Антонов Г. В. Выборочный метод в социологических исследованиях / Г. В. Антонов // Научный диалог. – 2013. – № 11 (23) : История. Социология. Философия. – С. 96–109.

УДК 303.5:311.21

Выборочный метод в социологических исследованиях

Г. В. Антонов

Представлена авторская точка зрения на процедуру применения выборочного метода в прикладном социологическом исследовании. Автор опирался на многолетний опыт проведения социологических опросов и результаты контент-анализа документов, осуществленного в рамках научных исследований и по заказам государственных организаций и частных лиц. Рассматриваются основные методы и приёмы математической статистики и теории вероятностей, применяемые при переносе результатов исследования с выборочной совокупности на генеральную. При этом автор пытался по возможности избежать излишнего теоретизирования, подробного вывода формул и т. п. Ставилась задача представить читателю готовые образцы, продемонстрировав порядок и технику применения самых необходимых формальных операций. Изложение материала сопровождается примерами и подробными комментариями, доступными для понимания материала даже неподготовленным читателем, например, начинающим социологом (студентом, аспирантом и т. п.), а также специалистом, впервые приступающим к проведению подобных исследований. Особенное внимание уделяется вопросам практического применения рассмотренной методики формирования выборочной совокупности. Материал статьи может быть использован в учебном процессе в рамках преподавания дисциплин «Методика и техника конкретного социологического исследования» и «Социальная статистика».

Ключевые слова: выборочный метод; генеральная совокупность; выборка; доверительный интервал; уровень доверия; репрезентативность; предельная ошибка; случайный отбор.

Сразу следует отметить, что ***выборочный метод*** ни в коем случае не является «изобретением» именно социологов, и даже нельзя сказать, что при проведении прикладных социологических исследований он используется чаще, чем в каких-то других случаях. Данный приём применяется настолько во многих отраслях научной и практической деятельности человека, что простое перечисление всех возможных примеров займёт не одну страницу. Нельзя также утверждать, что применение выборочного метода в прикладной социологии принципиально отличается от его применения в других научных или практических сферах. По одним и тем же правилам и формулам «отбираются» как люди для проведения социологических опросов, так и изделия на заводе для последующего контроля качества, платёжные документы в фирме для изучения их налоговыми органами, домохозяйства в демографических обследованиях, экспериментальные группы (реальные или гипотетические) в естественных науках и т. д. Суть в любом случае одна: если нет возможности непосредственно обследовать все интересующие объекты по причине слишком большого их количества, то обследуется лишь часть таких объектов, причём, как правило, очень небольшая. Смысл данной процедуры состоит в том, что сделанные выводы переносятся и на ту часть объектов, которая обследована не была.

Точно так же обстоит дело и в подавляющем большинстве социологических исследований, когда непосредственному изучению может быть подвергнута далеко не вся совокупность, условно составляющая объект исследования (так называемая *генеральная совокупность*), а лишь некоторая её часть (то есть *выборочная совокупность*, или *выборка*). Проблема в таком случае состоит в обеспечении *репрезентативности* изучаемой выборочной совокупности, то есть возможности переноса полученных на основе её изучения результатов на всю интересующую исследователя генеральную совокупность. При этом в качестве подлежащих переносу на генеральную совокупность суждений о выборке обычно выступают утверждения о численном значении каких-либо параметров (например, утверждение о среднем арифметическом значении некоторого признака). Подобный перенос осуществляется с помощью методов математической статистики, позволяющих указать точность получаемых выводов, которая обычно выражается посредством *доверительных интервалов*. Доверительный интервал представляет собой диапазон значений вокруг оценки какого-либо параметра по выборочной совокупности, в котором с заданным *уровнем доверия* находится реальное значение этого параметра по генеральной совокупности. Уровень доверия это – оценочный показатель достоверности результата, полученного по выборочной совокупности, в сравнении его с аналогичным результатом, который мог бы быть получен при непосредственном изучении всей генеральной совокупности.

