**ЛЕКЦІЯ 6**

**ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ**

1. Наукова і статистична гіпотеза.
2. Нульова та альтернативна гіпотеза.
3. Статистичний критерій та кількість ступенів свободи.
4. Рівень статистичної значущості.
5. Правила прийняття статистичного висновку.

***1. Наукова і статистична гіпотеза***

Одна з основних завдань, що стоять перед психологом, який проводить емпіричне дослідження, – це з'ясування того, які висновки про властивості генеральної сукупності можна зробити з вибіркового спостереження.

***Наукова гіпотеза*** – *це наукове припущення, що висувається для пояснення будь-якого явища і потребує перевірки на досліді та теоретичного обґрунтування для того, щоб стати достовірною науковою теорією*.

***Статистична гіпотеза*** – *припущення на певному рівні статистичної значущості про властивості генеральної сукупності за оцінками вибірки*.

Суть перевірки статистичної гіпотези полягає в тому, щоб установити, чи узгоджуються експериментальні дані та висунута гіпотеза, чи можливо віднести розходження між гіпотезою та результатом статистичного аналізу експериментальних даних до випадкових причин.

Отже, *наукова* (первинна, змістова) гіпотеза слугує для організації експерименту, *статистична* (вторинна) – для здійснення процедури порівняння реєстрових параметрів, тобто вона необхідна на етапі математичної інтерпретації емпіричних даних.

Сформульовану статистичну гіпотезу позначають літерою Н (Hypothesis).

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 генеральної сукупності.

Надійність зв'язку визначається тим, наскільки ймовірно, що виявлений у вибірці зв'язок буде знову виявлений (підтвердиться) на іншій аналогічної вибірці, сформованої 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 витратити багато сил і часу та й не завжди можливо багатократно проводити дослідження.

Але можна сформулювати питання по-іншому. Якщо в генеральної 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.

*Статистична гіпотеза* – це твердження щодо невідомого параметра генеральної сукупності, яке формулюється для перевірки надійності зв'язку і яке можна перевірити за результатами дослідження.

***2. Нульова та альтернативна гіпотеза***

***Нульова гіпотеза*** – це гіпотеза про відсутність відмінностей. Вона позначається H0 і називається нульовою тому, що містить **0**: *Х*1-*Х*2=0, де *Х*1 та *Х*2 – показники, між якими з’ясовують відмінність.

***Альтернативна гіпотеза*** – це гіпотеза про існування та значимість відмінностей. Позначається H1. Альтернативна гіпотеза – це те, що ми в більшості випадків хочемо довести.

Нульова гіпотеза Но – це гіпотеза про схожість.

Альтернативна гіпотеза Н1  – гіпотеза про відмінність.

Якщо, наприклад, дві вибірки отримано з нормально розподілених генеральних сукупностей, причому одна вибірка має параметри х1 і σ1, а друга *х*2 і σ 2, то нульова гіпотеза буде виходити з положення, що *х*1 = *х*2 і σ1= σ2, тобто різниця двох середніх х1 − *х*2 = 0 і різниця двох стандартних відхилень σ1-σ2= 0 (звідси і назва гіпотеза нульова).

Нульова та альтернативна гіпотези можуть бути *спрямованими* і *неспрямованими*.

Неспрямовані гіпотези говорять тільки про наявність чи відсутність відмінностей. *Спрямовані* гіпотези говорять не тільки про наявність чи відсутність відмінностей, але і вказують напрямок відмінностей.

Наприклад, Н0: рівень розумового розвитку гімназистів не перевищує рівня розумового розвитку учнів загальноосвітніх шкіл;

H1: рівень розумового розвитку гімназистів перевищує рівень розумового розвитку учнів загальноосвітніх шкіл.

У статистичних гіпотезах мова йде не про арифметичні (тобто числові), а про статистично значущі відмінності: тобто чи з однаковою частотою зустрічаються різні значення ознаки в обох емпіричних розподілах.

Альтернативна гіпотеза H1 приймається, (а не доводиться), якщо за допомогою міркувань ми можемо відкинути нульову гіпотезу H0, тобто альтернативна гіпотеза підтверджується не прямо, а опосередковано.

Якщо ми помітили, що в одній з груп індивідуальні значення досліджуваних за будь-якою ознакою, наприклад за соціальною активністю, вищі, а в іншій нижчі, то для перевірки значущості цих відмінностей нам необхідно сформулювати спрямовані гіпотези.

