



Профпа

ЧАСТЬ И ЦЕЛОЕ

# Выборочное наблюдение

МОИ  
первые  
ИГРЫ

# Определение выборочного наблюдения

- **Выборочное наблюдение** — это способ несплошного статистического наблюдения, при котором обследуются не все единицы изучаемой (генеральной) совокупности, а лишь часть ее (выборка), отобранная по определенным правилам и обеспечивающая получение данных, характеризующих совокупность в целом.

Причины применения:

Экономия

Невозможность проведения сплошного исследования



Вся изучаемая совокупность называется **генеральной совокупностью**

Часть генеральной совокупности, которая подвергается обследованию – называется **выборочной совокупностью (выборкой)**.

Для того, чтобы выборочная совокупность давала объективные результаты, она должна быть **репрезентативной** (каждая единица генеральной совокупности должна иметь равную возможность попасть в выборку).

Некоторые единицы могут попадать в выборку дважды, трижды или даже большее число раз. Такой отбор в выборку называется **повторным**.

При **бесповторном** отборе попавшая в выборку единица подвергается обследованию и в дальнейшей процедуре отбора не участвует.

Теоретической основой выборки являются теоремы закона больших чисел (Чебышева, Ляпунова, Бернулли и др.)



# Классификация ошибок выборочного наблюдения

## Ошибки выборочного наблюдения

### Ошибки регистрации

Систематические

Случайные

### Ошибки репрезентативности

Систематические

Случайные

- **Ошибки регистрации** являются следствием неправильного установления значения наблюдаемого признака или неправильной записи.
- **Ошибки репрезентативности** обусловлены тем, что выборочная совокупность не может по всем параметрам в точности воспроизвести генеральную совокупность.
- **Систематические ошибки** репрезентативности связаны с нарушением принципов формирования выборочной совокупности.
- **Случайные ошибки** репрезентативности обусловлены действием случайных факторов, не содержащих каких-либо элементов системности

- В 1936 г. в США проводился опрос, проведенный «Литэрари Дайджест» («Литературное обозрение») относительно исхода президентских выборов.
- Кандидатами на этих выборах были Ф. Д. Рузвельт и А. М. Ландон. Редакция журнала организовала план выборки следующим образом. В выборку попали более двух миллионов американцев, выбранных при помощи случайного отбора из списков, имеющихся в телефонных книгах. По всей стране попавшим в выборку лицам были разосланы открытки с просьбой назвать фамилию будущего президента. Затратив огромную сумму на рассылку, сбор и обработку полученных открыток, журнал информировал общественность, что на предстоящих выборах президентом США с большим перевесом будет избран А. М. Ландон. Результаты выборов опровергли этот прогноз.
- В то же время социологи Д. Гэллуп и Э. Роупер правильно предсказали победу Ф. Д. Рузвельта, основываясь только на четырех тысячах анкет.
- Ошибочный прогноз относительно возможного президента объясняется неправильным планом выборки, который не обеспечил полного отражения в ней всей генеральной совокупности: в телефонных книгах, которые использовались для организации выборки, были представлены лишь наиболее обеспеченные слои американского населения. Поскольку обеспеченные слои американцев составляли меньшую часть генеральной совокупности, то распространение мнения этой части населения на всю страну в целом оказалось ошибочным.



# Задачи выборочного метода

**Определение доверительного интервала, в котором находится характеристика генеральной совокупности**

**Определение минимального объема выборки**

**Определение доверительной вероятности того, что разность между характеристиками выборочной и генеральной совокупностей не превзойдет наперед заданного числа**

## Основные обозначения:

- N** – объем генеральной совокупности (количество единиц генеральной совокупности);
- n** – объем выборочной совокупности (количество единиц выборочной совокупности);
- $\bar{X}$  - генеральная средняя (средняя величина, которая имеет место в генеральной совокупности);
- $\tilde{x}$  - выборочная средняя;
- M** – численность единиц генеральной совокупности определенным вариантом (численность городского населения, число нерентабельных предприятий и т.д.);
- m** - численность единиц выборочной совокупности определенным вариантом;
- P** – генеральная доля, т.е. доля единиц, обладающих определенным вариантом (доля городского населения, доля нерентабельных предприятий), определяется как  $M/N$ ;
- w** – выборочная доля, т.е. доля единиц, обладающих определенным вариантом, определяется как  $m/n$ ;
- $\mu$  - средняя ошибка выборки;       $\Delta$  - предельная ошибка выборки

