Выборочный метод в исследовании

БАННЫХ Г.А.

К.С.Н.,

ДОЦЕНТ

Ангус Дитон

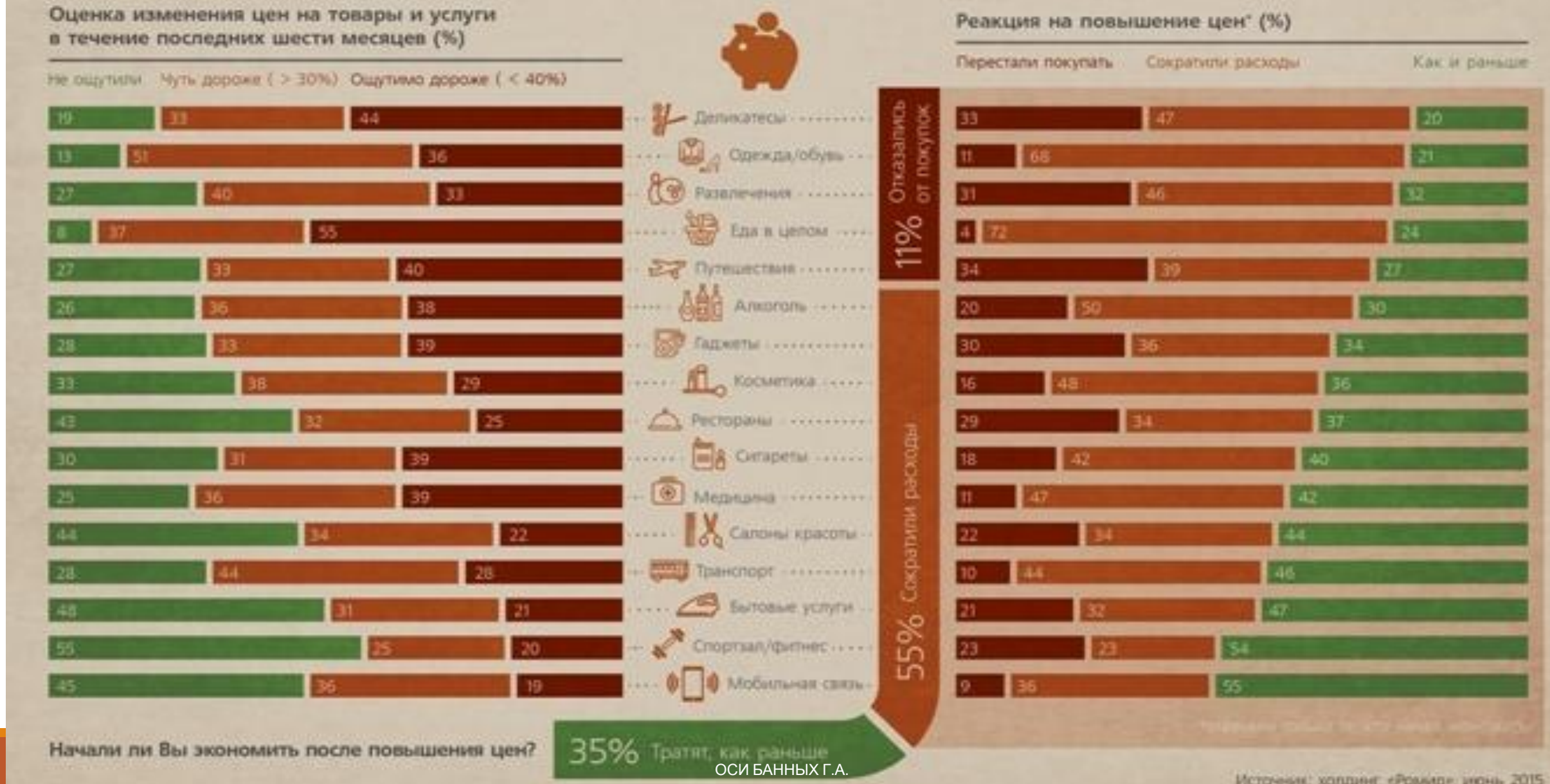
- восхождение от частного к общему

- Почти идеальная система спроса

- Парадокс Дитона

- изучение «счастья»

