

ОСНОВНІ ПОНЯТТЯ СТАТИСТИЧНОЇ ТА МАТЕМАТИЧНОЇ ОБРОБКИ ДАНИХ ПСИХОЛОГІЧНИХ ДОСЛІДЖЕНЬ

ЛЕКЦІЯ 1

План

1. Загальне поняття статистики, її галузі.
2. Застосування методів математичної статистики у психолого-педагогічних дослідженнях.
3. Класифікація психологічних завдань, що вирішуються за допомогою статистичних методів.
4. Статистична значимість.
5. Шкали вимірювання.
6. Наочне зображення статистичного розподілу.
7. Представлення даних.

Проникнення математичних методів у будь-яку науку є прогресивним явищем, що зумовлено:

- ▶ математичні методи дають змогу зробити процес дослідження явищ більш чітким, структурованим та раціональним;
- ▶ математичні методи необхідні для обробки великої кількості емпіричних даних (їхніх кількісних виразників), для їх узагальнення та організації в «емпіричну картину» дослідження.

Статистика

ВІД ЛАТ. ВІД «status» –
ПОЛОЖЕННЯ, СТАН ЯВИЩ

- ▶ дані, які характеризують масові суспільні явища;
- ▶ процес збирання, зберігання і обробки даних про масові суспільні явища, тобто галузь практичної діяльності, спрямованої на одержання, обробку, аналіз та інтерпретацію даних про явища і процеси суспільного життя;
- ▶ наука, яка вивчає величину, розміри і кількісну сторону масових суспільних явищ у нерозривному зв'язку з якісною стороною цих явищ, з їх соціально-економічним змістом.

“

Математична статистика – це розділ математики, який вивчає математичні методи обробки й використання статистичних даних для наукових і практичних висновків ”



Статистика містить 3 розділи:

Описова

дозволяє описувати, впорядковувати, підсумовувати та представляти дані того чи іншого розподілу в більш наочному вигляді (таблиці, графіки), обчислювати середні значення, дисперсію та ін. представлено розподілу

Індуктивна

перевірка того, чи можна розповсюдити результати, отримані в окремій вибірці, на всю популяцію, з якої взята ця вибірка

Кореляційний аналіз

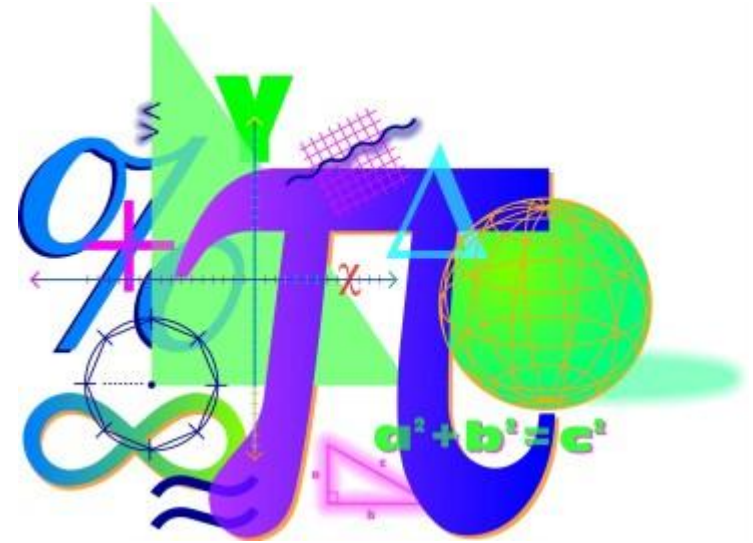
покликаний дізнатися, наскільки пов'язані між собою дві змінні. Це дозволяє прогнозувати можливі значення однієї з них, якщо ми знаємо іншу

Основні завдання математичної статистики

- ▶ оцінка ймовірності
- ▶ оцінка закону розподілу
- ▶ оцінка числових характеристик випадкової величини
- ▶ перевірка статистичних гіпотез

Методи статистичної обробки

це способи кількісних розрахунків, математичні формули і прийоми, які дозволяють узагальнювати емпіричні дані, виявляючи приховані в них закономірності