Якщо ми хочемо довести, що в групі А під впливом певних експериментальних впливів відбулися більш виражені зміни, ніж в групі Б, то нам теж необхідно сформулювати спрямовані гіпотези.

Якщо ж ми хочемо довести, що розрізняються форми розподілу ознаки в групах А і Б, то формулюються неспрямовані гіпотези.

Перевірка гіпотез здійснюється за допомогою критеріїв статистичної оцінки відмінностей.

**Спрямовані** (однобічні) статистичні гіпотези формулюються так:

Н0: *х*1≤*х*2 (*х*1 – не перевищує *х*2)

Н1: *х*1>*х*2 (*х*1 – перевищує *х*2)

**Неспрямовані**:

Н0: *х*1=*х*2 (*х*1 – не відрізняється від *х*2)

Н1: *х*1>*х*2 (*х*1 – відрізняється від *х*2)

***3. Статистичний критерій та кількість ступенів свободи***

Перевірку статистичних гіпотез здійснюють на основі статистичних критеріїв. **Статистичний критерій** – це ознака, на основі якої виконується оцінювання явища або процесу.

Критерій призначено для прийняття істинної гіпотези і відхилення хибної.

Залежно від вимог щодо певного розподілу вибірки статистичні критерії бувають:

**параметричні** – це критерії, в яких у формулу розрахунку включені параметри розподілу (Х і σ). До них відносяться t-критерій Стьюдента, критерій F – Фішера та ін.;

**непараметричні** – це критерії, що не включають у формулу розрахунку параметри розподілу і засновані на оперуванні частотами або рангами. До них відносяться критерій Q – Розенбаума, критерій знаків та ін.

***Статистичний критерій*** – це вирішальне правило, що забезпечує прийняття істинної чи відхилення помилкової гіпотези з високою ймовірністю.

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 Н0 вірна.

Однією з основних характеристик критерію є його **потужність**. ***Потужність статистичного критерію*** – це його здатність не припуститися помилки, тобто визнати відмінності недостовірними, в той час, як вони достовірні. Одні і ті ж завдання можуть бути вирішені за допомогою різних критеріїв, що володіють різною потужністю.

Кожний із критеріїв має свої переваги і недоліки, можливості та обмеження.

Вибір критерію доказу відмінностей залежить від таких особливостей груп:

* типу шкали, в якій виміряно досліджувану ознаку;
* характеру розподілу ознаки (нормальний або відмінний від нього);
* кількості порівнюваних вибірок і їх обсягу;
* якості вибірок (залежні чи незалежні).

Параметричні критерії вважаються більш потужними, ніж непараметричні, якщо ознака виміряна за інтервальною шкалою і нормально розподілена. Якщо розподіл ознаки відрізняється від нормального, використовують непараметричні критерії. Проте з їх допомогою не можна оцінити взаємодію двох і більше умов (факторів), що впливають на зміну ознаки. Це завдання розв’язують лише шляхом двофакторного дисперсійного аналізу.

***Залежними*** називаються групи, між якими є суттєві внутрішні зв'язки: вибірки, що складаються з батьків і дітей, братів і сестер (генетичний зв'язок); одна і та ж група, досліджувана двічі (наприклад, до і після педагогічного впливу).

***Незалежні*** групи – це групи, між якими немає істотних зв'язків (наприклад, групи мають однаковий вік різних професій, групи різного віку чи статі, які не є родичами).

Для характеристики вибірки використовується показник, що залежить від її обсягу, який називається ***числом ступенів свободи***.

***Число ступенів свободи (v)*** - *це число даних з вибірки, значення яких можуть бути випадковими.*

Нехай ми маємо вибірку *х1, х2, ..., хn*. Загальною характеристикою вибірки є сума всіх її значень *х1 + х2 + ... + хn*. Тоді кожне окреме значення вибірки *хi* можна дізнатися, якщо від суми всіх значень відняти інші n-1 значення. Число n-1 і є числом ступенів свободи.

Наприклад, відомо, що два новонароджених важать в сумі 7,5 кг, а один з них 4 кг. Тоді вага другого вже визначено вагою першого (7,5-4=3,5), тобто має одну ступінь свободи (2-1=1). Якщо три дитини разом важать 10,5 кг, то вага одного завжди точно визначається вагою двох інших, між якими можливі варіації. В цьому випадку є два ступені свободи (3-1=2) і т.д.

