Содержание

1 Выборочный метод: определение и истоки

# 2 Типы вероятностных выборок и их реализация

# 3 Размер вероятностной выборки

# 4 Целевой отбор

Библиографический список

1 Выборочный метод: определение и истоки

Задача построения выборки возникает всякий раз, когда необходимо собрать информацию о некоторой группе или большой совокупности людей. Выборку в той или иной форме используют в ориентированных на «жесткие» статистические методы опросах, в исследованиях политических и культурных элит и даже при отборе «случаев» для включенного наблюдения и качественного анализа.

Статистические (или квазистатистические) обследования населения и ресурсов, судя по всему, зародились одновременно с первыми формами централизованной социальной и политической организации: развитые аграрные общества и древние города-государства нуждались в такой информации и использовали ее при решении разнообразнейших управленческих задач — от фискальной политики до строительства общественных бань. Эти обследования иногда принимали форму сплошных переписей населения. (Об одной такой переписи, имевшей, правда, самые печальные последствия, рассказывает нам книга пророка Самуила: когда царь Давид (X в. до н. э.) осуществил перепись населения древнего Израиля, в стране разразилась страшная эпидемия (2 Цар. 24). Однако значительно чаще приходилось довольствоваться сведениями о какой-то части совокупности: об урожайности судили по пробному обмолоту, о партии товара — по образцу, а о прихожанах — по их духовному наставнику.

Выборка — это подмножество заданной совокупности (популяции), позволяющее делать более или менее точные выводы относительно совокупности в целом. Зачем нужно строить выборки? Прежде всего, из практических соображений, так как выборка экономит силы и средства исследователей. Проведение полномасштабной переписи или сплошного опроса населения требует значительных финансовых и трудовых затрат, которые к тому же могут пропасть впустую в случае, если в разработке методики исследования были допущены принципиальные просчеты.

Другая причина заинтересованности в выборках связана с тем, что выборочная процедура представляет собой удобную и экономичную форму индуктивного вывода. Третья причина заключается в том, что эта процедура реализует фундаментальный принцип рандомизации, т. е. случайного отбора (от англ. random — случайный, выбранный наугад).

Представление о том, что отбор наблюдений должен носить случайный, непредумышленный характер, в общем соответствует нашему интуитивному знанию об условиях вынесения объективного и непредвзятого суждения. Однако строгая, т. е. математико-статистическая, теория случайной выборки вплоть до конца XIX — начала XX вв. не пользовалась популярностью в среде профессиональных статистиков. Многим исследователям казалось, что в основе отбора должна лежать не «игра случая», а поиск типичных, характерных наблюдений. Это убеждение препятствовало применению в массовых обследованиях методов теории вероятности, достигшей высочайшего уровня развития уже в XVIII— первой половине XIX вв. Применимость выборочного метода для изучения случайно распределенных признаков, например дохода или размера семьи, была впервые обоснована в работах норвежца А. Киэра, англичан А. Боули и К. Пирсона, а также русского статистика А. И. Чупрова.

Следующим принципиально важным шагом в развитии выборочного метода стала осуществленная Р. Фишером разработка техники рандомизации в эксперименте и выборочном наблюдении.. Что же касается выборочного обследования, то оно часто используется как «замена» экспериментального метода. Нельзя провести эксперимент, в котором людям в случайном порядке присваиваются определенные значения переменных «пол» или «цвет кожи». Однако применение выборочного метода и статистического анализа, как мы увидим в дальнейшем, позволяет справляться с этими ограничениями и делать выводы о взаимосвязях между самыми разными переменными, включая вышеупомянутые. Но для того, чтобы такие выводы были обоснованы, нужно устранить любое систематическое влияние «посторонних», смешивающих факторов на изучаемые переменные. Единственным средством для достижения этой цели является абсолютно случайный характер отбора наблюдений. Лишь равенство шансов попадания в выборку для каждого наблюдения, т. е. отбор «наугад», гарантирует от намеренных или ненамеренных искажений. Пусть, например, в ходе опроса мы изучаем влияние пола и рода занятий респондента на его отношение к планированию семьи и ограничению рождаемости. Если используемая нами выборочная процедура ведет к тому, что работающие женщины имеют несколько меньшие шансы стать респондентами, чем домохозяйки и пенсионерки (последних, как известно, проще застать дома), наши результаты наверняка окажутся смещенными.

Поэтому наилучшей моделью отбора считается вероятностная, или случайная, выборка, в которой строго соблюдается принцип равенства шансов попадания в выборку и для всех единиц изучаемой совокупности, и для любых последовательностей таких единиц.

Именно с рассмотрения разных подходов к построению вероятностной выборки мы и начнем наше обсуждение, чтобы в дальнейшем перейти к не столь совершенным видам целевого, т. е. не основанного на вероятностях отбора, и их роли в практике социологических исследований.

Выше мы определили, что такое выборка. Сейчас нам необходимо строго определить еще несколько элементарных понятий. Переписью называют процедуру сбора информации о каждом члене изучаемой группы или популяции. Все члены интересующей исследователя группы (популяции) составляют генеральную совокупность. Выборочная процедура обеспечивает обоснованность и «законность» выводов о генеральной совокупности, сделанных на основании небольшой выборки.

# 2 Типы вероятностных выборок и их реализация

Первым шагом в построении любой модели отбора, включая вероятностную, является определение генеральной совокупности. Решение этой задачи далеко не всегда бывает очевидным. Прежде всего, генеральная совокупность, т. е. множество интересующих социолога объектов исследования, может быть задана и описана лишь на основе каких-то содержательных представлений. Если, например, нас интересуют политические пристрастия избирателей, естественно включить в генеральную совокупность лишь тех, кто уже достиг 18-летнего возраста. Изучение факторов, влияющих на формирование семейного бюджета горожан, потребует иного определения генеральной совокупности: интересующая исследователя популяция в данном случае будет состоять из городских семей.

Полезно также помнить о том, что идеальная генеральная совокупность, задаваемая теоретическим описанием предмета исследования, почти никогда не будет полностью совпадать с реальной совокупностью. Реальная генеральная совокупность подвержена постоянным колебаниям: «взрослое население города Воронежа на 00 час 15 ноября 1996 года» будет отличаться от «взрослого населения города Воронежа на 00 час 16 ноября 1996 года». Некоторые люди за день уедут из города, попадут в больницу, некоторые — вернутся домой из командировки и т. п. Поэтому столь важно при описании изучавшейся в исследовании генеральной совокупности указывать время и место проведения исследования. Следует также помнить, что идеальная генеральная совокупность — это теоретическая абстракция, более или менее совпадающая с реальной совокупностью. Выборка осуществляется из реальной популяции, переход от которой к идеальной совокупности обеспечивается не только правилами статистического вывода, но и некоторой долей теоретического воображения.

