слід зберегти. Як це робиться – див. нижче.

***Побудова графіка з таблиці результатів***

Таблиця результатів – це таблиця з кінцевими обчисленнями. Ви її отримаєте, виконавши операції, описані у Завданні 3. Однак, після їх виконання Ви повернетеся до цього підрозділу, щоб побудувати графік – динаміку середніх показників емоційного стану протягом тижня.

1. Ви маєте перед собою таблицю з описовими статистиками. Виділіть у ній стовпчик з *Середніми показниками (Means)*.
2. Після цього ввійдіть в меню *ГРАФІКИ (Graphs),* підменю *Графіки користувача (Custom Graphs)*, пункт *2-мірні графіки (2D Graphs)* – рис. 1.11.

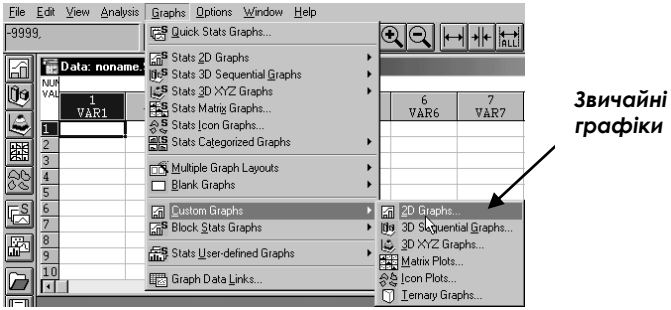



*Рис. 1.11. Меню ГРАФІКИ, підменю Звичайні графіки*

3. У вікні, що з'явилося (рис. 1.12), в полі *Тип графіка* можна обрати тип графіка (лінійний, гістограма та інші). Якщо нічого не вводити, то автоматично буде побудований лінійний графік. На одному графіку може бути зображено до 4 рядів змінних.

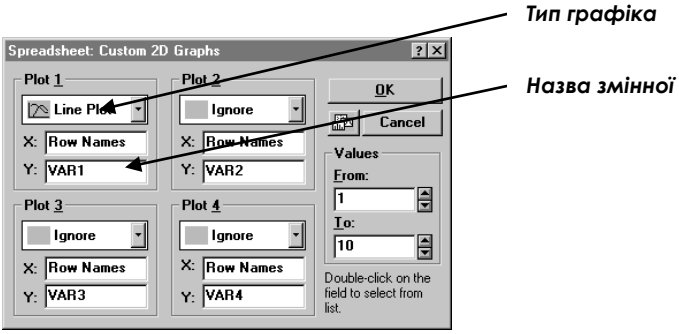



*Рис. 1.12. Вікно вибору типу графіка та змінних*

4. Після цього натисніть кнопку *ОК.* Перед Вами з'явиться побудований графік (рис. 1.13).

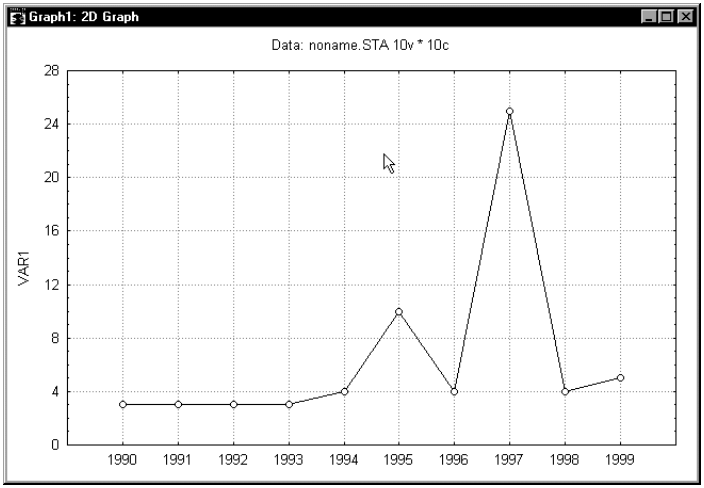



*Рис. 1.13. Побудований графік*

***Збереження графіка***

Для збереження графіка в окремому файлі слід по графіку 1 раз клацнути правою кнопкою миші. У контекстному меню (рис. 1.14) натисніть пункт *Зберегти графік (Save Graph)* та у наступному меню назвіть файл ***Graph*** та збережіть його у потрібну папку.

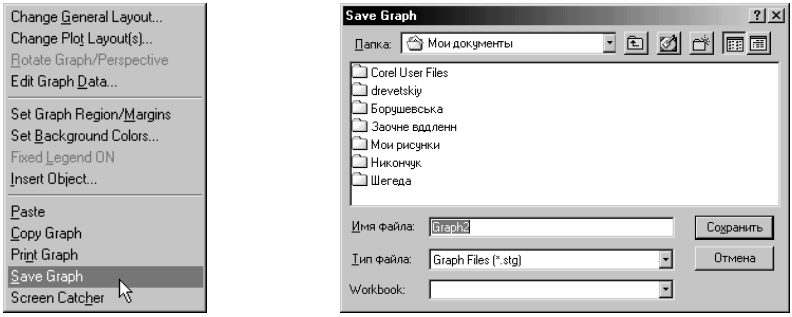



*Рис. 1.14. Збереження графіка*

При потребі тепер цей графік можна буде відкривати, копіювати та переносити в свою наукову роботу.

***Перенесення графіка в Excel та Word***

Наукова робота набирається за допомогою текстового редактора Microsoft Word. Статистичні графіки, створені з допомогою STATISTICA, можна переносити в цей редактор. Для цього слід знову клацнути правою кнопкою миші по графіку та в контекстному меню (рис. 13) вибрати пункт *Копіювати Графік (Copy Graph)*. Після цього запустіть Word (або Excel) та натисніть кнопку *Вставить* . В лист буде вставлений графік. Збережіть цей файл під назвою ***Графік***.

У зв'язку з несумісністю форматів графік буде відображатися з певними недоліками. Для того, щоб їх виправити, слід по графіку клацнути правою кнопкою миші та в контекстному меню обрати пункт *Объект STATISTICA Graph,* підпункт *Преобразовать…* У вікні, що з'явилося, слід натиснути кнопку *ОК*.

## ХІД РОБОТИ: ЗАВДАННЯ 3

### Обрахунок статистичних показників

У програмі STATISTICA є можливість пакетного обрахування статистичних показників. Достатньо задати набір необхідних для аналізу статистик і вони будуть обраховані для обраних змінних та виведені в окрему таблицю, яку можна перенести в файл Excel або Word.

1. Запустіть програму STATISTICA та створіть новий файл даних,

відповідно до Вашого варіанту, назвавши його ***data*** – цей крок ми вже зробили.

2. Після створення та збереження файлу даних слід перейти до іншого модуля програми – *Основні Статистики (Basic Statistics).*

Для переходу від модуля *Управління Даними (Data Management)* до модуля *Основні Статистики (Basic Statistics),* слід у головному меню обрати пункт *Аналіз (Analysis)*, підпункт *Інші Статистики (Other Statistics)*, як це показано на рис. 1.15.

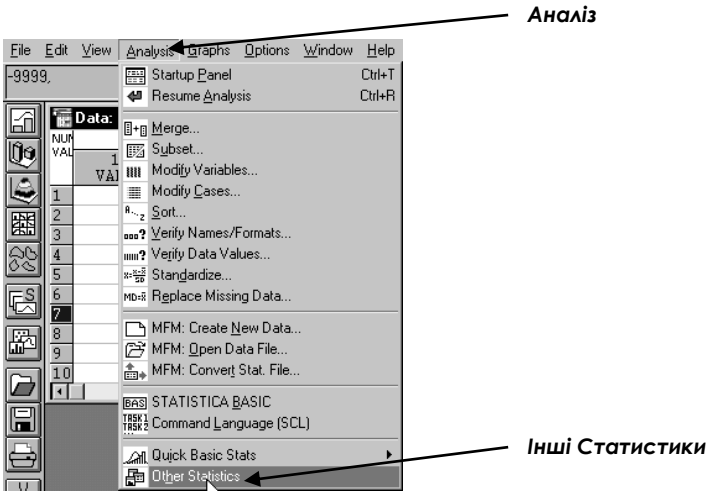



*Рис.1.15. Перехід до Вікна Модулів*

3. Після натискування підпункту *Інші Статистики* перед Вами з'явиться Вікно Модулів, де слід обрати модуль *Основні Статистики (Basic Statistics)*
4. Запустивши модуль *Основні Статистики,* Ви отримаєте доступ до діалогового вікна *Основні статистики і таблиці (Basic Statistics and Tables)*

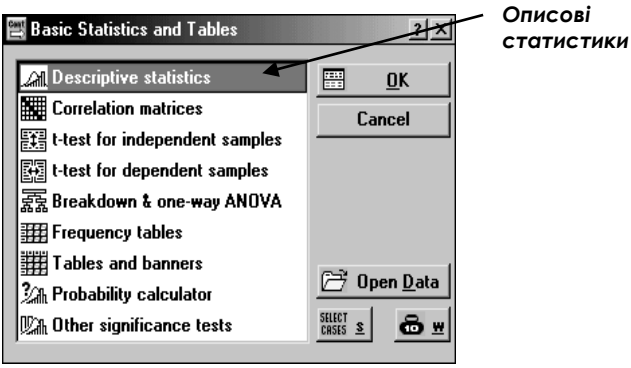



*Рис. 1.16. Діалогове вікно "Основні статистики і таблиці"*

5. У цьому вікні слід обрати пункт *Описові статистики (Descriptive statistics)* (рис. 1.16) та натиснути *ОК.*
6. Перед Вами (рис. 1.17) – вікно Описові статистки, в якому можна задати всі необхідні параметри для аналізу.

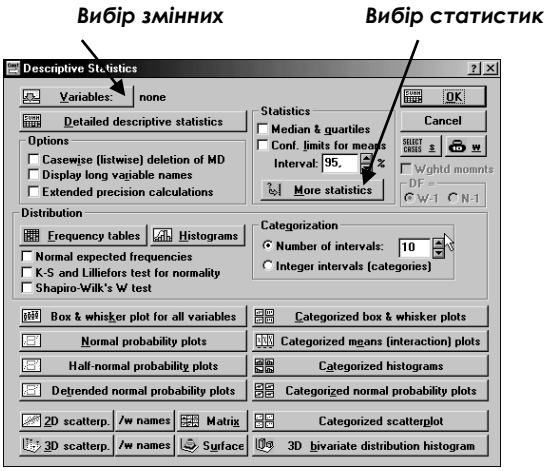



*Рис. 1.17. Вікно вибору параметрів аналізу*

По-перше, слід натиснути кнопку вибору змінних ( Variables: ). У вікні, що з'явилося (рис. 1.18), оберіть змінні для аналізу (у вас в файлі одна змінна, тому натисніть лише на неї). Якщо Ви хочете проаналізувати всі змінні, натисніть кнопку *Обрати всі (Select All).*

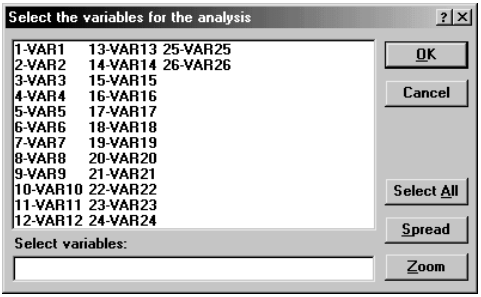



*Рис. 1.18. Вікно вибору змінних*

Після натискування кнопки *ОК Ви повернетеся до вікна описових статистик (рис. 1.17).*

Тепер треба обрати перелік необхідних статистик. Для цього натисніть кнопку *Вибір статистик (More statistics)* (рис. 1.17). З'явиться вікно *Статистики (Statistics)*, зображене на рис. 1.19. З допомогою лівої кнопки миші оберіть потрібні статистики, поставивши у відповідних комірках галочки.

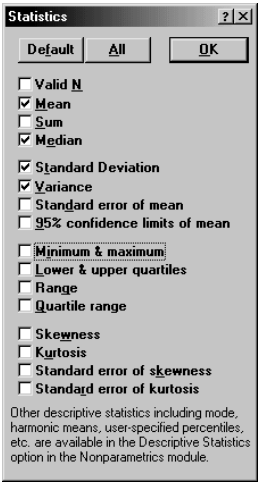


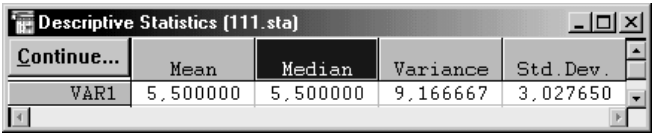

*Рис. 1.19.*

| Англійський термін | Український термін |
|---|---|
| Mean | Середнє |
| Median | Медіана |
| Standard Deviation | Стандартне відхилення |
| Variance | Дисперсія |

Після вибору статистик натисніть *ОК*. Ви повернетеся до вікна вибору параметрів аналізу, зображеного на рис. 1.17.

У цьому вікні теж слід натиснути *ОК*.

Ви отримали статистичну таблицю (рис. 1.20) з результатами аналізу. Кожен стовпчик містить відповідну характеристику (середнє, медіану, дисперсію та стандартне відхилення)



Descriptive Statistics (111.sta)

| Continue... | Mean | Median | Variance | Std.Dev. |
|---|---|---|---|---|
| VAR1 | 5,500000 | 5,500000 | 9,166667 | 3,027650 |

*Рис. 1.20. Результати аналізу*

Оскільки ці дані необхідні не лише для аналізу, але і для оформлення роботи, вони можуть бути або просто переписані у текст роботи, або перенесені з формату STATISTICA в формат офісних програм Excel, Word або Access.

**Перенести дані з програми STATISTICA в програми Excel та Word**

Для перенесення даних в Excel слід чинити так:

- виділити за допомогою миші всі дані в таблиці;
- по виділеному натиснути правою кнопкою миші;
- в контекстному меню обрати пункт *Копировать (Copy)*;
- запустити програму *Excel*;
- поставити курсор на комірку, в яку Ви хочете вставити скопійовані дані;
- натиснути кнопку *Вставить*
- зберегти файл в папку з своїм прізвищем, назвавши його *Описові статистики.*

Для перенесення даних у Word слід чинити так:

- перенести результати аналізу в *Excel;*
- з *Excel* аналогічно перенести дані в Word;
- зберегти файл в папку з своїм прізвищем, назвавши його *Описові статистики.*

## МАТЕРІАЛИ ДЛЯ ЕЛЕКТРОННОГО ЗВІТУ:

Звіти за лабораторні заняття з курсу “Математичні методи у психології”

здаються викладачу у електронному вигляді на дискетах формату 3,5”. Кожен звіт – це **окрема папка** з набором результатів виконаної роботи. Викладач у разі сумніву в авторстві виконаної роботи може вимагати короткого повторення виконаної роботи у своїй присутності.

На дискеті зі звітами обов’язково повинні бути зазначені № групи, № підгрупи та прізвище студента.

*Результати виконання завдання 2:*

Файл даних *data.sta*

Графік, вставлений у Excel *Графік з основної таблиці.xls*

Графік, вставлений у Excel *Графік з таблиці результатів.xls*

*Результати виконання завдання 3:*

Файл з описовими статистиками у форматі Excel *Описові статистики.xls*

*Результати виконання лабораторної роботи в цілому:*

Файл з описом результатів проведеного дослідження у форматі Word *Висновки.doc.* Файл повинен містити назву лабораторної роботи, тему дослідження, результати (графіки та описові статистики) та висновки

## ЛАБОРАТОРНЕ ЗАНЯТТЯ №2:
## АНАЛІЗ ЗВ'ЯЗКУ МІЖ ЗМІННИМИ

***Теоретична частина:***

**1. Міри зв'язку**

– Загальне поняття про зв'язок. Статистичний та функціональний зв'язок.
– Пошук зв'язку з допомогою діаграм розсіювання.
– Пошук зв'язку з допомогою коефіцієнта кореляції. Коефіцієнт кореляції Пірсона
– Область значень коефіцієнта кореляції, рівень його статистичної значимості та обмеження
– Міри зв'язку в непараметричній статистиці: коефіцієнт кореляції Спірмена, Коефіцієнт кореляції φ
– Множинна кореляція

**2. Регресійний аналіз**

– Задачі регресійного аналізу
– Визначення коефіцієнтів регресії
– Обчислення значень залежної змінної на основі регресії
– Обчислення похибки рівняння регресії
– Види рівнянь регресії
– Загальні принципи вибору рівняння регресії

***Література для підготовки до заняття:***


1. Власов К.П., Власов П.К., Киселева А.А. Методы исследований и организация экспериментов. – Х.: Гуманитарный центр, 2002. – 256 с.
2. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976. – 495 с.
3. Готтсданкер Р. Основы психологического эксперимента/ Пер. с англ. – М., 1982. – 464 с.
4. Джужа Н.В. Применение методов непараметрической статистики в психолого-педагогических исследованиях // Вопр. психол. – 1987. – №4. – С. 145-151.
5. Захаров В.П. Применение математических методов в социально-психологических исследованиях. – Л.: ЛГУ, 1985. – 64 с.
6. Музика О.Л. Курсові роботи з психології: навчально-методичний посібник. – Житомир, 1999. – 38 с.
7. Свалов Н.Н. Вариационная статистика. – М.: Лесная промышленность, 1977. –
8. Сидоренко Е.В. Методы математической обработки в психологии. –

СПб.: ООО “Речь”, 2000. – 350 с.
9. Солсо Р.Л., Джонсон Х.Х., Бил М.К. Экспериментальная психология: практический курс. – СПб.: прайм-ЕВРОЗНАК, 2001. – 528 с.
10. Теорія статистики: Навчальний посібник /Вашків П.Г., Пастер П.І., Сторожук В.П., Ткач Є.І. – К.: Либідь, 2001. – 320 с.

***Практична частина:***

**Завдання 1. Проведення психологічного дослідження (до початку заняття)**

*Проведіть дослідження зв’язку між емоційним станом та кількістю успішних ситуацій протягом дня.*

**Завдання 2. Графічний пошук зв’язку між явищами.**

*На основі отриманих даних побудуйте діаграму розсіювання для виявлення кореляційної залежності в середовищі STATISTICA 5.5.*

**Завдання 3. Коефіцієнт лінійної кореляції Пірсона**

*На основі отриманих даних обчисліть коефіцієнт лінійної кореляції та рівень його значимості (STATISTICA 5.5.)*

**Завдання 4. Коефіцієнт рангової кореляції Спірмена**

*На основі отриманих даних обчисліть коефіцієнт рангової кореляції Спірмена (STATISTICA 5.5.)*

**Завдання 5. Основи регресійного аналізу**

*На основі отриманих даних побудуйте лінію регресії та її рівняння*

## ХІД РОБОТИ: ЗАВДАННЯ 1

### ДОСЛІДЖЕННЯ ЗВ’ЯЗКУ МІЖ ЕМОЦІЙНИМ СТАНОМ ТА КІЛЬКІСТЮ СИТУАЦІЙ УСПІХУ

**Об’єкт дослідження**: емоційно-вольова та поведінкова сфера студентів.

**Предмет дослідження**: особливості зв’язку емоційним станом та ситуаціями успіху.

**Гіпотеза**: зв’язок між емоційним станом вранці та кількістю вдалих справ протягом дня сильніший, ніж зв’язок між кількістю вдалих справ протягом дня та емоційним станом ввечері.

**Методи дослідження**: методика „Емоційний градусник”

**Методи обробки даних**: аналіз середніх показників, кореляційний аналіз, регресійний аналіз

**Характеристики вибірки**: 10-15 студентів 1-5 курсів.

**Структура дослідження**:

***До лабораторного заняття:***

Знайдіть майбутніх учасників дослідження. Роздайте їм бланки методики та дайте інструкції з подальшої роботи (*див. Додаток 2*). По завершенню дослідження (через тиждень) зберіть заповнені бланки.

***На лабораторному занятті:***

1. Зведіть отримані сирі дані у статистичну таблицю, ввівши її в Statistica.

| | Учасник 1 | | | ... | Учасник 10 | | |
|---|---|---|---|---|---|---|---|
| | ЕС вранці | Вдалі справи | ЕС ввечері | | ЕС вранці | Вдалі справи | ЕС ввечері |
| Понеділок | | | | | | | |
| Вівторок | | | | | | | |
| Середа | | | | | | | |
| Четвер | | | | | | | |
| П'ятниця | | | | | | | |

2. Обчисліть описові статистики. Для цього зробіть кроки, описані у ***Лабораторному заняття 1, завдання 3***. Дійшовши до вікна вибору статистики, оберіть лише *Середні показники (Means)*. *(Інший варіант – проведіть роботу по обчисленню середніх показників по кожному з учасників дослідженя відразу в Excel)*. Збережіть цю таблицю в форматі Excel, а потім перегрупуйте:

| Учасник | 1 | 2 | 3 |
|---|---|---|---|
| | ЕС вранці (середній показник) | Кількість вдалих справ (середній показник) | ЕС ввечері (середній показник) |
| Учасник 1 | | | |
| Учасник 2 | | | |
| ... | | | |
| Учасник 10 | | | |

3. Створіть новий файл даних, і перенесіть в нього вже перегруповані дані.
4. Побудуйте діаграми розсіювання для стовпчиків 1 та 2, 2 та 3 *(Завдання 2).*
5. Обчисліть коефіцієнти кореляції Пірсона для стовпчиків 1 та 2, 2 та 3 *(Завдання 3).*

6. Обчисліть коефіцієнти кореляції Спірмена для стовпчиків 1 та 2, 2 та 3 *(Завдання 4).*
7. Побудуйте лінію регресії та рівняння регресії для стовпчиків 1 та 2, 2 та 3 *(Завдання 5).*
8. Зробіть висновки про особливості зміни емоційного стану студентів протягом робочого тижня. Чи підтвердилася гіпотеза дослідження? Які недоліки у проведеного дослідження?

***Після лабораторного заняття:***

Підготуйте та здайте звітні матеріали. Поінформуйте учасників дослідження про їх результати.

## ХІД РОБОТИ: ЗАВДАННЯ 2

***Знайти з допомогою графічного методу кореляційну залежність***

1. Створіть файл даних на 3 змінні та відповідну кількість рядків, збережіть його у свою папку під іменем ***korstat*** та заповніть його даними відповідно до Вашого варіанту.
2. Перейдіть до модуля *Основні Статистики* (див. роботу №1). Запустивши модуль, у вікні *Основні статистики і таблиці* запустіть модуль *Кореляційні матриці* (рис. 2.1).

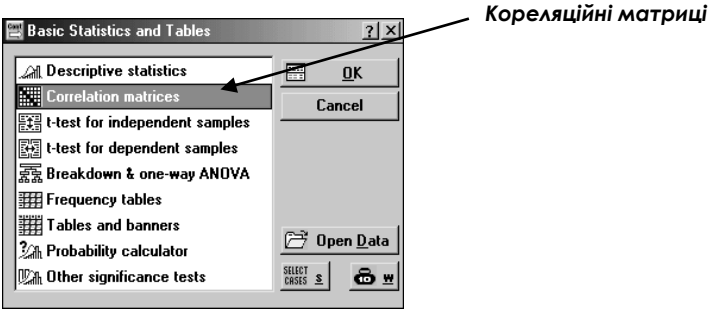



*Рис. 2.1. Кореляційні матриці*

3. Перед Вами з'явиться вікно *Обчислення кореляції за Пірсоном (Pearson Product-Moment Correlation),* в якому задаються всі необхідні параметри для обчислення коефіцієнтів кореляції та побудови діаграми розсіювання.

