5. ВИБІРКОВИЙ МЕТОД

5.1. Сутність та різновиди вибіркових спостережень

Вибіркове спостереження – це вид несуцільного спостереження, при якому обстежуються не всі елементи сукупності, а лише певним чином дібрана їх частина. Сукупність, з якої вибирають елементи для обстеження, називається ***генеральною сукупністю***, а сукупність, яку безпосередньо обстежують, – ***вибірковою сукупністю. Кінцева мета вибіркового спостереження – поширення його статистичних характеристик на генеральну сукупність.*** Іншими словами вибіркова сукупність представляє – ***репрезентує*** генеральну сукупність.

Вибіркове спостереження забезпечує, у порівнянні з суцільним спостереженням, економію трудових, фінансових ресурсів та часу. При обстеженні частини генеральної сукупності зменшуються помилки реєстрації, але можлива похибка репрезентативності. При вивченні деяких соціально-економічних явищ єдине можливе спостереження – це вибіркове. Наприклад, перевірка якості, яка пов’язана із руйнуванням продукції, перевірка чистоти та вологості зерна, наукове вивчення біологічних ресурсів…

За причинами виникнення ***похибки*** поділяються:

Тенденційні (***систематичні***) ***похибки*** виникають внаслідок порушення принципу випадковості: упереджений вибір елементів, недосконала основа вибірки тощо. Ці похибки небезпечні внаслідок систематичного викривлення результатів.

Випадкові похибки виникають унаслідок випадковості вибору елементів для дослідження. Випадкові похибки мають назву ***похибок репрезентативності,*** їх принципово неможливо уникнути. Однак завжди є можливість знайти ймовірність того, що випадкова похибка вибірки не перевищить заданого (припустимого) рівня, або знайти обсяг вибірки, який забезпечить потрібну точність результатів для взятої ймовірності.

Різновиди вибірок:

Простий випадковий добір – проводиться жеребкуванням чи за допомогою таблиць випадкових чисел. Для жеребкування на кожну одиницю генеральної сукупності необхідно заготувати відповідну фішку. Для використання таблиць випадкових чисел усі елементи генеральної сукупності мають бути пронумеровані. Якщо сукупність велика чи розповсюджена на території та у часі, так використовувати простий випадковий метод вибірки важко.

Механічний добір використовують, якщо чисельність елементів сукупності впорядкована. Добір інтервалів здійснюється через рівні інтервали. Механічна вибірка порівняно із простою випадковою ефективніша, її простіше здійснити. Однак при наявності циклічних коливань значень ознаки, цикл яких збігається з інтервалом добірки, можливий зсув вибіркових оцінок.

Іноді, вивчаючи безпосередні в часі процеси, проводять ***моменті спостереження*** – фіксація стану процесу на певні моменти часу, які вибирають за схемою випадкової або механічної вибірки.

Розшарований добір – це спосіб формування вибірки з урахуванням структури генеральної сукупності. Незалежний добір здійснюється у кожній складовій частині генеральної сукупності, яка попередньо структурується.

Серійна вибірка ґрунтується на розподілі генеральної сукупності на частини – серії. Серії відбираються за схемою простої випадкової, або механічної вибірки. Вибрані **серії підлягають суцільному обстеженню,** а залишок – серій не обстежується зовсім.

Поєднання різних видів вибірки можливе у рамках багатоступеневої вибірки.

5.2. Вибіркові оцінки середньої та частки

Кінцева мета вибіркового спостереження – це поширення його характеристик на генеральну сукупність. Для середньої величини (а також частки) визначають межі можливих значень їх у генеральній сукупності з певною ймовірністю – ***довірчі межі.***

У статистиці є два типи оцінок параметрів генеральної сукупності:

Точкова оцінка – це значення параметру генеральної сукупності приймається за даними вибірки: вибіркова середня та вибіркова частка.

Інтервальна оцінка – це інтервал значень параметра, розрахований за даними вибірки для певної ймовірності – ***довірчий інтервал.*** Чим точніше вибіркова оцінка, тим менший довірчий інтервал. Межі довірчого інтервалу визначають на основі точкової оцінки та ***граничної похибки вибірки.***

$$\Delta x = t \cdot \mu_x; \quad \Delta d = t \cdot \mu_d, \qquad (5.1)$$

$$x_i - \bar{x} = \Delta x,$$

$$t = \frac{x_i - \bar{x}}{\sigma},\ t = \frac{\Delta x}{\sigma}, \tag{5.2}$$

$$\Delta x = t \cdot \sigma,$$

довірчий інтервал для середньої

$$(\bar{x} - \Delta x) \le \bar{x} \le (\bar{x} + \Delta x),$$

$$(\bar{x} - t\mu_x) \le x_0 \le (\bar{x} + t\mu_x), \tag{5.3}$$

довірчий інтервал для частки

$$(\bar{d} - \Delta d) \le d_0 \le (\bar{d} + \Delta d),$$

$$(\bar{d} - t\mu_d) \le d_0 \le (d + t\mu_d),$$

(5.4)

де t – ***довірче число (квантиль розподілу ймовірностей), коефіцієнт довіри***, μ – ***середня*** або ***стандартна похибка вибірки,*** тобто середнє квадратичне відхилення вибіркових оцінок від значень параметра в генеральній сукупності.

Вибірки здійснюються ***повторним*** та ***безповторним методом.*** При повторному методі кожна вибрана одиниця сукупності повертається після обстеження до сукупності і може бути вибрана знову. При безповторному методі кожна обстежена одиниця не повертається до сукупності і принципово не може бути обстежена повторно.

При безповторному методі:

$$\mu_x = \sqrt{\frac{\sigma^2}{n-1}\left(1 - \frac{n}{N}\right)},$$

$$\mu_d = \sqrt{\frac{P(1-p)}{n-1}\left(1 - \frac{n}{N}\right)}. \tag{5.5}$$

При повторному методі:

$$\mu_x = \sqrt{\frac{\sigma^2}{n-1}},$$

$$\mu_d = \sqrt{\frac{p(1-p)}{n-1}}, \tag{5.6}$$

де N – кількість елементів генеральної сукупності; n – кількість елементів вибіркової сукупності; σ^2 – дисперсія ознаки вибірковій сукупності; p – частка вибіркової сукупності, якій властива ознака; g =1 – p – частка вибіркової сукупності, якій не властива ознака.

$$\sigma^2 = p(1-p) = pg \tag{5.7}$$

При суцільному спостереженні похибка репрезентативності відсутня $(\Delta \to 0)$, (при без повторному доборі).