# Теорема П.Л. Чебышева

- При достаточно большом числе независимых наблюдений можно с вероятностью, близкой к единице (т.е. почти с достоверностью), утверждать, что отклонение выборочной средней от генеральной будет сколько угодно малым.
- В теореме доказано, что **величина предельной ошибки  $\Delta$  не должна превышать  $t\mu$** , где  $t$  - число, связанное с вероятностью через таблицу закона нормального распределения.
- В свою очередь, величина  $\mu$ , выражающая среднее квадратическое отклонение выборочной средней от генеральной средней, зависит от колеблемости признака в генеральной совокупности  $\sigma$  и числа отобранных единиц  $n$ . Эта величина называется **средней ошибкой выборки** и выражается формулой:



# Теорема А.М. Ляпунова

Вероятность предельной ошибки при достаточно большом объеме выборки подчиняется закону нормального распределения

$$P(\Delta \leq t\mu) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-t}^t e^{-\frac{t^2}{2}} dt = F(t)$$

- Значения этого интеграла для различных значений коэффициента доверия  $t$  приводятся в специальных математических таблицах. В частности, при  $t = 1$ ,  $P = 0,683$ ;  $t = 2$ ,  $P = 0,954$ ;  $t = 3$ ,  $P = 0,997$ ;

это может быть прочитано так: в 68,3% случаев ошибка репрезентативности не выйдет за пределы  $\pm \mu$ .

- Зная выборочную среднюю величину признака  $\bar{x}$  и предельную ошибку выборки  $\Delta_{\bar{x}}$ , можно определить границы (пределы), в которых заключена генеральная средняя:

$\bar{x} - \Delta_{\bar{x}} \leq \mu \leq \bar{x} + \Delta_{\bar{x}}$   
или  $w - \Delta_w \leq p \leq w + \Delta_w$ . Зная выборочную долю признака ( $w$ ), можно определить границы, в которых заключена

Из теорем Чебышева, Ляпунова и закона больших чисел следует:

**Хотя каждая выборочная средняя отличается от генеральной, среднее значение по ним равно генеральной:**

$$\frac{\sum x_i}{n} = \bar{x}$$



# Условия проведения выборки

Выборка будет представлять всю совокупность с приемлемой точностью при выполнении двух условий.

Во-первых, она должна быть достаточно **многочисленной**, чтобы в ней могли проявиться закономерности, существующие в генеральной совокупности.

Во-вторых, элементы выборки должны быть отобраны **объективно**, независимо от воли исследователя, чтобы каждый из них имел одинаковые шансы быть отобранным или чтобы эти шансы были известны исследователю.

Кандидат Иванов

Кандидат Петров

Кандидат Сидоров

# Способы отбора

- По виду различают индивидуальный, групповой и комбинированный отбор. При **индивидуальном** отборе в выборочную совокупность отбираются отдельные единицы генеральной совокупности, при **групповом отборе** – группы единиц, а **комбинированный отбор** предполагает сочетание группового и индивидуального отбора.
- Способ отбора определяет конкретный механизм или процедуру выборки единиц из генеральной совокупности. В практике выборочных обследований наибольшее распространение получили следующие выборки:
  - **собственно-случайная;**
  - **механическая;**
  - **типическая (стратифицированная);**
  - **серийная;**
  - **комбинированная.**



# 1. Простая случайная

- Вы сняты с выборов

за отсутствием

разрешения на полеты!

## выборка

при которой  $n$  объектов случайно извлекаются из генеральной совокупности  $N$  объектов (например с помощью таблицы или датчика случайных чисел), причем каждая из возможных выборок имеют равную вероятность. Такие выборки называются **собственно-случайными**.

Случайная выборка - основа всех других способов отбора.

