Примітки:
1. Для побудови шкали даного типу, крім операції, зазначеної у відповідному рядку, повинні бути емпірично реалізовані операції всіх попередніх типів шкал.
2. Групи перетворень чисел для шкал даного типу входять у групи перетворень шкал усіх попередніх типів, але не навпаки.
3. Для шкал даного типу можна обґрунтовано застосовувати статистичні міри шкал усіх попередніх типів, але не можна застосовувати міри шкал наступних типів.

Дані положення відображено на схемі за допомогою відповідних стрілок.

Наведемо тепер таблицю коефіцієнтів зв'язку для ознак, вимірюваних за допомогою шкал різних рівнів (з огляду на те, що для інтервальних шкал і шкал відносин використовуються ті самі міри).

У клітинках таблиці 40, що представляють перетинання рядків (рівень виміру ознаки X) і стовпчиків (ознаки Y), наводяться відповідні статистичні міри зв'язку. Так, на перетинанні другого рядка (ознака X – порядкова) і третього стовпчика (ознака Y – метрична) зазначено міри зв'язку номінальних і порядкових (за допомогою стрілок), порядкових і порядкових шкал, а також метричних і порядкових (η_{yx}), тому що Y – метрична. Зазначимо, що використовувати η_{xy} в цій ситуації не можна, тому що ознака X не є метричною.

Таблиця 40. **Рівні вимірювання і міри зв'язку між ознаками**

Тип шкали X	Тип шкали Y		
	Номінальна	Порядкова	Метрична (інтервальна і шкала відносин)
Номінальна	$\Phi\ Q\ C\ T\ T_c$ $g_{yx}\ \ g_{xy}\ \ \delta\ \ \Delta$ $\lambda_{yx}\ \ \lambda_{xy}$ →	r_{pb} →	$r_{pb}\ \ \eta_{yx}$
Порядкова	→	$\tau\ \gamma$ $d_{xy}\, d_{yx}$ →	η_{yx}
Метрична	η_{xy}	η_{xy}	$r\ \rho\ R$

СТАТИСТИЧНІ ВИСНОВКИ: ОЦІНЮВАННЯ ТА ПЕРЕВІРКА ГІПОТЕЗ

Розділ V

Генеральна та вибіркова сукупність. Оцінка похибки вибірки 1

Об'єктом соціологічних досліджень зазвичай є різні соціальні спільноти. Якщо вивчаються всі індивіди даної сукупності, то йдеться про суцільне дослідження, якщо ж тільки частина, то – про вибіркове. Як правило, соціологічні дослідження мають вибірковий характер.

Це пов'язано передусім з тим, що економічні й часові обмеження не дозволяють провести суцільне дослідження (витрати на проведення Першого всеукраїнського перепису населення 2001 року складали понад 146 млн грн і вимагали залучення понад 250 тис. обліковців). Але навіть у тих випадках, коли суцільне дослідження практично може бути здійснене, найчастіше більш рентабельно проводити вибіркове. Його економічність дозволяє збільшити витрати на удосконалення інструменту дослідження і компенсувати тим самим падіння надійності за рахунок того, що дослідження не суцільне, а вибіркове (похибки вибірки дуже часто набагато нижчі, ніж похибки за рахунок недосконалого інструменту дослідження). У підсумку дослідник має можливість отримати більш повну і надійну інформацію.

Підстави, які дозволяють нам за вивченням частини судити про ціле, пов'язані з деякими ймовірнісними законами[1]. Якщо, наприклад, виймати з урни[2], в якій знаходяться добре перемішані білі й чорні камінці (50 білих і 50 чорних), 20 камінців, то імовірність того, що нам випадуть усі чорні камінці, дуже мала (імовірність того, що перший камінчик буде чорним,– $\frac{50}{100}$, що другий – $\frac{49}{99}$, оскільки в урні залишилось усього 99 камінців, з них 49 чорних; тоді ймовірність, що перші два – чорні, дорівнює $\frac{50}{100} \cdot \frac{49}{99}$. Продов-

[1] Див. додаток 1 («Про ймовірності»).
[2] Приклади з урною широко використовуються в теорії ймовірностей і беруть свій початок, мабуть, від прийнятої в Стародавній Греції процедури голосування.

жуючи ці міркування, отримуємо, що ймовірність того, що всі 20 камінців чорні, дорівнює $\frac{50 \cdot 49 \ldots 31}{100 \cdot 99 \ldots 81} = 0,00000009$). Імовірність того, що одних камінців буде набагато більше, ніж інших, також мала. Найчастіше буде зустрічатися така ситуація, коли число чорних камінців приблизно дорівнюватиме числу білих. Аналогічно, якщо в місті половина населення має одні ціннісні орієнтації, а половина – інші, то малоймовірно у вибірковому дослідженні (при випадковому відборі[1]) отримати, що осіб з одними ціннісними орієнтаціями набагато більше, ніж з іншими.