Если исследователь построил выборку, которая представляет интересующую его совокупность с приемлемой степенью точности, то полученная выборка является репрезентативной (представительной). В противоположном случае можно говорить о наличии существенной выборочной ошибки. Более строго выборочную ошибку определяют как расхождение между оценкой некоторого показателя, получаемой на основании исследования выборки, и истинным значением этого показателя в генеральной совокупности.

К счастью, существуют точные методы для учета и оценки случайной выборочной ошибки, связанной с не носящими систематического характера колебаниями изучаемой переменной в разных подвыборках из одной и той же генеральной совокупности. Подробнее эти методы мы будем обсуждать ниже. Значительно более серьезную проблему создает наличие систематических смещений, возникающих в результате нарушения случайного характера выборочной процедуры. Результаты такого «не вполне случайного» отбоpa могут выглядеть более или менее правдоподобно, однако сами по себе он: никогда не позволят обнаружить смещение или оценить его величину.

Последнее утверждение можно проиллюстрировать на примере классического опыта с рулеткой. Если нам скажут, что вчера десять раз подряд выпало «красное», мы сможем назвать такую серию событий крайне маловероятной. Однако этот субъективно подозрительный результат сам по себе не дает оснований для каких-то суждений о величине и характере ошибок, порождаемых выборочной процедурой, т. е. об исправности механизма самой рулетки.

Систематическая ошибка выборки не обязательно является результатом злого умысла. Например, в США во время войны во Вьетнаме (до введения контрактной системы набора на армейскую службу) правительство проводило специальные лотереи для отбора призывников. Фактически случайно отбирались даты рождения: все годные к несению строевой службы юноши, родившиеся в день, который определялся в ходе такого «розыгрыша», призывались в армию. В 1970 г. результаты отбора были подвергнуты острой критике. Проведенное специальной комиссией расследование показало, что в выборочной процедуре действительно присутствовало смещение. Билетики с напечатанными датами были заключены в специальные капсулы, которые затем опускали в лотерейный барабан в порядке следования месяцев, начиная с января. Из-за недостаточного перемешивания капсул внутри барабана капсулы с ноябрьскими и декабрьскими датами концентрировались в верхней части и, естественно, выпадали с заметно большей частотой.

Самым знаменитым примером смещенной выборочной процедуры в истории социологии стал предвыборный опрос, проведенный американским журналом «The Literary Digest» в 1936 г. Результаты опроса показывали, что Ф. Д. Рузвельт получит 40,9% голосов и уступит президентское кресло республиканцу А. Ф. Лэндону. В действительности Рузвельт получил 60,2% голосов избирателей. Расхождение в 19,3% в значительной степени объяснялось характером выборочной процедуры. Дело в том, что на практике для построения любой выборки используют какой-то список всех членов изучаемой совокупности, называемый основой выборки. В опросе, проведенном «The Literary Digest», в качестве основы выборки использовались телефонные справочники, а также регистрационные списки владельцев автомобилей. Во второй половине 1930-х гг. такие списки включали в себя почти исключительно представителей экономически благополучных классов. Беднейшие слои населения, избирательная активность которых, кстати, существенно увеличилась в годы Великой Депрессии, оказались недостаточно представлены в выборке, что и послужило причиной столь значительной ошибки. (Интересно отметить, что объем выборки в описываемом случае был просто огромным — свыше двух миллионов человек!)

Существует несколько типов вероятностной выборки, различающихся характером выборочной процедуры. Мы рассмотрим лишь пять: простую случайную, систематическую, стратифицированную, кластерную и многоступенчатую.

Процедура построения простой случайной выборки включает в себя следующие шаги.

Во-первых, нужно получить полный список членов генеральной совокупности и пронумеровать этот список. Такой список, напомним, называется основой выборки.

### Таблица 1

Таблица случайных чисел

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Номер столбца Номер строки | 1 | 2 | 3 | 4 | 5 |  | 6 | 7 | 8 | 9 | 10 |
| 1 | 98 | 08 | 62 | 48 | 26 |  | 45 | 24 | 02 | 84 | 04 |
| 2 | 33 | 18 | 51 | 62 | 32 |  | 41 | 94 | 15 | 09 | 49 |
| 3 | 80 | 95 | 10 | 04 | 06 |  | 96 | 38 | 27 | 07 | 74 |
| 4 | 79 | 75 | 24 | 91 | 40 |  | 71 | 96 | 12 | 82 | 96 |
| 5 | 18 | 63 | 33 | 25 | 37 |  | 98 | 14 | 50 | 65 | 71 |
| 6 | 74 | 02 | 94 | 39 | 02 |  | 77 | 55 | 73 | 22 | 70 |
| 7 | 54 | 17 | 84 | 56 | 11 |  | 80 | 99 | 33 | 71 | 43 |
| 8 | 11 | 66 | 44 | 98 | 83 |  | 52 | 07 | 98 | 48 | 27 |
| 9 | 48 | 32 | 47 | 79 | 28 |  | 31 | 24 | 96 | 47 | 10 |
| 10 | 69 | 07 | 49 | 41 | 38 |  | 87 | 63 | 79 | 19 | 76 |
| 11 | 09 | 18 | 82 | 00 | 97 |  | 32 | 82 | 53 | 95 | 27 |
| 12 | 90 | 04 | 58 | 54 | 97 |  | 51 | 98 | 15 | 06 | 54 |
| 13 | 73 | 18 | 95 | 02 | 07 |  | 47 | 67 | 72 | 52 | 69 |
| 14 | 75 | 76 | 87 | 64 | 90 |  | 220 | 97 | 18 | 17 | 49 |
| 15 | 67 | 35 | 86 | 33 | 26 |  | 50 | 10 | 39 | 42 | 61 |

Во-вторых, следует определить предполагаемый объем выборки, т. е. ожидаемое число опрошенных.

В-третьих, нужно извлечь из таблицы случайных чисел столько чисел, сколько нам требуется выборочных единиц. Если в выборке должно оказаться 100 человек, из таблицы берут 100 случайных чисел.

В-четвертых, нужно выбрать из списка-основы те наблюдения, номера которых соответствуют выписанным случайным числам.

Прежде чем мы перейдем к обсуждению возникающих на этом пути практических затруднений, рассмотрим упрощенный пример реализации описанной процедуры.