Методи статистичної обробки

Первинні

- ▶ методи, за допомогою яких отримують показники, що безпосередньо відображають результати емпіричних досліджень: наочне представлення даних у вигляді графіків і діаграм, обчислення заходів центральної тенденції (середнього значення, моди, медіани), заходів мінливості (розмаху, дисперсії, стандартного відхилення)

Вторинні

- ▶ методи статистичної обробки, які використовують первинні дані, дозволяють виявити приховані статистичні закономірності, зробити якісний аналіз даних: висування статистичних гіпотез, підготовка даних для застосування статистичних методів, перевірка гіпотез за допомогою статистичних критеріїв, формулювання висновків і т.д.

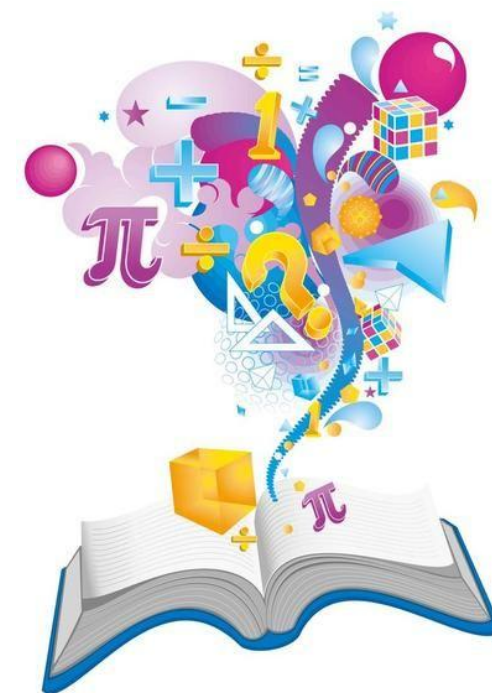
Використання математико-статистичного аналізу емпіричних даних дозволяє:

більш, чітко і лаконічно описувати досліджувані об'єкти, узагальнювати дані дослідження

виявляти наявність істотних відмінностей між групами, кількісно порівнюючи досліджувані ознаки

встановлювати приховані причини і суть психологічних явищ

підвищувати доказовість висновків, супроводжуючи їх статистичними підтвердженням



Виділяють **три** етапи процесу математизації психологічної науки:

I

- застосування математичних методів для аналізу і обробки результатів експериментів, спостережень і встановлення найпростіших кількісних закономірностей.

