

# ТЕМА 3. ФОРМИРОВАНИЕ ВЫБОРКИ В МАРКЕТИНГОВЫХ ИССЛЕДОВАНИЯХ

*Математика – единственный совершенный метод,  
позволяющий провести самого себя за нос.*

*Альберт Эйнштейн*

1. **Статистика в маркетинговых исследованиях.**
2. **Выборочный метод в маркетинговом исследовании.**

# 1. Статистика в маркетинговых исследованиях

## *Статистические величины в маркетинге*

- абсолютные данные;
- относительные данные;
- средние величины.

# Пример

Динамика объемов реализации в торговле 2010-2015 (млрд руб)

## Оборот розничной торговли



— с исключением сезонных и календарных факторов (январь 1995 - 100%)

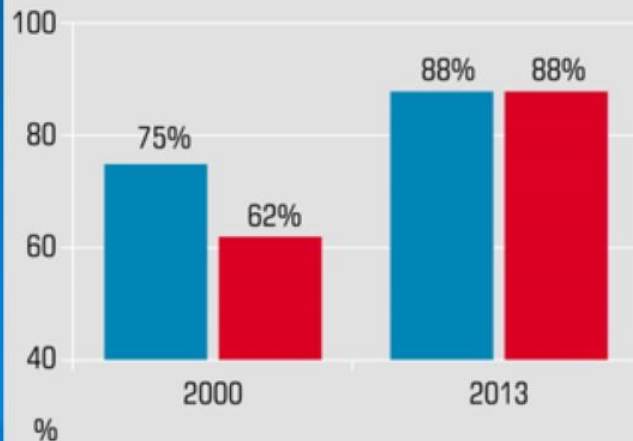
Источник: расчеты Минэкономразвития России



■ Розничная торговля ■ Оптовая торговля

В Белгородской и Калужской областях оборот розничной торговли на душу населения растет быстрее, чем в РФ

Динамика оборота розничной торговли на душу населения относительно среднего по России



■ Калужская обл. ■ Белгородская обл.

Источники: Росстат, расчеты «Эксперта»

# Абсолютные величины в маркетинговом исследовании



# Абсолютные величины в маркетинговом исследовании (пример)

## Зарплата учителей в разных странах мира

Страна	З/п учителя*, \$/год	З/п учителя*, \$/месяц	З/п учителя*, руб./месяц	
Мексика	68 343	5 695	375 830	
Германия	61 317	5 110	337 191	
США	42 695	3 558	234 787	
Испания	40 752	3 396	224 101	
Австралия	39 125	3 260	215 155	
Нидерланды	38 473	3 206	211 569	
Франция	30 651	2 554	168 557	
Юж. Корея	29 252	2 438	160 860	
Англия	27 768	2 314	152 704	
Япония	27 627	2 302	151 927	
Греция	17 760	1 480	97 663	
Россия	5 724	477	31 477	

\*Начальная з/п учителя стандартной квалификации в старших классах школы

По данным ЦБ РФ и организации экономического сотрудничества и развития (OECD)

mel.fm

# Относительные величины



# Относительные величины (пример 1)

## Отношение зарплаты учителей к минимальному размеру оплаты труда

Страна	З/п учителя, \$/год	МРОТ*, \$/год	Отношение з/п учителей к МРОТ
Мексика	68 343	1 714	39,9 
Испания	40 752	12 170	3,3 
Россия	5 724	1 940	3,0 
США	42 695	15 080	2,8 
Германия	61 317	23 750	2,6 
Юж. Корея	29 252	12 170	2,4 
Япония	27 627	13 606	2,0 
Нидерланды	38 473	24 273	1,6 
Греция	17 760	10 886	1,6 
Австралия	39 125	30 791	1,3 
Франция	30 651	23 350	1,3 
Англия	27 768	23 472	1,2 

\*Минимальный размер оплаты труда

По данным международного банка реконструкции и развития (IBRD) и организации экономического сотрудничества и развития (OECD)

mel.fm

## Отношение зарплаты учителей к зарплате главы государства

Страна	З/п президен. \$/год	З/п учителя, \$/год	Процентное соотношение
Испания	84 240	40 752	48% 
Мексика	244 900	68 343	28% 
Германия	244 000	61 317	25% 
Нидерланды	204 000	38 473	19% 
Греция	95 350	17 760	19% 
Юж. Корея	178 000	29 252	16% 
Франция	202 000	30 651	15% 
Япония	203 000	27 627	14% 
Англия	206 000	27 768	13% 
США	400 000	42 695	11% 
Австралия	396 000	39 125	10% 
Россия	136 000	5 724	4% 