При безповторному доборі середня або стандартна похибка вибірки μ відрізняється відсутністю множника $(1-\frac{n}{N}) \leq 1$ від стандартної похибки при безповторному доборі:

$$\mu = \sqrt{\frac{\sigma^2}{n}},$$
$$\mu = \sqrt{\frac{\sigma^2}{n}(1-\frac{n}{N})}. \qquad (5.8)$$

У тих випадках, коли обсяг вибірки незначний (2 ÷ 5 %), множник

$$(1-\frac{n}{N}) \approx 0{,}98 \div 0{,}95;$$
$$\sqrt{0{,}98 \div 0{,}95} = 0{,}99 \div 0{,}975,$$

та їм можливо знехтувати і користуватись формулою для повторного добору у випадку як повторного, так і безповторного добору.

5.3. Сталість, незсувність та ефективність оцінки параметра вибірки

Оцінка параметра вибірки вважається сталою, якщо при зростанні обсягу вибірки з певною мірою (імовірність $\to$ 1) значення оцінки наближається до свого теоретичного значення. Сталість оцінки гарантує дослідникові зростання точності з ростом обсягу вибірки:

$$n \to N,\ \bar{x}_0 \to \bar{x}\ .$$

Незсувність оцінки параметра вибірки позначає відсутність систематичної (тенденційної) помилки в оцінці параметра.

Незсувність оцінки дисперсії вибіркової сукупності гарантується введенням у знаменник коригуючого ***множника Шепарда*** (n – 1):

$$\sigma_b^2 = \frac{\sum_1^n (x_i - \bar{x}_b)^2}{n-1}, \qquad (5.9)$$

яким можливо знехтувати при $n \geq 30$, але для малих вибірок, де $n \leq 20$ його потрібно враховувати.

Окрім того, для малих вибірок діє закон розподілу Стьюдента, у якому квантіль розподілу «t» (довірче число) залежить не тільки від ймовірності, з якою обчислюється гранична похибка, а також від кількості елементів вибірки. Для t = 1 залежність відсотка ймовір-ностей від кількості елементів сукупності надано у таблиці.

Таблиця 5.1

Залежність відсотка ймовірностей від кількості елементів сукупності

Кількість елементів сукупності, *n*	**2**	**4**	**5**	**10**	**20**	**30 і більше**
Відсоток ймовірності	0,5	0,609	0,637	0,657	0,670	0,683

Оцінка параметра вибірки вважається ***ефективною,*** якщо серед усіх оцінок параметра, вона має найменшу дисперсію, тобто найменшу випадкову помилку і у цьому разі вважається найбільш точною.

5.4. Визначення обсягів вибірки

На практиці є потреба визначити ***обсяг вибірки.*** Задача формулюється таким чином: є обсяг генеральної сукупності N, гранична похибка Δx або Δp, імовірність граничної похибки t, %; треба визначити n – кількість елементів вибірки.

Для повторного добору використовується формула:

$$n = \frac{t^2\sigma^2}{\Delta^2},\quad \Delta = \Delta x, \text{ або } \Delta = \Delta p, \tag{5.10}$$

для безповторного добору:

$$n' = \frac{n}{1 + \frac{n}{N}}. \tag{5.11}$$

у порівнянні з попередніми формулами:

$$n = \frac{t^2\sigma^2}{\Delta^2} + 1, \text{ та} \tag{5.12}$$

$$n' = \frac{t^2\sigma^2 N}{\Delta^2 N + t^2\sigma^2} \tag{5.13}$$

або:

$$n' = \frac{t^2 \cdot p(1-p)N}{\Delta^2 N + t^2 \cdot p(1-p)}. \tag{5.14}$$

Для розв'язку задачі необхідно знати дисперсію сукупності – σ^2. Є декілька варіантів визначення дисперсії:

1. σ^2 використовують з попереднього дослідження близької сукупності.

2. $\sigma = \frac{1}{6} R = \frac{1}{6}(x_{\max} - x_{\min})$ – як закон розподілу наближається до нормального закону розподілу.

3. Використовують метод послідовних наближень.

4. Використання розмаху варіації «R» та кількості елементів сукупності «n».

Для приблизного знаходження дисперсії в залежності від кількості елементів сукупності використовується формула:

$$\sigma = \frac{x_{\max} - x_{\min}}{k}, \tag{5.15}$$

де k – коефіцієнт, який залежить від кількості елементів сукупності n, наведений у таблиці.

Таблиця 5.2

Залежність коефіцієнта «к» від кількості елементів сукупності «n»

n	2-5	6-15	16-49	50-200	201-1000	> 1000
k	2	3	4	5	6	7

Середня арифметична приблизно дорівнює:

$$\bar{x} = \frac{x_{\max} + x_{\min}}{2}. \tag{5.16}$$

Розшарований (типологічний) добір відрізняється тим, що у формулах граничної помилки замість дисперсії вибіркової сукупності (σ_b^2) використовується середня з внутрішньогрупових дисперсій:

$$\sigma_j^2 = \frac{\sum (x - \bar{x}_j)^2 \cdot f}{\sum f}, \tag{5.17}$$

$$\bar{\sigma}_{jb}^2 = \frac{\sum_1^m \sigma_{jb}^2 \cdot n_j}{\sum_1^m n_j \equiv n}, \tag{5.18}$$

$$(\bar{\sigma}_j^2 = \frac{\sum_1^m \sigma_j^2 f_j}{\sum_1^m f_j}), \tag{5.19}$$

де m – кількість груп, на яку поділена сукупність, j – номер групи, n_j – кількість елементів обраних у групі, n – загальна кількість елементів вибірки, σ_{jb}^2 – дисперсія вибіркової сукупності у «j»-ій групі.

