

Метод выборочного наблюдения социально-экономических явлений и процессов

ПЛАН

1. Понятие о выборочном наблюдении. Преимущества выборочного наблюдения.
2. Теоретические основы выборочного метода. Генеральная и выборочная совокупности, их обобщающие характеристики.
3. Ошибки выборочного наблюдения.
4. Определение необходимой численности выборки.
5. Методы, виды и способы отбора выборочных совокупностей.
6. Способы распространения выборочных данных на генеральную совокупность.
7. Малые выборки и их особенности.

Статистическое наблюдение можно организовать как сплошное и несплошное.

- Сплошное наблюдение предусматривает обследование всех единиц изучаемой совокупности явлений, несплошное лишь ее часть. К несплошному наблюдению относится и *выборочное наблюдение*.

Понятие выборочного наблюдения

Выборочное наблюдение

- это такое несплошное наблюдение, при котором статистическому наблюдению подвергаются не все единицы изучаемой совокупности, а лишь отобранные в определенном порядке

Цель выборочного наблюдения состоит в том, чтобы по характеристикам отобранной части единиц судить о характеристиках всей совокупности.

Основные причины, по которым во многих случаях выборочному наблюдению отдается предпочтение перед сплошным наблюдением следующие:

- достижение большей точности результатов обследования благодаря сокращению ошибок регистрации (за счет участия более квалифицированных участников)



- экономия трудовых и денежных средств и времени в результате сокращения объема работы



- возможность детального обследования каждой единицы наблюдения за счет расширения программы наблюдения



- сведение к минимуму или даже уничтожению и приведению в негодность обследуемых единиц совокупности



- уточнение результатов сплошного наблюдения

Преимущества выборочного наблюдения по сравнению со сплошным возможно обеспечить, если оно организовано и проведено в строгом соответствии с научными принципами выборочного наблюдения.

Научные принципы выборочного наблюдения

- *обеспечение случайности отбора единиц* (при отборе каждой из единиц изучаемой совокупности обеспечивается равная возможность попасть в выборку)
- *обеспечение достаточного числа отобранных единиц совокупности*

Соблюдение этих принципов позволяет получить такую совокупность единиц, которая по интересующим исследователя признакам представляет всю изучаемую совокупность, т.е. является *репрезентативной* (представительной).

Понятие генеральной и выборочной совокупностей

Генеральная совокупность — это совокупность, из которой производится отбор единиц совокупности

Выборочная совокупность — это совокупность отобранных в определенном порядке единиц

Основные понятия и характеристики выборочного наблюдения

Доля выборки

это отношение численности выборочной совокупности к численности генеральной совокупности

$$k = n/N$$

где N – численность генеральной совокупности;

n – численность выборочной совокупности.

**Генеральная
средняя ($\bar{x}$)**

это среднее значение признака всей совокупности

$$\bar{X} = \frac{\sum x_i}{N}$$

Выборочная средняя ($\tilde{x}$)	это среднее значение признака у единиц, которые подверглись выборочному наблюдению $\tilde{x} = \frac{\sum x_i}{n}$
Генеральная доля (p)	это доля единиц, обладающих тем или иным признаком в генеральной совокупности $p = M/N$ где M – численность единиц, обладающих определенным признаком в генеральной совокупности.
Выборочная доля или частота (w)	это доля единиц, обладающих тем или иным признаком в выборочной совокупности $w = m/n$ где m – численность единиц, обладающих определенным признаком в выборочной совокупности.

<i>Дисперсия количественного признака в генеральной совокупности</i>	$\sigma_{\bar{X}}^2 = \frac{\sum (x_i - \bar{X})^2}{N}$
<i>Дисперсия количественного признака в генеральной совокупности</i>	$\sigma_{\tilde{X}}^2 = \frac{\sum (x_i - \tilde{X})^2}{n}$
<i>Дисперсия доли признака в генеральной совокупности</i>	$\sigma_p^2 = p(1-p)$
<i>Дисперсия доли признака в выборочной совокупности</i>	$\sigma_W^2 = W \cdot (1-W)$

При проведении выборочного наблюдения даже теоретически нельзя получить абсолютно точные данные, как при сплошном. Обусловлено это тем, что обследованию подвергается не вся совокупность, а только ее часть. Поэтому при проведении выборочного наблюдения неизбежна некоторая свойственная ему погрешность (ошибки).

Ошибки, свойственные выборочному наблюдению, называются *ошибками репрезентативности*.