Однак здійснити випадковий відбір дуже важко. Навіть у випадку з камінцями потрібно забезпечити ретельне перемішування, однакові розміри камінців і гарантію, що той, хто витягує, не бачить кольору камінця (прикладом ідеально організованого випадкового відбору є тиражі спортлото). Експерименти, що були проведені з відбором камінців одного кольору, які лежали на столі, показали, що учасники випробування мимоволі обирають більш крупні камінці – вони, напевно, частіше потрапляють під руку[2]. Незрівнянно важче забезпечити випадковий відбір у соціологічних дослідженнях. Широко відомий приклад невдалого прогнозу результатів виборів президента США в 1936 р.: журнал «Літерарі Дайджест» за телефонними книгами відібрав понад два мільйони адресатів, отримавши тим самим, здавалося б, випадкову вибірку. За адресами були розіслані листівки з проханням відповісти – Рузвельту чи Ландону віддасть свій голос респондент. За результатами опитування журнал пророчив перемогу Ландона з великою перевагою. Цікаво, що соціологи Дж. Гелап и Ел. Роупер правильно передбачали перемогу Рузвельта, базуючись на аналізі в 500 разів меншого масиву – чотирьох тисяч анкет. Похибка в прогнозі «Літерарі Дайджест» пояснюється тим, що вибірка за телефонними книгами не була випадковою, вона не забезпечувала рівну імовірність потрапити у вибірку для всіх осіб, що мають виборче право, оскільки в 1936 р. телефони були переважно у забезпечених верств населення, які надавали перевагу Ландону.

Властивість вибірки відображати характеристики досліджуваної сукупності називається *репрезентативністю.* Іноді замість вибірки говорять *вибіркова* сукупність (*sample*), а досліджувану сукупність

[1] Випадковим відбором називають такий відбір, при якому всі елементи досліджуваної сукупності мають рівну імовірність потрапити у вибірку.
[2] *Нейтс Фрэнк.* Выборочный метод в переписях и обследованиях.– М., 1965.– С. 31.

називають *генеральною (population).* Можна сказати, що генеральна сукупність – це та, на яку дослідник має намір поширювати висновки, зроблені при вивченні вибірки. Відмінність характеристик вибіркової і генеральної сукупності називають похибкою репрезентативності. Можна виділити два види таких похибок – систематичні та випадкові.

Систематичні похибки – це похибки такого типу, як ті, що були допущені журналом «Літерарі Дайджест», тобто якесь постійне зміщення, яке не зменшується при збільшенні кількості опитаних (якщо б журнал опитав не два, а чотири мільйони власників телефонів, це не врятувало б його від похибки).

Випадкові похибки – це ті, які при повторних вимірюваннях змінюються за імовірнісними законами. Зокрема, якщо ми визначаємо якусь характеристику вибірки, наприклад, середнє арифметичне, то, вибираючи все нові й нові вибірки того ж розміру, будемо отримувати, що ця характеристика відхиляється то в один, то в інший бік від істинного значення (тобто від значення в генсукупності) приблизно з однаковою частотою і при збільшенні числа вибірок середня арифметична похибка наближається до нуля. Систематичну похибку можна усунути, змінюючи процедуру формування вибірки; випадкова похибка буде присутня завжди, при будь-якому вибірковому опитуванні. Проте систематична похибка значно небезпечніша, оскільки її неможливо оцінити за процедурою побудови вибірки. Випадкова ж похибка підпорядковується певним законам і піддається оцінці. Взагалі репрезентативність вибірки характеризується двома взаємопов'язаними параметрами – рівнем похибки й імовірністю. Вести мову про якусь вибірку, що вона репрезентативна, не зовсім точно, бо будь-яка вибірка має певний рівень репрезентативності (хоча цей рівень може нас зовсім не влаштовувати). Точніше кажучи, похибка репрезентативності даної вибірки з імовірністю P не перевищує Δ (імовірність P називають довірчою).

Для випадкової вибірки існують методи, що дозволяють оцінити цю похибку (ми розглянемо їх при викладенні способів перевірки гіпотез). Плануючи соціологічне дослідження, зазвичай вирішують інше завдання – задаються якимсь, що влаштовує дослідника, рівнем точності результату, тобто припустимою похибкою і довірчою імовірністю, і визначають для цих параметрів необхідний обсяг вибірки. Зокрема, обсяг вибірки для визначення частки деякої ознаки X в генсукупності визначається формулою[1]:

[1] *Кокрен У.* Методы выборочного исследования.– М., 1976.– С. 89.

$$n = \frac{1}{\frac{\Delta^2}{t^2 \nu(1-\nu)} + \frac{1}{N}}, \qquad \text{(V.1.1)}$$

де N – обсяг генеральної сукупності, n – обсяг вибірки, t – коефіцієнт, що відповідає довірчій імовірності (див. табл. I додатка 3; якщо $n > 60$, то при $P = 0{,}954$ $t = 2$, а при $P = 0{,}997$ $t = 3$ і т. д.), ν – частка ознаки X в генсукупності, Δ – величина припустимої похибки (в частках).

Якщо, наприклад, дослідник хоче отримати з імовірністю 0,95 ($t = 2$) дані про частку ознаки X в генсукупності (нехай $N = 10\,000$) з похибкою, що не перевищує 5 % ($\Delta = 0{,}05$), і йому відомо, що шукана частка складає приблизно 20 % ($\nu = 0{,}20$), то за формулою (V.1.1) отримаємо, що потрібно опитати 256 людей ($n = 256$).