Число ступенів свободи по-різному для залежних і незалежних вибірок:

- для ***залежних*** вибірок обсягом n число ступенів свободи визначається за формулою v = n-l;

- для **незалежних** вибірок обсягами n1 і n2 число ступенів свободи визначається за формулою v = (n1-1) + (n2-l) = n1-l + n2-l = n1 + n2-2.

При використанні багатовимірних методів аналізу даних (наприклад, дисперсійний аналіз) застосовують більш складні підрахунки числа ступенів свободи.

***4. Рівень статистичної значущості***

Питання про прийняття або відкиданні статистичної гіпотези не може бути вирішене зі стовідсотковою впевненістю. Завжди допускається ризик прийняття неправильного рішення. Мірою такого ризику є рівень значущості.

***Рівень значущості (р)*** - *це ймовірність помилки, яку ми допускаємо при прийнятті статистичної гіпотези.*

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 є кількісною оцінкою надійності зв'язку: чим менша ця вірогідність, тим надійніший зв'язок.

Припустимо, при порівнянні двох вибіркових середніх було отримано значення рівня статистичної значущості р=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.*

Історично склалося так, що при дослідженні психологічних явищ виділяють три рівні статистичної значущості критерію, який визначається співвідношенням його експериментального та критичного значення:

* нижчим рівнем статистичної значущості вважають 5%-й рівень (р≤0,05);
* достатнім рівнем статистичної значущості вважають 1%-й рівень (р≤0,01);
* вищим – 0,1%-й рівень (р≤0,001).

Тому в таблицях критичних значень вказують значення критеріїв, що відповідають рівням статистичної значущості р≤0,05 і р≤0,01, інколи – р≤0,001.

Рівень статистичної значущості 5% означає, що допускається 5 помилок у вибірці зі 100 елементів.

***5. Правила прийняття статистичного висновку***

* на основі отриманих експертних даних психолог підраховує за обраним ним статистичним методом так звану емпіричну статистику або емпіричні значення. Цю величину позначимо *Чемп.*
* емпіричні значення *Чемп* порівнюють з двома критичними величинами, які відповідають рівням значущості 5% і 1% для обраного статистичного методу і позначається *Чкр*.
* величини *Чкр* знаходять для даного статистичного методу за відповідними таблицями. Ці величини, як правило, різні і їх позначають *Чкр1* і *Чкр2*.



(Ч – від скороченого «число». Для кожного методу прийнято своє позначення цього числа. Наприклад, при використанні критерію Стьюдента коефіцієнт позначається як t).



* порівнюють емпіричне значення з двома знайденими в таблиці *tкр1* і *tкр*2. Для цього можна три числа розташувати на так званій «осі значущості» (в порядку зростання), на які виділено три зони:



* отримане число Ч (за одним із статистичних методів) обов’язково попаде в одну із зон:

1) Чемп належить зоні незначущості ⇒ приймається гіпотеза Но

2) Чемп належить зоні значущості ⇒ приймається гіпотеза Н1

3) Чемп належить зоні невизначеності ⇒ найкраще рішення: збільшити обсяг вибірки, існує ймовірність помилки.

4) Якщо Чемп = Чкр1 (або Чкр2 ), то можна віднести до зони незначущості (значущості).

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

* Помилка І роду.
* Помилка ІІ роду.

Розглянемо **помилку І роду**.

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

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

Інший варіант помилки – помилка ІІ роду.

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

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

Потужність критерію визначають емпіричним шляхом. Одна й та ж задача може бути вирішена різними критеріями, при цьому виявляється, що деякі критерії можуть виявити відмінності там, де інші їх не бачать.

**Приклад розв’язання задач:**

Нехай, що критичні значення критерія *Q*-Розенбаума відповідно дорівнюють 6 і 9та позначаються как *Q0,05*=6 и *Q0,01*=9. Прийнята наступна стандартна форма запису критичних значень:

****.

Припустимо, що емпіричне значення критерію дорівнює 11: *Qемп*=11.

Зона незначущості

Зона значущості

Зона невизначеності

0,01

0,05

11

9

6

Емпіричне значення критерії в даній задачі потрапляє в область між *Q0,05* и *Q0,01* і ми можемо вважати відмінності достовірними при р≤0,05