ИЗМЕНЕНИЕ РАСХОДОВ РОССИЯН



Источник: холдинг «Ромир», июнь 2015

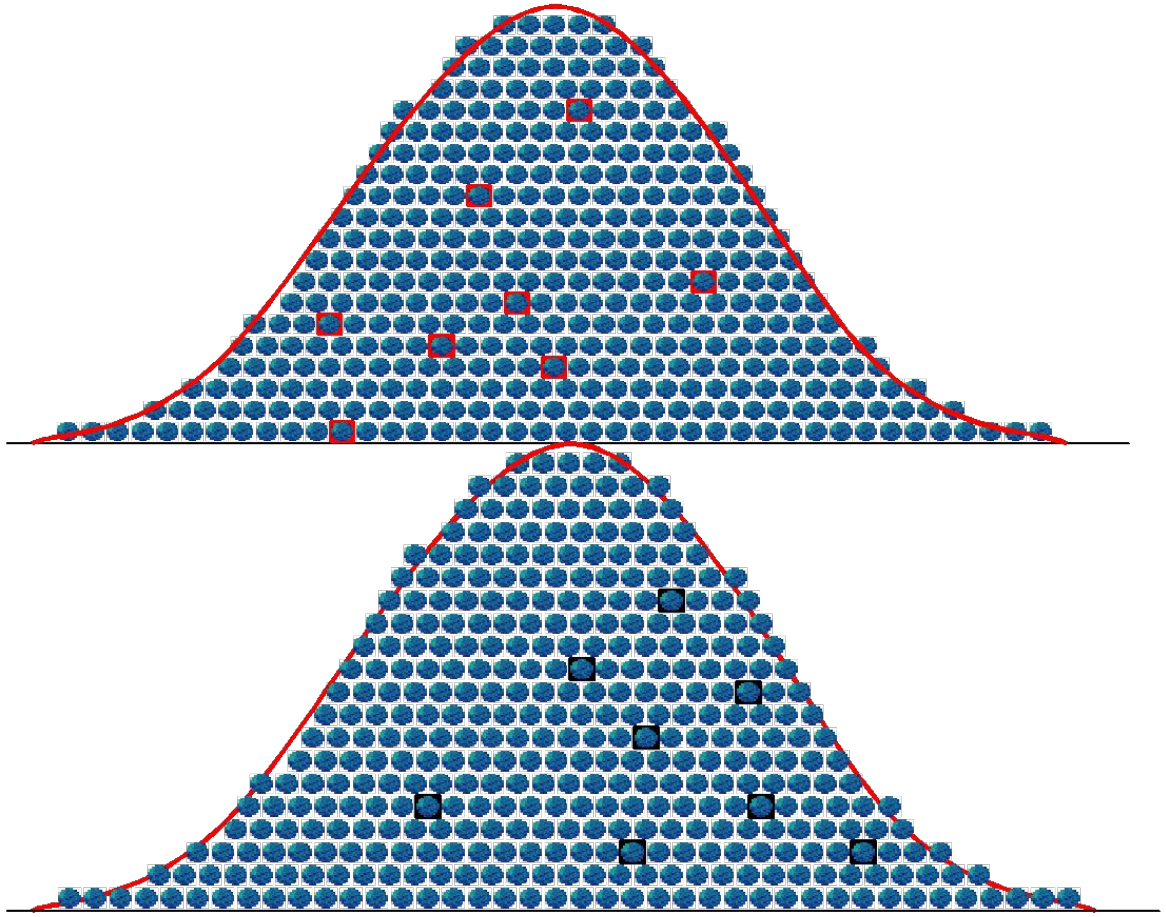
Исследования сплошные и выборочные

Что такое сплошные и выборочные исследования?

Понятие о генеральной совокупности

- **Все субъекты, подходящие для проведения исследования**
- **Самая большая генеральная совокупность – все человечество**
- **Понятие выборки - Sample**

Генеральная совокупность и выборка



Генеральная совокупность – это полная совокупность единиц (вся статистическая совокупность).
Выборочная совокупность (***выборка***) - это часть единиц генеральной совокупности, отобранная в случайном порядке.

Обозначения:
объем генеральной совокупности – N;
объем выборки - n

Генеральная совокупность

Суммарная численность объектов наблюдения (люди, домохозяйства, предприятия, населенные пункты и т.д.), обладающих определенным набором признаков (пол, возраст, доход, численность, оборот и т.д.), ограниченная в пространстве и времени.

Почему?

с целью экономии времени и средств в результате сокращения объема работы (10-25%)

- вследствие того, что исследуемая совокупность может быть полностью недоступна

вследствие того, что исследуемая совокупность может не иметь конечного объема

- ???

Почему исследователи любят выборочные исследования

Меньше работы

Меньше денег

Меньше времени

Меньше головной боли (?)

Допущение

Результаты, полученные на выборке, можно распространить на всю генеральную совокупность (с известным допущением, выражаемым через вероятности ά- и β-ошибок)

Оптимальный размер выборки

В интересах исследователя

В интересах реально существующей ситуации в генеральной совокупности

Выборка (Выборочная совокупность)

- Часть объектов из генеральной совокупности, отобранных для изучения, с тем чтобы сделать заключение обо всей генеральной совокупности.

- **Отобранное по строго заданному правилу число элементов генеральной совокупности, подлежащее непосредственному исследованию.**

Единицы анализа **– элементы выбранной, или обследуемой совокупности. Ими могут быть как индивиды, так и, например, учебные группы, бригады и т.д.**

Для того чтобы заключение, полученное путем изучения выборки, можно было распространить на всю генеральную совокупность, выборка

Репрезентативность выборки

Свойство выборки корректно отражать генеральную совокупность.

Одна и та же выборка может быть репрезентативной и нерепрезентативной для разных генеральных совокупностей.

ОСИ БАННЫХ Г.А.

ІНТЕР
КАРТОННИЙ ФЛЕШМОБ
Как поступить с Востоком Украины? www.podrobnosti.ua
Зачистить силой 45.46 %
Депортировать 2.51 %
Договориться 27.68 %
Отделить 24.35 %
ЧЕРНОЕ ЗЕРКАЛО

Обеспечение репрезентативности

Статистические данные для расчета распределения выборки по типам поселений и демографическим группам, тыс. человек

			11	12	13	14	21	22	23	24	
Код района	Район	Всего в возрасте 16 лет и старше	Мужчины 16-29	Мужчины 30-44	Мужчины 45-59	Мужчины 60 лет и старше	Женщины 16-29	Женщины 30-44	Женщины 45-59	Женщины 60 лет и старше	FIN
1	(1) Тракторозаводский р	116.78	14.26	14.67	13.79	9.31	14.62	14.98	17.34	17.81	FIN
2	(2) Краснооктябрьский р	133.43	15.72	15.39	15.55	11.14	16.62	17.07	20.25	21.69	FIN
3	(3) Центральный район	77.37	11.38	8.88	7.57	6.32	12.12	9.19	9.78	12.13	FIN
4	(4) Дзержинский район	148.45	22.71	17.57	17.13	10.53	19.48	19.14	22.13	19.78	FIN
5	(5) Ворошиловский рай	68.08	8.67	7.65	7.46	5.51	9.13	8.59	10.05	11.03	FIN
6	(6) Советский район	70.51	9.99	7.82	7.77	5.37	10.51	8.66	9.98	10.42	FIN
7	(7) Кировский район	87.78	11.44	11.19	10.15	6.97	11.06	11.56	12.27	13.15	FIN
8	(8) Красноармейский ра	146.18	17.50	18.50	16.95	11.85	18.11	19.36	21.39	22.51	FIN
9	(9) Волжский	273.00	35.77	33.61	32.55	20.37	35.57	35.42	41.79	37.91	FIN
10	(9) Камышин	101.34	15.80	12.14	11.39	8.17	12.09	12.03	14.45	15.29	FIN
FIN	FIN	FIN	FIN	FIN	FIN	FIN	FIN	FIN	FIN	FIN	FIN
Всего жителей, тыс. человек		1223	163	147	140	96	159	156	179	182	FIN

Для репрезентативной выборки выборочное распределение должно соответствовать генеральной совокупности по основным контролируемым признакам (в данном случае – район проживания, пол и возраст)

Обеспечение репрезентативности

Статистические данные для расчета распределения выборки по типам поселений и демографическим группам, тыс. человек

			11	12	13	14	21	22	23	24	
Код района	Район	Всего в возрасте 16 лет и старше	Мужчины 16-29	Мужчины 30-44	Мужчины 45-59	Мужчины 60 лет и старше	Женщины 16-29	Женщины 30-44	Женщины 45-59	Женщины 60 лет и старше	FIN
1	(1) Тракторозаводский р								17.34	17.81	FIN
2	(2) Краснооктябрьский р									21.69	FIN
3	(3) Центральный район								9.78	12.13	FIN