Незручність користування формулою полягає в тому, що вона потребує хоча б приблизної інформації про частку ознаки в генеральній сукупності, тобто якраз про те, що дослідникові потрібно визначити. Щоб позбутися цієї незручності, зазначимо, що при $\nu = 0{,}5$ добуток $\nu\,(1 - \nu)$ максимальний, отже, n також максимальне. Тому якщо в (V.1.1) замість ν підставити 0,5, отримаємо формулу, якою можна користатися за будь-яких значень частки ознаки в генеральній сукупності (обсяг вибірки при цьому буде виходити з деяким запасом). Поклавши також значення довірчої імовірності рівним 0,954, тобто $t = 2$, отримаємо

$$n = \frac{1}{\Delta^2 + \frac{1}{N}}. \qquad \text{(V.1.1')}$$

Скориставшись цією формулою, визначимо, як об'єм вибірки залежить від об'єму генеральної сукупності і від величини припустимої в дослідженні похибки.

З табл. 41 видно, що для забезпечення заданої репрезентативності при дослідженні міста з населенням 100 тис. жителів потрібно опитати 398 чоловік, а при дослідженні усієї країни практично стільки ж – 400 чол. Для забезпечення одного й того ж рівня репрезентативності (5 %) потрібно опитати такі частки генсукупності: для $N = 500$ – 222 особи, тобто приблизно 44 % генсукупності, для $N = 5000$ – 7,4 %, а для N, що дорівнює 4 мільйонам (наприклад, населення Санкт-Петербурга) – соту долю відсотка. Тому деколи характеристики вибірки, що трапляються в публікаціях типу «було опитано 15 % генсукупності» або «опитувався кожен двадцятий школяр», нічого не говорять про репрезентативність вибірки.

Таблиця 41. **Залежність обсягу вибірки від обсягу генсукупності за припустимоъ похибки 5 % (довірча імовірність – 0,954)**

Обсяг генсукупності	100	200	300	400	500	600	700	800	900	1000	1500	2000
Обсяг вибірки	80	133	171	200	222	240	255	267	277	286	316	333

Обсяг генсукупності	2500	3000	4000	5000	10 000	100 000	Нескінченна
Обсяг вибірки	345	353	364	370	385	398	400

Взагалі з таблиці видно, що, починаючи з якогось моменту, збільшення обсягу генеральної сукупності не має істотного впливу на збільшення обсягу вибірки, тому при великих генеральних сукупностях (скажімо, при $N > 5000$) величиною $\frac{1}{N}$ в формулі (V.1.1') можна знехтувати. Тоді формула (V.1.1') набуває вигляду: $n = \frac{1}{\Delta^2}$, звідки $\Delta = \sqrt{\frac{1}{n}}$. Табл. 42 і рис. 24, що показують зв'язок між обсягом і похибкою вибірки, можуть використовуватися для прийняття рішення про необхідний обсяг вибірки.

При плануванні обсягу вибірки потрібно мати на увазі таке. Наведені вище формули дозволяють отримати задану точність при аналізі вибірки в цілому, тобто якщо ми не будемо розчленовувати її на частини. Якщо, наприклад, потрібно визначити частку осіб,

Таблиця 42. **Залежність між обсягом і похибкою вибірки**

(для $P = 0{,}954$, $p = q = \frac{1}{2}$ і нескінченно великою генеральною сукупністю)

n	100	200	300	400	500	600	700	800	900	1000
Δ	10,0	7,1	5,8	5,0	4,5	4,1	3,8	3,5	3,3	3,2

n	1100	1300	1500	2000	2500	3000	4000	5000	10 000	100 000
Δ	3,0	2,8	2,6	2,2	2,0	-1,8	1,6	1,4	1,0	0,3

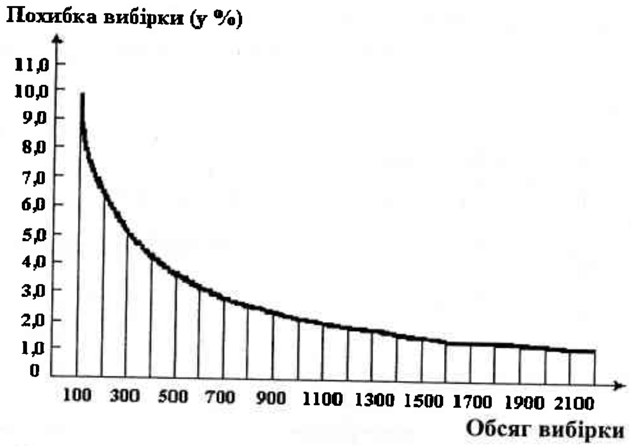

Рис. 24. Залежність між обсягом і похибкою вибірки

(для $P = 0{,}954$, $p = q = \frac{1}{2}$ і нескінченно великої генеральної сукупності)

що живуть у шлюбі, для великого міста, то, опитавши 400 випадково відібраних людей, ми визначимо шукану частку з похибкою, що не перевищує 5 % (з імовірністю 0,954). Але якщо ми хочемо визначити цю частку не для усього масиву в цілому, а для жінок і для чоловіків, то нам необхідно, щоб у вибірці було 400 чоловіків і 400 жінок, тобто 800 осіб. Чим більше буде дробитися масив при аналізі інформації, тим більший обсяг вибірки буде потрібен. Тому зазвичай обсяг вибірки більше залежить від глибини аналізу інформації, ніж від запланованої похибки вибірки.