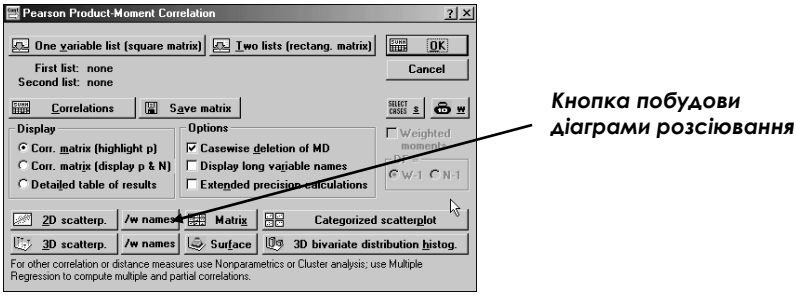



*Рис. 2.2. Вікно Обчислення кореляції за Пірсоном*

4. Для знаходження кореляційної залежності графічним методом достатньо натиснути *Кнопку побудови діаграми розсіювання (/w names).*

Вікно, що з'явиться після цього, запропонує Вам обрати змінні для аналізу (рис. 2.3). Вибір слід здійснити так, як показано на рис. 2.3 – по одній змінній в кожному вікні.

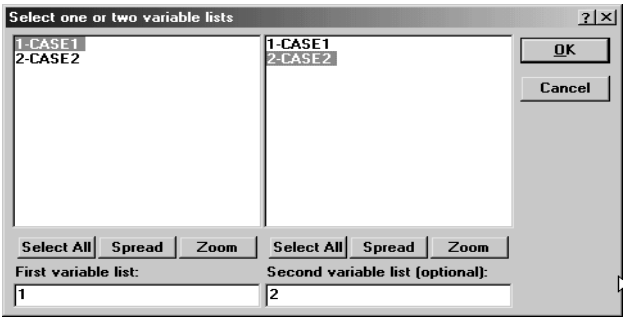



*Рис. 2.3. Вікно вибору змінних для кореляційного аналізу*

Після того, як вибір зроблено, натисніть ОК. Перед Вами з'явиться кореляційний графік.

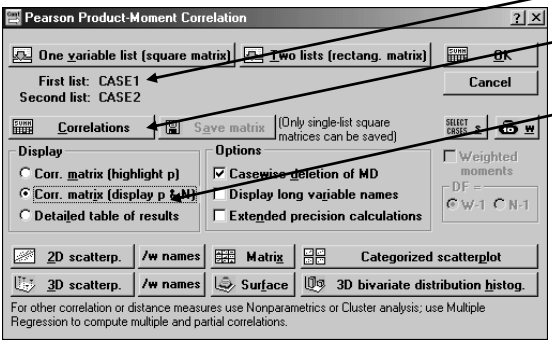
5. Збережіть отриманий графік у Word або Excel під ім'ям ***Діаграма розсіювання***

## ХІД РОБОТИ: ЗАВДАННЯ 3

***Обчислити коефіцієнт лінійної кореляції та рівень його значимості***

Графічний метод дозволяє наочно представити форму кореляційного зв'язку, оцінити приблизно його силу та побачити розподіл аналізованих даних. Для знаходження точного значення коефіцієнта лінійної кореляції Пірсона слід зробити так:

1. Після побудови та збереження кореляційного графіка слід його закрити. Зробивши це, Ви автоматично повернетеся до вікна *Обчислення кореляції за Пірсоном* (рис. 2.4).




***Перелік змінних***

***Обчислення коефіцієнта кореляції***

***Обрахування рівня значимості***

*Рис. 2.4. Обчислення коефіцієнта лінійної кореляції*

2. Для обчислення коефіцієнта кореляції та рівня його значимості слід відмітити в розділі *Display* пункт *Corr. matrix (display p & N)*, як це показано на рис. 2.4.

   Якщо Ви проводили графічний аналіз, то перелік змінних буде задано автоматично.

   Для отримання результатів аналізу достатньо натиснути кнопку *Кореляції (Corellations).*

3. Результат буде представлено у вигляді таблиці (рис. 2.5)

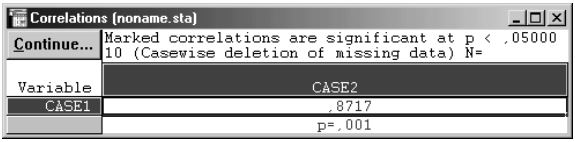



*Рис. 2.5. Результат обчислення коефіцієнта лінійної кореляції*

4. Збережіть отриманий результат в Excel та назвіть файл ***Кореляція***

## ХІД РОБОТИ: ЗАВДАННЯ 4

Коефіцієнт лінійної кореляції широко застосовується в експериментальних дослідженнях, однак він має ряд суттєвих особливостей, що не завжди дозволяють його використовувати. До цих особливостей відноситься вимога **нормальності розподілу** даних, з якої випливає **незастосовність** цього коефіцієнта **до малих вибірок**.

У зв'язку з цим розроблено цілий ряд статистичних методів, які для свого застосування не вимагають нормальності вибірки (непараметричні методи), а тому можуть бути застосовані до малих вибірок. До непараметричних статистичних методів належить і коефіцієнт рангової кореляції Спірмена, який ми й розглянемо.

Дані, які досліджуються за цим коефіцієнтом, мають бути ранговими, однак STATISTICA рангує їх автоматично.

Для обрахування коефіцієнта рангової кореляції Спірмена слід чинити так:

1. Файл даних у Вас вже створено. Тепер слід так, як це раніше, перейти до *Вікна Модулів* (рис. 2.6).
2. Оберіть у цьому вікні модуль *Непараметричні Тести (Nonparametrics/Distrib.)*

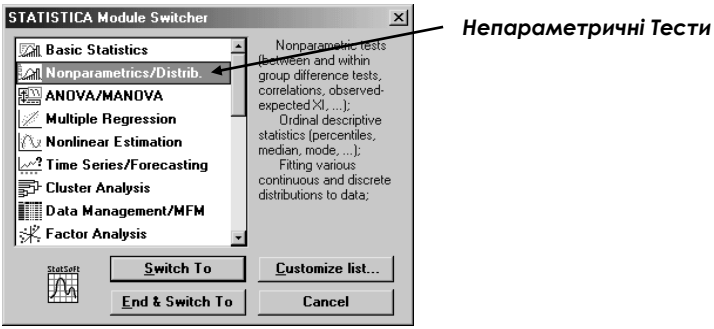



*Рис. 2.6. Вікно модулів. Непараметичні тести*

3. Перейшовши до модуля *Непараметричні тести*, Ви отримаєте доступ до діалогового вікна *Непараметричні статистики* (рис. 2.7). Оберіть у цьому вікні пункт *Кореляції (Correlation (Spearmen, Kendall tau, gamma))*, як це показано на рис. 2.7 та натисніть *ОК*.

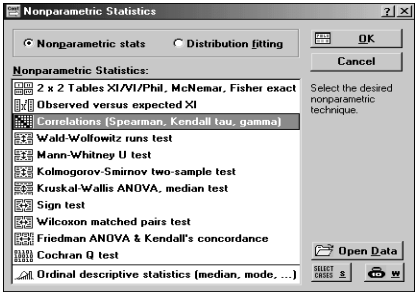



*Рис. 2.7. Непараметричні статистики*

4. Після вибору пункту *Кореляції*, Ви перейдете до наступного діалогового вікна *Непараметричні Кореляції (Nonparametric Corellations)*, в якому слід у пункті *Змінні (Variables)* задати змінні для аналізу (див. Обчислення коефіцієнта лінійної кореляції), а у пункті *Кореляція (Correlation)* – тип коефіцієнта (програма автоматично задає коефіцієнт рангової кореляції Спірмена) (рис. 2.8).

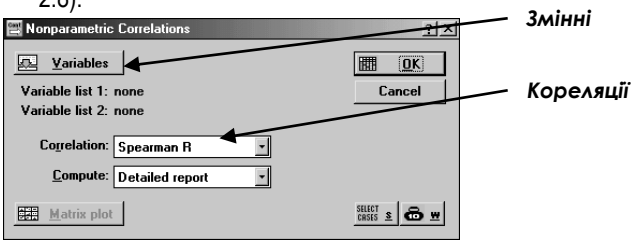



*Рис. 2.8. Непараметричні Кореляції*

5. Коли всі необхідні параметри у вікні *Непараметричні Кореляції* задані, натисніть кнопку ОК. Ви отримаєте таблицю з результатами кореляційного аналізу (рис. 2.9).

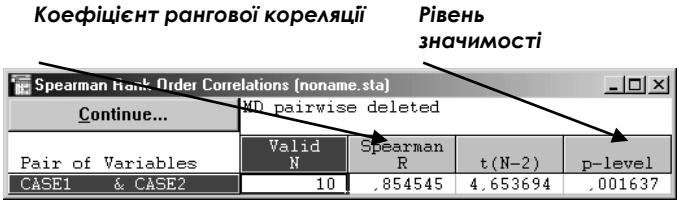



*Рис. 2.9. Результат обчислення коефіцієнта рангової кореляції Спірмена*

6. Збережіть отриманий результат у попередній файл ***Кореляція***, поруч з лінійним коефіцієнтом кореляції. Чи відрізняються отримані результати?

## ХІД РОБОТИ: ЗАВДАННЯ 5

Як ми вже з'ясували на лекціях, часто виникає потреба не просто виявити певний рівень зв'язку та його напрям, але і встановити його **форму**. Для чого? Приміром, для можливості прогнозувати майбутні події на основі вже зібраних даних. Для цього використовується регресійний аналіз. У Statistica є цілий модуль, присвячений складному множинному регресійному аналізу, однак ми будемо користуватися спрощеним варіантом.

1. Після збереження результатів попередньої роботи позакривайте всі робочі вікна, залишивши лише вихідну таблицю з введеними даними.
2. Оберіть у головному меню пункт *Графіки (Graphs)*, підменю *Статистичні 2-мірні графіки (Stats 2D Gaphs)*, підменю *Точкові графіки (Scatterplots…)* – рис. 2.10.

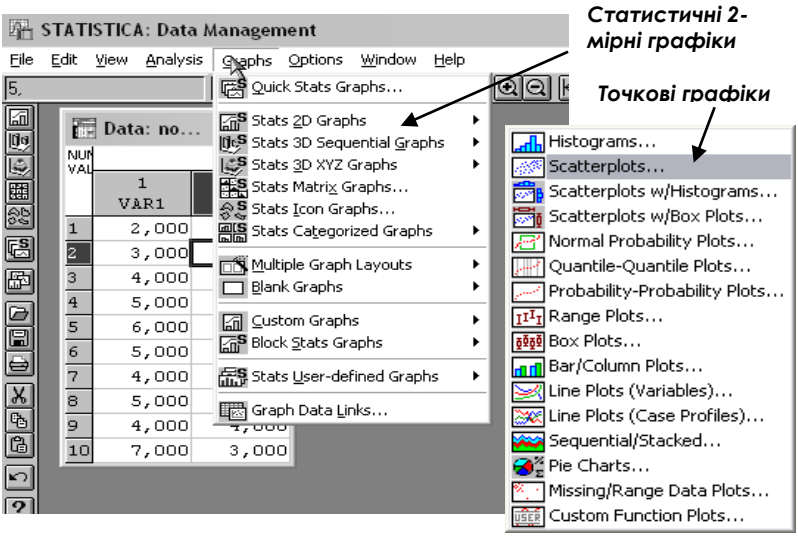



*Рис. 2.10. Пошук пункту меню "Точкові графіки"*

3. Після цього перед Вами з'явиться вікно побудови *2-мірних точкових графіків (2D Scatterplots).* У цьому вікні потрібно задати параметри регресійного аналізу (рис. 2.11). Нас цікавитиме один параметр – тип лінії регресії. Оберіть той тип лінії, який Ви вважаєте найприйнятнішим (наприклад, *лінійну – linear*)/

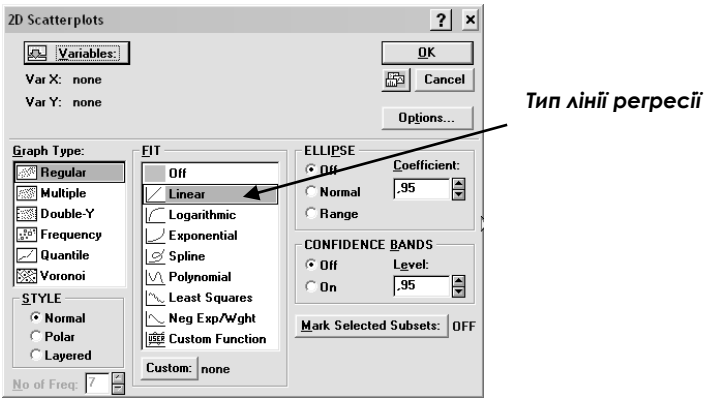



*Рис. 2.11. Вікно вибору параметрів регресійного аналізу*

4. Після натискання кнопки *OK* перед Вами з'явиться вікно *Вибору змінних* (*Select Variables*). Оберіть у лівій колонці першу змінну, а в правій колонці – другу, як показано на рис. 2.12. Натисніть кнопку *OK.*

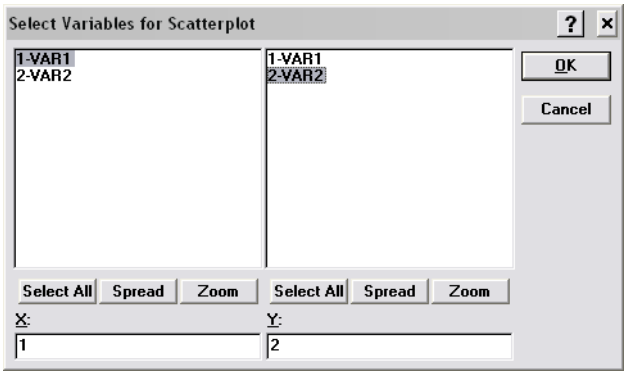



*Рис. 2.12. Вікно вибору змінних*

5. Вітаю Вас, Ви отримали свою першу в житті лінію регресії ☺, а також її формулу (рис. 2.13). Збережіть її у форматі Word під назвою ***Регресійний аналіз***. Повернувшись до рис. 2.11, Ви можете обрати іншу лінію регресії (логарифмічну, експоненційну тощо) і подивитися, яка лінія краще описуватиме Ваші дані.

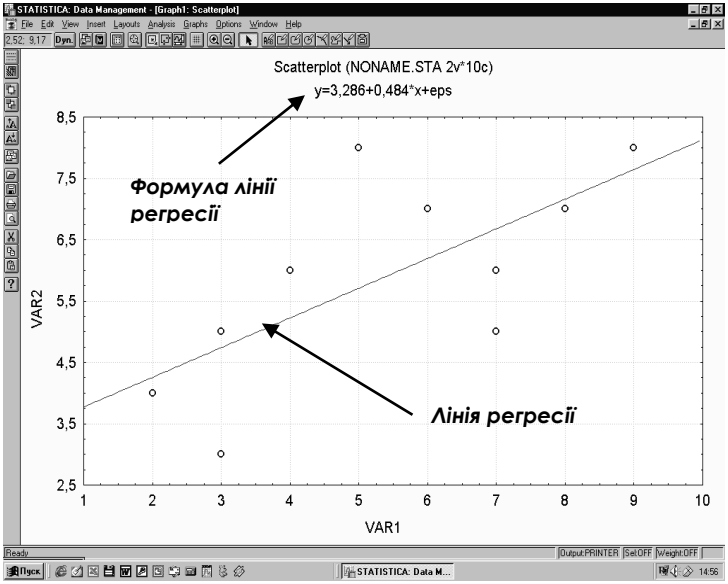



*Рис. 2.13. Лінія регресії та її формула*

## МАТЕРІАЛИ ДЛЯ ЕЛЕКТРОННОГО ЗВІТУ:

*Результати виконання завдання 2:*

Файл даних *korstat.sta*

Файл з 2-ма діаграмами розсіювання (змінні 1-2 та 2-3) – *Діаграми розсіювання.doc*

*Результати виконання завдання 3-4:*

2 коефіцієнти кореляції Пірсона (змінні 1-2 та 2-3), вставлені у Excel *Кореляція.xls*

2 коефіцієнти кореляції Спірмена (змінні 1-2 та 2-3), вставлені у той же файл

*Результати виконання завдання 5:*

2 лінії регресії та їх формули (змінні 1-2 та 2-3) у файлі *Регресійний аналіз.doc*

*Результати виконання лабораторної роботи в цілому:*

Файл з описом результатів проведеного дослідження у форматі Word *Висновки.doc.* Файл повинен містити назву лабораторної роботи, тему дослідження, результати (графіки та описові статистики) та висновки – між якими змінними зв'язок тісніший – між ЕС вранці (1) і кількістю вдали справ (2), чи кількістю вдалих справ (2) та ЕС ввечері (3).

## ЛАБОРАТОРНЕ ЗАНЯТТЯ №3:
## МЕТОДИ СТАТИСТИЧНОГО ВИСНОВКУ:
## ВИЯВЛЕННЯ ВІДМІННОСТЕЙ У РІВНІ ПРОЯВУ ОЗНАКИ

***Теоретична частина:***

- Задача зіставлення і порівняння
- U-критерій Манна-Уітні
- t-критерій Стьюдента для незалежних змінних
- S-критерій Джонкіра
- Алгоритм вибору критерію порівняння

**Література для підготовки до заняття:**


1. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976. – 495 с.
2. Готтсданкер Р. Основы психологического эксперимента/ Пер. с англ. – М., 1982. – 464 с.
3. Мартин Д. Психологические эксперименты: Секреты механизмов психики. – СПб.: прайм-ЕВРОЗНАК, 2002. – 480 с.
4. Музика О.Л. Курсові роботи з психології: навчально-методичний посібник. – Житомир, 1999. – 38 с.
5. Солсо Р.Л., Джонсон Х.Х., Бил М.К. Экспериментальная психология: практический курс. – СПб.: прайм-ЕВРОЗНАК, 2001. – 528 с.


***Практична частина:***

**Завдання 1. Проведення психологічного дослідження (до початку заняття)**

*Проведіть дослідження гендерних відмінностей у рівні розвитку комунікативних та організаторських здібностей.*

**Завдання 2. Використання t-критерію Стьюдента для незалежних змінних**

*На основі отриманих результатів оцініть статистичну значимість відмінностей між даними двох вибірок з допомогою t-критерію Стьюдента для незалежних змінних (STATISTICA 5.5.)*

**Завдання 3. Використання U-критерію Манна-Уітні**

*На основі отриманих результатів оцініть статистичну значимість відмінностей між даними двох вибірок з допомогою U-критерію Манна-Уітні (STATISTICA 5.5.)*

## ХІД РОБОТИ: ЗАВДАННЯ 1

### ДОСЛІДЖЕННЯ ГЕНДЕРНИХ ВІДМІННОСТЕЙ У РІВНІ РОЗВИТКУ КОМУНІКАТИВНИХ ТА ОРГАНІЗАТОРСЬКИХ ЗДІБНОСТЕЙ

**Об'єкт дослідження**: комунікативні та організаторські здібності студентів.

**Предмет дослідження**: гендерні відмінності у рівні розвитку комунікативних та організаторських здібностей студентів.

**Гіпотеза**: гендерних відмінностей у рівні розвитку комунікативних та організаторських здібностей не існує.

**Методи дослідження**: методика КОС-2

**Методи обробки даних**: описові статистики (середнє арифметичне), t-критерій Стьюдента для незалежних змінних, U-критерій Манна-Уітні.

**Характеристики вибірки**: 20 студентів одного курсу, з них 10 чоловіків, 10 жінок.

**Структура дослідження**:

***До лабораторного заняття:***

Знайдіть майбутніх учасників дослідження. Роздайте їм бланки методики "КОС-2" та дайте інструкції з подальшої роботи (*див. Додаток 3*). По завершенню дослідження зберіть заповнені бланки.

***На лабораторному занятті:***

1. Зведіть отримані дані у статистичну таблицю:

| | Комунікативні здібності | | Організаторські здібності | |
|---|---|---|---|---|
| | Чоловіки | Жінки | Чоловіки | Жінки |
| 1. | | | | |
| 2. | | | | |
| 3. | | | | |
| ... | | | | |
| 9. | | | | |
| 10. | | | | |

2. Обчисліть і збережіть у окремий файл середні арифметичні по здібностях у чоловіків та жінок (файл *Описові статистики.xls*). Пригадати, як це робиться, можна в *лабораторному заняття 1, завданні 3*. Чи спостерігається якась тенденція?
3. Обчисліть значення t-критерію Стьюдента окремо по комунікативних здібностях (чоловіки та жінки)та окремо – по організаторських здібностях (чоловіки та жінки) (*див. Завдання 2*). Які можна зробити висновки?

4. Обчисліть значення U-критерію Манна-Уітні окремо по комунікативних здібностях (чоловіки та жінки)та окремо – по організаторських здібностях (чоловіки та жінки) (*див. Завдання 3*). Які можна зробити висновки?
5. Зробіть загальні висновки. Чи підтвердилася гіпотеза дослідження? Який критерій більш потужний? Які недоліки у проведеному дослідженні?

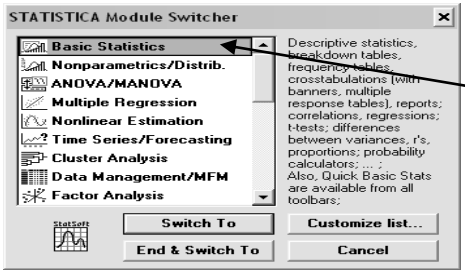
***Після лабораторного заняття:***

Підготуйте та здайте звітні матеріали. Поінформуйте учасників дослідження про їх результати.

## ХІД РОБОТИ: ЗАВДАННЯ 2

***Обчислення t-критерію Стьюдента для незалежних змінних***

1. Створіть за допомогою модуля *Data Management* файл даних на 4 змінні та 10 випадків (4 стовпчики та 10 рядків), збережіть його у свою папку під іменем ***vidminnosti*** та заповніть його даними. ЗБЕРЕЖІТЬ ВВЕДЕНІ ДАНІ.
2. Перейдіть від модуля *Data Management* до модуля *Basic Statistics*.