Пусть нам предстоит построить случайную выборку объемом в 12 человек из совокупности, содержащей 60 членов. Можно предположить, что мы хотим оценить калорийность ежедневного рациона питания 60 студентов-социологов, обучающихся на втором курсе университета, чтобы исследовать возможное влияние энергетической ценности рациона на академическую успеваемость. Для этого можно пронаблюдать за питанием небольшой выборки, состоящей из двенадцати студентов. В качестве основы выборки мы используем список всех 60 студентов. Присвоим всем студентам в списке двузначные номера — от «01» до «60» (если бы максимальный номер в списке был трехзначным, мы бы присваивали трехзначные номера, используя нули в отсутствующих разрядах — например, «067», «003»). Далее нам предстоит последовательно выписать двенадцать двузначных чисел из таблицы случайных чисел. Отметим, что таблицы случайных чисел фактически состоят из случайных цифр, которые обычно сгруппированы для удобства в блоки, состоящие из двузначных либо пятизначных чисел. Объединение цифр в последовательности и блоки условно и не имеет особого статистического смысла. Поэтому в случаях, когда нужны, например трехзначные числа, а таблица состоит из пятизначных, пользуются каким-то несложным правилом, скажем, используют только три первые цифры каждого пятизначного числа, а оставшиеся две игнорируют. Соответственно двузначные числа можно объединять.

Чтобы решить, с какого места в таблице начинать отсчет номеров, достаточно задаться произвольными номерами строки и столбца. В нашем примере мы начнем с пересечения второй строки и третьего столбца. Первым номером в нашем списке окажется 51. Далее можно двигаться по любому правилу: подряд, через строку, через два столбца и т. п. Мы будем выписывать нужные нам двенадцать двузначных номеров подряд по строке, двигаясь по горизонтали и переходя при необходимости на следующую строку.

Если при этом будут попадаться числа, превосходящие по величине самый большой номер в нашем списке (60), мы будем их пропускать. То же относится и к повторяющимся числам. В результате мы получим последовательность:

51, 32, 41, 15, 09, 49, 10, 04, 06, 38, 27, 07.

Нам остается выписать из списка-основы фамилии, стоящие под этими номерами. Если вы располагаете персональным компьютером, то вместо таблицы можно воспользоваться «генератором случайных чисел», имеющимся в большинстве статистических программ.

Простая случайная выборка — это не только наглядное воплощение идеи случайного отбора, но и своего рода эталон, с которым сравниваются другие вероятностные процедуры. Здесь необходимо заметить, что вопреки часто высказываемому и неверному мнению простую случайную выборку не следует рассматривать как самую примитивную форму вероятностного отбора. Напротив, более сложные модели случайных выборок используют в тех случаях, когда простую нельзя применить из-за практических или финансовых ограничений. О качестве этих более сложных процедур отбора также судят посредством сравнения с простой случайной выборкой.

Самые очевидные ограничения для использования простой выборки возникают в случае большого объема генеральной совокупности. Прежде всего исследователь сталкивается с проблемами поиска полной и несмещенной основы выборки. При обследованиях небольших групп и первичных коллективов эти проблемы обычно легко решаются: достаточно воспользоваться членскими списками, списками личного состава и т. п., внеся в них необходимые уточнения. В широкомасштабных опросах общественного мнения и социологических обследованиях чаще применяют другие основы: переписные листы, списки избирателей, домовые книги, карточки паспортных столов милиции (а также картотеки РЭУ, ДЭЗ и т. п.), нехозяйственные книги сельских советов. Все эти «готовые» основы выборки обладают определенными преимуществами и недостатками. Решая практическую задачу планирования выборочного исследования, социолог обычно оценивает возможные основы по нескольким параметрам.

Во-первых, списки, пригодные для составления основы выборки, могут храниться либо централизованно, либо децентрализованно, «вразброс», в различных территориальных органах власти, статистических учреждениях и т.п. Естественно, что в первом случае затраты на получение доступа к основе будут значительно ниже, чем во втором. Фактически при децентрализованном хранении исследователь должен самостоятельно составить единый список-основу, собрав необходимые данные в результате обхода (или объезда) всех соответствующих институций.

Во-вторых, используемые в качестве основы выборки списки могут обладать различной степенью точности. Точность списка, в свою очередь, зависит от его полноты, частоты его обновления. Эти качества (полнота списка и высокая частота его пересмотра) редко встречаются одновременно. Как правило, самыми полными оказываются именно те основы, которые реже всего обновляются. Таковы, конечно, данные переписей или эпизодически составляемые именные распределительные списки (типа списков на получение приватизационных чеков). К сожалению, чем больше времени отделяет планируемое вами исследование от последней переписи, тем больше вероятность возникновения ошибок и смещений в основе выборки.

Очень существенными достоинствами обладают списки паспортных столов милиции, жилищно-эксплуатационных контор и других местных административных органов.

Качество основы выборки оценивают уже на стадии планирования исследования. Особое внимание уделяют таким потенциальным угрозам валидности, как неполнота выборочной основы, «склеивание» единиц отбора, «пустые» элементы в списке. О неполноте говорят в тех случаях, когда список, используемый для построения выборки, не содержит в себе некоторые единицы, безусловно относящиеся к целевой совокупности. Например, списки жильцов могут не содержать сведений о тех жильцах, которые еще не зарегистрировались по новому месту жительства. В некоторых случаях проблему неполной основы можно решить за счет использования дополнительных основ. В нашем примере со списками жильцов такой дополнительной основой могут стать «листки прибытия-убытия», которые хранятся в паспортных столах отделений милиции (с помощью последних ведется учет прописки граждан). Примером «склеивания» может служить ситуация, когда генеральная совокупность, определяемая объектом исследования, состоит из индивидов, а реальной основой отбора служит список квартир или домовладений, содержащий лишь сведения об ответственных квартиросъемщиках либо о собственниках недвижимости. «Пустые» цементы в основе выборки встречаются в тех случаях, когда исходный список содержит имена или адреса, за которыми не стоят реально существующие (или практически доступные) выборочные единицы. Эта проблема часто возникает при использовании устаревших списков, содержащих информацию о временно уехавших, выбывших, умерших и т. п.

Описанные выше трудности составления валидной, т.е. соответствующей объекту исследования (целевой совокупности), основы выборки носят и статистический, и «экономический» характер. Довольно часто исследователь сталкивается с ситуацией, когда временные и финансовые затраты на осуществление простой случайной выборки становятся неприемлемо высокими. Наиболее разумным выходом здесь является использование других, «компромиссных», процедур случайного отбора.

Систематическая выборка по качеству часто приближается к простой случайной. Систематическая выборка, как и простая случайная, требует полного списка или заданного упорядочения совокупности (см. ниже). Техника осуществления систематического отбора элементарна: сначала случайным образом отбирается первая единица, затем отбору подлежит каждый k-й элемент. Число k в данном случае называют шагом отбора. Можно, например, отбирать каждый 25-й или каждый 200-й элемент. Чтобы определить шаг отбора, нужно поделить известный объем генеральной совокупности (N) на предполагаемый объем выборки (n).