Практично визначення обсягу вибірки здійснюється в ході приблизно такого діалогу із замовником (якщо замовника немає і ви самі плануєте витрати, можна вважати, що це діалог із вашим внутрішнім голосом):

Ви: У соціології зазвичай прийнято використовувати імовірність 0,954. Чи влаштовує вас, якщо похибка з імовірністю 0,954 буде відрізнятися від генеральної сукупності не більш ніж в 5 %, за будь-якою ознакою анкети?

Замовник: Цілком влаштовує. Мене цікавить відсоток населення, що підтримує приєднання України до НАТО, точність в 5 % є достатньою.

Ви: Чи достатньо вам отримати оцінку відсотка прибічників приєднання до НАТО для України в цілому чи вас цікавлять також якісь соціальні групи?

Замовник: Мене цікавить цей відсоток серед чоловіків і жінок, серед людей з незакінченою середньою, середньою і вищою освітою, серед людей шести вікових груп (до 25, 26–35, 36–45, 46–55, 56–65 і 66 та старші), а також серед жителів західного, центрального, східного і південного регіонів.

Ви (розмірковуючи уголос): Отже, ви хочете рахувати таблиці двомірного розподілу ознаки «ставлення до НАТО» і ознаки стать (2 альтернативи); освіта (3 альтернативи); вік (6 альтернатив) і регіон (4 альтернативи). Максимальне число альтернатив – в ознаці «вік». Щоб з імовірністю 0,954 отримати 5 % похибку, потрібно (див. формулу (V.1.1) або табл. 37, *б*) опитати 400 респондентів. Щоб отримати 5 % похибку для кожної з вікових груп, потрібно опитати $400 \cdot 6 = 2400$ респондентів, тоді вибірка буде тим більш репрезентативна для груп за статтю, освітою і регіоном. Отже, обсяг вибірки – 2400 респондентів.

Замовник: У мене немає грошей на таке дослідження, є гроші тільки на 1600 респондентів.

Ви: Тоді похибка не перевищить 5 % для груп за статтю, освітою і регіоном, але в кожній віковій групі у вас в середньому буде 267 респондентів, тобто похибка становитиме близько 6,1 %. Вас це влаштовує?

Замовник: Забагато.

Ви: А не могли б ви згрупувати населення не в 6, а в 5 вікових груп, тоді у вас в кожній групі буде близько 320 респондентів і похибка вибірки (формула V.1.1') становитиме 5,6 %?

Замовник: Це, напевне, можна. А який обсяг вибірки потрібний для 5 груп, щоб похибка була не більше 5 %?

Ви: $400 \cdot 5 = 2000$.

Замовник: Гаразд, домовилися. Я буду аналізувати не 6, а 5 груп за віком. Якщо я знайду гроші ще на 400 респондентів, то обсяг вибірки буде 2000, а якщо ні – буде 1600 і я миритимуся з похибкою в 5,6 %.

Звичайно, це дещо ідеалізований діалог; реально замовник може не знати толком, чого він хоче, або знає, але не вміє висловити, а ви недостатньо добре знаєте специфіку його проблем і не розумієте, що йому потрібно, і т. д. Хоча суть діалогу не змінюється.

Але основна складність при плануванні вибірки полягає, мабуть, у тому, що в багатьох випадках, особливо при великомасштабних дослідженнях, випадкову вибірку сформувати дуже складно. Так, у проведеному нами репрезентативному дослідженні працюючого населення м. Києва[1] ($P = 0{,}954$, Δ не більше 3–5 %), дані якого вже використовувалися в прикладах, було застосовано наступну процедуру, що моделює випадковий відбір. З виборчих списків для кожної дільниці міста (всього їх більше семисот[2]) відбиралися з певним кроком адреси виборців: наприклад, адреса 1-а, 100-а, 200-а і т. д. (якщо крок – 100).

[1] Дослідження проводилися у 1985 р. відділом конкретних соціологічних досліджень Інституту філософії АН УРСР (керівник дослідження В. Ф. Черноволенко, проект вибірки був розроблений В. І. Паніотто).

[2] Списки всіх виборчих дільниць району зберігались в райвиконкомах 12 районів Києва, що дуже спрощувало роботу.

При цьому могла бути допущена систематична похибка такого роду: прізвища в списках розташовані за алфавітом; починаючи з 1-го номера, ми майже завжди включаємо до вибіркового списку прізвище, що починається з літери А, тобто в списку частка осіб, що мають прізвища на літеру А, буде вищою, ніж в генсукупності. Якщо у осіб якоїсь національності (наприклад, вірменської) прізвища частіше починаються на А, ніж у осіб інших національностей, то їх частка у вибірці вища, ніж в генсукупності.

Тому про всяк випадок ми «стохастизували»[1] вибір першого прізвища в списку, зробивши його випадковим, рівномірно розподіленим усередині кроку вибірки (можна, наприклад, починати з номера, що дорівнює цілій частині числа $\frac{k}{7}+1$, де k – номер виборчої дільниці, що змінюється від 1 до 700: тоді в 7 виборчих дільницях прізвища будуть відбиратися з 1-го номера, в 7 – з 2-го і т. д. до сотого номера.