***Модуль "Основні статистики" (Basic Statistics)***

3. Перед Вами з'явиться вікно *Основні статистики та таблиці* (*Basic Statistics and tables)*. У цьому вікні слід обрати розділ, який відповідає за проведення аналізу з допомогою t-критерію Стьюдента для незалежних зміних – *T-test for independent samples (рис. 3.2).*

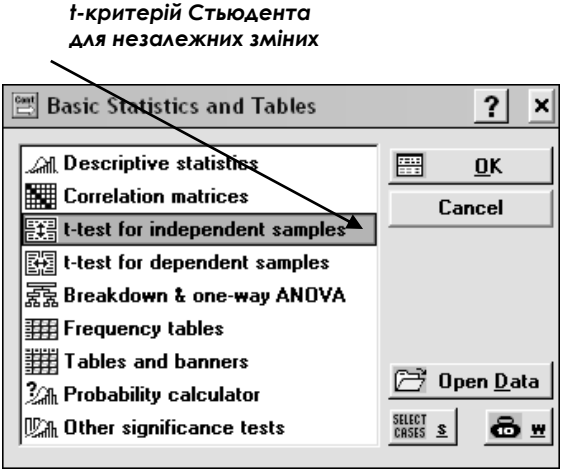



*Рис. 3.2. Вікно вибору t-критерію Стьюдента для незалежних зміних*

4. Перед Вами з'явиться вікно обчислення *t-критерію Стьюдента для незалежних зміних (T-test for independent samples) – рис. 3.3.*

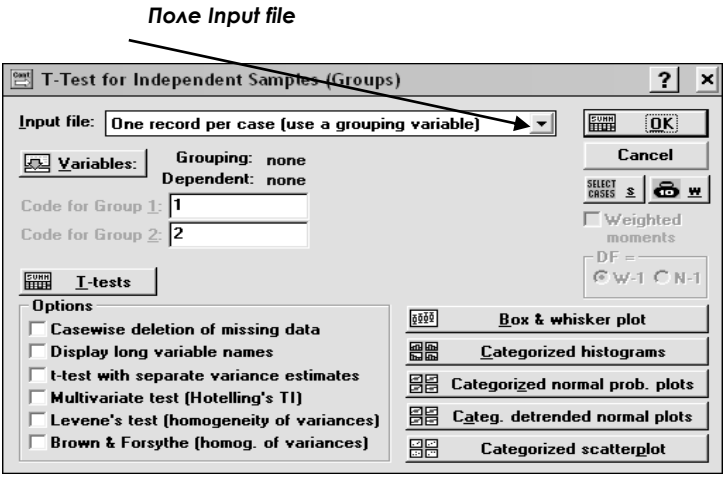



*Рис. 3.3. Вікно обчислення t-критерію Стьюдента для незалежних зміних*

5. У цьому вікні у полі *Input file* потрібно обрати позицію *“Each variable contains the data for one group”.* Відразу після цього вікно набуде іншого вигляду (рис. 3.4).

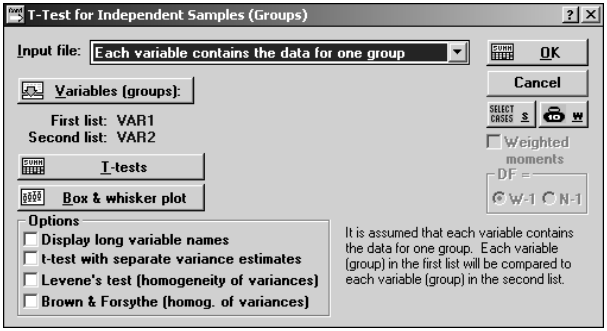



*Рис. 3.4. Новий вигляд вікна обчислення t-критерію Стьюдента для незалежних зміних*

6. Далі натисніть кнопку ОК, і у наступному вікні оберіть змінні для аналізу (рис. 3.5). Оберіть 1 та 3 змінні у лівому вікні, та 2 і 4 – у правому.

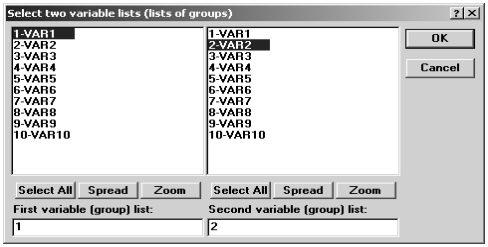



*Рис. 3.5. Вікно вибору змінних для аналізу*

7. Знову натисніть кнопку ОК, і перед Вами з’явиться повна таблиця результатів аналізу (рис. 3.6). Оберіть з неї ті коефіцієнти, які Вам потрібні.

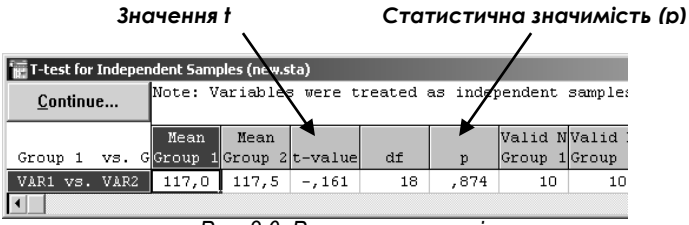



*Рис. 3.6. Результати аналізу*

8. Збережіть отримані результати у файлі Excel під назвою *Стьюдент.xls*

## ХІД РОБОТИ: ЗАВДАННЯ 3

### *Обчислення U-критерію Манна-Уітні*

1. Файл даних вже створено. Перейдіть тепер до *Вікна модулів*, і в ньому оберіть модуль *Непараметричні тести (Nonparametrics/Distrib) – Рис. 3.7.*

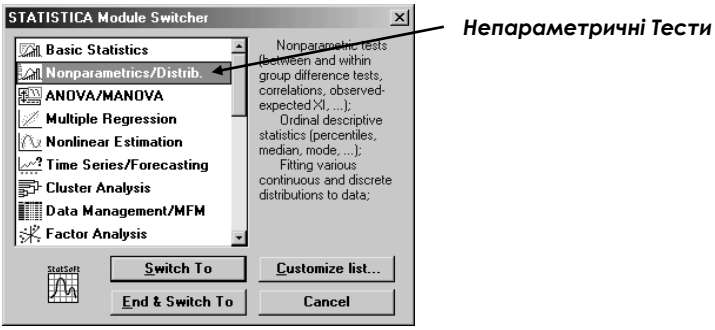



*Рис. 3.7. Перехід до вікна Непараметричні тести*

2. У вікні, що з'явилося, оберіть пункт *U-критерій Манна-Уітні (Mann-Whitney U Test)*, як це показано на рис. 3.8. Натисніть ОК.

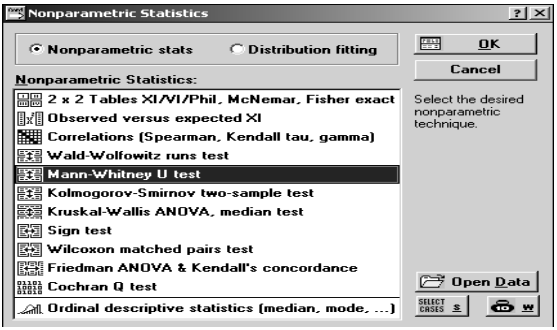



*Рис. 3.8. Вибір U-критерію Манна-Уітні*

3. Перед Вами з'явиться вікно (рис. 3.9). В ньому слід натиснути кнопку *Змінні (Variables)*. У новому вікні оберіть змінні для аналізу – у лівій колонці змінну 1, у правій – змінну 2 (рис. 3.10). Натисніть ОК.

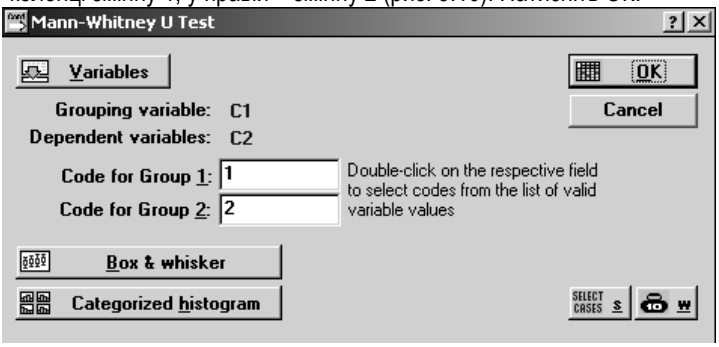



*Рис. 3.9. Вікно обчислення критерію*

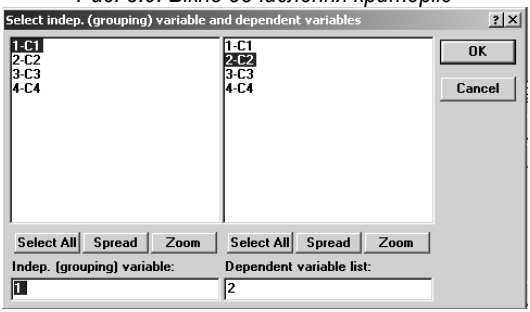



*Рис. 3.10. Вікно вибору змінних*

4. Ви повернулися до вікна обчислення критерію (рис. 3.9). Натисніть ОК, і отримаєте результат (рис. 3.11).

Mann-Whitney U Test (larisa.sta)

Continue...

By variable C1
Group 1: 1 Group 2: 2

| variable | Rank Sum Group 1 | Rank Sum Group 2 | U | Z | p-level | Z adjusted | p-level | Valid N Group 1 | Valid N Group 2 |
|---|---|---|---|---|---|---|---|---|---|
| C2 | -- | -- | 0,00 | 0,00 | 1,000000 | 0,00 | 1,000000 | 0 | 0 |

*Рис. 3.11. Результат обчислення U-критерію Манна-Уітні*

5. Для обчислення цього критерію по організаторським здібностям слід натиснути у вікні результатів кнопку *Продовжити (Continue...)*, і виконати кроки 3-4, з новими змінними (3-ю та 4-ю).
6. Отримані результати скопіюйте і збережіть у файл *Манн-Уітні.xls*.

## МАТЕРІАЛИ ДЛЯ ЕЛЕКТРОННОГО ЗВІТУ:

*Результати виконання завдання 2:*

Файл даних *vidminnosti.sta*

Файл з результатами обчислення t-критерію *Стьюдент.xls*

*Результати виконання завдання 3:*

Файл з результатами обчислення U-критерію *Манн-Уітні.xls*

*Результати виконання лабораторної роботи в цілому:*

Файл з описом результатів проведеного дослідження у форматі Word *Висновки.doc.* Файл повинен містити назву лабораторної роботи, тему дослідження, результати (графіки та описові статистики) та висновки – чи виявилися відмінності між чоловіками та жінками в сфері комунікативних та організаторських здібностей.

## ЛАБОРАТОРНЕ ЗАНЯТТЯ №4:
## МЕТОДИ СТАТИСТИЧНОГО ВИСНОВКУ:
## ОЦІНКА ДОСТОВІРНОСТІ ЗСУВУ ЗНАЧЕНЬ

***Теоретична частина:***

- Задача дослідження змін
- Критерій знаків (G)
- t-критерій Стьюдента для залежних вибірок
- Критерій $\chi_r^2$ Фрідмана
- L-критерій тенденцій Пейджа
- Алгоритм вибору критерію оцінки змін

**Література для підготовки до заняття:**


1. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976. – 495 с.
2. Готтсданкер Р. Основы психологического эксперимента/ Пер. с англ. – М., 1982. – 464 с.
3. Мартин Д. Психологические эксперименты: Секреты механизмов психики. – СПб.: прайм-ЕВРОЗНАК, 2002. – 480 с.


***Практична частина:***

**Завдання 1. Проведення психологічного дослідження (до початку заняття)**

*Проведіть дослідження впливу рівню шуму в приміщенні на лабільність нервових процесів.*

**Завдання 2. Використання t-критерію Стьюдента для залежних змінних**

*На основі отриманих результатів оцініть статистичну значимість змін лабільності нервових процесів з допомогою t-критерію Стьюдента для залежних змінних (STATISTICA 5.5.)*

**Завдання 3. Використання G-критерію знаків**

*На основі отриманих результатів оцініть статистичну значимість змін лабільності нервових процесів з допомогою G-критерію знаків (STATISTICA 5.5.)*

### ХІД РОБОТИ: ЗАВДАННЯ 1

### ДОСЛІДЖЕННЯ ВПЛИВУ ШУМУ НА ЛАБІЛЬНІСТЬ НЕРВОВИХ ПРОЦЕСІВ

**Об'єкт дослідження**: лабільність нервових процесів.

**Предмет дослідження**: шум як фактор зміни лабільності нервових процесів.

**Гіпотеза**: підвищення рівня шуму в приміщенні призводить до зниження рівня лабільності нервових процесів.

**Методи дослідження**: експеримент, теппінг-тест.

**Методи обробки даних**: описові статистики (середнє арифметичне), t-критерій Стьюдента для залежних змінних, G-критерій знаків.

**Характеристики вибірки**: 10 студентів одного курсу.

**Структура дослідження**:

***До лабораторного заняття:***

Знайдіть майбутніх учасників дослідження. Роздайте їм бланки методики "Теппінг-тест" та дайте інструкції з подальшої роботи (*див. Додаток 4*).

***На лабораторному занятті:***

1. Зведіть отримані дані у статистичну таблицю:

| | Лабільність нервових процесів | |
|---|---|---|
| | У тихому приміщенні | У шумному приміщенні |
| 1 | | |
| ... | | |
| 10 | | |

2. Обчисліть і збережіть у окремий файл середні арифметичні показники лабільності нервових процесів (файл *Описові статистики.xls*). Пригадати, як це робиться, можна в *лабораторному заняття 1, завданні 3*. Чи спостерігається якась тенденція?
3. Обчисліть значення t-критерію Стьюдента (*див. Завдання 2*). Які можна зробити висновки?
4. Обчисліть значення G-критерію знаків (*див. Завдання 3*). Які можна зробити висновки?
5. Зробіть загальні висновки. Чи підтвердилася гіпотеза дослідження? Який критерій більш потужний? Які недоліки у проведеному дослідженні?

***Після лабораторного заняття:***

Підготуйте та здайте звітні матеріали. Поінформуйте учасників дослідження про їх результати.

## ХІД РОБОТИ: ЗАВДАННЯ 2

***Обчислення t-критерію Стьюдента для залежних змінних***

1. Створіть з допомогою модуля *Data Management* файл даних на 2 змінні та 10 випадків (2 стовпчики та 10 рядків), збережіть його у свою папку під іменем ***zminy*** та заповніть його даними. ЗБЕРЕЖІТЬ ВВЕДЕНІ ДАНІ.

2. Перейдіть від модуля *Data Management* до модуля *Basic Statistics*.

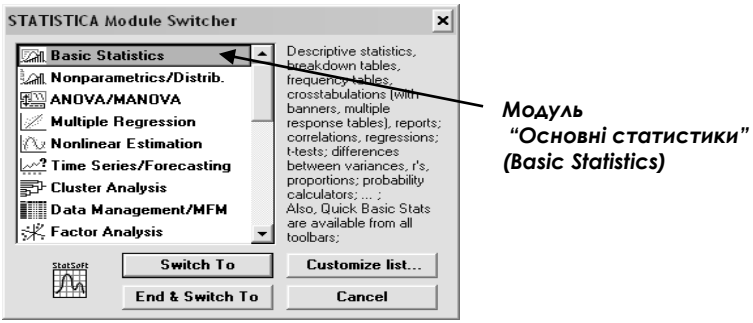



*Рис. 4.1. Вікно модулів*

3. Перед Вами з'явиться вікно *Основні статистики та таблиці* (*Basic Statistics and tables)*. У цьому вікні слід обрати розділ, який відповідає за проведення аналізу з допомогою t-критерію Стьюдента для незалежних зміних – *T-test for independent samples (рис. 4.2).*

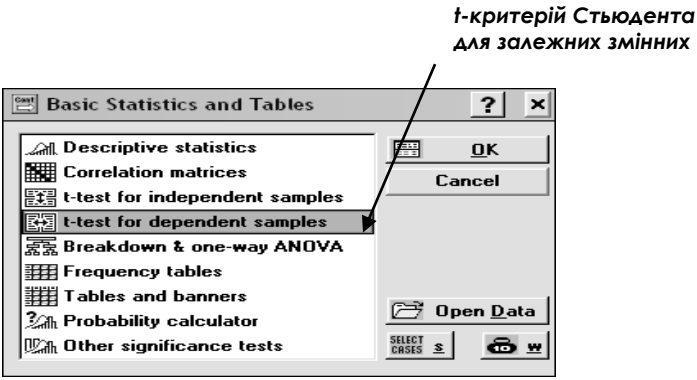



*Рис. 4.2. Вікно вибору t-критерію Стьюдента для залежних зміних*

4. Перед Вами з'явиться вікно обчислення *t-критерію Стьюдента для залежних зміних (T-test for dependent samples) – рис. 4.3.*

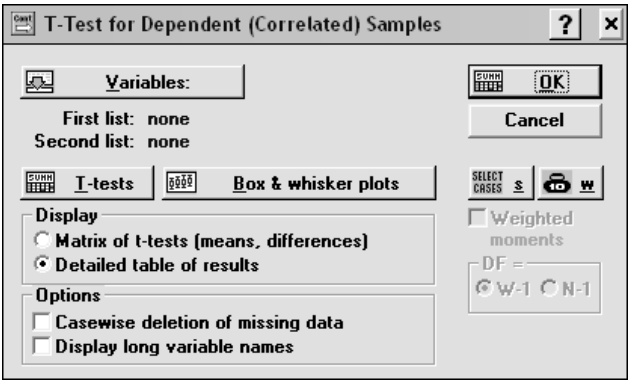



*Рис. 4.3. Вікно обчислення t-критерію Стьюдента для залежних зміних*

5. Далі натисніть кнопку *Вибору змінних (Variables)*, і у наступному вікні оберіть для аналізу показники (як показано на рис. 4.5).

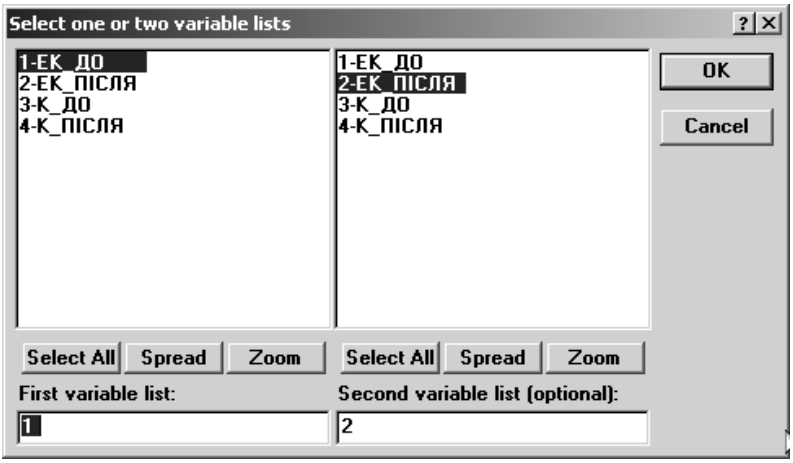



*Рис. 4.5. Вікно вибору змінних для аналізу*

6. Знову натисніть кнопку ОК, потім ще раз ОК, і перед Вами з'явиться повна таблиця результатів аналізу цих показників (рис. 4.6)

| Variable | Mean | Std.Dv. | N | Diff. | Std.Dv. Diff. | t | df | p |
|---|---|---|---|---|---|---|---|---|
| ЕК_ДО | ,512757 | ,305389 | | | | | | |
| ЕК_ПІСЛЯ | ,496301 | ,263208 | 10 | ,016456 | ,392605 | ,132543 | 9 | ,897471 |

*Рис.4.6. Результати аналізу*

7. Збережіть отриману таблицю у файлі формату Excel під назвою *Стьюдент - залежні змінні.*
8. Зробіть висновки про значимість зміни показників лабільності нервових процесів.

## ХІД РОБОТИ: ЗАВДАННЯ 3

***Обчислення G-критерію знаків***

1. Файл даних вже створено. Перейдіть тепер до *Вікна модулів*, і в ньому оберіть модуль *Непараметричні тести (Nonparametrics/Distrib) – Рис. 4.7.*

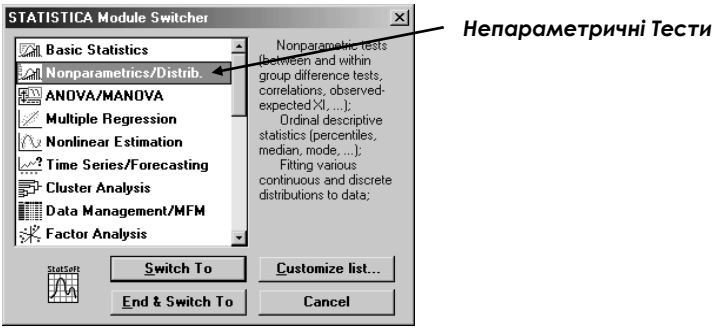



*Рис. 4.7. Перехід до вікна Непараметричні тести*

2. У вікні, що з'явилося, оберіть пункт *Критерій знаків (Sign test)*, як це показано на рис. 4.8. Натисніть ОК.

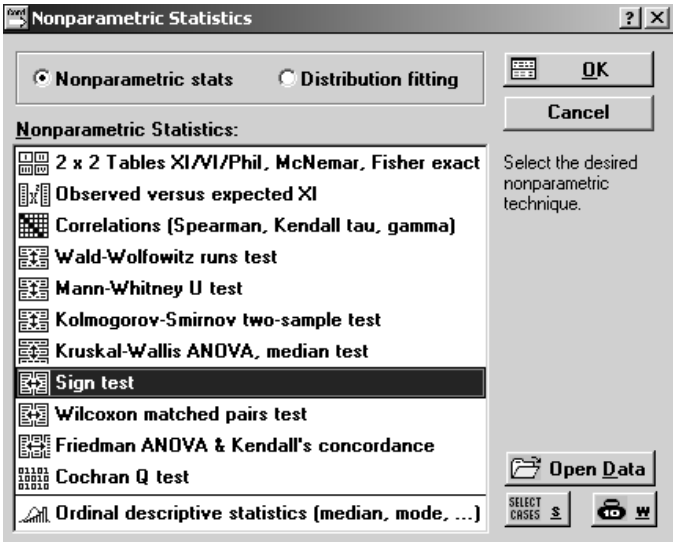
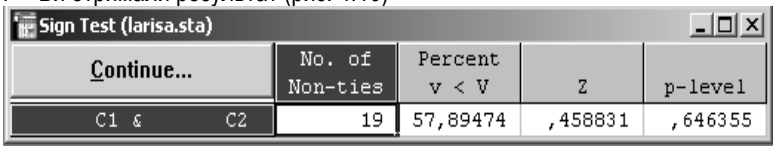



*Рис. 4.8. Вибір критерію знаків*

3. Перед Вами з'явиться вікно обчислень критерію знаків (рис. 4.9)

*Рис. 4.9. Вікно обчислення критерію знаків.*

4. Натисніть кнопку *Змінні (Variables)*, щоб обрати змінні для аналізу, а потім, повернувшись до вікна обчислень натисніть ОК.
5. Ви отримали результат (рис. 4.10)



Sign Test (larisa.sta)

| Continue... | No. of Non-ties | Percent v < V | Z | p-level |
|---|---|---|---|---|
| C1 & C2 | 19 | 57,89474 | ,458831 | ,646355 |

*Рис. 4.10. Результат обчислення критерію знаків*

6. Збережіть отриману таблицю у файлі формату Excel під назвою *Критерій знаків*.
7. Зробіть висновки. Чи відрізняються результати обчислень за критерієм знаків та критерієм Стьюдента?

## МАТЕРІАЛИ ДЛЯ ЕЛЕКТРОННОГО ЗВІТУ:

*Результати виконання завдання 2:*

Файл даних *zminy.sta*

Файл з результатами обчислення t-критерію *Стьюдент-залежні змінні.xls*

*Результати виконання завдання 3:*

Файл з результатами обчислення G-критерію знаків *Критерій знаків.xls*

*Результати виконання лабораторної роботи в цілому:*

Файл з описом результатів проведеного дослідження у форматі Word *Висновки.doc.* Файл повинен містити назву лабораторної роботи, тему дослідження, результати (графіки та описові статистики) та висновки – чи відбулися зміни в лабільності нервових процесів під впливом шуму? Який недолік у структурі експерименту не дає можливості бути впевненими у висновках?