Для примера обратимся к такой всем известной вещи, как прогноз погоды. Скажем, кто-то утверждает, что завтра в полдень температура воздуха будет находиться в пределах от +20 до +21°C. Вопрос: какова вероятность того, что данный прогноз сбудется? Очевидно – низкая, причём вне зависимости от времени года, поскольку даже если на улице весна или лето, то температура воздуха вполне может быть и +19, и +15, и +22, и +25°C и т. д. А теперь представим, что кто-то делает другой прогноз: завтра в полдень температура воздуха буде находиться в пределах от –50 до +50°C. Ясно, что такой прогноз практически в любом регионе нашей страны в любое время года сбудется почти со 100 %-й вероятностью. Температура –51 или +51°C, конечно, возможна, но такое бывает крайне редко. Итак, в первом случае прогноз делается очень точный, конкретный, поскольку предсказываемый температурный интервал составляет всего лишь 1°C, но при этом вероятность того, что он сбудется, – низкая. Во втором случае, напротив, прогноз даётся крайне расплывчатый, неопределённый (предсказанный размах температур составляет 100°C), но зато вероятность его истинности почти 100 %. Именно приведённый в данном примере размах температур называется в статистике *доверительным интервалом*, тогда как вероятность попадания в него – *уровнем доверия*. Понятно, что чем шире доверительный интервал (то есть чем неопределённее прогноз), тем уровень доверия (вероятность того, что он сбудется) выше. И наоборот: сужая доверительный интервал, мы снижаем уровень доверия. Аналогичным образом следует интерпретировать эти параметры и при использовании выборочного метода в социологических исследованиях. Мы как бы предсказываем, что если среднее значение какого-либо признака по выборочной совокупности приняло определённое значение, то значение этого же признака по генеральной совокупности будет с определённой вероятностью в определённом интервале находиться вокруг его среднего значения по выборке.

Допустим, объектом исследования является население крупного города в трудоспособном возрасте, например, 500 тыс. человек. Выборочная же совокупность даже в очень крупных исследованиях редко превышает 2,5 тыс. человек, а в большинстве случаев бывает гораздо меньше. Но даже 2,5 тыс. человек составляют всего 0,5 % от рассматриваемой в данном примере генеральной совокупности. Закономерно возникает вопрос: на основании чего осуществляется перенос результатов с выборки, составляющей в лучшем случае половину процента генеральной совокупности, на всю генеральную совокупность?

Рассмотрим ещё один пример. Что такое 0,5 %? Представим себе, что перед нами находится ёмкость, внутри которой содержится 10 тыс. шариков разного цвета. Мы наугад, случайным образом, достаем из неё 50 шариков (то есть 0,5 % от 10000), и оказывается, что все они чёрные. Можем ли мы сделать вывод о том, что и оставшиеся 9950 шариков тоже чёрные? На первый взгляд – нет, однако даже если шарики в ёмкости всего двух цветов (пусть чёрные и белые) и распределены они по признаку «цвет» равномерно (поровну), то *вероятность* вытащить наугад подряд 50 чёрных составляет приблизительно $(\frac{1}{2})^{50} \approx 8{,}9 \times 10^{-16}$. Для тех, кто не знаком с основами теории вероятностей, поясняем, что *вероятность* – это отношение числа интересующих нас событий к общему числу возможных событий. Поэтому вероятность того, что 1-й вытащенный шарик будет чёрным, составляет $\frac{1}{2}$ или $\frac{1}{2}$, вероятность того, что 2-й будет также чёрным, – $\frac{4999}{9999}$, то есть *приблизительно* $\frac{1}{2}$ и т. д., наконец, вероятность того, что 50-й шарик, как и все предыдущие, будет тоже чёрным, составляет $\frac{4951}{9951} \approx 0{,}4975$, то есть опять-таки примерно $\frac{1}{2}$. Следовательно, для расчёта вероятности вытащить наугад 50 подряд чёрных шариков нам необходимо $\frac{1}{2}$ перемножить 50 раз или возвести в 50-ю степень. Результат $8{,}9 \times 10^{-16}$ означает, что такое событие практически невозможно. Оно может состояться, но только при условии, что в ёмкости если не все шарики, то подавляющее большинство – чёрные.

Что означает «подавляющее» большинство? Допустим, в ёмкости 90 % шариков чёрные, а 10 % – белые. Вероятность вытащить наугад 50 чёрных составит приблизительно $0{,}9^{50}\approx0{,}005$ – шансы уже значительно выше, но тоже невелики. Наконец, если распределение по цвету составляет 95 % против 5 %, то вероятность рассматриваемого события будет уже $0{,}95^{50}\approx0{,}077$, то есть где-то в 8 случаях из 100 мы можем вытащить такую комбинацию. Итак, если подобное событие произошло, то мы вправе утверждать, что в ёмкости осталось не менее 95 % чёрных шариков (то есть почти все!). Если же цветов больше, чем два, то вероятность наступления указанного события при более или менее равномерном распределении по цвету будет на несколько порядков ниже, чем $8{,}9\times10^{-16}$. Следовательно, перенос результатов с выборочной совокупности на генеральную очень даже возможен, нужно только соблюдать определённые правила при отборе объектов в выборку. Об этих правилах будет сказано далее, а кроме этого, необходимо найти основание для оценки параметров точности переносимых результатов.