Пусть, например, нужно отобрать 200 человек из 20000 владельцев телефонов:

1) определим шаг отбора: N/n = 20000 : 200 = 100;

2) с помощью таблицы случайных чисел найдем первую выборочную единицу. Если, скажем, выпал номер «053», то из списка владельцев телефонов выпишем того, кто значится под этим номером;

3) с установленным шагом отбираем номера: 153, 253, 353, 453 и т. д. до исчерпания списка.

Иногда генеральная совокупность (и соответственно основа выборки) слишком велика либо исследователю известен не полный список, а лишь правило упорядочения элементов в генеральной совокупности. Предположим, что мы хотим составить представление о весе и формате книг, содержащихся в некой библиотеке, при том, что мы не располагаем полным каталогом, а лишь видим, как книги расставлены на стеллажах. При условии, что объем библиотечного собрания нам приблизительно известен, мы можем воспользоваться процедурой систематического отбора и отобрать, скажем, каждую 55-ю книгу. Очень важно отобрать «стартовую» единицу сугубо случайным образом. Именно в этом пункте кроется основная слабость систематического отбора. Если в способе упорядочения единиц совокупности имеет место некая цикличность, т. е. неизвестная нам «система» (систематический паттерн), а случайность в выборе «старта» должным образом не обеспечена, то полученная выборка может также оказаться смещенной (если о систематическом паттерне мы знаем заранее, то он не представляет собой угрозы валидности и может быть учтен в ходе отбора). Если воспользоваться примером с отбором книг в библиотеке, то легко представить себе такую гипотетическую ситуацию: исследователь выбирает в качестве стартовой первую книгу на нижней полке ближайшего стеллажа и далее двигается с шагом 250 единиц. Если на каждом стеллаже размещается около 500 книг, то приблизительно половина его выборки будет взята с нижних полок. Однако известно, что на нижних полках многих библиотек нередко размещают книги больших форматов — художественные альбомы, атласы и т. п. Если в нашем примере это правило упорядочения будет соблюдено хотя бы в половине случаев (т. е. половина нижних полок будет отведена под «неформатные» издания, под так называемые фолио), любые выборочные оценки «направленности» библиотечного собрания или формата представленных в нем книг окажутся невалидными. Аналогией примеру с библиотечными книгами может служить случай систематической выборки городских квартир. Если в результате осуществляемого непосредственно «в поле» интервьюерами систематического отбора в выборке будут сверхпредставлены квартиры, расположенные на первых и последних этажах, возникнет систематическая выборочная ошибка. На первых и последних этажах в российских городах часто живут люди из групп, имеющих более низкий социально-экономический статус и соответственно ограниченные финансовые ресурсы: квартиры, расположенные на «крайних» этажах и соприкасающиеся с системами коммунального водо- и теплоснабжения, обычно стоят дешевле, так как названные системы в России традиционно являются источником неприятностей и дисфункций в структуре жизнеобеспечения.

Стратифицированный отбор и соответственно стратифицированная выборка используются в тех случаях, когда из каких-то содержательных соображений важно обеспечить представительность вероятностной выборки по каким-то конкретным важным для исследовательских целей критериям. В литературе существует определенная путаница вокруг проблемы стратификации («страта» — это социальная, возрастная или иная группа, буквально «слой»).

Применительно к стратифицированному отбору часто высказывают все те неверные и предрассудочные мнения, которые в начале XX века высказывались относительно квотной выборки (см. ниже) и ее воображаемых преимуществ перед случайным отбором. В действительности стратифицированный отбор имеет определенные практические преимущества до тех пор, пока сохраняется его вероятностный, случайный характер. Как только стратифицированная выборка превращается в более или менее специально отобранную квотную выборку, воспроизводящую некоторые известные пропорции генеральной совокупности (например, 51% женщин, 30% горожан и т. п.), любые статистические, т. е. строгие, оценки параметров генеральной совокупности становятся невозможными.

Стратификацией, строго говоря, называют процедуру, при которой отбор осуществляют как бы из нескольких «параллельных» подсовокупностей, заданных на одной и той же генеральной совокупности. Это абстрактное определение можно прояснить с помощью примера. Пусть у нас есть генеральная совокупность взрослых горожан, относительно которой мы располагаем какой-то существенной с точки зрения исследовательских гипотез информацией. Наличие такой предварительной информации — необходимое условие стратифицированного отбора. Предположим, мы знаем, что в генеральной совокупности 60% рабочих и 40% служащих. Это соотношение может оказаться весьма существенным с точки зрения наших исследовательских гипотез, если оно задает одну из независимых переменных, как, например, при изучении влияния рода занятий на частоту посещения футбольных матчей. Даже при отсутствии значительной систематической погрешности небольшие смещения в реализации случайной выборочной процедуры могут привести к ситуации, когда в нашей конкретной выборке соотношение рабочих и служащих будет существенно (на 5—7%) отклоняться от ожидаемой «правильной» пропорции, имеющей место в генеральной совокупности. Соответственно под угрозой окажется точность наших оценок взаимосвязи между главной независимой переменной (профессиональным статусом) и интересом к футболу. Такого рода неточность может быть устранена при использовании еще одной случайной выборки из генеральной совокупности, но здесь вступают в силу экономические соображения, так как исследовательский бюджет обычно ограничен. В описанной ситуации желательно заранее обеспечить представленность обеих интересующих нас групп, т. е. страт, сохранив вероятностный характер отбора. Этого можно добиться, если осуществить некую независимую процедуру случайного отбора для каждой социальной группы в отдельности (в нашем примере для рабочих и служащих) и затем объединить полученные случайные подвыборки в одну (заметьте, что для нашего примера объем подвыборки рабочих, в согласии с заранее известной пропорцией, будет в 1,5 раза больше объема подвыборки служащих). Полученная в результате выборка будет и стратифицированной (по профессиональному статусу), и вероятностной.

На практике две случайные процедуры отбора в подвыборки-страты можно технически объединить в одну, если мы располагаем априорной информацией о принадлежности каждой выборочной единицы к той или иной страте. Для этого достаточно вести параллельный отбор из списка-основы в несколько подвыборок (по числу страт). Собственно выборочная процедура может быть и простой случайной, и систематической (соответственно мы получим либо простую, либо систематическую стратифицированную выборку).

