

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

МАТЕМАТИЧНІ МЕТОДИ В СОЦІОЛОГІЇ

Контент лекцій
для студентів спеціальності 054 «Соціологія»

Харків-2023

Тема 1. Основні принципи вимірювання. Графічне відображення соціологічної інформації

Предмет, завдання і структура курсу. Поняття вимірювання в соціології. Методологічні основи вимірювань в соціології. Методи вимірювання, призмінювані в ході соціологічних досліджень. Рівні і шкали вимірювання: номінальна, порядкова, полупорядкова, інтервальна і шкала відносин; їх особливості. Можливості вимірювання для кожної з шкал. Види помилок вимірювання. Типи змінних: кількісні, якісні, дискретні, безперервні змінні. Частота, відносна частота, частість.

Типи перехресної класифікації соціологічної інформації і їх інтерпретація. Особливості побудови та аналізу динамічних рядів. Заходи вимірювання соціологічних даних: відносні і абсолютні. Кумулятивні динамічні ряди. Можливості використання співпадаючих тенденцій розвитку соціальних процесів в ході математико-статистичного аналізу. Призначення графіка. Загальні принципи побудови графіків (гістограма, полігон, кумулята). Відображення двох, трьох і більше полігонів на одному графіку. Можливості «EXCEL» при побудові різних видів графіків. Нерівні інтервали угруповання. Графіки якісних даних: діаграма смуг; кругова діаграма - гартовская діаграма; статистична карта; тимчасові діаграми; мно-гозначний графік.

Весь табличний матеріал можна представити в графічній формі, яка (звичайно) більш наочно, ніж таблиця, виражає картину загального розподілу. Таблиця частот дозволяє з більшою легкістю інтерпретувати результати. Головне призначення графіка – дати найбільш точне уявлення про форму частотного розподілу – уявлення, зрозуміле навіть абсолютно некваліфікованому читачеві.

Існує безліч видів графічного представлення, кожне з яких корисне в своєму конкретному додатку. Деякі з них вельми складні, але тут розглядаються тільки основні і найпростіші: 1) гістограма, 2) кумулята, 3) полігон розподілу, 4) діаграма смуг, 5) статистична карта, 6) часова діаграма.

Тема 2. Середні величини узагальнення характеристик соціологічної інформації

Основні числові характеристики аналізу одновимірного розподілу: максимум; мінімум; особливості вибору поняття середнього. Середнє арифметичне: визначення і правила його обчислення. Обчислення середньої для згрупованих і негрупірованих даних. Зважування середнього. Застосування середнього.

Принцип порядкового розташування. Обчислення медіани для згрупованих і негрупірованих даних. Медіана дискретних даних. Квантілі і інші заходи усереднення в якості нормують критеріїв. Мода чи розподіл усіх середнє, правила обчислення моди. Оцінка моди і бимодальність. Значення моди і медіани в соціологічних дослідженнях. Працює з середніх. Вибір середнього в залежності від мети усереднення, виду розподілу і технічних міркувань особливостей обчислення того чи іншого середнього. Мінімум, максимум і проміжні заходи як допоміжні характеристики аналізу даних. Порівняльні характеристики середніх.

Мода - що найчастіше спостерігається величина в розподілі; точка найбільшої щільності; значення моди визначається переважаючою частотою, а не значеннями змінної в розподілі.

Медіана – це величина, розташована точно в середній точці множини (а не в області зміни змінною); половина подій має значення більше її, а половина – менше. Значення медіани визначається її розташуванням у множині даних і не залежить від значення окремих величин.

Середнє арифметичне - це така величина в даній сукупності, яка спостерігалася б в тому випадку, якби всі величини були рівні; суми відхилень від середнього в будь-яку сторону рівні; отже, алгебраїчна сума відхилень рівна нулю.

Середнє можна обчислити навіть в тому випадку, коли значення окремих величин невідомі при умові, що відома сума всіх величин і N .

Тема 3. Показники змін значень ознак. Варіація якісних змінних

Дисперсія і середнє квадратичне відхилення. Заходи протяжності, проміжні діапазони і вимір розмаху варіації. Відхилення від середнього як міра варіації. Вибір норми і побудова заходи варіації. Середнє лінійне відхилення (). Коефіцієнт варіації. Квадратичні відхилення як міра варіації. Обчислення середнього квадратичного відхилення (). Характеристики середнього квадратичного відхилення. Варіація якісних змінних. Коефіцієнт якісної варіації. Елементарне нормування. Інші коефіцієнти якісної варіації: процентні відносини, пропорції, ступеня (коефіцієнти), індекс. Нормировка допомогою подкласифікації.

Протягом майже сторіччя σ була одним із найбільш поширених засобів статистичного аналізу не тільки як міра дисперсії, але і як складова частина більш складних обчислень. Широке використання сигми в якійсь мірі пояснюється двома властивими їй достоїнствами. По-перше, σ двічі відображає величину кожної змінної розподілу: а) точка відліку, від якої заміряються відхилення (\bar{X}) сама є репрезентацією всіх змінних; б) кожна величина як така представлена квадратичним відхиленням. По-друге, зведення відхилень в квадрат автоматично знижує проблему знаку відхилення.