Таким основанием зачастую как раз и выступает доверительный интервал с заданным уровнем доверия. Например, среднее арифметическое значение возраста респондентов в выборке составляет 40 лет, при этом нижняя и верхняя границы доверительного интервала с уровнем доверия 0,95 ($p=0{,}95$) составляют соответственно 38 и 42 года. Это значит, что среднее арифметическое значение возраста по генеральной совокупности с вероятностью 95 % будет находиться в диапазоне 38–42 года (то есть в пределах ±2 года или ±5% от значения данного параметра по выборке). Вероятность 95 % при доверительном интервале ±5% означает, что если из рассматриваемой генеральной совокупности случайным образом сформировать 100 одинаковых по объёму выборок, то в 95 из них результат не будет отличаться от гипотетического результата по генеральной совокупности более, чем на 5 %. Соответственно, оставшиеся 5 выборок из 100 будут давать результат, выходящий за границы установленного в данном случае доверительного интервала, следовательно, ошибочный. Приемлемый это уровень точности или нет, решает исследователь в зависимости от поставленных им цели и задач, а также сроков проведения исследования, финансовых возможностей и т. д. В любом случае ему остаётся только надеяться, что сформированная им выборка окажется из числа тех 95, которые дают правильный результат. Обычно уровень доверия 0,95 считается достаточной степенью достоверности результата. Если установить больший уровень доверия, то доверительный интервал станет шире, поскольку чем выше степень неопределённости утверждения (доверительный интервал), тем больше вероятность того, что оно – истинно (уровень доверия). Ясно, что доверительный интервал 38–42 года это – суждение гораздо более определённое, чем, скажем, интервал 35–45 лет (или ±12,5 % от среднего по рассматриваемой в приведённом примере выборке). Поэтому вероятность истинности (уровень доверия) последнего по отношению к генеральной совокупности будет заметно выше и может составить уже, например, 98 % (p=0,98).

Ширина доверительного интервала при заданном уровне доверия зависит, во-первых, от объёма выборочной совокупности, а во-вторых, от разброса или «кучности» данных. Увеличение объёма выборочной совокупности повышает точность оценки, тогда как увеличение разброса непосредственно наблюдаемых значений (или, что то же самое, уменьшение «кучности» данных) снижает её. Под разбросом в данном случае понимается степень удалённости крайних значений признака от средних. Скажем, разброс значений в паре чисел 20 и 40 гораздо выше, чем в паре 29 и 31, хотя среднее арифметическое значение в обоих случаях равно 30. Если применительно к приведённому ранее примеру при прочих равных условиях объём выборочной совокупности будет составлять не 2,5 тыс. человек, а 1,5 или 1 тыс., то придётся либо снижать уровень доверия, либо делать шире доверительный интервал. Следует отметить, что объём выборки даже в 1 тыс. человек считается достаточно большим и в большинстве случаев обеспечивает высокую надёжность получаемых выводов. При уменьшении объёма выборочной совокупности до 100 единиц и ниже метод доверительных интервалов может не работать, а полученные по выборочной совокупности выводы могут совершенно не соответствовать выводам, которые могли бы быть получены при непосредственном обследовании всей генеральной совокупности. Здесь возникает следующий вопрос: каким должен быть оптимальный объём выборки для того, чтобы, с одной стороны, выводы были достаточно точны, а с другой, проведение исследования было экономически целесообразно и не слишком растянуто во времени?

Прежде чем ответить на этот вопрос, вернёмся к условиям отбора объектов в выборку, без соблюдения которых рассмотренные закономерности работать не будут. Вот эти условия: (1) все элементы генеральной совокупности (например, люди или семьи) должны быть одинаково доступны; (2) все они должны иметь равные шансы попасть выборку; (3) их отбор должен производиться строго случайным образом. Несмотря на кажущуюся простоту этих условий, они могут быть реализованы далеко не всегда, и, следовательно, далеко не всегда при неограниченно большой генеральной совокупности можно обойтись выборочной совокупностью относительно малого объёма. Выборка, сформированная в полном соответствии с указанными условиями, называется *случайной*, или *вероятностной*. Существует несколько разновидностей случайной выборки, но все они имеют один общий недостаток: сама возможность их использования напрямую зависит от объёма генеральной совокупности. Дело в том, что если генеральная совокупность большая (десятки, сотни тысяч или даже миллионы человек), то, как уже говорилось, крайне редко все образующие её единицы бывают одинаково доступными.