## ЛАБОРАТОРНЕ ЗАНЯТТЯ №5:
## ФАКТОРНИЙ АНАЛІЗ

***Теоретична частина:***

1. Мета факторного аналізу.
2. Підготовка даних до факторного аналізу.
3. Інтерпретація результатів факторного аналізу.
4. Задача про кількість факторів.
5. Особливості факторного аналізу.
6. Місце факторного аналізу в структурі експерименту.

**Література для підготовки до заняття:**


1. Франселла Ф. Баннистер Д. Новый метод иследования личности – М., 1987. – 236 с
2. Докторов Б.З. Об использовании методов факторного анализа в работах советских исследователей //Вопр. психол. – 1969. – №2. – С. 142-146.
3. Левандовский Н.Г. О корректности применения факторного анализа и о критериях факторизации // Вопр. психол. – 1980. – №5. – С. 138-143.
4. Петренко В.Ф. Психосемантика сознания. – М.: Изд-во Моск. ун-та, 1988. – 208 с.
5. Беспалько И.Г. О некоторых неясных вопросах психологической интерпретации факторов в факторном анализе // Психол. журн. – 1987. – Т.8. – №3. – С. 137-145.


***Практична частина:***

**Завдання 1.** **Проведення психологічного дослідження (до початку заняття).**

*Провести до заняття порівняльне дослідження сприймання студентами слів "вбивця" та "кіллер"*[21].

**Завдання 2.** **Проведення факторного аналізу.**

*Зробити факторний аналіз результатів дослідження.*

**Завдання 3.** **Підготовка висновків за результатами**

*Виявити головні фактори у сприйманні слова "вбивця" та слова "кіллер". З'ясувати, чим вони між собою різняться.*

---

[21] *Музика О.Л. Психологічні механізми ціннісно-мовленнєвої взаємодії / Національна освіта: традиції та інновації у контексті ідей Івана Огієнка: Збірник наукових праць / За ред. проф. М.В. Левківського. – Київ-Житомир, ЖДПУ, 2002. – С. 101-106.*

## ХІД РОБОТИ: ЗАВДАННЯ 1

### ДОСЛІДЖЕННЯ СПРИЙМАННЯ СЛІВ “ВБИВЦЯ” ТА “КІЛЛЕР”

1. *Провести асоціативний експеримент, де словами-стимулами окремо будуть спочатку “вбивця”, а потім “кіллер”. Загальна вибірка – 10-15 осіб.*
2. *Зробити частотний аналіз асоціацій до кожного слова-стимула та виділити по 15 асоціацій з максимальною частотою (15 асоціацій до слова “вбивця” та 15 асоціацій до слова “кіллер”).*
3. *Попросити учасників дослідження оцінити ступінь близькості кожної з асоціацій до відповідного слова-стимула.*
4. *Створити 2 матриці результатів (1 матриця – сприймання слова “вбивця”; 2 матриця – сприймання слова “кіллер”).*

***Приклади матриць результатів:***

*Матриця 1: Сприймання слова “вбвиця”*

| Учасники дослідження | Асоціації | | | |
|---|---|---|---|---|
| | жах | крик | ... | біль |
| 1 | *4* | *5* | ... | *9* |
| 2 | *4* | *5* | ... | *8* |
| 3 | *7* | *8* | ... | *9* |
| ... | ... | ... | ... | ... |
| 20 | *7* | *9* | | *4* |

*Матриця 2: Сприймання слова “кіллер”*

| Учасники дослідження | Асоціації | | | |
|---|---|---|---|---|
| | фільм | робота | ... | осуд |
| 1 | *8* | *7* | ... | *7* |
| 2 | *6* | *6* | ... | *5* |
| 3 | *4* | *5* | ... | *6* |
| ... | ... | ... | ... | ... |
| 20 | *8* | *9* | | *9* |

## ХІД РОБОТИ: ЗАВДАННЯ 2

***Створення файлу даних***

Запустіть STATISTICA 5.5, модуль *Data Management* та створіть спочатку файл даних з матрицею асоціацій на слово “вбивця” (grid_v.sta) на потрібну Вам кількість стовпчиків (асоціацій) та рядків (досліджуваних). Як це

робиться, можна пригадати, повернувшись до лабораторної роботи №1, пункт *Створення та збереження файлу даних.*

***Перейменування стовпчиків та рядків***

Для спрощення аналізу слід відразу (до початку аналізу) правильно назвати змінні (асоціації), та випадки (досліджуваних). Спосіб перейменування даних можна пригадати в роботі №1.

Оскільки програма англомовна, то назви змінних краще робити латинськими літерами, довжиною не більше 8 символів. Використовуйте скорочення при введенні асоціацій.

Після того, як всі асоціації введені, збережіть зміни ( ).

***Перехід до модуля Факторного аналізу***

Для переходу від модуля *Управління Даними (Data Management)* до модуля *Факторний Аналіз (Factor Analysis)* слід в головному меню обрати пункт *Аналіз (Analysis)*, підпункт *Інші Статистики (Other Statistics)*.
Після натискування підпункту *Інші Статистики* перед Вами з'явиться вікно модулів, де слід обрати модуль *Факторний аналіз (Factor Analysis)*.

***Проведення факторного аналізу***

Автоматично після запуску модуля Факторний аналіз запускається діалогове вікно факторного аналізу (рис. 5.1)

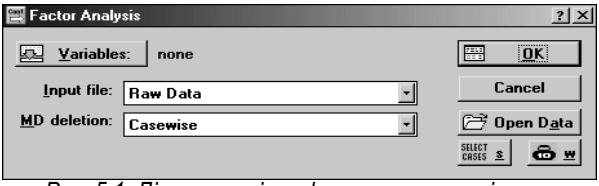



*Рис. 5.1. Діалогове вікно факторного аналізу*

Зліва вверху в цьому вікні розміщена кнопка *Змінні* ( Variables: ). Її слід натиснути, щоб обрати ті змінні, по яких Ви будете проводити факторний аналіз.

Після натискування на цю кнопку з'явиться нове діалогове вікно, схоже на представлене на рис. 5.2.

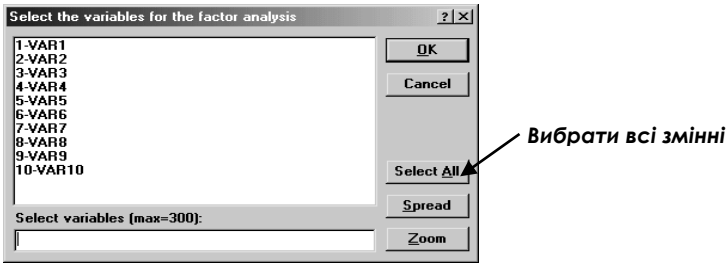



*Рис. 5.2. Вікно вибору змінних для факторного аналізу*

Оскільки нам для аналізу будуть потрібні всі змінні, то натисніть кнопку *Вибрати Всі (Select All)*. Після того натисніть кнопку *ОК.*

Ви повернулися до першого діалогового вікна (рис. 5.1). Для початку факторного аналізу натисніть кнопку ОК. Вікно, що з'явилося, вимагає обрати тип факторного аналізу, максимальну кількість факторів та деякі додаткові параметри (рис. 5.3).

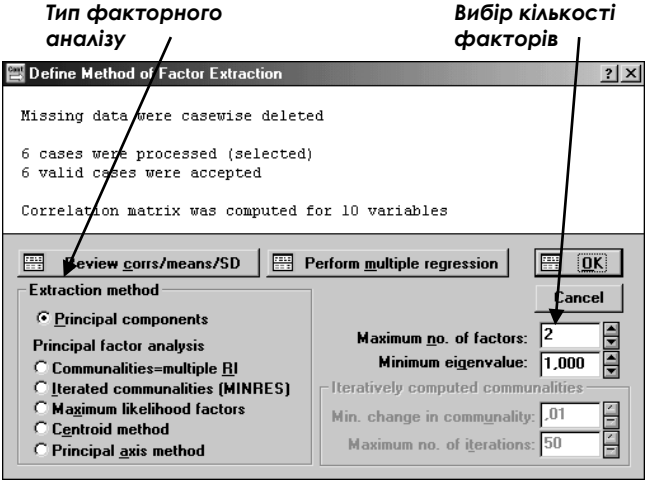



*Рис. 5.3. Вікно вибору методу факторного аналізу*

Коли Ви грунтовніше ознайомитеся з факторним аналізом, то зможете поекспериментувати з його видами та іншими параметрами. А поки що залишимо все без змін, лише ввівши кількість факторів рівну 2 *(Maximum no. of factors = 2).*

Для завершення процедури факторного аналізу натисніть *ОК*.

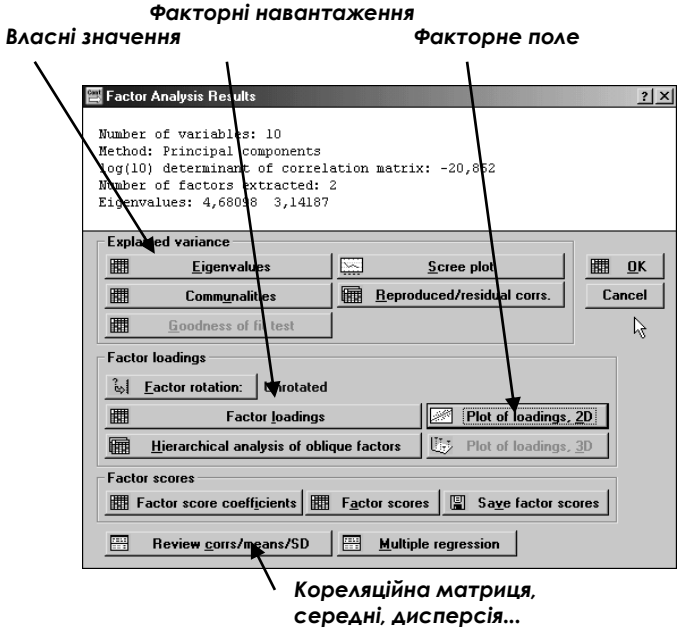



*Рис. 5.4. Вікно результатів факторного аналізу*

Перед Вами (рис. 5.4) повинно з'явитися вікно *Результати Факторного Аналізу (Factor Analysis Results).*

Для опису результатів факторного аналізу необхідний цілий комплекс числових результатів факторного аналізу: кореляційна матриця; власні значення факторів (у відсотках від загальної дисперсії); факторні навантаження; факторний графік;

Отримати ці результати можна, натиснувши на відповідні кнопки вікна результатів факторного аналізу.

Форми отриманих результатів показано на рис. 5.5. Щоразу після натискування відповідної кнопки перед Вами з'являтимуться обрані результати. Щоб продовжити аналіз, необхідно після копіювання обраних показників у Excel натискати у верхньому лівому кутку кнопку *Продовжити (Continue…)*

**Correlations (111.sta)**

Continue... Casewise deletion of MD N=26

| Variable | CASE1 | CASE2 | CASE3 | CASE4 | CASE5 | CASE6 | CASE7 | CASE8 | CASE9 | CASE10 |
|---|---|---|---|---|---|---|---|---|---|---|
| CASE1 | 1,00 | -,01 | -,04 | -,23 | ,27 | ,61 | ,22 | -,21 | -,65 | -,87 |
| CASE2 | -,01 | 1,00 | -,65 | -,57 | -,68 | -,20 | ,34 | ,52 | ,46 | -,06 |
| CASE3 | -,04 | -,65 | 1,00 | ,72 | ,52 | -,07 | -,69 | -,75 | -,46 | ,03 |
| CASE4 | -,23 | -,57 | ,72 | 1,00 | ,50 | -,25 | -,78 | -,77 | -,37 | ,22 |
| CASE5 | ,27 | -,68 | ,52 | ,50 | 1,00 | ,36 | -,40 | -,65 | -,67 | -,37 |
| CASE6 | ,61 | -,20 | -,07 | -,25 | ,36 | 1,00 | ,18 | -,28 | -,69 | -,61 |
| CASE7 | ,22 | ,34 | -,69 | -,78 | -,40 | ,18 | 1,00 | ,81 | ,27 | -,29 |
| CASE8 | -,21 | ,52 | -,75 | -,77 | -,65 | -,28 | ,81 | 1,00 | ,77 | ,17 |
| CASE9 | -,65 | ,46 | -,46 | -,37 | -,67 | -,69 | ,27 | ,77 | 1,00 | ,61 |
| CASE10 | -,87 | -,06 | ,03 | ,22 | -,37 | -,61 | -,29 | ,17 | ,61 | 1,00 |

**Eigenvalues (111.sta)**

Continue... Extraction: Principal components

| Value | Eigenval | % total Variance | Cumul. Eigenval | Cumul. % |
|---|---|---|---|---|
| 1 | 4,680984 | 46,80984 | 4,680984 | 46,80984 |
| 2 | 3,141868 | 31,41868 | 7,822852 | 78,22852 |

**Factor Loadings (Unrotated) (111.sta)**

Continue... Extraction: Principal components (Marked loadings are > ,700000)

| Variable | Factor 1 | Factor 2 |
|---|---|---|
| CASE1 | -,282453 | ,859092 |
| CASE2 | ,722386 | ,156354 |
| CASE3 | -,798999 | -,353618 |
| CASE4 | -,749467 | -,549028 |
| CASE5 | -,819710 | ,150287 |
| CASE6 | -,335172 | ,777668 |
| CASE7 | ,679539 | ,556153 |
| CASE8 | ,942047 | ,088712 |
| CASE9 | ,821148 | -,497831 |
| CASE10 | ,267376 | -,872083 |
| Expl.Var | 4,680984 | 3,141868 |
| Prp.Totl | ,468098 | ,314187 |

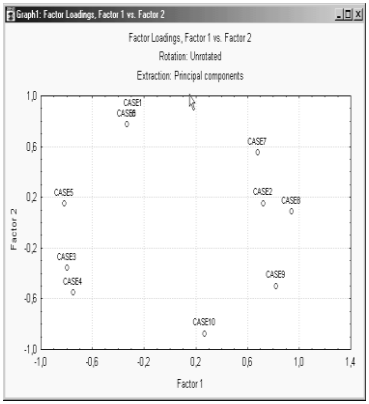



*Рис. 5.5. Основні результати факторного аналізу (кореляційна матриця, власні значення факторів, факторні навантаження та факторний графік)*

***Збереження результатів факторного аналізу***

Факторний аналіз проведено, однак, одним із елементів експериментальної роботи є грамотне представлення її результатів. Таке представлення переслідує декілька цілей:

1. Полегшення інтерпретації факторних структур.
2. Можливість перенесення числових результатів до тексту наукової роботи.
3. Переконання читачів в достовірності отриманих Вами даних.

Саме тому всі отримані дані, наведені на рис. 34, повинні бути представлені як в тексті (факторний графік), так і в додатках (інші статистики).

Найзручніший спосіб **збереження результатів** – це перенесення їх у файл Excel. Збережіть файл у папку з своїм прізвищем, назвавши його "*Факторний аналіз"*.

***Новий факторний аналіз***

Після завершення факторного аналізу матриці по слову "вбивця" проведіть новий факторний аналіз матриці по слову "кіллер". Файл з даними назвіть grid_k.sta.

## ХІД РОБОТИ: ЗАВДАННЯ 3

Інтерпретацію слід проводити, спираючись на всі отримані в результаті аналізу показники, однак, найкорисніше користуватися факторним полем. З'ясуйте: Які виділилися головні фактори у сприйманні студентами слова "вбивця"? Які виділилися головні фактори у сприйманні студентами слова "кіллер"? Чим ці фактори між собою різняться? Як Ви поясните різницю у сприйманні слів "вбивця" та "кіллер"?

Підготуйте свою інтерпретацію у форматі Word та збережіть у свою папку під назвою "Факторний аналіз".

## МАТЕРІАЛИ ДЛЯ ЕЛЕКТРОННОГО ЗВІТУ:

*Результати виконання завдання 2:*

Файли даних *grid_v.sta, grid_k.sta*

Файл з результатами проведення факторного аналізу *Факторний аналіз.xls* (формат Excel)

*Результати виконання завдання 3:*

Файл з інтерпретацією результатів факторного аналізу *Факторний аналіз.doc* (формат Word)

## ЛАБОРАТОРНЕ ЗАНЯТТЯ №6:
## БАГАТОМІРНЕ ШКАЛЮВАННЯ

***Теоретична частина:***

1. Загальна мета багатомірного шкалювання. Підгототовка даних до БШ
2. Логіка багатомірного шкалювання
3. Проблема розмірності при багатомірному шкалюванні
4. Інтерпретація результатів багатомірного шкалювання
5. Багатомірне шкалювання та факторний аналіз. Реплікуюче БШ.

**Література для підготовки до заняття:**


1. Крылов В.Ю. О многомерном шкалировании в неметрических пространствах // Психол. журн. – 1987. – Т.8. – №5. – С. 140-142.
2. Лосик Г.В. Исследование восприятия гласных методом многомерного шкалирования // Психол. журн. – 1992. – Т.13. – №2.
3. Терехина А.Ю. Многомерное шкалирование в психологии // Психол. журн. – 1983. – Т.4. – №1.
4. Шепард Р.Н. Многомерное шкалирование и безразмерное представление различий // Психол. журн. – 1980. – Т.1. – №4. – С. 72-84.


***Практична частина:***

**Завдання 1.** **Підготовка даних до багатомірного шкалювання.**
*Провести до заняття порівняльне дослідження сприймання студентами слів "вбивця" та "кіллер"*[22]*. Якщо виконувалося завдання 1 до лабораторної роботи №5 (Факторний аналіз), дослідження можна не проводити – слід скористатися вже готовими результатами.*

**Завдання 2.** **Проведення багатомірного шкалювання.**
*Зробити багатомірне шкалювання результатів дослідження.*

**Завдання 3.** **Підготовка висновків за результатами**
*Виявити головні виміри у сприйманні слова "вбивця" та слова "кіллер". Проаналізувати групи та конфігурації груп асоціацій. Порівняти результати БШ з результатами ФА.*

---

[22] *Музика О.Л. Психологічні механізми ціннісно-мовленнєвої взаємодії / Національна освіта: традиції та інновації у контексті ідей Івана Огієнка: Збірник наукових праць / За ред. проф. М.В. Левківського. – Київ-Житомир, ЖДПУ, 2002. – С. 101-106.*

## ХІД РОБОТИ: ЗАВДАННЯ 1

***Див. лабораторне заняття №5.***

## ХІД РОБОТИ: ЗАВДАННЯ 2

***Створення файлу даних***

Запустіть STATISTICA 5.5, модуль *Data Management* та створіть спочатку файл даних з матрицею асоціацій на слово “вбивця” (grid_v.sta) на потрібну Вам кількість стовпчиків (асоціацій) та рядків (досліджуваних). Як це робиться, можна пригадати, повернувшись до лабораторної роботи №1, пункт *Створення та збереження файлу даних.*

**Якщо Ви виконували завдання 2 до роботи №5, то цей файл у Вас вже є.**

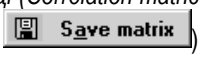
***Перейменування стовпчиків та рядків***

Для спрощення аналізу слід відразу (до початку аналізу) правильно назвати змінні (асоціації), та випадки (досліджуваних). Спосіб перейменування даних можна пригадати в роботі №1. Оскільки програма англомовна, то назви змінних краще робити латинськими літерами, довжиною не більше 8 символів. Використовуйте скорочення при введенні асоціацій.

Після того, як всі назви введені, збережіть зміни ( ).

**Якщо Ви виконували завдання 2 до роботи №5, то ця частина роботи у Вас теж є.**

***Створення вихідної матриці***

Як ми вже зазначали в теоретичній частині, для багатомірного шкалювання необхідно, щоб дані були представлені у вигляді матриці схожості (або кореляційної матриці). Поки що наші дані цю вимогу не задовольняють. Для цього слід відкрити файл даних grid_v.sta, перейти до модуля *Основні статистики (Basic Statistics),* підмодуль *Кореляційні матриці (Correlation matrices)* та натиснути кнопку *Зберегти матрицю (Save matrix* - 

) – рис 6.1.

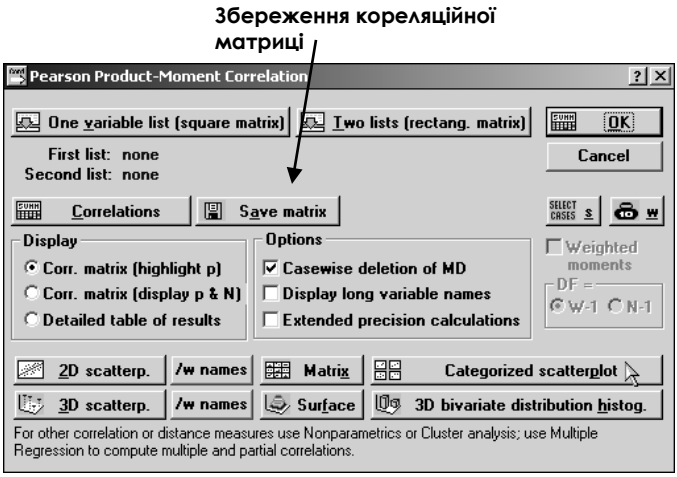



*Рис. 6.1. Збереження кореляційної матриці*

У вікні, що з'явилося слід обрати усі змінні для аналізу (Select All) та OK - рис. 6.2.

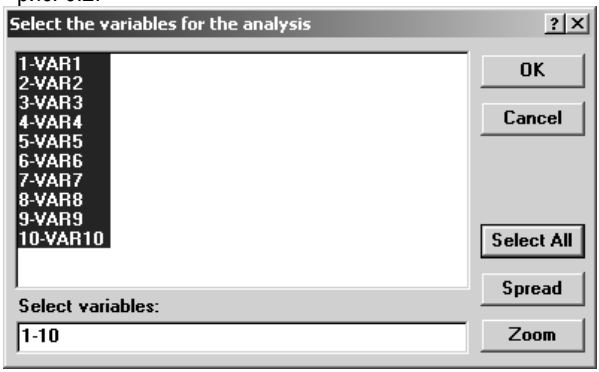



*Рис. 6.2. Обирання змінних для кореляційного аналізу*

У новому вікні слід вказати назву нашої кореляційної матриці, а також місце для її збереження. Збережіть її у свою папку, назвавши ***cor_v.sta***. – рис. 6.3.

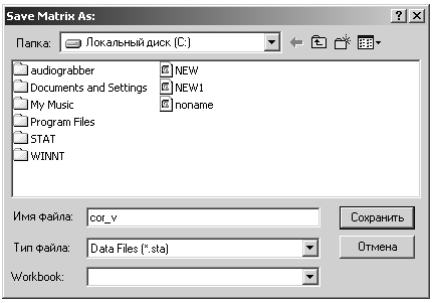
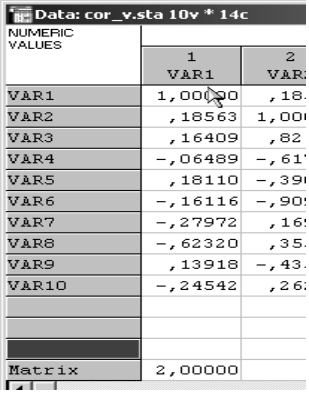



*Рис. 6.3. Вікно збереження кореляційної матриці*

Вихідну кореляційну матрицю для слова “вбивця” створено! Тепер слід закрити усі вікна і програму Statistica.