II

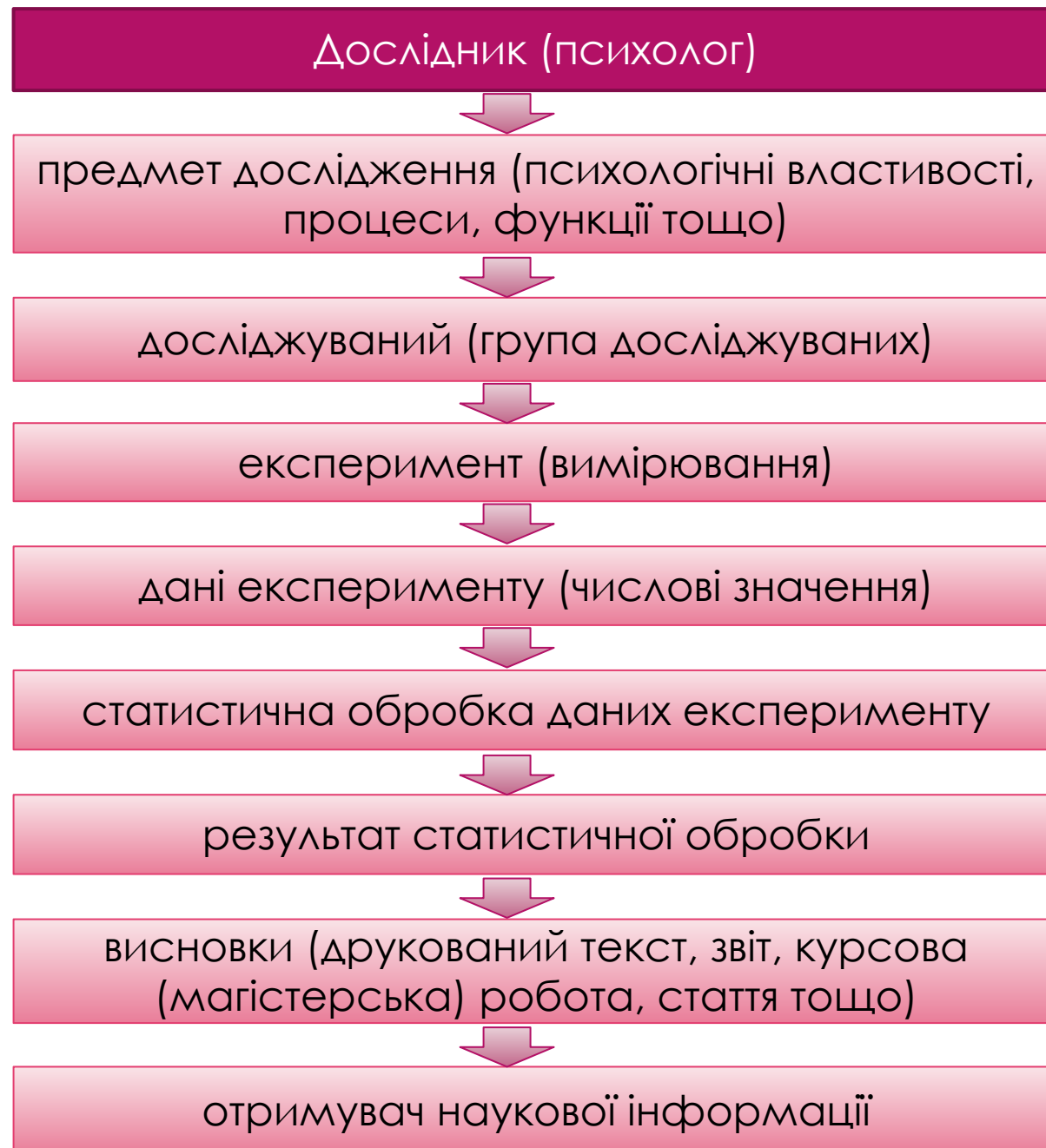
- полягає в спробі моделювання психічних процесів і явищ за допомогою готового математичного апарату, розробленого раніше для інших наук.

III

- характеризується розробкою спеціалізованого математичного апарату для дослідження і моделювання психічних процесів і явищ, формування математичної психології як самостійного розділу теоретичної психології.

Особливості застосування математичних методів обробки в психології :

- ▶ чим ближче до реальності експериментальні дані, тим надійніше результат математичного дослідження;
- ▶ при використанні математичних методів для аналізу і обробки результатів експериментів і спостережень більшу частину успіху дослідження становлять визначення типу розв'язуваної задачі і вибір методу вирішення;
- ▶ важливу частину рішення задачі займає інтерпретація отриманого результату.



Статистичний метод має складові:



масове спостереження

статистичне зведення

групування

обчислення середніх величин та індексів

побудова графіків, діаграм

Статистичне спостереження – це процес науково організованого планомірного збору даних

За часом реєстрації фактів

- ▣ поточне
- ▣ періодичне
- ▣ одночасне

За кількістю досліджуваних

- ▣ суцільне
- ▣ несуцільне
- ▣ вибіркове
- ▣ спостереження
основного масиву
- ▣ монографічне

За способами

- ▣ безпосереднє
- ▣ документальне
- ▣ опитування

Принципи математичного спостереження

Формулювання мети дослідження

- слід визначити мету дослідження, інакше буде зібрано багато непотрібної інформації, і мало – потрібної

Визначення об'єкта дослідження

- слід визначити, яке коло явищ досліджується і в якому аспекті.

Розробка програми дослідження

- слід визначити в якій послідовності і які факти вивчатимуться.