Оскільки нас цікавило все працююче населення, а не тільки особи, старші 18 років і занесені в списки для голосування, опитувалася не обов'язково особа, що була зазначена у виборчому списку. Список використовувався лише як вибірка адрес. При відвідуванні сім'ї, що проживає за даною адресою, інтерв'юер переписував усіх жильців у певному порядку і за спеціальною процедурою[2], стохастизуючою вибір, визначав, кого потрібно опитати[3]. Можна показати, що отримана вибірка є випадковою, для оцінки похибки вибірки можна застосовувати формулу (V.1.1).

Але що робити у разі, коли такий список неможливо скласти, наприклад, при побудові вибірки, репрезентативної для країни? У таких випадках удаються до багатоступінчастого відбору: наприклад, у першій для України репрезентативній вибірці, що була розроблена в 1983 р. Л. Фінкелем, В. Паніотто і Н. Чуріловим для Держтелерадіо УРСР[4], спочатку (перший ступінь) з множини всіх областей країни відбиралися області (область у даному випадку – це одиниця відбору на першому кроці). На другому ступені з обраних областей відбиралися райони. На третьому з кожного району – на-

[1] Від слова стохастичний, тобто випадковий.

[2] Використовувалася деяка модифікація процедури Кіша. (*Кокрен У.* Методы выборочного исследования.– М., 1976.– С. 384, 385; *Петренко Е. С., Ярошенко Т. М.* Социально-демографические показатели в социологических исследованиях.– М., 1979.– С. 96.)

[3] Для реалізації випадкових відборів іноді використовуються спеціальні таблиці випадкових чисел (див. табл. додатка 3).

[4] Див.: *Паниотто В. И., Рукавишников В. О., Чурилов Н. Н.* Опросы населения (методический опыт).– М.: Финансы и статистика, 1984.

селені пункти. На четвертому з населених пунктів – особи, що підлягають опитуванню (на перших трьох ступенях вибиралися одиниці відбору, на останньому – одиниці дослідження). При такій побудові вибірки формула (IV.1.1) непридатна для оцінки похибки, потрібні формули, що дозволяють оцінити похибку, яка виникає на кожному ступені[1]. Щоб оптимізувати процедуру побудови вибірки, дослідники на різних ступенях відбору використовують спеціальні процедури. Наприклад, проводять класифікацію первинних одиниць відбору (областей, районів або населених пунктів), об'єднують їх в однорідні групи (страти), а потім відбирають певне число одиниць відбору з кожної страти. Залежно від процедур, що використовуються для відбору на кожному ступені, вибірки називають стратифікованими (районованими) або кластерними (гніздовими), на різних ступенях відбору можуть використовуватися різні типи відбору.

Взагалі будь-яке вибіркове дослідження може бути охарактеризовано такими параметрами: числом ступеней, способом виділення кластерів, способом їх групування (стратифікації) і способом відбору кластерів на кожному ступені. Багато що залежить також від наявної статистики і від територіально-адміністративної системи в тій чи іншій країні. Наприклад, побудова вибірок у країнах колишнього Радянського Союзу відрізняється від США з добре розвиненою системою територіальних карт різного рівня[2]. Очевидно, що в багатьох випадках спосіб організації вибірки дуже далекий від одноступінчастого випадкового або навіть багатоступінчастого випадкового відбору. Порівнюючи величину похибки з тією, яка була б, якщо б вибірка будувалася як одноступінчаста випадкова (тобто з похибкою, розрахованою за формулою (V.1.1)), можна оцінити відхилення побудованої вибірки від випадкової і внести корективи в результати, що були отримані за допомогою формул, заснованих на припущенні про випадковості вибірки. Співвідношення похибки вибірки даної конструкції (даного дизайну – якщо використовувати пряме запозичення англійського терміна) і похибки випадкової вибірки називається ***дизайн-ефектом***. Читачеві, що стикається з викладенням результатів соціологічного опитування, слід звертати увагу на те, чи наведена величина похибки вибірки з дизайн-ефектом, чи без дизайн-ефекту. На жаль, у більшості публі-

[1] *Кокрен У.* Методы выборочного исследования.– М., 1976, гл. *Kalton G.* Introduction to survey sampling.– Sage Publications Inc., 1983.– P. 1–94.

[2] Рекомендації по побудові вибірок у колишньому Радянському Союзі див.: *Swafford M., Kosolapov M., Kish L., Heeringa S.* Sample Design for Republics of the Former Soviet Union.– National Council for Soviet and East European Research, 1995. В розділі 5 описано вибірку для України Київського міжнародного інституту соціології.

кацій таких даних в Україні в 2003 р., коли ми писали цю книгу (не кажучи вже про попередні роки), взагалі не сказано, чи врахований дизайн-ефект вибірки. Це означає, що автори не враховують дизайн-ефекту і тому реальна похибка вибірки може бути в півтора-два рази вища за ту, про яку повідомляється в публікації.