Понятие ошибки репрезентативности

Ошибка репрезентативности

- это расхождение между выборочной характеристикой и характеристикой генеральной совокупности

Ошибки репрезентативности



```
graph TD; A[Ошибки репрезентативности] --> B[Систематические (возникают в результате нарушения научных принципов отбора единиц совокупности)]; A --> C[Случайные (возникают в результате несплошного характера наблюдения)]; B --> D[Преднамеренные]; B --> E[Непреднамеренные]; C --> F[Средняя (стандартная) ошибка выборки]; C --> G[Предельная ошибка выборки];
```

Систематические (возникают в результате нарушения научных принципов отбора единиц совокупности)

Случайные (возникают в результате несплошного характера наблюдения)

Преднамеренные

Непреднамеренные

Средняя
(стандартная)
ошибка выборки

Предельная
ошибка выборки

- Теоретическим обоснованием появления случайных ошибок выборки является теория вероятностей и ее предельные теоремы.
- **Сущность предельных теорем** состоит в том, что в массовых явлениях совокупное влияние различных случайных причин на формирование закономерностей и обобщающих характеристик будет сколь угодно малой величиной или практически не зависит от случая.
- Так как случайная ошибка выборки возникает в результате случайных различий между границами выборочной и генеральной совокупностей, то при достаточно большом объеме выборки она будет сколь угодно мала. Этот вывод, опирающийся на доказательства предельных теорем, позволяет предполагать, что характеристики выборочного наблюдения могут достаточно хорошо представлять характеристики генеральной совокупности.
- Случайные ошибки могут быть доведены до незначительных размеров, а главное, размеры и пределы их можно определить с достаточной точностью на основании закона больших чисел.

Понятие средней (стандартной) ошибки выборки

Средняя ошибка выборки - представляет собой такое расхождение между средними выборочной и генеральной совокупностями $(\tilde{x} - \bar{x})$, которое не превышает $\pm \sigma$

**Средняя ошибка выборки
зависит от:**

- **объема выборки** (чем больше численность при прочих равных условиях, тем меньше величина средней ошибки выборки)

- **степени варьирования признака** (чем меньше вариация признака, а следовательно, и дисперсия, тем меньше ошибка выборки, и наоборот)

Формула для определения величины средней ошибки выборки для количественного признака

Средняя ошибка выборки

Ф. 1

$$\mu_{\tilde{x}} = \sqrt{\frac{\sigma_{\tilde{x}}^2}{n}} = \frac{\sigma_{\tilde{x}}}{\sqrt{n}}$$

Формула для определения величины средней ошибки выборки для альтернативного признака

Средняя ошибка выборки

Ф. 2

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

Понятие предельной ошибки выборки

Предельная ошибка выборки - максимально возможное расхождение выборочной и генеральной средних $(\tilde{x} - \bar{x})$, т.е. максимум ошибки при заданной вероятности ее появления.

О величине предельной ошибки можно судить с определенной вероятностью, на величину которой указывает *коэффициент доверия t* .

Формула для определения величины предельной ошибки выборки

Предельная ошибка выборки

Ф. 3 $\Delta_{\tilde{x}} = \mu_{\tilde{x}} \cdot t$ или $\Delta_w = \mu_w \cdot t$

где Δ – предельная ошибка выборки;

t – коэффициент доверия, зависящий от вероятности,
с которой гарантируется предельная ошибка выборки.

Чем больше пределы, в которых допускается возможная ошибка, тем с большей вероятностью можно установить ее величину. Предельная ошибка выборки позволяет определять предельные значения характеристик генеральной совокупности при заданной вероятности и их доверительные интервалы.

Формула для определения интервальной оценки генеральной средней

Интервальная оценка

Ф. 4
$$\tilde{x} - \Delta_{\tilde{x}} \leq \bar{x} \leq \tilde{x} + \Delta_{\tilde{x}}$$

Это означает, что с заданной вероятностью можно утверждать, что значение генеральной средней можно ожидать в пределах от $\tilde{x} - \Delta_{\tilde{x}}$ до $\tilde{x} + \Delta_{\tilde{x}}$

Формула для определения интервальной оценки генеральной доли

Интервальная оценка

Ф. 5
$$\omega - \Delta_w \leq p \leq \omega + \Delta_w$$

При подготовке выборочного наблюдения с заранее заданным значением допустимой ошибки выборки очень важно правильно определить объем (численность) выборочной совокупности. Согласно одному из принципов выборочного наблюдения, объем выборки должен быть достаточным, чтобы обеспечить репрезентативность выборки.