По данным 24/WSJ и организации экономического сотрудничества и развития (OECD)

mel.fm

# Относительные величины (пример 2)

## Месячная зарплата учителя в Айфонах

Страна	Стоимость iPhone 6s (16 GB), \$	З/п учителя в Айфонах (в месяц)
Мексика	762,46	7,5 
Германия	823,1	6,2 
США	649	5,5 
Австралия	774,07	4,2 
Нидерланды	834,24	3,8 
Испания	956,75	3,5 
Франция	834,24	3,1 
Англия	782,41	3,0 
Юж. Корея	890,05	2,7 
Япония	932,62	2,5 
Греция	879,02	1,7 
Россия	730,27	0,7 

Цены по данным apple.com

mel.fm

## Дневная зарплата учителя в Биг Маках

Страна	Стоимость Биг Мака, \$	З/п учителя в Биг Маках (в день)
Мексика	2,81	68 
Германия	3,86	44 
Испания	3,76	30 
Австралия	3,74	29 
Нидерланды	3,71	29 
Япония	3,12	25 
США	4,93	24 
Юж. Корея	3,59	23 
Франция	4,41	19 
Англия	4,22	18 
Греция	3,6	14 
Россия	1,53	10 

Цены по данным The Economist

mel.fm



# Статистические величины в маркетинговом исследовании



## Значение средних величин для маркетинговых исследований

- характеризуют общий уровень рынка, или его сегмента.
- погашают индивидуальные отклонения от проявляющейся тенденции или колеблемости рыночного развития.
- отражают типичный уровень и при этом позволяет абстрагироваться от индивидуальных значений.
- в значении средней величины проявляются основные факторы, а не второстепенные или случайные.

## Средняя арифметическая простая (невзвешенная)

$$\bar{x}_{\text{Ар.}} = \frac{\sum_{i=1}^n x_i}{n}$$

## Средняя арифметическая взвешенная

$$\bar{x}_{\text{Ар.}} = \frac{\sum_{i=1}^n x_i F_i}{\sum_{i=1}^n F_i},$$

где  $\bar{x}_{\text{Ар.}}$  — средняя арифметическая величина;  
 $x_i$  — индивидуальная величина (вариант) осредняемого признака;  
 $F_i$  — «вес»  $i$ -го варианта;  
 $n$  — число вариантов (количество единиц).

## Средняя гармоническая

$$\bar{x}_{\text{Гарм. 1}} = \frac{n}{\sum_i \frac{1}{x_i}}; \quad \bar{x}_{\text{Гарм. 2}} = \frac{\sum_i W_i}{\sum_i \frac{W_i}{x_i}},$$

- где  $\bar{x}_{\text{Гарм. 1}}$  — средняя гармоническая невзвешенная (простая);  
 $\bar{x}_{\text{Гарм. 2}}$  — средняя гармоническая взвешенная;  
 $x_i$  — осредняемая величина;  
 $W_i$  — веса осредняемой величины;  
 $n$  — число осредняемых величин.

## 2. Выборочный метод в маркетинговом исследовании

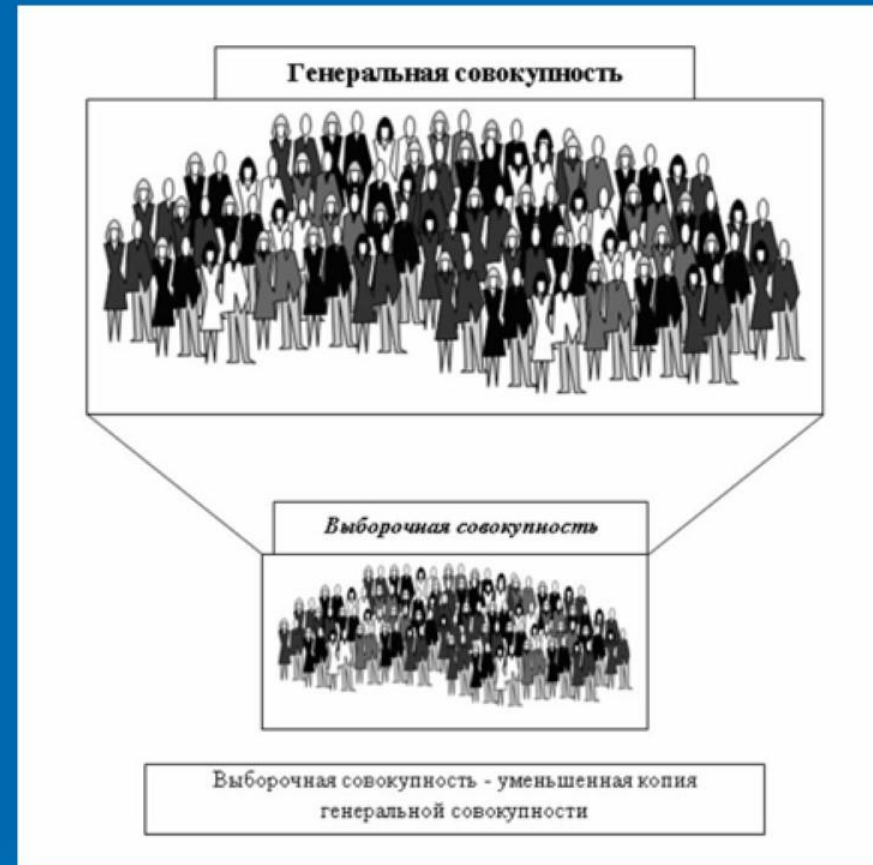
Выборочный метод (*выборка*) – статистический прием, который базируется на некоторых положениях закона больших чисел.