Змінна

В процесі проведення психологічного дослідження проводяться різні вимірювання. Як об'єкти вимірювання виступають психічні процеси, особливості властивості індивідів. Вимірювані психологічні явища позначаються поняттями «**ознака**» або «**змінна**». В якості змінної можуть виступати швидкість виконання тесту, соціометричний статус, кількість допущених помилок і т.д.



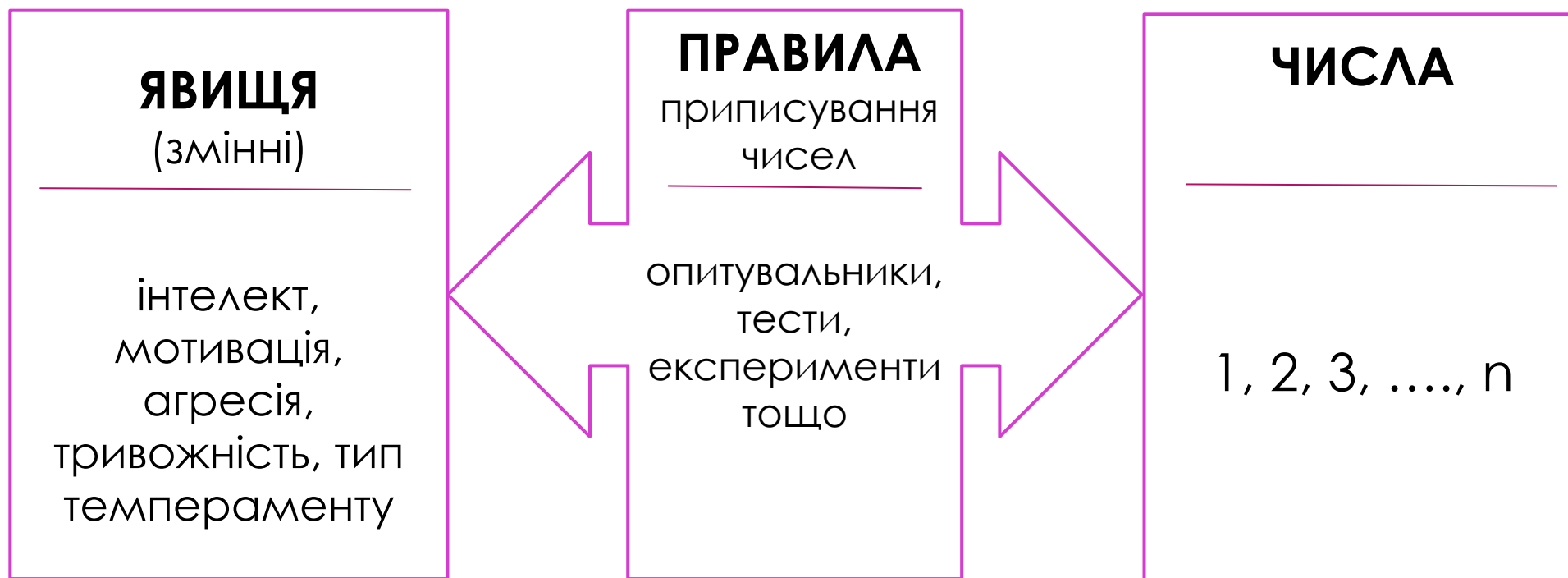
Змінна – це будь-яка реальність, яка може бути піддана виміру.

Психологічні змінні є випадковими величинами, оскільки заздалегідь невідомо, які значення вони приймуть.

У психології розглядають такі види змінних:

- ▶ кількісні та якісні змінні (дискретні і безперервні)
- ▶ незалежні і залежні змінні

