

**Національний університет цивільного захисту України**  
**Кафедра психології діяльності в особливих умовах**

# **МАТЕМАТИЧНІ МЕТОДИ В ПСИХОЛОГІЇ**

*Курс лекцій*

Розробник:  
старший викладач кафедри психології  
діяльності в особливих умовах,  
кандидат психологічних наук  
Боснюк В.Ф.

**Харків 2016**

## ЗМІСТ

<b>ЛЕКЦІЯ 1. Математика і психологія</b>	4
1.1. Передумови запровадження математичних методів в психології.	4
1.2. Проблеми, з якими стикаються студенти вивчаючи «Математичні методи в психології».	7
1.3. Викладання «Математичних методів в психології» в умовах кредитно-модульної системи.	10
Завдання на самостійну підготовку.	11
<b>ЛЕКЦІЯ 2. Вимірювання в психології. Вимірювання і шкали</b>	12
2.1. Проблема вимірювання у психології.	12
2.2 Типи та характеристики вимірювальних шкал.	14
Завдання на самостійну підготовку.	20
<b>ЛЕКЦІЯ 3. Первинна описова статистика</b>	21
3.1. Міри центральної тенденції.	21
3.2. Міри мінливості даних.	23
3.3. Міри положення.	23
Завдання на самостійну підготовку.	24
<b>ЛЕКЦІЯ 4. Вступ до проблеми статистичного висновку</b>	24
4.1. Ідея перевірки статистичної гіпотези.	25
4.2. Статистичний критерій та рівень статистичної значущості.	26
Завдання на самостійну підготовку.	30
<b>ЛЕКЦІЯ 5. Вибір методу статистичного висновку</b>	30
5.1. Класифікація методів статистичного висновку.	30
5.2. Генеральна сукупність та вибірка дослідження	35
Завдання на самостійну підготовку.	37
<b>ЛЕКЦІЯ 6. Методи кореляційного аналізу</b>	37
6.1. Сутність методів встановлення статистичних взаємозв'язків. Основні властивості коефіцієнта кореляції.	38
6.2. Кореляція метричних змінних.	39
6.3. Кореляція рангових змінних.	41
6.4. Кореляція дихотомічних змінних.	42
Завдання на самостійну підготовку.	43
<b>ЛЕКЦІЯ 7. Параметричні методи порівняння двох вибірок досліджуваних</b>	44
7.1. Теоретичні засади та сфера застосування t-критерію Стьюдента.	44
7.2. Статистичний критерій t-Стьюдента для однієї вибірки.	47
7.3. Статистичний критерій t-Стьюдента для незалежних вибірок.	48
7.4. Статистичний критерій t-Стьюдента для залежних вибірок.	49
Завдання на самостійну підготовку.	50
<b>ЛЕКЦІЯ 8. Непараметричні методи порівняння вибірок досліджуваних</b>	50
8.1. Статистичний критерій U-Манна-Уїтні.	51
8.2. Статистичний критерій G-знаків.	53

8.3. Статистичний критерій $\phi$ – кутове перетворення Фішера.....	54
Завдання на самостійну підготовку.....	56

## ЛЕКЦІЯ 1 МАТЕМАТИКА І ПСИХОЛОГІЯ

### План лекції

- 1.1. Передумови запровадження математичних методів в психології
- 1.2. Проблеми, з якими стикаються студенти вивчаючи «Математичні методи в психології»
- 1.3. Викладання «Математичних методів в психології» в умовах кредитно-модульної системи

Час проведення: 2 учбові години.

### Література

1. Климчук В.О. Викладання курсу “Математичні методи у психології” в умовах кредитно-модульної системи // Соціальна психологія. – 2008. – №2 (28). – С. 180-189.
2. Климчук В.О. Математичні методи у психології. Навчальний посібник. – К.: Освіта України, 2009. – 288 с.
3. Кричевець А.Н. Математика для психологов: Учебник / А.Н. Кричевець, Е.В. Шикин, А.Г. Дьячков / Под ред. А.Н. Кричевца. М., 2003. – 198 с.
4. Телейко, А. Б. Математико-статистичні методи в соціології та психології: навч. посібник / А. Б. Телейко, Р. К. Чорней. – Київ : МАУП, 2007. – 418 с

### **1.1. Передумови запровадження математичних методів в психології**

Існує думка, неодноразово висловлена визначними вченими: область знань стає наукою, лише застосовуючи математику. З цією думкою, можливо, не погодяться багато гуманітаріїв. А дарма: саме математика дозволяє кількісно порівнювати явища, перевіряти правильність словесних тверджень і тим самим добиратися до істини або наближатися до неї. Математика робить осяжними довгі і часом туманні словесні описи, прояснює і економить думка.

Математичні методи дозволяють обґрунтовано прогнозувати майбутні події, замість того, щоб гадати на кавовій гущі або як-небудь інакше. Загалом, користь від застосування математики велика, але і праці на її освоєння потрібно багато. Однак вона окупатиметься сповна.

Психологія в своєму науковому становленні неминуче повинна була пройти і пройшла шлях математизації, хоча не у всіх країнах і не в повній мірі. Точна дата початку шляху математизація, мабуть, не знає жодна наука. Однак для психології як умовну дату початку цього шляху можна прийняти 18 квітня 1822 року. Саме тоді в Королівському німецькому науковому товаристві

Йоганн Фрідріх Гербарт прочитав доповідь «Про можливості і необхідність застосовувати в психології математику». Основна ідея доповіді зводилася до згаданої вище думки: якщо психологія хоче бути наукою, подібно фізиці, їй необхідно застосовувати математику.

Через два роки після цієї доповіді І. Ф. Гербарт видав книгу «Психологія як наука, заново заснована на досвіді, метафізиці та математиці». Ця книга примітна у багатьох відношеннях. Вона стала першою спробою створення психологічної теорії, що спирається на те коло явищ, які безпосередньо доступні кожному суб'єкту, а саме на потік уявлень, що змінюють один одного в свідомості. Гербарт за відсутності цих даних, як він сам писав, мав придумувати гіпотетичні моделі боротьби спливаючих і зникаючих в свідомості уявлень. Гербарт, маніпулюючи числовими значеннями параметрів, намагався описати можливі характеристики зміни уявлень.

Мабуть, І. Ф. Гербарту першому належить думка про те, що властивості потоку свідомості – це величини і, отже, вони в подальшому розвитку наукової психології підлягають вимірюванню. Йому також належить ідея «порога свідомості», і він перший вжив вислів «математична психологія».

У І. Ф. Гербарта в Лейпцігському університеті знайшовся учень і послідовник, який пізніше став професором філософії і математики, – Моріц-Вільгельм Дробиш. Він сприйняв, розвинув і по-своєму реалізував ідею вчителя. У словнику Брокгауза і Ефрона про Дробиш сказано, що ще в 30-х роках XIX століття він займався дослідженнями з математики та психології і публікувався на латинській мові. В 1842 р. М.В.Дробиш видав в Лейпцигу німецькою мовою монографію під недвозначною назвою: «Емпірична психологія згідно природничо-наукового методу».

Ця книга М.В. Дробиша дає чудовий приклад первинної формалізації знань в області психології свідомості. Там немає математики в сенсі формул, символіки і розрахунків, але там є чітка система понять про характеристики потоку уявлень в свідомості як взаємозалежні величини. Уже в передмові М.В. Дробиш написав, що ця книга передувє іншій, вже готовій, – мається на увазі книга з математичної психології. Оскільки його колеги-психологи недостатньо підготовлені в математиці, він визнав за необхідне спочатку продемонструвати емпіричну психологію без будь-якої математики, а лише на установлених природничих основах.

Не зрозуміло, подіяла ця книга на тодішніх філософів і богословів, що займалися психологією. Швидше за все – ні. Але вона, безсумнівно, вплинула на лейпцігських вчених з гуманітарною освітою. Через вісім років, в 1850 р в Лейпцигу вийшла в світ друга основна книга М.В. Дробиша – «Першооснови математичної психології».

Треба визнати, що по частині новацій в математичній психології Дробиш поступається зробленому своїм вчителем Гербартом. Правда, Дробиш до двох борюючих в свідомості уявлень додав третє, а це сильно ускладнило рішення. Але головне в іншому. Більшу частину обсягу книги складають приклади чисельного моделювання. На жаль, ні сучасники, ні

нащадки не зрозуміли і не оцінили науковий подвиг, здійснений М.В. Дробиш: у нього ж не було комп'ютера для чисельного моделювання. А в сучасній психології математичне моделювання – це продукт другої половини ХХ століття.

Основні методи тодішньої математики – диференційне й інтегральне вичислення, використання рівнянь порівняно нескладних залежностей. Цього було цілком достатньо для виявлення і опису найпростіших психофізичних законів і різних реакцій людини, але вони не годилися для вивчення складних психічних явищ і сутностей. Не дарма В.Вундт категорично заперечував можливості емпіричної психології в дослідженні вищих психічних функцій.

Математичні засоби для вивчення складних багатовимірних об'єктів, в тому числі вищих психічних функцій – інтелект, здібності, особистість, стали створювати англомовні вчені. З'явилося поняття «регресія», і були отримані рівняння, що виражають цю залежність. Був вдосконалений коефіцієнт, раніше запропонований французом Браве. Цей коефіцієнт кількісно виражає співвідношення двох змінюючих змінних, тобто кореляція. Тепер цей коефіцієнт – один з найважливіших засобів багатовимірного аналізу даних.

Ще будучи студентом Кембриджа, Френсіс Гальтон зауважив, що рейтинг успішності складання іспитів з математики, – а це був випускний іспит, – вимірюється від декількох тисяч до декількох сотень балів. Пізніше, пов'язавши це з розподілом талантів, Гальтон прийшов до думки про те, що спеціальні випробування дозволяють прогнозувати подальші життєві успіхи людей. Так в 80-х рр. ХІХ століття народився гальтоновський метод тестів.

Ідею тестів підхопили і розвинули французи – А. Біне, В. Анрі та інші, створивши перші тести для селекції соціально відсталих дітей. Це послужило початком психометричної тестології, що, в свою чергу, спричинило за собою розвиток психологічних вимірювань.

Великі масиви числових результатів вимірювань за тестами – бали, стали об'єктом численних досліджень, в тому числі математико-психологічних. Особлива роль тут належить англійському інженеру, що працював в Америці, – Чарльзу Спірмену.

По-перше, Ч. Спірмену, який вважав, що для обчислення кореляції між рядами цілочисельних балів, або рангів, потрібна спеціальна міра. Він перепробувавши різні варіанти зупинився, нарешті, на тій формі коефіцієнта кореляції рангів, яка з тих пір носить його ім'я.

По-друге, маючи справу з великими масивами числових результатів тестів і кореляцій між цими результатами, Ч. Спірмен припустив, що ці кореляції зовсім не виражає взаємовплив результатів, а виражають їх спільну мінливість під впливом загальної латентної психічної причини, або фактору, наприклад інтелект. Відповідно до цього Спірмен запропонував теорію «генерального» фактора, що визначає спільну мінливість змінних тестових результатів, а також розробив метод виявлення цього фактора за кореляційною матрицею. Це був перший метод факторного аналізу, який створений в психології і для психологічних цілей.

У однофакторній теорії Ч. Спірмена швидко знайшлися опоненти. Протилежну багатфакторну теорію, що пояснює кореляції запропонував Леон Терстоун. Йому ж належить перший метод мультифакторного аналізу, заснований на застосуванні лінійної алгебри. Після Ч. Спірмена і Л. Терстоуна факторний аналіз, не тільки став одним з найважливіших математичних методів багатовимірного аналізу даних в психології, а й вийшов далеко за її межі, перетворившись в загальнонауковий метод аналізу даних.

З кінця 20-х років XX століття математичні методи все ширше проникають в психологію і творчо використовуються в ній. Інтенсивно розвивається психологічна теорія вимірювань. На основі апарату ланцюгів Маркова розробляються стохастичні моделі навчання в психології поведінки. Створений в області біології Рональда Фішера дисперсійний аналіз стає основним математичним методом в генетичній психології. Математичні моделі з теорії автоматичного регулювання та Шеннон-Фано-Хартлі теорія інформації широко застосовуються в інженерній і загальній психології. В результаті сучасна наукова психологія в багатьох своїх галузях математизується.

## **1.2. Проблеми, з якими стикаються студенти вивчаючи «Математичні методи в 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 в психології”.

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

Так, О.Ю. Артем'єва та Є.М. Мартинов 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 ймовірно-статистичної мови). Проблемою ж дослідними вважаються механічне копіювання викладачами стандартних математичних курсів без врахування специфіки формалізації саме у психології.

Аналізуючи роботи Д. Кемпбелла 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 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 необхідністю проводити обчислення, або “калькулятофобія”.

Аналізуючи роботу А.Д. Наследова 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 операціоналізувати, і як наслідок – проблеми із підбором адекватних методів вимірювання психологічних явищ, шкал.

С.С. Паповян, 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 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 потім 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 в 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.

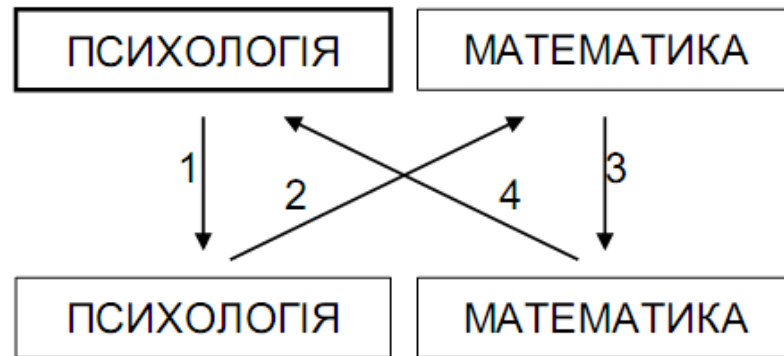
Четверта проблема – страх складних статистичних показників. Не всі студенти готові до обчислення вручну навіть  $t$ -критерія Стюдента. З іншого боку іноді зустрічається нелюбов до простих методів. Тоді студенти намагаються знайти найскладніший метод для аналізу найпростішого експерименту.

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.



Психолого-психологічна інтерпретація (1) – операціоналізація понять, встановлення співвідношення між предметом дослідження та методами збору даних або як влучно пише А.Д. Наследов, 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.

Розроблена класифікація проблем дає змогу системно підійти до їх вирішення, адже вирішуючи кожен 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.

### **1.3. Викладання «Математичних методів в психології» в умовах кредитно-модульної системи**

Завдяки специфіці кредитно-модульній системі організації навчального процесу 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 самотійним 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 самотійної роботи. Деканат же має змогу обчислити загальний обсяг самотійної роботи студентів з сіх дисциплін, і відповідно вирішити, чи реальний він для виконання. Такий підхід дозволяє

із годин, які відводяться для самостійної роботи, зробити реальний механізм 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, початковий 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 модульною 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.

Ще два резерви кредитно-модульної системи описує О.Л. Музика. 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.

### **Завдання на самостійну підготовку**

1. Історична логіка розвитку математичного апарату психології.
2. Особливості застосування математичної статистики та математичних методів у психології.
3. Методологічні основи використання математики у психології. Етапи математизації психології.

4. Класифікація математичних моделей (С. Паповян). Математичні моделі вимірювання. Математичні моделі структур і процесів. Теоретичні і емпіричні математичні моделі.
5. Математичні методи у системі методів психології.

## **ЛЕКЦІЯ 2**

### **ВИМІРЮВАННЯ В ПСИХОЛОГІЇ. ВИМІРЮВАННЯ І ШКАЛИ**

#### **План лекції**

- 2.1. Проблема вимірювання у психології
- 2.2. Типи та характеристики вимірювальних шкал

Час проведення: 2 учбові години.

#### **Література**

1. Климчук В.О. Математичні методи у психології. Навчальний посібник. – К.: Освіта України, 2009. – 288 с.
2. Кричевец А.Н. Математика для психологов: Учебник / А.Н. Кричевец, Е.В. Шикин, А.Г. Дьячков / Под ред. А.Н. Кричевца. М., 2003. – 198 с.
3. Наследов А.Д. Математические методы психологического исследования. – СПб.: Речь, 2004. – 392 с.
4. Телейко, А.Б. Математико-статистичні методи в соціології та психології: навч. посібник / А.Б. Телейко, Р.К. Чорней. – Київ : МАУП, 2007. – 418 с.