Якщо групи однакові за кількістю елементів, так $n_j = \dfrac{n}{m,}$, якщо ні, то:

$$n_j = \frac{n}{N} N_{j,} \quad \text{àáî} \quad n_j = \frac{N_j \sigma_{jb}}{\sum_1^m N_j \sigma_{jb}} n, \tag{5.20}$$

де N_j – загальна кількість елементів у «j» групі;

$$\begin{gathered} \sum_1^m N_j \equiv N, \\ \sum_1^m n_j \equiv n. \end{gathered} \tag{5.21}$$

У зв’язку з загальним правилом складання дисперсій $\sigma^2 > \bar{\sigma}_j^2$, тому помилка при типологічному доборі завжди менша ніж при випадковому.

Завдання для самоконтролю

1. Яке спостереження має назву вибіркового? У чому його сутність та мета?
2. Назвіть переваги та недоліки вибіркового спостереження.
3. Чому при вибірковому спостереженні принципово присутні помилки? Назвіть їх різновиди.
4. Як оцінити розмір похибки репрезентативності?
5. Які показники генеральної та вибіркової сукупності порівнюються між собою?
6. Перелічите та охарактеризуйте існуючи методи добору.
7. Як здійснюється простий випадковий добір?
8. Як здійснюється механічний добір?
9. Як здійснюється розшарований добір?
10. Як здійснюється серійний добір?
11. Що таке точкова оцінка параметрів генеральної сукупності?
12. Що таке інтервальна оцінка параметрів генеральної сукупності?
13. Як обчислюється гранична похибка вибірки для середнього та для частки?

14. Від чого залежить і як знаходиться довірче число (квантиль розподілу ймовірностей, коефіцієнт довіри)?
15. У чому різниця між повторним та безповторним методами добору?
16. Порівняйте між собою повторну та безповторну вибірки.
17. У яких випадках є можливість користуватися формулою стандартної похибки при повторному доборі для безповторного добору?
18. Як обчислюється гранична похибка вибірки при розшарованому (типовому) доборі?
19. Як обчислюється гранична похибка вибірки при серійному доборі?
20. Що таке мала вибірка і в чому її особливість?
21. Як знаходиться необхідна чисельність вибірки?
22. Як визначається дисперсія майбутньої вибірки?
23. Що таке сталість оцінки параметра вибірки?
24. Що таке незсувність оцінки параметра вибірки?
25. Що таке ефективність оцінки параметра вибірки?

Задачі

5.1. З метою контролю за дотриманням норм витрат металу проведено вибіркове спостереження партії заготовок деталей, які надходять з ливарного цеху. Вага відібраних 10 зразків становила, г: 520, 565, 550, 540, 525, 485, 515, 595, 500, 555.

Визначте:

а) середню вагу перевірених заготовок деталей, граничну похибку та довірчі межі для середньої з імовірністю 0,95;

б) чи відповідають фактичні витрати металу на виготовлення деталей установленій нормі – 525 г?

Висновки зробіть з тією ж імовірністю.

Розв'язок:

а) Середня вага – це вибіркова середня:

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{1}{10}(520+565+550+540+525+485+515+595+500+555) =$$
$$= 535 \text{ г.}$$

Довірчій імовірності F(x) = 0,95; $k = n - 1 = 9$ за таблицями розподілу Стьюдента відповідає квантиль $t = 1{,}83$. Гранична похибка вибірки з вказаною імовірністю

$$\Delta\bar{x} = t\sqrt{\frac{\sigma^2}{n-1}} = 1{,}83\sqrt{\frac{960}{10-1}} = 18{,}9 \text{ г},$$

$$\text{де } \sigma^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2 = 960.$$

Таким чином, довірчий інтервал складає $\bar{x} - \Delta\bar{x} \le \bar{x}_0 \le \bar{x} + \Delta\bar{x}$, де $\bar{x}$ – вибіркова серія, $\bar{x}_0$ – середня генеральної сукупності.

x_1 = 535 – 18,9 = 516,1 = 516 г, x_2 = 535 + 18,9 = 553,9 = 554 г.

$$516 \text{ г} \le \bar{x}_0 \le 554 \text{ г}.$$

б) фактичні витрати металу відповідають встановленій нормі (525 г), так як це значення належить довірчому інтервалу.

5.2. За результатами вибіркового обстеження 100 домогосподарств, які ведуть індивідуальну забудову, у 24 з них основним джерелом коштів був кредит під заставу нерухомості.

Визначте частку індивідуальних забудовників, які брали кредит під заставу нерухомості, та довірчі межі частки з імовірністю 0,954.

Чи погоджуються вибіркові дані з твердженням, що кожний третій індивідуальний забудовник брав кредит під заставу нерухомості?

Розв'язок:

а) *n* = 100 – обсяг вибірки; *m* = 24 – число забудовників, які скористались кредитом під заставу.

Частка забудовників, які скористались кредитом $p = \frac{m}{n} = 0{,}24$.

Гранична похибка $\Delta p = t\sqrt{\frac{pq}{n}}$, де $q = 1 - p = 0{,}76$.

Довірчій імовірності $F(x) = 0{,}954$; $n = 100$ за таблицями розподілу Стьюдента відповідає квантиль $t = 2{,}0$.

$$\Delta p = 2{,}0\sqrt{\frac{0{,}76 \cdot 0{,}24}{100}} = 0{,}0854,$$

$0{,}24 - 0{,}0854 \le p_0 \le 0{,}24 + 0{,}0854$,

$0{,}1546 \le p_0 \le 0{,}3254$.

Твердження, що кожний третій забудовник брав такий кредит є перебільшенням вказаної частки (її верхнім значенням). З вибіркових даних $\frac{1}{7} < p_0 < \frac{1}{3}$, тобто частка це від кожного сьомого до кожного третього.

5.3. За результатами контрольної перевірки податковими службами 400 бізнесових структур 140 з числа перевірених у податкових деклараціях за минулий рік вказали не всі доходи, які підлягають оподаткуванню.