Вимірювання



Види психологічного вимірювання

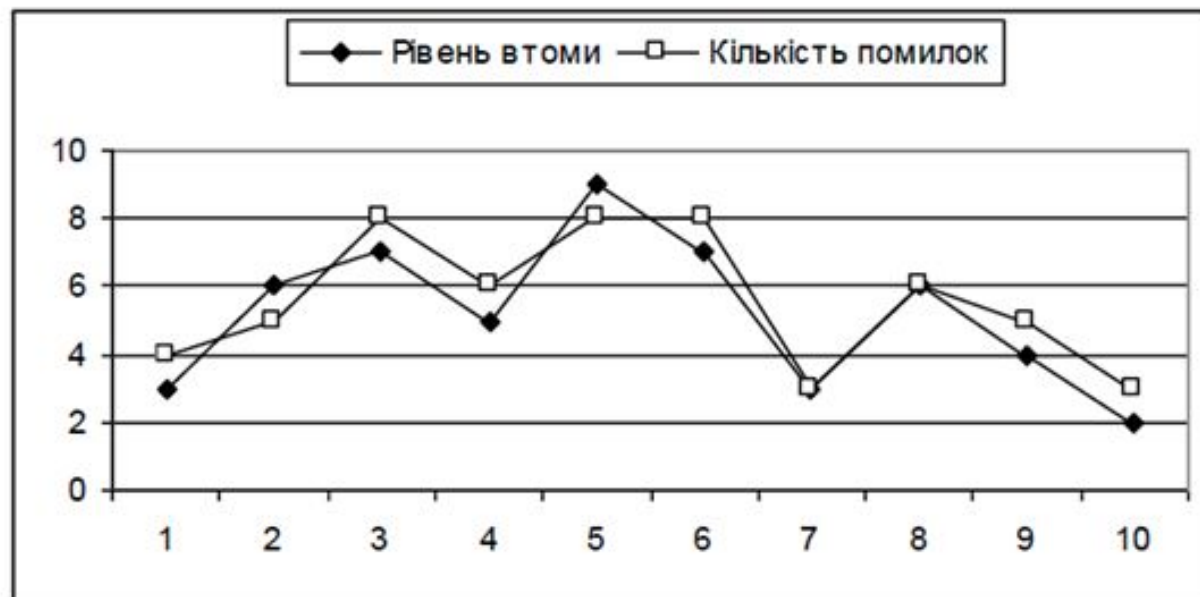
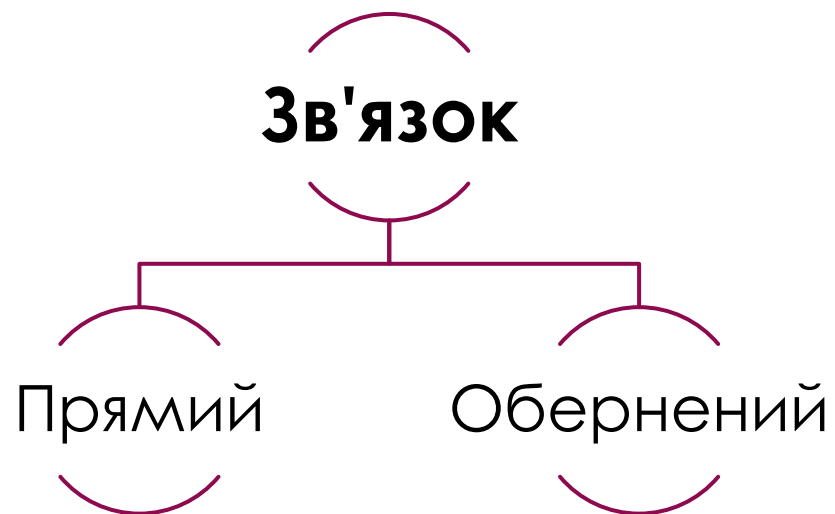
- Нормативне
- Критеріальне
- Іпсативне



Психологічні завдання, які вирішуються за допомогою методів математичної статистики

- ▶ завдання, що вимагають встановлення подібності або відмінності
- ▶ завдання, що вимагають угруповання і класифікації даних
- ▶ завдання, що ставлять за мету аналіз джерел варіативності одержуваних психологічних ознак
- ▶ завдання, які передбачають можливість прогнозу на основі наявних даних

Задача пошуку зв'язку



Зображення прямого зв'язку між рівнем в томи та кількістю помилок

Задача аналізу впливу



Якщо між явищами існує зв'язок, то чи означає це, що одне з них впливає на інше?

А якщо так, то що на що впливає?