#### **2.1. Проблема вимірювання у психології**

В своїй роботі психолог досить часто стикається з проблемою вимірювання індивідуально-психологічних особливостей, таких, наприклад, як креативність, нейротизм, імпульсивність, властивості нервової системи і т.п. Для цього у психодіагностиці розробляються спеціальні вимірювальні процедури, втому числі і тести. Крім того в психології широко використовуються експериментальні методи і моделі дослідження психічних феноменів в пізнавальній і особистісній сферах. Це можуть бути моделі процесів пізнання (сприйняття, пам'яті, мислення) або особливості мотивації, ціннісних орієнтацій, особистості і т.п. Головне полягає втому, що в ході експерименту характеристики, що вивчаються, можуть отримувати кількісне вираження. Кількісні дані, отримані внаслідок ретельно спланованого експерименту за допомогою спеціальних вимірювальних процедур, в подальшому використовуються для математико-статистичної обробки.

Вимірювання може бути визначене як приписування чисел об'єктам або подіям, яке здійснюється за певними правилами. Ці правила повинні

встановлювати відповідність між деякими властивостями об'єктів, що розглядаються, з одного боку, і ряду чисел – з іншою. Загалом можна сказати, що вимірювання – це процедура, за допомогою якої об'єкт, що вимірюється, порівнюється з деяким еталоном і отримує чисельне вираження в певному масштабі або шкалі.

У кожному конкретному випадку вимірювання є операцією, за допомогою якої експериментальним даним надається форма зв'язного числового повідомлення. Саме закодована в числовій формі інформація дозволяє використати математичні методи і виявляти те, що без звернення до числової інтерпретації могло б залишитися прихованим; крім того, числове представлення об'єктів або подій дозволяє оперувати складними поняттями в більш скороченій формі. Саме це і є причиною використання вимірювань в будь-якій науці, в тому числі і психології.

Будь-який різновид вимірювання передбачає наявність одиниць вимірювання. Одиниця вимірювання – це та «вимірювальна паличка», як говорив С. Стівенс, яка є умовним еталоном для здійснення тих або інших вимірювальних процедур. У природничих науках і техніці існують стандартні одиниці вимірювання, наприклад, градус, метр, ампер і т.д.

Психологічні змінні за одиничними виключеннями не мають власних вимірювальних одиниць. Тому в більшості випадків значення психологічної ознаки визначається за допомогою спеціальних вимірювальних шкал.

Вимірювальна шкала – основне поняття, введене в психологію в 1950 р. С. Стівенсом; його трактування шкали і сьогодні використовується в науковій літературі.

Важливо розуміти, що приписування чисел об'єктам за певними принципами і правилами визначає тип шкали. Створення шкали можливе, оскільки існує ізоморфізм формальних систем і систем дій, що здійснюються над реальними об'єктами.

Розрізняють декілька типів шкал. Операції, а саме способи вимірювання об'єктів задають тип шкали. Шкала, в свою чергу, характеризується видом перетворень, які можуть бути віднесені до результатів вимірювання. Якщо не дотримуватися цього правила, то структура шкали порушиться, а дані вимірювання не можна буде осмислено інтерпретувати.

Тип шкали однозначно визначає сукупність статистичних методів, які можуть бути застосовані для обробки даних вимірювання.

Шкала (лати. *scala* – сходи) – інструмент для вимірювання безперервних властивостей об'єкта; являє собою числову систему, де відносини між різними властивостями об'єктів виражені властивостями числового ряду.

П. Суппес і Дж. Зінес дали класичне визначення шкали: «Нехай  $A$  – емпірична система з відносинами (ЕСВ),  $R$  – повна числова система з відносинами (ЧСВ),  $f$  – функція, яка гомоморфно відображає  $A$  в підсистемі  $R$  (якщо в області немає двох різних об'єктів з однаковою мірою, що є

відображенням ізоморфізму). Назвемо шкалою впорядковану трійку  $[A; R; f]$ ».

Звичайно як числова система  $R$  вибирається система дійсних чисел або її підсистема. Безліч  $A$  – це сукупність об'єктів, що вимірюються з системою відносин, визначеною на цій безлічі. Відображення  $f$  – правило приписування кожному об'єкту певного числа.

У цей час визначення Суппеса і Зінеса уточнене. По-перше, у визначення шкали вводиться  $G$  – група допустимих перетворень. По-друге, безліч  $A$  розуміється не тільки як числова система, але і як будь-яка формальна знакова система, яка може бути поставлена у відношення гомоморфізму з емпіричною системою. Таким чином, шкала – це четвірка  $[A; R; f; G]$ . Згідно з сучасними уявленнями, внутрішньою характеристикою шкали виступає саме група  $G$ , а  $f$  є лише прив'язкою шкали до конкретної ситуації вимірювання.

У цей час під вимірюванням розуміється конструювання будь-якої функції, яка ізоморфно відображає емпіричну структуру в символічній структурі. Як вже відмічено вище, зовсім не обов'язково такою структурою повинна бути числова. Це може бути будь-яка структура, за допомогою якої можна виміряти характеристики об'єктів, замінивши їх іншими, більш зручними у використанні (в тому числі – числами).

У психології різні шкали використовуються для вивчення різних характеристик соціально-психологічних явищ. Спочатку виділялися чотири типи числових систем, що визначали відповідно чотири рівні, або шкали вимірювання:

- 1) шкала найменувань – номінальна;
- 2) шкала порядку – рангова;
- 3) шкала інтервалів;
- 4) шкала відношень – пропорційна.

Перші дві шкали отримали назву неметричних, дві інші – метричних. Відповідно до цього в психології говорять і про два підходи до психологічних вимірювань: метричний (більш точний) і неметричний (менш точний). Ряд фахівців виділяють також абсолютну шкалу і шкалу різниць.

## 2.2. Типи та характеристики вимірювальних шкал

**Шкала найменувань.** Шкала найменувань виходить шляхом привласнення «імен» об'єктам. При цьому треба розділити безліч об'єктів на непересічні підмножини.

Інакшими словами, об'єкти порівнюються один з одним і визначається їх еквівалентність-нееквівалентність. Внаслідок цієї процедури утвориться сукупність класів еквівалентності. Об'єкти, що належать одному класу, еквівалентні один одному і відмінні від об'єктів, що відносяться до інших класів. Еквівалентним об'єктам привласнюються однакові імена.

Операція порівняння є первинною для побудови будь-якої шкали. Для побудови такої шкали треба, щоб об'єкт був рівний або подібний сам собі ( $x=x$  для всіх значень  $x$ ), тобто на безлічі об'єктів повинне бути реалізоване відношення рефлексивності. Для психологічних об'єктів, наприклад, випробуваних або психічних образів, це відношення реалізовується, якщо абстрагуватися від часу. Але оскільки операції попарного (зокрема) порівняння безлічі всіх об'єктів емпірично реалізуються неодноразово, то в ході емпіричного вимірювання навіть ця найпростіша умова не виконується.

Потрібно запам'ятати: будь-яка шкала – це ідеалізація, модель реальності, навіть така найпростіша, як шкала найменувань.

На об'єктах повинне бути реалізоване відношення симетрії ( $R(X=Y) \rightarrow R(Y=X)$ ) і транзитивності  $R(X=Y, Y=Z) \rightarrow R(X=Z)$ . Але на безлічі результатів психологічних експериментів ці умови можуть порушуватися.

Крім того, багаторазове повторення експерименту (накопичення статистики) приводить до «перемішування» складу класів: щонайбільше ми можемо дістати оцінку, що вказуватиме на ймовірність приналежності об'єкта до класу.

Таким чином, немає підстав говорити про шкалу найменувань (номінативну шкалу) як найпростішу шкалу, початковий рівень вимірювання в психології.

Існують більш «примітивні» (з емпіричної, але не з математичної точки зору) види шкал: шкали, засновані на відношеннях толерантності; шкали «розмиті» класифікації і т.п.

Про шкалу найменувань можна говорити в тому випадку, коли емпіричні об'єкти просто «маркуються» числом.

Отже, якщо об'єкти в якомусь відношенні еквівалентні, то ми маємо право віднести їх до одного класу. Головне, як говорив Стівенс, не приписувати один і той же символ різним класам або різні символи одному і тому ж класу.

Незважаючи на тенденцію «завищувати» потужність шкали, психологи дуже часто застосовують шкалу найменувань в дослідженнях. «Об'єктивні» вимірювальні процедури при діагностиці особистості приводять до типологізації: віднесенню конкретної особистості до того або іншого типу. Прикладом такої типології є класичні темпераменти: холерик, сангвінік, меланхолік і флегматик. Найпростіша номінальна шкала називається дихотомічною. При вимірюваннях за дихотомічною шкалою ознаки, що вимірюються, можна кодувати двома символами або цифрами, наприклад 0 і 1, або 2 і 6, або буквами А і Б, а також будь-якими двома відмінними один від одного символами. Ознака, виміряна за дихотомічною шкалою, називається альтернативною. У дихотомічній шкалі всі об'єкти, ознаки або властивості, що вивчаються, розбиваються на два непересічних класи, при цьому дослідник ставить питання про те, «чи виявилася» цікава для нього ознака у випробуваного чи ні.

Дослідник, що користується шкалою найменувань, може застосовувати наступні інваріантні статистики: відносні частоти, моду, кореляції випадкових подій.

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

Порядкова шкала можлива у випадку, якщо на множині реалізоване одне бінарне відношення – порядок (відношення «більше» і «менше»). Побудова шкали порядку – процедура більш складна, ніж створення шкали найменувань. Вона дозволяє зафіксувати ранг, або місце, кожного значення змінної по відношенню до інших значень. Цей ранг може бути результатом встановлення порядку між якимись стимулами або їх атрибутами самим випробуваним (первинний показник методик ранжування, або рейтингових процедур), але може і встановлюватися експериментатором як повторний показник (наприклад, при ранжуванні частот позитивних відповідей випробуваних на питання, що відносяться до різних тем).

Класи еквівалентності, виділені за допомогою шкали найменувань, можуть бути впорядковані за деяким критерієм. Розрізняють шкалу суворого порядку (сувора впорядкованість) і шкалу слабого порядку (слаба впорядкованість). У першому випадку на елементах безлічі реалізуються відношення «більше» і «менше», а у другому – «не більше або однаково» і «менше або однаково».

Значення величин можна замінювати квадратами, логарифмами, нормалізувати і т.д. При таких перетвореннях значень величин, визначених за шкалою порядку, місце об'єктів на шкалі не змінюється, тобто не відбувається інверсій.

Ще Стівенс висловлював точку зору, що результати більшості психологічних вимірювань щонайбільше відповідають лише шкалам порядку.

Шкали порядку широко використовуються в психології пізнавальних процесів, експериментальній психосемантиці, соціальної психології: ранжування, оцінювання, в тому числі педагогічне, дають порядкові шкали. Класичним прикладом використання порядкових шкал є тестування особистісних рис, а також здібностей. Більшість же фахівців в області тестування інтелекту вважають, що процедура вимірювання цієї властивості дозволяє використати інтервальну шкалу і навіть шкалу відношень.

Як би там не було, ця шкала дозволяє ввести лінійну впорядкованість об'єктів на деякій осі ознаки. Тим самим вводиться найважливіше поняття – властивість, що вимірюється, або лінійна властивість, тоді як шкала найменувань використовує «вироджений» варіант інтерпретації поняття «властивість»: «точкова» властивість (властивість є – властивості немає). 56

У порядковій (ранговій) шкалі повинно бути не менше трьох класів (груп): наприклад, відповіді на питальник: «так», «не знаю», «немає»; або – низький, середній, високий; і т.п., з тим розрахунком, щоб можна було



розставити вимірянні ознаки по порядку. Саме тому ця шкала і називається порядковою, або ранговою, шкалою.

Від класів просто перейти до чисел, якщо вважати, що нижчий клас отримує ранг (код або цифру) 1, середній – 2, вищий – 3 (або навпаки). Чим більше число класів, на які розбита вся експериментальна сукупність, тим ширші можливості статистичної обробки отриманих даних і перевірки статистичних гіпотез.

При кодуванні порядкових змінних їм можна приписувати будь-які цифри (коди), але в цих кодах (цифрах) обов'язково повинен зберігатися порядок, або, інакше говорячи, кожна подальша цифра повинна бути більше (або менше) попередньої.

Для інтерпретації даних, отриманих за допомогою порядкової шкали, можна використати більш широкий спектр статистичних засобів (в доповнення до тих, які допустимі для шкали найменувань).

Як характеристику центральної тенденції можна використати медіану, а як характеристику варіації – процентилі. Для встановлення зв'язку двох вимірювань допустима порядкова кореляція (т-Кендела і р-Спірмена).

Числові значення порядкової шкали не можна додавати, віднімати, ділити і множити.

**Шкала інтервалів.** Шкала інтервалів є першою метричною шкалою. Власне, починаючи з неї, доцільно говорити про вимірювання у вузькому значенні цього слова – про введення міри на безлічі об'єктів.

Шкала інтервалів визначає величину відмінностей між об'єктами у вияві властивості. За допомогою шкали інтервалів можна порівнювати два об'єкти. При цьому з'ясовують, на скільки більш або менш виражена певна властивість у одного об'єкта, ніж у іншого.

Шкала інтервалів дуже часто використовується дослідниками. Класичним прикладом застосування цієї шкали у фізиці є вимірювання температури за Цельсієм. Шкала інтервалів має масштабну одиницю, але положення нуля на ній довільне, тому немає сенсу говорити, у скільки разів більше або менше ранкова температура повітря, виміряна за шкалою Цельсія, ніж денна.

Інтервальна шкала дозволяє застосовувати практично всю параметричну статистику для аналізу даних, отриманих з її допомогою. Крім медіани і моди для характеристики центральної тенденції використовується середнє арифметичне, а для оцінки варіації – дисперсія. Можна обчислювати коефіцієнти асиметрії і ексцесу і інші параметри розподілу. Для оцінки величини статистичного зв'язку між змінними застосовується коефіцієнт лінійної кореляції Пірсона і т.д.

Більшість фахівців з теорії психологічних вимірювань вважають, що тести вимірюють психічні властивості за допомогою шкали інтервалів. Передусім, це стосується тестів інтелекту і досягнень. Числові значення одного тесту можна переводити в числові значення іншого тесту за допомогою лінійного перетворення:  $x' = ax + b$ .

Ряд авторів вважають, що відносити тести інтелекту до шкал інтервалів немає підстав. По-перше, кожний тест має «нуль» – будь-який індивід може отримати мінімальний бал, якщо не вирішить жодної задачі у відведений час. По-друге, тест має максимум шкали – бал, який випробуваний може отримати, вирішивши всі задачі за мінімальний час. По-третє, різниця між окремими значеннями шкали неоднакова. Принаймні, немає ніяких теоретичних і емпіричних підстав стверджувати, що 100 і 120 балів за шкалою IQ відрізняються на стільки ж, на скільки 80 і 100 балів.

Швидше усього, шкала будь-якого тесту інтелекту є комбінованою шкалою, з природним мінімумом і\або максимумом, але порядкової. Однак ці міркування не заважають тестологам розглядати шкалу IQ як інтервальну, перетворюючи «сирі» значення в шкальні за допомогою відомої процедури «нормалізації» шкали.

**Шкала відношень.** Шкалу відношень називають також шкалою рівних відношень. Особливістю цієї шкали є наявність абсолютного (твердо фіксованого) нуля, який означає повну відсутність якої-небудь властивості або ознаки. Шкала відношень є найбільш інформативною шкалою, що допускає будь-які математичні операції і використання різноманітних статистичних методів.

Шкала відношень по суті дуже близька інтервальній, оскільки якщо суворо фіксувати початок відліку, то будь-яка інтервальна шкала перетворюється в шкалу відношень.

Шкала відношень показує дані про вираженість властивостей об'єктів, коли можна сказати, у скільки разів один об'єкт більше або менше іншого.