Расчет необходимой численности выборки строится с помощью *формул 2 и 3*.

Формула для определения необходимой численности выборки для средней

Численность выборки

$$n = \frac{t^2 \cdot \sigma_{\tilde{x}}^2}{\Delta_{\tilde{x}}^2}$$

Формула для определения необходимой численности выборки для доли

Численность выборки

$$n = \frac{\omega \cdot (1 - \omega)}{\Delta_w^2}$$

В теории выборочного наблюдения разработаны различные методы, способы и виды отбора единиц из генеральной совокупности.

Методы отбора

- повторный

Каждая единица, отобранная в случайном порядке, после ее обследования возвращается в генеральную совокупность и в последующем отборе может снова попасть в выборку.

При таком отборе вероятность попасть в выборку для каждой единицы генеральной совокупности не меняется независимо от числа отобранных единиц.

- бесповторный

Каждая единица, отобранная в случайном порядке, после ее обследования в генеральную совокупность не возвращается.

Вероятность попасть в выборку для каждой единицы генеральной совокупности увеличивается по мере производства отбора.

Так как бесповторный отбор охватывает все новые и новые совокупности, а повторный отбор на всем протяжении одну и ту же совокупность, то бесповторный отбор дает более точные результаты, чем повторный.

Виды отбора

Индивидуальный

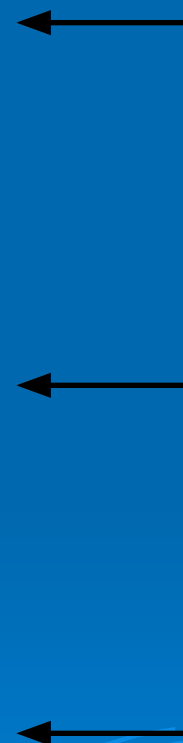
(в выборочную совокупность отбираются отдельные единицы генеральной совокупности)

Групповой

(в выборочную совокупность отбираются качественно однородные группы или серии изучаемых единиц)

Комбинированный

(происходит сочетание первого и второго видов отбора)



Способы отбора:

- **собственно-случайный**
- **механический**
- **типический**
- **серийный**
- **комбинированный**

Понятие собственно-случайного отбора

Собственно-случайный отбор – это отбор, при котором наблюдению подвергается часть совокупности, отобранная из всей совокупности в случайном порядке

Собственно-случайный отбор бывает повторным и бесповторным.

Основные формулы, используемые при собственно-случайном отборе

	Выборка повторная	Выборка бесповторная
<i>Средняя ошибка выборки для средней</i>	$\mu_{\tilde{x}} = \sqrt{\frac{\sigma_{\tilde{x}}^2}{n}}$	$\mu_{\tilde{x}} = \sqrt{\frac{\sigma_{\tilde{x}}^2}{n} \left(1 - \frac{n}{N}\right)}$
<i>Средняя ошибка выборки для доли</i>	$\mu_{\omega} = \sqrt{\frac{\omega \cdot (1 - \omega)}{n}}$	$\mu_{\omega} = \sqrt{\frac{\omega \cdot (1 - \omega)}{n} \left(1 - \frac{n}{N}\right)}$
<i>Численность выборки при определении среднего размера признака</i>	$n = \frac{t^2 \cdot \sigma_{\tilde{x}}^2}{\Delta_{\tilde{x}}^2}$	$n = \frac{t^2 \cdot \sigma_{\tilde{x}}^2 \cdot N}{N \cdot \Delta_{\tilde{x}}^2 + t^2 \cdot \sigma_{\tilde{x}}^2}$



***Численность
выборки при
определении доли
признака***

$$n = \frac{t^2 \cdot \omega \cdot (1 - \omega)}{\Delta_w^2}$$

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

Понятие механического отбора

Механический отбор - применяется в тех случаях, когда генеральная совокупность каким-либо образом упорядочена, т. е. имеется определенная последовательность в расположении единиц (например, номера домов, списки избирателей)

При проведении механического отбора устанавливается **шаг отсчета**, т.е. расстояние между отбираемыми единицами (N/n – величина обратная доли выборки) и **начало отсчета** - номер той единицы, которая должна быть обследована первой.

- Механический отбор всегда бывает ***бесповторным***. При этом отборе применяются те же формулы, что и при собственно-случайном бесповторном отборе.
- Механический отбор имеет преимущество перед случайным отбором, его не только легче организовать, но при нем единицы выборочной совокупности равномернее распределяются в генеральной совокупности.

