ВИБІРКОВЕ СПОСТЕРЕЖЕННЯ

Поняття про вибіркове спостереження

Суцільне спостереження - вивчення всіх одиниць сукупності - *недоцільно (неможливо).*

Генеральна сукупність - *в*сі одиниці явища.

Вибіркова сукупність - окрема частина одиниць, явища, відібраних із генеральної сукупності для безпосереднього спостереження.

УМОВИ ВІДБОРУ

а) ***випадковість***, тобто кожна одиниця повинна мати таку ж ймовірність потрапити у вибірку (так, наприклад, відібрані найкращі або найгірші одиниці не відображають дійсний розподіл ознаки в генеральній сукупності);

б) **однорідність сукупності**, так як за інших обставин результати вибірки будуть не точними і не можуть в повній мірі репрезентувати генеральну сукупність.

Підходи до формування вибірки

1) відбір при жеребкуванні заздалегідь занумерованих одиниць генеральної сукупності;

1.а. *повторна вибірка* - відібрана з генеральної сукупності занумерована одиниця фіксується і знов повертається на своє місце, після чого пачка номерів одиниць генеральної сукупності ретельно перемішується; цей спосіб відбору на практиці є обмеженим із-за недоцільності, а іноді й неможливості повторного обстеження;

1.б. *безповторна вибірка*, коли відібраний із пачки номер одиниці генеральної сукупності відкладається в сторону і не повертається назад в пачку; цей спосіб відбору характеризується підвищеним ступенем точності, надійності вибірки і найчастіше використовується на практиці.

2) використання *таблиць випадкових чисел.*

Із таблиці випадкових чисел відбирають *n* чисел із любого рядка або стовпця таблиці, кількість яких не перевищує *N* чисел генеральної сукупності; потім відбирають будь-яким способом ті одиниці заздалегідь занумерованої сукупності із *n* чисел, які відповідають відібраним числам таблиці, що і складає вибіркову сукупність.

Різновиди вибірки

а) за способом організації вибіркового обстеження:

- проста випадкова вибірка;
- механічна вибірка;
- районована (типова) вибірка;
- серійна вибірка;
- ступенева вибірка.

б) за ступенем охоплення одиниць обстежуваної сукупності.

- великі (при $n \geq 30$);
- малі (при $n < 30$).

Характеристики генеральної та вибіркової сукупностей

Генеральна сукупність

Варіанти x	x_1	x_2	…	x_i	…	x_M	x
Частоти F	F_1	F_2	…	F_i	…	F_M	$\sum_{i=1}^{M} F_i = N$

1) *генеральна середня* : $\bar{x} = \dfrac{\sum x_i F_i}{N}$

2) *генеральна дисперсія :* $\sigma^2 = \dfrac{\sum (x_i - \bar{x})^2 F_i}{N}$

3) *Генеральне СКВ*

4) *частка ознаки одиниць генеральної сукупності р, тобто частка одиниць М, яка володіє даним значенням ознаки в загальному обсягу N генеральної сукупності:* *p=M/N.*

Характеристики генеральної та вибіркової сукупностей

Вибіркова сукупність

Варіанти x	x_1	x_2	…	x_i	…	x_M	x
Частоти F	f_1	f_2	…	f_i	…	f_M	$\sum_{i=1}^{M} f_i = N$

1) *вибіркова середня* : $\tilde{x} = \dfrac{\sum x_i f_i}{n}$

2) вибіркова *дисперсія :* $\sigma_в^2 = \dfrac{\sum (x_i - \tilde{x})^2 f_i}{n}$

3) *Вибіркове СКВ*

4) *частка ознаки одиниць вибіркової сукупності w,* тобто відношення кількості одиниць вибіркової сукупності *m,* яка володіє даною ознакою, до обсягу вибіркової сукупності n*:*

$$w = m/n$$

5) *частка вибірки* $w_в$, як відношення обсягу вибірки до обсягу генеральної сукупності $w_в = n/N$

Закон великих чисел

- *Суть* закону великих чисел полягає в тому, що при збільшенні чисельності одиниць сукупності поступово зменшується елемент випадковості в узагальнених характеристиках сукупності. На основі закону можна стверджувати, що при достатньо великому обсязі вибірки (n≥30) вибіркові характеристики мало відрізняються від генеральних, внаслідок чого використовуються наближені рівняння для середньої, частки, дисперсії, середньому квадратичному відхиленні:

$$p \approx \omega$$

$$\bar{x} \approx \tilde{x} \qquad \sigma \approx \sigma_{в}$$

Проста випадкова вибірка

1)визначення помилки вибіркового спостереження;

2)визначення меж генеральних характеристик на основі вибіркових із заданою довірчою ймовірністю (ступенем надійності);

3)визначення довірчої ймовірності тому, що генеральні характеристики можуть відрізнятися від вибіркових не більш певної заданої величини;

4)знаходження необхідної чисельності вибірки, яка б з практичною достовірністю забезпечувала задану точність вибіркових характеристик.

Вирішення зазначених задач може проводитись як по відношенню до генеральної середньої арифметичної $\bar{x}$, так і до частки $\bar{p}$.