Це можливе лише тоді, коли крім визначення рівності, рангового порядку, рівності інтервалів відомо рівність відношень. Шкала відношень відрізняється від шкали інтервалів тим, що на ній визначене положення «природного» нуля. Класичний приклад – шкала температур Кельвіна.

Саме в шкалі відношень відбуваються точні і надточні вимірювання в таких науках, як фізика, хімія, мікробіологія і інш. Вимірювання за шкалою відношень проводяться і в близьких до психології науках, таких, як психофізика, психофізіологія, психогенетика.

Вимірювання маси, часу реакції і виконання тестового завдання – області застосування шкали відношень.

Відмінністю цієї шкали від абсолютної є відсутність «природної» масштабної одиниці.

### **Інші шкали.**

*Дихотомічна класифікація* часто розглядається як варіант шкали найменувань. Це вірно, за винятком одного випадку, коли ми вимірюємо властивість, що має усього лише два рівні вираженості: «є-немає», так звана «точкова» властивість. Прикладів таких властивостей багато: наявність або відсутність у випробуваного якої-небудь спадкової хвороби (дальтонізм, хвороба Дауна, гемофілія і інш.), абсолютного слуху і інш. У цьому випадку дослідник має право провести «відцифровку» даних, приписуючи кожному з

типів цифру «1» або «0», і працювати з ними, як зі значеннями шкали інтервалів.

*Шкала різниць*, на відміну від шкали відношень, не має природного нуля, але має природну масштабну одиницю вимірювання. Їй відповідає аддитивна група дійсних чисел. Класичним прикладом цієї шкали є історична хронологія. Вона схожа зі шкалою інтервалів. Різниця лише в тому, що значення цієї шкали не можна множити (ділити) на константу. Тому вважається, що шкала різниць – єдина з точністю до зсуву. У психології шкала різниць використовується в методиках парних порівнянь.

*Абсолютна шкала* є розвитком шкали відношень і відрізняється від неї тим, що володіє природною одиницею вимірювання. У цьому її схожість зі шкалою різниць. Число вирішених задач («сирий» бал), якщо задачі еквівалентні, – один з виявів абсолютної шкали.

У психології абсолютні шкали не використовуються. Дані, отримані за допомогою абсолютної шкали, не перетворюються, шкала тотожна сама собі. Будь-які статистичні заходи допустимі.

У літературі, присвяченій проблемам психологічних вимірювань, згадуються і інші типи шкал: ординальна (порядкова) з природним початком, логінтервальна, впорядкована метрична і інш.

Все написане вище відноситься доодномірних шкал. Шкали можуть бути і багатомірними: шкальована ознака в цьому випадку має ненульові проєкції на два (або більше) відповідних параметри. Векторні властивості, на відміну від скалярних, є багатомірними.

Шкалювання – метод моделювання реальних процесів за допомогою числових систем. У науках соціальних – антропології, соціології, психології та ін. шкалювання є одним з найважливіших інструментів аналізу математичного явища, що вивчається, а також способом організації емпіричних даних, що отримуються за допомогою спостереження, вивчення документів, опитування, анкетування, експерименту або тестування. До основних процедур шкалювання відносяться:

- 1) парне порівняння об'єктів;
- 2) віднесення об'єктів до категорій і т.д.

Більшість соціальних і психологічних об'єктів неможливо суворо фіксувати відносно місця і часу їх існування, через що вони не піддаються прямому вимірюванню. Тому виникає питання про специфіку числової системи, що може співвідвестися з емпіричними даними такого роду. Різні методи шкалювання якраз служать особливими прийомами трансформації якісних характеристик в деяку числову змінну.

Загальний процес шкалювання складається з конструювання за певними правилами самої шкали і включає два етапи:

1) на етапі збору даних, від методів якого залежить вигляд соціально-психологічної інформації, створюється емпірична система досліджуваних об'єктів і фіксуються типи відношень між ними;

2) на етапі аналізу даних, від методів якого залежить обсяг інформації, будується числова система, що моделює відношення емпіричної системи об'єктів; іноді цей етап означається як вибір і реалізація методу шкалювання.

Є два типи задач, що вирішуються за допомогою шкалювання:

1) числове відображення сукупності об'єктів за допомогою їх усередненої групової оцінки; в цьому випадку відображення здійснюється за допомогою шкали оцінок;

2) числове відображення внутрішніх характеристик індивідів за допомогою фіксації їх відношення до деякого соціально-психологічного явища; в цьому випадку відображення здійснюється за допомогою шкали установок.

Отже, підсумовуючи все вище викладене, зазначимо, що шкали розрізняються не тільки за математичними властивостями, але і різними способами збору інформації. У кожній шкалі застосовуються певні методи аналізу даних. В залежності від типу задач, що вирішуються з допомогою шкалювання, будуються або шкали оцінок, або шкали для вимірювання соціальних установок. У практиці психологічних досліджень кожна шкала – незалежно від рівня вимірювання – має спеціальну назву, пов'язану з найменуванням властивості об'єкта, що вивчається.

Шкала оцінок – методичний прийом, що дозволяє розподілити сукупність об'єктів, що вивчаються, по мірі вираженості загальної для них властивості. Такий розподіл засновується на суб'єктивних оцінках даної властивості, усереднених по групі експертів.

У психології шкали оцінок стали застосовуватися одними з перших. Найпростіший приклад такої шкали – звичайна шкільна система балів.

Шкала установок – прийом, що дозволяє порівнювати індивідів по величині, інтенсивності і стійкості їх відношення до явища, що вивчається.

У соціальній психології шкала установок застосовується як один з головних засобів аналізу, бо тут об'єкт вимірювання, передусім, – особистісні якості. Побудова шкали установок пов'язана з підбором таких думок, що виражають весь спектр можливих відношень суб'єкта до певного соціально-психологічного явища. За шкалою оцінок кожна думка оцінюється групою експертів і отримує свій усереднений бал.

### **Завдання на самостійну підготовку**

1. Проблема вимірювання у психології. Об'єкти психологічного дослідження, їх властивості і ознаки. Внутрішньоіндивідуальна і міжіндивідуальна мінливість психіки.

2. Особливості вимірювання психічних явищ. «Опосередкованість» вимірювання психічних явищ. Поняття одиниці вимірювання у психології.

3. Вимірювальні шкали. Типи вимірювальних шкал (С. Стівенс). Основні властивості метричних і неметричних шкал вимірювання.

4. Характеристики психологічних даних, представлених у номінальній, порядковій, інтервальній та абсолютній шкалах вимірювання.

## ЛЕКЦІЯ 3 ПЕРВИННА ОПИСОВА СТАТИСТИКА

### План лекції

- 3.1. Міри центральної тенденції
- 3.2. Міри мінливості даних
- 3.3. Міри положення

Час проведення: 2 учбові години.

### Література

1. Ермолаев О.Ю. Математическая статистика для психологов. – М.: Моск. психологосоциальный ин-т: Флинта, 2002. – 321 с.
2. Климчук В.О. Математичні методи у психології. Навчальний посібник. – К.: Освіта України, 2009. – 288 с.
3. Наследов А.Д. Математические методы психологического исследования. – СПб.: Речь, 2004. – 392 с.
4. Суходольский Г.В. Математические методы психологии. СПб., – 2003. – 245 с.
5. Тарасов С.Г. Основы применения математических методов в психологии. СПб., – 1998. – 275 с.

### 3.1. Міри центральної тенденції

Всі методи кількісної обробки прийнято розділяти на первинні та вторинні.

Первинна статистична обробка націлена на впорядкування інформації про об'єкт і предмет вивчення. На цій стадії «сирі» відомості групуються за тими чи іншими критеріями, що заносяться у зведені таблиці. Первинно оброблені дані дають можливість дослідникові зрозуміти характер всієї сукупності даних в цілому: про їх однорідність-неоднорідність, компактність-розкиданість, чіткість-розмитість і т.д. Ця інформація чітко спостерігається в візуальних формах представлення даних і дає відомості про їх розподіл.

*До основних методів первинної статистичної обробки відносяться:* обчислення мір центральної тенденції, мір розкиду (мінливості) даних та квантилі розподілу.

Первинний статистичний аналіз всієї сукупності отриманих у дослідженні даних надає можливість охарактеризувати її в гранично стислому вигляді і відповісти на два головних запитання: 1) яке значення найбільш характерне для вибірки; 2) чи великий розкид даних щодо цього характерного значення, тобто яка варіативність даних. Для вирішення першого запитання обчислюються міри центральної тенденції, для вирішення другого – міри мінливості (або розкиду).

**Мірами центральної тенденції** називають чисельні показники типових властивостей емпіричних даних. Ці показники дають відповіді на питання про те, наприклад, «який середній рівень інтелекту студентів педагогічного університету?», «яке типове значення показника відповідальності певної групи осіб?». Існує порівняно невелика кількість таких показників-мір і в першу чергу: мода, медіана, середнє арифметичне. Кожна конкретна міра центральної тенденції має свої особливості, що роблять її цінною для характеристики об'єкта дослідження в певних умовах.

**Мода** ( $M_o$ ) – це значення, яке найбільш часто зустрічається у вибірці, тобто число з найбільшою частотою. Якщо всі значення в групі зустрічаються з однаковою частотою, то вважається, що моди немає. Якщо два сусідніх значення мають однакову частоту і більше частоти будь-якого іншого значення, мода є середнє цих двох значень. Якщо те ж саме відноситься до двох несуміжних значень, то існує дві моди, а група оцінок є бімодальною.

**Середнє арифметичне значень** (вибіркове середнє або середнє)  $M$  – це результат ділення суми всіх значень ( $X$ ) на їх кількість ( $N$ ).

$$M = \frac{\sum x_i}{N}$$

**Медіана** ( $M_e$ ) – це значення, яке приходить на середину упорядкованої послідовності емпіричних даних. Медіана не обов'язково повинна збігатися з конкретним значенням. Збіг відбувається у випадку непарного числа значень (відповідей), розбіжність – при парному їх числі. В останньому випадку медіана обчислюється як середнє арифметичне двох центральних значень у впорядкованому ряді даних.

Особливості мір центральної тенденції:

- мода вибірки обчислюється просто, її можна визначити «на око». Для дуже великих груп даних мода є досить стабільною мірою центру розподілу;
- медіана займає проміжне положення між модою і середнім з погляду її підрахунку. Ця міра особливо легко визначається у разі ранжированих даних;
- середнє арифметичне передбачає використання всіх значень вибірки, причому всі вони впливають на значення цієї міри.

Зазвичай вибіркове середнє застосовується при прагненні до найбільшої точності у визначенні центральної тенденції. Медіана обчислюється в тому випадку, коли у серії є «нетипові» дані, різко впливаючі на середнє. Мода використовується в ситуаціях, коли не потрібна висока точність, але важлива швидкість визначення міри центральної тенденції.

Обчислення всіх трьох показників проводиться також для оцінки розподілу даних. При нормальному розподілі даних середнє арифметичне значення, медіана і мода однакові або дуже близькі.

### 3.2. Міри мінливості даних

**Міри мінливості даних** – це статистичні показники, що дозволяють судити про ступінь однорідності отриманої множини даних, її компактності. Найбільш поширені в психологічних дослідженнях показники: розмах, дисперсія, стандартне відхилення.

**Розмах (R)** – це інтервал між максимальним і мінімальним значеннями ознаки. Він визначається легко і швидко, але чутливий до випадкових даних особливо при малій кількості досліджуваних. Це найпростіша міра відхилення.

Розмах визначається лише крайніми значеннями ознаки

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

**Дисперсія (D)** обчислюються тільки для інтервальних та абсолютних шкал. Дисперсія є середнім арифметичним значенням квадратів відхилень окремих значень ознаки від їхнього середнього арифметичного значення. Дисперсія має розмірність рівну квадратові розмірності ознаки, і обчислюється за формулою:

$$D = \frac{\sum_{i=1}^k n_i (x_i - M)^2}{n} = \sum_{i=1}^k w_i (x_i - M)^2$$

**Стандартне відхилення ( $\delta$ )** застосовуються для інтервальних та абсолютних даних. Щоб її визначити необхідно з дисперсії розрахувати квадратний корінь. Його позитивне значення і приймається за міру мінливості, іменовану середньоквадратичним або стандартним відхиленням:

$$\delta = \sqrt{D}$$

**Асиметрія (A)** характеризує ступінь несиметричності розподілу відносно його середнього. Позитивна асиметрія вказує на відхилення вершини розподілу в бік від'ємних значень, негативна – у бік додатних.

**Ексцес (E)** характеризує відносну опуклість або згладженість розподілу вибірки порівняно з нормальним розподілом. Позитивний ексцес позначає відносно загострений розподіл, негативний – відносно згладжений.

«Стандартом» розподілів служить нормальний розподіл з нульовою асиметрією і ексцесом.

### 3.3. Міри положення

У психології також широко використовуються **міри положення**, які називаються квантилями розподілу. **Квантиль** – це точка на числовій осі вимірної ознаки, яка ділить всю сукупність упорядкованих даних на дві групи з відомим співвідношенням їх чисельності. З одним з квантилів ми вже знайомі – це медіана. Це значення ознаки, яке ділить всю сукупність вимірювань на дві групи з рівною кількістю. Крім медіани часто використовуються проценти та квартили.

*Проценти́лі* – це 99 точок – значень ознаки ( $P_1 \dots P_{99}$ ), які ділять упорядковану (за зростанням) множину даних на 100 частин, які рівні за чисельністю. Визначення конкретного значення процентиля аналогічно визначенню медіани. Наприклад, при визначенні 10-го процентиля,  $P_{10}$ , спочатку всі значення ознаки упорядковуються за зростанням. Потім відраховується 10% досліджуваних, що мають найменшу вираженість ознаки.  $P_{10}$  буде відповідати тому значенню ознаки, яке відокремлює ці 10% досліджуваних від решти 90%.

*Кварти́лі* – це 3 точки – значення ознаки ( $P_{25}, P_{50}, P_{75}$ ), які ділять упорядковану (за зростанням) множину даних на 4 рівні за чисельністю частини. Перший квартиль відповідає 25-му процентилю, другий – 50-му процентилю або медіані, третій квартиль відповідає 75-му процентилю.

Проценти́лі і кварта́лі використовуються для визначення частоти виникнення тих чи інших значень (або інтервалів) вимірної ознаки або для виділення підгруп і окремих досліджуваних, які найбільш типові або нетипові для даної множини спостережень.

### **Завдання на самостійну підготовку**

1. Поняття міри центральної тенденції. Поняття мода, середнє арифметичне значення, медіана.
2. Квантілі розподілу. Проценти́лі та кварта́лі.
3. Міри мінливості. Поняття розмах, дисперсія, стандартне відхилення. Властивості дисперсії. Стандартизація.

## **ЛЕКЦІЯ 4**

### **ВСТУП ДО ПРОБЛЕМИ СТАТИСТИЧНОГО ВИСНОВКУ**

#### **План лекції**

- 4.1. Ідея перевірки статистичної гіпотези
- 4.2. Статистичний критерій та рівень статистичної значущості

Час проведення: 2 учбові години.

#### **Література**

1. Климчук В.О. Математичні методи у психології. Навчальний посібник. – К.: Освіта України, 2009. – 288 с.
2. Наследов А.Д. Математические методы психологического исследования. – СПб.: Речь, 2004. – 392 с.
3. Сидоренко Е.В. Методы математической обработки в психологии / Е.В. Сидоренко. – СПб.: Речь, 2006. – 350 с.
4. Тарасов С.Г. Основы применения математических методов в психологии. СПб., – 1998. – 275 с.



#### 4.1. Ідея перевірки статистичної гіпотези

Зазвичай дослідження проводиться для перевірки гіпотези, яка є наслідком теоретичних уявлень. Ця гіпотеза містить твердження про зв'язок абстрактних категорій, що відноситься до властивостей більш широкої сукупності об'єктів – генеральної сукупності.

Припущення, яке перевіряється із застосуванням наукового методу, будемо називати науковою гіпотезою. Слід зазначити, що не всяка гіпотеза, а лише та, яка допускає для своєї перевірки застосування наукового методу, може претендувати на науковість. Крім того, можна науково перевіряти гіпотези щодо будь-яких дрібних проблем, що володіють незначною науковою або практичною значимістю. Сам факт застосування наукового методу зовсім не гарантує, що перевіряема гіпотеза представляє науковий інтерес.