Визначте частку бізнесових структур, які приховують частину доходів від сплати податків, та довірчі межі частки з імовірністю 0,954.

Чи погоджуються вибіркові дані з твердженням, що 40 % бізнесових структур не сплачують податки у повному обсязі?

Розв'язок:

Обсяг вибіркової сукупності n = 400, з них порушників m = 140.

$$p = \frac{m}{n} = \frac{140}{400} = 0{,}35\,.$$

Довірчій імовірності $F(x)$ = 0,954 та n = 400 за таблицями розподілу Стьюдента відповідає квантиль t = 2,0.

Гранична похибка $\Delta p = t\sqrt{\dfrac{pq}{n}} = 2{,}0\,\dfrac{\sqrt{0{,}35\cdot 0{,}65}}{20} = 0{,}0477\,.$

$$p_1 = 0{,}35 - 0{,}0477 = 0{,}302, \qquad p_2 = 0{,}35 + 0{,}0477 = 0{,}398.$$

$$0{,}302 \le p_0 \le 0{,}398.$$

Твердження, що 40 % бізнесових структур не сплачують податків у повному обсязі не вірне, тому що це число не входить в довірчі межі.

5.4. Хімічний аналіз 10 партій молока дав такі показники кислотності (у градусах Тернера): 18; 21; 17; 19; 20; 23; 16; 22; 24; 21.

Визначте:

а) середній рівень кислотності молока та граничну похибку вибірки для середньої з імовірністю 0,95;

б) частку молока, що відповідає стандартові (не більше 21°), та похибку вибірки для частки з імовірністю 0,95;

в) скільки партій молока необхідно перевірити, щоб похибка вибірки для частки нестандартного молока зменшилась у 2 рази?

Розв'язок:

а) $$\bar{x} = \frac{1}{n}\sum x_i = \frac{1}{10}(18+21+17+19+20+23+16+22+24+21) = 20{,}1\,.$$

$$\sigma^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2 = 6{,}09\,; \qquad \sigma = 2{,}468.$$

Довірчій імовірності $F(x)$ = 0,95; k = 9 за таблицями розподілу Стьюдента відповідає квантиль t = 1,83.

$$\Delta\bar{x} = 1{,}83 \cdot \frac{2{,}468}{3} = 1{,}5\,.$$

$x_1 = 20{,}1 - 1{,}5 = 18{,}6°$, $x_2 = 20{,}1 + 1{,}5 = 21{,}6°$.

$$18{,}6\,° \le \bar{x}_0 \le 21{,}6°.$$

б) m = 7 – число проб з $x \le 21°$

$$p = \frac{m}{n} = 0{,}7 \qquad \Delta p = t\sqrt{\frac{pq}{n}} = 1{,}83\sqrt{\frac{0{,}7 \cdot 0{,}3}{10}} = 0{,}27\,.$$

$x_1 = 0{,}7 - 0{,}27 = 0{,}43$, $x_2 = 0{,}7 + 0{,}27 = 0{,}97$,

$$0{,}43 \le p_0 \le 0{,}97.$$

в) Припускаючи, що при збільшенні об'єму вибірки з $n_1 = 10$ до n_2 p не зміниться, маємо

$$(\Delta p)_1 = t\sqrt{\frac{pq}{n_1}}\ ;\ (\Delta p)_2 = t\sqrt{\frac{pq}{n_2}}\ ; \quad \frac{(\Delta p_1)}{(\Delta p_2)} = \sqrt{\frac{n_1}{n_2}} = 2\,,$$

$n_2 = 4 \cdot n_1 = 40$.

5.5. З різних вагонів вугілля, яке надійшло на електростанцію, з метою визначення його зольності взято 100 проб. Результати аналізу такі:

Таблиця 5.1

Зольність, %	**До 14**	**14-16**	**16-18**	**18-20**	**20 і більше**	**Разом**
Число проб	8	17	36	25	14	100

Визначте:

а) середню зольність вугілля та довірчий інтервал для середньої з ймовірністю 0,954;

б) з тією ж ймовірністю визначте довірчий інтервал частки вугілля, зольність якого менша 16 %.

Розв'язок:

а) Середня зольність

$$\bar{x} = \frac{1}{n}\sum x_i \cdot n_i = \frac{1}{100}\left(13 \cdot 8 + 15 \cdot 17 + 17 \cdot 36 + 19 \cdot 25 + 21 \cdot 14\right) = 17{,}1\%,$$

де x_i – середини інтервалів ($x_1 = 13$, $x_5 = 21$, оскільки інтервали відкриті).

Довірчій імовірності $F(x) = 0{,}954$; $n = 100$ за таблицями розподілу Стьюдента відповідає квантіль $t = 2{,}0$.

$$\Delta\bar{x} = \frac{t\sigma}{\sqrt{n}} = \frac{2{,}0 \cdot 2{,}24}{10} = 0{,}45\%\text{, де } \sigma^2 = 5{,}04\,; \quad \sigma = 2{,}24\ \%.$$

Довірчий інтервал $16{,}65\ \% \le \bar{x}_0 \le 17{,}55\ \%$.

б) Дано: m = 8 + 17 = 25; n = 100, $p = \frac{m}{n} = 0{,}25$;

$$\Delta p = t\sqrt{\frac{pq}{n}} = \frac{2{,}0}{10}\sqrt{0{,}25 \cdot 0{,}75} = 0{,}087.$$

$0{,}163 \le p_0 \le 0{,}337$, або $16{,}3\ \% \le p_0 \le 33{,}7\ \%$.

5.6. За даними 20 %-го вибіркового обстеження 100 сімей переселенців із зони радіаційного контролю, кількість дітей у сім’ях становить:

Таблиця 5.2

Кількість дітей	**0**	**1**	**2**	**3**	**4**	**Разом**
Кількість сімей	11	32	30	20	7	100

Визначте:

а) середню кількість дітей у сім’ях переселенців та довірчий інтервал для середньої з імовірністю 0,954;

б) з тією ж імовірністю визначте граничну похибку та довірчий інтервал для частки сімей, які мають трьох і більше дітей.