Оскільки d_{cp} вимірює відхилення без зведення в квадрат, то воно по абсолютній величині менше, ніж σ , яке непропорційно збільшує великі відхилення внаслідок зведення в квадрат.

Варіацію якісних змінних не можна вимірювати тим же способом, який був розроблений для кількісних. У якісних змінних не існує «нуля» відліку, і, отже, вони не мають величини. Не існує середнього значення діапазону і проміжних інтервалів. Отже, не існує і арифметичних відхилень.

Число відмінностей, як показник варіації, порівнянне тільки з максимальним можливим числом відмінностей. Це максимальне число відмінностей буде спостерігатися в тому випадку, коли всі частоти різних ознак рівні.

Тема 4. Нормальний розподіл як модель варіації. Критерії лінійного взаємозв'язку

Поняття нормального частотного розподілу. Закон розподілу. Приклади соціальних явищ, описуваних нормальним, біноміальним і джена пуассонівська розподілом. Параметрична і непараметрическая статистика. Особливості емпіричних розподілів (асиметрія). Критерій лінійної взаємозв'язку. Інші заходи характеристик одновимірного розподілу. Як правило, емпіричні розподіли порівнюють з деяким еталоном, який називається ідеальним розподілом. Це порівняння необхідне для:

1) можливості спрогнозувати подальшу поведінку і розвиток того або іншого феномена, у випадку якщо відмінності між емпіричними і теоретичними розподілами невелики. 2) виявлення причин, що впливають на вияв відмінностей між теоретичними і емпіричними розподілами, якщо такі спостерігаються.

До ідеальних розподілів можна віднести: а) закон нормального розподілу. Адекватне застосування кількісних методів, що вийшли в практику соціологічних досліджень, в тій або іншій мірі спирається на припущення, що ознака (або сукупність ознак), що вивчається підкоряється певному статистичному закону розподілу. Таким що найчастіше зустрічається розподілом є нормальний або гаусовський закон розподілу. Нормальний розподіл ознаки спостерігається в тих випадках, коли на величину його значень впливає безліч випадкових незалежних або слабо залежних чинників, кожний з яких грає в загальній сумі приблизно однакову і малу роль (т. е. відсутні домінуючі чинники).

По визначенню, нормальна крива складається з нескінченного числа точок, унімодальна, симетрична і неограничена в обох напрямках. Отже, мода, медіана і середнє рівні по величині і ділять розподіл на дві рівні частини. Графічне зображення цього розподілу являє собою рівну, колоколообразную криву з характерним крилом, яке ніде не торкається базової лінії.

Тема 5. Заходи взаємозв'язку для інтервального рівня виміру.

Кореляційний аналіз. Рівняння регресії

Особливості сприйняття принципу взаємної спряженості. Види функціональної залежності соціальних ознак. Способи вимірювання зв'язку між різною кількістю соціальних ознак. Кореляційне поле як форма графічного представлення кореляційної залежності. Види кореляційного розсіювання. Скедастичність (варіабельність). Кореляційна таблиця як форма подання згрупованих даних. Техніка групування. Функція кореляційної таблиці.

Групи зображених на кореляційному полі точок навколо гіпотетичної лінії наводять на думку про певний «закон зв'язку» між двома змінними. Більш того спостерігаючи ширину розброса, можна зробити принаймні попередній висновок про те, наскільки добре події відповідають цьому гіпотетичному закону. Отже, їх не можна як-небудь описати або зв'язати з чим би те не було, не відтворюючи кореляційного поля. А оскільки ці «візуальні критерії» не точні в математичному значенні, їх зіставлення неможливе, навіть якщо в розпорядженні дослідника є всі кореляційні поля у всій їх складності.

Необхідність загальної міри кореляції. Відхилення від середнього арифметичного. Що пояснюється відхилення і необ'яснюєміе залишкові відхилення. Вимірювання лінійної кореляції. Обчислення нахилу лінії регресії. Коефіцієнт детермінації. Порівняння і. Розрахункові формули: негрупуваних і згруповані дані. Призначення кореляційної таблиці. Коефіцієнт кореляції як міра тісноти, типу і напрямку зв'язку між двома ознаками. Рівняння регресії.

Прямий зв'язок означає, що при збільшенні значення однієї ознаки в середньому збільшується значення іншого. Зворотний зв'язок - що при збільшенні однієї ознаки в середньому меншає значення іншого.

За формою кореляційний зв'язок може бути прямолінійним або криволінійним. Гіпотезу про форму зв'язку встановлюють по кореляційному полю.

Тема 6. Корреляційне відношення. Нелінійна регресія. Множинна і приватна кореляції

Всякий індекс кореляції є мірою того, у якому ступені одну перемінну можна пророчити чи пояснити через іншу. З багатьох уживаних у статистику показників індекс r Пірсона є однією з найбільш розповсюджених, майже банальних мір. Однак застосовність r ґрунтується в основному на наступних двох умовах: 1) відношення між перемінними лінійно; 2) двовимірний розподіл гомоскедастично, тобто має постійну умовну дисперсію. Узагалі говорячи, обидві ці умови виконуються, якщо маргінальні розподіли нормальні. Тому одним з вимог, пропонованих до критерію r , часто є нормальність маргінальних розподілів.