Рассмотрим эту процедуру на примере составления систематической выборки населения, стратифицированной по этнической принадлежности. Пусть мы осуществляем выборку взрослых жителей небольшого промышленного центра, при этом полученная выборка должна отражать существующую этнодемографическую ситуацию: 80% русских, 10% украинцев и 10% представителей других национальностей. Основываясь на информации, хранящейся в паспортных столах милиции (или на избирательных списках), мы в идеальном случае можем составить полный список-основу, включающий 100000 известных административным органам постоянных жителей. Если предварительно мы предполагаем включить в нашу выборку около 1000 человек, нам нужно отобрать из картотек паспортных столов (или избирательных списков) каждого сотого. То есть доля генеральной совокупности f, включенная в выборку, составит 1/100:

f = объем выборки (и) / объем целевой совокупности (N).

Выборка объемом в 1000 человек будет включать в себя 800 русских, 100 украинцев и 100 представителей других национальностей. Причем шаг систематического отбора (К) для всех трех подсовокупностей будет равен 100.

Определение шага отбора (К):

80000 человек в «русской» страте: 800 русских в выборке = 100;

10000 человек в «украинской» страте: 100 украинцев в выборке = 100;

10000 человек в страте «другие национальности»: 100 представителей других национальностей в выборке = 100.

Таким образом, мы будем выписывать из реальных картотек (списков) каждого сотого русского, каждого сотого украинца и т.п. (естественно, украинцы и представители других национальностей будут встречаться в списках в среднем в 10 раз реже русских).

Выборка в описанном нами примере является пропорциональной, так как она представляет все страты в той пропорции, в которой они содержатся в генеральной совокупности. Пропорциональный стратифицированный отбор особенно важен для целей дескриптивной, описательной статистики, т. е. когда перед исследователем стоит задача, основываясь на выборке, описать, как распределены те или иные параметры в разных группах генеральной совокупности. Именно так обычно можно сформулировать цель предвыборного опроса, маркетингового исследования покупательских предпочтений и т. п. Еще одним преимуществом стратифицированного вероятностного отбора является уменьшение такого источника общей ошибки измерения, как дисперсия выборки. Не вдаваясь здесь в статистические тонкости, заметим, что стратификация уменьшает так называемую стандартную ошибку лишь в том случае, если интересующая исследователя переменная значительно варьирует между стратами, т. е. когда заранее выделенные страты (например, возрастные группы) сильно отличаются по уровню измеряемой переменной (например, по частоте посещения дискотек). При этом различия внутри страт должны быть относительно невелики, т. е. межгрупповой разброс значений переменной должен значительно превосходить внутригрупповой.

Иногда, однако, основной задачей исследования является сравнение различных, обычно важных с точки зрения некоторой теории, групп внутри выборки с целью описания некоторого соотношения, имеющего место в генеральной совокупности. Некоторые из таких «теоретически релевантных» групп могут быть весьма малочисленными. Для того чтобы сделать такие малочисленные группы-субпопуляции статистически сопоставимыми с другими группами и, следовательно, получить статистически значимые выводы о существующих (несуществующих) межгрупповых различиях, можно использовать два метода.

Первый метод заключается в увеличении объема выборки. В этом случае пропорционально возрастает объем «редкой» страты, но столь же быстро (а иногда и быстрее) растут расходы на проведение исследования. Если, например, пожилые люди старше 85 лет составляют лишь 1/20 часть целевой совокупности горожан-пенсионеров, то в исследовании эффективности социальной работы с пожилыми людьми нам понадобится выборка объемом 4000 пенсионеров, чтобы получить 200 наблюдений, относящихся к редкой подсовокупности тех, кто старше 85.

Другой, более дешевый, метод заключается в непропорциональной стратификации, т. е. в непропорциональном отборе из различных подсовокупностей. Нередко возникает необходимость сделать «распространенные» и «редкие» страты равно представленными в выборке. Если вернуться к обсуждавшемуся выше примеру исследования городского населения, можно, в частности, представит; ситуацию, когда необходимо сравнить кулинарные предпочтения русских и украинцев. Очевидно, не вполне корректно сравнивать 800 русских и 100 украинцев. В этом случае можно прибегнуть к непропорциональному систематическому отбору из названных страт: если отбирать каждого 200-го русского и каждого 25-го украинца, мы получим две вполне сопоставимые, равные по объему, — 400 и 400 человек — подвыборки (однако эти равные подвыборки будут непропорционально репрезентировать доли соответствующих подсовокупностей, в чем можно убедиться, самостоятельно произведя подсчеты по описанным выше формулам).

Выбор между пропорциональной и непропорциональной стратификацией исследователь осуществляет, исходя из содержательных и экономических соображений. Нужно, однако, иметь в виду некоторые «послевыборочные» последствия непропорционального отбора, с которыми социологи сталкиваются на стадии анализа. В частности, для получения более точных оценок распределения исследуемых переменных иногда приходится применять так называемое взвешивание (иногда употребляют термин «перевзвешивание»). Взвешивание используют также для того, чтобы исключить влияние некоторых типов систематического смещения в основе выборки и других типов систематической ошибки измерения. Например, взвешивание полезно для исключения смещений, возникающих из-за дублирования в списке-основе или, наоборот, из-за наличия систематических «пропусков» для какой-то одной группы (скажем, если в списке пропущено много пожилых людей, постоянно проживающих с детьми, но прописанных по другому адресу). Так как необходимость взвешивания чаще всего вызвана нарушением исходных соотношений, пропорций между входящими в целевую совокупность группами, мы опишем общую идею этой процедуры на примере непропорционального стратифицированного отбора.

Напомним, что к непропорциональной стратифицированной выборке прибегают в тех случаях, когда точность оценок для выборки в целом или для отдельных подгрупп (субпопуляций) внутри выборки оказывается недостаточной. В этом случае доли генеральной совокупности (f) будут различны для разных страт. Последнее утверждение равносильно признанию разной вероятности попадания в выборку для единиц, принадлежащих к разным стратам. Как совместить неравные вероятности отбора с данным нами выше определением вероятностной (случайной) выборки, в котором подчеркивалось равенство шансов попадания в выборку для всех входящих в генеральную совокупность единиц-«случаев»? Некоторые статистики считают предложенное нами выше определение не вполне точным и предпочитают говорить о вероятностной выборке как о выборке, где каждая единица отбора имеет «известную, ненулевую вероятность быть включенной в выборку», хотя шансы для различных единиц не обязательно равны. Существующее многообразие определений вероятностной выборки восходит к давней дискуссии о правомерности выводов, основанных на априорных («до») и апостериорных («после испытания») вероятностях. Мы, однако, сохраним наше определение случайной выборки, внеся в него некоторое уточнение: когда шансы попадания в выборку неравны, как при непропорциональном отборе из страт, они могут быть выровнены при помощи взвешивания на стадии анализа, т.е. на собственно послевыборочной стадии исследования (конечно, если отбор внутри страт сохраняет свой случайный и равновероятный характер). Для этого нужно внести определенные поправки в полученные данные, а именно — приписать некоторым наблюдениям (классам наблюдений) больший «вес», компенсирующий меньшие шансы попадания в выборку (и наоборот).