Понятие типического отбора

Типичный отбор - представляет собой отбор, при котором генеральная совокупность разбивается на качественно однородные типические группы, затем из каждой группы собственно-случайной или механической выборкой производится отбор единиц в выборочную совокупность

Из всех типических групп можно отбирать число единиц, пропорциональное их численностям и непропорциональное. В зависимости от этого различают **пропорциональный и непропорциональный** типический отбор.

Объем выборки из типической группы при отборе пропорциональном численности единиц типических групп определяется по формуле:

$$n_i = n \frac{N_i}{N}$$

где n_i - объем выборки из i -ой типической группы;

N_i – объем i -ой типической группы в генеральной совокупности

- Типический отбор бывает **повторным и бесповторным**.
- Разбивка на типические группы дает возможность избежать влияния межгрупповой вариации на точность выборки. Так как в типическую выборку должны попасть представители всех групп, то средняя ошибка типической выборки зависит только от **средней из внутригрупповых дисперсий** $\sigma_{\tilde{x}}^2$ или $\omega(1 - \omega)$,
а не общей дисперсии $\sigma_{\tilde{x}}^2$ или $\omega(1 - \omega)$

Основные формулы, используемые при типичном отборе

	Выборка повторная	Выборка бесповторная
<i>Средняя ошибка выборки для средней</i>	$\mu_{\tilde{x}} = \sqrt{\frac{\overline{\sigma}_i^2}{n}}$	$\mu_{\tilde{x}} = \sqrt{\frac{\overline{\sigma}_i^2}{n} \left(1 - \frac{n}{N}\right)}$
<i>Средняя ошибка выборки для доли</i>	$\mu_{\omega} = \sqrt{\frac{\omega \cdot (1 - \omega)}{n}}$	$\mu_{\omega} = \sqrt{\frac{\omega \cdot (1 - \omega)}{n} \left(1 - \frac{n}{N}\right)}$
<i>Численность выборки при определении среднего размера признака</i>	$n = \frac{t^2 \cdot \overline{\sigma}_i^2}{\Delta_{\tilde{ч}}^2}$	$n = \frac{t^2 \cdot \overline{\sigma}_i^2 \cdot N}{N \cdot \Delta_{\tilde{x}}^2 + t^2 \cdot \overline{\sigma}_i^2}$



*Численность
выборки при
определении
доли признака*

$$n = \frac{t^2 \cdot \overline{\omega \cdot (1-\omega)}}{\Delta_w^2}$$

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

Понятие серийного отбора

Серийный номер - такой отбор, когда в случайном порядке отбираются не единицы, подлежащие обследованию, а группы единиц (серии, гнезда). Внутри отобранных серий обследованию подвергаются все единицы, т.е. применяется сплошное наблюдение

- Поскольку внутри серий обследуются все без исключения единицы, средняя ошибка выборки при отборе равновеликих серий зависит от величины только

$\delta_{\tilde{x}}$ *межгрупповой дисперсии*
или δ_{ω}

- Серийный отбор бывает *повторным и бесповторным*.

Основные формулы, используемые при серийном отборе

	Выборка повторная	Выборка бесповторная
<i>Средняя ошибка выборки для средней</i>	$\mu_{\tilde{x}} = \sqrt{\frac{\delta_{\tilde{x}}^2}{r}}$	$\mu_{\tilde{x}} = \sqrt{\frac{\delta_{\tilde{x}}^2}{r} \left(1 - \frac{r}{R}\right)}$
<i>Средняя ошибка выборки для доли</i>	$\mu_{\omega} = \sqrt{\frac{\delta_{\omega}^2}{r}}$	$\mu_{\omega} = \sqrt{\frac{\delta_{\omega}^2}{r} \left(1 - \frac{r}{R}\right)}$
<i>Численность выборки при определении среднего размера признака</i>	$r = \frac{t^2 \cdot \delta_{\tilde{x}}^2}{\Delta_{\tilde{x}}^2}$	$r = \frac{t^2 \cdot \delta_{\tilde{x}}^2 \cdot R}{R \cdot \Delta_{\tilde{x}}^2 + t^2 \cdot \delta_{\tilde{x}}^2}$



*Численность
выборки при
определении
доли признака*

$$r = \frac{t^2 \cdot \delta_w^2}{\Delta_w^2}$$

$$r = \frac{t^2 \cdot \delta_\omega^2 \cdot R}{R \cdot \Delta_w^2 + t^2 \cdot \delta_\omega^2}$$

где R – общее число серий;
 r – число отобранных серий;

$\delta_{\tilde{x}}^2$ – межгрупповая дисперсия средних, определяемая по формуле

$$\delta_{\tilde{x}}^2 = \frac{\sum (\tilde{x}_i - \tilde{x})^2}{r}$$

где $\tilde{x}_i$ – средняя i -ой серии;

$\tilde{x}$ – средняя по всей выборочной совокупности.

