

## 7.3. Виды выборочного наблюдения

Различные виды выборочного наблюдения классифицируются по семи признакам.




## Виды выборочного наблюдения по признакам





В зависимости от определения единицы отбора различают выборку **единицами и сериями (гнездами)**.


В первом случае совпадают единица совокупности и единица отбора и применяемый метод называется **простой случайной выборкой**.




**При серийной ( гнездовой) выборке** единицей отбора является серия (гнездо), представленная группой единиц совокупности, составляющих некоторое единство, - семья, школьный класс, студенческая группа и т. д.

### **Преимущества:**

- экономия затрат на обследование;
- возможность исследования взаимосвязей и процессов, действующих в пределах серии.



**Повторная выборка** является математической моделью выборки, при которой каждая отобранная единица возвращается в исходную генеральную совокупность и существует потенциальная возможность повторного отбора одних и тех же единиц.




**Метод типической  
(стратифицированной,  
районированной, расслоенной)  
выборки -**

генеральная совокупность  
подразделяется на качественно  
различные типы (страты, слои) и отбор  
производится строго самостоятельно из  
каждого типа.


# Вида стратифицированной выборки:

- 1) структура выборки пропорциональна структуре генеральной совокупности;
- 2) равномерное размещение единиц выборочной совокупности по типам;
- 3) отбор с долями, пропорциональными величине дисперсии (оптимальная выборка, метод Неймана) по каждому типу.



*Метод Неймана - он предложил его в 1934 г.  
В действительности впервые этот метод  
был сформулирован и опубликован русским  
статистиком А. А. Упровым еще в 1923 г;.  
Метод оптимального размещения выборки  
был описан также в книге А. Г. Ковалевского  
«Основы теории выборочного метода»,  
опубликованной в Саратове в 1924 г.*





При **многоступенчатой выборке** в целях сокращения затрат на проведение работ и обеспечения удобства обследования выборочная совокупность формируется поэтапно.


На всех ступенях, кроме последней, отбор производится сериями.

На каждой последующей ступени единицей отбора выступает более мелкая серия, а на последней ступени единицей отбора может быть единица совокупности либо серия более мелкая, чем на предыдущей ступени.

# Пример


Выборочная совокупность населения может быть сформирована в такой последовательности:

- 1) **первая ступень**- отбор субъектов РФ (единицей отбора является серия- субъект РФ);
- 2) **вторая ступень**- в каждом отобранном субъекте РФ производится отбор муниципальных образований, т. е. административных районов (единицей отбора выступает более мелкая серия- район);



3) **третья ступень**- в каждом отобранном районе производится отбор населенных пунктов (в сельской местности) или микрорайонов в крупных городах (единицей отбора является еще более мелкая серия-населенный пункт, микрорайон);

4) **четвертая ступень** - отбор необходимого числа жителей в отобранных населенных пунктах (единица отбора соответствует единице совокупности - человек).




**Принцип многофазной выборки** состоит в своеобразном совмещении нескольких наблюдений по одной проблеме в рамках одного обследования.



В течение 1 фазы предусматривается большой объем выборки при краткой программе обследования.

Во 2 фазе из отобранных в первой фазе единиц проводится «Подвыборка» и предусматривается существенное расширение программы наблюдения, которая обязательно включает вопросы программы 1 фазы и т.д.



**Взаимопроникающие выборки** могут быть получены делением общего объема выборки на равновеликие подвыборки путем проведения нескольких циклов механического отбора из уже сформированной выборки.

# Пример применения

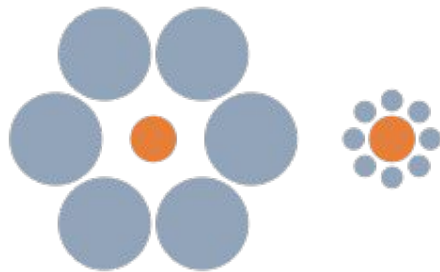
*Если необходимо годовой объем выборки распределить на 12 равных по объему подвыборок, с тем чтобы проводить наблюдение ежемесячно, каждый раз по новому кругу единиц.*

# Комбинированные выборки:

- 1) совмещаются сплошное и выборочное наблюдение;
- 2) сочетаются несколько видов выборочного наблюдения;
- 3) совмещаются выборочный и какой-либо иной вид несплошного наблюдения.