Результатом приписывания веса каждому наблюдению является увеличение точности оценок для исследуемых параметров. Вес каждой единицы (респондента) в k-й страте равен отношению числа таких элементов в генеральной совокупности к объему выборки для k-й страты, т.е.:



При расчете среднего или других параметров каждое наблюдавшееся значение просто умножается на весовой коэффициент «своей» страты.

В частности, среднее значение какого-то параметра совокупности (например, средний доход или среднее количество хронических заболеваний) будет равняться просто взвешенной сумме средних значений для отдельных страт:



Формула расчета стандартной ошибки для стратифицированной выборки также включает в себя весовые коэффициенты, w:



Стандартные компьютерные программы, используемые при статистическом анализе данных, всегда содержат элементарные процедуры взвешивания.

Вернемся к нашему примеру с непропорциональным стратифицированным отбoром русского и украинского населения. Предположим, мы выяснили, что в среднем каждая украинская семья заготавливает на зиму 50 кг варенья, тогда как среднее значение для русской страты составило 40 кг.

Для украинской страты весовой коэффициент составит:

wукр. = 10000 : 400 = 25.

Соответственно для русского населения:

wрусск. = 80000 : 400 = 200.

С учетом этих весовых коэффициентов уточненная оценка среднего запаса варенья в выборке составит:

х = 25 • 50 • 400 + 200 • 40 • 400 /100000 = 37 кг.

Если бы мы не учли в своих расчетах сверхпредставительность украинцев в нашей непропорциональной стратифицированной выборке, то оценка среднего запаса варенья для всей совокупности оказалась бы завышенной (45 кг).

Четвертый тип вероятностной выборки, используемой социологами, — это кластерная выборка. «Кластеры» (дословно с англ. — гроздья) — это естественные группировки единиц наблюдения. Например, популяция избирателей имеет тенденцию жить в городах и деревнях, генеральная совокупность военнослужащих естественным образом группируется по воинским частям и подразделениям, а совокупность студентов — по университетам, институтам и колледжам. Способность к образованию локальных группировок, которую обнаруживают генеральные совокупности, изучаемые социологами, при соблюдении ряда условий позволяет уменьшить расходы на получение единицы информации.

Цель использования кластерной выборки таким образом заключается в повышении эффективности затрат на проведение исследования. При фиксированном бюджете и объеме выборки социолог получает возможность снизить общие расходы на проведение личных интервью преимущественно за счет уменьшения транспортных расходов.

В общем случае кластерная выборка основана на первоначальном отборе группировок (кластеров) и затем — на изучении всех единиц внутри кластеров. Возможными примерами кластеров, используемых в больших общенациональных опросах, являются сельские районы, городские квартиры, избирательные участки. При изучении специфических популяций используются иные кластеры: больницы — при изучении пациентов, школы — при изучении школьников и т. п.

Корректное применение кластерной процедуры основано на неукоснительном соблюдении четырех необходимых условий:

1) кластеры должны быть однозначно и явно заданы: каждый член генеральной совокупности должен принадлежать к одному (и только одному) кластеру;

2) число членов генеральной совокупности, входящих в каждый кластер, должно быть известно или поддаваться оценке с приемлемой степенью точности;

3) кластеры должны быть не слишком велики и географически компактны, иначе кластерная выборка теряет всякий финансовый смысл;

4) выбор кластеров должен быть осуществлен таким способом, который минимизирует рост выборочной ошибки (последний процесс, в свою очередь, является неизбежным следствием кластеризации).

Для того чтобы уяснить, как именно кластерная процедура влияет на рост выборочной ошибки, рассмотрим ее на простейшем примере. Допустим, мы изучаем труд и занятость жителей небольшого сельского района. Для того чтобы составить полный список-основу для случайной выборки, нам пришлось бы предварительно посетить все сельские советы, а в некоторых случаях — и весьма отдаленные деревни. Располагая ограниченными ресурсами, мы решаем использовать имеющуюся в нашем распоряжении карту района, на которой отмечены все населенные пункты, включая самые небольшие хутора. Известна и численность населения для каждого пункта. Естественными границами кластеров-поселений являются шоссе и проселочные дороги. Составив список всех 40 деревень и хуторов, мы можем теперь без труда осуществить простую случайную выборку кластеров. Для отдельного поселения вероятность попадания в выборку составит 1/40. Если, например, мы собираемся опросить 200 человек, нам, скорее всего, потребуется отобрать 1—2 кластера-поселения. Отметим здесь, что естественные различия в величине кластеров никак не влияют на процедуру кластерного отбора.

Что при этом происходит с выборочной ошибкой и, следовательно, с получаемыми в нашем исследовании статистическими параметрами генеральной совокупности сельского населения района (т. е. с оценками возраста, дохода и т. п.)? Чтобы ответить на этот вопрос, мы должны ввести еще одно статистическое понятие «независимых наблюдений» (степеней свободы).

Предположим, мы хотим оценить соотношение работающих и пенсионеров в обследуемом нами районе. Мы отобрали, условно, три деревни по 30 домовладений каждая (итого 90 домовладений). Однако в ходе опроса выясняется, что в двух деревнях, не входящих ни в одно сельхозобъединение или кооператив, живут исключительно старики-пенсионеры, а в одной, построенной недавно для переселенцев из Средней Азии, живут только молодые семьи с детьми. Таким образом, каждая деревня является населенной либо только работающими семейными парами, либо исключительно «пенсионерской». В результате мы можем заранее предсказать результат обследования каждой деревни (кластера), посетив лишь один дом. Если в первом доме интервьюер обнаружит чету пенсионеров, во всех остальных домах тоже будут жить пенсионеры. Если в первом доме живут люди трудоспособного возраста, посещение остальных 29 домовладений приведет к тому же результату. Фактически для каждой деревни мы будем располагать одним независимым наблюдением и, посетив 90 семей в трех деревнях, получим лишь три независимых, информативных наблюдения относительно распределения работающих и пенсионеров в выборке. Соответственно наши оценки величины данного соотношения в генеральной совокупности окажутся более неточными, чем в случае 90 независимых наблюдений. Причина возникающей ошибки заключается в том, что использованные вами кластеры (деревни) оказались гомогенными, однородными по исследуемому признаку трудовой занятости, хотя по другим признакам, например, по политической активности, они вполне могут быть гетерогенными, неоднородными. В принципе можно показать, что рост выборочной ошибки для кластерной выборки (в сравнении с простой случайной) является функцией двух нерешенных — величины кластеров и гомогенности исследуемого признака внутри каждого кластера.

