

Базовые понятия выборочного исследования

Шабденова Айжан Базархановна

Генеральная и выборочная совокупность

Генеральной совокупностью считают все население или ту его часть, которую социолог намерен изучить, совокупность людей, обладающих одним или несколькими свойствами, подлежащими изучению. Часто генеральная совокупность (еще называемая популяцией) настолько крупная, что опрос каждого представителя чрезвычайно обременителен и дорогостоящ. Это те, на кого направлен теоретический интерес социолога (в том смысле, что узнать о каждом представителе генеральной совокупности ученый может только косвенно — на основе информации о выборочной совокупности).

Выборочная совокупность — уменьшенная модель генеральной совокупности. Иначе говоря, это множество людей, которых социолог опрашивает. В выборку, или выборочную совокупность, входят только те, кого социолог намеревается непосредственно опросить. Представим, что предметом его исследования, т.е. темой, выступает экономическая активность пенсионеров. Все пенсионеры — будут составлять генеральную совокупность. По специальным формулам социолог рассчитал, что ему достаточно опросить 2,5 тыс. пенсионеров. Это и станет его выборочной совокупностью.

Основное правило ее составления гласит: **каждый элемент генеральной совокупности должен иметь одинаковые шансы попасть в выборку.**

Соотношение генеральной и выборочной совокупности



Кого именно относить к генеральной совокупности, определяют цели исследования, а кого включать в выборочную совокупность, решают математические методы.

Репрезентативной выборкой в социологии считается такая выборочная совокупность, основные характеристики которой полностью совпадают (представлены в той же пропорции или с той же частотой) с такими же характеристиками генеральной совокупности. Только для этого типа выборки результаты обследования части единиц (объектов) можно распространять на всю генеральную совокупность.

Преимущества выборочного метода

- Меньше стоимость
- Короче сроки
- Шире область применения
- Больше достоверность полученных данных

Основные проблемы выборочного исследования

- Цели обследования: чрезвычайно полезна четкая формулировка целей обследования
- Совокупность, из которой производится отбор: нужно, чтобы в ходе работы обследователь был в состоянии без особых колебаний определять, принадлежит ли сомнительный объект к совокупности или нет.
- Собираемые данные: все собираемые данные соответствуют целям обследования и никаких важных данных не пропущено.
- Желательная степень точности
- Способы сбора информации.
- Основа выборки.

Типы выборок по Г. Силласте



Методы невероятностной (целенаправленной) выборки (1)

Отбор в такой выборке осуществляется не по принципам случайности, а по субъективным критериям – доступности, типичности, равного представительства и т.д..

2.1. 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 30 тысяч рублей, с доходом от 30 до 60 тысяч рублей и с доходом свыше 60 тысяч рублей) 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.

2.2. Метод снежного кома

Выборка строится следующим образом. У каждого респондента, начиная с первого, просят 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.)

Методы невероятной (целенаправленной) выборки (2)

2.3 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.

2.4 Выборка типичных случаев

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

Методы вероятностной (случайной) выборки

Случайная (вероятностная) *выборка* — это выборка, для которой каждый элемент генеральной совокупности имеет определенную, заранее заданную вероятность быть отобранным. Это позволяет исследователю рассчитать, насколько правильно выборка отражает генеральную совокупность, из которой она выделена (спроектирована). Такую выборку иногда называют еще *случайной*.

Вероятностные методы включают:

1.1 Случайная выборка (простой случайный отбор)

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.

1.2 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 – $N=n*k$

1.3 Стратифицированная (районированная)

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

1.4 Серийная (гнездовая или кластерная) выборка

При серийной выборке единицами отбора выступают не сами объекты, а группы (кластеры или гнёзда). Группы отбираются случайным образом. Объекты внутри групп обследуются сплошняком.

Простой случайный отбор

Простой случайный отбор предполагает, что вероятность быть включенным в выборку известна и является одинаковой для всех единиц совокупности. Он реализуется двумя методами:

- ◆ отбор вслепую (другое название — метод лотереи или жребия),
- ◆ отбор не вслепую (происходит с помощью таблицы случайных чисел).

Вероятностную выборку целесообразно применять только при наличии соответствующих условий. *Первое условие* осуществления вероятностной выборки — наличие *полного списка* всех элементов генеральной совокупности (отсутствие или недоступность которого чаще всего и препятствует ее реализации) от 1 до N , где N — общее число всех элементов. *Второе условие* вероятностной выборки — *хорошая перемешанность* элементов генеральной совокупности.