Розв’язок:

а) Дано: $\frac{n}{N} = 0{,}2 = 20\%$, n = 100.

$$\bar{x} = \frac{1}{n}\sum x_i \cdot n_i = \frac{1}{100}\left(0 \cdot 11 + 1 \cdot 32 + 2 \cdot 30 + 3 \cdot 20 + 4 \cdot 7\right) = 1{,}8.$$

$$\sigma^2 = \frac{1}{n}\sum n_i x_i^2 - \bar{x}^2 = \frac{1}{100}\left(0^2 \cdot 11 + 1^2 \cdot 32 + 2^2 \cdot 30 + 3^2 \cdot 20 + 4^2 \cdot 7\right) - 1{,}8^2 = 1{,}20;$$

$$\sigma = 1{,}0954.$$

$\Delta = \mu t$, де $\mu = \sqrt{\frac{\sigma^2}{n} \cdot (1 - \frac{n}{N})} = \sqrt{\frac{1{,}2}{100} \cdot (1 - 0{,}2)} = 0{,}098 \cdot$

Довірчій імовірності $F(x)$ = 0,954; n = 100 за таблицями розподілу Стьюдента відповідає квантиль t = 2,0.

$\Delta\bar{x} = 2{,}0 \cdot 0{,}098 = 0{,}196$; x_1 = 1,8 – 0,196 = 1,604; x_2 = 1,8 + 0,196 =
= 1,998 = 2,0.

$$1{,}6 \le \bar{x}_0 \le 2{,}0.$$

б) m = 20 + 7 = 27. $p = \frac{m}{n} = \frac{27}{100} = 0{,}27$, t=2.

$$\Delta p = t\sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)} = 2\sqrt{\frac{0{,}27\cdot 0{,}73}{100}\left(1-0{,}2\right)} = 0{,}08\ .$$

$$0{,}19 \le p_0 \le 0{,}35.$$

5.7. За даними 5 %-го вибіркового обстеження, верстати за терміном служби розподіляються так:

Таблиця 5.3

Термін служби, років	**До 4**	**4-8**	**8-12**	**12 і більше**	**Разом**
Кількість верстатів	25	40	20	15	100

Визначте:

а) середній термін служби верстатів та довірчий інтервал для середньої з імовірністю 0,954;

б) з тією ж імовірністю визначте граничну похибку та довірчий інтервал частки верстатів, що мають термін служби понад 12 років.

Розв'язок:

Дано: $\frac{n}{N} = 0{,}05 = 5\%$, n = 100.

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{2\cdot 25 + 6\cdot 40 + 10\cdot 20 + 14\cdot 15}{100} = 7.$$

$$\sigma^2 = \frac{\sum\left(\bar{x}-x_i\right)^2 f_i}{\sum f_i} = \frac{(7-2)^2\cdot 25 + (7-6)^2\cdot 40 + (7-10)^2\cdot 20 + (7-14)^2\cdot 15}{100} = 835.$$

$$\mu_x = \sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)} = \sqrt{\frac{835}{100}\left(1-\frac{100}{2000}\right)} = 2{,}82.$$

Довірчій імовірності $F(x)$ = 0,954; n = 100 за таблицями розподілу Стьюдента відповідає квантиль t = 2,0.

$\Delta\bar{x} = t\cdot\mu = 2{,}0\cdot 2{,}82 = 5{,}64$; x_1 =7-5,64=1,36; x_2 =7+5,64=12,64.

$$1{,}36 \le \bar{x}_0 \le 12{,}64$$

б) m = 15, $p = \frac{m}{n} = \frac{15}{100} = 0{,}15$, t = 2,

$$\Delta p = t\sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)} = 2\sqrt{\frac{0{,}15\cdot 0{,}85}{100}\left(1-0{,}05\right)} = 0{,}07\ ,$$

$$0{,}08 \le p_0 \le 0{,}22.$$

5.8. На молочній фермі з поголов’ям 2250 корів проведена контрольна перевірка добового надою та жирності молока (16 проб). За даними перевірки маємо:

Таблиця 5.4

Показник	Вибіркова сукупність	Середній рівень	Середнє квадратичне відхилення
Середній добовий надій молока на одну корову, кг	25	18	4,5
Середня жирність молока, %	16	3,8	0,4

а) з імовірністю 0,95 визначте відносні похибки вибірки для середнього надою молока та його жирності;

б) порівняйте похибки, зробіть висновки.

Розв’язок:

$$\mu_x = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)},$$

$$V_\mu = \frac{\mu_x}{\bar{x}} \cdot 100\%$$ – відносна похибка.

1) знаходимо відносну похибку для середнього добового надою молока на одну корову:

$$\mu_x = \sqrt{\frac{4{,}5^2}{25}\left(1 - \frac{25}{2250}\right)} = \sqrt{0{,}81 \cdot 0{,}99} = 0{,}895.$$

$$V_\mu = \frac{0{,}895}{18} 100\% = 4{,}97\%.$$

2) знаходимо відносну похибку для середньої жирності молока:

$$\mu_x = \sqrt{\frac{0{,}4^2}{16}\left(1 - \frac{16}{2250}\right)} = \sqrt{0{,}01 \cdot 0{,}993} = 0{,}996.$$

$$V_\mu = \frac{0{,}996}{3{,}8} 100\% = 26{,}21\%.$$

Відносна похибка для середнього добового надою молока = 4,97 %, а його жирності = 26,21 %. Ми можемо сказати, що середній надій молока на добу легше визначити з малою похибкою, що не можна сказати про його жирність. Ще можна сказати, що обстеживши лише 16 надоїв, ми не можемо казати про взагалі весь надій за добу, тому що результат буде з великою похибкою.

5.9. За даними 20 %-го вибіркового обстеження домогосподарств витрати населення області на побутові послуги становили:

Таблиця 5.5

Тип поселення	Число обстежених домогосподарств	Витрати одного члена, гр.од. на місяць	Дисперсія витрат
Місто	36	120	324
Село	64	50	196

Для кожного типу поселення визначте: відносні похибки вибірки з імовірністю 0,954, порівняйте похибки, зробіть висновки.