Итак, процедуры формирования выборки потому и называются случайными или вероятностными, что отбор респондентов в них производится строго случайным образом, единицы генеральной совокупности имеют равную вероятность попасть в выборку, а получаемые при этом значения распределяются строго в соответствии с законами теории вероятностей. Одним из таких законов является закон *нормального распределения*, суть которого состоит в том, что при достаточно большом числе единиц анализа ($n > 100$) и абсолютно случайном их отборе 68,27 % значений интересующего исследователя признака расположено в пределах ±1 *стандартное отклонение* от среднего арифметического значения данного признака, 95,45 % значений расположено в пределах ±2 стандартных отклонения, 99,73 % – в пределах ±3 стандартных отклонения, 99,99 % – в пределах ±4 стандартных отклонения и т. д. Стандартное (*среднеквадратическое*) отклонение представляет собой корень квадратный из *дисперсии* (центрального *момента распределения* второго порядка), то есть величины, характеризующей степень разброса значений рассматриваемого признака. Например, мы каким-то образом выяснили, что средний рост всех людей на Земле составляет 170 см. Допустим также, что стандартное отклонение по признаку «рост» равно 10 см. Тогда насколько большой будет вероятность «натолкнуться» случайным образом на человека ростом 160–180 см? Очевидно, что – 68,27 %, поскольку именно таков процент на нашей планете людей с ростом 170±10 см (согласно закону нормального распределения). Соответственно, вероятность встретить человека с ростом 150–190 см составляет уже 95,45 % и т. д. И если формируемая таким образом выборка будет достаточно большой, то приблизительно 68 % респондентов в ней будет обладать ростом 170±10 см, 95–96% – ростом 170±20 см и т. д., то есть распределение по признаку «рост» по выборочной совокупности будет примерно соответствовать аналогичному распределению в генеральной.

Уровень доверия 95,45 % (р=0,9545) – это процент значений, попадающих в интервал ±2 стандартных отклонения. Если рассчитать выборочные средние по всем возможным одинаковым по объёму выборкам, какие только могут быть сформированы из одной и той же генеральной совокупности, а затем рассчитать среднее арифметическое всех выборочных средних, то полученное значение будет в точности соответствовать генеральной средней (*центральная предельная теорема*). Соответственно, выборочное среднее по одной случайно сформированной выборке будет с определённой вероятностью находиться в определённом интервале вокруг генеральной средней. Сам этот интервал и есть величина случайной ошибки выборки. Поскольку данная величина ошибки является предельно (то есть максимально) допустимой, она называется *предельной ошибкой выборки*. Количество численных значений одного стандартного отклонения в статистике обозначается буквой t. Тогда при $t = 1$ уровень доверия составляет 68,27 %, при $t = 2$ – 95,45 %, при $t = 3$ – 99,73 %, при $t = 4$ – 99,99 %. На графике нормальное распределение выглядит как плавная «колоколообразная» кривая, где

68,27 % это – доля площади под кривой в интервале от t до $-t$ при условии, что всю площадь между кривой и осью абсцисс мы принимаем за 100 %. Соответственно 95,45 % – это доля площади под кривой в интервале от $-2t$ до $2t$; 99,73 % – в интервале от $-3t$ до $3t$ и т. д.

Средняя *величина случайной ошибки выборки*, то есть среднее арифметическое всех отклонений выборочного среднего от генерального среднего, зависит только от двух параметров: разброса значений признака по генеральной совокупности и объёма выборки. Если выборка абсолютно случайна, а её объём достаточно большой (n > 100), то разброс значений какого-либо признака по генеральной совокупности, как уже говорилось, должен соответствовать разбросу значений этого признака по выборке. Поэтому средняя величина случайной ошибки выборки определяется по формуле:

$$\mu = \frac{\sigma}{\sqrt{n}}\ , \tag{1}$$

где σ – стандартное отклонение, n – объём выборки.