Застосування наукового методу для перевірки гіпотези передбачає певну послідовність дій дослідника. Дослідження починається з операціоналізації абстрактних категорій – визначення операцій, за допомогою яких може бути виміряна ця категорія явищ (агресивність, зовнішній вигляд і т.д.). Потім дослідник формує вибірку і проводить відповідні вимірювання. Результати вимірювання за допомогою описових статистик преведить до виду, що допускає статистичну перевірку наукової гіпотези.

Будь-яке дослідження зводиться до виявлення зв'язку між змінними. Зв'язок цей може виражатися в величині і напрямку відмінностей між порівнюваними групами або в знаку і величині коефіцієнта кореляції. Тобто зв'язок характеризується своєю силою і напрямком. Однак є ще одна не менш важлива характеристика зв'язку – її надійність, «істинність».

Надійність зв'язку безпосередньо пов'язана з репрезентативністю вибірки, з тим, наскільки впевнено статистики вибірки дозволяють судити про відповідні параметри в генеральній сукупності. Адже зв'язок, виявлений у вибірці, цікавить дослідника лише в тій мірі, в якій вона дозволяє судити про зв'язок, який існує в генеральній сукупності.

*Надійність зв'язку* визначається тим, наскільки ймовірно, що виявлений у вибірці зв'язок буде знову виявлений (підтвердиться) на іншій аналогічній вибірці, сформованій з тієї ж генеральної сукупності.

Очевидний спосіб перевірки надійності виявленого в дослідженні зв'язку – це багаторазове проведення аналогічного дослідження на різних вибірках. Однак для цього необхідно багато затратити багато сил і часу та й не завжди можливо багатократно проводити дослідження. Але можна сформулювати питання по-іншому. Якщо в генеральній сукупності зв'язку немає, то наскільки ймовірно є випадкове отримання даного результату дослідження? Інакше кажучи, наскільки ймовірним є те, що отриманий результат є випадковим, а насправді зв'язку в генеральній сукупності немає? Питання, сформульоване таким чином, дозволяє отримати відповідь з

використанням методів статистики. Відповідне перевіряєме твердження – про відсутність зв'язку – називається статистичною гіпотезою.

*Статистична гіпотеза* – це твердження щодо невідомого параметра генеральної сукупності, яке формулюється для перевірки надійності зв'язку і яке можна перевірити за результатами дослідження. Зазвичай виділяють основну (нульову) і альтернативну статистичні гіпотези. Основна (нульова) гіпотеза ( $H_0$ ) – містить твердження про відсутність зв'язку в генеральній сукупності. Альтернативна гіпотеза ( $H_1$ ) – приймається при відхиленні  $H_0$  і містить твердження про наявність зв'язку. При цьому нульова і альтернативна гіпотези представляють собою, в термінах теорії ймовірності, «повну групу несумісних подій»: якщо вірна одна з них, то інша є помилковою, і навпаки, відхилення однієї з них неминуче тягне прийняття іншої.

Відзначимо, що статистична перевірка наукової гіпотези слідує Аристотелевій логіці доказу «від протилежного». Дослідник зазвичай зацікавлений у встановленні зв'язку між явищами, що вивчаються, відповідно, його наукова гіпотеза зазвичай містить твердження про наявність такого зв'язку. Але методами статистики за результатами вибіркового дослідження перевіряється гіпотеза про відсутність відмінностей. І наукова гіпотеза підтверджується в тій мірі, в якій за результатами вибіркового дослідження можливе відхилення основної статистичної гіпотези.

#### **4.2. Статистичний критерій та рівень статистичної значущості**

Статистична значущість (Significant level), або р-рівень значущості (p-level), – основний результат перевірки статистичної гіпотези. Говорячи технічною мовою, це ймовірність отримання даного результату вибіркового дослідження за умови, що насправді для генеральної сукупності вірна нульова статистична гіпотеза – тобто зв'язку немає. Інакше кажучи, це ймовірність того, що виявлений зв'язок носить випадковий характер, а не є властивістю сукупності. Саме статистична значущість, р-рівень значущості є кількісною оцінкою надійності зв'язку: чим менша ця вірогідність, тим надійніший зв'язок.

Припустимо, при порівнянні двох вибірових середніх було отримано значення рівня статистичної значущості  $p = 0,05$ . Це означає, що перевірка статистичної гіпотези про рівність середніх у генеральній сукупності показала, що якщо вона вірна, то ймовірність випадкової появи виявлених відмінностей становить не більше 5%. Інакше кажучи, якби дві вибірки багаторазово формувалися з однієї і тієї ж генеральної сукупності, то в 1 з 20 випадків виявлялося б такаж або більша різниця між середніми цих вибірок. Тобто існує 5%-ва ймовірність того, що виявлені відмінності носять випадковий характер, а не є властивістю сукупності.

Відносно наукової гіпотези рівень статистичної значущості – це кількісний показник ступеня недовіри до висновку про наявність зв'язку,

обчислений за результатами вибіркової, емпіричної перевірки цієї гіпотези. *Чим менше значення р-рівня, тим вища статистична значущість результату дослідження, що підтверджує наукову гіпотезу.*

Корисно знати, що впливає на рівень значущості. Рівень значущості при інших рівних умовах вище (значення р-рівня менше), якщо:

- величина зв'язку (відмінності) більша;
- мінливість ознаки (ознак) менша;
- обсяг вибірки (вибірок) більша.

Чим більше гіпотез перевіряється, тим вищий шанс отримати результат чисто випадково – р-рівень збільшується пропорційно кількості перевірених гіпотез!

Наприклад, якщо результат вважається значущим при  $p < 0,05$  і перевіряється 20 гіпотез, то одна з гіпотез підтвердиться напевно, незалежно від дійсного стану справ. Єдиний шанс внести ясність – перевірити ці гіпотези на паралельній (ідентичній) вибірці.

*Статистичний критерій* (Statistical Test) – це інструмент визначення рівня статистичної значущості. Дослідник за формулою обчислює емпіричне значення критерію. Отримане емпіричне значення дозволяє визначити р-рівень – значення ймовірності того, що нульова статистична гіпотеза вірна.

Крім формули емпіричного значення, критерій задає формулу для визначення числа ступенів свободи. *Число ступенів свободи* (degrees of freedom – позначається як df або k) – це кількість можливих напрямків мінливості ознаки. Як правило, число ступенів свободи лінійно залежить від обсягу вибірки, від числа ознак або їх градації – чим більші ці показники, тим більше число ступенів свободи. У зв'язку з тим, що для кожного випадку визначення df має свою специфіку, зараз підкреслимо лише наступне. Кожна формула для розрахунку емпіричного значення критерію обов'язково супроводжується правилом (формулою) для визначення числа ступенів свободи.

Призначення критерію – перевірка статистичної гіпотези шляхом визначення р-рівня значущості (ймовірність того, що  $H_0$  вірна).

Критерій включає в себе:

- формулу розрахунку емпіричного значення критерію за вибірковими статистикам;
- правило (формула) визначення числа ступенів свободи;
- теоретичного розподілу для даного числа ступенів свободи;
- правило співвіднесення емпіричного значення критерію з теоретичним розподілом для визначення ймовірності того, що  $H_0$  вірна.

Для перевірки статистичних гіпотез застосовуються різні критерії. При цьому одному теоретичному розподілу можуть відповідати різні формули критеріїв – в залежності від статистичної гіпотези. Але принцип перевірки є загальним для всього цього різноманіття: обчислене за формулою емпіричне значення критерію порівнюється з теоретичним розподілом для заданого

числа ступенів свободи, що дозволяє визначити ймовірність того, що  $H_0$  вірна.

Безлічі розробленим статистичним критеріям відповідають безліч можливих формулювань статистичних гіпотези. Вибір критерію являє собою окрему проблему, яка буде розглядатися пізніше. А зараз будемо виходити з того, що дослідник вже вирішив проблему вибору критерію, і розглянемо загальну послідовність перевірки гіпотези.

При обробці даних на комп'ютері за допомогою статистичної програми (наприклад, SPSS) досліднику досить вказати програму, який критерій необхідно застосувати до заданої вибірки вихідних даних. Далі програма сама обчислює емпіричне значення критерію і зіставляє його з теоретичним розподілом. Як результат дослідник отримує значення  $p$ -рівня значущості разом з емпіричним значенням критерію і числом ступенів свободи.

Коли розрахунки проводяться «вручну», дослідник здійснює більш складну послідовність дій для перевірки гіпотези, що включає застосування спеціальних таблиць критичних значень критерію:

1. Вибір критерію в залежності від виду вихідних даних і статистичної гіпотези: теоретичного розподілу, формули розрахунку емпіричного значення критерію і числа ступенів свободи.

2. Розрахунок за вихідними даними (або за наявним статистикам) емпіричного значення критерію і числа ступенів свободи.

3. Застосування «Таблиці критичних значень критерію» дозволяє визначити значення  $p$ -рівня для даного числа ступенів свободи.

*Таблиця критичних значення* містять значення (квантилі) теоретичного розподілу, що відповідають найбільш важливими критичним значенням  $p$ -рівня (0,1; 0,05; 0,01 і т.д.).  $P$ -рівень значущості за емпіричним значенням критерію згідно таблиці критичних значення визначається наступним чином. Для даного числа ступенів свободи по таблиці визначаються найближчі критичні значення  $p$ -рівня. Далі значення  $p$ -рівня визначається у вигляді нерівності за правилом, яке демонструється на рис. 4.1 (значущість зростає зліва направо, відповідно до зменшення  $p$ -рівня):

- якщо емпіричне значення критерію ( $K_e$ ) знаходиться між двома критичними значеннями, то  $p$ -рівень менший того критичного  $p$ , яке знаходиться лівіше;

- якщо  $K_e$  знаходиться лівіше крайнього лівого критичного значення (зазвичай це відповідає критичному  $p = 0,1$ , рідше –  $p = 0,05$ ), то  $p$ -рівень більший, ніж крайнє праве критичне значення  $p$ ;

- якщо  $K_e$  знаходиться правіше крайнього правого критичного значення, то  $p$ -рівень менший крайнього правого критичного значення  $p$ .

Наприклад, якщо емпіричне значення критерію ( $K_e$ ) знаходиться між  $K_{0,05}$  і  $K_{0,01}$  то  $p < 0,05$ . Якщо  $K_e$  знаходиться лівіше  $K_{0,1}$  то  $p > 0,1$ . Якщо  $K_e$  знаходиться правіше  $K_{0,001}$  то  $p < 0,001$ .

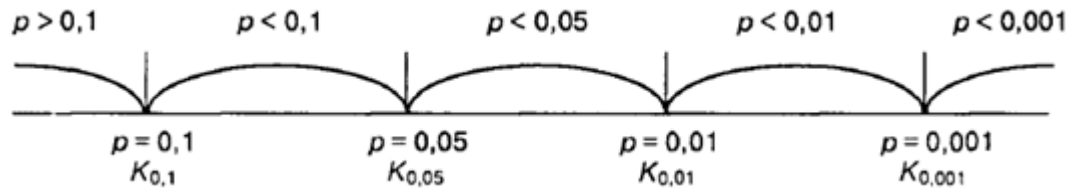


Рис. 4.1. Схема визначення р-рівня

До цих пір під перевіркою статистичної гіпотези ми мали на увазі процедуру визначення надійності зв'язку (р-рівень, як показник статистичної значущості). Однак в остаточному підсумку перевірка статистичної гіпотези повинна закінчуватися прийняттям статистичного рішення про те, яка ж гіпотеза вірна: нульова – про відсутність зв'язку або альтернативна – про її наявність. Відповідно, від цього залежить і остаточний, змістовний висновок дослідження: підтверджена чи ні вихідна наукова гіпотеза.

Цілком очевидно, що підставою для прийняття дослідником рішення про те, яка гіпотеза вірна, є р-рівень – ймовірність того, що вірна все-таки нульова гіпотеза. Чим менший р-рівень, тим з більшою впевненістю можна відхилити  $H_0$  на користь  $H_1$ , тим самим підтвердивши вихідну змістовну гіпотезу. Не менш очевидно й те, що приймаючи рішення, дослідник завжди допускає ймовірність його помилковості: адже дослідження проводиться на вибірці, а висновок робиться щодо генеральної сукупності. При відхиленні  $H_0$  на користь  $H_1$  дослідник ризикує, що зв'язку насправді в генеральній сукупності немає. І навпаки, рішення на користь  $H_0$  зовсім не виключає наявності зв'язку.

Питання про те, при якому ж р-рівні дослідник може відхилити  $H_0$ , вирішується переважно виходячи з неформальних критеріїв, прийнятих на основі практичного досвіду в різних областях дослідження. Традиційна інтерпретація різних рівнів значимості приведена в табл. 4.1. Відповідно до неї прийнятним для відхилення  $H_0$  визнається рівень  $p < 0,05$ . Така відносно висока ймовірність помилки 1 роду може бути рекомендована для невеликих вибірок (коли висока ймовірність помилки II роду). Якщо обсяги вибірок близько 100 і більше об'єктів, то поріг відхилення  $H_0$  доцільно приймати рішення про наявності зв'язку (відмінностей) при  $p < 0,01$ .

Таблиця 4.1.

Традиційна інтерпретація рівнів статистичної значущості

Рівень значущості	Рішення	Можливий статистичний висновок
$p > 0,1$	Приймається $H_0$	«Статистично достовірні відмінності не виявлені»
$p \leq 0,1$	Існують сумніви в істинності $H_0$ , невизначеність	«Відмінності виявлені на рівні статистичної тенденції»
$p \leq 0,05$	Значущість, відхиляється $H_0$	«Виявлені статистично достовірні (значущі) відмінності»

$p \leq 0,01$	Висока значущість, відхиляється $H_0$	«Відмінності виявлені на високому рівні статистичної значущості»
$p \leq 0,001$	Дуже висока значущість, відхиляється $H_0$	«Відмінності виявлені на дуже високому рівні статистичної значущості»

### **Завдання на самостійну підготовку**

1. Поняття гіпотези. Наукові та статистичні гіпотези.
2. Надійність зв'язку. Статистична значущість
3. Статистичний критерій та ступінь свободи. Перевірка гіпотез за допомогою статистичних критеріїв.
4. Види гіпотез. Нульова та альтернативна гіпотези. Спрямовані та неспрямовані гіпотези.

## **ЛЕКЦІЯ 5 ВИБІР МЕТОДУ СТАТИСТИЧНОГО ВИСНОВКУ**

### **План лекції**

- 5.1. Класифікація методів статистичного висновку
- 5.2. Генеральна сукупність та вибірка дослідження

Час проведення: 2 учбові години.

### **Література**

1. Ермолаев О.Ю. Математическая статистика для психологов. – М.: Моск. психологосоциальный ин-т: Флинта, 2002. – 321 с.
2. Климчук В.О. Математичні методи у психології. Навчальний посібник. – К.: Освіта України, 2009. – 288 с.
3. Кричевец А.Н. Математика для психологов: Учебник / А.Н. Кричевец, Е.В. Шикин, А.Г. Дьячков / Под ред. А.Н. Кричевца. М., 2003. – 198 с.
4. Наследов А.Д. Математические методы психологического исследования. – СПб.: Речь, 2004. – 392 с.
5. Сидоренко Е.В. Методы математической обработки в психологии / Е.В. Сидоренко. – СПб.: Речь, 2006. – 350 с.
6. Суходольский Г.В. Математические методы психологии. СПб., – 2003. – 245 с.

### **5.1. Класифікація методів статистичного висновку**

Приступаючи до операціоналізації змістовної гіпотези – до визначення того, як буде вимірюватись досліджуване явище, дослідник вже повинен уявляти собі, який метод статистичного висновку буде відповідати

одержаним в процесі дослідження вихідним даним. В іншому випадку він ризикує опинитися в драматичній ситуації, коли дані вже зібрані, але неможливо визначити метод їх аналізу.

Будь-яка змістовна гіпотеза наукового дослідження стосується зв'язку між явищами (властивостями, подіями) – незалежно від того, чи міститься в формулюванні гіпотези вказівка на зв'язок або на відмінності (між групами, умовами, подіями). Наприклад, формулювання «чоловіки і жінки відрізняються за комунікативною компетентністю» тотожне формулюванню «комунікативна компетентність пов'язана з статтю». Крім того, незалежно від свого формулювання, одна і та ж змістовна гіпотеза може бути перевірена за допомогою самих різних статистичних методів.