**Малой выборкой считается выборка  
объемом менее 30 единиц.**



## 7.4. Ошибки репрезентативности

**Ошибка репрезентативности (ошибка выборки) выборочных показателей** - это разница между выборочным показателем и значением данного показателя, которое было бы получено при сплошном наблюдении всех единиц генеральной совокупности.

**Средняя ошибка репрезентативности ( $\mu_{\tilde{x}}$ )**  
обратно пропорциональна корню квадратному  
из объема выборки ( $n$ ) и прямо  
пропорциональна дисперсии ( $\sigma^2$ ) изучаемого  
показателя (средней величины) по  
генеральной совокупности.


Ошибка репрезентативности уменьшается с  
ростом доли выборки ( $n$ ) в генеральной  
совокупности ( $N$ ), т. е. отношения  $n/N$ .

При  $n \rightarrow N$ , ошибка репрезентативности должна  
стремиться к нулю.

# Формулы средних ошибок репрезентативности для определяемых по выборочной совокупности средних величин $\bar{x}$

$$\mu_{\tilde{x}} = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)} \quad \text{— для бесповторной выборки;}$$

$$\mu_{\tilde{x}} = \sqrt{\frac{\sigma^2}{n}} \quad \text{— для повторной выборки.}$$



Если вычисляется средняя ошибка репрезентативности для показателей доли ( $w$ ), то используется **дисперсия доли**, определяемая как  $w(1 - w)$ , и формулы средних ошибок репрезентативности имеют вид:

$$\mu_w = \sqrt{\frac{w(1-w)}{n} \times \left(1 - \frac{n}{N}\right)} \quad \text{— для бесповторной выборки;}$$

$$\mu_w = \sqrt{\frac{w(1-w)}{n}} \quad \text{— для повторной выборки.}$$



В таблице условные обозначения:

$r$  – число серий в выборке;

$R$  - число серий в генеральной  
совокупности.

# Формулы средних ошибок репрезентативности

Вид выборки	Для выборочной средней ( $\bar{x}$ )	Для выборочного показателя доли ( $w$ )
Повторная, отбор единицами	$\sqrt{\frac{\sigma_x^2}{n}}$	$\sqrt{\frac{w(1-w)}{n}}$
Бесповторная, отбор единицами	$\sqrt{\frac{\sigma_x^2}{n} \left(1 - \frac{n}{N}\right)}$	$\sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$
Серийная, нерайонированная	$\sqrt{\frac{\sigma_{xR}^2}{r} \left(1 - \frac{r}{R}\right)}$	$\sqrt{\frac{w(1-w)}{r} \left(1 - \frac{r}{R}\right)}$
Районированная, отбор единицами	$\sqrt{\frac{\sigma_x^2}{n} \left(1 - \frac{n}{N}\right)}$	$\sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$
Районированная, отбор сериями	$\sqrt{\frac{\sigma_{xR}^2}{r} \left(1 - \frac{r}{R}\right)}$	$\sqrt{\frac{w(1-w)}{r} \left(1 - \frac{r}{R}\right)}$