### Преимущества

- дает возможность экономить значительные средства;
- позволяет ускорить получение нужных сведений;
- обеспечивает сбор сведений, которые иным способом получить невозможно.

**Выборка или выборочная совокупность** – часть генеральной совокупности элементов, которая охватывается маркетинговым исследованием.

**Генеральная совокупность** – вся совокупность объектов, принимающих участие в исследовании.



# Требования, предъявляемые к выборке



## Этапы построения выборки

1. Определение исследуемой совокупности.
2. Выбор основы для построения выборки.
3. Выбор способа построения выборки.
4. Определение требуемого размера выборки.
5. Реализация процесса построения выборки.





# Этап 1. Определение исследуемой совокупности (target population)

**Элемент исследуемой совокупности** – объект, о котором или от которого должна быть получена информация.

При опросах это, как правило, потенциальный респондент.

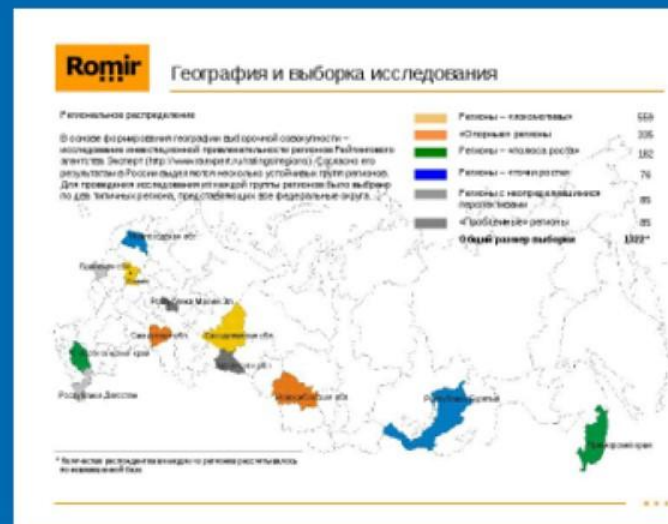
**Единицы отбора** – элементы или группы элементов, которые доступны для отбора в ходе построения выборки.

# Пример



# Этап 2. Основа для построения выборки

- списки элементов исследуемой совокупности или правила нахождения этих элементов.

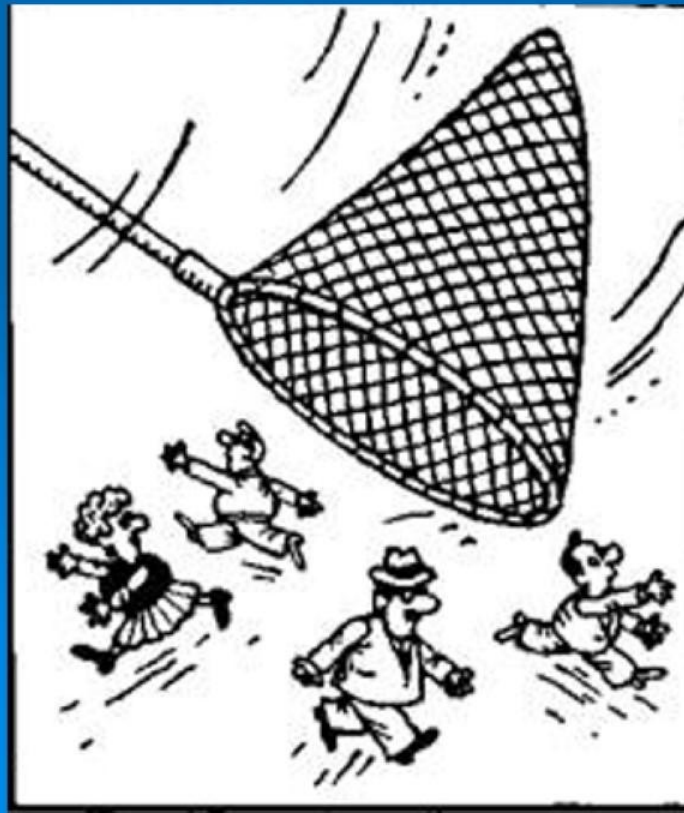


## Этап 3. Способы построения выборки

- вероятностные методы построения выборки,
- детерминированные методы построения выборки.