Окрім типів шкал, в яких виміряні або представлені досліджувані ознаки, на вибір методів статистичної перевірки гіпотези впливає кількість порівнюваних груп, залежність або незалежність порівнювальних вибірок і ряд інших причин. До цього можна додати велику кількість різноманітних статистичних критеріїв і варіантів їх застосування, які розроблені для самих різних дослідницьких ситуацій. Не дивно, що проблема вибору методу статистичного висновку, або проблема вибору критерію, часто стає складною навіть для досвідченого дослідника.

Проте, все безліч змістовних гіпотез може бути зведене до відносно невеликого числа типових дослідних ситуацій. Кожна така ситуація відповідає своїй структурі вихідних даних і оптимальному методу статистичної перевірки.

Відповідно до статистичних гіпотез статистичні критерії діляться на параметричні й непараметричні.

*Параметричні критерії* використовуються в завданнях перевірки параметричних гіпотез і включають у свій розрахунок показники розподілу, наприклад, середні, дисперсії тощо. Це такі відомі класичні критерії, як *z*-критерій, *t*-критерій Стюдента, *F*-критерій Фішера та ін. *Непараметричні критерії* перевірки гіпотез засновані на операціях з іншими даними, зокрема, частотами, рангами тощо. Це *λ*-критерій Колмогорова-Смірнова, *T*-критерій Вілкоксона, *U*-критерій Манна-Вітні та багато інших.

*Параметричні критерії* дозволяють прямо оцінити рівень основних параметрів генеральних сукупностей, різниці середніх і відмінності в дисперсіях. Критерії спроможні виявити тенденції зміни ознаки при переході від умови до умови, оцінити взаємодію двох і більш факторів у впливі на зміни ознаки. *Параметричні* критерії вважаються дещо більш потужними, ніж непараметричні, за умов, якщо ознака виміряна за інтервальною шкалою і нормально розподілена. Проте з інтервальною шкалою можуть виникнути певні проблеми, якщо дані, представлено не в стандартизованих оцінках. До того ж перевірка розподілу «на нормальність» вимагає досить складних розрахунків, результат яких заздалегідь невідомий. Найчастіше розподіли ознак відрізняються від нормального, тоді доводиться звертатися до непараметричних критеріїв.

*Непараметричні* критерії позбавлені перерахованих вище обмежень. Проте вони не дозволяють здійснити пряму оцінку рівня таких важливих параметрів, як середнє або дисперсія, з їхньою допомогою неможливо оцінити взаємодію двох і більше умов або факторів, що впливають на зміну ознаки. Непараметричні критерії дозволяють вирішити деякі важливі завдання, які супроводжують дослідження в психології і педагогіці: виявлення відмінностей у рівні досліджуваної ознаки, оцінка зсуву значень досліджуваної ознаки, виявлення відмінностей у розподілах ознак.

Застосування критеріїв для прийняття (відхилення) статистичних гіпотез завжди здійснюються з довірчою ймовірністю, інакше кажучи, на певному рівні значущості.

Таблиця 5.1.

## Порівняльна характеристика статистичних критеріїв

Параметричні	Непараметричні
1. Дають змогу прямо оцінити різниці середніх, отриманих у двох вибірках (t-критерій Стьюдента)	1. Дають змогу оцінити лише середні тенденції, відповісти на питання, чи частіше у вибірці А трапляються вищі, а у вибірці Б – нижчі значення ознаки (критерії $Q$ , $U$ , $\varphi^*$ та ін.)
2. Дають змогу прямо оцінити відмінності в дисперсіях (F-критерій Фішера)	2. Уможливають оцінювання лише відмінності в діапазонах варіативності ознаки (критерій $\varphi^*$ )
3. Уможливають виявлення тенденції зміни ознаки при переході від умови до умови при нормальному розподілі ознаки (однофакторний дисперсійний аналіз)	3. Уможливають виявлення тенденції зміни ознаки при переході від умови до умови при будь-якому розподілі ознаки (критерії тенденцій $L$ і $S$ )
4. Уможливають оцінювання взаємодії двох і більше факторів у їх впливі на зміни ознаки (двофакторний дисперсійний аналіз)	4. Ця можливість відсутня
5. Експериментальні дані повинні відповідати умовам: а) значення ознаки виміряні за інтервальною шкалою; б) розподіл ознаки є нормальним; в) у дисперсійному аналізі слід дотримуватися вимоги рівності дисперсій у комірках комплексу	5. Експериментальні дані можуть не відповідати жодній із цих умов: а) значення ознаки можуть бути пред'явлені у будь-якій шкалі; б) розподіл ознаки може бути будь-яким, збіг його з теоретичним законом розподілу не потребує перевірки; в) вимога рівності дисперсій відсутня



6. Математичні розрахунки дуже складні	6. Математичні розрахунки здебільшого прості і займають мало часу
7. Якщо умови, перераховані в п. 5, задовольняються, параметричні критерії виявляються дещо потужнішими, ніж непараметричні	7. Якщо умови, перераховані в п. 5, не задовольняються, непараметричні критерії виявляються потужнішими, ніж параметричні, оскільки вони менш чутливі до «засмічень»

Розглянемо класифікацію задач і методів їх розв'язання з використанням параметричних і непараметричних статистичних критеріїв (див. Табл. 5.2, 5.3).

Таблиця 5.2.

Класифікація задач і методів їх розв'язання з використанням параметричних статистичних критеріїв

Задача	Умови	Методи (критерії)
1. Оцінка відповідності емпіричного розподілу нормальному закону	зіставлення емпіричною розподілу з теоретичним	Критерії асиметрії і ексцесу, $\chi^2$ – критерій Пірсона
2. Пряма оцінка рівня середньої величини	а) дисперсія відома:	z-критерій
	б) дисперсія невідома.	t-критерій Стюдента
3. Пряма оцінка рівня дисперсії		$\chi^2$ – критерій Пірсона
4. Оцінка відмінностей у рівні середніх значень ознаки	а) 2 незв'язані вибірки однакової чисельності;	t-критерій Стюдента
	б) 2 незв'язані вибірки різної чисельності;	
	в) 2 зв'язані вибірки:	
	г) 3 і більше вибірок.	Однофакторний дисперсійний аналіз Фішера
5. Оцінка істотності різниць дисперсій ознаки	а) 2 незв'язані вибірки однакової чисельності;	F-критерій Фішера
	б) 2 зв'язані вибірки;	t -критерій Стюдента
	в) 3 і більше вибірок однакової чисельності;	q-критерій Кохрана
	г) 3 і більше вибірок різної чисельності.	M-критерій Бартлета
6. Аналіз змін ознаки під впливом контрольованих умов	а) під впливом одного фактора;	Однофакторний дисперсійний аналіз Фішера
	б) під впливом двох факторів.	Двохфакторний дисперсійний аналіз Фішера

Таблиця 5.3.

**Класифікація задач і методів їх розв'язання з використанням  
непараметричних статистичних критеріїв**

<b>Задача</b>	<b>Умови</b>	<b>Методи (критерії)</b>
1. Виявлення відмінностей у рівні досліджуваної ознаки	а) 2 вибірки;	Q-критерій Розенбаума, U-критерій Манна-Уїтні, ф-критерій кутового перетворення Фішера
	б) 3 і більше вибірок.	S-критерій Джонкіра, H-критерій Крускала-Уолліса
2. Оцінка зсуву значень досліджуваної ознаки	а) 2 виміри на одній і тій же вибірці;	T-критерій Вілкоксона, G-критерій знаків, ф-критерій кутового перетворення Фішера
	б) 3 і більше вимірів на одній і тій же вибірці	$\chi^2$ -критерій Фрідмана, L-критерій Пейджа
3. Виявлення відмінностей у розподілі досліджуваної ознаки	а) при зіставленні емпіричного розподілу з теоретичним;	$\chi^2$ -критерій Пірсона, $\lambda$ -критерій Колмогорова-Смірнова, m-біноміальний критерій
	б) при зіставленні двох емпіричних розподілів.	$\chi^2$ -критерій Пірсона, $\lambda$ -критерій Колмогорова-Смірнова, ф-критерій кутового перетворення Фішера
4. Аналіз змін ознаки під впливом контрольованих умов	а) під впливом одного фактора;	S-критерій Джонкіра, L-критерій Пейджа, Однофакторний дисперсійний аналіз Фішера
	б) під впливом двох факторів одночасно.	Двохфакторний дисперсійний аналіз Фішера

*Алгоритм прийняття рішення про задачу і метод її вирішення на стадії, коли дані вже отримані*

1. По першій колонці Табл. 5.2 і Табл. 5.3 визначити, яка із задач стоїть у вашому дослідженні.

2. По другій колонці Табл. 5.2 і Табл. 5.3 визначити, які умови вирішення вашої задачі, наприклад, скільки вибірок досліджено або на яку кількість груп ви можете розділити досліджену вибірку.

3. По третій колонці Табл. 5.2 і Табл. 5.3 визначити, які методи буде найбільш доцільно використовувати у вашій дослідницькій ситуації.

*Алгоритм прийняття рішення про задачу і метод її вирішення на стадії планування дослідження*

1. Визначте, яку модель ви вважаєте найбільш оптимальною для доведення ваших наукових передбачень.

2. Уважно ознайомтесь з описом методу, прикладами і задачами для самостійного вирішення у спеціальній літературі.

3. Якщо ви переконані у правильності вибору методу, зверніть особливу увагу на обмеження його застосування (друга колонка Табл. 5.2 і Табл. 5.3 або відповідний розділ спеціальної літератури) і вирішіть, чи зможете ви зібрати дані, які будуть відповідати цим обмеженням (великі об'єми вибірок, наявність кількох вибірок та ін.)

4. Проведіть дослідження, а потім проведіть обробку отриманих даних за попередньо вибраним алгоритмом, якщо не вдалося виконати обмеження.

5. Якщо обмеження виконати не вдалося, зверніться до алгоритму прийняття рішення про задачу і метод її вирішення на стадії, коли дані вже отримані.

## **5.2. Генеральна сукупність та вибірка дослідження**

Генеральною сукупністю вважається множина усіх об'єктів, властивості яких цікавлять дослідника. Вибіркою прийнято вважати підмножину генеральної сукупності, яка реально вивчається. Вибірка повинна адекватно відображати властивості сукупності, що цікавлять дослідника. Результати вибіркового дослідження мають бути поширені на усю сукупність.

У більшості статистичних досліджень потрібно з деякої генеральної сукупності сформулювати вибірку згідно заданим показникам адекватності. На підставі властивостей випадкової вибірки і висунутого критерію ефективності необхідно оцінити вибіркові статистичні характеристики. По вибіровим характеристикам належить методом математичної індукції зробити висновок про характеристики усієї генеральної сукупності досліджуваних об'єктів.

Таким чином, множина усіх можливих уявних спостережень, які могли б бути проведені при даному комплексі умов, прийнято вважати генеральною сукупністю. Оскільки у визначенні йдеться про уявно можливі спостереження, то генеральна сукупність вважається абстрактним поняттям, яке не слід змішувати з реальними сукупностями, що підлягають статистичному дослідженню. Закон або функція розподілу вірогідності повністю характеризують генеральну сукупність.

Вибірку можна тлумачити як деякий емпіричний аналог (модель) генеральної сукупності, з яким найчастіше на практиці доводиться мати справу, оскільки дослідження усієї генеральної сукупності буває або дуже

трудомістким, або принципово неможливим. Задачі математичної статистики фактично зводяться до обґрунтування висновку про об'єктивні властивості генеральної сукупності по результатам вибірки.

Достовірність висновків, що отримуються в результаті статистичної обробки даних, у багатьох випадках залежить від успішного вирішення питання *репрезентативності вибірки*, тобто повноти і адекватності подання властивостей генеральної сукупності моделлю у вигляді вибірки.

Співвідношення між властивостями генеральної сукупності і характеристиками вибірки можуть задовольняти умовам «репрезентативності», а можуть і не задовольнити їм. Слово «репрезентативний» означає типовий, характерний, адекватний, представлений рисами, подібний до чого-небудь.

Репрезентативність досягається завдяки побудові вибірки, яка найкращим чином представляє генеральну сукупність і дозволяє обґрунтовано переносити висновки, отримані при аналізі вибірки, на генеральну сукупність.

Репрезентативність являє собою таке подання інформації у вигляді, достатньому для отримання обґрунтованого розв'язку, заради якого воно (представлення) було створено. Наприклад, у вибірковому методі статистики репрезентативність інформації являє собою відповідність (адекватність) характеристик вибірки характеристикам генеральної сукупності.

Таким чином, найважливішою властивістю статистичної обробки даних є проблема репрезентативності вибірки. Вибірка повинна адекватно відображати властивості генеральної сукупності.

Виникає закономірне питання, як сформулювати репрезентативну вибірку? З точки зору статистики репрезентативність вибірки означає, що представлений в вибірці розподіл досліджуваних ознак відповідає (з певною часткою похибки) їх розподілу в генеральній сукупності.

Наведемо два методи, що забезпечують репрезентативність вибірки.

*Перший метод формування простої випадкової вибірки.* У цьому випадку вибірка складається з елементів, відібраних з генеральної сукупності таким чином, щоб кожен елемент цієї сукупності мав би рівні можливості (рівну ймовірність) потрапити до вибірки. Отримана таким чином вибірка називається простою випадковою вибіркою.

*Другий метод ґрунтується на понятті стратифікованої випадкової вибірки.* Для цього необхідно розбити елементи генеральної сукупності на страти (групи) за деякими характеристиками. Наприклад, при обстеженні попиту на деякий товар генеральну сукупність бажано розбити на групи, що розрізняються за величиною доходу, соціальної приналежності або навіть за місцем проживання (місто, село). Якщо проведена подібна розбивка сукупності і випадкова вибірка формується окремо з кожної групи (страти), то отримана в результаті вибірка носить назву стратифікована випадкова вибірка.

Як визначається обсяг вибірки? Підкреслимо, що він залежить насамперед від завдань дослідження. Тим не менш, можна сформулювати найбільш загальні рекомендації:

- Найбільший обсяг вибірки необхідний при розробці діагностичної методики – від 200 до 1000-2500 чоловік.

- Якщо необхідно порівнювати дві вибірки, їх загальна чисельність повинна бути не менше 50 осіб; чисельність порівнюваних вибірок повинна бути приблизно однаковою.

- Якщо вивчається взаємозв'язок між будь-якими властивостями, то обсяг вибірки повинен бути не менше 30-35 осіб.

- Чим більша мінливість досліджуваної властивості, тим більший повинен бути обсяг вибірки. Мінливість можна зменшити, збільшуючи однорідність вибірки, наприклад, за статтю, віком і т.д. При цьому, відповідно, зменшуються можливості генералізації висновків.

Окрім того *вибірки можуть бути залежними або незалежними*. Незалежні вибірки характеризуються тим, що ймовірність відбору будь-якого досліджуваного однієї вибірки не залежить від відбору будь-якого досліджуваного з іншої вибірки. Навпаки, залежні вибірки характеризуються тим, що кожен досліджуваний однієї вибірки поставлений у відповідність за певним критерієм досліджуваному з іншої вибірки.

### **Завдання на самостійну підготовку**

6. Статистичні критерії. Види статистичних критеріїв. Параметричні і непараметричні критерії. Порівняльна характеристика параметричних і непараметричних статистичних критеріїв.

7. Алгоритм прийняття рішення про задачу і метод її вирішення на стадії, коли дані вже отримані або на стадії планування дослідження.

8. Нормальний закон розподілення даних і його застосування. Характеристики параметрів нормального закону розподілення.

9. Дайте порівняльну характеристику основним стратегіям формування вибірок.

10. Які існують рекомендації щодо вибору оптимального обсягу вибірки?

11. Характеристика залежних і незалежних вибірок дослідження

## **ЛЕКЦІЯ 6 МЕТОДИ КОРЕЛЯЦІЙНОГО АНАЛІЗУ**

### **План лекції**

6.1. Сутність методів встановлення статистичних взаємозв'язків. Основні властивості коефіцієнта кореляції

6.2. Кореляція метричних змінних

6.3. Кореляція рангових змінних

#### 6.4. Кореляція дихотомічних змінних

Час проведення: 4 учбові години.