Багатомірне шкалювання можна проводити і тоді, коли є лише попарні оцінки відстаней між аналізованими об’єктами. У такому випадку вище описані кроки можна пропустити, створивши вихідну матрицю з допомогою модуля *Data Management* із виконанням певних умов:

1. Кількість стрічок = кількість стовпчиків +4.
2. Назви стрічок повинні бути точно такі ж, як назви стовпчиків.
3. Останні 4 стрічки повинні бути пусті і без назв (за виключеням 4-ї).
4. Остання 4-та стрічка повинна називатися matrix. В цій стрічці у першому стовчпику повинно стояти число 2 (матриця схожостей) або число 3 (матриця відмінностей) – рис. 6.4.



Data: cor_v.sta 10v * 14c

| NUMERIC VALUES | 1 VAR1 | 2 VAR |
|---|---|---|
| VAR1 | 1,00000 | ,18 |
| VAR2 | ,18563 | 1,00 |
| VAR3 | ,16409 | ,82 |
| VAR4 | -,06489 | -,61 |
| VAR5 | ,18110 | -,39 |
| VAR6 | -,16116 | -,90 |
| VAR7 | -,27972 | ,16 |
| VAR8 | -,62320 | ,35 |
| VAR9 | ,13918 | -,43 |
| VAR10 | -,24542 | ,26 |
| | | |
| | | |
| | | |
| Matrix | 2,00000 | |

*Рис 6. 4. Фрагмент матриці схожостей, підготовленої до БШ*

***Проведення багатомірного шкалювання***

Запустіть програму Statistica та у вікні модулів оберіть модуль *Багатомірне шкалювання (Multidimensional Scaling)* – рис. 6.5.

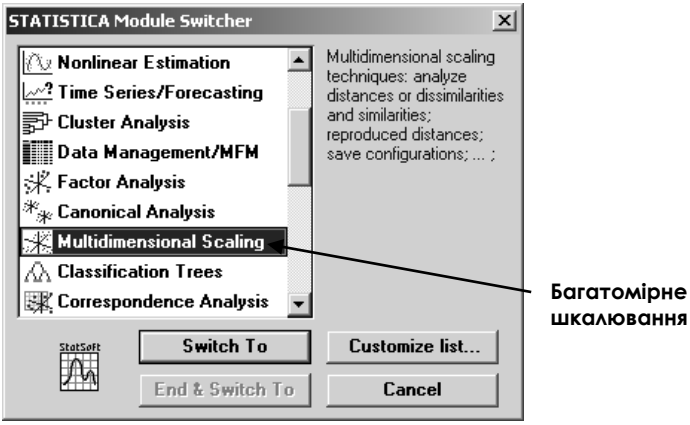



*Рис. 6.5. Запуск Багатомірного шкалювання*

Після цього слід у головному меню обрати пункт *File*, підпункт *Open data…* – рис. 6.6 зліва.

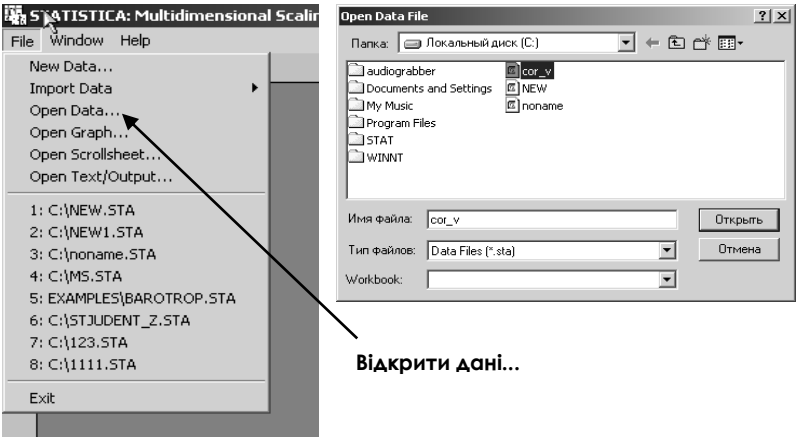



*Рис 6.6. Відкриття файлу даних*

У новому вікні (рис. 6.6, справа) слід знайти свій файл cor_v.sta та відкрити його. Після цього знову у головному меню слід обрати пункт *Analysis*, підпункт *Resume Analysis* – рис. 6.7.

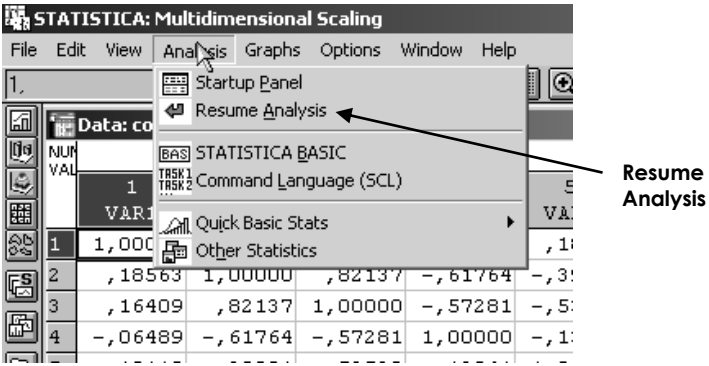



*Рис. 6.7. Команда для початку аналізу*

У новому вікні (рис. 6.8) задаються вихідні параметри для багатомірного шкалювання – змінні (кнопка Variables:), мірність простору (*Number of dimension*) тощо.

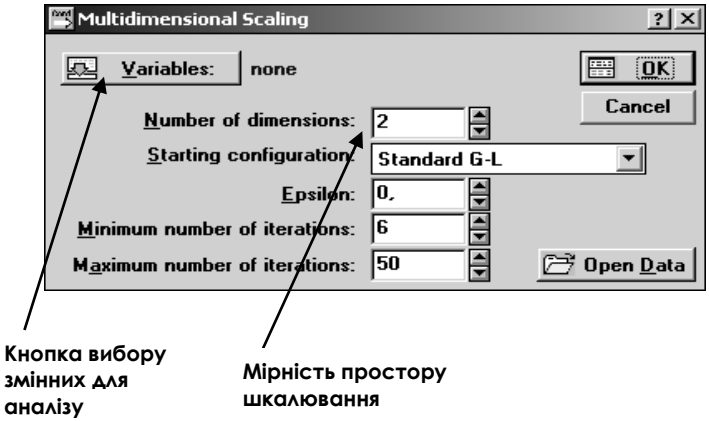



*Рис. 6.8. Вихідні параметри для БШ*

Ви повинні обрати усі змінні для аналізу (кнопка *Select All – рис. 6.9*), а кількість вимірів простору БШ поставити рівною 2-м.

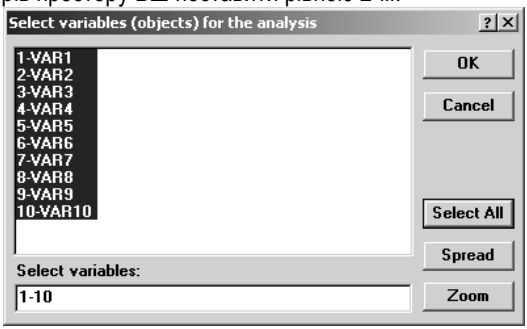
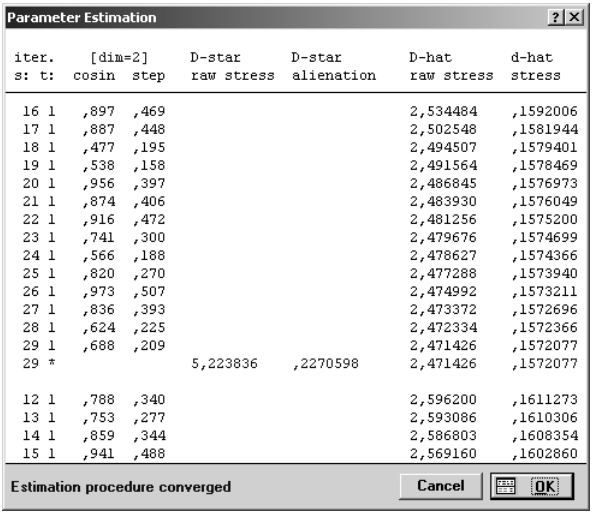



*Рис. 6.9. Вибір змінних для аналізу*

Після того, як усі параметри будуть задані у вікні *Вихідних параметрів* (рис. 6.8) можна натискати *ОК*.

Після цього з'явиться ітераційне вікно, у якому Statistica продемонструє весь хід обчислень – рис. 6.10. Натисніть внизу у цьому вікні кнопку *ОК*.



Parameter Estimation

| iter. s: | t: | [dim=2] cosin | step | D-star raw stress | D-star alienation | D-hat raw stress | d-hat stress |
|---|---|---|---|---|---|---|---|
| 16 | 1 | ,897 | ,469 | | | 2,534484 | ,1592006 |
| 17 | 1 | ,887 | ,448 | | | 2,502548 | ,1581944 |
| 18 | 1 | ,477 | ,195 | | | 2,494507 | ,1579401 |
| 19 | 1 | ,538 | ,158 | | | 2,491564 | ,1578469 |
| 20 | 1 | ,956 | ,397 | | | 2,486845 | ,1576973 |
| 21 | 1 | ,874 | ,406 | | | 2,483930 | ,1576049 |
| 22 | 1 | ,916 | ,472 | | | 2,481256 | ,1575200 |
| 23 | 1 | ,741 | ,300 | | | 2,479676 | ,1574699 |
| 24 | 1 | ,566 | ,188 | | | 2,478627 | ,1574366 |
| 25 | 1 | ,820 | ,270 | | | 2,477288 | ,1573940 |
| 26 | 1 | ,973 | ,507 | | | 2,474992 | ,1573211 |
| 27 | 1 | ,836 | ,393 | | | 2,473372 | ,1572696 |
| 28 | 1 | ,624 | ,225 | | | 2,472334 | ,1572366 |
| 29 | 1 | ,688 | ,209 | | | 2,471426 | ,1572077 |
| 29 | * | | | 5,223836 | ,2270598 | 2,471426 | ,1572077 |
| 12 | 1 | ,788 | ,340 | | | 2,596200 | ,1611273 |
| 13 | 1 | ,753 | ,277 | | | 2,593086 | ,1610306 |
| 14 | 1 | ,859 | ,344 | | | 2,586803 | ,1608354 |
| 15 | 1 | ,941 | ,488 | | | 2,569160 | ,1602860 |

Estimation procedure converged



*Рис. 6.10. Ітераційне вікно*

Після цього з'явиться кінцеве вікно Результати (Results) – рис. 6.11.

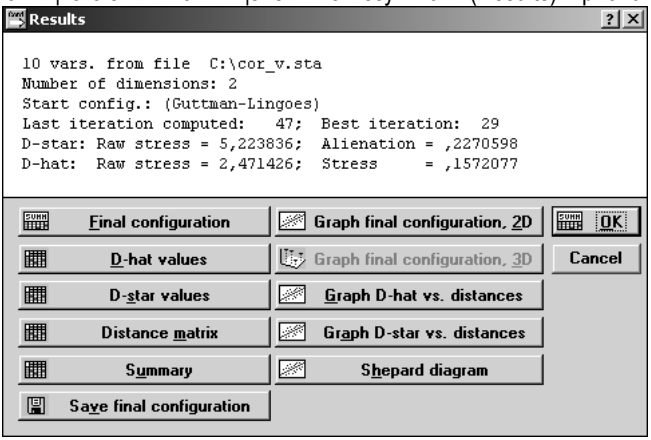



*Рис. 6.11. Результати багатомірного шкалювання*

Нас у цьому вікні цікавитимуть такі результати:

– 2-мірна діаграма розсіювання.

ї – таблиця кінцевої конфігурації.

– оновлена матриця відстаней.

– діаграма Шепарда.

Натискаючи на відповідні кнопки у вікні (рис. 6.11), ми отримуватимемо відповідні результати, наведені нижче. Зберігати їх слід покроково у файл Microsoft Excel під назвою ***Багатомірне шкалювання.xls***.

Для повернення від певного результату до рис. 6.11 необхідно натиснути кнопку *Continue* на цьому результаті (рис. 6.12)

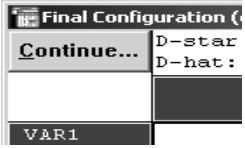



*Рис. 6.12. Кнопка Continue*

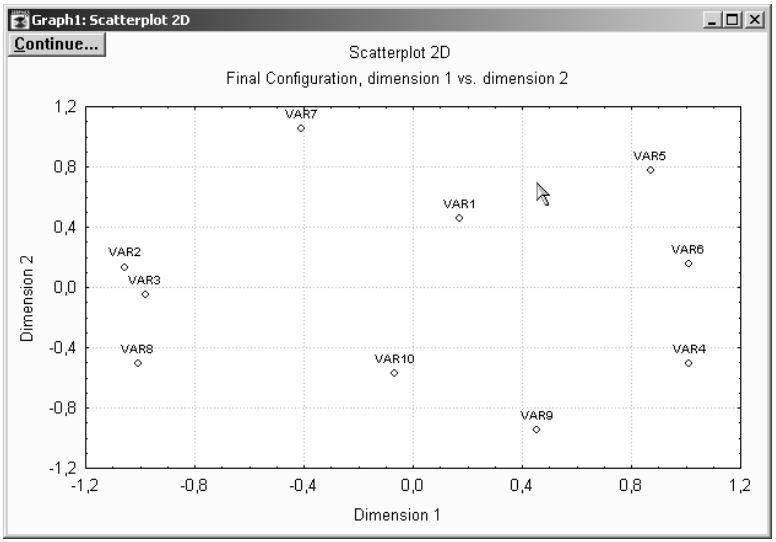



*Рис. 6.13. 2-мірна діаграма розсіювання*

Final Configuration (cor_v.sta)

Continue...

D-star: Raw stress = 5,223836; Alienation = ,2270598
D-hat: Raw stress = 2,471426; Stress = ,1572077

| | DIM. 1 | DIM. 2 |
|---|---|---|
| VAR1 | ,17085 | ,455019 |
| VAR2 | -1,05490 | ,134809 |
| VAR3 | -,97998 | -,052407 |
| VAR4 | 1,01238 | -,504425 |
| VAR5 | ,87144 | ,777737 |
| VAR6 | 1,00939 | ,155139 |
| VAR7 | -,40780 | 1,055875 |
| VAR8 | -1,00764 | -,507516 |
| VAR9 | ,45388 | -,946832 |
| VAR10 | -,06762 | -,567399 |

*Рис. 6.14. Таблиця кінцевої конфігурації*

Distances in Final Configuration (cor_v.sta)

Continue...

D-star: Raw stress = 5,223836; Alienation = ,2270598
D-hat: Raw stress = 2,471426; Stress = ,1572077

| | VAR1 | VAR2 | VAR3 | VAR4 | VAR5 | VAR6 | V |
|---|---|---|---|---|---|---|---|
| VAR1 | 0,000000 | 1,266883 | 1,257733 | 1,276207 | ,771349 | ,890547 | ,8 |
| VAR2 | 1,266883 | 0,000000 | ,201650 | 2,163852 | 2,030801 | 2,064386 | 1,1 |
| VAR3 | 1,257733 | ,201650 | 0,000000 | 2,042993 | 2,029018 | 2,000165 | 1,2 |
| VAR4 | 1,276207 | 2,163852 | 2,042993 | 0,000000 | 1,289884 | ,659570 | 2,1 |
| VAR5 | ,771349 | 2,030801 | 2,029018 | 1,289884 | 0,000000 | ,637696 | 1,3 |
| VAR6 | ,890547 | 2,064386 | 2,000165 | ,659570 | ,637696 | 0,000000 | 1,6 |
| VAR7 | ,834180 | 1,125658 | 1,247272 | 2,109841 | 1,309127 | 1,679204 | 0,0 |
| VAR8 | 1,521618 | ,644061 | ,455949 | 2,020025 | 2,276587 | 2,123093 | 1,6 |
| VAR9 | 1,430138 | 1,856438 | 1,689958 | ,712492 | 1,774400 | 1,234069 | 2,1 |

*Рис. 6.15. Матриця оновлених відстаней*

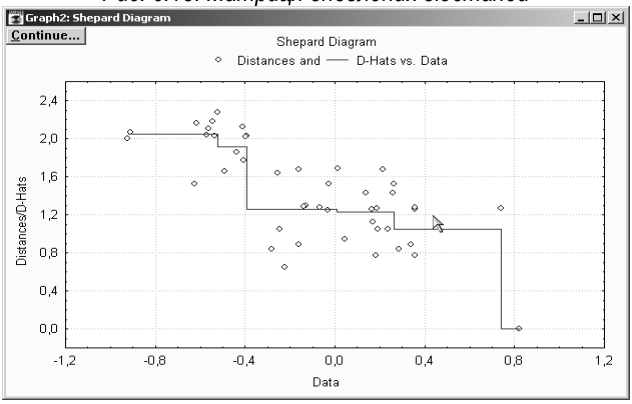



*Рис. 6.16. Діаграма Шепарда*

Найзручніший спосіб **збереження результатів** – це перенесення їх у файл Excel. Зберігати файл слід у свою папку назвавши його “***Багатомірне шкалювання***”.

***Нове багатомірне шкалювання***

Після завершення БШ матриці по слову “вбивця” проведіть багатомірне шкалювання матриці по слову “кіллер”. Файл з даними назвіть grid_k.sta, а вихідну кореляційну матрицю – cor_k.sta.

## ХІД РОБОТИ: ЗАВДАННЯ 3

Інтерпретацію слід проводити, спираючись всі показники, отримані під час аналізу, однак, найкорисніше користуватися діаграмою розсіювання. З'ясуйте:

1. Які виділилися головні виміри у сприйманні студентами слова "вбивця"?
2. Які виділилися головні виміри у сприйманні студентами слова "кіллер"?
3. Чим ці виміри між собою різняться?
4. Які особливості конфігурацій та групування асоціацій у діаграмах, отриманих по матрицям "вбивця" та "кіллер"
5. Як Ви поясните різницю у сприйманні слів "вбивця" та "кіллер"?
6. Порівняйте результати багатомірного шкалювання та факторного аналізу. Що є між ними спільного, а що – відмінного?

Підготуйте власну інтерпретацію у форматі Word та збережіть у свою папку під назвою "***Багатомірне шкалювання***".

## МАТЕРІАЛИ ДЛЯ ЕЛЕКТРОННОГО ЗВІТУ:

*Результати виконання завдання 2:*

Файли даних *grid_v.sta, grid_k.sta*

Вихідні матриці *cor_v.sta* та *cor_k.sta.*

Файл з результатами проведення багатомірного шкалювання *Багатомірне шкалювання.xls* (формат Excel)

*Результати виконання завдання 3:*

Файл з інтерпретацією результатів багатомірного шкалювання та порівняння його з факторним аналізом *Багатомірне шкалювання.doc* (формат Word)

## ЛАБОРАТОРНЕ ЗАНЯТТЯ №7:
## КЛАСТЕРНИЙ АНАЛІЗ

***Теоретична частина:***

1. Мета та методи кластерного аналізу.
2. Підготовка даних до кластерного аналізу.
3. Деревоподібна кластеризація: пошук відстаней між об'єктами, стратегії кластеризації та інтерпретація.
4. Представлення результатів кластерного аналізу.
5. Кластерний аналіз за методом В.Ю. Крилова та Т.В. Острякової.

**Література для підготовки до заняття:**


1. Крылов В.Ю., Острякова Т.В. Новые методы кластерного анализа на основе психологической теории развития понятий Л.С. Выготского // Психологический журнал. – 1995. – Т16. – №1. – С. 130-137.
2. Общая психодиагностика. Основы психодиагностики, немедицинской психотерапии и психологического консультирования. /Под ред. А.А. Бодалева, В.В. Столина. – М.: Изд-во Моск.ун-та, 1987. – 304 с
3. Петренко В.Ф. Психосемантика сознания. – М.: Изд-во Моск. ун-та, 1988.


***Практична частина:***

**Завдання 1.** **Підготовка даних до багатомірного шкалювання.**

*Провести до заняття порівняльне дослідження сприймання студентами слів "вбивця" та "кіллер"[23]. Якщо виконувалося завдання 1 до лабораторної роботи №5 (Факторний аналіз), дослідження можна не проводити – слід скористатися вже готовими результатами.*

**Завдання 2.** **Проведення кластерного аналізу.**

*Зробити кластерний аналіз результатів дослідження.*

**Завдання 3.** **Підготовка висновків за результатами**

*Виявити головні кластери у сприйманні слова "вбивця" та слова "кіллер". Проаналізувати отримані кластери. Порівняти результати КА з результатами БШ та ФА.*

---

[23] *Музика О.Л. Психологічні механізми ціннісно-мовленнєвої взаємодії / Національна освіта: традиції та інновації у контексті ідей Івана Огієнка: Збірник наукових праць / За ред. проф. М.В. Левківського. – Київ-Житомир, ЖДПУ, 2002. – С. 101-106.*

## ХІД РОБОТИ: ЗАВДАННЯ 1

***Див. лабораторне заняття №5***

## ХІД РОБОТИ: ЗАВДАННЯ 2

***Створення файлу даних***

Запустіть STATISTICA 5.5, модуль *Data Management* та створіть спочатку файл даних з матрицею асоціацій на слово “вбивця” (grid_v.sta) на потрібну Вам кількість стовпчиків (асоціацій) та рядків (досліджуваних). Як це робиться, можна пригадати, повернувшись до лабораторної роботи №1, пункт *Створення та збереження файлу даних.*

**Якщо Ви виконували завдання 2 до роботи №5, то цей файл у Вас вже є.**

***Перейменування стовпчиків та рядків***

Для спрощення аналізу слід відразу (до початку аналізу) правильно назвати змінні (асоціації), та випадки (досліджуваних). Спосіб перейменування даних можна пригадати в роботі №1. Оскільки програма англомовна, то назви змінних краще робити латинськими літерами, довжиною не більше 8 символів. Використовуйте скорочення при введенні асоціацій.

Після того, як всі назви введені, збережіть зміни ( ).

**Якщо Ви виконували завдання 2 до роботи №5, то ця частина роботи у Вас теж є.**

***Проведення кластерного аналізу***

Перш за все необхідно перейти до модуля *Кластерного аналізу (Cluster analysis)*. Це можна зробити, обравши пункт *Analysis*, підпункт *Other Statistics* (рис. 7.1, зліва). З’явиться вікно модулів, в якому можна обрати потрібний розділ (рис. 7.1, справа).