## Вероятностные методы построения выборки

- основаны на использовании теории вероятности и базируются на том, что репрезентация будет достигнута в случае, если каждой единице генеральной совокупности будет обеспечено равновероятное попадание в выборку.



# Виды выборок

- Простая случайная выборка.
- Систематическая выборка.
- Стратифицированная выборка.
- Кластерная выборка.

Жребий (к/ф «Гараж»)

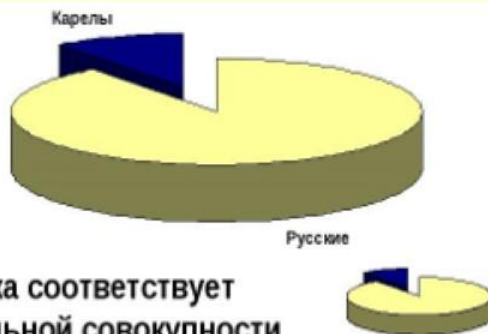
Кубики

Генератор случайных чисел

Программа PERI



1 2 3 4 5 6 7 8 9 10 11 12 13 14  
15 16...



Выборка соответствует  
генеральной совокупности  
по структуре



# Недостатки вероятностных методов построения выборки

- предположение о том, что все элементы генеральной совокупности являются равнодоступными: и в «техническом» смысле, и в «психологическом».
- граждане, с относительно высокими доходами, менее доступны для исследователей, чем те, чьи доходы невысоки.
- Менее образованные граждане идут на контакт с социологами-маркетологами менее охотно, чем лица с высшим образованием.

- Никто из сотрудников исследовательских компаний (а также сотрудников заказчиков) не желает общаться с бомжами, алкоголиками, наркоманами, психо-, социопатами и прочими маргиналами.
- Некоторые граждане боятся отвечать на вопросы, даже самые невинные («квестофобия»).
- Есть люди, которые просто не желают участвовать в исследовании.



**Все выборки в маркетинге и социологии являются смещенными в сторону средне и малообеспеченных, более образованных, контактных и вменяемых граждан.**

**Они и репрезентируют общее мнение генеральной совокупности.**

### **Способ преодоления**

- **метод «квот» — пропорциональное представительство носителей существенных признаков (пол, возраст, доход, образование и т.п.) генеральной совокупности в выборке.**

## Детерминированные методы построения выборки

- отбор осуществляется не по принципам случайности, а по субъективным критериям – доступности, типичности, равного представительства и т. д.

### Виды выборок

- Нерепрезентативная.
- Поверхностная.
- Квотная.
- «Снежный ком».

# Этап 4. Определение необходимого размера выборки

## Подходы к определению объема выборки

### 1. Произвольный подход

- основан на применении «правила большого пальца».
- например, бездоказательно принимается, что для получения точных результатов выборка должна составлять 5 % от совокупности.

## 2. Объем выборки может быть установлен исходя из заранее оговоренных условий.

- выборка обычно составляет 1000 – 1200 человек.

### Минимальный и типичный размеры выборки

Задачи исследования	Минимальный размер выборки	Типичный размер выборки
Выявление проблем (например, оценка потенциала выборки)	500	1000-2500
Решение проблем (например, относительно цены товара)	200	300-500
Тестирование товара	200	300-500
Тестирование рекламы (в расчете на 1 рекламный продукт)	150	200-300

# Расчеты репрезентативной выборки

Объем генеральной совокупности	500	1000	2000	3000	4000	5000	10000	100000
Объем выборки	222	286	333	350	360	370	385	358

**3. Главный аргумент при определении объема выборки – стоимость проведения опроса.**

**4. Объем выборки определяется на основе статистического анализа.**

- предполагает определение минимального объема выборки с учетом требований к надежности и достоверности получаемых результатов.

## 5. Подход основан на расчете доверительного интервала.

Достоверность (уровень достоверности) – показатель вероятности того, что истинное значение изучаемого параметра генеральной совокупности попадет в доверительный интервал.

- Чем выше задаваемый уровень достоверности, тем больше должна быть выборка.

**Доверительный интервал** – диапазон, в который попадет истинное значение изучаемого параметра генеральной совокупности при данном уровне достоверности.