Распределение населения по районам

Код района	Район	Всего в возрасте 16 лет и старше	Мужчины 16-29	Мужчины 30-44	Мужчины 45-59	Мужчины 60 лет и старше	Женщины 16-29	Женщины 30-44	Женщины 45-59	Женщины 60 лет и старше	FIN
1	(1) Тракторозаводский р								1.4%	1.5%	FIN
2	(2) Краснооктябрьский р								1.7%	1.8%	FIN
3	(3) Центральный район								0.8%	1.0%	FIN
4	(4) Дзержинский район	12.1%	1.9%	1.4%	1.4%	0.9%	1.6%	1.6%	1.8%	1.6%	FIN
5	(5) Ворошиловский рай	5.6%	0.7%	0.6%	0.6%	0.5%	0.7%	0.7%	0.8%	0.9%	FIN
6	(6) Советский район	5.8%	0.8%	0.6%	0.6%	0.4%	0.9%	0.7%	0.8%	0.9%	FIN
7	(7) Кировский район	7.2%	0.9%	0.9%	0.8%	0.6%	0.9%	0.9%	1.0%	1.1%	FIN
8	(8) Красноармейский ра	12.0%	1.4%	1.5%	1.4%	1.0%	1.5%	1.6%	1.7%	1.8%	FIN
9	(9) Волжский	22.3%	2.9%	2.7%	2.7%	1.7%	2.9%	2.9%	3.4%	3.1%	FIN
							1.0%	1.0%	1.2%	1.3%	FIN
							FIN	FIN	FIN	FIN	FIN

Доля женщин в возрасте 60 лет и старше, проживающих в Тракторозаводском районе Волгограда (17.81 тыс. чел.), составляет 1.5% в общем объеме объекта исследования (1227 тыс. чел.)

Для репрезентативной выборки выборочное распределение должно соответствовать генеральной совокупности по основным контролируемым признакам (в данном случае – район проживания, пол и возраст)

Ошибки репрезентативности

(представительности) возникают в результате того, что состав отобранной для обследования части единиц совокупности недостаточно полно отображает состав всей изучаемой совокупности (иначе говоря не все типы явления представлены в выборке).

РЕПРЕЗЕНТАТИВНОСТЬ И ОШИБКА ВЫБОРКИ

В то же время,

Важно понимать, что репрезентативность выборки и ошибка выборки – разные явления. Репрезентативность, в отличие от ошибки никак не зависит от размера выборки.

Ошибка выборки (доверительный интервал)

Отклонение результатов, полученных с помощью выборочного наблюдения от истинных данных генеральной совокупности.

Статистическая ошибка

Зависит от размера выборки. Чем больше размер выборки, тем она ниже.

Систематическая ошибка

зависит от различных факторов, оказывающих постоянное воздействие на исследование и смещающих результаты исследования в определенную

KEEP
CALM
AND
INCREASE
SAMPLE SIZE

Систематические ошибки

= неконтролируемые перекосы в распределении выборочных наблюдений («потеря» объекта). При возрастании объема выборки не уменьшаются

1936 г. выборы президента (Рузвельт – Лэндон):

«Литерэри Дайджест»: 2 376 523 человека
- Прогноз: Лендон – 57% , Рузвельт – 43%.
- Результат: Рузвельт – 62,5%, Лэндон – 37,5%

George Gallup
(1901 – 1984)

Типичные систематические ошибки

Давление доступных объектов

Иллюзия постоянства *(пренебрежение группой неопределившихся респондентов)*

Недостаточный учет аномальных и труднодоступных единиц исследования *(больные, личный состав ВС…)*

Недостаточный учет отсутствующих в месте сбора данных *(нет дома)*

Отказы от ответа *(нет мнения, негативная установка, внешние обстоятельства)*

Ошибка выборки

- это погрешность, с которой выборочные данные могут быть перенесены на всю аудиторию. Обычно ошибка выборки не превышает 5% на 95% доверительном интервале.

.

Средняя ошибка выборки

Средняя ошибка выборки (μ) – это среднее (по выборкам) отклонение выборочной оценки от истинного значения генеральной характеристики.

В каждой конкретной выборке фактическая ошибка выборки может быть меньше средней ошибки, равна ей или больше ее. Причем каждое из этих расхождений имеет различную вероятность.

Предельная ошибка выборки

Предельная ошибка выборки (Δ) – это максимально возможная при данной вероятности ошибка выборки.

То есть мы с заданной вероятностью (Рдов) гарантируем, что оценка ,полученная по нашей конкретной выборке, будет отличаться от значения генеральной характеристики не больше, чем на величину предельной ошибки Δ.

Доверительная вероятность

Вероятность, с которой мы гарантируем, что ошибка нашей выборки не превысит предельную ошибку, называется ***доверительной вероятностью*** - Рдов.
Предельная ошибка рассчитывается по формуле:

$$\Delta = t \cdot \mu,$$

где t- коэффициент доверия, значение которого определяется доверительной вероятностью (Рдов). Чем больше Рдов, тем больше t.

Понятие об ά- и β-ошибке

Результаты тестирования нулевой гипотезы

		Нулевая гипотеза истинна	
		да	нет
Результат проверки истинности нулевой гипотезы	да	Истиннополо жительные	β-ошибка
	нет	ά-ошибка	Истинноот рицательные

ά- и β-ошибки в выборочном исследовании

Вероятность ошибки – *P value*, величина p.

Какую величину p мы можем допустить?

- $p_{ά}$ <0.05
- $p_{β}$ <0.2 (0.1)

Что такое $p_{ά}$ <0.05 $p_{β}$ <0.2

- 0.05 0.2
- 5% 20%
- 1/20 1/5

Величина $p_{ά}$ для выборки

Величина p <0.05 говорит о том, что я готов найти различия по какому-нибудь параметру менее чем в 5% выборок одинаковой численности, формируя их одну за другой случайным образом из генеральной совокупности, или что меньше, чем одна выборка из двадцати не будет адекватно представлять по этому параметру генеральную совокупность

или

более чем 95% (более чем 19 из 20) всех таких выборок будут адекватно представлять генеральную совокупность.

Закон больших чисел

– ***методологическая основа выборочного метода.***
Теоретической основой выборочного метода является закон больших чисел:
С увеличением объема выборки вероятность появления больших ошибок и пределы максимально возможной ошибки уменьшаются (т.е. чем больше обследуется единиц, тем меньше будет величина расхождений выборочных и генеральных характеристик).

Расчеты репрезентативной выборки с допущением 5%-ной ошибки В.Паниотто

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

Вопросы, предваряющие процедуру выборки

Как сделать репрезентативную выборку?

или

Как избежать ошибки отбора *selection bias*

и

Насколько я готов ошибиться исходя из того, что обследована будет не генеральная совокупность а выборка из нее?