Требования к вероятностному отбору:

- 1. Должна существовать возможность указать множество различных выборок S_1, S_2, \dots, S_v , которые могут быть получены при применении данного метода отбора к некоторой конкретной совокупности. Это значит, что мы можем точно указать, какие единицы отбора принадлежат к S_1 , к S_2 и т. д.
- 2. Для каждой из возможных выборок S_i , задана известная нам вероятность ее извлечения.
- 3. Мы извлекаем одну из выборок S_i , с помощью некоторого процесса, при котором вероятность извлечения каждой выборки принимает соответствующее значение.
- 4. Должен быть установлен метод вычисления оценки по выборке и для каждой конкретной выборки он должен приводить к единственному значению. Мы можем принять, например, в качестве оценки среднее значение результатов наблюдений отдельных единиц в выборке.

К методу отбора такого типа применяется термин вероятностный отбор. Это, конечно, не единственный способ, которым можно извлечь выборку.

На практике мы редко получаем вероятностную выборку, записывая S_i , как было указано ранее. Для больших совокупностей, где принятый порядок отбора дал бы миллиарды возможных выборок, это оказалось бы невыносимо трудоемкой работой. Обычно отбор производится путем указания вероятностей включения в выборку отдельных единиц и затем извлечения единиц по одной или по несколько сразу до тех пор, пока не будет образована выборка нужного объема и типа.

Нормальное распределение

При обследованиях выборки часто достаточно велики, так что получаемые по ним оценки имеют приблизительно нормальное распределение.

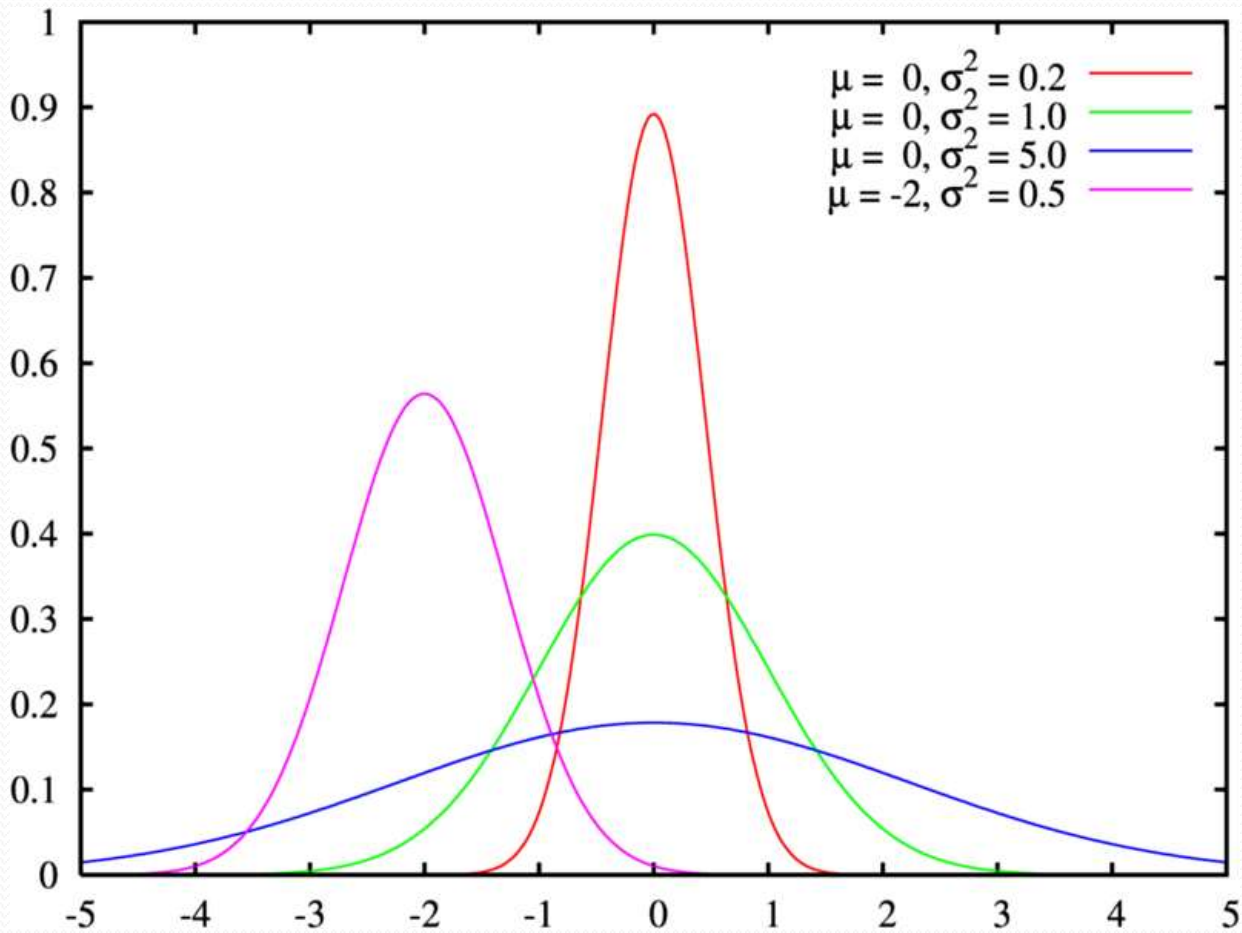
Что такое нормальное распределение?