Викладення різних типів вибірки і методів оцінки їх похибки потребує спеціального розгляду і не може бути зроблене в межах цієї книги. Тому весь подальший матеріал цього розділу подано в припущенні, що із генеральної сукупності вибирається випадкова одноступінчаста вибірка, для якої дизайн-ефект дорівнює 1.

2 Вибірковий розподіл

Статистичний висновок – це певне твердження про досліджувану генеральну сукупність на основі вивчення вибірки (тобто міркування від окремого до загального, індукція). Математична статистика розглядає, певна річ, не будь-які твердження про генеральну сукупність, а лише ті, що стосуються числових характеристик, розглянутих вище (середні, міри варіації, коефіцієнти кореляції і т. д.). Числові характеристики, що описують генеральну сукупність, називаються *параметрами.* Ті ж самі характеристики, але розраховані для вибірки, називаються *статистиками.* Ми будемо позначати параметри через M^{Γ}, $(\sigma^{\Gamma})^2$, r^{Γ} і т. д. (генеральне середнє, дисперсія, коефіцієнт кореляції), а статистики через $M^{В}$, $(\sigma^{В})^2$, $r^{В}$ і т. д. (вибіркове середнє, дисперсія, коефіцієнт кореляції)[1]. Таким чином, статистичний висновок – це твердження про параметри генеральної сукупності на основі вивчення статистики. Такі твердження мають імовірнісний характер і підрозділяються на три види: статистичне оцінювання точкове, статистичне оцінювання інтервальне і перевірка гіпотез.

Статистичне оцінювання полягає в тому, що дослідник за вибіркою шукає показник, найбільш близький до оцінюваного параметра, чи інтервал, в межах якого з великою імовірністю лежить цей параметр. Іншим різновидом статистичного висновку є перевірка гіпотез: дослідник заздалегідь формулює якесь твердження про параметри генеральної сукупності (гіпотезу), потім оцінює ступінь відповідності результатів, отриманих у вибірковому дослідженні, сформульованій гіпотезі й приймає рішення про істинність чи хибність

[1] У літературі можна зустріти також позначення параметрів грецькими, а статистик – латинськими літерами. Наприклад, μ, δ^2, π и M^2, s^2, *p* для середнього, дисперсії і частки ознаки відповідно.

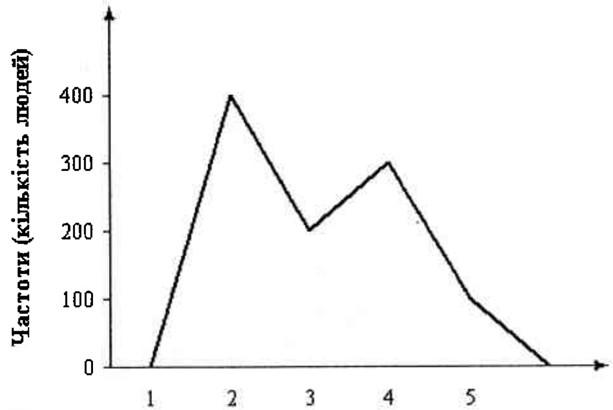

Рис. 25. Приклад полігону розподілу (оцінки з математики)

гіпотези. Методологію перевірки статистичних гіпотез ми розглянемо докладніше, що дозволить уточнити відмінність і подібність видів статистичного висновку. У даний час для нас важливо, що статистичне оцінювання і перевірка гіпотез ґрунтуються на ідеї так званого *вибіркового розподілу* (звертаємо увагу читача на важливість цього поняття для розуміння сутності статистичного висновку).

Розгляд вибіркового розподілу почнемо з прикладу. Припустімо, що відомий розподіл оцінок 1000 абітурієнтів якогось вузу на іспиті з математики. Нехай 400 осіб отримали «2», 200 – «3», 300 – «4» і 100 – «5»; полігон розподілу наведено на рис. 25. Легко підрахувати, що середній бал дорівнює $M^{\Gamma} = 3{,}1$, а дисперсія $(\sigma^{\Gamma})^2 = 0{,}11$. Наскільки ймовірно отримати у вибірці значення, що суттєво відрізняється від генерального середнього? Визначимо, наприклад, імовірність того, що для вибірки з 5 осіб вибіркове середнє $M^{В}$ буде відрізнятися від генерального не менше ніж на 0,5, тобто $\left|M^{\Gamma} - M^{В}\right| \geq 0{,}5$. З цією метою станемо формувати вибірки за п'ятою особою багаторазово і обчислювати для кожної з них середній бал. Тоді відношення таких вибірок, для яких $\left|M^{\Gamma} - M^{В}\right| \geq 0{,}5$, до загального числа вибірок, що робляться, дасть частоту, близьку до шуканої імовірності. При збільшенні числа вибірок ця частота необмежено наближається до імовірності того, що $\left|M^{\Gamma} - M^{В}\right| \geq 0{,}5$.

Побудуємо полігон розподілу вибіркових середніх (на осі абсцис відкладаються значення $M^{В}$, по осі ординат – частоти): при

□ *Вправа 101.* Розрахуйте $N = \sum_{i=2}^{3} \sum_{j=3}^{5} N_{ij}$ для таблиці нижче:

2	3	5	2	1
2	6	3	3	1
5	1	1	5	3
1	5	3	4	4

Відповідь: 3 + 3 + 1 + 1 + 5 + 3 = 16.