Объем выборки – общее число единиц наблюдения, включенных в выборочную совокупность

Чем больше объем выборки – тем выше точность ее результатов

Чем более однородна генеральная совокупность – тем меньший объем выборки можно использовать

Однородная совокупность – в которой контролируемый признак распределен равномерно (не образует пустот и сгущений)

Объем выборки зависит от:

- целей и задач, гипотез и методов исследования;
- объема ГС;
- степени однородности ГС;
- требующейся точности полученной информации
- характера предполагаемых распределений ответа…;

Стратегии расчета объема выборки

предварительная – объем выборки определяется до проведения основного исследования;

последовательная – объем выборки не рассчитывается заранее, а ставится в зависимость от конечных результатов исследования (например, в телефонном опросе постоянно проводится расчет распределений ответов на все вопросы и требуемого объема выборки);

комбинированная – предварительно рассчитывают верхние допустимые значения для последовательной стратегии.

• Случайные

Типы выборки

- Простая случайная
- Стратифицированная
- Кластерная (многоступенчатая, территориальная)

Случайные стратифицированные многоступенчатые кластерные

Методы отбора

- Простой случайный
- С вероятностью, пропорциональной размеру (ВПР)
- Систематический
- Систематический ВПР
- Контролируемый
- По карточкам Киша

Методы Отбора смешанные:
- по маршруту
- по дню рождения

• Неслучайные

Типы выборки

- Добровольцев (стихийная)
- Квотная
- Целевая (экспертная)
- Типичных представителей
- Доступная

Методы отбора

- Стихийный
- По квотам
- Направленный
- В местах скопления
- «Снежный ком»

Методы формирования выборки

Типы выборок

вероятностные

- Случайная выборка (простой случайный отбор)
- Механическая (систематическая) выборка
- Стратифицированная (районированная)
- Серийная (гнездовая или кластерная) выборка

невероятностные

- Квотная выборка
- Метод снежного кома
- Стихийная выборка
- Выборка типичных случаев

Случайный и неслучайный отбор

- «Уличный опрос».
  - Девушка-интервьюер стоит на перекрестке пр. Ленина и ул.Комсомольской и и «пристает» к прохожим с вопросами.
  - Она проводит интервью со всеми, кто соглашается с ней общаться.
- Какой тип выборки в данном случае используется?

- Это не «случайный», это «стихийный» отбор.
  - Случайный отбор – лучший (хотя и не всегда возможный) способ формирования выборки.
  - Стихийный отбор (отбор «кого попало») наихудший способ отбора.

Вероятностный отбор

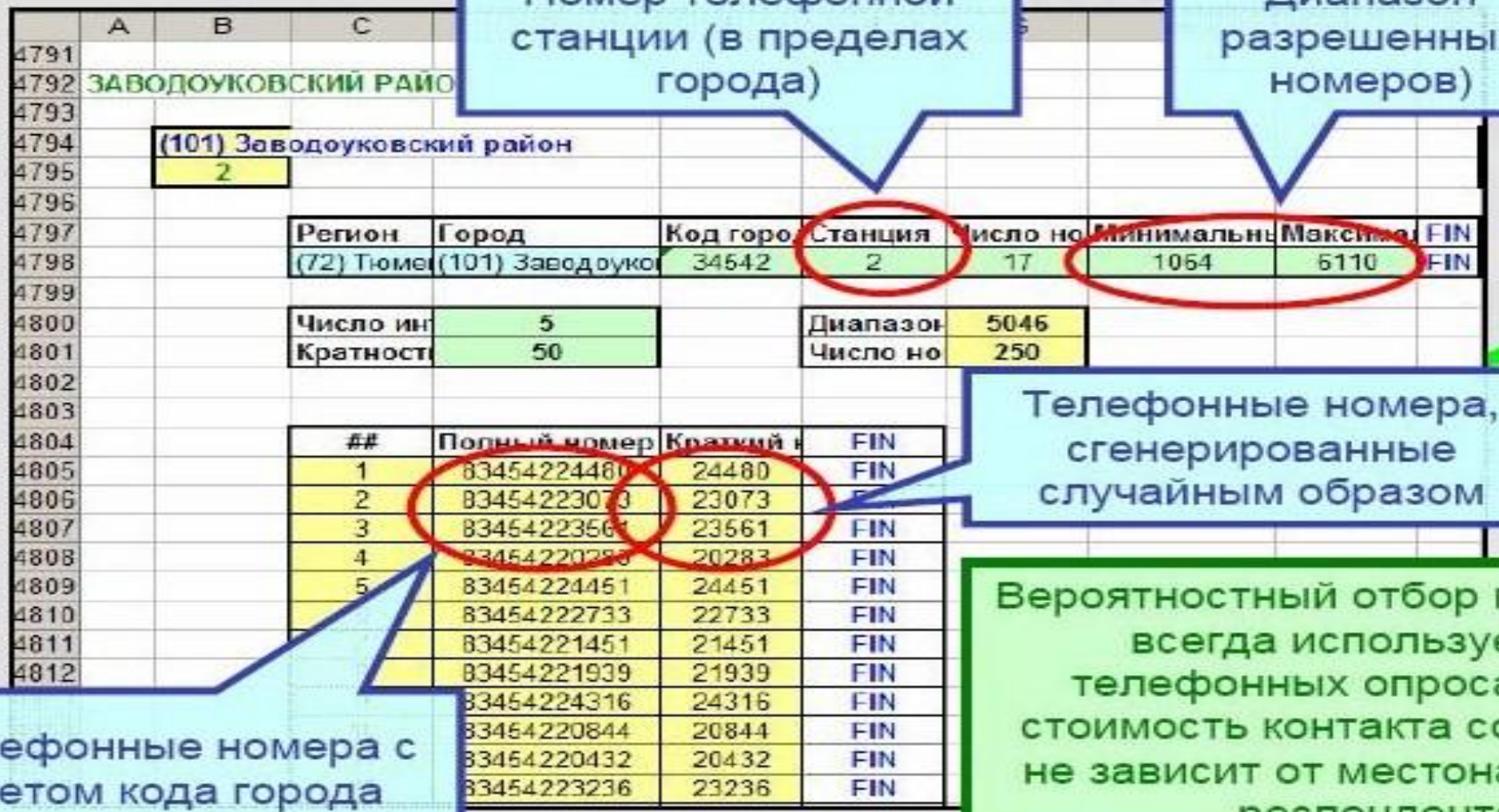
- Вероятностный отбор - это способ формирования выборки, при котором каждая единица генеральной совокупности имеет равную вероятность быть включенной в выборку.
  - При вероятностном отборе имеется возможность количественной оценки статистической погрешности выборки.
  - Является единственным приемлемым методом формирования выборки в тех случаях, когда отсутствуют или недостаточны предварительные данные о распределении объектов, входящих в генеральную совокупность, по интересующим исследователя характеристикам.

Вероятностные выборки строятся с использованием таблиц или генераторов случайных чисел.