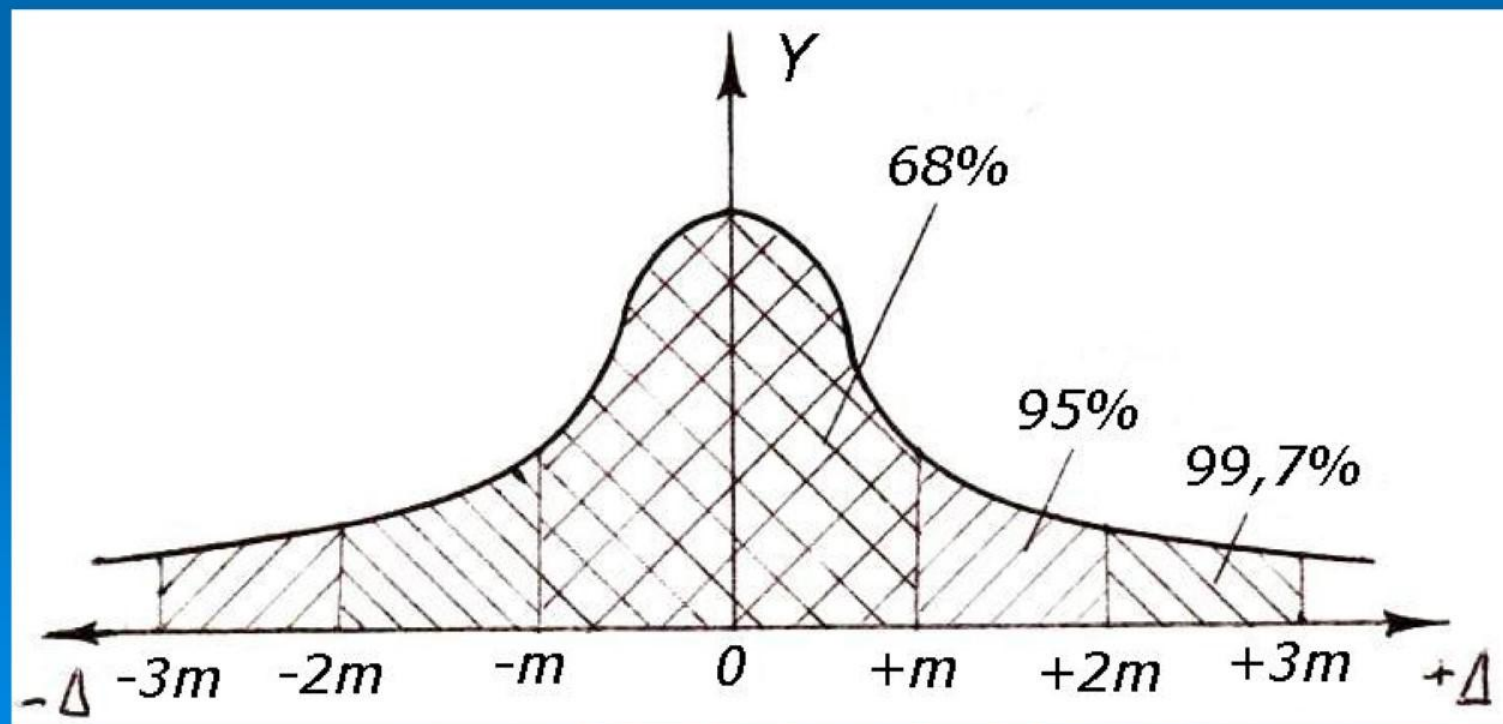
- Чем меньше доверительный интервал, тем больше должна быть выборка.

**Доверительный интервал** – диапазон, крайние точки которого характеризуют процент определенных ответов на какой-то вопрос.



# Значение нормированного отклонения оценки $z$ от среднего значения в зависимости от доверительной вероятности (а) полученного результата

$a, \%$	60	70	80	90	90	95	97	99,0	99,7
$z$	0,84	1,03	1,29	1,44	1,65	1,96	2,18	2,58	3,0



## Доверительный интервал

- «допустимая погрешность» выборки. Ее исследователи задают (назначают) сами или рассчитывают по специальным формулам в зависимости от заданного объема выборки. Измеряется «допустимая погрешность» в процентных пунктах – п.п.
- индикатор степени отличия оценки, истинной для совокупности в целом, которая ожидается для типичной выборки, – среднеквадратическая ошибка.

