

Выборочное наблюдение



Понятие выборочного наблюдения.

Выборочное наблюдение – это такой вид статистического наблюдения, при котором обследованию подвергается не вся изучаемая совокупность, а лишь часть ее единиц, отобранных в определенном порядке.

- Вся исследуемая совокупность называется генеральной;
- Единицы подлежащие наблюдению составляют выборочную совокупность или выборку.

Условные обозначения

	Генеральные параметры	Выборочные показатели
Объем совокупности	N	n
Средняя величина	μ	\bar{x}
Относительная величина	π	p
Дисперсия	σ^2	s^2

Цель выборочного наблюдения

- Определение параметров генеральной совокупности на основе показателей выборочной совокупности.

Выборочный метод обладает следующими достоинствами:

- относительно небольшие (по сравнению со сплошным наблюдением) материальные, трудовые, стоимостные затраты на сбор данных;
- оперативность получения результатов;
- широкая область применения;
- высокая достоверность результатов.

- Выборочные оценки отличаются от генеральных параметров за счет ошибки наблюдения и ошибки выборки:

Выборочная оценка = генеральный параметр \pm
 \pm ошибка наблюдения \pm ошибка выборки

Различают два вида отбора

– **повторный**

соответствует схеме «возвращенного шара».

- бесповторный. Бесповторная выборка соответствует схеме «невозвращенного шара».

Возможны три способа отбора:

- случайный;
- отбор единиц по определенной схеме;
- сочетание первого и второго способов.

Различают следующие виды выборочного наблюдения:

- Типическая (расслоенная или стратифицированная)
- Серийная (гнездовая)
- Многоступенчатая
- Многофазовая

Определение ошибки выборки

- средняя (стандартная),
- предельная
- относительная

При случайном и механическом отборах

- средняя ошибка выборки для средней величины ($\mu_{\bar{x}}$) при повторном отборе:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$$

- При бесповторном отборе:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$$

На практике величина дисперсии признака в генеральной совокупности σ^2 как правило неизвестна, поэтому ее заменяют выборочной дисперсией s^2 . Это возможно, поскольку доказано, что соотношение σ^2 и s^2 определяется равенством:

$$\sigma^2 = s^2 \frac{n}{n-1}$$

При большой численности выборочной совокупности сомножитель

$$\frac{n}{n-1}$$

стремится к единице и им можно пренебречь.

Величина дисперсии доли в генеральной совокупности определяется по формуле:

$$\sigma_{\text{доли}}^2 = p(1 - p)$$

где p – доля единиц, обладающих каким-либо значением признака в генеральной совокупности.

При расчете средней ошибки выборочной доли дисперсия доли в генеральной совокупности, как правило, тоже неизвестна, поэтому ее заменяют дисперсией доли в выборочной совокупности:

$$s_{доли}^2 = w(1 - w)$$

где w – доля единиц, обладающих каким-либо значением признака в выборочной совокупности.

Формула для расчета средней ошибки
выборочной доли для повторного отбора

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

Формула для расчета средней ошибки
выборочной доли для бесповторного отбора

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

Предельная ошибка выборки

$$\Delta = t\mu$$

где t – коэффициент доверия, который определяется по таблице значений интегральной функции Лапласа при заданной доверительной вероятности.

Наиболее часто употребляемые уровни доверительной вероятности и соответствующие им значения t :

$P(t)$	0,683	0,95	0,954	0,99	0,997
t	1,00	1,96	2,00	2,58	3,00

- Зная величину выборочной средней (\bar{x}) или доли (w), а также предельную ошибку выборки (Δ), можно определить доверительные интервалы, в которых находятся значения генеральных параметров:

$$\bar{x} - \Delta \leq \bar{X} \leq \bar{x} + \Delta$$

$$w - \Delta \leq p \leq w + \Delta$$

- **Пример:**

Для определения среднего срока пользования краткосрочным кредитом в банке была произведена 5%-ая механическая выборка, в которую попали 200 счетов. По результатам выборки установлено, что средний срок пользования кредитом составляет 60 дней при среднеквадратическом отклонении 20 дней. В 8 счетах срок пользования кредита превышал 6 месяцев. Необходимо с вероятностью 0,99 определить пределы, в которых находится срок пользования краткосрочными кредитами банка и доля краткосрочных кредитов со сроком пользования более полугода.

Нахождения необходимой численности выборки

- на практике расчет объема выборки производят по формуле для повторного отбора:

$$n_{\text{повт}} = \frac{t^2 \sigma^2}{\Delta^2}$$

- Если полученный объем выборки превышает 5% численности генеральной совокупности, расчеты корректируют «на бесповторность»:

$$n_{\text{бесповт}} = \frac{t^2 \sigma^2 N}{t^2 \sigma^2 + \Delta^2 N}$$

- При решении задачи определения объема выборки величина допустимой предельной ошибки и уровень вероятности, гарантирующей точность оценок будущей выборки, задаются исследователем

Для оценки величины генеральной дисперсии можно использовать:

1. выборочную дисперсию по данным прошлых или пробных обследований;
2. дисперсию, найденную из соотношения для среднего квадратического отклонения:

$$\sigma = \frac{1}{3} \bar{x}$$

3. дисперсию, определенную из соотношения для асимметричного распределения:

$$\sigma = \frac{1}{5} (x_{\max} - x_{\min})$$

4. дисперсию, вычисленную из соотношения для нормального распределения:

$$\sigma = \frac{1}{6} (x_{\max} - x_{\min})$$

- **Пример.**
- Определить численность выборки по следующим данным.
- Для определения средней цены говядины на рынках города предполагается произвести выборочную регистрацию цен. Известно, что цены на говядину колеблются от 40 до 70 крон за кг. Сколько торговых точек необходимо обследовать, чтобы с вероятностью 0,954 ошибка выборки при определении средней цены не превышала 2 кроны за кг?

Относительная ошибка выборки

- характеризует относительную погрешность выборочного наблюдения

$$\Delta_{\text{относит}} = \frac{\Delta}{\bar{x}} 100\%$$

Расчет объема выборки при заданном уровне относительной ошибки

$$n_{\text{повт}} = \frac{t^2 v^2}{\Delta_{\text{относит}}^2}$$

$$n_{\text{бесповт}} = \frac{t^2 v^2 N}{t^2 v^2 + \Delta_{\text{относит}}^2 N}$$

- **Пример.**
- В городе зарегистрировано 30 тыс безработных. Для определения средней продолжительности безработицы организуется выборочное обследование. По данным прошлых лет известно, что коэффициент вариации продолжительности безработицы составляет 40%. Какое число безработных необходимо охватить выборочным наблюдением, чтобы с вероятностью 0,997 утверждать, что полученная предельная ошибка выборки не превышает 5% средней продолжительности безработицы?