Фрагмент таблицы равномерно распределенных случайных чисел

10	09	73	25						
37	54	20	48						
08	42	26	89						
99	01	90	25						
12	80	79	99	70	80	15	73	61	47
66	06	57	47	17	34	07	27	68	50
31	06	01	08	05	45	57	18	24	06
85	26	97	76	02	02	05	16	56	92
63	57	33	21	35	05	32	54	70	48
73	79	64	57	53	03	52	96	47	78

Вероятностный отбор



	A	B	C
4791			
4792	ЗАВОДОУКОВСКИЙ РАЙО		
4793			
4794		(101) Заводоуковский район	
4795		2	
4796			

Регион	Город	Код горо	Станция	Число но	Минимальны	Максима	FIN
(72) Тюме	(101) Заводоуко	34542	2	17	1064	6110	FIN

Число ин	5
Кратност	50

Диапазон	5046
Число но	250

##	Полный номер	Краткий н	FIN
1	83454224480	24480	FIN
2	83454223073	23073	
3	83454223561	23561	FIN
4	83454220283	20283	FIN
5	83454224451	24451	FIN
	83454222733	22733	FIN
	83454221451	21451	FIN
	83454221939	21939	FIN
	83454224316	24316	FIN
	83454220844	20844	FIN
	83454220432	20432	FIN
	83454223236	23236	FIN

Направленный отбор

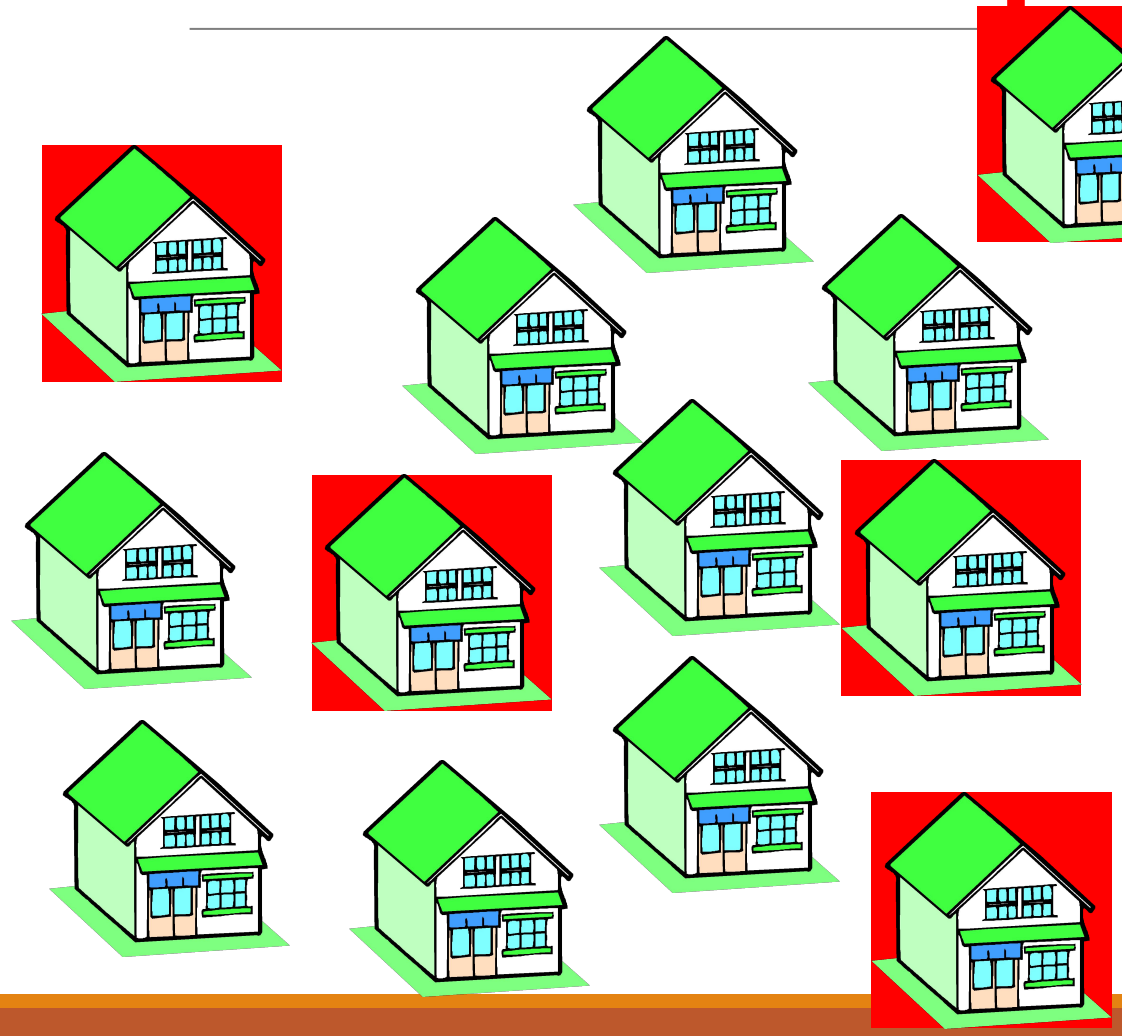
- «Направленным отбором» называют такой способ отбора, при котором при формировании выборки не соблюдаются условия случайности.
  - Случайный отбор возможен лишь в том случае, если имеется «основа выборки» - полный список единиц отбора, входящих в генеральную совокупность. В противном случае использование направленного отбора неизбежно.
  - Часто в исследованиях используется «функциональная выборка» («целевая выборка»), которая формируется из объектов, удовлетворяющих некоторым критериям («молодежь», «потребители определенной ТМ»). Такой вид отбора также является направленным.
- Оценка погрешности при направленном отборе.
  - Для направленных выборок принципиально невозможно оценить случайную погрешность оценки. На практике для оценки погрешности используются те же формулы, что и для одноступенчатой вероятностной выборки.

Простая случайная выборка

Жребий
Кубики
Генератор случайных чисел

Такая выборка предполагает однородность генеральной совокупности, одинаковую вероятность доступности всех элементов, наличие полного списка всех элементов. При отборе элементов, как правило, используется таблица случайных чисел.