Статистичні таблиці

Додаток 3

Наведені нижче таблиці частково запозичені з інших видань, частково розраховані авторами. Під час добору таблиць ми керувалися, по-перше, міркуваннями зручності (наведені в деяких виданнях таблиці потребують іноді перерахунку при використанні – тут цей недолік був усунутий); по-друге, міркуваннями відповідності таблиць поширеним у соціологічних дослідженнях формам представлення і кількісним характеристикам інформації (наприклад, таблиця χ^2 часто наводиться у статистичній літературі лише для числа ступенів свободи, що не перевищують 30, в той час як в соціологічних дослідженнях нерідко трапляються таблиці 7×10, 10×10 тощо з 50–80 і навіть більшим числом ступенів свободи). Наведені нижче таблиці складено, як правило, для числа ступенів свободи чи обсягу вибірки від 1 до 1000 та для рівнів значущості 5 %, 1 та 0,1 %. Сподіваємося, що така стандартизація полегшить користування таблицями.

Виходячи з висловлених у розділі V міркувань щодо більшої небезпечності похибок I роду, ніж II роду, *усі таблиці наводяться для двостороннього критерію* (що зменшує вірогідність похибок I роду і збільшує вірогідність похибок II роду). З цих самих міркувань рекомендуємо читачеві у випадку, якщо в таблиці не наведено критичне значення для отриманої ним статистичної характеристики, брати з деяким запасом найближче критичне значення (або вдатися до інтерполяції).

Зазначимо також, що під час роботи над таблицями було виявлено помилки друку в деяких виданнях[1].

[1] Статистические методы..., с. 300, останній рядок: замість 7,251 має бути 6,251, а замість 6,815 – 7,815; *Венецкий И. Г., Венецкая В. И.* Основные математико-статистические понятия и формулы в экономическом анализе.– М., 1979.– С. 413, останній рядок: замість 1,77 має бути 2,77.

Таблиця А. **Нормальний розподіл**[1]

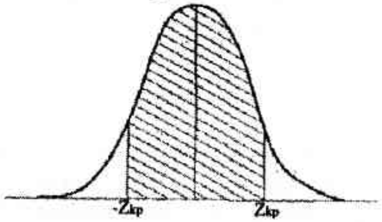

Рис. 35. Частки площі (*P*) під нормальною кривою в межах від $-z_{кр}$ до $+z_{кр}$

$z_{кр}$	P	$z_{кр}$	P	$z_{кр}$	P
0,00	0,00000	0,30	0,23582	0,60	0,45149
01	00798	31	24344	61	45814
02	01596	32	25103	62	46474
03	02393	33	25860	63	47131
04	03191	34	26614	64	47783
05	03988	35	27366	65	48431
06	04784	36	28115	66	49075
07	05581	37	28862	67	49714
08	06376	38	29605	68	50350
09	07171	39	30346	69	50981
0,10	0,07966	0,40	0.31084	0,70	0,51607
11	08759	41	31819	71	52230
12	09552	42	32552	72	52848
13	10348	43	33280	73	53461
14	11134	44	34006	74	54070
15	11924	45	34729	75	54675
16	12712	46	35448	76	55275
17	13499	47	36164	77	55870
18	14285	48	36877	78	56461
19	15069	49	37587	79	57047
0,20	0,15852	0,50	0,38292	0,80	0,57629
21	16633	51	38995	81	58206
22	17413	52	39694	82	58778
23	18191	53	40389	83	59346
24	18967	54	41080	84	59909
25	19741	55	41768	85	60468
26	20514	56	42452	86	61021
27	21284	57	43132	87	61570
28	22052	58	43809	88	62114
29	22818	59	44481	89	62653

[1] *Венецкий И. Г., Венецкая В. И.* Основные математико-статистические понятия...– С. 402–404.

Продовження табл. А

$z_{кр}$	P	$z_{кр}$	P	$z_{кр}$	P
0,90	0,63188	1,35	0,82298	1,80	0,92814
91	63718	36	82617	81	92970
92	64243	37	82931	82	93124
93	64763	38	83241	83	93275
94	65278	39	83547	84	93423
95	65789	1,40	0,83849	85	93569
96	66294	41	84146	86	93711
97	66795	42	84439	87	93852
98	67291	43	84728	88	93989
99	67783	44	85013	89	94124
1,00	0,68269	45	85294	1,90	0,94257
01	68750	46	85571	91	94387
02	69227	47	85844	92	94514
03	69699	48	86113	93	94639
04	70166	49	86378	94	94762
05	70628	1,50	0,86639	95	94882
06	71086	51	86696	96	95000
07	71538	52	87149	97	95116
08	71986	53	87398	98	95230
09	72429	54	87644	99	95341
1,10	0,72867	55	0,87886	2,00	0,95450
11	73300	56	88124	01	95557
12	73729	57	88358	02	95662
13	74152	58	88589	03	95764
14	74571	59	88817	04	95865
15	74986	1,60	0,89040	05	95964
16	75395	61	89260	06	96060
17	75800	62	89477	07	96155
18	76200	63	89690	08	96247
19	76595	64	89899	09	96338
1,20	0,76986	65	90106	2,10	0,96427
21	77372	66	90309	11	96514
22	77754	67	90508	12	96599
23	78130	68	90704	13	96683
24	78502	69	90897	14	96765
25	78870	1,70	0,91087	15	96844
26	79233	71	91273	16	96923
27	79592	72	91457	17	96999
28	79945	73	91637	18	97074
29	80295	74	91814	19	97148
1,30	0,80640	75	91988	2,20	0,97219
31	80980	76	92159	21	97289
32	81316	77	92327	22	97358
33	81648	78	92492	23	97425
34	81975	79	92655	24	97491