Случайная выборка осуществляется методом жеребьевки: все единицы совокупности нумеруются, номера записываются на карточки, а потом отбираются.

На практике осуществляется с помощью таблиц случайных чисел.

И запомни, Золушка!

# Формулы предельных ошибок выборки

- Если отбор единиц из генеральной совокупности произведен бесповторным способом, то в формулы средней ошибки выборки вносится поправка:

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

Наименование ошибки	Способ отбора	
	повторный	бесповторный
<b>Средняя ошибка :</b>		
для средней	$\mu = \sqrt{\frac{\sigma^2}{n}}$	$\mu = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$
для доли	$\mu = \sqrt{\frac{w(1-w)}{n}}$	$\mu = \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$
<b>Предельная ошибка :</b>		
для средней	$\Delta_{\bar{x}} = t \cdot \sqrt{\frac{\sigma^2}{n}}$	$\Delta_{\bar{x}} = t \cdot \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$
для доли	$\Delta_p = t \cdot \sqrt{\frac{w(1-w)}{n}}$	$\Delta_p = t \cdot \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$



**Пример.** Для определения среднего срока службы изделий было обследовано 250 изделий. При этом средний срок службы был установлен на уровне 41,9 месяца. Среднее квадратическое отклонение равно 6,2 месяцам.

**С вероятностью 0,9973 определить, в каких пределах находится средний срок службы всех изделий.**

$P=0,9973$ ,  $t=3$  (из таблицы интеграла вероятностей закона нормального распределения).

$$\Delta_{\bar{x}} = 3 \cdot \sqrt{\frac{6,2^2}{250}} = 1,2 \text{ мес}$$

$$41,9 - 1,2 \leq \bar{x} \leq 41,9 + 1,2$$

$$40,7 \text{ мес} \leq \bar{x} \leq 43,1 \text{ мес}$$

**Определить вероятность того, что предельная ошибка среднего срока службы не превысит 1 месяц.**

$$t = \frac{\Delta_{\bar{x}}}{\sqrt{\frac{\sigma^2}{n}}} = \frac{1}{\sqrt{\frac{6,2^2}{250}}} = 2,55$$

$$p = 0,9892$$

## Пример. Определение минимального объема выборки

Сколько следует прохронометрировать операций, чтобы с вероятностью 0,9973 можно было бы утверждать, что разность между средней продолжительностью операций в выборочной и генеральной совокупности не превысит 1 секунды, если по результатам предыдущего испытания установлено, что средняя продолжительность операции равна 30 секундам, а среднее квадратическое отклонение равно 7 секундам?

$$n = ?$$

$$\Delta = 1$$

$$n = \frac{t^2 \cdot \sigma^2}{\Delta_x^2} = \frac{3^2 \cdot 7^2}{1^2} = 441$$

*Ответ:* нужно прохронометрировать не менее 441 операции.



## 2. Механическая (систематическая) выборка

применяется в случаях, когда генеральная совокупность каким-либо образом **упорядочена**, т.е. имеется определенная последовательность в расположении единиц (табельные номера работников, списки избирателей, телефонные номера респондентов, номера домов и квартир и т. п.).

- Для проведения механической выборки устанавливается пропорция отбора, которая определяется соотношением объемов выборочной и генеральной совокупностей. Отбор единиц осуществляется в соответствии с установленной пропорцией через равные интервалы. Например, при пропорции 1 : 50 (2%-ная выборка) отбирается каждая 50-я единица, при пропорции 1 : 20 (5 %-ная выборка) - каждая 20-я единица и т.

д.  
22:48

Выборь невесты Царя Алексея Михайловича

- Для определения средней ошибки механической выборки

### 3. Типическая (стратифицированная) выборка

заключается в том, что генеральная совокупность объема  $N$  подразделяется на части совокупности или слои (страты) объема  $N_1, N_2, \dots, N_r$ , так что  $N_1 + N_2 + \dots + N_r = N$ .

Страты - однородные объекты с точки зрения статистических характеристик. Например:

**население**

сельское

городское

из каждой типической группы отбирается некоторое количество единиц.