Кластерная выборка

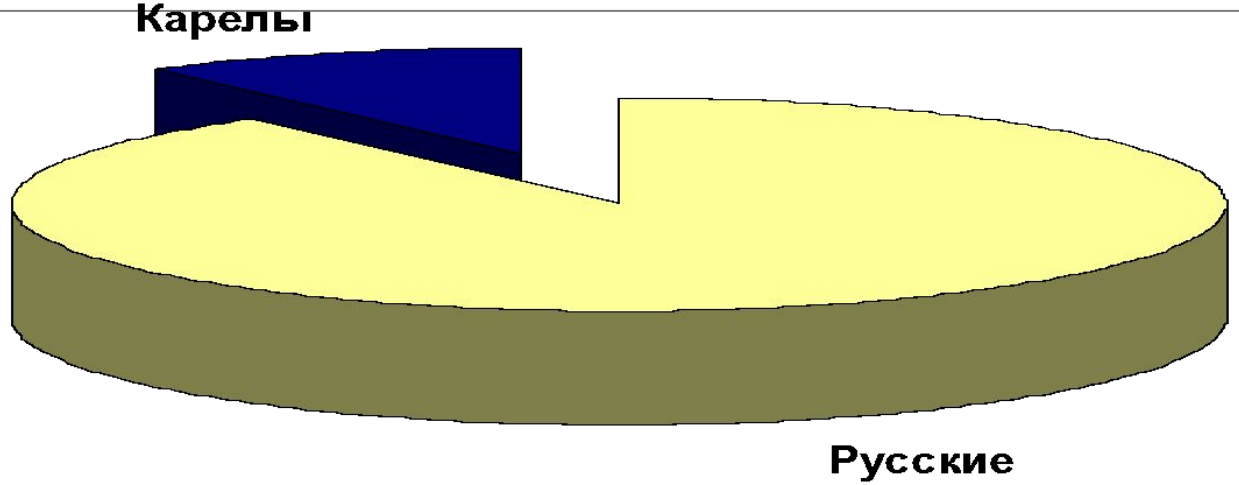


Классы школы

Цеха завода

Дома микрорайона

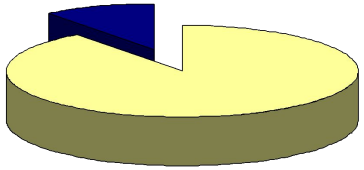
Стратифицированная выборка



Выборка соответствует генеральной совокупности по структуре

Применяется в случае неоднородности генеральной совокупности. Генеральная совокупность разбивается на группы (страты). В каждой страте отбор осуществляется случайным

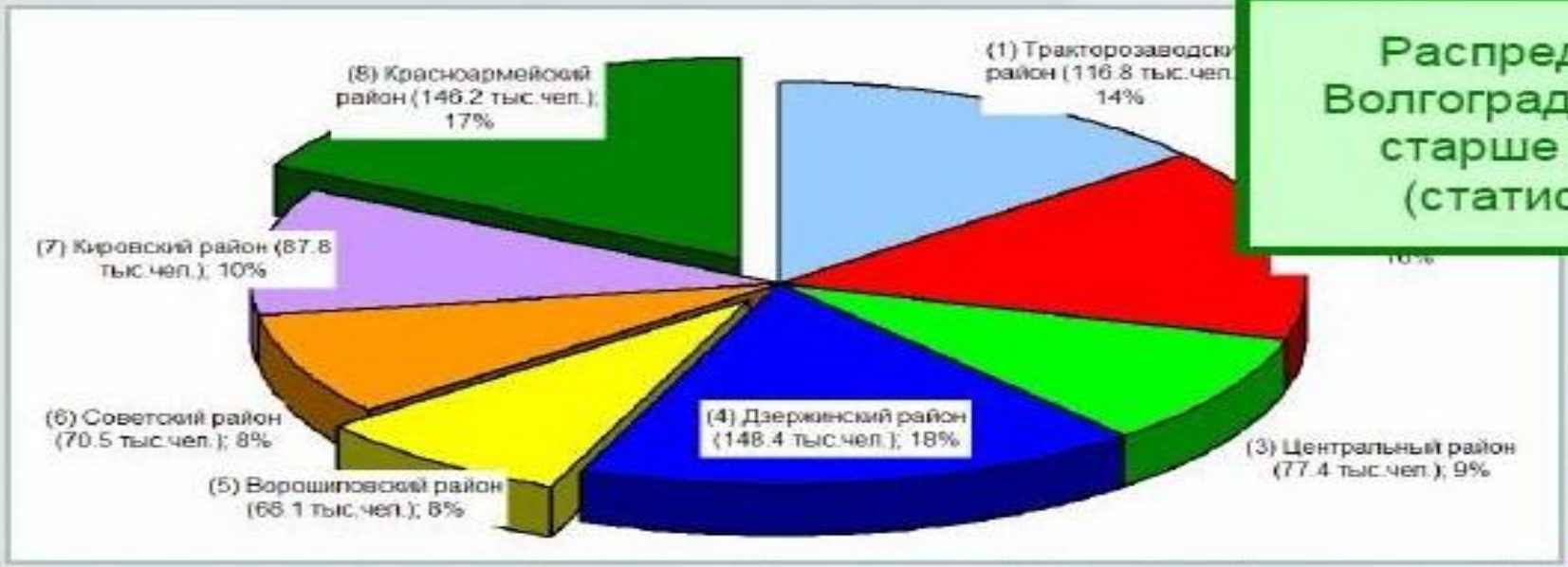
или механическим образом.



Стратифицированный отбор

- ☑ Процедура районированного (стратифицированного, расслоенного) отбора.
  - ☒ До извлечения выборки производится разделение генеральной совокупности на «страты», различающиеся по контролируемым (важным с точки зрения цели исследования) признакам.
  - ☒ Выборка (случайная или направленная) извлекается независимо из отдельных страт.
  - ☒ Объем выборки из каждой страты определяется исследователем исходя из специфики исследования (как правило, из соображений минимизации статистической погрешности выборки).
  - ☒ На этапе обработки результатов исследования проводится перевзвешивание полученных данных (различным стратам присваиваются различные «весовые коэффициенты»).
- ☑ Стратифицированный отбор – единственный вид отбора, который при определенных условиях может обеспечить ошибку выборки ниже, чем при вероятностном отборе.

Пример стратифицированного отбора



Распределение населения Волгограда в возрасте 16 лет и старше по районам города (статистические данные)

Район	Всего в возрасте 16 лет и старше, тыс. человек	Доля в общей совокупности, %	Объем выборки при пропорциональном отборе	Максимальная статистическая погрешность	Объем выборки при стратифицированном отборе	Максимальная статистическая погрешность выборки
(1) Тракторозаводский район	116.8	14%	220	6.6%	200	6.9%
(2) Краснооктябрьский район	133.4	16%	252	6.2%	200	6.9%
(3) Центральный район	77.4	9%	146	8.1%	200	6.9%
(4) Дзержинский район	148.4	17%	280	5.9%	200	6.9%
(5) Ворошиловский район	68.1	8%	128	8.7%	200	6.9%
(6) Советский район	70.5	8%	133	8.5%	200	6.9%
(7) Кировский район	87.8	10%	166	7.6%	200	6.9%
(8) Красноармейский район	146.2	17%	276	5.9%	200	6.9%
Общий итог	**848.6**	**100%**	**1600**	**2.5%**	**1600**	**2.5%**

Механическая (систематическая) выборка

Разновидность случайной выборки, упорядоченная по какому-либо признаку (алфавитный порядок, номер телефона, дата рождения и т.д.).