Розв'язок:

$$\mu_x = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)},$$

$$V_\mu = \frac{\mu_x}{\bar{x}} \cdot 100\%$$ – відносна похибка.

1) для міста:

$$\mu_x = \sqrt{\frac{324}{36}\left(1 - \frac{36}{180}\right)} = \sqrt{9 \cdot 0{,}8} = 2{,}68.$$

$$V\mu = \frac{2{,}68}{120}100\% = 2{,}23\%.$$

2) для села:

$$\mu_x = \sqrt{\frac{196}{64}\left(1 - \frac{64}{320}\right)} = \sqrt{3{,}0625 \cdot 0{,}8} = 1{,}57.$$

$$V\mu = \frac{1{,}56}{50}100\% = 3{,}12\%.$$

Ми можемо зробити висновок, що домогосподарства села обстежуються більш точніше, ніж міста. Оскільки $V\mu$ для міста дорівнює 2,23 % при обстеженні 36 домогосподарств, $V\mu$ для села дорівнює 3,12 % при обстеженні 64 домогосподарств.

5.10. За результатами технічного аналізу проб руди трьох родовищ вміст заліза у руді становить:

Таблиця 5.6

Родовище	Середній вміст заліза, %	Стандартна (середня) похибка вибірки, %
A	52	1,56
B	58	2,61
C	55	1,98

Визначте відносні похибки вибірки з імовірністю 0,954, порівняйте їх.

Розв'язок:

$$V_\mu = \frac{\mu}{\bar{x}} 100\%,$$

А) $V_\mu = \dfrac{1{,}56}{52} 100\% = 3\%.$

В) $V_\mu = \dfrac{2{,}61}{58} 100\% = 4{,}5\%.$

С) $V_\mu = \dfrac{1{,}98}{55} 100\% = 3{,}6\%.$

5.11. За даними вибіркового обстеження машинобудівних підприємств зафіксовано значне недовикористання робочих місць:

Таблиця 5.7

Галузь машинобудування	Число спостережень	Незайнятих робочих місць у першу зміну, %
Верстатобудівна	76	40
Приладобудівна	56	30
Автомобільна	94	15
Сільськогосподарського машинобудування	128	20

Для кожної галузі машинобудування визначте відносну похибку вибірки незайнятих робочих місць, порівняйте їх, зробіть висновки.

Розв'язок:

Відносна похибка частки (долі) p

$$V_p = \frac{\mu_p}{p} = \sqrt{\frac{q}{np}}, \qquad \text{де } q = 1 - p.$$

Таблиця

Розрахунок відносної похибки вибірки

i	n_i	p_i	q_i	Vp_i	%
1	76	0,40	0,60	0,14	14
2	56	0,30	0,70	0,204	20,4
3	94	0,15	0,85	0,246	24,6
4	128	0,20	0,80	0,177	17,7

За інших однакових умов, чим менша частка p, тим більша її похибка V_p.

5.12. За даними 2 %-го вибіркового обстеження 400 актів прийому будівельних об’єктів частка дефектів будівельно-монтажних робіт становила:

бетонні та залізобетонні роботи – 0,10;
цегляна кладка – 0,20;
штукатурні роботи – 0,30;
малярні роботи – 0,25.

Для кожного виду будівельних робіт визначте відносні похибки вибірки з імовірністю 0,954, порівняйте їх, зробіть висновки.

Розв’язок:

$$\frac{n}{N} = 0{,}02 = 2\% \text{ ; } n = 400\text{; N} = 20000\text{;}$$

$$V_{\mu} = \frac{\mu_{\partial}}{p}\text{;}$$

$$\mu = \sqrt{\frac{pq}{n}\left(1 - \frac{n}{N}\right)}\text{,}$$

Таблиця

Розрахунок відносної похибки вибірки

i	***p***	***q***	μ	***V***	**%**
1	0,10	0,90	0,015	0,15	15
2	0,20	0,80	0,02	0,1	10
3	0,30	0,70	0,023	0,077	7,7
4	0,25	0,75	0,021	0,028	2,8

5.13. Результати вибіркового опитування 400 мешканців міста з метою визначення їхньої думки відносно подальшої долі екологічно шкідливого об’єкта такі:

Таблиця 5.9

Захід щодо об’єкта	**Число відповідей**
Негайно закрити	200
Перепрофілювати на виробництво іншої продукції	120
Збудувати нові очисні споруди	80

Для кожної відповіді респондентів визначте відносну похибку вибірки з імовірністю 0,954, порівняйте їх, зробіть висновки.

Розв’язок:

$$V_{\mu} = \sqrt{\frac{q}{np}}\text{,} \qquad \text{де } n = 400\text{,}$$

Таблиця

Розрахунок відносної похибки вибірки

i	m_i	$p_i = m_i/n$	q_i	Vp_i	%
1	200	0,5	0,5	0,05	5
2	120	0,3	0,7	0,076	7,6
3	80	0,2	0,8	0,10	10

Порівняно великі похибки V_p обумовлені малим об'ємом вибірки n.

5.14. За даними опитування із 225 респондентів основними джерелами інформації про ринок цінних паперів вважають:

радіо та телебачення – 170;

газети та журнали – 90.

Для кожного джерела інформації визначте його частку та відносну похибку вибірки з імовірністю 0,954. Порівняйте похибки вибірки.

Розв'язок:

$$n = 225;$$

$$m_1 = 170; \qquad \text{Частка } p_1 = \frac{m_1}{n} = \frac{170}{225} = 0{,}756.$$

$$q_1 = 0{,}244.$$

$$V_{p1} = \sqrt{\frac{q}{np_1}} = \sqrt{\frac{0{,}244}{225 \cdot 0{,}756}} = 3{,}8\%.$$

$$m_2 = 90; \text{ частка } p_2 = \frac{m_2}{n} = \frac{90}{225} = 0{,}4.$$

$$q_2 = 0{,}6.$$

$$V_{p2} = \sqrt{\frac{q}{np_2}} = \sqrt{\frac{0{,}6}{225 \cdot 0{,}4}} = 8{,}15\%.$$

5.15. З метою визначення потенціалу споживчого ринку планується анкетування населення (на одну квартиру – одна анкета). Визначте мінімально необхідний обсяг вибірки, щоб гранична похибка вибірки (з імовірністю 0,954) для середньомісячного розміру покупки не перевищувала 5 гр.од. За даними пробних обстежень дисперсія середньомісячного розміру покупок становить 1875.