Змінні

Незалежні

- ▶ явище, яке знаходиться під контролем дослідника, і яке він може змінювати відповідно до наперед визначеної експериментальної процедури

Залежні

- ▶ явище, яке змінюється під впливом незалежної змінної

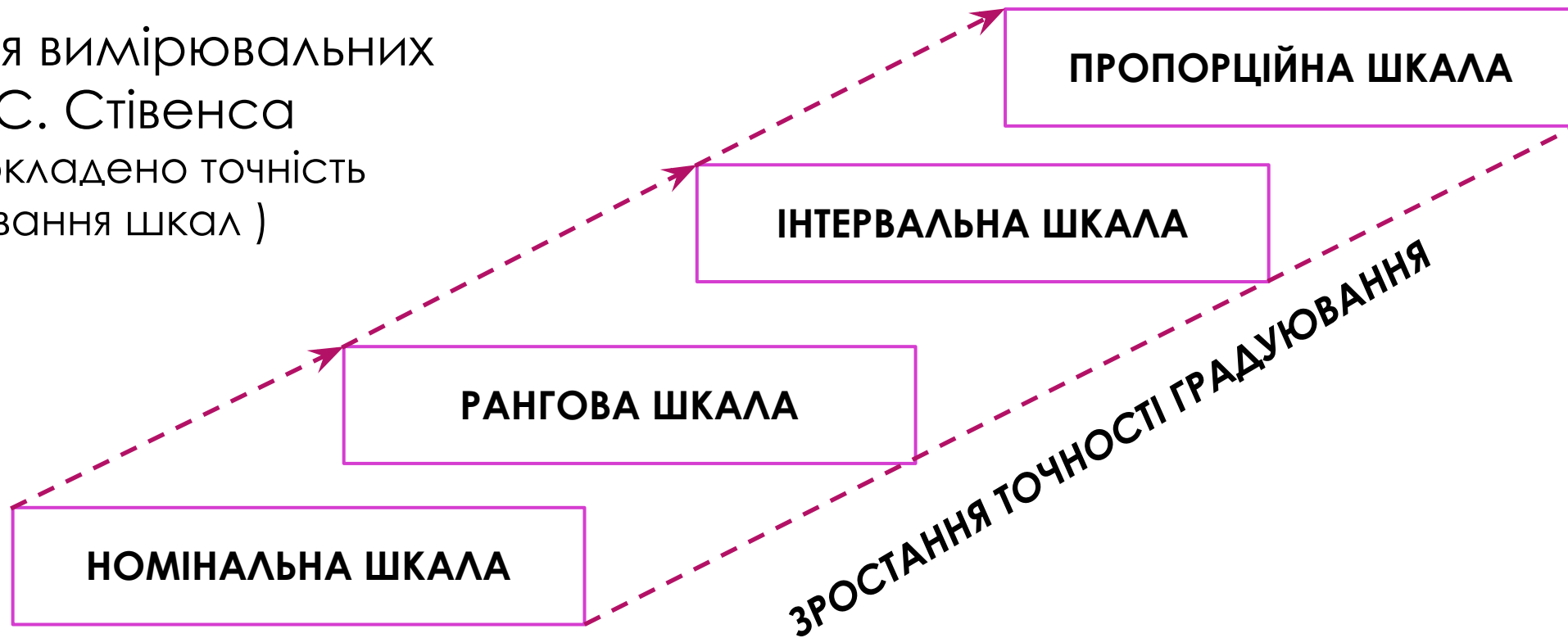


Статистична значимість

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

Вимірювальні шкали

Класифікація вимірювальних
шкал С. Стівенса
(в основу покладено точність
градуювання шкал)



Номінальна шкала

Шкала	Характеристика	Математичні операції	Приклад
Номінальна	<p>У межах цієї шкали об'єкти класифікуються, а класи позначаються номерами. Число тут слугує лише назвою певного класу, а тому нічого не говорить про властивості об'єкту, крім того, що він належить до певного класу.</p>	Встановлення рівності	Колір очей, номери гравців футбольної команди, типи темпераменту тощо