Первый элемент отбирается случайно, затем, с шагом ‘n’ отбирается каждый ‘k’-ый элемент.

Размер генеральной совокупности, при этом – **N=n*k**

Систематический отбор

- ☑ Систематический отбор - способ формирования выборочной совокупности, при котором выбор первого объекта репрезентации (как правило, случайный) полностью определяет набор объектов репрезентации, включаемых в выборку.
  - ☒ Отбор единиц осуществляется через один и тот же интервал (шаг) в исходном списке. Номер первого объекта в списке, включаемого в выборку, определяется случайным образом или принимается равным половине шага.

МАРШРУТНОЕ ЗАДАНИЕ	Организатор опроса выдает Вам маршрутное задание, в котором указаны адреса отобранных к опросу домов. Выполнять маршрутное задание можно в любом направлении. Не так важно, в каком порядке Вы проходите дома, важно, чтобы дом был указан в маршрутном задании.
ОТБОР КВАРТИРЫ	**В ПЕРВОМ ВЫБРАННОМ** многоквартирном доме или общежитии интервьюер обращается в квартиру №5.
	ЕСЛИ В КВАРТИРЕ ИНТЕРВЬЮ БЫЛО ВЗЯТО, то следующая квартира, куда надо обращаться, вычисляется согласно описываемому ниже шагу. Так, в одно, двух и трехэтажных домах следующая квартира вычисляется прибавлением цифры 3 (№5, №8, №11…), в четырех и пятиэтажных домах шаг равен 5 (№5, №10, №15…) и в домах свыше 5 этажей шаг равен 7 (т.е. №5, №12, №19…). Когда дом закончится, следующий дом надо рассматривать как продолжение предыдущего.

Требование строгой случайности нарушается. Выбор первого объекта и шага отбора полностью определяет выборку

Квотная выборка

Изначально выделяется некоторое количество групп объектов

(например, мужчины в возрасте 20-30 лет, 31-45 лет и 46-60 лет; лица с доходом до 13 тысяч , с доходом от 13 до 30 тысяч и с доходом свыше 30 тысяч)

Для каждой группы задается количество объектов, которые должны быть обследованы. Количество объектов, которые должны попасть в каждую из групп, задается, чаще всего, либо пропорционально заранее известной доле группы в генеральной совокупности, либо одинаковым для каждой группы. Внутри групп объекты отбираются произвольно. Квотные выборки используются в исследованиях достаточно часто.

Квотный отбор

- ☑ Формирование квотной выборки.
  - ☒ До начала опроса на основании статистических данных рассчитывается распределение объектов в выборке, различающихся по контролируемым признакам (пол, возраст, образование и т.д.).
  - ☒ Общая выборка делится на «квотные планы» (по 10-15 респондентов), которые и выдаются интервьюерам в качестве задания.
  - ☒ После «объединения» индивидуальных заданий, выполненных различными интервьюерами, получается выборка, репрезентативная по контролируемым признакам.
- ☑ Достоинства и недостатки.
  - ☒ Является, по сути, единственным методом отбора, который позволяет контролировать репрезентативность.
  - ☒ Правильность отбора зависит от правильности «внешних» статистических данных, которые используются для расчета выборки.

:100 000

=

Выборка - 1450 человек по всей РФ

Городские жители - 1059, сельские - 391чел.

Женщины - 769, мужчины - 681 чел.

В возрасте до 18 лет - 363 чел.

Пенсионеры - 334 чел.

Между 18 и пенсионным возрастом - 753 чел.

Стихийный отбор

- Отбор «кого попало» - это не «случайный», а «стихийный отбор» - худший способ формирования выборки.
  - Используется, как правило, в разведывательных исследованиях, а также в случаях, когда данных для того, чтобы сформировать квотную или вероятностную выборку, недостаточно.
- Виды стихийного отбора.
  - «Отбор первого встречного» - включение в выборку тех элементов, которые легче всего достижимы для интервьюера / исследователя. Используется в экспресс-исследованиях и пилотажных исследованиях.
  - «Снежный ком» - отбор «знакомых знакомых». Используется в тех случаях, когда объект исследования мал по объему и сложнодостижим, а также в качественных исследованиях с направленной выборкой.
  - «Прессовый отбор» - «дистанционный анкетный опрос», когда решение об участии в исследовании принимает сам респондент.

Стихийная выборка

Опрашиваются наиболее

доступные респонденты.

Типичные примеры стихийных выборок – опросы в газетах/журналах, анкеты, отданные респондентам на самозаполнение, большинство интернет-опросов. Размер и состав стихийных выборок заранее не известен, и определяется только одним параметром – активностью респондентов.

Метод снежного кома

У каждого респондента, начиная с

первого, просятся контакты его друзей,

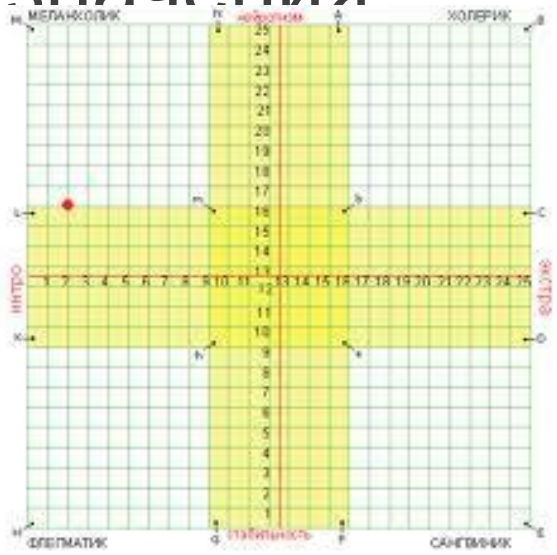
коллег, знакомых, которые подходили бы

под условия отбора и могли бы принять участие в исследовании. Таким образом, за исключением первого шага, выборка формируется с участием самих объектов исследования.

Метод часто применяется, когда необходимо найти и опросить труднодоступные группы респондентов (например, респондентов, имеющих высокий доход, респондентов, принадлежащих к одной профессиональной группе, респондентов, имеющих какие-либо схожие хобби/увлечения и т.д.)