Нормальное распределение — это распределение вероятностей, которое задается функцией плотности распределения:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

где параметр μ — **среднее значение** (математическое ожидание) случайной величины и указывает координату максимума кривой плотности распределения, а σ^2 — **дисперсия**.

Примеры нормальных распределений



Как оценить стандартную ошибку выборки?

Предположим, что мы получили выборку методом, обеспечивающим несмещенную оценку, и вычислили соответствующее значение выборочной оценки μ и ее среднее квадратичное отклонение σ (часто называемое иначе ее стандартной ошибкой или стандартным отклонением). Если упростить, то ошибка выборки показывает, насколько среднее значение в выборке может быть больше или меньше среднего значения в генеральной совокупности.

Насколько хороша наша оценка? Из свойств нормального распределения вытекает, что с вероятностями:

- 0,32 (или приблизительно в одном случае из трех) абсолютное значение ошибки превосходит σ ;
- 0,05 (или в одном случае из двадцати) абсолютное значение ошибки превосходит $1,96 \sigma$ (*принято условно говорить 2σ*);
- 0,01 (или в одном случае из ста) абсолютное значение ошибки превосходит $2,58 \sigma$.

Таким образом, 95% доверительный интервал обозначает, что среднее в выборочной совокупности μ отличается от среднего значения в генеральной совокупности более чем на σ в 95% случаев.

99% доверительный интервал обозначает, что только в 1% случаев среднее выборочной совокупности значительно отличается от среднего в генеральной совокупности.

Ошибка выборки: пример

Например, если при определении срока службы некоторых приборов на большом предприятии при обычной нагрузке вероятностная выборка показала, что среднее время работы прибора для этой выборки = 394 дням при среднем квадратичном отклонении (стандартной ошибке) = 4,6 дня, то среднее время работы приборов для всей их совокупности в 99 случаях из ста заключено между

$$394 - (2,58) \cdot (4,6) = 382 \text{ дням}$$

И

$$394 + (2,58) \cdot (4,6) = 406 \text{ дням.}$$

- Эти границы, 382 дня и 406 дней, называются нижней (lower) и верхней (upper) доверительными границами, или доверительным интервалом. Доверительные границы (интервал) могут быть выражены в абсолютном значении, но чаще используют процентное отношение к величине измеряемого признака.

Сколько элементов нужно отобразить, или поправка на конечную совокупность

Известно, что для выборки объема n из бесконечной совокупности дисперсия среднего равна σ^2/n . Если совокупность конечна, то необходимо ввести множитель $(N-n)/N$ – поправку на конечность совокупности, где N – количество объектов в генеральной совокупности. По другому этот множитель можно записать как $(N/n - 1)$.

Если доля отбора n/N низка, то данный множитель практически равен единице, то есть влияние объема совокупности очень мало.

Например, если для двух совокупностей дисперсия одинакова, то выборка объемом в 500 единиц из генеральной совокупности в 20000 единиц обеспечивает ту же точность данных, что и выборка 500 единиц из 10000.

Определение объема выборки в зависимости от ошибки выборки

На практике необходимый объем выборки определяется тем, какую точность данных мы хотим получить. Чем выше точность данных (и, соответственно, ниже ошибка выборки), тем большую выборку необходимо использовать.