Рангова шкала

Шкала	Характеристика	Математичні операції	Приклад
Рангова	У межах цієї шкали об'єкти розташовуються в порядку спадання чи зростання у них певної якості. При цьому кожній градації якості приписується свій порядковий номер (ранг). Фактично, об'єкти лише впорядковуються.	Встановлення відношень «більше-менше»	Місця в олімпіаді, ранжування досліджуваних за проявами індивідуальних рис, рейтинг успішності студентів тощо

Інтервальна шкала

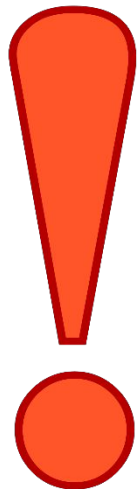
Шкала	Характеристика	Математичні операції	Приклад
Інтервальна	У межах цієї шкали існує одиниця вимірювання, за допомогою якої можна не лише впорядкувати об'єкти, але й приписати їм числа так, щоб однакові різниці між числами виражали однакові відмінності у проявах вимірюваної якості.	Встановлення рівності інтервалів (різниць)	Календарний час, шкала температур за Цельсієм тощо

Пропорційна шкала

Шкала	Характеристика	Математичні операції	Приклад
Пропорційна	У межах цієї шкали числа мають такі ж властивості, як і в шкалі інтервалів, але, крім того, відношення чисел виражають кількісні відношення ступенів прояву якості.	Встановлення відношень	Зріст, вага, рівень інтелекту, мотивація тощо

Наочне зображення статистичного розподілу

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



Статистичні таблиці – мають **підмет** і **присудок**.

Статистичний підмет – це та сукупність, про яку йдеться в таблиці. Як правило, розміщують у лівій частині таблиці.

Статистичний присудок – це ті ознаки або показники, які характеризують статистичний підмет. Присудок розміщують у заголовках стовпців.

Статистичні таблиці

Прості

- підмет задається переліком окремих об'єктів (назви підприємств, міст, країн тощо)

Складні

- в підметі одиниці сукупності групуються за однією якоюсь ознакою

Комбінаційні

- в підметі групуються одиниці за двома і більше ознаками, пов'язаними між собою.

Правила складання статистичних таблиць

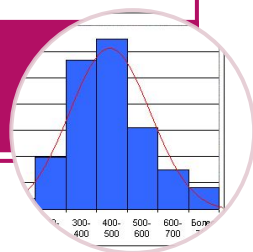
1. Таблиці мають бути наочними і виразними, їх не слід робити громіздкими та перевантажувати деталями. Іноді замість однієї загальної таблиці слід зробити декілька часткових.
2. Назва таблиці має розкривати її зміст у стислій і виразній формі. Назви рядків і стовпчиків пишуться коротко і без скорочень.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



Графіки

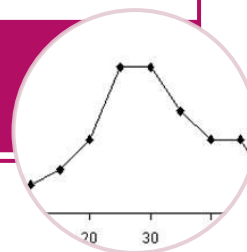
- це послідовність стовпців, кожний з яких спирається на один розрядний інтервал, а висота його відображає кількість випадків або частот у цьому розряді.

Гістограма



- використовується для графічного зображення дискретних та атрибутивних рядів розподілу. В полігоні кожний стовпчик закінчується точкою над серединою свого розрядного інтервалу на такій самій висоті

Полігон



- призначена для графічного подання рядів розподілу з накопиченими частотами. Це може бути стовпчата діаграма (для дискретного та атрибутивного рядів розподілу – лінійний графік)

Кумулята



Загальні правила побудови графіків

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

Ранжування даних

Нехай маємо деяку сукупність даних обсягом n . Впорядкуємо її за зростанням.

Припустимо, одержано таку послідовність даних

$$X_1, X_2, X_3, \dots, X_{n-1}, X_n$$

Можливі 2 випадки: коли всі значення в сукупності різні і коли деякі значення повторюються

Ідеальний випадок і випадок з повторенням

Ідеальний

- ▶ У ситуації, коли всі значення сукупності різні, найменшому значенню (елементу) X_1 призначимо ранг 1, наступному за розміром елементу X_2 - ранг 2 і так до кінця. Найбільший елемент X_n матиме ранг n .