Ясно, что оценка гомогенности часто становится важной практической задачей в планировании кластерной выборки. Основная проблема здесь заключается в том, что соответствующими данными о распределении признаков внутри кластеров исследователь располагает после завершения собственно полевой стадии. Практически при проектировании выборки обычно основываются на уже существующих данных предыдущих исследований, переписей и т. п.

Таблица 2

Значения мер гомогенности р для кластеров, состоящих из домовладений (для основных социально-демографических параметров)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Параметр | Значение р для кластера, имеющего средний размер п | | | |
| п = 3 | п = 9 | n = 27 | n = 62 |
| Доля домовладений: — находящихся в личной собственности; | ,170 | ,171 | ,161 | ,096 |
| — наемных, с низкой квартплатой; | ,235 | ,169 | ,107 | ,062 |
| — наемных, с высокой квартплатой; | ,430 | ,349 | ,243 | ,112 |
| Среднее количество жильцов | ,230 | ,186 | ,142 | ,066 |
| Доля среди жильцов: |  |  |  |  |
| — белых мужчин | ,100 | ,088 | ,077 | ,058 |
| — безработных мужчин | ,060 | ,070 | ,045 | ,034 |
| — мужчин в возрасте 25—34 лет | ,045 | ,026 | ,018 | ,008 |

Мера гомогенности р ведет себя так же, как соответствующий коэффициент корреляции. Величина р — это корреляция между значениями признака для всех возможных парных сочетаний элементов, входящих в кластер. Эта величина обычно положительна и возрастает с ростом гомогенности элементов внутри кластера. Если наблюдения внутри кластера абсолютно независимы (как в примере случайного распределения между разными кластерами), то р = 0. При использовании территориальной кластерной выборки городского населения, например при отборе кварталов или многоэтажных домов, р для признаков экономического статуса может быть весьма высоким из-за «пороговых» эффектов: в престижном кооперативном доме маловероятно встретить семьи с очень низкими доходами (верхний порог) и, наоборот, лишь немногие состоятельные люди обитают в коммуналках, подобно герою «Золотого теленка» Александру Ивановичу Корейко (нижний порог).

Ориентировочное представление о типичных значениях р и их изменении для кластеров разной величины для общенационального выборочного исследования дает табл. 2. В таблице показаны величины р для имеющих разные размеры кластеров, составленных из соседних городских домовладений (квартир и домов). Данные таблицы основаны на выборке городского населения США (N> 100000).

Еще одной немаловажной практической проблемой в планировании кластерной либо стратифицированной выборки является сравнение эффективности затрат на исследование при разных среднем размере кластера и количестве кластеров (заметим, что и кластеры, и страты часто обозначают общим термином — «первичные единицы отбора»). Функция, описывающая зависимость расходов от вышеперечисленных двух переменных, выглядит так:

Сt = ас1 + пс2,

где Ct — общая стоимость исследования,

а — количество «первичных единиц отбора»,

с1 — средние затраты на обследование первичной единицы отбора, планируемые для данного исследования,

n — общий размер планируемой выборки,

с2 — средние затраты на проведение одного интервью.

Дальнейшим обобщением идей случайного отбора из субпопуляций и естественных группировок, лежащих в основе, соответственно стратифицированной и кластерной выборок, является многофазная (многоступенчатая) выборка. Построение такой выборки представляет собой довольно сложную статистическую задачу, подходы к решению которой мы рассмотрим лишь в самом обобщенном виде.

В простейшем случае многофазная выборка состоит из двух фаз случайного отбора. На первой — как при кластерном отборе — выбираются «первичные единицы отбора», например, районы, избирательные участки, предприятия. На второй фазе производится случайный отбор единичных членов генеральной совокупности — отдельных респондентов, семей и т. п. Так как «первичные единицы отбора» могут существенно отличаться по величине (как, например, отличаются друг от друга городские квартиры или дома с разной численностью проживающих), то результатом первой фазы может стать неравная вероятность попадания в выборку для членов генеральной совокупности, относящихся к разным «первичным единицам отбора». В этом случае исследователь имеет возможность выравнивания вероятностей на последующих фазах (например, из «первичной единицы отбора», где проживает 1000 семей, он выберет 10, а из «первичной единицы», где живет 500 семей, будет отобрано 20).

Рассмотрим многофазную процедуру на простейшем примере с равной вероятностью отбора.

Пусть нам необходимо осуществить выборку размером 2000 человек из генеральной совокупности населения крупного города, где проживает 4 млн. человек. Каждая «первичная единица отбора» — городской квартал — содержит 1000 единиц (т. е. отдельных респондентов). На первой фазе мы отберем из 100000 кварталов («первичных единиц отбора») 400, так что для каждого квартала вероятность попадания в выборку составит:

400:100000 = 0,004.

На следующей стадии из 1000 жителей каждого квартала мы отберем 50, так что для каждого респондента суммарная накопленная вероятность попадания в двухфазную выборку составит:

0,004 X (50:1000) = 0,0002.

Решение об использовании многофазной выборки обычно принимается после анализа «баланса» затрат и приобретений. Снижение затрат на сбор данных. достигаемое в этом случае, сопровождается увеличением сложности выборочной процедуры. С ростом числа фаз (в больших общенациональных обследованиях нередко используют 4 или 5 «ступенек» отбора — от области до квартала) точность получаемых оценок имеет тенденцию снижаться. Поэтому исследователям нередко приходится сочетать многофазный отбор со стратификацией на завершающих стадиях выборочной процедуры, что обычно ведет к улучшению характеристик выборки. Отсюда понятно, почему многофазная выборка в значительной мере остается «прерогативой» крупных исследовательских организаций, которые обладают значительными финансовыми ресурсами и могут воспользоваться услугами профессионалов-статистиков при проектировании выборки.

# 3 Размер вероятностной выборки

Вопрос об оптимальном размере вероятностной выборки всегда был спорным и, в значительной мере, остается таковым. Мы обсудим лишь основные принципы, лежащие в основе современного подхода к оптимизации размера выборки.

Решение относительно размера выборки принимают с учетом целого ряда факторов, среди которых самую существенную роль играют два: 1) ценность и новизна получаемой в результате опроса информации и 2) затраты на проведение опроса (включая временные) при заданном размере выборки.