Розв'язок:

$$\Delta = \frac{t\sigma}{\sqrt{n}}.$$

$$\Delta = 5; \quad \sigma^2 = 1875; \; F(x) = 0{,}954; \; t = 2.$$

$$n_{\min} = \frac{t^2\sigma^2}{\Delta^2} = \frac{2^2 \cdot 1875}{5^2} = 300.$$

5.16. На лісовому масиві в 400 га передбачається визначити загальний запас деревини. Пробні площі становлять 0,1 га. За даними попередніх обстежень дисперсія виходу деревини з 0,1 га становила 6. Скільки пробних площ необхідно обстежити, щоб похибка вибірки з імовірністю 0,954 не перевищила 0,5 м^3?

Розв'язок:

Скористаємось формулою для безповторної вибірки:

$$\Delta = t\sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}$$ в припущенні, що $n \ll N$.

$$\Delta = t\sqrt{\frac{\sigma^2}{n}}, \quad n = \frac{t^2 \cdot \sigma^2}{\Delta^2}.$$

$$t = 2; \quad \sigma^2 = 6; \quad \Delta = 0{,}5;$$

$$n = \frac{2^2 \cdot 6}{0{,}5^2} = 96.$$

5.17. Проектується опитування підприємств з приводу оцінки економіко-правових умов їхньої діяльності. Визначте обсяг вибіркової сукупності, щоб з імовірністю 0,954 відносна похибка вибірки не перевищила 10 %. За результатами попереднього опитування 80 % підприємців вважали умови діяльності несприятливими.

Розв'язок:

Дано: $t = 2$; $V_p \leq 10$ %. Визначте n.

$q = 0{,}8$; $p = 1 - q = 1 - 0{,}8 = 0{,}2$;

$\sigma^2 = pq = 0{,}2 \cdot 0{,}8 = 0{,}16$ – за результатами попереднього опитування.

$$\Delta = t\sqrt{\frac{\sigma^2}{n}}, \qquad n = \frac{t^2\sigma^2}{\Delta^2} = \frac{2^2 \cdot 0{,}16}{0{,}1^2} = 64.$$

5.18. На винограднику 15 000 кущів. З метою визначення ступеня ураженості виноградника шкідниками проведено вибіркове обстеження. Виноградник поділено на 100 однакових ділянок по 100 кущів кожна і методом випадкової вибірки відібрано 10 ділянок. Результати їх суцільного обстеження (серійна вибірка) такі:

Таблиця 5.10

№ ділянки	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Ураженість, %	8	14	9	11	7	9	6	12	13	11

Визначте:

а) ступінь ураженості винограднику в середньому, похибку вибірки та межі довірчого інтервалу середньої з імовірністю 0,95;

б) скільки ділянок необхідно обстежити, щоб похибка вибірки для ступеня ураженості винограднику з тією ж імовірністю не перевищувала 5 %?

Розв'язок:

x – ступінь ураженості.

а) $\bar{x} = \frac{1}{n}\sum x_i \cdot n_i = \frac{1}{10}\left(8+14+9+11+7+9+6+12+13+11\right) =$

$= \frac{100}{10}\% = 10\%;$

$$\sigma^2 = \frac{\sum_1^n (x_i - \bar{x})^2}{n} = \frac{(10-8)^2 + (10-14)^2 + (10-9)^2 + \ldots + (10-11)^2}{10} =$$

= 6,2;

$$\mu_x = \sqrt{\frac{\sigma^2}{n}(1-\frac{n}{N})} = \sqrt{\frac{6,2}{10}(1-\frac{10}{100})} = \sqrt{6,2\times 0,9} = \sqrt{0,558} = 0,75;$$

$$\Delta = \mu_x t = 0,75*2 = 1,5\ \%;$$

$$\bar{x} - \Delta x \le \bar{x}_0 \le \bar{x} + \Delta x;$$

$$10\ \% - 1,5\ \% \le \bar{x}_0 \le 10\% + 1,5\%;$$

$$8,5\ \% \le \bar{x}_0 \le 11,5\ \%.$$

б) $n = \frac{t^2\sigma^2}{\Delta^2} = \frac{2^2 \cdot 6,2}{0,05^2} = \frac{4 \cdot 6,2}{0,0025} = 9920$ кущів.

5.19. Проведено вибіркове обстеження втрат зерна озимої пшениці через несвоєчасне збирання врожаю. Кількість пробних ділянок визначалась пропорційно посіву відповідного сорту пшениці (розшарована вибірка). Результати обстеження такі:

Таблиця 5.11

Сорт пшениці	Кількість пробних ділянок	Втрати зерна, ц/га	Дисперсія втрат зерна
A	10	2	6,4
B	6	7	7,8
C	4	9	10,3

Визначте:

а) середні втрати зерна у розрахунку на одну пробну ділянку та довірчі межі середніх втрат з імовірністю 0,954;

б) мінімально достатній обсяг вибірки, при якому похибка вибірки з тією самою імовірністю не перевищить 1 ц/га.

Розв'язок:

x – втрати зерна.

а) $\bar{x} = \frac{1}{n}\sum x_i \cdot n_i = \frac{1}{20}(2 \cdot 10 + 6 \cdot 7 + 4 \cdot 9) = 4{,}9$

$$\sigma^2 = \frac{1}{n}\sum \sigma_i^2 \cdot n_i = \frac{1}{20}(10 \cdot 0{,}4 + 6 \cdot 7{,}8 + 4 \cdot 10{,}3) = 7{,}6;$$

$\sigma = 2{,}76$,

$$\Delta = \frac{t \cdot \sigma}{\sqrt{n-1}} = \frac{2 \cdot 2{,}76}{\sqrt{20-1}} = 1{,}265, \quad 3{,}64 \le \bar{x}_0 \le 6{,}17.$$

б) $\Delta = \frac{t \cdot \sigma}{\sqrt{n-1}}$, $\qquad n = 1 + \left(\frac{t\sigma}{\Delta}\right)^2 = 1 + \frac{2^2 \cdot 7{,}6}{1^2} = 31{,}4 = 32$.