Повторення

- ▶ У цій ситуації спочатку призначаємо кожному елементу ранги так, ніби повторень немає. Назвемо ці ранги попередніми.

Впорядковані значення	X_1	X_2	X_3	...	X_{n-1}	X_n
Попередні ранги	1	2	3	...	$n-1$	n

Якщо значення не має повторень, то його попередній ранг і є рангом. Для однакових елементів попередні ранги потрібно модифікувати.

Розглянемо деяку групу однакових значень (нехай решта даних у сукупності мають інші значення) з їх попередніми рангами.)

Група однакових значень	X_i	X_{i+1}	...	X_{i+k}
Попередні ранги	i	$i+1$...	$i+k$

Усім елементам цієї групи призначимо однаковий ранг, який дорівнює середньому значенню їх попередніх рангів:

$$\frac{i + (i + k)}{2}$$

Ранжування

Процедура ранжування не складна, а скоріше механічна. Проте дослідники, виконуючи ранжування вручну, часто помиляються.

Для контролю правильності ранжування прийнято перевіряти таку властивість: якщо сукупність має обсяг n , то загальна сума рангів її елементів дорівнює:

$$\frac{n(n + k)}{2}$$

Приклад ранжування даних

Психолог за допомогою тесту оцінив рівень роздратованості у групі вчителів початкових класів за 7-бальною шкалою.

Результати дослідження:

Номер учителя у списку	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Рівень роздратованості	1	5	2	5	3	3	7	1	2	2	2	4	5	1

Проранжуємо цей список. Для цього запишемо послідовність рівнів роздратованості в порядку зростання

1, 1, 1, 2, 2, 2, 2, 3, 3, 5, 5, 5, 7

Призначимо цим значенням попередні ранги

Приклад ранжування даних

Призначимо цим значенням попередні ранги

Рівень роздратованості	1	1	1	2	2	2	2	3	3	4	5	5	5	7
Попередній ранг	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Після цього підправимо ранги в кожній підгрупі однакових значень. Для підгрупи елементів з рівнем 1 маємо попередні ранги 1, 2, 3. Тому ранг елементів підгрупи дорівнює

$$\frac{1 + 3}{2} = 2$$

Приклад ранжування даних

Рівень роздратованості	1	1	1	2	2	2	2	3	3	4	5	5	5	7
Попередній ранг	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Попередні ранги підгрупи з рівнем роздратованості 2 дорівнюють 4, 5, 6 та 7. Крайнім позиціям відповідають ранги 4 і 7. Отже спільний ранг підгрупи дорівнює

$$\frac{4 + 7}{2} = 5,5$$

Приклад ранжування даних

Рівень роздратованості	1	1	1	2	2	2	2	3	3	4	5	5	5	7
Попередній ранг	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Аналогічно знаходимо ранги для підгруп з рівнями 3 і 5

$$\frac{8 + 9}{2} = 8,5$$

$$\frac{11 + 13}{2} = 12$$

Приклад ранжування даних

Оскільки ранги 4 і 7 мають лише по одному індивідууму, їх ранг збігаються з попередніми рангами відповідно 10 і 14.

Запишімо результати ранжування:

Рівень роздратованості	1	2	3	4	5	7
Ранг	2	5,5	8,5	10	12	14

Приклад ранжування даних

- ▶ Перевіримо правильність ранжування, враховуючи кількість повторень рангів.

$$\text{Сума рангів} = 2 \cdot 3 + 5,5 \cdot 4 + 8,5 \cdot 2 + 10 + 10 \cdot 3 + 14 = 6 + 22 + 17 + 10 + 36 + 14 = 105$$

$$\frac{n(n+k)}{2} = \frac{14(14+1)}{2} = 7 \cdot 15 = 105$$

Оскільки контрольна рівність

$$\text{Сума рангів} = \frac{n(n+1)}{2}$$

Приклад ранжування даних

Остаточні результати ранжування зведено в таблицю

Номер учителя у списку	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Рівень роздратованості	1	5	2	5	3	3	7	1	2	2	2	4	5	1
Ранг	2	12	5,5	12	8,5	8,5	14	2	5,5	5,5	5,5	10	12	2