#### Література

1. Климчук В.О. Математичні методи у психології. Навчальний посібник. – К.: Освіта України, 2009. – 288 с.
2. Кричевец А.Н. Математика для психологов: Учебник / А.Н. Кричевец, Е.В. Шикин, А.Г. Дьячков / Под ред. А.Н. Кричевца. М., 2003. – 198 с.
3. Наследов А.Д. Математические методы психологического исследования. – СПб.: Речь, 2004. – 392 с.
4. Сидоренко Е.В. Методы математической обработки в психологии / Е.В. Сидоренко. – СПб.: Речь, 2006. – 350 с.
5. Телейко, А.Б. Математико-статистичні методи в соціології та психології: навч. посібник / А.Б. Телейко, Р.К. Чорней. – Київ: МАУП, 2007. – 418 с.

#### **6.1. Сутність методів встановлення статистичних взаємозв'язків. Основні властивості коефіцієнта кореляції**

Міри зв'язку виявляють співвідношенням, як правило, між двома змінними, які виміряні на одній вибірці. Ці зв'язки визначають через обчислення коефіцієнтів кореляції.

Термін «кореляція» вперше застосував французький палеонтолог Ж. Кюв'є, який вивів «закон кореляції частин і органів тварин» (цей закон дозволяє відновлювати за знайденими частинам тіла вигляд всієї тварини). У статистику зазначений термін ввів англійський біолог і статистик Ф. Гальтон.

Кореляційний аналіз для двох випадкових величин складається з:

- побудови кореляційного поля і складання кореляційної таблиці;
- обчислення вибірових коефіцієнтів кореляції і кореляційних відносин;
- перевірки статистичної значущості зв'язку.

Основне призначення кореляційного аналізу – виявлення зв'язку між двома або більше досліджуваними змінними, яка розглядається як спільна узгоджена зміна двох досліджуваних характеристик. Кореляційний зв'язок не можна вважати свідченням причинно-наслідкового зв'язку. Коефіцієнт кореляції характеризується за формою, напрямком і силою.

За формою кореляційний зв'язок може бути лінійним або нелінійним. Більш зручним для виявлення та інтерпретації кореляційного зв'язку є лінійна форма. Для лінійного кореляційного зв'язку можна виділити два основних напрямки: позитивний («прямий зв'язок») і негативний («зворотний зв'язок»).

Сила зв'язку безпосередньо вказує, наскільки синхронно проявляється спільна мінливість досліджуваних змінних. Наочне уявлення про характер і

зв'язку дає діаграма розсіювання – графік, осі якого відповідають значенням двох змінних, а кожен досліджуваний є точкою.

В якості числової характеристики ймовірнісної зв'язку коефіцієнти кореляції використовують значення в діапазоні від -1 до +1. Після проведення розрахунків дослідник, як правило, відбирає тільки найбільш сильні кореляції, які в подальшому інтерпретуються:

- сильний або тісний зв'язок при  $r \geq 0,7$ ;
- середній при  $0,5 \leq r < 0,7$ ;
- помірний при  $0,3 \leq r < 0,5$ ;
- слабкий при  $0,2 \leq r < 0,3$ ;
- дуже слабкий при  $r < 0,2$ .

Критерієм для відбору «досить сильних» кореляцій може бути як абсолютне значення самого коефіцієнта кореляції так і відносна величина цього коефіцієнта, яка визначається за рівнем статистичної значущості (від 0,01 до 0,1), який залежить від розміру вибірки. У малих вибірках для подальшої інтерпретації коректніше відбирати сильні кореляції на підставі рівня статистичної значущості. Для досліджень, які проведені на великих вибірках, краще використовувати абсолютні значення коефіцієнтів кореляції.

Таким чином, завдання кореляційного аналізу зводиться до встановлення напрямку (позитивного чи негативного) і форми (лінійної, нелінійної) зв'язку між ознаками, вимірюванню її тісноти, і перевірки рівня статистичної значущості отриманих коефіцієнтів кореляції.

На даний час розроблено безліч різних коефіцієнтів кореляції. Найбільш вживаними є коефіцієнти кореляції *r*-Пірсона, *r*-Спірмена і *r*-Кендалла. Як правило, комп'ютерні статистичні програми в меню «Кореляції» пропонують саме ці три коефіцієнта як основні, а для вирішення інших дослідницьких завдань пропонуються методи порівняння груп.

Вибір методу обчислення коефіцієнта кореляції, в першу чергу, залежить від:

- типу шкали, в якій виміряні змінні (номінальна, рангова, інтервальна, абсолютна);
- виду нормальності розподілу даних.

Вивчення зв'язку між ознаками, які приймають випадкові значення, починається з оцінювання його лінійності.

## 6.2. Кореляція метричних змінних

Для вивчення взаємозв'язку двох метричних змінних, виміряних на одній і тій же вибірці, застосовується *коефіцієнт кореляції r-Пірсона*. Сам коефіцієнт характеризує наявність тільки лінійного зв'язку між ознаками. Коефіцієнт лінійної кореляції є параметричним методом і його коректне застосування можливо тільки в тому випадку, якщо дані двох змінних відрізняється від нормального виду в незначній мірі.

При обробці даних «вручну» необхідно обчислити коефіцієнт кореляції, а потім визначити  $p$ -рівень значущості (з метою перевірки даних користуються таблицями критичних значень, які складені для цього критерію). Величина коефіцієнта лінійної кореляції Пірсона не може перевищувати  $+1$  і бути менше ніж  $-1$ . Ці два числа  $+1$  і  $-1$  є межами для коефіцієнта кореляції. Коли при розрахунку виходить величина, більша або менша  $+1$   $-1$ , це свідчить, що сталася помилка в обчисленнях.

При обчисленнях на комп'ютері статистична програма (SPSS, Statistica) супроводжує обчислений коефіцієнт кореляції більш точним значенням  $p$ -рівня.

Для статистичного рішення про прийняття або відхилення  $H_0$  зазвичай встановлюють  $p = 0,05$ , а для великого обсягу вибірки досліджуваних (100 і більше)  $p = 0,01$ . Якщо  $p < 0,05$ ,  $H_0$  відхиляється і робиться змістовний висновок, що виявлений статистично достовірний (значущий) зв'язок між досліджуваними змінними (позитивний чи негативний в залежності від знака кореляції). Коли  $p > 0,05$ ,  $H_0$  не відхиляється, змістовний висновок обмежений констатацією, що зв'язок (статистично достовірний) не виявлений.

Якщо зв'язок не виявлено, але є підстави вважати, що зв'язок насправді існує, слід перевірити можливі причини недостовірності зв'язку.

*Нелінійність зв'язку* – для встановлення цього необхідно проаналізувати графік двовимірного розсіювання. Якщо зв'язок нелінійний, але монотонний необхідно перейти до рангових кореляцій. Якщо зв'язок не монотонний, то доречно поділити вибірку досліджуваних на частини, в яких зв'язок монотонний, і обчислити кореляції окремо для кожної частини вибірки, або поділити вибірку на контрастні групи і далі порівнювати їх за рівнем вираженості ознаки.

*Наявність викидів і виражена асиметрія розподілу однієї або обох змінних.* Для цього необхідно проаналізувати гістограми розподілу частот обох ознак. При наявності викидів або асиметрії виключити викиди або перейти до рангових кореляцій.

*Неоднорідність вибірки* (проаналізувати графік двовимірного розсіювання). Спробувати розділити вибірку на частини, в яких зв'язок може мати різні напрямки.

Якщо ж зв'язок статистично достовірний, то перш ніж робити змістовний висновок, необхідно виключити можливість помилкової кореляції:

- *зв'язок обумовлений викидами.* При наявності викидів перейти до рангових кореляцій або виключити викиди;
- *зв'язок обумовлений впливом третьої змінної.* Якщо є таке явище, необхідно обчислити кореляцію не тільки для всієї вибірки, а й для кожної групи окремо. Якщо «третья» змінна метрична – обчислити приватну кореляцію.



### 6.3. Кореляція рангових змінних

Якщо до кількісних даних неприйнятний коефіцієнт кореляції  $r$ -Пірсона, то для перевірки гіпотези про зв'язок двох змінних після попереднього рангування можуть бути застосовані коефіцієнти кореляції  $r$ -Спірмена або  $\tau$ -Кендалла.

Для коректного обчислення обох коефіцієнтів (Спірмена і Кендалла) результати вимірювань повинні бути представлені в шкалі рангів або інтервалів. Принципових відмінностей між цими критеріями не існує, але прийнято вважати, що коефіцієнт Кендалла є більш «змістовним», так як він більш повно і детально аналізує зв'язку між змінними, перебираючи всі можливі відповідності між парами значень. Коефіцієнт Спірмена більш точно враховує саме кількісну ступінь зв'язку між змінними.

*Коефіцієнт рангової кореляції Спірмена* є непараметрическим аналогом класичного коефіцієнта кореляції Пірсона, але при його розрахунку враховуються не середнє арифметичне і дисперсія змінних, а ранги.

Застосування коефіцієнта кореляції Спірмена для перевірки гіпотез подібне до використання коефіцієнта  $r$ -Пірсона. У комп'ютерних програмах (SPSS, Statistica) рівні значущості для однакових коефіцієнтів  $r$ -Пірсона і  $r$ -Спірмена на завжди збігаються.

Перевага коефіцієнта  $r$ -Спірмена в порівнянні з коефіцієнтом  $r$ -Пірсона – більший чутливості до зв'язку. Ми використовуємо його в наступних випадках:

- наявність істотного відхилення розподілу хоча б однієї змінної від нормального вигляду (асиметрія, викиди);

- поява криволінійного (монотонного) зв'язку.

Обмеженням для застосування коефіцієнта  $r$ -Спірмена є:

- для кожної змінної не менше 5 спостережень;
- коефіцієнт при великій кількості однакових рангів по змінним дає огрублене значення.

*Коефіцієнт рангової кореляції  $\tau$ -Кендалла* є самостійним оригінальним методом, що спирається на обчислення співвідношення пар значень двох вибірок, що мають однакові або різні тенденції (зростання або спадання значень). Цей коефіцієнт називають ще коефіцієнтом конкордації. Таким чином, основною ідеєю даного методу є те, що про направлення зв'язку можна судити, попарно порівнюючи між собою досліджуваних: якщо у пари досліджуваних змінна  $X$  збігається за напрямком зі змінною  $Y$ , це свідчить про позитивний зв'язок, якщо не збігається – про негативний зв'язок.. У цьому методі одна змінна представляється у вигляді монотонної послідовності в порядку зростання величин, іншій змінній присвоюються відповідні рангові місця. Кількість інверсій (порушень монотонності в порівнянні з першим рядом) використовується у формулі для кореляційних коефіцієнтів. При підрахунку  $\tau$ -Кендалла «вручну» дані спочатку впорядковуються за змінною  $X$ . Потім для кожного досліджуваного підраховується, скільки разів його

ранг по  $Y$  виявляється меншим, ніж ранг досліджуваних, що знаходяться нижче. Результат записується в стовпець «Збіги». Сума всіх значень стовпчика «Збіг» і  $n$  – загальне число збігів, підставляється в формулу для обчислення коефіцієнта Кендалла, який є більш простим в обчислювальному відношенні, але при зростанні вибірки, на відміну від  $r$ -Спірмена, обсяг обчислень зростає не пропорційно, а в геометричній прогресії. При обчисленнях на комп'ютері в статистичній програмі (SPSS, Statistica) коефіцієнт Кендалла обраховується аналогічно коефіцієнтам  $r$ -Спірмена і  $r$ -Пірсона. Обчислений коефіцієнт кореляції  $\tau$ -Кендалла характеризується більш точним значенням  $p$ -рівня.

Застосування коефіцієнта Кендалла є кращим, якщо у вихідних даних є викиди.

Особливістю рангових коефіцієнтів кореляції є те, що максимальні за модулем значення кореляцій (+1, -1) не обов'язково відповідають прямим або обернено пропорційним зв'язкам між  $X$  і  $Y$ . Рангові кореляції досягають свого максимального за модулем значення, якщо більшому значенню однієї змінної завжди відповідає більше значення іншої змінної (+1), або більшому значенню однієї змінної завжди відповідає менше значення іншої змінної (-1).

Перевірка статистичної гіпотези, порядок прийняття статистичного рішення і формулювання змістовного висновку така ж, як і для  $r$ -Спірмена або  $r$ -Пірсона.

Якщо статистично достовірний зв'язок не виявлено, але є підстави вважати, що зв'язок насправді існує, слід спочатку перейти від коефіцієнта  $r$ -Спірмена до коефіцієнта  $\tau$ -Кендалла (або навпаки), а потім перевірити можливі причини недостовірності зв'язку:

- *нелінійність зв'язку*: для цього необхідно проаналізувати графік двовимірного розсіювання. Якщо зв'язок не монотонний, то ділимо вибірку на частини, в яких зв'язок монотонний, або ділимо вибірку на контрастні групи і далі порівнювати їх за рівнем вираженості ознаки;

- *неоднорідність вибірки*: необхідно проаналізувати графік двовимірного розсіювання, спробувати розділити вибірку на частини, в яких зв'язок може мати різні напрямки.

Якщо ж зв'язок статистично достовірний, то перш ніж робити змістовний висновок, необхідно виключити можливість помилкової кореляції (по аналогії з метричними коефіцієнтами кореляції).

#### 6.4. Кореляція дихотомічних змінних

При порівнянні двох змінних, виміряних в дихотомічній шкалі, мірою кореляційної зв'язку служить так званий *коефіцієнт  $\phi$* , який представляє собою коефіцієнт кореляції для дихотомічних даних.

Величина коефіцієнта  $\phi$  лежить в інтервалі між +1 та -1. Він може бути як позитивним, так і негативним, характеризуючи напрямок зв'язку двох

дихотомічних ознак. Проте інтерпретація  $\phi$  може породжувати специфічні проблеми. Неправильно вважати, що інтерпретуються значення  $r = 0,60$  і  $\phi = 0,60$  однакові. Коефіцієнт  $\phi$  можна обчислити методом кодування, а також використовуючи так звані чотириохклітинні таблиці або таблицю спряженості.

Для застосування коефіцієнта кореляції  $\phi$  необхідно дотримуватися таких умов:

- порівнювані ознаки повинні бути виміряні в дихотомічній шкалі;
- число ознак в порівнюваних змінних  $X$  і  $Y$  має бути однаковим.

Даний вид кореляції розраховують в комп'ютерній програмі SPSS на підставі визначення подібності/відмінності. Деякі статистичні процедури, такі як факторний аналіз, кластерний аналіз, багатовимірне шкалювання, побудовані на застосуванні цих принципів, а іноді самі представляють додаткові можливості для обчислення цих характеристик.

У тих випадках коли одна змінна вимірюється в дихотомічній шкалі (змінна  $X$ ), а інша в шкалі інтервалів або відносин (змінна  $Y$ ), використовується *бісеріальний коефіцієнт кореляції*, наприклад, при перевірці гіпотез про вплив статі дитини на показник зросту і ваги. Цей коефіцієнт змінюється в діапазоні від  $-1$  до  $+1$ , але його знак для інтерпретації результатів не має значення. Для його застосування необхідно дотримуватися таких умов:

- порівнювані ознаки повинні бути виміряні в різних шкалах: одна  $X$  – в дихотомічній шкалі; інша  $Y$  – в шкалі інтервалів або відносин;
- змінна  $Y$  повинна мати нормальний закон розподілу;
- число варіюючих ознак в порівнюваних змінних  $X$  і  $Y$  має бути однаковим.

Якщо ж змінна  $X$  виміряна в дихотомічній шкалі, а змінна  $Y$  в ранговій, необхідно використовувати *рангово-бісеріальний коефіцієнт кореляції*, який тісно пов'язаний з  $\tau$ -Кендалла. Інтерпретація результатів така ж сама.