Предельная же ошибка выборки непосредственно зависит от приемлемого для исследователя уровня точности (t) и определяется по формуле:

$$\Delta = t \times \mu, \tag{2}$$

Стандартное отклонение рассчитывается по формуле:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(\bar{x} - x_i)^2 p_i}{\sum_{i=1}^{n} p_i}}\ , \tag{3}$$

где $\bar{x}$ – среднее арифметическое значение интересующего признака, x_i – каждое непосредственно наблюдаемое значение данного признака, p_i – весовой коэффициент, то есть число раз, которое данное численное значение признака встречается в совокупности.

Среднее арифметическое значение рассчитывается по формуле:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i p_i}{\sum_{i=1}^{n} p_i}, \qquad (4)$$

где x_i – каждое непосредственно наблюдаемое значение данного признака, p_i – весовой коэффициент.

Например, требуется рассчитать среднее арифметическое значение и стандартное отклонение по признаку «возраст» в условной совокупности из трёх человек, возрасты которых составляют 18, 18 и 19 лет. Тогда $\bar{x} = \frac{18 \times 2 + 19 \times 1}{2 + 1} \approx 18{,}33$ года; $\sigma = \sqrt{\frac{(18{,}33 - 18)^2 \times 2 + (18{,}33 - 19)^2 \times 1}{2 + 1}} \approx 0{,}47$ года.

Формула (1) используется только для признаков, значения которых получены по интервальной шкале (или шкале более высокого уровня). Поскольку стандартное отклонение – это корень квадратный из дисперсии ($\sigma = \sqrt{s}$), то формулу (1) можно преобразовать:

$$\mu = \sqrt{\frac{s}{n}}, \qquad (5)$$

где s – дисперсия признака, по которому производится расчёт.

Дисперсия количественного признака рассчитывается по формуле:

$$s = \frac{\sum_{i=1}^{n} (\bar{x} - x_i)^2 p_i}{\sum_{i=1}^{n} p_i}, \qquad (6)$$

где $\bar{x}$ – среднее арифметическое значение интересующего признака, x_i – каждое непосредственно наблюдаемое значение данного признака, p_i – весовой коэффициент.

Формулу (5) использовать более предпочтительно, чем формулу (1), поскольку дисперсия, в отличие от стандартного отклонения, может быть рассчитана и для качественных признаков, то есть тех, численные значения которых получены по шкале низкого уровня. Дисперсия качественного признака рассчитывается по формуле:

$$s = p\,(1 - p), \tag{7}$$

где p – доля единиц в совокупности, обладающих интересующим исследователя признаком, или, другими словами, вероятность случайным образом «натолкнуться» на данный признак.

Например, имеется некий регион или организация, где представительство жителей или сотрудников по признаку «пол» следующее: 55 % женщин и 45 % мужчин. Тогда доля женщин составляет 0,55, а мужчин, соответственно, – 0,45. Подставляя долю женщин в формулу (7), получаем численное значение дисперсии: s=0,55(1–0,55)=0,55×0,45≈0,248. Если в формулу мы подставим показатель количества не женщин, а мужчин, то результат будет тот же: s=0,45(1–0,45)=0,45×0,55 (от перемены мест множителей произведение не меняется).

Формула (5) используется только в том случае, если генеральную совокупность мы считаем бесконечной. Эмпирически установлено, что бесконечной генеральную совокупность можно считать в том случае, если её объём составляет 100 тыс. единиц и более. В прикладной социологии это, как правило, город, регион, страна в целом, большие по численности категории населения (все студенты, все пенсионеры и т. д.); гораздо реже – организации, поскольку организаций таких размеров очень мало. Если же генеральная совокупность составляет менее 100 тыс. единиц, то в формулу (5) следует ввести поправку на объём генеральной совокупности, что повышает точность оценки:

$$\mu = \sqrt{\frac{s}{n}\left(1 - \frac{n}{m}\right)}, \tag{8}$$

где m – объём генеральной совокупности.

Очевидно, что если m ≥ 100000, то дробь n/m в формуле (8) примет такое значение, которое при вычитании из единицы даст практически опять же единицу. Следовательно, скобкой в этом случае можно пренебречь, что даёт нам формулу (5). Из формул (2) и (5) можно получить формулу предельной ошибки выборки (доверительного интервала):

$$\Delta = t \times \sqrt{\frac{s}{n}} \ . \qquad (9)$$

Из формулы (9) получаем формулу объёма выборочной совокупности:

$$n = \frac{t^2 \times s}{\Delta^2}, \qquad (10)$$

где t – достаточный по мнению исследователя уровень доверия, Δ – приемлемая по его же мнению величина доверительного интервала или предельной ошибки, а s – дисперсия признака генеральной совокупности, по которому производится расчёт объёма выборки.