## Формулы для расчета размера выборки

- Формула для расчета необходимого объема выборки при известном среднем отклонении (дисперсии) и заданных уровнях достоверности и точности.

$$N = \frac{G^2 \times Z^2}{D^2}$$

где  $N$  – искомый объем выборки;

$G$  – дисперсия признака, ожидаемое среднее отклонение получаемых результатов от ожидаемого среднего значения;

$Z$  – коэффициент уровня достоверности (2 – для 0,95; 3 – для 0,99);

$D$  – желаемый уровень точности, доверительный интервал, допустимая погрешность.

## Пример

Мы изучаем поведение покупателей в продовольственном магазине и хотим определить среднюю сумму чека. Из бесед с владельцем магазина мы узнаем, что она может быть в районе 500—700 руб., а среднее отклонение может составить 200 руб. В ходе опроса мы хотели бы определить среднее значение с точностью до 20 руб.

$$N = \frac{G^2 \times Z^2}{D^2}$$

где  $N$  – искомый объем выборки;

$G$  – дисперсия признака, ожидаемое среднее отклонение получаемых результатов от ожидаемого среднего значения (200);

$Z$  – коэффициент уровня достоверности (2 для 0,95);

$D$  – желаемый уровень точности, доверительный интервал, допустимая погрешность (20).

$$200^2 \times 2^2 / 20^2 = 40000 \times 4/400 = 400.$$

Если бы мы хотели узнать среднюю с точностью до 10 руб. и уровнем достоверности 0,95, то сколько нам надо опросить?

**1600 покупателей (40 000 X 4/100).**

Если бы мы хотели узнать среднюю с точностью до 10 руб. и уровнем достоверности 0,99, то сколько нам надо опросить?

**3600 человек (40 000 X 9/100).**

Если нас устроила бы точность +/-50 руб., уровень достоверности 0,95, то сколько нам надо опросить?

**64 человека (40 000 X 4/2500).**

- «Каноническая» формула расчета объема выборки.

$$n = \frac{1}{(d / p(1 - p)t) + 1 / N}$$

где  $n$  – искомый объем генеральной выборки;

$d$  – квадрат ошибки выборки, он же – допустимая погрешность (задается по нашему желанию);

$p$  – ожидаемая доля ответов «ДА»;

$t$  – квадрат значения статистики Стьюдента, он же – коэффициент уровня достоверности;

$N$  – размер генеральной совокупности.

## Пример.

Мы изучаем отношение взрослых горожан к идее запрета на курение в общественных местах в г. Вологде. Из предыдущих исследований известно, что доля курящих во взрослом населении города составляет 40% (ожидаемая доля ответов «Да» = 0,4). Квадрат значения статистики Стьюдента пусть будет равен 4 (коэффициент уровня достоверности = 0,95). Допустимая погрешность, ошибка выборки — 4 п.п. (0,04).

Какова выборка?



- «Каноническая» формула расчета объема выборки.

$$n = \frac{1}{(d / p(1 - p)t) + 1 / N}$$

где  $n$  – искомый объем генеральной выборки;

$d$  – квадрат ошибки выборки, он же – допустимая погрешность (задается по нашему желанию) – 0,04;

$p$  – ожидаемая доля ответов «ДА» (0,4);

$t$  – квадрат значения статистики Стьюдента, он же – коэффициент уровня достоверности (4 при 0,95);

$N$  – размер генеральной совокупности (500000).

$$n = 1 / ((0,0016 / 0,4 \times 0,6 \times 4) + 0,000005) = 599$$

**человек.**

Какова выборка в Москве (8 млн)?

$$n = 1 / ((0,0016 / 0,4 \times 0,6 \times 4) + 0,000000125) = 600$$

**человек.**

Уровень достоверности 0,99 (квадрат коэффициента Стьюдента = 9) и допустимую погрешность в 3 п.п.

$$n = 1 / ((0,0009 / 0,4 \times 0,6 \times 9) + 0,000005) = 2375$$

**человек.**

Квадрат значения статистики Стьюдента 4. Допустимая погрешность, ошибка выборки пусть теперь будет 5 п.п. (0,05).

$$n = 1 / ((0,0025 / 0,4 \times 0,6 \times 4) + 0,000005) = 385 \text{ человек.}$$

- Формула для расчета выборки для опроса с простыми альтернативными вопросами по принципу «Да/Нет»

$$n = \frac{z^2 (p \times q)}{e^2}$$

где  $n$  – объем выборки;

$z$  – нормированное отклонение, определяемое исходя из выбранного уровня доверительности;

$p$  – найденная вариация для выборки;

$q = (100 - p)$ ;

$e$  – допустимая ошибка.

- Формула для расчета выборки для опроса с простыми альтернативными вопросами по принципу «Да/Нет»

$$n = \frac{z^2 (p \times q)}{e^2}$$

где  $n$  – объем выборки;

$z$  – нормированное отклонение, определяемое исходя из выбранного уровня доверительности;

$p$  – найденная вариация для выборки;

$q = (100 - p)$ ;

$e$  – допустимая ошибка.