### Завдання на самостійну підготовку

1. Характеристика кореляційного аналізу даних
2. Алгоритми розрахунку на матеріалі психологічних даних:
  - лінійної кореляції Пірсона;
  - рангової кореляції Спірмена;
  - « $\tau$ » Кендалла;
  - бісеріального коефіцієнту кореляції;
  - рангово-бісеріального коефіцієнту кореляції;
  - $\chi^2$  - Пірсона;
  - $\phi$ -коефіцієнта погодженості;
  - коефіцієнта множинної кореляції;
  - коефіцієнта приватної кореляції.

## ЛЕКЦІЯ 7

### ПАРАМЕТРИЧНІ МЕТОДИ ПОРІВНЯННЯ ДВОХ ВИБІРОК ДОСЛІДЖУВАНИХ

#### План лекції

- 7.1. Теоретичні засади та сфера застосування t-критерію Стьюдента
- 7.2. Статистичний критерій t-Стьюдента для однієї вибірки
- 7.3. Статистичний критерій t-Стьюдента для незалежних вибірок
- 7.4. Статистичний критерій t-Стьюдента для залежних вибірок

Час проведення: 4 учбові години.

#### Література

1. Ермолаев О.Ю. Математическая статистика для психологов. – М.: Моск. психологосоциальный ин-т: Флинта, 2002. – 321 с.
2. Климчук В.О. Математичні методи у психології. Навчальний посібник. – К.: Освіта України, 2009. – 288 с.
3. Кричевец А.Н. Математика для психологов: Учебник / А.Н. Кричевец, Е.В. Шикин, А.Г. Дьячков / Под ред. А.Н. Кричевца. М., 2003. – 198 с.
4. Наследов А.Д. Математические методы психологического исследования. – СПб.: Речь, 2004. – 392 с.
5. Сидоренко Е.В. Методы математической обработки в психологии / Е.В. Сидоренко. – СПб.: Речь, 2006. – 350 с.
6. Суходольский Г.В. Математические методы психологии. СПб., – 2003. – 245 с.
7. Тарасов С.Г. Основы применения математических методов в психологии. СПб., – 1998. – 275 с.
8. Телейко, А.Б. Математико-статистичні методи в соціології та психології: навч. посібник / А.Б. Телейко, Р.К. Чорней. – Київ: МАУП, 2007. – 418 с.

#### 7.1. Теоретичні засади та сфера застосування t-критерію Стьюдента

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

Критерій розроблений англійським математиком Вільямом С. Госсетом (1876-1937). «Стьюдент» – це псевдонім, під яким В. Госсет друкував свої роботи, працюючи в Дубліні на пивоварні Артура Гіннеса. Інший дослідник, який свого часу працював на Гіннеса, опублікував статтю, що містила конфіденційну комерційну інформацію, – з того часу працівникам підприємства було заборонено публікуватися (незалежно від змісту

публікації). Тож В. Госсет, щоб уникнути розголосу і обійти заборону, друкував свої роботи під псевдонімом.

В. Госсет тісно співпрацював із К. Пірсоном (багато його робіт надруковано у Пірсонівському журналі «Біометрика»), і саме К. Пірсон відправив його роботу «The probable error of a mean», у якій власне і дається опис t-статистики Р.Фішеру. Завдяки пропозиціям Р. Фішера t-критерій набув сучасного вигляду.

У вітчизняній традиції прийнято говорити про t-критерій Стьюдента. В англomовній літературі і в статистичних програмах вживають термін «Т-тест».

Критерій t-Стьюдента використовується у трьох випадках:

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

Застосовувавши критерій Стьюдента, ми дізнаємося, наскільки статистично значимі відмінності між двома вибірками і, відповідно, наскільки впевнено можна робити висновки про ці відмінності. Ситуації застосування критерію Стьюдента за змістом не відрізняються від ситуацій застосування критерію Манна-Уїтні, критерій знаків чи критерію Т-Вілкоксона – основна відмінність лежить у царині формальних характеристик числових даних. Зокрема, критерій Стьюдента значно потужніший за свої непараметричні аналоги, але вимагає дотримання двох вимог: 1) відповідності числових даних параметрам нормального розподілу; 2) рівності дисперсій двох вибірок, які порівнюються між собою.

#### ***Алгоритм застосування t-критерію Стьюдента:***

1. Перевірка нормальності розподілу даних у вибірках, що порівнюються.
2. Перевірка рівності (гомогенності) дисперсій незалежних вибірок або перевірка наявності прямого кореляційного зв'язку між залежними вибірками.
3. Власне обчислення t-критерію (схема обчислення критерію різниться залежно від того, порівнюються залежні чи незалежні вибірки; однаковий чи різний обсяг вибірок).
4. Порівняння отриманого емпіричного значення t-критерію із критичним табличним значенням і висновки про підтвердження чи спростування нульової гіпотези про відсутність відмінностей.

#### ***Обчислення t-критерію Стьюдента (без застосування комп'ютера)***

1. Перевірка нормальності розподілу даних у вибірках, що порівнюються. На нормальність перевіряють обидві вибірки. Для цього розраховують для кожної вибірки емпіричні значення асиметрії та ексцесу та порівнюють їх із критичними значеннями.

Емпіричне значення асиметрії розраховується за такою формулою:

$$A = \frac{\sum_i z_i^3}{n}$$

Формула для емпіричної величини ексцесу має такий вигляд:

$$E = \frac{\sum_i z_i^4}{n} - 3$$

Для обчислення величин  $z$  слід знайти середнє арифметичне ( $\bar{X}$ ) та стандартне відхилення ( $\sigma$ ) і потім розрахувати значення  $z$ , за формулою:

$$z_i = \frac{x_i - \bar{X}}{\sigma}$$

Розрахунок стандартного відхилення роблять за формулою:

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

Для розрахунків критичних значень асиметрії та ексцесу користуються формулами:

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

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

У всіх випадках  $n$  – обсяг вибірки.

Далі емпіричні значення асиметрії та ексцесу порівнюють із критичними значеннями. Рішення про нормальність розподілу даних приймають, якщо отримані емпіричні значення менші за критичні.

Перевірка нормальності розподілу є обов'язковою процедурою для  $t$ -критерію Стюдента, оскільки він був розроблений виключно для аналізу даних, розподілених нормально. У випадку проведення обчислень на комп'ютері у програмах Statistica, SPSS є статистичні критерії для перевірки нормальності – критерій Колмогорова-Смирнова, критерій Лілліфорса,  $W$ -критерій Шапіро-Вілка. Останнім часом більшість закордонних дослідників орієнтуються на критерій Лілліфорса, який власне є удосконаленим критерієм Колмогорова-Смирнова, або на критерій Шапіро-Вілка. Зверніть увагу, що перевірка нормальності полягає у аналізі відмінностей між частотним розподілом, побудованим на основі ваших даних, і нормальним розподілом, тому позитивним результатом є підтвердження нульової гіпотези – отримання статистично не значимого результату, при якому  $p$  більше 0,05 ( $p > 0,05$ ). Якщо перевірка нормальності дала негативний результат ( $p < 0,05$ ) – скористайтеся непараметричними критеріями.

1. Перевірка рівності (гомогенності) дисперсій незалежних вибірок.

Для перевірки гомогенності дисперсій А.Д. Наследов пропонує використовувати критерій F-Фішера (який теж є параметричним). Перш за все слід сформулювати дві статистичні гіпотези:

$H_0: \delta_1^2 = \delta_2^2$  (дисперсії обох вибірок рівні)

$H_1: \delta_1^2 \neq \delta_2^2$  (дисперсії обох вибірок нерівні)

Для проведення обчислень необхідно скористатися формулою F-критерію Фішера:

$$F_{\text{емп}} = \frac{\sigma_1^2}{\sigma_2^2}; df_{\text{чис}} = n_1 - 1; df_{\text{знам}} = n_2 - 1$$

Критичні значення ( $F_{\text{крит}}$ ) шукають у таблиці критичних значень F-критерію Фішера для перевірки ненаправлених альтернатив.

Якщо  $F_{\text{емп}} \geq F_{\text{крит}}$  – приймають рішення про підтвердження гіпотези про нерівність дисперсій на рівні  $p < 0,05$  і, відповідно, слід застосовувати не критерій Стюдента у стандартній формі, а критерій Велша (модифікований критерій Стюдента для незалежних вибірок із нехомогенними дисперсіями).

Для перевірки гомогенності дисперсій незалежних вибірок у програмі Statistica є два критерії: критерій Левена та критерій Брауна-Форсайта. Ці критерії доступні в діалоговому вікні обчислення критерію Стюдента. Зауважимо, що критерій Брауна-Форсайта рекомендують використовувати для порівняння дисперсій вибірок різного обсягу.

## 7.2. Статистичний критерій t-Стюдента для однієї вибірки

### *Алгоритм обчислення t-критерію Стюдента для однієї вибірки*

1. Сформулюйте статистичні гіпотези:

$H_0$  – відмінності між  $\bar{X}$  та  $A$  випадкові і незначимі.

$H_1$  – відмінності між  $\bar{X}$  та  $A$  достовірні, значимі.

2. Обчисліть емпіричне значення t-критерію ( $t_{\text{емп}}$ ) за формулою:

$$t_{\text{емп}} = \frac{|\bar{X} - A|}{\sigma / \sqrt{n}}; df = n - 1$$

$\bar{X}$  – середнє арифметичне;

$A$  – величина, з якою середнє арифметичне порівнюється;

$\sigma$  – стандартне відхилення;

$n$  – обсяг вибірки;

$df$  – число ступенів свободи.

3. Врахувавши число ступенів свободи ( $df$ ) за таблицею критичних значень критерію t-Стюдента, знайдіть  $t_{\text{крит}}$  і порівняйте отримані значення:

- якщо  $t_{\text{емп}} \geq t_{\text{табл}}$  – приймають гіпотезу  $H_1$ ;

- якщо  $t_{\text{емп}} < t_{\text{табл}}$  – приймають гіпотезу  $H_0$ .

### 7.3. Статистичний критерій t-Стюдента для незалежних вибірок

#### *Обчислення t-критерію Стюдента для незалежних вибірок (гомогенні дисперсії)*

Формальним критерієм незалежності вибірок є відсутність кореляції між ними. З точки зору змісту – незалежними є ті вибірки, між якими не можна виявити жодних зв'язків. Прикладом незалежних вибірок можуть бути рандомізовані (сформовані випадковим чином) контрольна та експериментальна група, два класи у школі, дві студентські групи тощо.

1. Сформулюйте статистичні гіпотези:

$H_0$  – відмінності між  $\bar{X}_1$  та  $\bar{X}_2$  випадкові і незначимі.

$H_1$  – відмінності між  $\bar{X}_1$  та  $\bar{X}_2$  достовірні, значимі.

2. Обчисліть емпіричне значення t-критерію ( $t_{\text{емп}}$ ) за формулою, якщо обсяги вибірок однакові:

$$t_{\text{емп}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{n}}}; df = 2n - 2$$

Якщо обсяги вибірок різні то емпіричне значення t-критерію ( $t_{\text{емп}}$ ) обчислюється за формулою:

$$t_{\text{емп}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}; df = n_1 + n_2 - 2$$

$\bar{X}_1$  та  $\bar{X}_2$  – середні арифметичне;

$\sigma_1^2, \sigma_2^2$  – стандартні відхилення відповідних вибірок;

$n, n_1, n_2$  – обсяги вибірок;

$df$  – число ступенів свободи.

3. Врахувавши число ступенів свободи ( $df$ ) за таблицею критичних значень критерію t-Стюдента, знайдіть  $t_{\text{крит}}$  і порівняйте отримані значення:

- якщо  $t_{\text{емп}} \geq t_{\text{табл}}$  – приймають гіпотезу  $H_1$ ;

- якщо  $t_{\text{емп}} < t_{\text{табл}}$  – приймають гіпотезу  $H_0$ .

#### *Обчислення t-критерію Велша для незалежних вибірок (модифікований t-критерій Стюдента для вибірок із негомогенними дисперсіями)*

Процедура обчислення критерію Велша відрізняється лише формулою для обчислення величини  $t$  та числа ступенів свободи:

$$t_{\text{емп}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$df = \frac{(\sigma_1^2/n_1 + \sigma_2^2/n_2)^2}{(\sigma_1^2/n_1)^2/(n_1 - 1) + (\sigma_2^2/n_2)^2/(n_2 - 1)}$$



$\bar{X}_1, \bar{X}_2$  – середнє арифметичне;  
 $\delta_1^2, \delta_2^2$  – стандартні відхилення відповідних вибірок;  
 $n_1, n_2$  – обсяги вибірок;  
 $df$  – число ступенів свободи.

#### 7.4. Статистичний критерій t-Ст'юдента для залежних вибірок

Формальним критерієм залежності вибірок є наявність кореляції між ними. З точки зору змісту – залежними є ті вибірки, між членами якої яких можна встановити однозначну відповідність. Прикладом залежних вибірок може бути експериментальна група до та після експериментального впливу. До певної міри залежними можна вважати вибірки, одна з яких сформована із чоловіків, а друга – з їх дружин, або одна вибірка – брати, друга – їх сестри.

##### **Обчислення t-критерію Ст'юдента для залежних вибірок**

1. Сформулюйте статистичні гіпотези:

$H_0$  – відмінності між  $\bar{X}_1$  та  $\bar{X}_2$  випадкові і незначимі.

$H_1$  – відмінності між  $\bar{X}_1$  та  $\bar{X}_2$  достовірні, значимі.

2. Оскільки тут ми маємо справу із залежними вибірками, фактично – із парами значень, то одиницею аналізу є різниця між цими парами значень. Саме тому спершу слід обчислити величини  $d_i$  – різниці між усіма парами значень.

3. Обчисліть емпіричне значення t-критерію ( $t_{\text{емп}}$ ) за формулою:

$$t_{\text{емп}} = \frac{|\bar{X}_d|}{\sigma_d / \sqrt{n}}; df = n - 1$$

$\bar{X}_d$  – середнє арифметичне різниці пар значень;

$\delta_d$  – стандартне відхилення різниці пар значень;

$n$  – обсяг вибірок;

$df$  – число ступенів свободи.

3. Врахувавши число ступенів свободи ( $df$ ) за таблицею критичних значень критерію t-Ст'юдента, знайдіть  $t_{\text{крит}}$  і порівняйте отримані значення:

- якщо  $t_{\text{емп}} \geq t_{\text{табл}}$  – приймають гіпотезу  $H_1$ ;

- якщо  $t_{\text{емп}} < t_{\text{табл}}$  – приймають гіпотезу  $H_0$ .

##### **Обмеження застосування**

Оскільки t-критерій Ст'юдента є параметричним, його застосування супроводжується рядом обмежень. Загальне обмеження для усіх варіантів критерію Ст'юдента – нормальність розподілу. Якщо дані розподілені ненормально, застосовувати критерій Ст'юдента не можна.

Обмеження, яке стосується виключно t-критерію Ст'юдента для незалежних вибірок, – гомогенність дисперсій обох вибірок. Якщо дисперсії негомогенні, користуватися цим варіантом критерію Ст'юдента не можна, – слід обрати критерій Велша.

Останнє обмеження стосовно t-критерію Стюдента для залежних вибірок – наявність значимого прямого кореляційного зв'язку між залежними вибірками. Якщо такої прямої кореляції немає, користуватися цим варіантом критерію Стюдента теж не можна.

Стосовно обсягів вибірок, жорстких обмежень немає. Однак дуже важко отримати нормальний розподіл на малих вибірках, тому досить часто дослідники для малих вибірок не проводять перевірки нормальності, а одразу застосовують непараметричні критерії (Манна-Уїтні, Вілкоксона тощо). Подібний крок відносно великих вибірок і критерію Стюдента є неправильним, адже великий обсяг вибірки зовсім не є гарантією нормальності розподілу даних.

### **Завдання на самостійну підготовку**

1. Особливості використання та алгоритм розрахунку критерію t-Стюдента для однієї вибірки.
2. Особливості використання та алгоритм розрахунку критерію t-Стюдента для залежних вибірок.
3. Особливості використання та алгоритм розрахунку критерію t-Стюдента для незалежних вибірок.