# Формулы средних ошибок репрезентативности

Вид выборки	Для выборочной средней ( $\bar{x}$ )	Для выборочного показателя доли ( $w$ )
Повторная, отбор единицами	$\sqrt{\frac{\sigma_{\bar{x}}^2}{n}}$	$\sqrt{\frac{w(1-w)}{n}}$
Бесповторная, отбор единицами	$\sqrt{\frac{\sigma_{\bar{x}}^2}{n} \left(1 - \frac{n}{N}\right)}$	$\sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$
Серийная, нерайонированная	$\sqrt{\frac{\sigma_{\bar{x}r}^2}{r} \left(1 - \frac{r}{R}\right)}$	$\sqrt{\frac{w(1-w)}{r} \left(1 - \frac{r}{R}\right)}$
Районированная, отбор единицами	$\sqrt{\frac{\sigma_{\bar{x}}^2}{n} \left(1 - \frac{n}{N}\right)}$	$\sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$
Районированная, отбор сериями	$\sqrt{\frac{\sigma_{\bar{x}r}^2}{r} \left(1 - \frac{r}{R}\right)}$	$\sqrt{\frac{w(1-w)}{r} \left(1 - \frac{r}{R}\right)}$

# Средняя из дисперсий по районам (типам)


$$\overline{\sigma^2} = \frac{\sum \sigma_{\text{р}i}^2 n_i}{\sum n_i},$$

где  $\sigma_{\text{р}i}^2$  — дисперсия средней по  $i$ -му району;  $n_i$  — объем выборочной совокупности по  $i$ -му району;  $i$  — порядковый номер района.

# Дисперсия средней при серийной выборке

$$\sigma_{\bar{x}(r)}^2 = \frac{\sum_{j=1}^r (\bar{x}_j - \bar{x})^2}{r},$$


где  $\sigma_{\bar{x}(r)}^2$  – дисперсия средней при серийной выборке;  $j$  – порядковый номер серии;  $\bar{x}$  – средняя по всей выборке;  $\bar{x}_j$  – средняя по  $j$ -й серии.



**Надежность оценки** - это вероятность того, что ошибка не превысит установленных исследователем границ.

Если, например, оцениваются параметры, от которых зависит работа авиадвигателя или траектория ракеты, то потребуются очень высокая надежность, скажем 0,99999.

Если речь идет о результатах социологического исследования мнений населения о деятельности губернатора, то достаточно надежности на уровне 0,9.



При распределениях значений признака в генеральной совокупности, близких к закону Гаусса-Лапласа («нормальное распределение»), которые в основном приняты в статистических методах, вероятность того, что **ошибка отдельной выборки не превысит одной средней ошибки, составляет 0,6817.**

Такая надежность в 68% мала для большинства статистических задач.



Поэтому в расчетах используется предельная величина ошибки репрезентативности

$\Delta_{\bar{x}}$ , вычисляемую по формуле

$$\Delta_{\bar{x}} = t \mu_{\bar{x}},$$

$$\Delta_{\bar{x}} = t \mu_{\bar{x}}$$

где  $t$  — кратность средней ошибки;  $\Delta_{\bar{x}}$  — предел ошибки средней;  $\Delta_w$  — предел ошибки доли.

Если задано значение доверительной вероятности  $F(t)$  – вероятности того, что ошибка репрезентативности не выйдет за установленные границы, то величину  $t$  определяют по таблицам нормального распределения Гаусса-Лапласа (приложение).

Среди наиболее часто употребляемых на практике значений  $F(t)$  представлены, в частности, следующие:

$$F(2) = 0,9545;$$


$$F(2,24) = 0,9749;$$

$$F(2,5) = 0,9876;$$

$$F(3) = 0,9973;$$

$$F(3,15) = 0,9984.$$





При  $n < 30$  (малая выборка) следует находить  $F(t)$  по таблицам распределения Стьюдента.


## Относительная ошибка репрезентативности

равна отношению предельной ошибки  
репрезентативности к средней величине  
изучаемого признака (не должна  
превышать 5%)

$$K_{\Delta} = \frac{\Delta_{гр}}{\bar{x}}$$


## 7.5. Оценка параметров генеральной совокупности на основе материалов проведенного выборочного наблюдения

Весь цикл работ по проведению статистического исследования с использованием выборочного метода можно подразделить на три последовательно выполняемых блока процедур:

- 
- 1) проектирование выборочного наблюдения;
  - 2) собственно выборочное наблюдение (сбор первичных данных);
  - 3) разработка материалов выборочного наблюдения.