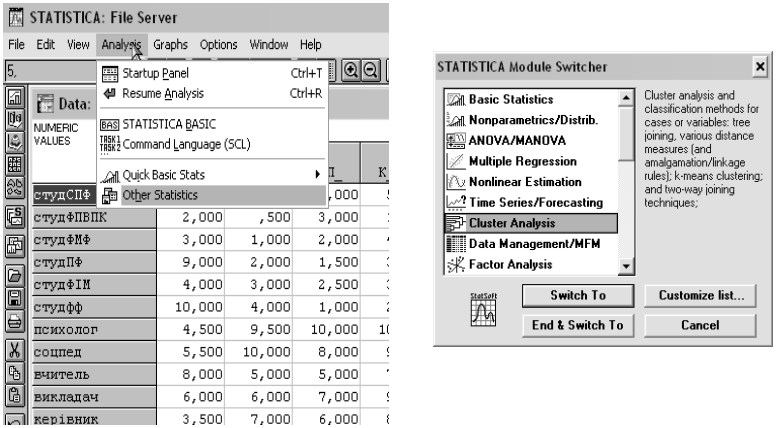



*Рис. 7.1. Перехід до модуля кластерного аналізу*

Автоматично після запуску модуля Кластерний аналіз запускається діалогове вікно (рис. 7.2)

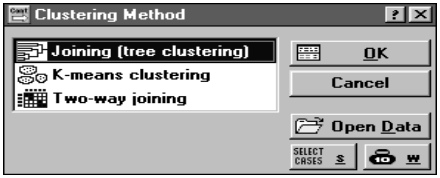



*Рис. 7.2. Діалогове вікно кластерного аналізу*

У цьому вікні слід вибрати один із методів кластерного аналізу. Нас цікавитиме Деревоподібна кластеризація, а тому слід обрати пункт *Joining (tree clustering)* і натиснути *OK*.

Відразу з'явиться наступне вікно, в якому Вам пропонується обрати параметри кластерного аналізу (рис. 7.3):

- обрати змінні для аналізу (кнопка Variables:)
- обрати спосіб кластеризації (по стовпчиках – Variables, по стрічках - Cases)
- обрати стратегію кластеризації
- вибрати спосіб вимірювання відстаней між об'єктами

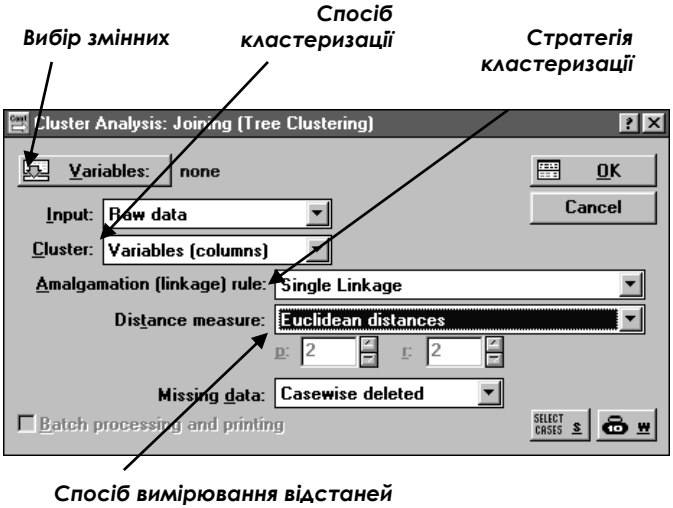



*Рис. 7.3. Параметри кластерного аналізу*

Зліва вгорі в цьому вікні розміщена кнопка *Змінні* (Variables:). Її слід натиснути, щоб обрати ті змінні, по яких Ви будете проводити факторний аналіз.

Після натискування на цю кнопку з'явиться нове діалогове вікно, представлене на р

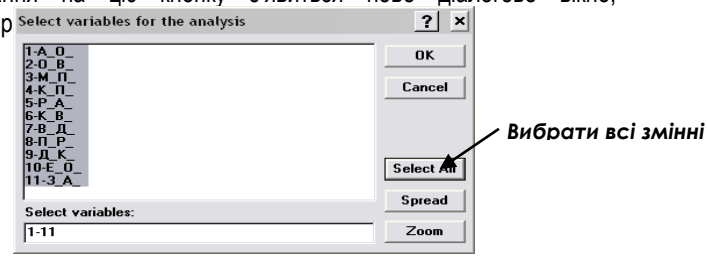



*Рис. 7.4. Вікно вибору змінних для факторного аналізу*

Оскільки нам для аналізу будуть потрібні всі змінні, то натисніть кнопку *Вибрати Всі (Select All)*. Після того натисніть кнопку *ОК.*

Ви повернулися до вікна параметрів кластерного аналізу (рис. 7.3). В більшості випадків усі інші параметри не змінюють, – так вчинимо і ми.

Для початку кластерного аналізу натисніть кнопку ОК. Перед вами (рис. 7.5) з'явиться вікно *Результати Об'єднання (Joining Results). Це і є результати кластерного аналізу. Натискаючи відповідні кнопки Ви отримаєте необхідні для аналізу таблиці та графіки.*

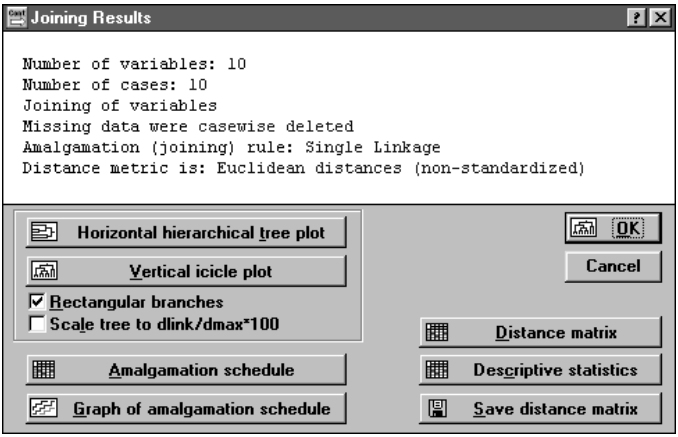



*Рис. 7.5. Результати кластерного аналізу*

Для опису результатів кластерного аналізу нам будуть необхідні такі результати:

- таблиця відстаней (Distance matrix);
- таблиця об'єднань (Amalgamation schedule);
- горизонтальне чи вертикальне кластерне дерево (Horizontal hierarchical tree plot або Vertical icicle plot).

*Таблиця відстаней* має такий вигляд:

Euclidean distances (claster.sta)

| Continue... | А_О | О_В_ | М_П_ | К_П_ | Р_А_ | К_В_ | В_Д_ | П_Р_ | Д_К_ | Е_О_ | З_А_ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| А_О_ | ,0 | 13,3 | 14,8 | 13,4 | 13,5 | 11,5 | 14,4 | 13,6 | 12,8 | 12,8 | 12,8 |
| О_В_ | 13,3 | ,0 | 6,2 | 7,0 | 7,3 | 5,9 | 16,8 | 12,9 | 21,7 | 21,7 | 21,7 |
| М_П_ | 14,8 | 6,2 | ,0 | 6,8 | 6,7 | 8,0 | 14,9 | 12,4 | 20,4 | 20,4 | 20,4 |
| К_П_ | 13,4 | 7,0 | 6,8 | ,0 | 4,0 | 8,8 | 13,4 | 11,0 | 19,6 | 19,6 | 19,6 |
| Р_А_ | 13,5 | 7,3 | 6,7 | 4,0 | ,0 | 8,9 | 12,3 | 9,0 | 18,9 | 18,9 | 18,9 |
| К_В_ | 11,5 | 5,9 | 8,0 | 8,8 | 8,9 | ,0 | 16,4 | 13,8 | 18,7 | 18,7 | 18,7 |
| В_Д_ | 14,4 | 16,8 | 14,9 | 13,4 | 12,3 | 16,4 | ,0 | 6,6 | 12,9 | 12,9 | 12,9 |
| П_Р_ | 13,6 | 12,9 | 12,4 | 11,0 | 9,0 | 13,8 | 6,6 | ,0 | 16,0 | 16,0 | 16,0 |
| Д_К_ | 12,8 | 21,7 | 20,4 | 19,6 | 18,9 | 18,7 | 12,9 | 16,0 | ,0 | ,0 | ,0 |
| Е_О_ | 12,8 | 21,7 | 20,4 | 19,6 | 18,9 | 18,7 | 12,9 | 16,0 | ,0 | ,0 | ,0 |
| З_А_ | 12,8 | 21,7 | 20,4 | 19,6 | 18,9 | 18,7 | 12,9 | 16,0 | ,0 | ,0 | ,0 |

*Рис. 7.6. Таблиця відстаней*

З цієї таблиці можна отримати інформацію про відстані між окремими об'єктами – чим більше числове значення в певній комірці, тим більша відстань між об'єктами.

Однак, на основі цієї таблиці важко робити певні узагальнення та висновки. Більш наочною є *таблиця об'єднань* (рис. 7.7):

Amalgamation Schedule (claster.sta)

Continue... Single Linkage
Euclidean distances

| linkage distance | Obj. No. 1 | Obj. No. 2 | Obj. No. 3 | Obj. No. 4 | Obj. No. 5 | Obj. No. 6 | Obj. No. 7 | Obj. No. 8 | Obj. No. 9 | Obj. No. 10 | Obj. No. 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ,0000000 | Д_К_ | Е_О_ | | | | | | | | | |
| ,0000000 | Д_К_ | Е_О_ | З_А_ | | | | | | | | |
| 4,031129 | К_П_ | Р_А_ | | | | | | | | | |
| 5,937171 | О_В_ | К_В_ | | | | | | | | | |
| 6,244998 | О_В_ | К_В_ | М_П_ | | | | | | | | |
| 6,614378 | В_Д_ | П_Р_ | | | | | | | | | |
| 6,689544 | О_В_ | К_В_ | М_П_ | К_П_ | Р_А_ | | | | | | |
| 9,041571 | О_В_ | К_В_ | М_П_ | К_П_ | Р_А_ | В_Д_ | П_Р_ | | | | |
| 11,51086 | А_О_ | О_В_ | К_В_ | М_П_ | К_П_ | Р_А_ | В_Д_ | П_Р_ | | | |
| 12,81601 | А_О_ | О_В_ | К_В_ | М_П_ | К_П_ | Р_А_ | В_Д_ | П_Р_ | Д_К_ | Е_О_ | З_А_ |

*Рис. 7.7. Таблиця об'єднань*

Як бачимо, в цій таблиці зміст кластерів представлений дещо по іншому – зліва розміщені *відстані* (*linkage distance*), зверху кількість об'єктів у кластері, а в самій таблиці розміщені вже об'єкти. Тепер уже видно, що до

першого кластеру входять 2 об'єкти, до другого – 3, до третього – 5 і т.д.

Ще наочніше можна представити результати кластерного аналізу на кластерному дереві (рис. 7.8):

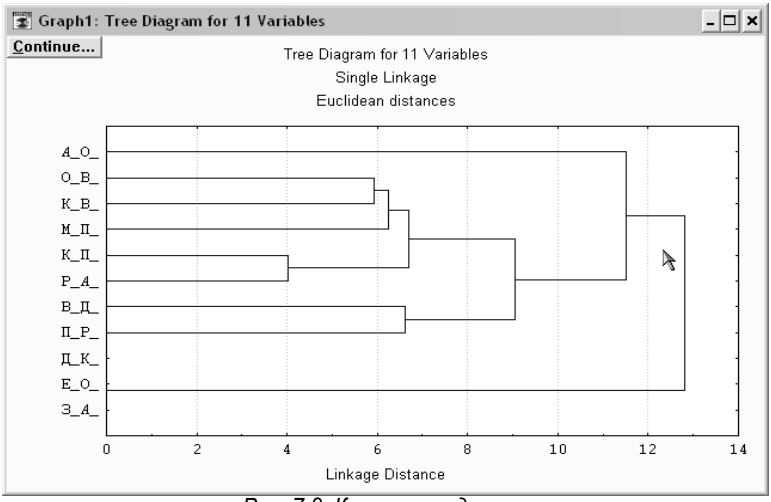



*Рис. 7.8. Кластерне дерево*

***Збереження результатів кластерного аналізу***

Кластерний аналіз проведено, однак, одним із елементів експериментальної роботи є грамотне представлення її результатів. Таке представлення переслідує декілька цілей:

1. Полегшення інтерпретації кластерних структур.
2. Можливість перенесення числових результатів до тексту наукової роботи.
3. Переконання читачів роботи в достовірності отриманих Вами даних.

Саме тому всі отримані дані повинні бути представлені як в тексті (кластерне дерево), так і в додатках (таблиці відстаней та об'єднань).

Для зручності роботи збережуть усі вище названі параметри до файлу Кластерний аналіз.xls (формат Excel).

## ХІД РОБОТИ: ЗАВДАННЯ 3

Інтерпретацію слід проводити, спираючись на всі отримані в результаті аналізу показники, однак, найкорисніше користуватися кластерним деревом. З'ясуйте:

1. Які виділилися кластери у сприйманні студентами слова "вбивця"?
2. Які виділилися кластери у сприйманні студентами слова "кіллер"?
3. Чим ці кластери між собою різняться?
4. Які особливості дендрограм, отриманих по матрицям "вбивця" та "кіллер"
5. Як Ви поясните різницю у сприйманні слів "вбивця" та "кіллер"?
6. Порівняйте результати кластерного аналізу з багатомірним шкалюванням та факторним аналізом. Що є між ними спільного, а що – відмінного?

Підготуйте свою інтерпретацію у форматі Word та збережіть у свою папку під назвою "***Кластерний аналіз***".

## МАТЕРІАЛИ ДЛЯ ЕЛЕКТРОННОГО ЗВІТУ:

*Результати виконання завдання 2:*

Файли даних *grid_v.sta, grid_k.sta*

Файл з результатами проведення кластерного аналізу *Кластерний аналіз.xls* (формат Excel)

*Результати виконання завдання 3:*

Файл з інтерпретацією результатів кластерного та порівняння його з факторним аналізом та багатомірним шкалюванням *Кластерний аналіз.doc* (формат Word)

# ПОКАЖЧИК СТАТИСТИЧНИХ КРИТЕРІЇВ

L-критерій тенденцій Пейджа – § 3.3.
S-критерій Джонкіра – § 3.2.
t-критерій Стьюдента для залежних вибірок – § 3.3.
t-критерій Стьюдента для незалежних вибірок – § 3.2.
U-критерій Манна-Уітні – § 3.2.
Бісеріальний коефіцієнт кореляції – § 2.2.
Бісеріальний коефіцієнт рангової кореляції – § 2.2.
Дисперсія – § 1.4.
Коефіцієнт т-Кендалла – § 2.2.
Коефіцієнт варіації – § 1.4.
Коефіцієнт кореляції φ – § 2.2.
Коефіцієнт кореляції Пірсона – § 2.1.
Коефіцієнт осциляції – § 1.4.
Коефіцієнт рангової кореляції Спірмена – § 2.2.
Коефіцієнти регресії – § 2.3.
Критерій $\varphi^*$ – кутове перетворення Фішера – § 3.4.
Критерій $\chi_r^2$ Фрідмана – § 3.3.
Критерій λ Колмогорова-Смірнова – § 3.4.
Критерій $\chi^2$-Пірсона – § 3.4.
Критерій знаків (G) – § 3.3.
Критерій Кайзера – § 5.1.
Критерій кам'янистого насипу – § 5.1.
Критерій Фішера F – § 4.2.
Медіана – § 1.3.
Множинна кореляція – § 2.2.
Мода – § 1.3.
Розмах – § 1.4.
Середнє арифметичне – § 1.3.
Стандартне відхилення – § 1.4.
Тетрахоричний коефіцієнт кореляції – § 2.2.
Точковий бісеріальний коефіцієнт кореляції – § 2.2.

# РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ІНТЕРНЕТ-ДЖЕРЕЛА

## ПІДРУЧНИКИ ТА ПОСІБНИКИ


1. Артемьева Е.Ю., Мартынов Е.М. Вероятностные методы в психологии. – М., 1975. – 206 с.
2. Бешелев С.Д., Гурвич Ф.Г. Математико-статистические методы экспертных оценок. – М., 1980. – 263 с.
3. Битинас В.П. Многомерный анализ в педагогике и педагогической психологии. – Вильнюс, 1971. – 347 с.
4. Венецкий И.Г. Вариационные ряды и их характеристики. – М.: Статистика, 1970. – 158 с.
5. Власов К.П., Власов П.К., Киселева А.А. Методы исследований и организация экспериментов. – Х.: Гуманитарный центр, 2002. – 256 с.
6. Выборочный метод в социологии / Под ред. А. Ермолаева. – М., 2000. – 26 с.
7. Гайда В.К. Приемы измерений и статистические способы обработки их результатов в психологическом исследовании // Практикум по общей, экспериментальной и прикладной психологии / В.Д. Балин, В.К. Гайда, В.К. Гербачевский и др. / Под общей ред. А.А. Крылова, С.А. Маничева. – СПб.: Питер, 2000. – С. 5-33
8. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М., 1976. – 495 с.
9. Готтсданкер Р. Основы психологического эксперимента/ Пер. с англ. – М., 1982. – 464 с.
10. Гудвин Дж. Иследование в психологии: методы и планирование. – 3-е изд. – СПб.: Питер, 2004. – 558 с.
11. Гусев А.Н., Измайлов Ч.А., Михалевская М.Б. Измерение в психологии: общий психологический практикум – М.: Смысл, 1997. – 286 с.
12. Дейвисон М. Многомерное шкалирование. Методы наглядного представления данных: Пер. с англ. В.С. Каменского с предсл. С.А. Айвазяна и В.С. Каменского. – М.: Финансы и статистика, 1988. – 254 с.
13. Дронов С.В. Многомерный статистический анализ. – Барнаул: Изд-во Алтайского гос. ун-та, 2003. – 213 с.
14. Дюк В.А. Компьютерная психодиагностика. - СПб.: Братство, 1994. – 363 с.
15. Захаров В.П. Применение математических методов в социально-психологических исследованиях. – Л.: ЛГУ, 1985. – 64 с.
16. Мартин Д. Психологические эксперименты: Секреты механизмов

психики. – СПб.: прайм-ЕВРОЗНАК, 2002. – 480 с.
17. Математическая психология: методология, теории, модели / Под ред. В.Ю. Крылова. – М.: Наука, 1965. – 236 с.
18. Мельников В.М., Ямпольский Л.Т. Введение в экспериментальную психологию личности. – М., 1985. – 319 с.
19. Музика О.Л. Курсові роботи з психології. Навчальний посібник. – 2-ге вид., перероб і доп. – Житомир: Вид-во ЖДУ ім. І. Франка, 2007. – 104 с.
20. Наследов А.Д. Математические методы психологического исследования. – СПб.: Речь, 2004. – 392 с.
21. Общая психодиагностика. Основы психодиагностики, немедицинской психотерапии и психологического консультирования. /Под ред. А.А. Бодалева, В.В. Столина. – М.: Изд-во Моск.ун-та, 1987. – 304 с.
22. Паниотто В.И. Структура межличностных отношений: Методика и математические методы исследования. – К., 1975. – 127 с.
23. Петренко В.Ф. Психосемантика сознания. – М.: Изд-во Моск. ун-та, 1988. – 208 с.
24. Похилько В.И., Федотова Е.О. Техника репертуарных решеток в экспериментальной психологии личности //Вопр. психол. – № 3, 1984. – С.151-157.
25. Свалов Н.Н. Вариационная статистика. – М.: Лесная промышленность, 1977. – 177 с.
26. Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО “Речь”, 2000. – 350 с.
27. Солсо Р.Л., Джонсон Х.Х., Бил М.К. Экспериментальная психология: практический курс. – СПб.: прайм-ЕВРОЗНАК, 2001. – 528 с.
28. Справочник по прикладной статистике. В 2-х т. Т.1. /Под ред. Э. Ллойда, У. Ледермана, Ю.Н. Тюрина. – М.: Финансы и статистика, 1989. – 511 с.
29. Суходольский Г.В. Математические методы в психологии. – Харьков: Изд-во “Гуманитарный центр”, 2004. – 284 с.
30. Суходольский Г.В. Основы математической статистики для психологов. – Л.: ЛГУ, 1972. – 428 с.
31. Теорія статистики: Навчальний посібник /Вашків П.Г., Пастер П.І., Сторожук В.П., Ткач Є.І. – К.: Либідь, 2001. – 320 с.
32. Толстова Ю.Н. Основы многомерного шкалирования. – М.: Университет. – 157 с.
33. Урбах В.Ю. Статистический анализ в биологических и медицинских исследованиях. – М.: Медицина, 1975. – 295 с.

34. Франселла Ф. Баннистер Д. Новый метод иследования личности – М., 1987. – 236 с.
35. Хазанов Ю.С. Статистика. – М.: Статистика, 1974. – 192 с.
36. Циба В.Т. Основи теорії кваліметрії: навчальний посібник. – К.: ІЗМН, 1997. – 160 с.