## **ЛЕКЦІЯ 8 НЕПАРАМЕТРИЧНІ МЕТОДИ ПОРІВНЯННЯ ВИБІРОК ДОСЛІДЖУВАНИХ**

### **План лекції**

- 8.1. Статистичний критерій U-Манна-Уїтні
- 8.2. Статистичний критерій G-знаків
- 8.3. Статистичний критерій  $\phi$  – кутове перетворення Фішера

Час проведення: 2 учбові години.

### **Література**

1. Ермолаев О.Ю. Математическая статистика для психологов. – М.: Моск. психологосоциальный ин-т: Флинта, 2002. – 321 с.
2. Климчук В.О. Математичні методи у психології. Навчальний посібник. – К.: Освіта України, 2009. – 288 с.
3. Кричевец А.Н. Математика для психологов: Учебник / А.Н. Кричевец, Е.В. Шикин, А.Г. Дьячков / Под ред. А.Н. Кричевца. М., 2003. – 198 с.
4. Наследов А.Д. Математические методы психологического исследования. – СПб.: Речь, 2004. – 392 с.
5. Сидоренко Е.В. Методы математической обработки в психологии / Е.В. Сидоренко. – СПб.: Речь, 2006. – 350 с.

6. Суходольский Г.В. Математические методы психологии. СПб., – 2003. – 245 с.
7. Тарасов С.Г. Основы применения математических методов в психологии. СПб., – 1998. – 275 с.
8. Телейко, А.Б. Математико-статистичні методи в соціології та психології: навч. посібник / А.Б. Телейко, Р.К. Чорней. – Київ: МАУП, 2007. – 418 с.

### 8.1. Статистичний критерій U-Манна-Уїтні

#### *Теоретичні засади та сфера застосування*

U-критерій вперше був запропонований Ф. Вілкоксоном (1945) для аналізу значимості відмінностей між двома однаковими за обсягом незалежними вибірками. Дослідники Г.Б. Манн та Д.Р. Уїтні (1947) модифікували його для вибірок різного обсягу. О. В. Сидоренко зазначає, що існує кілька способів обчислення U-критерія і, відповідно, кілька варіантів таблиць статистичної значимості, тож працюючи за формулою із одного джерела, варто користуватися саме тією таблицею статистичної значимості, яка у цьому ж джерелі наводиться. Також можна зустріти різні назви цього критерію: критерій Манна-Уїтні-Вілкоксона (скорочено – MWW), критерій Вілкоксона-Манна-Уїтні, критерій рангових сум Вілкоксона.

U-критерій Манна-Уїтні використовується для порівняння між собою результатів досліджень у *двох незалежних вибірках*. Формальним критерієм незалежності вибірок є відсутність кореляції між ними.

Застосувавши критерій Манна-Уїтні ми дізнаємося, наскільки статистично значимі відмінності між двома незалежними вибірками і, відповідно, наскільки впевнено можна робити висновки про ці відмінності. Також за допомогою цього критерію можна робити висновок про незначимість відмінностей між вибірками – це необхідно, наприклад, коли ми хочемо переконатися, що показники контрольної та експериментальної групи до експерименту не відрізняються, або хочемо довести, що середні показники IQ учнів 10-А і 10-Г класу однакові.

#### *Алгоритм розрахунку U-критерій Манна-Уїтні*

1. Обчисліть середні арифметичні для обох вибірок ( $\bar{X}_1$  та  $\bar{X}_2$ );
2. Сформулюйте нульову і альтернативну статистичні гіпотези ( $H_0$  та  $H_1$ ):  
 $H_0$  – відмінності між  $\bar{X}_1$  та  $\bar{X}_2$  випадкові (або їх взагалі не існує).  
 $H_1$  – відмінності між  $\bar{X}_1$  та  $\bar{X}_2$  достовірні, значимі.
3. Дані обох груп об'єднайте у один ряд, розташували їх у порядку зростання. Щоб не заплутатися, де яка група, дані однієї групи кодують 1, а другої групи – 2 (або, якщо зручніше, X та Y відповідно).
4. Значення об'єднаного ряду прорангуйте (присвойте кожному числу його ранг).
5. Знайдіть суми рангів для обох груп ( $R_1$  та  $R_2$ ).

6. Обчисліть два значення ( $U_1$  та  $U_2$ ) за формулами:

$$U_1 = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 \cdot n_2 + \frac{n_2 \cdot (n_2 + 1)}{2} - R_2$$

$n_1$  – кількість досліджуваних в 1 групі;

$n_2$  – кількість досліджуваних в 2 групі;

$R_1, R_2$  – суми рангів для першої та другої групи.

7. Після обчислень порівняйте  $U_1$  та  $U_2$  і оберіть найменше значення ( $U_{\text{емп}}$ ).

8. Зверніться до таблиці критичних значень критерію U-Манна-Уїтні. У комірці, що відповідає обсягам першої та другої вибірок ( $n_1$  та  $n_2$ ), знайдіть критичне значення U-критерія ( $U_{\text{крит}}$ ).

9. Порівняйте обчислене вами  $U_{\text{емп}}$  та знайдене у таблиці  $U_{\text{крит}}$ . Якщо  $U_{\text{емп}} \leq U_{\text{крит}}$  приймається гіпотеза  $H_1$  на рівні  $p \leq 0,05$ ; якщо  $U_{\text{емп}} > U_{\text{крит}}$  – приймається гіпотеза  $H_0$ .

### ***Особливості обчислення та інтерпретації***

При обчисленні U-критерія ми зіштовхуємося із необхідністю рангування. Якщо жодне число у наборі даних не повторюється, рангування здійснюється без особливих проблем. Однак може статися, що 2 і більше досліджуваних матимуть однакові показники. Тоді цим досліджуваним уявно проставляються їх ранги (наприклад, 4 та 5), а потім знаходиться їх середнє арифметичне (4,5), і тоді обом уже реально присвоюють ранги, тотожні цьому числу (4,5 та 4,5). Далі рангування продовжують з того рангу, на якому закінчилося уявне рангування (тобто наступному досліджуваному буде присвоєно ранг 6). Такі ранги називають зв'язаними.

Прийняття гіпотези  $H_0$  дає підстави стверджувати, що значимих відмінностей між вибірками (за параметром, що досліджувався) не виявлено. Прийняття гіпотези  $H_1$  на рівні  $p \leq 0,05$  дає підстави для такого висновку: за параметром, що досліджувався, між групами виявлено відмінності, статистично значимі на рівні  $p \leq 0,05$ . Зі статистичної точки зору це означає, що у 5% випадків ми можемо помилитися і виявлені нами відмінності не проявляться. Інакше кажучи, якщо ми повторно проведемо дослідження із вибіркою 100 осіб, то 5 із них не підпадуть під виявлену закономірність.

Розроблені також таблиці критичних значень для рівнів  $p \leq 0,01$ ;  $p \leq 0,001$ , а також для вибірок обсягом до 60 осіб (див. Сидоренко Е.В., с. 316).

### ***Обмеження застосування***

Критерій Манна-Уїтні є непараметричним, тобто при його застосуванні знімається вимога нормальності розподілу і вимога рівності дисперсій. В цьому плані критерій менш вимогливий, ніж його параметричний аналог – t-критерій Стюдента для незалежних вибірок, застосування якого вимагає і перевірки нормальності розподілу, і порівняння дисперсій за додатковими

статистичними критеріями. Водночас U-критерій менш чутливий, ніж t-критерій.

Застосовуючи критерій Манна-Уїтні, слід також знати деякі обмеження, які стосуються обсягів вибірок. По-перше, мінімальний обсяг вибірок – по 3 особи. Якщо в одній вибірці 2 особи, у іншій має бути мінімум 5. Максимальний обсяг вибірок – по 60 осіб. Однак, як зазначає О.В. Сидоренко, коли обсяги вибірок наближаються до 20, рангування стає доволі громіздким – варто подумати про статистичні програми (Statistica, SPSS).

Ще однією проблемою є зв'язані ранги, які впливають на результат обчислень. Якщо зв'язаних рангів небагато, їх вплив незначний і просто ігнорується. Якщо ж доля зв'язаних рангів є великою (більше половини), результати можуть бути суттєво спотворені. У цьому випадку буде доцільним або звернутися до обчислення у статистичній програмі (Statistica, SPSS), або подумати про t-критерій Ст'юдента для незалежних вибірок.

## 8.2. Статистичний критерій G-знаків

### *Теоретичні засади та сфера застосування*

Критерій знаків вперше було описано фізиком, математиком, публіцистом і сатириком Дж. Арбетнотом у роботі «Аргументи щодо Божественного Провидіння, отримані на основі регулярності народження людей обох статей» (1710), у якій він перевіряв, чи перевищує 50% доля народження хлопчиків у Лондоні. Пізніше виявилось, що критерій знаків є частинним випадком іншого критерію – біноміального – тому, як зазначає О.В. Сидоренко, інколи критерій знаків називають критерієм Мак-Немара.

Критерій знаків використовується для порівняння між собою результатів досліджень у двох залежних вибірках. Формальним критерієм залежності вибірок є наявність кореляції між ними.

Застосувавши критерій знаків, ми дізнаємося, наскільки статистично значимі відмінності між двома залежними вибірками і, відповідно, наскільки впевнено можна робити висновки про ці відмінності. Наприклад, можна зробити висновок про ефективність тренінгової роботи – слід перевірити статистичну значимість відмінностей середніх показників до та після тренінгу в експериментальній та контрольній групах.

Важливою перевагою критерія знаків є можливість його застосування не лише до кількісних, але й до якісних показників.

### *Алгоритм розрахунку критерію знаків*

Застосовувати критерій знаків варто за таким алгоритмом:

1. Обчисліть середні арифметичні для обох вибірок ( $\bar{X}_1$  та  $\bar{X}_2$ );
2. Сформулюйте нульову і альтернативну статистичні гіпотези ( $H_0$  та  $H_1$ );

$H_0$  – відмінності між  $\bar{X}_1$  та  $\bar{X}_2$  випадкові (або їх взагалі не існує).

$H_1$  – відмінності між  $\bar{X}_1$  та  $\bar{X}_2$  достовірні, значимі.

3. Дані обох груп записуємо поруч, щоб кожному досліджуваному відповідала його пара значень.

4. Порівнюємо пару значень кожного досліджуваного. Якщо спостерігається зростання показника – ставимо знак «+», якщо спадання – знак «-», показник не змінився – знак «=».

5. Підраховуємо кількість тих знаків, які нас цікавлять (якщо нас цікавить зростання показників – це будуть знаки «+», а якщо спадання – знаки «-»). Отримана кількість знаків і буде емпіричним значенням критерія знаків ( $G_{\text{емп}}$ ).

6. Зверніться до таблиці критичних значень критерію знаків. У комірці, що відповідає обсягу вибірки ( $n$ ), знайдіть критичне значення ( $G_{\text{крит}}$ ).

7. Порівняйте обчислене вами  $G_{\text{емп}}$  та знайдене у таблиці  $G_{\text{крит}}$ . Якщо  $G_{\text{емп}} < G_{\text{крит}}$  – приймається гіпотеза  $H_0$ ; якщо  $G_{\text{емп}} \geq G_{\text{крит}}$  – приймається  $H_1$  на рівні  $p \leq 0,05$ .

### ***Особливості обчислення та інтерпретації***

Прийняття гіпотези  $H_0$  дає підстави стверджувати, що значимих відмінностей між вибірками (за параметром, що досліджувався) не виявлено. Прийняття гіпотези  $H_1$  на рівні  $p \leq 0,05$  дає підстави для такого висновку: за параметром, що досліджувався, між групами виявлено відмінності, статистично значимі на рівні  $p \leq 0,05$ . Зі статистичної точки зору це означає, що у 5% випадків ми можемо помилитися і виявлені нами відмінності не проявляться.

Розроблені також таблиці критичних значень для рівня  $p \leq 0,01$  та для вибірок обсягом до 300 осіб (див. Сидоренко Е.В., с. 323).

### ***Обмеження застосування***

Критерій знаків є непараметричним, тобто при його застосуванні знімається вимога нормальності розподілу і вимога рівності дисперсій. В цьому плані критерій менш вимогливий, ніж його параметричний аналог – t-критерій Стюдента для залежних вибірок, застосування якого вимагає і перевірки нормальності розподілу, і порівняння дисперсій за додатковими статистичними критеріями. Водночас критерій знаків менш чутливий, ніж t-критерій. Існує непараметричний аналог критерія знаків – *T-критерій Вілкоксона*.

Вимоги до вибірок: мінімальний обсяг вибірок – 5 осіб, максимальний – 300.

## **8.3. Статистичний критерій $\phi$ – кутове перетворення Фішера**

### ***Теоретичні засади та сфера застосування***

Непараметричний критерій  $\phi$  використовують для порівняння двох вибірок за частотою виявлення певного ефекту. Критерій  $\phi$  оцінює достовірність відмінностей між відсотковими частками (іншими словами, пропорціями) двох вибірок, в яких зареєстровано ефект, що становить інтерес.

Він входить до числа багатофункціональних статистичних критеріїв. Це означає, що дані можуть бути представлені у будь-якій шкалі, починаючи від номінальної (шкали найменувань). Це також означає, що вибірки можуть бути як незалежними, так і залежними.

### **Алгоритм розрахунку критерію $\phi$ -Фішера**

1. Визначити значення ознаки, які будуть критерієм розподілу досліджуваних на тих, що мають «ефект», і тих, хто немає «ефекту».
2. Перевірити, чи виконуються умови застосування критерію  $\phi$ .
3. Визначити відсоткові частки досліджуваних, у яких «ефект є». Записати ці відсотки поруч з відповідними їм значеннями в лівому стовпці чотирикоміркової таблиці розміром  $2 \times 2$ . Перший стовпець – «ефект є», другий – «ефекту немає»; перший рядок – вибірка 1 (вибірка з більшою відсотковою часткою), другий – вибірка 2 (вибірка з меншою відсотковою часткою). У відповідний спосіб заповнити таблицю.

4. Сформулювати основну і альтернативну гіпотези.

$H_0$ : частка тих, хто має досліджуваний ефект, у вибірці 1 не більша, ніж у вибірці 2;

$H_1$ : частка тих, хто має досліджуваний ефект, у вибірці 1 більша, ніж у вибірці 2.

5. За таблицею визначити кути  $\phi_1$  та  $\phi_2$  для порівнюваних відсоткових часток.

6. Обчислити емпіричне значення  $\phi$  за формулою

$$\phi_{\text{емп}}^* = (\phi_1 - \phi_2) \sqrt{\frac{n_1 n_2}{n_1 + n_2}},$$

де,  $n_1$ ;  $n_2$  – обсяг вибірки відповідно першої та другої.

7. За таблицею визначити рівень значущості отриманого  $\phi_{\text{емп}}$ . На цій підставі дійти висновку про прийняття чи відхилення основної гіпотези. Якщо рівень значущості отриманого  $\phi_{\text{емп}}$  менший від 0,01 (тобто 1%), гіпотеза  $H_0$  відхиляється, якщо щонайменше 0,05 – приймається. В інших випадках гіпотеза  $H_0$  відхиляється на рівні значущості, що відповідає отриманому  $\phi_{\text{емп}}$  (тобто на рівні значущості 1-5 %)

### **Обмеження критерію $\phi$ -Фішера**

Перед застосуванням кутового перетворення Фішера потрібно перевірити такі умови-обмеження.

1. Жодна з часток, що зіставляються не повинна дорівнювати нулю.
2. Верхня межа в критерії  $\phi$  відсутня – вибірки можуть бути якими завгодно великими.
3. Нижня межа – 2 спостереження в одній з вибірок. При цьому необхідно, щоб вибірки задовольняли наступним умовам:
  - якщо в одній вибірці лише 2 спостереження, то в іншій повинно бути не менше 30;
  - якщо в одній вибірці лише 3 спостереження, в іншій має бути не менше 7;

- якщо в одній вибірці лише 4 спостереження, друга має містити не менше 5;
- при  $n_1, n_2 \geq 5$  можливі будь-які порівняння.

### **Завдання на самостійну підготовку**

1. Особливості використання та алгоритм розрахунку критерію U-Манна-Уїтні.
2. Особливості використання та алгоритм розрахунку критерію Т-Вілкоксона.
3. Особливості використання та алгоритм розрахунку критерію G-знаків.
4. Особливості використання та алгоритм розрахунку критерію  $\phi$ -Фішера.