Задача 1. при великій кількості одиниць вибіркової сукупності (n ≥30) середня квадратична помилка безповторної вибірки:

$$\mu_x = \sqrt{\frac{\sigma_в^2}{n}\left(1-\frac{n}{N}\right)} \qquad \mu_w = \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}$$

На основі теореми Ляпунова *гранична помилка вибірки* $\Delta = t\mu$ *, де t - коєфіцієнт довіри*

Імовірність *F(x)*	Коефіцієнт довіри ***t***
0,683	1
0,954	2
0,997	3
0,999	4

Задача 2. Величини генеральної середньої та частки можуть бути представлені інтервальною оцінкою у вигляді визначення довірчого інтервалу із заданого рівня довірчої ймовірності Р: $\tilde{x} - \Delta_x \leq \bar{x} \leq \tilde{x} + \Delta_x$ $\quad w - \Delta_w \leq p \leq w + \Delta_w$

Ймовірність того, що величина генеральної середньої або частки вийде за довірчі межі, дорівнює (α = 1 – Р) і називається рівнем значущості (істотності).

Для ймовірності Р = 0,950 або Р = 0,954 рівень значущості дорівнює відповідно 0,050 (або 5,0 %) та 0,046 (або 4,6 %), і перевищення меж у довірчих інтервалах (5.9), (5.10), що має таку ймовірність, практично неможливе.

Задача 3. Довірча ймовірність *P,* яку необхідно обчислити за теоремою Ляпунова, є функцією від коефіцієнта *t: P* = Ф(*t*) *,* де Ф(*t*) – інтеграл Лапласа .

$$\Phi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt$$

Значення *t* у свою чергу може бути визначено через граничну та стандартну помилки $t = \Delta / \mu$, обчисленими відносно середньої або частки.

Нарешті, за найденим значенням *t* із довідкових таблиць знаходиться інтеграл Лапласа, який відповідає розшукуваній ймовірності *P,* яка порівнюється із заданою величиною.

Задача 4. Чисельність вибірки

Для середньої - $$n = \frac{t^2 \sigma_{\text{в}}^2 N}{\Delta_x^2 N + t^2 \sigma_{\text{в}}^2}$$

$$n = \frac{t^2 w(1-w)N}{\Delta_w^2 N + t^2 w(1-w)}$$

Механічна вибірка

Це вибірка, при якій генеральна сукупність обсягом N одиниць, розташованих певному порядку (за зростанням або зменшенням, за алфавітом географічним положенням тощо), розділяються на n рівних частин і з кожної частини обстежується одна одиниця.

Відношення N/n – називається інтервалом вибірки.

За початок відрахунку приймають або початкову одиницю, або середину першого інтервалу.

Помилки вибірки при механічному відборі одиниць обчислюють за формулами простої випадкової безповторної вибірки.

Районована (типова) вибірка

Спосіб відбору на основі розподілу кількості відібраних одиниць n між районами (групами), які є в генеральній сукупності.

Якщо генеральна сукупність розділяється на ***m*** частин, груп, районів тощо, тобто N = N1+N2+...+Ni+Nm, то і вибіркова сукупність повинна формуватися із ***m*** частин так, щоб ***n = n1+n2+...+ni+…+nm***.

Розподіл між районами:

- пропорційний (до питомої ваги району в генеральній сукупності) $n_i = n\frac{N_i}{N}$
- непропорційний (однакова кількість одиниць з кожного району) $n_i = \frac{n}{k}$

  k – кількість відокремлюваних районів
- оптимальним, яке враховує і чисельність району ***Ni,*** і середнє квад-ратичне відхилення ознаки в районі

$$n_i = n\frac{\sigma_i N_i}{\sum \sigma_i N_i}$$

На практиці в більшості випадків застосовують перший і третій способи розподілення між районами.

Серійна вибірка

Відбору підлягають окремі серії (групи) одиниць генеральної сукупності. У відібраних методом випадкового безповторного або механічного відбору серіях проводять суцільне спостереження всіх одиниць, що до них увійшли.

Оскільки при серійній вибірці кожна серія виступає як самостійна одиниця спостереження, то дисперсія усередині серій у разі визначення середньої помилки та чисельності вибірки має бути виключена і враховується тільки міжсерійна дисперсія

При рівних серіях середня квадратична помилка безповторної вибірки та її чисельність

$$\mu = \sqrt{\frac{\delta^2}{r}\left(1 - \frac{r}{R}\right)}; \qquad r = \frac{t^2\delta^2 R}{R\Delta^2 + t^2\delta^2},$$

де r – кількість відібраних серій; R – загальна кількість серій в генеральній сукупності.

Міжсерійна дисперсія

$$\delta^2 = \frac{\sum\left(\tilde{x}_i - \tilde{x}_0\right)^2}{r}; \qquad \delta^2 = \frac{\sum\left(w_i - \overline{w}\right)^2}{r},$$

Метод моментних спостережень

це фіксація стану безперервного процесу на певні моменти часу.

Середня помилка частки w визначається як для простої повторної випадкової вибірки

$$\mu_w = \sqrt{\frac{w(1-w)}{n}}$$

Гранична помилка вибірки Δ= μ $t_{м.в.}$ розраховується з використанням значення коефіцієнта довіри t, який визначається прийнятим рівнем довірчої ймовірності P.

Для визначення *чисельності* моментних спостережень рекомендується до використання формула повторної випадкової вибірки, де не враховується чисельність одиниць *N* генеральної сукупності

$$n = \frac{\sigma^2 t^2}{\Delta^2}$$

Зазвичай приймають максимальне значення дисперсії = 0,25. Якщо довірча ймовірність *P*=0,954, то при *t* = 2,0 чисельність вибірки за методом моментних спостережень:

$$n = 0,25 \cdot \frac{2^2}{\Delta^2} = \frac{1}{\Delta^2}.$$