Размер выборки	Доверительный интервал, \pm %
100	9,8
500	4,38
1000	3,1
1500	2,53
2000	2,19
3000	1,79
5000	1,39

Применение ошибки выборки для сравнения результатов исследований.

Предположим, у нас есть результаты двух исследований, проведенных в различных странах, в которых респондентам задавался один и тот же вопрос: «Голосовали ли Вы на прошлых выборах».

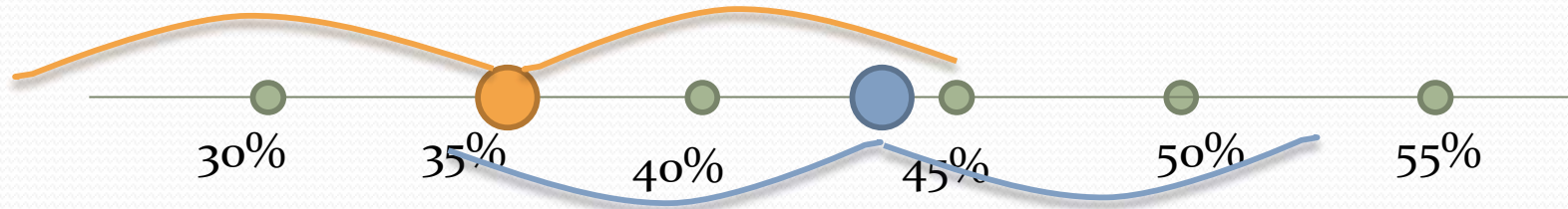
В обоих исследованиях было опрошено 100 человек.

Результаты первого исследования показали, что 35% респондентов принимали участие в выборах.

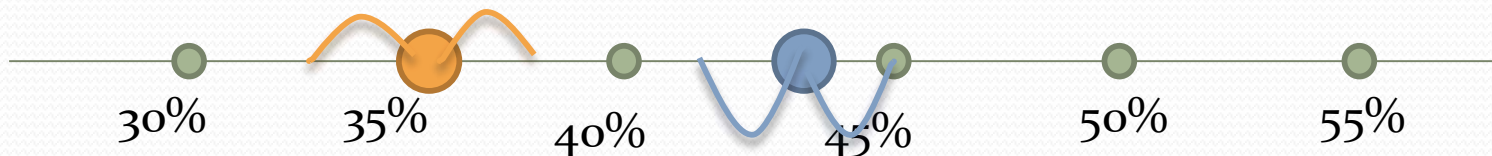
Второе исследование показало, что 43% респондентов принимало участие в выборах.

Различается ли уровень участия респондентов в выборах в этих двух странах на статистически значимом уровне?

Для того, чтобы ответить на данный вопрос, надо посмотреть на доверительный интервал. Для выборки объемом 100 человек он составляет 9,8%. Что это значит? Это значит, что в генеральной совокупности реальное значение может быть как на 9,8% больше, так и на 9,8% меньше, чем полученное нами.



Теперь предположим. Что объем выборки в обеих странах был 1500 человек. Соответственно, ошибка выборки составляет 2,53%.



Видно, что если в первом случае интервалы пересекаются, то есть мы не можем с уверенностью сказать, есть ли различия или нет, то во втором случае интервалы не пересекаются и различия в поведении есть.

Ошибки выборки: типология

По происхождению:

- 1) теоретические, возникающие до процесса отбора на стадии формирования концептуального представления об объекте исследования и выработки стратегии отбора;
- 2) 2) процедурные, связанные с построением выборочной модели;
- 3) 3) ошибки на этапе реализации вплоть до непосредственного контакта с единицей наблюдения.

Ошибки выборки: типология

По характеру воздействия на выборочную оценку:

- 1) случайная – имеет вероятностную природу, она органически присуща выборочному наблюдению, если отбор организован по строго случайному принципу. Величина случайной ошибки зависит от: плана построения выборки, объема выборочной совокупности, степени вариации признаков.
- 2) систематическая (смещение выборки) носит неслучайный характер и представляет собой некоторую постоянную или закономерно меняющуюся величину.

Ошибки смещения бывают обычно следствием:

- а) неверных статистических данных о параметрах контрольных признаков генеральной совокупности;
- б) слишком малого объема выборочной совокупности;
- в) неверного применения способа отбора единиц анализа (ошибки обработки данных и др.).



Спасибо за
внимание