# Третий блок процедур включает:

- 1) расчет обобщающих показателей (средних величин, дисперсий, характеристик доли единиц с определенными значениями признака во всей совокупности и ряда других параметров );
- 2) расчет предельных значений ошибок репрезентативности для средних величин и показателей доли;
- 3) определение границ, в которых находятся средние величины, показатели доли, абсолютные итоги по генеральной совокупности.



На основе вычисленных абсолютных и относительных значений предельных ошибок репрезентативности делается вывод о том, *какова степень точности выборочных показателей,*

т. е. в какой мере они пригодны для использования в качестве характеристик генеральной совокупности.



Далее необходимо рассчитать, в каких границах при заданной доверительной вероятности находятся значения изучаемых показателей по генеральной совокупности.

$\bar{x} = \tilde{x} \pm \Delta_{\tilde{x}}$  — для средних величин;

$p = \omega \pm \Delta_{\omega}$  — для показателей доли;

$X = \bar{x}N$  — для абсолютных итогов.




## **7.6. Определение необходимого объема выборочной совокупности**

Путем алгебраических преобразований из формул предельных ошибок выборки выводятся формулы для расчета объема выборки (таблица).


# Формулы для расчета необходимого объема выборки n

Вид выборки	Вид показателя	Бесповторная выборка	Повторная выборка
Простая случайная выборка	Средняя	$\frac{t^2 \sigma_{\bar{x}}^2 N}{\Delta_{\bar{x}}^2 N + t^2 \sigma_{\bar{x}}^2}$	$\frac{t^2 \sigma_{\bar{x}}^2}{\Delta_{\bar{x}}^2}$
	Доля	$\frac{t^2 w(1-w)}{\Delta_w^2 N + t^2 w(1-w)}$	$\frac{t^2 w(1-w)}{\Delta_w^2}$
Стратифицированная выборка, пропорциональная	Средняя	$\frac{t^2 \overline{\sigma_{\bar{x}}^2} N}{\Delta_{\bar{x}}^2 N + t^2 \overline{\sigma^2}}$	$\frac{t^2 \overline{\sigma_{\bar{x}}^2}}{\Delta_{\bar{x}}^2}$
	Доля	$\frac{t^2 N \overline{w(1-w)}}{\Delta_w^2 N + t^2 \overline{w(1-w)}}$	$\frac{t^2 \overline{w(1-w)}}{\Delta_w^2}$
Серийная нестратифицированная выборка	Средняя	$\frac{t^2 \sigma_{\bar{x}(r)}^2 R}{\Delta_{\bar{x}}^2 R + t^2 \sigma_{\bar{x}(r)}^2}$	$\frac{t^2 \sigma_{\bar{x}(r)}^2}{\Delta_{\bar{x}}^2}$




Если объем выборки завышен, то необоснованно завышается стоимость работ, увеличиваются сроки их выполнения.

При заниженном объеме выборки материалы выборочного наблюдения могут оказаться дефектными и непригодными для использования по назначению.




Следует принимать во внимание качество и достоверность используемой исходной информации:

- сведений о составе и объеме генеральной совокупности,
- предлагаемых экспертами и заказчиками уровней доверительной вероятности и предельной ошибки репрезентативности,
- косвенных оценок дисперсии.



Необходимо исходить из таких условий и обстоятельств, как:

- объем финансирования,
- сроки выполнения работ,
- степень сложности системы итоговых показателей, которую необходимо получить по генеральной совокупности,
- степень риска,
- появления непредвиденных обстоятельств, препятствующих неукоснительному выполнению принятых решений.



Например, если существует вероятность того, что часть отобранных единиц окажется недоступной для обследования, то необходимо предусмотреть порядок их замены другими единицами.

Целесообразно проведение пилотных (пробных) обследований.