Продовження табл. А

$z_{кр}$	P	$z_{кр}$	P	$z_{кр}$	P
2,25	0,97555	2,70	0,99307	3,15	0,99837
26	97618	71	99327	16	99842
27	97679	72	99347	17	99848
28	97739	73	99367	18	99853
29	97798	74	99386	19	99858
2,30	0,97855	75	99404	3,20	0,99863
31	97911	76	99422	21	99867
32	97966	77	99439	22	99872
33	98019	78	99456	23	99876
34	98072	79	99473	24	99880
35	98123	2,80	0,99489	25	99855
36	98172	81	99505	26	99889
37	98221	82	99520	27	99892
38	98269	83	99535	28	99896
39	98315	84	99549	29	99900
2,40	0,98360	85	99563	3,30	0,99903
41	98405	86	99576	31	99907
42	98448	87	99590	32	99910
43	98490	88	99602	33	99913
44	98531	89	99615	34	99916
45	98571	2,90	0,99627	35	99919
46	98611	91	99639	36	99922
47	98649	92	99650	37	99925
48	98686	93	99661	38	99928
49	98723	94	99672	39	99930
2,50	0,98758	2,95	0,99682	3,40	0,99933
51	98793	96	99692	41	99935
52	98826	97	99702	42	99937
53	98859	98	99712	43	99940
54	98891	99	99721	44	99942
55	98923	3,00	0,99730	45	99944
56	98953	01	99739	46	99946
57	98983	02	99747	47	99948
58	99012	0,3	99755	48	99950
59	99040	04	99763	49	99952
2,60	0,99068	05	99771	3,50	0,99953
61	99095	06	99779	51	99955
62	99121	07	997P6	52	99957
63	99146	08	99793	53	99958
64	99171	09	99800	54	99960
65	99195	3,10	0,99806	55	99961
66	99219	11	99813	56	99963
67	99241	12	99819	57	99964
68	99263	13	99825	58	99966
69	99285	14	99831	59	99967

Продовження табл. А

$z_{кр}$	P	$z_{кр}$	P	$z_{кр}$	P
3,60	0,99968	3,74	0,99982	3,88	0,99990
61	99969	75	99982	89	99990
62	99971	76	99983	3,90	0,99990
63	99972	77	99984	91	99991
64	99973	78	99984	92	99991
65	99974	79	99985	93	99992
66	99975	3,80	0,99986	94	99992
67	99976	81	99986	95	99992
68	99977	82	99987	96	99992
69	99978	83	99987	97	99993
3,70	0,99978	84	99988	98	99993
71	99979	85	99988	99	99993
72	99980	86	99989		
73	99981	87	99989		

Таблиця Б. **Значення x_0^2 залежно від числа ступенів свободи (*f*) та рівня значущості**[1]

f	Рівень значущості, %			*f*	Рівень значущості, %		
	5	1	0,1		5	1	0,1
1	3,84	6,63	10,83	24	36,41	42,98	51,18
2	5,99	9,21	13,81	25	37,65	44,31	52,62
3	7,81	11,34	16,27	26	38,88	45,64	54,05
4	9,49	13,28	18,46	27	40,11	46,96	55,48
5	11,07	15,09	20,52	28	41,34	48,28	56,89
6	12,59	16,81	22,46	29	42,56	49,59	58,30
7	14,07	18,47	24,32	30	43,77	50,89	59,70
8	15,51	20,09	26,12	40	55,76	63,69	73,40
9	16,92	21,67	27,88	50	67,50	76,15	86,66
10	18,31	23,21	29,59	60	79,08	88,38	99,61
11	19,67	24,72	31,26	70	90,53	100,42	112,33
12	21,03	26,22	32,91	80	101,88	112,33	124,84
13	22,37	27,69	34,53	90	113,15	124,12	137,21
14	23,68	29,14	36,12	100	124,34	135,81	149,45
15	25,00	30,58	37,70	200	233,99	249,44	267,50
16	26,30	32,00	39,25	300	341,40	359,91	–
17	27,59	33,41	40,79	400	447,63	468,72	–
18	28,87	34,80	42,31	500	553,13	576,49	–
19	30,14	36,19	43,82	600	658,09	683,52	–
20	31,41	37,57	45,31	700	762,66	789,97	–
21	32,67	38,93	46,80	800	866,91	895,98	–
22	33,92	40,29	48,27	1000	1074,68	1106,97	–
23	35,17	41,63	49,73				

[1] Частину таблиці було запозичено з книжки: *Закс Л.* Статистическое оценивание.– М., 1976.– С. 132–134; частину з: *Оуэн Д. Б.* Сборник статистических таблиц.– М., 1966, с. 49–55.