## Пример

Предприятием, выпускающим покрышки, проводится опрос автолюбителей, использующих радиальные покрышки.

На вопрос: «Используете ли Вы радиальные покрышки?» возможны только 2 ответа: «Да» или «Нет».

Принимая вариацию равной 50 %, точность равной 10 % при 95 %-м уровне доверительности, каков размер выборки?

$$n = \frac{z^2 (p \times q)}{e^2}$$

$$n = 1,96^2 (50 \times 50) / 10^2 = 96.$$

При уровне доверительности равном 99 %, и  $e = \pm 3$  %, чему равен объем выборки?

$$n = 1067.$$

- **Формула случайной бесповторной выборки при организации устных опросов**

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

- где
- $n$  — число единиц, которые следует отобрать для обследования (опроса);
  - $t$  — коэффициент доверия, вычисляемый по таблицам в зависимости от вероятности с которой можно гарантировать, что предельная ошибка не превысит  $t$ -кратную среднюю ошибку (при вероятности 0,990 он равен 3, а при вероятности 0,999—3,28, чаще всего опираются на вероятность 0,954, при которой  $t$  составляет 2);
  - $\sigma^2$  — выборочная дисперсия (ее обычно определяют на основе эксперимента, пробного обследования, или же по аналогам);
  - $\Delta$  — предельная (задаваемая) ошибка выборки;
  - $N$  — численность генеральной (изучаемой) совокупности.

## Пример

В процессе маркетингового исследования рыночной ситуации требуется определить, сколько потребуется анкет для получения обоснованного вывода о спросе на товар Q?

Известны следующие характеристики:

массив покупателей (генеральная совокупность) – 100 тыс. чел. ( $N = 100000$ );

выборочная дисперсия (определенная по прошлым исследованиям –  $\sigma^2$ ) составила:  $\pm 900$  руб.;

коэффициент доверия:  $t = 2$ ; задаваемая (предельная) ошибка выборки:  $\Delta = 1$ .



$$n = (2^2 \times 900 \times 100000) / (1 \times 100000 + 4 \times 900) = 3475 \approx 3500$$

Округленно численность выборки составляет 3500 чел. или 3,5%. Рекомендуется увеличить численность выборки до 5%, поскольку часть анкет может быть возвращена незаполненными.

- **Формула расчета для малой выборки**  
**Рекомендуемый объем выборки при «малой» генеральной совокупности**

<b>Объем генеральной совокупности</b>	<b>Рекомендуемый объем выборки</b>
<b>500</b>	220
<b>1000</b>	280
<b>2000</b>	330
<b>3000</b>	350
<b>5000</b>	370
<b>7000</b>	385
<b>10 000</b>	400

Если  $V$  выборки превышает 5 % совокупности, то последняя считается малой, и в вышеприведенные формулы вводится поправочный коэффициент.

$$n_1 = n \sqrt{\frac{N - n}{N - 1}}$$

где  $n_1$  – объем выборки для малой совокупности,  
 $n$  – объем выборки,  
 $N$  – объем генеральной совокупности.

## Пример

Изучается мнение членов совокупности, состоящей из 1000 компаний, относительно строительства химического комбината в границах города X. Вследствие отсутствия информации о вариации принимается наихудший случай: 50:50. Исследователь вынес решение использовать уровень доверительности равный 95 %. Заказчик исследования указал, что его устроит точность результатов плюс минус 5 %.

$$n_1 = \frac{z^2(p \times q)}{e^2} X \sqrt{\frac{N-n}{N-1}}$$

$$N = 1000$$

$$p=50, q=100-50.$$

Уровень доверительности равный 95 %  $\Rightarrow z=1,96$

$$e = 5 \%$$

$$n_1 = 1,96^2 \times (50 \times 50) / 5^2 \times \sqrt{((1000-384)/(1000-1))} = \\ = 384 \times 0,79 = 303$$