Однак дуже часто вихідні дані приводять до нелінійних моделей. Це особливо справедливо для соціологічних наук, де багато розподілів, такі як дохід чи розмір родини, сильно скошені, і лінії регресії тому значною мірою нелінійні.

Порівняння статистичних показників r^2 і η^2 :

$r^2 = 0$, якщо x і y незалежні.

$r^2 = \eta_{x/y}^2 = 1$, тоді і тільки тоді, коли **мається** стругаючи лінійний зв'язок між x і y .

$r^2 \leq \eta_{x/y}^2$, тоді і тільки тоді, коли **мається** стругаючи нелінійна функціональна залежність x і y .

$r^2 = \eta_{x/y}^2 < 1$, тоді і тільки тоді, коли регресія x і y строго лінійна, але немає **функціональної** залежності.

$r^2 < \eta_{x/y}^2 < 1$ означає, що немає **функціональної** залежності й існує нелінійна крива регресії.

Тема 7. Кореляція рангів: коефіцієнт. Коефіцієнт взаємної спряженості

Вимірювання взаємозв'язку ознак за допомогою рангів. Метод кореляції рангів () □ коефіцієнт Спірмена. Техніка обчислення коефіцієнта Спірмена в разі об'єднаних рангів. Кореляція між впорядкованими змінними. Аналіз і корисність □ інтерпретація коефіцієнтів рангової кореляції. Міра відповідності для трьох і більше рангових рядів □ коефіцієнт множинної кореляції ознак для порядкового рівня вимірювання.

Ранги іноді можуть поєднуватися. Експерт у складному становищі може оцінити дві картини однаково, кількісні значення і міри можуть також бути однаковими. У подібних випадках два чи більш спостереження можуть, очевидно, одночасно претендувати на той самий ранг. Тому що число рангів і число спостережень повинне збігатися, те просто неможливо двом спостереженням привласнити той самий ранг, тому їм повинні бути привласнені об'єднані ранги. У таких випадках обом спостереженням приписується значення середнього арифметичних із двох об'єднаних рангів. Особливості обчислення коефіцієнта множинної кореляції для номінального рівня вимірювання. Необхідність і функції параметра

Хі-квадрат. Механізм обчислення очікуваних частот. Коефіцієнт взаємної спряженості: основні характеристики і принципи обчислення.

Соціологічний та статистичний сенс кореляції: випадковий зв'язок; безглузда кореляція; кореляція як доказ наявності причини і наслідки; кореляція як міра загальних факторів; помилкова кореляція. Причини використання великого числа критеріїв кореляції: вид залежності; форма представлення даних; необхідний ступінь точності; оборотність індексів. «Пастки» при інтерпретації кореляції: зсув у часі; заходи кореляції, що застосовуються до неоднорідним даними; визначеність індексу; кореляція між синоптичними заходами. Загальні соціологічні міркування інтерпретації кореляції.

Тема 8. Надійність вимірювання соціальної інформації

Основні показники поняття надійності. Компоненти надійного вимірювання. Правильність вимірювання виявлення систематичних помилок. Стійкість вимірювання: показник абсолютної стійкості W , середня квадратична помилка, відносні показники помилок. Обґрунтованість вимірювання як завершальний етап підтвердження надійності вимірювання.

Вибіркове розподіл. Розподіл помилок вибірки. Побудова довірчого інтервалу (велика вибірка). Деякі актуальні проблеми вибірки. Нестабільність генеральної сукупності. Гетерогенність соціального універсуму. Перевірка статистичних гіпотез і порівняти оцінки. Принцип перевірки нуль-гіпотези.

Тому що соціальна інформація є метричною величиною, те необхідно з'ясувати надійність проведеного виміру якого-небудь соціального процесу. До надійності відносять три поняття: правильність; стійкість; обґрунтованість. Відповівши на ці три питання: про правильність проведеного дослідження, його стійкості й обґрунтованості отриманих результатів – можна стверджувати, що вимір надійний.

Правильність установлює загальну прийнятність даного способу виміру і виявляється у виявленні систематичних помилок. Систематичні помилки – це помилки, що виявляються постійно чи відповідно до визначеного закону.

Стійкість характеризує збіг результатів виміру при повторних застосуваннях вимірювальної процедури. Описується випадковими помилками.

Обґрунтованість – зв'язана з доказом того, що був обмірюваний саме той об'єкт і та властивість, що і були потрібні, – є самою важкою процедурою, тому що правильність і стійкість визначаються за допомогою математичних формул, а обґрунтованість доводиться логічним шляхом. При доказі обґрунтованості необхідно точно визначити: 1) об'єкт, 2) засоби, 3) суб'єкт виміру. Перевірку надійності необхідно почати з перевірки правильності і стійкості, якщо вони підтверджуються, те підтверджується обґрунтованість.