Выборка типичных случаев

Отбираются единицы генеральной совокупности, обладающие средним (типичным) значением признака. При этом возникает проблема выбора признака и определения его типичного значения



Многоступенчатая выборка

- Многоступенчатая выборка – это выборка, которая состоит из нескольких «подвыборок», вложенных друг в друга по принципу «матрешки».
  - Как правило, используется для сложных, географически распределенных объектов.
  - На каждом шаге отбора используется собственная «единица отбора».
- Пример построения общероссийской многоступенчатой выборки для face-to-face опроса:
  - Отбор регионов России, в которых будет проходить опрос.
  - В каждом регионе – отбор конкретных населенных пунктов.
  - В каждом населенном пункте – отбор улиц (начальных точек маршрутов).
  - На каждом маршруте – отбор домохозяйств.
  - В каждом домохозяйстве – отбор респондентов к опросу.

Многоэтапная выборка

Первый этап – зонная

Второй этап – кластерная

Третий этап – стратифицированная по полу

Обоснование системы выборки единиц наблюдения

Метод основного массива – опрос 60-70% генеральной совокупности.

Если опрос проводится в рамках коллектива, представительным (репрезентативным) будет сплошное анкетирование.

В масштабах организации следует опросить 25%.

При массовых опросах (генеральная совокупность в пределах 5000 чел.) достаточной является 10%-я выборка.

Стратегия предварительного расчета (случайная выборка)

Необходимо знать:

Желаемую точность оценивания (5%)

Величину риска получаемого ответа (0,95)

Степень изменчивости ответа (50% на 50% при дихотомическом вопросе, т.е. макс. разброс берем, если нет информации о распределении ответов на вопрос)

V гс	500	1000	2000	3000	4000	5000	10 т.	100 т.	более
V вс	222	286	333	350	360	370	385	398	400

Зависимость объема выборки от распределения дихотомического ответа

Распределение ответов, %	50	40	30	20	10
	50	60	70	80	90
Vвыборки	384	369	323	246	139

Расчет квотной выборки

1 вариант – как для случайной (при больших объемах) (см. таблицу выше)

2 вариант – на основе теории малых выборок (если не предполагается анализ по группам): количество градаций вопросов х 25 (минимальный статистически значимый размер группы):

Например:

- *пол (2 градации);*
- *возраст (2 градации: до 30, после 30);*
- *удовлетворенность трудом (5-балльная шкала)*

V выборки = 2х2х5х25 = 500 человек

Стратегия последовательного расчета выборки

Объем выборки не рассчитывается заранее – а ставится в зависимости от результатов исследования: в зависимости от разброса оценок □ необходимый V выборки

Зависимость объема выборки от коэффициента вариации (= ср.квадр.отклонение от ср.арифм. в %)

Кв	10	20	30	40	50	60	70	80	90	100	110	120
Vв	15	61	138	246	384	553	753	984	1245	1537	1860	2213

* Удобна – если можно производить необходимые расчеты в ходе самого опроса

Комбинированная стратегия

Рассчитываем выборку по предварительной стратегии получаем верхние допустимые значения для последовательной стратегии (т.е. ту величину объема, при которой опрос прекращается)

Ошибка выборки

= отклонение средних характеристик ВС от средних характеристик ГС

На практике ошибка выборки определяется путем сравнения известных характеристик ГС с характеристиками выборки

Например (ВЦИОМ):

Группы населения	Данные гос.стат.	Выборочные данные
Пол		
1. Мужской	45,1	45,6
2. Женский	54,9	54,4
Возраст		
3.До 29 лет	26,9	25,4
4. 30-54 года	44,5	45,9
5. 55 лет и старше	28,6	28,6
Образование		
6. В/о, незакон.в/о	13,7	14,3
7. Ср. и ср.спец.	*47,6*	*43,0*
8. Неполное среднее	38,7	42,7
Тип поселения: 9-11	9,3/26,4/38,5/25,8	9,3/28,6/36,6/25,5

Ошибки выборки:

Случайные: вероятность того, что выборочная средняя выйдет (или не выйдет) за пределы заданного интервала вероятностные, уменьшаются при возрастании объема выборочной совокупности

Эксперименты Дж.Гэллапа (1935 г. – отношение американцев к запрету спиртного)

Выборки	Одобряют	Не одобряют	Нет мнения
442 чел.	**31**	**62**	**7**
884 чел.	**29**	**63**	**8**
1327 чел.	**30**	**63**	**7**
2585 чел.	31	61	8
5255 чел.	33	59	8
8253 чел.	32	60	8
12494 чел.	32	61	7

Порядок определения выборочной совокупности

Проектирование выборки включает:

Определение границ и состава генеральной совокупности (сплошное или выборочное исследование).
Определение типа выборки (случайная, квотная, гнездовая, простая или многоступенчатая).
Определение объема выборки.
Расчет предельной ошибки выборки (репрезентативности).
Определение технологии отбора объектов.
Репрезентативность (представительность) – свойство выборки отражать характеристики генеральной совокупности.

Формулы для определения объема выборки:

$$n = 2\sqrt{N}$$

$$n = (0{,}25\ t^2\ N) / (A^2N + 0{,}25t^2),$$

где n – объем выборочной совокупности,
N – объем генеральной совокупности,
t^2 – число, определяемое по специальным таблицам с учетом предельной ошибки выборки и доверительной вероятности,
A – предельная ошибка выборки.

Выборка считается репрезентативной, если она отражает характеристики генеральной совокупности не ниже, чем на 95%.

Контроль и ремонт выборки

Ремонт В. = процесс устранения погрешностей, т.е. расхождения ВС и ГС различными способами:

Коррекция ВС *(напр., формирование повторной выборки или выборки большего размера, или следующих номер абонента…)*

Коррекция распределений демографических характеристик респондентов *(взвешивание исходных данных…)*

Коррекция резко выделяющихся ответов респондентов (брак?)

Коррекция пропущенных ответов

Основные выводы

- Фактически, единственным «методически строгим» способом формирования выборки является случайный отбор.
- На практике в большинстве случаев случайный отбор реализовать невозможно.
  - Вероятностный отбор возможет лишь в случаях, когда «единицы отбора» не могут «отказаться» от участия в исследовании.
  - Даже в случае, если выборка сформирована случайным образом, на этапе ее реализации возникают неизбежные смещения.
- Как правило, в ходе опросов используются комбинированные многоступенчатые выборки, сочетающие в себе черты случайного и направленного отбора.
  - Случайный отбор используется при отборе регионов, точек опроса, маршрутов, телефонных номеров и т.д.
  - Квотный отбор используется при отборе конкретных респондентов для обеспечения репрезентативности по контролируемым признакам.
  - Стихийный отбор используется для изучения малых целевых групп, в разведывательных и качественных исследованиях.

СПАСИБО ЗА ВНИМАНИЕ!