5.20. За даними 20 %-го вибіркового обстеження домогосподарств, витрати часу жінками на домашні справи значною мірою залежать від наявності дітей до 12 років:

Таблиця 5.12

Кількість дітей	Кількість домогосподарств	Витрати часу на домашні справи, годин на тиждень	Середнє квадратичне відхилення витрат часу, год.
Без дітей	200	26	10
Одна дитина	120	30	20
Дві і більше	80	35	15

Визначте: середній тижневий розмір витрат часу на домашні справи для всієї сукупності обстежених домогосподарств та довірчі межі для середньої з імовірністю 0,954.

Розв'язок:

x – витрати часу.

$$\bar{x} = \frac{1}{n}\sum x_i \cdot n_i = \frac{1}{400}(26 \cdot 200 + 30 \cdot 120 + 35 \cdot 80) = 29.$$

$$\sigma^2 = \frac{1}{n}\sum \sigma_i^2 \cdot n_i = \frac{1}{400}(100 \cdot 200 + 400 \cdot 120 + 225 \cdot 80) = 215\ ;$$

$$\sigma = \sqrt{215} = 14{,}7\ .$$

Розрахунок внутрішньогрупових дисперсій виконано в таблиці.

Таблиця

Допоміжна таблиця

z	n_i	x_i	σ_i	σ_i^2
0	200	26	10	100
1	120	30	20	400
2	80	35	15	225
$\sum$	400			

$$\Delta = \frac{t\sigma}{\sqrt{n}}\sqrt{\left(1 - \frac{n}{N}\right)} = \frac{2 \cdot 14{,}66}{\sqrt{400}} \cdot \sqrt{1 - 0{,}8} = 1{,}31\ ;$$

$x_1 = 29 – 1{,}31 = 27{,}7$; $x_2 = 29 + 1{,}31 = 30{,}3$.

27,7 год. $\le x_0 \le$ 30,3 год.

5.21. Вибірково обстежено посіви озимої пшениці, визначена частка зимової загибелі посівів:

Таблиця 5.14

Сорт пшениці	**Обстежена площа, га**	**Частка зимової загибелі посівів, %**
А	6	10
В	4	15

Визначте:

а) середню частку зимової загибелі посівів;

б) граничну похибку вибірки для частки та довірчий інтервал з імовірністю 0,954;

в) з якою імовірністю можна стверджувати, що частка зимової загибелі посівів не перевищує 22 %?

Розв'язок:

а) x – частка загибелі.

$$\bar{x} = \frac{1}{n}\sum x_i \cdot n_i = \frac{120}{10} = 12\%.$$

$$\sigma^2 = \frac{1}{n}\sum x_i^2 \cdot n_i - \bar{x}^2 = \frac{1}{10}\left(6 \cdot 10^2 + 4 \cdot 15^2\right) - 12^2 = 6,$$

$$\sigma = 2{,}45\ \%.$$

б)

$$\Delta = \frac{t \cdot \sigma}{\sqrt{n}} = \frac{2 \cdot 2{,}45}{\sqrt{10}} = 1{,}55\%, \qquad 10{,}45\ \% \le x_0 \le 13{,}55\ \%.$$

в) з якою імовірністю можна стверджувати, що частка зимової загибелі посівів не перевищує 22 %:

$$\bar{\tilde{o}} = 12\%, \Delta\tilde{o} \le 10; \text{ то } \Delta x = \mu_x t \text{, тоді } t = \frac{\Delta x}{\mu} = \frac{10}{0{,}147} = 68{,}03\ \%.$$

Отже, з імовірністю 68,03 % можна стверджувати, що частка зимової загибелі посівів не перевищує 22 %.

5.22. З метою перевірки ефективності нового пристрою для вдосконалення технології обробки деталей робітників бригади розділили на 2 групи: на верстатах з новим пристроєм працювало 4 робітника, за традиційною технологією обробки деталей – 6. Результати експерименту такі: у І групі середній виробіток за годину становить 84 деталі при σ^2 = 10, у ІІ – 77 деталей при σ^2 = 8.

Сформулюйте нульову та альтернативну гіпотези. За допомогою *t*-критерію Стьюдента з рівнем істотності α = 0,05 зробіть висновок, чи дійсно нова технологія впливає на продуктивність праці.

Розв'язок:

Нульова гіпотеза H_0: $\bar{x}_1 = \bar{x}_2$, де $\bar{x}$ – продуктивність праці, тобто різниця між продуктивністю в 1-й та 2-й групах неістотна, і вони рівні;

Альтернативна гіпотеза H_a: $\bar{x}_1 > \bar{x}_2$ – нова технологія більш ефективна.

t – статистика для розподілу Стьюдента.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

$n_1 = 4$; $x_1 = 84$; $\sigma_1^2 = 10$; $n_2 = 6$; $x_2 = 77$; $\sigma_2^2 = 8$.

$$\bar{\sigma} = \frac{\sigma_1^2 \cdot n_1 + \sigma_2^2 \cdot n_2}{n_1 + n_2 - 2} = \frac{10{,}4 + 8{,}6}{4 + 6 - 2} = \frac{88}{8} = 11.$$

$$t = \frac{84 - 77}{\sqrt{11 \cdot \left(\frac{1}{4} + \frac{1}{6} \right)}} = 3{,}27.$$

З табл. 2 при $F(x)$ = 0,95; k = 8 $\Rightarrow$ $t_{кр}$ = 1,86.

Так як $t_{факт} > t_{кр}$, то нульова гіпотеза H_0: $\bar{x}_1 > \bar{x}_2$ відхиляється, приймається альтернативна H_a: $\bar{x}_1 > \bar{x}_2$ – нова технологія ефективніша.