δ_{ω}^2 – межгрупповая дисперсия доли, определяемая по формуле ,

$$\delta_{\omega}^2 = \frac{\sum (\omega_i - \bar{\omega})^2}{r}$$

где

ω_i - доля признака i -ой серии;

$\bar{\omega}$ - общая доля признака во всей выборочной совокупности.

Конечной целью любого выборочного наблюдения является распространение его характеристик на генеральную совокупность.

Способы распространения данных выборочного наблюдения на генеральную совокупность



способ прямого пересчета

Применяется в том случае, если целью выборочного наблюдения является определение объема признака генеральной совокупности, когда известна лишь численность ее единиц (пример 1)



способ поправочных коэффициентов

Применяется в тех случаях, когда целью выборочного метода является уточнение результатов сплошного наблюдения (пример 2)

Пример 1. Для определения качества продукции проверено 200 изделий из 10000, или 2%. В результате проверки установлено с вероятностью 0,997, что средний процент бракованной продукции выборочной партии составил 4%, а пределы возможных отклонений 0,5%. Способом прямого пересчета необходимо определить количество бракованных изделий в генеральной совокупности.

Количество бракованных изделий = (Число изделий · Процент бракованных изделий) : 100

Генеральная средняя будет находится в пределах:

$$\tilde{x} - \Delta \leq \bar{x} \leq \tilde{x} + \Delta \quad 4 - 0,5 \leq \bar{x} \leq 4 + 0,5 \quad 3,5 \leq \bar{x} \leq 4,5$$

Количество бракованных изделий во всей партии (штук):

$$(10000 \times 3,5) : 100 \leq \text{Количество бракованных изделий} \leq (10000 \times 4,5) : 100$$

$$350 \leq \text{Количество бракованных изделий} \leq 450$$

Пример 2. При проведении сплошного учета скота в регионе было зарегистрировано 10000 голов. С целью проверки данных сплошного учета проведены контрольные обходы части обследованных хозяйств и выявлено, что если данные сплошного учета скота в хозяйствах, попавших в выборку, показали 300 голов, то данные выборки в этих же хозяйствах – 305 голов. Следовательно, «процент недоучета» (коэффициент) при сплошном наблюдении составил:

$$(305 \div 300) \times 100 = 1,67\%$$

Количество голов скота необходимо умножить на этот коэффициент:

$$10000 \times 1,0167 = 10167 \text{ голов}$$

Понятие малой выборки

Малая выборка - несплошное статистическое обследование, численность единиц которого не превышает 30.

Для определенного способа отбора единиц величина стандартной ошибки зависит от объема выборки и степени колеблемости изучаемого признака в генеральной совокупности. Причем, чем меньше объем выборки, тем большую величину стандартной ошибки следует ожидать, а это снижает точность оценки параметров генеральной совокупности.

Для оценки возможных пределов ошибки малой выборки применяется отношение Стьюдента, определяемое по формуле:

$$t = \frac{\tilde{x} - \bar{x}}{\mu_{M.B}}$$

$\mu_{M.B}$ где — величина среднего квадратического отклонения малой выборки, которая определяется по формуле:

$$\mu_{M.B} = \sqrt{\frac{\sigma^2}{n-1}} = \frac{\sigma}{\sqrt{n-1}}$$

Величина σ вычисляется на основе данных выборочного наблюдения. Она равна:

$$\sigma = \sqrt{\frac{\sum (x_i - \tilde{x})^2}{n}}$$

Таким образом, теоретическое распределение отношения Стьюдента t имеет дело с величинами, определяемыми непосредственно по данным выборки. Для отдельных значений t и n доверительная вероятность малой выборки определяется по специальным таблицам Стьюдента, которые приводятся в учебниках по математической статистике.

Предельная ошибка малой выборки определяется по формуле:

$$\Delta_{M.V} = t \cdot \mu_{M.V}$$

Порядок расчетов тот же, что и при больших выборках.