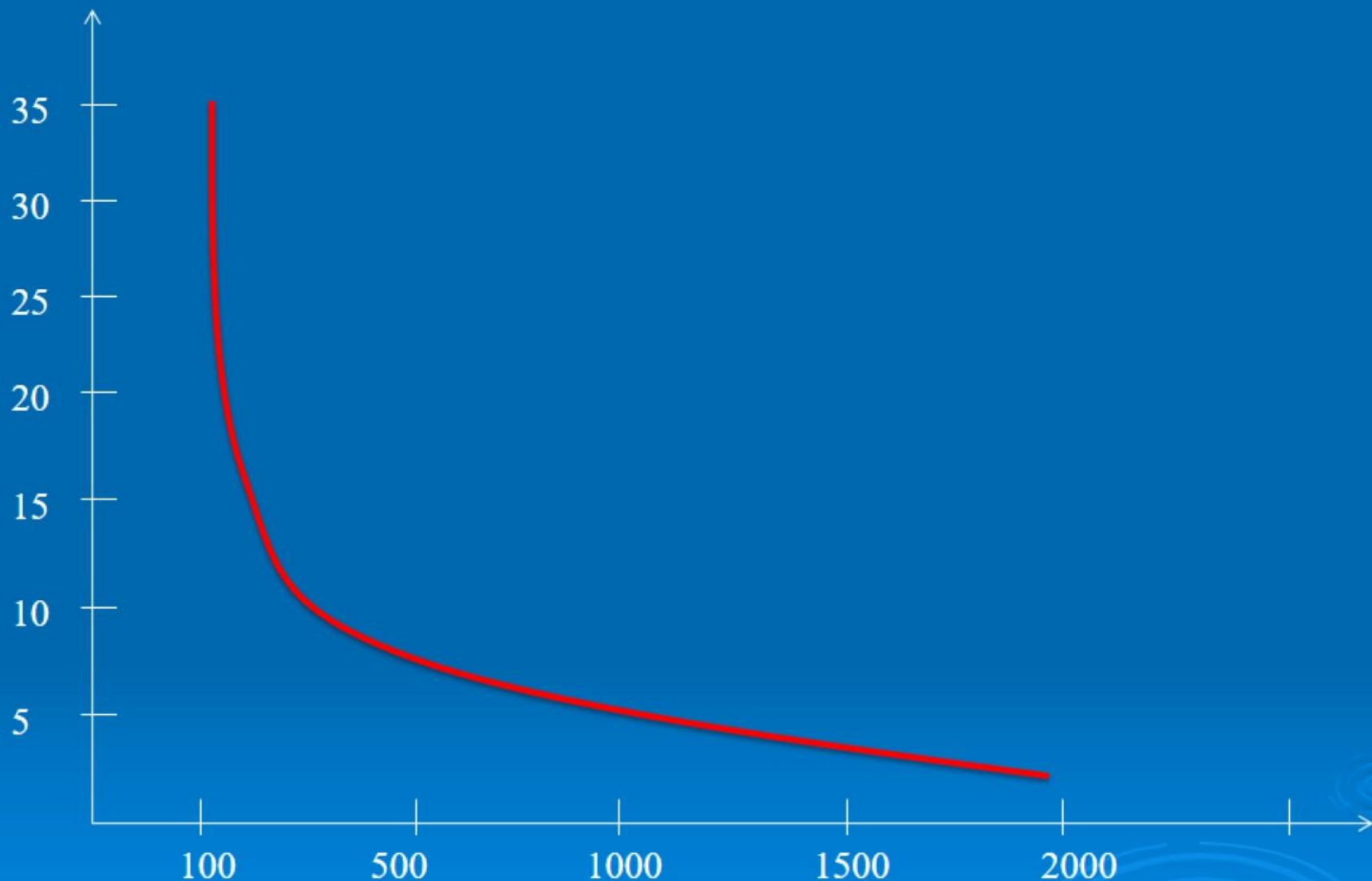
Диапазон погрешности рассчитывается по следующей формуле:

$$\sqrt{\frac{1 p(p - 1)}{N}}$$

где:

$N$  – объем выборки,

$p$  и  $(p - 1)$  – величина типологических групп, определенная по какому-либо параметру, например, по возрасту или полу.



## Зависимость погрешности от объема выборки

# Этап 5. Реализация процесса построения выборки

- как будут отбираться респонденты?





## Карты Л. Киша

- респонденты разного возраста отбираются с вероятностью, пропорциональной их доле в исследуемой совокупности, причем каждый интервьюер действует строго однозначным, полностью контролируемым образом.

# Правила использования карт отбора респондентов

№ карты	Доля отобранных по карте респондентов	Порядковый (по увеличению возраста) номер члена семьи, которого следует опросить, если число подходящих для опроса членов семьи составляет:					
		1	2	3	4	5	6 и более
1	1/6	1	1	1	1	1	1
2	1/12	1	1	1	1	2	2
3	1/12	1	1	1	2	2	2
4	1/6	1	1	2	2	3	3
5	1/6	1	2	2	3	4	4
6	1/12	1	2	3	3	3	5
7	1/12	1	2	3	4	5	5
8	1/6	1	2	3	4	5	6

# Пример. Правильно ли сформирована выборка???

## С чем вы готовы мириться, приобретая новый дешевый (бюджетный) автомобиль?