Обращаем внимание, что формулы (9) и (10) также применяются при бесконечной генеральной совокупности, так как выводились они из формулы (5). Следует отметить, что величина Δ измеряется обычно в процентах, но в формулу подставляются не проценты, а абсолютное значение (например, при заданном доверительном интервале ±5 % в формулу подставляют значение 0,05). Ясно, что если нет вообще никакой информации о структуре генеральной совокупности, то применение каких бы то ни было формул для расчёта объёма выборки становится невозможным. В этом случае следует рассуждать логически. В формуле (10), как уже было сказано, два параметра из трёх (t и Δ) в любом случае задаются самим исследователем. Остаётся дисперсия (s). Если признак, по которому нам необходимо сделать расчёт, но значения которого нам неизвестны, – качественный, то нужно подставить в формулу максимально возможное значение дисперсии. Это значение тем выше, чем более равномерно распределены в совокупности единицы по интересующему признаку. Если в приведённом ранее примере с формулой (7) соотношение по полу будет не 55 % и 45 %, а 50 % и 50 %, то дисперсия станет равной 0,5(1–0,5)=0,25. Это и есть её максимально возможное значение. Так как большинство признаков, с которыми работают социологи, – качественные, то данным численным значением дисперсии можно пользоваться в большинстве случаев. Подставляя его в формулу (10), а также наиболее часто используемые параметры точности (уровень доверия – 0,95 или $t \approx 2$, доверительный интервал – ±5 %), получаем: $n = \frac{2^2 \times 0{,}25}{0{,}05^2} = \frac{1}{0{,}0025} = 400$.

Итак, если генеральная совокупность бесконечная (≥100000) и о ней ничего или почти ничего неизвестно (что, кстати, бывает крайне редко, поскольку сведения о любом населённом пункте или регионе всегда можно найти в статистических сборниках), то, скорее всего, достаточно будет опросить 400 человек, чтобы получить достоверные данные. Но иногда при строго случайном отборе и обычных параметрах точности объём выборки может значительно превышать 400 человек и составлять, например, 1 тыс. или даже 2 тыс. человек. Это делается для того, чтобы при заданном уровне доверия оценить различия внутри выборочной совокупности между группами, выделяемыми по значимым для исследователя критериям.

Если генеральная совокупность у нас меньше 100 тыс. человек, то формулы расчёта предельной ошибки и объёма выборки несколько усложняются. Их можно получить из формул (2) и (8):

$$\Delta = t \times \sqrt{\frac{s}{n}\left(1 - \frac{n}{m}\right)}, \qquad (11)$$

Отсюда:

$$n = \frac{t^2 \times s}{\Delta^2 + \dfrac{t^2 \times s}{m}}\ . \qquad (12)$$

Если признаков генеральной совокупности, по которым возможен расчёт дисперсии, несколько, то используют признак, дисперсия которого наибольшая, либо – максимально возможное значение дисперсии (0,25), если все признаки качественные.

Многие начинающие исследователи, проводя социологический опрос, иногда пребывают в полной уверенности, что реализуют случайную выборку, тогда как на самом деле занимаются целенаправленным отбором. Это распространённое заблуждение происходит из-за неверного толкования термина «случайная» выборка. Рассуждают примерно так: случайный отбор это – отбор «случайных» людей на улице. Что в этом случае происходит? Допустим, нас интересует мнение жителей города по какому-либо вопросу. Для того, чтобы это мнение узнать, мы днём выходим на улицы и опрашиваем случайных прохожих в количестве, которое считаем необходимым. Это случайная выборка или нет? Ответ – однозначно нет, поскольку в этом случае мы не сможем распространить полученные результаты на весь объект исследования. Распространить же мы их сможем только на людей в городе, гуляющих по улицам в рабочее время, а это – далеко не весь город. Налицо явная ошибка смещения, причём существенная, которая может привести к заметным искажениям результатов исследования.

Литература

1. *Антонов Г. В.* Особенности применения выборочного метода в прикладной социологии / Г. В. Антонов // Вестник Волгоградского государственного университета. – 2005. – Сер. 7. – Вып. 4. – С. 111–117.

2. *Основы* прикладной социологии / под. ред. Ф. Э. Шереги, М. К. Горшкова. – Москва : Интерпракс, 1996. – С. 31–38.

© **Антонов Георгий Вячеславович (2013),** кандидат социологических наук, доцент кафедры социологии, Волгоградский государственный университет (Волгоград), antonovgv@mail.ru.