Отбор может быть как **пропорциональным** объёму типических групп, так и **непропорциональным**



# Объем типической выборки

При отборе, пропорциональном объему типических групп, число наблюдений по каждой группе определяется по формуле:

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

$n_i$  - объем выборки из  $i$ -й типической группы.

$n$  - общий объем выборки.

$N_i$  - объем  $i$ -й типической группы в генеральной совокупности.

$N$  - объем генеральной совокупности.

## 4. Серийная выборка

- Приемы **серийного** отбора используются для формирования *серийных или гнездовых выборок*. Они удобны в том случае, если необходимо обследовать сразу "блок" или серию объектов (например, партию товара, продукцию определенной серии или предприятия территориально-административной единицы).
- Вся совокупность делится на серии, после чего механическим или собственно случайным способом отбирается некоторое количество серий. Все единицы совокупности, входящие в отобранные серии, подвергаются сплошному контролю.



# Объем выборки

Число наблюдений  $n$ , образующих выборку, называется **объемом выборки**. Если объем выборки  $n$  достаточно велик ( $n \rightarrow \infty$ ), выборка считается **большой**, в противном случае она называется выборкой **ограниченного объема**.

Выборка считается **малой**, если при измерении одномерной случайной величины  $X$  объем выборки не превышает 30 ( $n \leq 30$ ), а при измерении одновременно нескольких ( $k$ ) признаков в многомерном пространстве отношение  $n$  к  $k$  не превышает 10 ( $n/k < 10$ ).

Если мы работаем с обычной выборкой, то используется таблица «Интеграла вероятностей закона нормального распределения». В случае малой выборки необходимо пользоваться таблицей «Распределение Стьюдента», при этом число степеней свободы :

$$K=n-1$$

## выборки.

- Для определения необходимой численности выборки исследователь должен задать уровень точности выборочной совокупности с определенной вероятностью. В частности, необходимая численность случайной повторной выборки определяется по формуле: 
$$n = \frac{t^2 \cdot \sigma^2}{\Delta^2}$$
- которая вытекает из формулы предельной ошибки  $\Delta = t \sqrt{\frac{\sigma^2}{n}}$
- Эта формула показывает, что с увеличением предполагаемой ошибки выборки значительно уменьшается необходимый объем выборки. Так, увеличение допустимой ошибки выборки в 2 раза уменьшает необходимый ее объем в 4 раза. Необходимая численность выборки прямо пропорциональна дисперсии признака и величине  $\sigma^2$ .



- Вы сняты с выборов за отсутствием у вас разрешения на полеты!

Вид выборочного наблюдения	Повторный отбор	Бесповторный отбор
<b>Собственно-случайная выборка, механическая выборка:</b>		
<b>а) при определении среднего размера признака</b>	$n = \frac{t^2 \cdot \sigma_{\tilde{x}}^2}{\Delta_{\tilde{x}}^2}$	$n = \frac{t^2 \cdot \sigma_{\tilde{x}}^2 \cdot N}{\Delta_{\tilde{x}}^2 \cdot N + t^2 \cdot \sigma_{\tilde{x}}^2}$
<b>б) при определении доли признака</b>	$n = \frac{t^2 \cdot w(1-w)}{\Delta_w^2}$	$n = \frac{t^2 \cdot w(1-w) \cdot N}{\Delta_w^2 \cdot N + t^2 \cdot w(1-w)}$

**Пример.** В городе 2000 семей. Предполагается провести выборочное обследование методом случайной бесповторной выборки для нахождения среднего размера семьи.

**Определить необходимую численность выборки** при условии, что с вероятностью 0,954 ошибка выборки не превысит 1 человека при среднем квадратическом отклонении 3 человека.

**Решение**

$$n = \frac{t^2 \cdot \sigma^2 \cdot N}{N \cdot \Delta_{\bar{x}}^2 + t^2 \cdot \sigma^2} = \frac{4 \cdot 9 \cdot 2000}{2000 \cdot 1 + 4 \cdot 9} = 36$$

**Т.е. необходимо обследовать не менее 36 семей.**