**ПЕРІОДИЧНІ ВИДАННЯ**

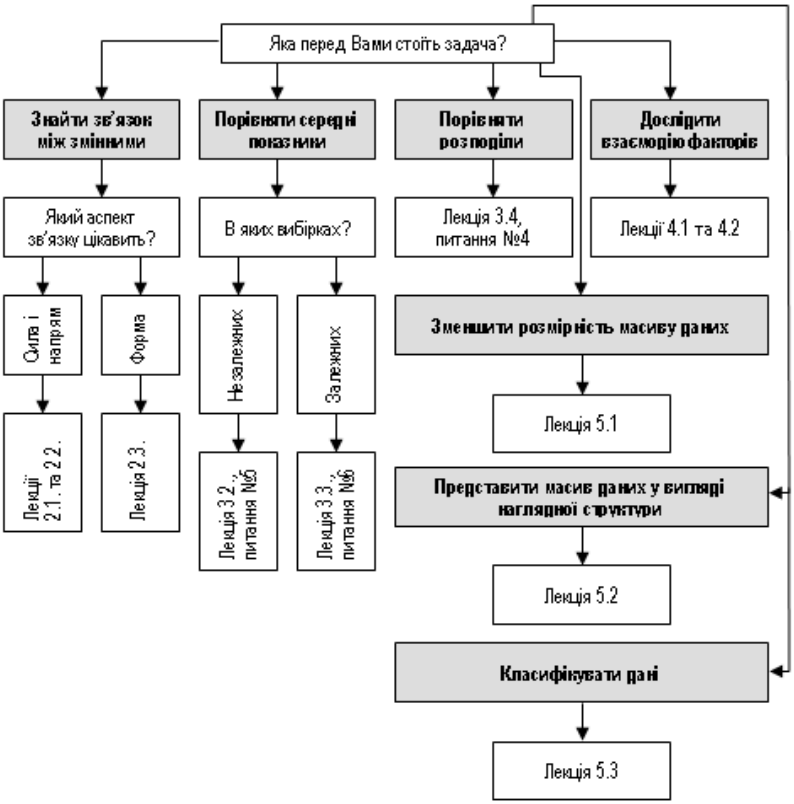
37. Балл Г.А. Проблемы взаимодействия психологии с формализированными научными дисциплинами // Психол. журн. – 1989. – Т.10. – №6. – С. 34-40.
38. Беспалько И.Г. О некоторых неясных вопросах психологической интерпретации факторов в факторном анализе // Психол. журн. – 1987. – Т.8. – №3. – С. 137-145.
39. Джужа Н.В. Применение методов непараметрической статистики в психолого-педагогических исследованиях // Вопр. психол. – 1987. – №4. – С. 145-151.
40. Докторов Б.З. Об использовании методов факторного анализа в работах советских исследователей //Вопр. психол. – 1969. – №2. – С. 142-146.
41. Ивлиев Ю.А. Новые математические методы в психологии, их разработка и применение (проблемное исследование) // Психол. журн. – 1988. – Т.9. – №1.
42. Ивлиев Ю.А. О методе математического моделирования субъективных отношений // Психол. журн. – 1987. – Т.8. – №5. – С. 137-140.
43. Климчук В.А. Проблемы преподавания дисциплины “Математические методы в психологии” // Тезисы юбилейной научной конференции “Тенденции развития современной психологической науки” (Москва, 31 января – 1 февраля 2007 г.). – М., 2007. – С. 37.
44. Климчук В.О. Багатомірне шкалювання: використання у психологічних дослідженнях // Практична психологія та соціальна робота. – 2007. – №6. – С. 17-21.
45. Климчук В.О. Відображення у свідомості студентів-психологів термінологічного апарату психології здібностей // Науковий часопис НПУ імені М.П. Драгоманова, 2005. – №6 (30). – Ч.І. – С. 369-374.
46. Климчук В.О. Викладання курсу “Математичні методи у психології” в умовах кредитно-модульної системи // Соціальна психологія. – 2008. – №2 (28). – С. 180-189.
47. Климчук В.О. Використання потенціалу кредитно-модульної системи у

вирішенні проблем викладання курсу “Математичні методи у психології” // Вища освіта України – Додаток 3 (т.2). – Тематичний випуск “Вища освіта України у контексті інтеграції до європейського освітнього простору”. – К.: Гнозис, 2006. – С. 156-162.
48. Климчук В.О. Життєвий шлях особистості. Математичне моделювання за допомогою багатомірного шкалювання // Соціальна психологія – 2006. – №4(18). – С. 182-191.
49. Климчук В.О. Особливості інформатизації викладання дисципліни “Математичні методи у психології” в умовах кредитно-модульної системи // Інформаційне забезпечення начального процесу у вищій школі: Тези доп. наук.-практ. конф. 19-21 травня 2007. – О.: Наука і техніка, 2007. – С. 209-210.
50. Климчук В.О. Кластерний аналіз: використання у психологічних дослідженнях // Практична психологія та соціальна робота. – 2006. – №4. – С. 30-36.
51. Климчук В.О. Факторний аналіз: використання у психологічних дослідженнях // Практична психологія та соціальна робота. – 2006. – №8. – С. 43-48.
52. Коренев Г.В. Математическая модель психомоторного акта // Психол. журн. – 1981. – Т.2. – №3. – С. 124-140.
53. Крылов В.Ю. Актуальные проблемы математической психологии // Психол. журн. – 1992. – Т.13. – №6. – С. 13-25.
54. Крылов В.Ю. Математические методы в психологии // Психол. журн. – 1980. – Т.1. – №6. – С.26-35.
55. Крылов В.Ю. Метод математического моделирования структуры субъективных отношений // Психол. журн. – 1986. – Т.7. – №4. – С. 36-46.
56. Крылов В.Ю. О многомерном шкалировании в неметрических пространствах // Психол. журн. – 1987. – Т.8. – №5. – С. 140-142.
57. Крылов В.Ю., Острякова Т.В. Новые методы кластерного анализа на основе психологической теории развития понятий Л.С. Выготского // Психологический журнал. – 1995. – Т16. – №1. – С. 130-137.
58. Куравский Л.С., Малых С.Б. Применение марковских моделей для анализа эволюции психологических характеристик в популяции // Вопр. психол. – 2003. – №4. – С. 51-62.
59. Левандовский Н.Г. О корректности применения факторного анализа и о критериях факторизации // Вопр. психол. – 1980. – №5. – С. 138-143.

60. Лосик Г.В. Исследование восприятия гласных методом многомерного шкалирования // Психол. журн. – 1992. – Т.13. – №2.
61. Налимов В.В. О возможности метафорического использования математических представлений в психологии // Психол. журн. – 1981. – Т.2. – №3 – С. 39-48.
62. Орехов А.Н., Ильясов И.И. Опыт построения математизированной теории психических процессов // Психол. журн. – 1985. – Т.6. – №1.
63. Паповян С.С. Эксперимент в прикладной социальной психологии: актуальные проблемы статистического анализа данных // Психол. журн. – 1985. – Т.6. – №6. – С. 90-101.
64. Прісняковa Л. Можливості математичного моделювання психологічних процесів // Психологія і суспільство. – 2003. – №2. – С. 90-96.
65. Савченко Т.Н. Развитие математической психологии: история и перспективы // Психол. журн. – 2002. – Т.23. – №5. – С. 32-41.
66. Степанский В.И., Осницкий А.К., Гренадер В.Б. Применение нормированных коэффициентов в исследовании вероятностного прогнозирования // Вопр. психол. – 1980. – №1. – С. 116-121.
67. Тарасов В.Б., Чернышев А.П. О применении “нечеткой” математики в инженерной психологии // Психол. журн. – 1981. – Т.2. – №4. – С. 110-123.
68. Терехина А.Ю. Многомерное шкалирование в психологии // Психол. журн. – 1983. – Т.4. – №1.
69. Фридман Л.М. Моделирование в психологии и психология моделирования //Вопр. психол. – 1977. – №2. – С. 15-29.
70. Фридман Л.М. О путях развития математической психологии // Вопр. психол. – 1970. – №4. – С. 13-24.
71. Шепард Р.Н. Многомерное шкалирование и безразмерное представление различий // Психол. журн. – 1980. – Т.1. – №4. – С. 72-84.

# ДОДАТКИ

## ДОДАТОК 1: ЗАГАЛЬНИЙ АЛГОРИТМ ВИБОРУ МЕТОДУ СТАТИСТИЧНОЇ ОБРОБКИ



# ДОДАТОК **2: СТАТИСТИЧНІ ТАБЛИЦІ**

**Таблиця 1.**

***Величини кута φ для різних відсоткових долей***[24]

| % | %, останній десятковий знак | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| | Значення φ | | | | | | | | | |
| 0,0 | 0,000 | 0,020 | 0,028 | 0,035 | 0,040 | 0,045 | 0,049 | 0,053 | 0,057 | 0,060 |
| 0,1 | 0,063 | 0,066 | 0,069 | 0,072 | 0,075 | 0,077 | 0,080 | 0,082 | 0,085 | 0,087 |
| 0,2 | 0,089 | 0,092 | 0,094 | 0,096 | 0,098 | 0,100 | 0,102 | 0,104 | 0,106 | 0,108 |
| 0,3 | 0,110 | 0,111 | 0,113 | 0,115 | 0,117 | 0,118 | 0,120 | 0,122 | 0,123 | 0,125 |
| 0,4 | 0,127 | 0,128 | 0,130 | 0,131 | 0,133 | 0,134 | 0,136 | 0,137 | 0,139 | 0,140 |
| 0,5 | 0,142 | 0,143 | 0,144 | 0,146 | 0,147 | 0,148 | 0,150 | 0,151 | 0,153 | 0,154 |
| 0,6 | 0,155 | 0,156 | 0,158 | 0,159 | 0,160 | 0,161 | 0,163 | 0,164 | 0,165 | 0,166 |
| 0,7 | 0,168 | 0,169 | 0,170 | 0,171 | 0,172 | 0,173 | 0,175 | 0,176 | 0,177 | 0,178 |
| 0,8 | 0,179 | 0,180 | 0,182 | 0,183 | 0,184 | 0,185 | 0,186 | 0,187 | 0,188 | 0,189 |
| 0,9 | 0,190 | 0,191 | 0,192 | 0,193 | 0,194 | 0,195 | 0,196 | 0,197 | 0,198 | 0,199 |
| 1 | 0,200 | 0,210 | 0,220 | 0,229 | 0,237 | 0,246 | 0,254 | 0,262 | 0,269 | 0,277 |
| 2 | 0,284 | 0,291 | 0,298 | 0,304 | 0,311 | 0,318 | 0,324 | 0,330 | 0,336 | 0,342 |
| 3 | 0,348 | 0,354 | 0,360 | 0,365 | 0,371 | 0,376 | 0,382 | 0,387 | 0,392 | 0,398 |
| 4 | 0,403 | 0,408 | 0,413 | 0,418 | 0,423 | 0,428 | 0,432 | 0,437 | 0,442 | 0,446 |
| 5 | 0,451 | 0,456 | 0,460 | 0,465 | 0,469 | 0,473 | 0,478 | 0,482 | 0,486 | 0,491 |
| 6 | 0,495 | 0,499 | 0,503 | 0,507 | 0,512 | 0,516 | 0,520 | 0,524 | 0,528 | 0,532 |
| 7 | 0,536 | 0,539 | 0,543 | 0,547 | 0,551 | 0,555 | 0,559 | 0,562 | 0,566 | 0,570 |
| 8 | 0,574 | 0,577 | 0,581 | 0,584 | 0,588 | 0,592 | 0,595 | 0,599 | 0,602 | 0,606 |
| 9 | 0,609 | 0,613 | 0,616 | 0,620 | 0,623 | 0,627 | 0,630 | 0,633 | 0,637 | 0,640 |
| 10 | 0,644 | 0,647 | 0,650 | 0,653 | 0,657 | 0,660 | 0,663 | 0,666 | 0,670 | 0,673 |
| 11 | 0,676 | 0,679 | 0,682 | 0,686 | 0,689 | 0,692 | 0,695 | 0,698 | 0,701 | 0,704 |
| 12 | 0,707 | 0,711 | 0,714 | 0,717 | 0,720 | 0,723 | 0,726 | 0,729 | 0,732 | 0,735 |
| 13 | 0,738 | 0,741 | 0,744 | 0,747 | 0,750 | 0,752 | 0,755 | 0,758 | 0,761 | 0,764 |
| 14 | 0,767 | 0,770 | 0,773 | 0,776 | 0,778 | 0,781 | 0,784 | 0,787 | 0,790 | 0,793 |
| 15 | 0,795 | 0,798 | 0,801 | 0,804 | 0,807 | 0,809 | 0,812 | 0,815 | 0,818 | 0,820 |
| 16 | 0,823 | 0,826 | 0,828 | 0,831 | 0,834 | 0,837 | 0,839 | 0,842 | 0,845 | 0,847 |
| 17 | 0,850 | 0,853 | 0,855 | 0,858 | 0,861 | 0,863 | 0,866 | 0,868 | 0,871 | 0,874 |
| 18 | 0,876 | 0,879 | 0,881 | 0,884 | 0,887 | 0,889 | 0,892 | 0,894 | 0,897 | 0,900 |
| 19 | 0,902 | 0,905 | 0,907 | 0,910 | 0,912 | 0,915 | 0,917 | 0,920 | 0,922 | 0,925 |
| 20 | 0,927 | 0,930 | 0,932 | 0,935 | 0,937 | 0,940 | 0,942 | 0,945 | 0,947 | 0,950 |
| 21 | 0,952 | 0,955 | 0,957 | 0,959 | 0,962 | 0,964 | 0,967 | 0,969 | 0,972 | 0,974 |
| 22 | 0,976 | 0,979 | 0,981 | 0,984 | 0,986 | 0,988 | 0,991 | 0,993 | 0,996 | 0,998 |

[24] *Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО “Речь”, 2000. – С. 330.*

***Величини кута φ для різних відсоткових долей*** *(продовження)*

| % | %, останній десятковий знак | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| | Значення φ | | | | | | | | | |
| 23 | 1,000 | 1,003 | 1,005 | 1,007 | 1,010 | 1,012 | 1,015 | 1,017 | 1,019 | 1,022 |
| 24 | 1,024 | 1,026 | 1,029 | 1,031 | 1,033 | 1,036 | 1,038 | 1,040 | 1,043 | 1,045 |
| 25 | 1,047 | 1,050 | 1,052 | 1,054 | 1,056 | 1,059 | 1,061 | 1,063 | 1,066 | 1,068 |
| 26 | 1,070 | 1,072 | 1,075 | 1,077 | 1,079 | 1,082 | 1,084 | 1,086 | 1,088 | 1,091 |
| 27 | 1,093 | 1,095 | 1,097 | 1,100 | 1,102 | 1,104 | 1,106 | 1,109 | 1,111 | 1,113 |
| 28 | 1,115 | 1,117 | 1,120 | 1,122 | 1,124 | 1,126 | 1,129 | 1,131 | 1,133 | 1,135 |
| 29 | 1,137 | 1,140 | 1,142 | 1,144 | 1,146 | 1,148 | 1,151 | 1,153 | 1,155 | 1,157 |
| 30 | 1,159 | 1,161 | 1,164 | 1,166 | 1,168 | 1,170 | 1,172 | 1,174 | 1,177 | 1,179 |
| 31 | 1,182 | 1,183 | 1,185 | 1,187 | 1,190 | 1,192 | 1,194 | 1,196 | 1,198 | 1,200 |
| 32 | 1,203 | 1,205 | 1,207 | 1,209 | 1,211 | 1,213 | 1,215 | 1,217 | 1,220 | 1,222 |
| 33 | 1,224 | 1,226 | 1,228 | 1,230 | 1,232 | 1,234 | 1,237 | 1,239 | 1,241 | 1,243 |
| 34 | 1,245 | 1,247 | 1,249 | 1,251 | 1,254 | 1,256 | 1,258 | 1,260 | 1,262 | 1,264 |
| 35 | 1,266 | 1,268 | 1,270 | 1,272 | 1,274 | 1,277 | 1,279 | 1,281 | 1,283 | 1,285 |
| 36 | 1,287 | 1,289 | 1,291 | 1,293 | 1,295 | 1,297 | 1,299 | 1,302 | 1,304 | 1,306 |
| 37 | 1,308 | 1,310 | 1,312 | 1,314 | 1,316 | 1,318 | 1,320 | 1,322 | 1,324 | 1,326 |
| 38 | 1,328 | 1,330 | 1,333 | 1,335 | 1,337 | 1,339 | 1,341 | 1,343 | 1,345 | 1,347 |
| 39 | 1,349 | 1,351 | 1,353 | 1,353 | 1,357 | 1,359 | 1,361 | 1,363 | 1,365 | 1,367 |
| 40 | 1,369 | 1,371 | 1,374 | 1,376 | 1,378 | 1,380 | 1,382 | 1,384 | 1,386 | 1,388 |
| 41 | 1,390 | 1,392 | 1,394 | 1,396 | 1,398 | 1,400 | 1402 | 1,404 | 1,406 | 1,408 |
| 42 | 1,410 | 1,412 | 1,414 | 1,416 | 1,418 | 1,420 | 1,422 | 1,424 | 1,426 | 1,428 |
| 43 | 1,430 | 1,432 | 1,434 | 1,436 | 1,438 | 1,440 | 1,442 | 1,444 | 1,446 | 1,448 |
| 44 | 1,451 | 1,453 | 1,455 | 1,457 | 1,459 | 1,461 | 1,463 | 1,465 | 1,467 | 1,469 |
| 45 | 1,471 | 1,473 | 1,475 | 1,477 | 1,479 | 1,481 | 1,483 | 1,485 | 1,487 | 1,489 |
| 46 | 1,491 | 1,493 | 1,495 | 1,497 | 1,499 | 1,501 | 1,503 | 1,505 | 1,507 | 1,509 |
| 47 | 1,511 | 1,513 | 1,515 | 1,517 | 1,519 | 1,521 | 1,523 | 1,525 | 1,527 | 1,529 |
| 48 | 1,531 | 1,533 | 1,535 | 1,537 | 1,539 | 1,541 | 1,543 | 1,545 | 1,547 | 1,549 |
| 49 | 1,551 | 1,553 | 1,555 | 1,557 | 1,559 | 1,561 | 1,563 | 1,565 | 1,567 | 1,569 |
| 50 | 1,571 | 1,573 | 1,575 | 1,577 | 1,579 | 1,581 | 1,583 | 1,585 | 1,587 | 1,589 |
| 51 | 1,591 | 1,593 | 1,595 | 1,597 | 1,599 | 1,601 | 1,603 | 1,605 | 1,607 | 1,609 |
| 52 | 1,611 | 1,613 | 1,615 | 1,617 | 1,619 | 1,621 | 1,623 | 1,625 | 1,627 | 1,629 |
| 53 | 1,631 | 1,633 | 1,635 | 1,637 | 1,639 | 1,641 | 1,643 | 1,645 | 1,647 | 1,649 |
| 54 | 1,651 | 1,653 | 1,655 | 1,657 | 1,659 | 1,661 | 1,663 | 1,665 | 1,667 | 1,669 |
| 55 | 1,671 | 1,673 | 1,675 | 1,677 | 1,679 | 1,681 | 1,683 | 1,685 | 1,687 | 1,689 |
| 56 | 1,691 | 1,693 | 1,695 | 1,697 | 1,699 | 1,701 | 1,703 | 1,705 | 1,707 | 1,709 |
| 57 | 1,711 | 1,713 | 1,715 | 1,717 | 1,719 | 1,721 | 1,723 | 1,725 | 1,727 | 1,729 |
| 58 | 1,731 | 1,734 | 1,736 | 1,738 | 1,740 | 1,742 | 1,744 | 1,746 | 1,748 | 1,750 |
| 59 | 1,752 | 1,754 | 1,756 | 1,758 | 1,760 | 1,762 | 1,764 | 1,766 | 1,768 | 1,770 |
| 60 | 1,772 | 1,774 | 1,776 | 1,778 | 1,780 | 1,782 | 1,784 | 1,786 | 1,789 | 1,791 |
| 61 | 1,793 | 1,795 | 1,797 | 1,799 | 1,801 | 1,803 | 1,805 | 1,807 | 1,809 | 1,811 |
| 62 | 1,813 | 1,815 | 1,817 | 1,819 | 1,821 | 1,823 | 1,826 | 1,828 | 1,830 | 1,832 |

***Величини кута φ для різних відсоткових долей*** *(продовження)*

| % | %, останній десятковий знак | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| | Значення φ | | | | | | | | | |
| 63 | 1,834 | 1,836 | 1,838 | 1,840 | 1,842 | 1,844 | 1,846 | 1,848 | 1,850 | 1,853 |
| 64 | 1,855 | 1,857 | 1,859 | 1,861 | 1,863 | 1,865 | 1,867 | 1,869 | 1,871 | 1,873 |
| 65 | 1,875 | 1,878 | 1,880 | 1,882 | 1,884 | 1,886 | 1,888 | 1,890 | 1,892 | 1,894 |
| 66 | 1,897 | 1,899 | 1,901 | 1,903 | 1,905 | 1,907 | 1,909 | 1,911 | 1,913 | 1,916 |
| 67 | 1,918 | 1,920 | 1,922 | 1,924 | 1,926 | 1,928 | 1,930 | 1,933 | 1,935 | 1,937 |
| 68 | 1,939 | 1,941 | 1,943 | 1,946 | 1,948 | 1,950 | 1,952 | 1,954 | 1,956 | 1,958 |
| 69 | 1,961 | 1,963 | 1,965 | 1,967 | 1,969 | 1,971 | 1,974 | 1,976 | 1,978 | 1,980 |
| 70 | 1,982 | 1,984 | 1,987 | 1,989 | 1,991 | 1,993 | 1,995 | 1,998 | 2,000 | 2,002 |
| 71 | 2,004 | 2,006 | 2,009 | 2,011 | 2,013 | 2,015 | 2,018 | 2,020 | 2,022 | 2,024 |
| 72 | 2,026 | 2,029 | 2,031 | 2,033 | 2,035 | 2,038 | 2,040 | 2,042 | 2,044 | 2,047 |
| 73 | 2,049 | 2,051 | 2,053 | 2,056 | 2,058 | 2,060 | 2,062 | 2,065 | 2,067 | 2,069 |
| 74 | 2,071 | 2,074 | 2,076 | 2,078 | 2,081 | 2,083 | 2,085 | 2,087 | 2,090 | 2,092 |
| 75 | 2,094 | 2,097 | 2,099 | 2,101 | 2,104 | 2,106 | 2,108 | 2,111 | 2,113 | 2,115 |
| 76 | 2,118 | 2,120 | 2,122 | 2,125 | 2,127 | 2,129 | 2,132 | 2,134 | 2,136 | 2,139 |
| 77 | 2,141 | 2,144 | 2,146 | 2,148 | 2,151 | 2,153 | 2,156 | 2,158 | 2,160 | 2,163 |
| 78 | 2,165 | 2,168 | 2,170 | 2,172 | 2,175 | 2,177 | 2,180 | 2,182 | 2,185 | 2,187 |
| 79 | 2,190 | 2,192 | 2,194 | 2,197 | 2,199 | 2,202 | 2,204 | 2,207 | 2,209 | 2,212 |
| 80 | 2,214 | 2,217 | 2,219 | 2,222 | 2,224 | 2,227 | 2,229 | 2,231 | 2,234 | 2,237 |
| 81 | 2,240 | 2,242 | 2,245 | 2,247 | 2,250 | 2,252 | 2,255 | 2,258 | 2,260 | 2,263 |
| 82 | 2,265 | 2,268 | 2,271 | 2,273 | 2,276 | 2,278 | 2,281 | 2,284 | 2,286 | 2,289 |
| 83 | 2,292 | 2,294 | 2,297 | 2,300 | 2,302 | 2,305 | 2,308 | 2,310 | 2,313 | 2,316 |
| 84 | 2,319 | 2,321 | 2,324 | 2,327 | 2,330 | 2,332 | 2,335 | 2,338 | 2,341 | 2,343 |
| 85 | 2,346 | 2,349 | 2,352 | 2,355 | 2,357 | 2,360 | 2,363 | 2,366 | 2,369 | 2,372 |
| 86 | 2375 | 2,377 | 2,380 | 2,383 | 2,386 | 2,389 | 2,392 | 2,395 | 2,398 | 2,401 |
| 87 | 2,404 | 2,407 | 2,410 | 2,413 | 2,416 | 2,419 | 2,422 | 2,425 | 2,428 | 2,431 |
| 88 | 2,434 | 2,437 | 2,440 | 2,443 | 2,447 | 2,450 | 2,453 | 2,456 | 2,459 | 2,462 |
| 89 | 2,465 | 2,469 | 2,472 | 2,475 | 2,478 | 2,482 | 2,485 | 2,488 | 2,491 | 2,495 |
| 90 | 2,498 | 2,501 | 2,505 | 2,508 | 2,512 | 2,515 | 2,518 | 2,522 | 2,525 | 2,529 |
| 91 | 2,532 | 2,536 | 2,539 | 2,543 | 2,546 | 2,550 | 2,554 | 2,557 | 2,561 | 2,564 |
| 92 | 2,568 | 2,572 | 2,575 | 2,579 | 2,583 | 2,S87 | 2,591 | 2,594 | 2,598 | 2,602 |
| 93 | 2,606 | 2,610 | 2,614 | 2,618 | 2,622 | 2626 | 2,630 | 2,634 | 2,638 | 2,642 |
| 94 | 2,647 | 2651 | 2,655 | 2,659 | 2,664 | 2,668 | 2,673 | 2,677 | 2,681 | 2,686 |
| 95 | 2,691 | 2,295 | 2,700 | 2,705 | 2,709 | 2,714 | 2,719 | 2,724 | 2,729 | 2,734 |
| 96 | 2,739 | 2,744 | 2,749 | 2,754 | 2,760 | 2,765 | 2,771 | 2,776 | 2,782 | 2,788 |
| 97 | 2,793 | 2,799 | 2,805 | 2,811 | 2,818 | 2,824 | 2,830 | 2,837 | 2,844 | 2,851 |
| 98 | 2,858 | 2,865 | 2,872 | 2,880 | 2,888 | 2,896 | 2,904 | 2,913 | 2,922 | 2,931 |
| 99,0 | 2,941 | 2,942 | 2,943 | 2,944 | 2,945 | 2,946 | 2,948 | 2,949 | 2,950 | 2,951 |
| 99,1 | 2,952 | 2,953 | 2,954 | 2,955 | 2,956 | 2,957 | 2,958 | 2,959 | 2,960 | 2,961 |
| 99,2 | 2,963 | 2,964 | 2,965 | 2,966 | 2,967 | 2,968 | 2,969 | 2,971 | 2,972 | 2,973 |
| 99,3 | 2,974 | 2,975 | 2,976 | 2,978 | 2,979 | 2,980 | 2,981 | 2,983 | 2,984 | 2,985 |

***Величини кута φ для різних відсоткових долей*** *(продовження)*

| % | %, останній десятковий знак | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| | Значення φ | | | | | | | | | |
| 99,4 | 2,987 | 2,988 | 2,989 | 2,990 | 2,992 | 2,993 | 2,995 | 2,996 | 2,997 | 2,999 |
| 99,5 | 3,000 | J,002 | 3,003 | 3,004 | 3,006 | 3,007 | 3,009 | 3,010 | 3,012 | 3,013 |
| 99,6 | J,015 | 3,017 | 3,018 | 3,020 | 3,022 | J,023 | 3,025 | 3,027 | 3,028 | 3,030 |
| 99,7 | 3,032 | 3,034 | 3,036 | 3,038 | 3,040 | 3,041 | 3,044 | 3,046 | 3,048 | 3,050 |
| 99,8 | 3,052 | 3,054 | 3,057 | 3,059 | 3,062 | 3,064 | 3,067 | 3,069 | 3,072 | 3,075 |
| 99,9 | 3,078 | 3,082 | 3,085 | 3,089 | 3,093 | 3,097 | 3,101 | 3,107 | 3,113 | 3,122 |
| 100 | 3,142 | | | | | | | | | |

**Таблиця 2**

***Критичні значення критерію F Фішера***[25]

*(верхнє число в комірці – p=0,05, нижнє число – p=0,01)*

| | | Ступені свободи в чисельнику ($df_1$) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 12 | 24 | 4 |
| Ступені свободи в знаменнику ($df_2$) | 1 | 161,45 | 199,50 | 215,72 | 224,57 | 230,17 | 233,97 | 238,89 | 243,91 | 249,04 | 254,32 |
| | | 4032,10 | 4999,03 | 5403,49 | 5623,14 | 5764,08 | 5859,39 | 5981,34 | 6105,83 | 6234,16 | 6366,48 |
| | 2 | 18,51 | 19,00 | 19,16 | 19,25 | 19,30 | 19,33 | 19,37 | 19,41 | 19,45 | 19,50 |
| | | 98,49 | 99,01 | 99,17 | 99,25 | 99,30 | 99,33 | 99,36 | 99,42 | 99,46 | 99,50 |
| | 3 | 10,13 | 9,55 | 9,28 | 9,12 | 9,01 | 8,94 | 8,84 | 8,74 | 8,64 | 8,53 |
| | | 34,12 | 30,81 | 29,45 | 28,71 | 28,24 | 27,91 | 27,49 | 27,05 | 26,60 | 26,12 |
| | 4 | 7,71 | 6,94 | 6,59 | 6,39 | 6,26 | 6,16 | 6,04 | 5,91 | 5,77 | 5,63 |
| | | 21,20 | 18,00 | 16,69 | 15,98 | 15,52 | 15,21 | 14,80 | 14,37 | 13,93 | 13,46 |
| | 5 | 6,61 | 5,79 | 5,41 | 5,19 | 5,05 | 4,95 | 4,82 | 4,68 | 4,53 | 4,36 |
| | | 16,26 | 13,27 | 12,06 | 11,39 | 10,97 | 10,67 | 10,27 | 9,89 | 9,47 | 9,02 |
| | 6 | 5,99 | 5,14 | 4,76 | 4,53 | 4,39 | 4,28 | 4,15 | 4,00 | 3,84 | 3,67 |
| | | 13,74 | 10,92 | 9,78 | 9,15 | 8,75 | 8,47 | 8,10 | 7,72 | 7,31 | 6,88 |
| | 7 | 5,59 | 4,74 | 4,35 | 4,12 | 3,97 | 3,87 | 3,73 | 3,57 | 3,41 | 3,23 |
| | | 12,25 | 9,55 | 8,45 | 7,85 | 7,46 | 7,19 | 6,84 | 6,47 | 6,07 | 5,65 |
| | 8 | 5,32 | 4,46 | 4,07 | 3,84 | 3,69 | 3,58 | 3,44 | 3,28 | 3,12 | 2,93 |
| | | 11,26 | 8,65 | 7,59 | 7,01 | 6,63 | 6,37 | 6,03 | 5,67 | 5,28 | 4,86 |
| | 9 | 5,12 | 4,26 | 3,86 | 3,83 | 3,48 | 3,37 | 3,23 | 3,07 | 2,90 | 2,71 |
| | | 10,56 | 8,02 | 6,99 | 6,42 | 6,06 | 5,80 | 5,47 | 5,11 | 4,73 | 4,31 |
| | 10 | 4,6 | 4,10 | 3,71 | 3,48 | 3,33 | 3,22 | 3,07 | 2,91 | 2,74 | 2,54 |
| | | 10,04 | 7,56 | 6,55 | 5,99 | 5,64 | 5,39 | 5,06 | 4,71 | 4,33 | 3,91 |
| | 11 | 4,84 | 3,98 | 3,59 | 3,36 | 3,20 | 3,09 | 2,95 | 2,79 | 2,61 | 2,40 |
| | | 9,65 | 7,20 | 6,22 | 5,67 | 5,32 | 5,07 | 4,74 | 4,40 | 4,02 | 3,60 |
| | 12 | 4,75 | 3,88 | , 3,49 | 3,26 | 3,11 | 3,00 | 2,85 | 2,69 | 2,50 | 2,30 |
| | | 9,33 | 6,93 | 5,93 | 5,41 | 5,06 | 4,82 | 4,50 | 4,16 | 3,78 | 3,36 |
| | 14 | 4,60 | 3,74 | 3,34 | 3,11 | 2,96 | 2,85 | 2,70 | 2,53 | 2,35 | 2,13 |
| | | 8,86 | 6,51 | 5,56 | 5,03 | 4,69 | 4,46 | 4,14 | 3,80 | 3,43 | 3,00 |
| | 16 | 4,49 | 3,63 | 3,24 | 3,01 | 2,85 | 2,74 | 2,59 | 2,42 | 2,24 | 2,01 |
| | | 8,53 | 6,23 | 5,29 | 4,77 | 4,44 | 4,20 | 3,89 | 3,55 | 3,18 | 2,75 |
| | 18 | 4,41 | 3,55 | 3,16 | 2,93 | 2,77 | 2,66 | 2,51 | 2,34 | 2,15 | 1,92 |
| | | 8,28 | 6,01 | 5,09 | 4,58 | 4,25 | 4,01 | 3,71 | 3,37 | 3,01 | 2,57 |
| | 20 | 4,35 | 3,49 | 3,10 | 2,87 | 2,71 | 2,60 | 2,45 | 2,28 | 2,08 | 1,84 |
| | | 8,10 | 5,85 | 4,94 | 4,43 | 4,10 | 3,87 | 3,56 | 3,23 | 2,86 | 2,42 |

[25] *Солсо Р.Л., Джонсон Х.Х., Бил М.К. Экспериментальная психология: практический курс. – СПб.: прайм-ЕВРОЗНАК, 2001. – С. 514-515.*

# ДОДАТОК **3**

МЕТОДИКА “ЕМОЦІЙНИЙ ГРАДУСНИК”[26]
(модифікована відповідно до мети експерименту)

| ДЕНЬ | | ЕМОЦІЙНИЙ СТАН | | | | |
|---|---|---|---|---|---|---|
| | | **Позитивний** | **Більш позитивний, ніж негативний** | **Невизначений** | **Більш негативний, ніж позитивний** | **Негативний** |
| *Нарахований бал* | | *5* | *4* | *3* | *2* | *1* |
| Дні тижня | Понеділок | | | | | |
| | Вівторок | | | | | |
| | Середа | | | | | |
| | Четвер | | | | | |
| | П’ятниця | | | | | |
| | Субота | | | | | |
| | Неділя | | | | | |

**Інструкція:**

Вам пропонується взяти участь у психологічному дослідженні. Його метою є вивчення змін у емоційному стані студентів протягом тижня. В бланку, який Ви отримали, є рядки – дні тижня, і стовпчики – рівні емоційного стану. Протягом тижня Вам необхідно ввечері, перед сном, фіксувати рівень свого емоційного стану – ставити крапку напроти відповідних днів тижня та рівнів емоційного стану.

Після закінчення дослідження поверніть бланк досліднику, і через деякий час Ви зможете поцікавитися результатами.

[26] *Фетискин Н.П. Эмоциональное обеспечение учебной и трудовой деятельности. – Кострома. 1990.*

# ДОДАТОК **4**

МЕТОДИКА “ЕМОЦІЙНИЙ ГРАДУСНИК” ТА БЛАНК ФІКСАЦІЇ УСПІХІВ

| ДЕНЬ | | ЕМОЦІЙНИЙ СТАН ВРАНЦІ | | | | | КІЛЬКІСТЬ ВДАЛИХ СПРАВ | ЕМОЦІЙНИЙ СТАН ВРАНЦІ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ++ | + | I+ | I | II | | ++ | + | I+ | I | II |
| *Нарахований бал* | | *5* | *4* | *3* | *2* | *1* | | | | | | |
| Дні тижня | Понеділок | | | | | | | | | | | |
| | Вівторок | | | | | | | | | | | |
| | Середа | | | | | | | | | | | |
| | Четвер | | | | | | | | | | | |
| | П’ятниця | | | | | | | | | | | |
| | Субота | | | | | | | | | | | |
| | Неділя | | | | | | | | | | | |

## Інструкція:

Вам пропонується взяти участь у психологічному дослідженні. Його мета – перевірити, чи є зв’язок між емоційним станом вранці та ввечері і кількістю успішних, вдалих справ протягом дня. В бланку, який Ви отримали, є рядки – дні тижня, і стовпчики – рівні емоційного стану (вранці та ввечері) та кількість вдалих справ. Протягом тижня Вам необхідно вранці, до початку занять, зафіксувати свій емоційний стан, а ввечері, перед сном, зафіксувати кількість вдалих справ і знову – рівень свого емоційного стану.

По завершенню дослідження поверніть бланк досліднику, і через деякий час Ви зможете поцікавитися результатами.

# ДОДАТОК **5**

## КОМУНІКАТИВНІ ТА ОРГАНІЗАТОРСЬКІ СХИЛЬНОСТІ (КОС-2)[27]

Інструкція: *Перед вами 40 запитань, на кожне з яких потрібно дати відповідь «так» або «ні». Час виконання методики 10-15 хв.*

1. Чи є у Вас потяг до спілкування і знайомства з різними людьми?
2. Чи подобається Вам займатися громадською роботою?
3. Чи довго турбує Вас почуття образи на своїх товаришів?
4. Чи завжди Вам важко орієнтуватися в критичній ситуації, що склалася?
5. Чи багато у Вас друзів, з якими Ви постійно спілкуєтеся?
6. Чи часто Вам вдається схилити більшість своїх товаришів до прийняття ними Вашої думки?
7. Чи правда, що Вам приємніше і простіше проводити час за книгами або за яким-небудь іншим заняттям, ніж з людьми?
8. Якщо виникли певні перешкоди у здійсненні Ваших намірів, чи легко Вам відмовитися від своїх намірів?
9. Чи легко Ви встановлюєте контакти з людьми, які значно старші за Вас?
10. Чи любите Ви придумувати і організовувати зі своїми товаришами різноманітні ігри і розваги?
11. Чи важко Вам включатися в нові для Вас компанії (колективи)?
12. Чи часто Ви відкладаєте на інші дні справи, які слід було б виконати сьогодні?
13. Чи легко Вам вдається встановлювати контакти і спілкуватися з незнайомими людьми?
14. Чи прагнете Ви, щоб Ваші товариші діяли у відповідності з Вашою думкою?
15. Чи важко Вам освоюватися в новому колективі?
16. Правда, що у Вас не буває конфліктів з товаришами через невиконання ними своїх обіцянок, зобов'язань, обов'язків?
17. Чи прагнете Ви, якщо обставини сприяють цьому, познайомитися і поспілкуватися з незнайомою людиною?
18. Чи часто у вирішенні важливих справ берете ініціативу на себе?
19. Чи дратують Вас оточуючі, чи виникає у Вас бажання побути наодинці?
20. Чи правда, що Ви, як правило, погано орієнтуєтеся в незнайомій для Вас обстановці?
21. Чи подобається Вам постійно бути серед людей?
22. Чи виникає у Вас відчуття дискомфорту, якщо Вам не вдається завершити розпочату справу?
23. Чи важко Вам проявити ініціативу, щоб познайомитися з іншою людиною?
24. Чи правда, що Ви втомлюєтеся від постійного спілкування з друзями?
25. Чи подобається Вам брати участь в колективних іграх?
26. Чи часто Ви проявляєте ініціативу під час вирішення проблем, які зачіпають інтереси Ваших друзів?
27. Чи правда, що Ви відчуваєте себе невпевнено серед незнайомих людей?
28. Чи правда, що Ви не прагнете до того, щоб довести правильність .своїх дій чи слів?

---

[27] *Уніят С., Комінко С. Вибір професії, або задача з багатьма невідомими. – Тернопіль: Підручники і посібники, 1997. – с.37-40.*

29. Чи вважаєте Ви, що Вам .не важко внести пожвавлення у незнайоме товариство?
30. Чи брали Ви участь у громадському житті школи (підприємства)?
31. Чи прагнете Ви обмежити коло своїх знайомих?
32. Чи правда, що Ви не прагнете відстоювати свою думку чи рішення, якщо його не зразу прийняли друзі?
33. Чи відчуваєте Ви себе комфортно в незнайомій компанії?
34. Чи охоче Ви організуєте вечірку для своїх товаришів?
35. Чи правда, що Ви губитеся, коли треба говорити перед великою кількістю людей?
36. Чи часто Ви спізнюєтеся на ділові зустрічі, побачення?
37. У Вас багато друзів?
38. Чи часто Ви буваєте в центрі уваги своїх друзів?
39. Ви відчуваєте дискомфорт під час спілкування з малознайомими людьми?
40. Чи правда, що Ви не дуже впевнено почуваєте себе в оточенні великої групи своїх друзів?

***Обробка результатів:***

*Комунікативні схильності (КС):*

Так — 1, 5, 9, 13, 17, 21, 25, 29, 33, 37;
Ні — 3,7, 11, 15, 19, 23, 27, 31, 35, 39.

*Організаторські схильності (ОС):*

Так — 2, 6, 10, 14, 18, 22, 26, 30, 34, 38;
Ні — 4, 8, 12, 16, 20, 24, 28, 32, 36, 40.

$$K_{KC} = \frac{KC}{20}$$

$$K_{OC} = \frac{OC}{20}$$

*Шкала оцінок комунікативних схильностей*

| **Ккс** | 0,10-0,45 | 0,46-0,55 | 0,56-0,65 | 0,66-0,75 | 0,76-1,00 |
|---|---|---|---|---|---|
| Рівень | Низький | Нижче середнього | Середній | Високий | Дуже високий |

*Шкала оцінок організаторських схильностей*

| **Кос** | 0,20-0,55 | 0,56-0,65 | 0,66-0,70 | 0,71-0,80 | 0,81-1 |
|---|---|---|---|---|---|
| Рівень | Низький | Нижче середнього | Середній | Високий | Дуже високий |

# ДОДАТОК **6**

СТРУКТУРА ЕКСПЕРИМЕНТУ

Для проведення експерименту Вам буде потрібен помічник (можна декілька), який буде створювати в приміщенні шумовий фон. Сам експеримент проходить в 2 етапи: 1 етап: учасник дослідження виконує теппінг-тест у тихому приміщенні. По завершенню роботи – 20-хвилинна перерва. 2 етап: учасник дослідження виконує теппінг-тест повторно, але одночасно з виконанням ним тесту помічник експериментатора створює підвищений рівень шуму (будь-якими способами, які Вам підкаже фантазія).

ТЕППІНГ-ТЕСТ[28]

Інструкція:

*Сядьте зручно на стілець, так, щоб Ваша рука була на рівні стола. Візьміть олівець в руку. Постарайтесь поставити якомога більше крапок у кожному квадраті за 5 секунд (всього 30 секунд на виконання). Перехід з одного квадрата в інший потрібно здійснювати швидким і коротким рухом за командою "Раз!", яку дає експериментатор (викладач). Працювати потрібно весь час, на максимальному вольовому зусиллі.*

Бланк:

| 1 | 2 | 3 |
|---|---|---|
| | | |
| | | |
| 6 | 5 | 4 |

Обробка результатів:

Коли робота завершена, підраховують суму крапок у всіх шести квадратах. По цій сумі роблять висновок про лабільність нервових процесів.

| Кількість крапок | < 151 | 152 – 183 | 184 – 210 | > 211 |
|---|---|---|---|---|
| Лабільність | Низька | Середня | Висока | Дуже висока |

[28] *Уніят С., Комінко С. Вибір професії, або задача з багатьма невідомими. – Тернопіль: Підручники і посібники, 1997. – с.37-40.*

# ЗМІСТ

**ПЕРЕДМОВА**..... 3

**МЕТОДИЧНІ ПРОБЛЕМИ ВИКЛАДАННЯ “МАТЕМАТИЧНИХ МЕТОДІВ У ПСИХОЛОГІЇ”**..... 5

**МОДУЛЬ 1. КЛАСИЧНІ МАТЕМАТИЧНІ МЕТОДИ**..... 14

**Розділ І. ОСНОВНІ ПОНЯТТЯ МАТЕМАТИЧНОЇ СТАТИСТИКИ**

§ 1.1. Основні поняття математичної статистики..... 14
§ 1.2. Статистичне спостереження. Поняття про репрезентативність. Графічне представлення даних ..... 22
§ 1.3. Міри центральної тенденції..... 30
§ 1.4. Міри мінливості..... 37

**Розділ ІІ. АНАЛІЗ ЗВ’ЯЗКУ МІЖ ЗМІННИМИ**

§ 2.1. Міри зв’язку в параметричній статистиці..... 41
§ 2.2. Міри зв’язку в непараметричній статистиці..... 49
§ 2.3. Регресійний аналіз ..... 64

**Розділ ІІІ. МЕТОДИ СТАТИСТИЧНОГО ВИСНОВКУ**

§ 3.1. Статистичні гіпотези..... 74
§ 3.2. Виявлення відмінностей у рівні прояву ознаки..... 83
§ 3.3. Оцінка достовірності зсуву значень..... 95
§ 3.4. Виявлення відмінностей у розподілі ознак..... 107

***ЗАВДАННЯ ДЛЯ САМОКОНТРОЛЮ*** ..... 124

**МОДУЛЬ 2. БАГАТОМІРНИЙ МАТЕМАТИКО-СТАТИСТИЧНИЙ АНАЛІЗ**..... 129

**Розділ IV. ДИСПЕРСІЙНИЙ АНАЛІЗ**

§ 4.1. Факторні експерименти..... 129
§ 4.2. Двофакторний дисперсійний аналіз..... 135

**Розділ V. МЕТОДИ БАГАТОМІРНОГО СТАТИСТИЧНОГО АНАЛІЗУ**

§ 5.1. Факторний аналіз..... 144
§ 5.2. Багатомірне шкалювання..... 158
§ 5.3. Кластерний аналіз..... 170

***ЗАВДАННЯ ДЛЯ САМОКОНТРОЛЮ*** ..... 187

**ЛАБОРАТОРНИЙ ПРАКТИКУМ (на базі програми STATISTICA 5.5)**… 192

***ПЕРЕДМОВА*** ..... 192

***МОДУЛЬ 1: КЛАСИЧНІ МАТЕМАТИЧНІ МЕТОДИ***

***Лабораторне заняття № 1****: Описові статистики*..... 193
*Опис дослідження*..... 194
*Створення та збереження файлу даних*..... 195
*Побудова графіків та діаграм*..... 198
*Обчислення описових статистик*..... 202
*Матеріали для електронного звіту*..... 206

***Лабораторне заняття № 2****: Аналіз зв'язку між змінними*..... 208
*Опис дослідження*..... 209
*Побудова діаграми розсювання*…… 211
*Обчислення коефіцієнта лінійної кореляції Пірсона*..... 213
*Обчислення коефіцієнта рангової кореляції Спірмена*..... 214
*Побудова лінії регресії*..... 206
*Матеріали для електронного звіту*..... 218

***Лабораторне заняття № 3****: Методи статистичного висновку. Виявлення відмінностей у рівні прояву ознаки*…….....…….. 220
*Опис дослідження*..... 221
*Обчислення показника t (критерій Стьюдента для незалежних змінних)*..... 222
*Обчислення показника U (критерій Манна-Уітні)*..... 225
*Матеріали для електронного звіту*..... 227

***Лабораторне заняття № 4:*** *Методи статистичного висновку. Оцінка достовірності зсуву значень*..... 228
*Опис дослідження*..... 228
*Обчислення показника t (критерій Стьюдента для залежних змінних)*..... 229
*Обчислення показника G (критерій знаків)*..... 232
*Матеріали для електронного звіту*..... 234

***МОДУЛЬ 2: БАГАТОМІРНИЙ МАТЕМАТИКО-СТАТИСТИЧНИЙ АНАЛІЗ***

***Лабораторне заняття № 5:*** *Факторний аналіз*..... 235
*Опис дослідження*..... 236
*Проведення факторного аналізу*..... 236
*Матеріали для електронного звіту*..... 241

***Лабораторне заняття № 6:*** *Багатомірне шкалювання*..... 242
*Опис дослідження*..... 243
*Проведення багатомірного шкалювання*..... 243
*Матеріали для електронного звіту*..... 252

***Лабораторне заняття № 7:*** *Кластерний аналіз*..... 253
*Опис дослідження*..... 254
*Проведення кластерного аналізу*..... 254
*Матеріали для електронного звіту*..... 260

**АЛФАВІТНИЙ ПОКАЖЧИК СТАТИСТИЧНИХ КРИТЕРІЇВ**..... 261

**РЕКОМЕНДОВАНА ЛІТЕРАТУРА**..... 262

**ДОДАТКИ**
Додаток 1. Загальний алгоритм вибору методу обробки даних..... 267
Додаток 2. Статистичні таблиці: критерій кутового перетворення Фішера ($\varphi^*$) та критерій Фішера (F)..... 268
Додаток 3. Методика “Емоційний градусник”..... 273
Додаток 4. Методика “Емоційний градусник” та бланк фіксації успіхів..... 274
Додаток 5. Методика “КОС-2”..... 275
Додаток 6. Методика “Теппінг-тест” та структура експерименту..... 277



View publication stats