Некоторые исследователи полагают, что принятие решения о размере выборки может основываться на сугубо статистическом подходе. При этом в расчет принимают допустимую величину ошибки в оценке исследуемого параметра (например, дохода). Существуют статистические формулы, связывающие размер выборки с вероятностью ошибки и величиной доверительного интервала, задающего пределы этой ошибки. Так как использование этих формул требует принятия определенных предположений о том, как распределена интересующая исследователя величина, возникает необходимость в предварительной информации, относящейся к тому самому параметру, который мы решили изучить. Трудности, возникающие при использовании классического статистического подхода к определению размера вероятностной выборки, можно описать одной фразой, принадлежащей известному специалисту по массовым опросам С. Судману: «Очевидно, что формула, описывающая зависимость размера выборки от предполагаемой ширины доверительного интервала и приемлемой вероятности ошибки, попросту заменяет проблему определения размера выборки другой, не менее трудной проблемой — определения ширины доверительного интервала».

Во многих важных случаях можно руководствоваться сложившейся практикой, т.е. размером выборки, использовавшейся в аналогичных исследованиях. Кроме того, нужно помнить о простейших «правилах левой руки» для определения размера выборки.

Размер выборки растет

— при необходимости опубликовать данные для отдельных подгрупп (размеры подвыборок при этом суммируются, и выборка в целом растет пропорционально числу подгрупп);

— при проведении общенациональных обследований, когда велика генеральная совокупность (заданная доля генеральной совокупности/будет определять тем больший объем выборки, чем больше генеральная совокупность);

— если уже имеющаяся информация по ключевым вопросам (например, о намерениях избирателей голосовать за ту или иную партию) явно недостаточна, и степень неопределенности значительна

Размер выборки уменьшается

— при исследовании организаций, институтов и прочих «первичных единиц отбора», если сравнительно невелика величина генеральной совокупности, из которой производится отбор (например, совокупности сотрудников рекламных агентств, школьников, пациентов и т. п.);

— при проведении локальных и региональных исследований;

— если уже существующая информация относительно полна, и все еще остающаяся степень неопределенности незначительна.

«Типичные» размеры выборок для общенациональных опросов варьируют в пределах 1000—2500 респондентов (в зависимости от числа анализируемых подгрупп), для региональных опросов и опросов специальных популяций — от 200 до 500 (при анализе многочисленных подгрупп размер региональной или специальной выборки обычно возрастает как минимум до 1000 человек). Указанные значения, разумеется, могут служить лишь самым общим ориентиром для определения оптимального размера выборки.

# 4 Целевой отбор

социологическое исследование вероятностная выборка

Иногда социологи вынуждены применять не основанные на вероятностях выборки. Отбор в этом случае базируется не на принципе рандомизации, а на следовании тем или иным субъективным критериям — доступности, типичности, равного представительства и т. п. Многие из этих критериев при систематическом использовании позволяют добиться достаточно высокого качества социологических данных. Часто такой отбор называют целевым, так как он в большой степени определяется целями исследования. Кроме того, в конкретной исследовательской ситуации может оказаться, что осуществление случайной выборки — это практически невыполнимое или экономически неэффективное мероприятие (затраты на построение выборки превышают ценность получаемой в результате исследования информации). Наконец, использование вероятностного отбора лишено всякого смысла, если речь идет об исследовании уникальных событий, групп или ситуаций — полетов на Луну, войн или любовных историй.

Основной недостаток неслучайных процедур отбора связан с тем, что не существует строгих статистических методов, позволяющих обобщить результаты, полученные в ходе исследования выборки. Оценка точности и валидности этих результатов (и основанных на них выводов) остается делом субъективного суждения, опыта, теоретических предпочтений.

Самый распространенный тип не основанной на вероятности выборки — это выборка доступных случаев. Такого рода выборка может считаться корректной лишь тогда, когда используется в экспериментальном (или квазиэкспериментальном) исследовании. Так, в большинстве психологических экспериментов испытуемыми являются студенты. Это позволяет экономить скудные финансовые ресурсы, отпускаемые на сугубо академические изыскания. Для того чтобы исключить влияние посторонних, смешивающих факторов, экспериментатор в случайном порядке распределяет выборку доступных случаев (т. е. доступных испытуемых) по двум группам — экспериментальной и контрольной. В нашем обсуждении роли рандомизации в эксперименте подчеркивалось ее значение для получения точных и обоснованных выводов. Однако случайное приписывание испытуемых-добровольцев к экспериментальной и контрольной группам, строго говоря, не является достаточным основанием для обобщения результатов эксперимента для всей генеральной совокупности, из которой осуществлялась выборка доступных случаев. Точнее, в ситуации отбора доступных случаев невозможно с полной уверенностью сказать, что, собственно, являлось генеральной совокупностью в процессе исследования, так как последняя не была определена с самого начала. Поэтому, в частности, шутливое определение предмета психологии гласит, что это наука, изучающая студентов-второкурсников гуманитарных факультетов. В социологии выборкой доступных случаев чаще всего приходится довольствоваться при изучении таких специальных популяций, которые практически не поддаются локализации. Речь идет, прежде всего, об относительно малочисленных группах, находящихся вне сферы институционального (например, административного) контроля. Для таких групп трудно найти какую-то основу выборки — скажем, посетители стрелковых тиров едва ли состоят на каком-нибудь государственном учете. «Просеивание» большой случайной выборки из генеральной совокупности с целью рекрутирования сколько-нибудь значительного числа респондентов в специальную выборку требует непомерных затрат. Поэтому социологам иногда приходится уподобляться орнитологам и отбирать членов экзотических популяций в местах их «естественного обитания» или вероятного скопления. Многие исследования посетителей массовых библиотек проводятся в библиотеках, посетителей выставок — в музеях, ветеранов войны — в клубах ветеранов и т. п. В этой ситуации исследователю приходится прилагать дополнительные усилия для получения высококачественной информации. Следует заметить, что некоторая статистическая «небезупречность» получаемых таким образом результатов, при должной методической культуре исследователей, иногда окупается, и мы узнаем нечто принципиально новое об относительно «закрытых» областях человеческого поведения. Однако если целью исследования является описание распределения признаков во вполне определенной генеральной совокупности (покупателей зубной пасты, избирателей, читателей газет), то социолог, использующий выборку доступных случаев, понапрасну тратит деньги заказчика (и пренебрегает профессиональной этикой). Квалифицированному заказчику в этом случае также не стоит принимать всерьез рассуждения о принципиально новых, нестатистических и даже «мягких» методах проведения массовых опросов.

Значительно реже социологи используют две другие разновидности целевого отбора — отбор «критических случаев» и отбор «типичных случаев». В обоих случаях исследователь полагается на какие-то теоретические представления или предыдущий опыт, чтобы отобрать ограниченное число «симптоматических», характерных наблюдений, позволяющих сделать более широкие обобщения и предсказания. Иногда это удается, но следует помнить о том, что опыт и теоретические суждения обычно бывают субъективны. В печально знаменитых президентских выборах 1948 г. в Америке (Г. Трумэн против Т. Дьюи) ошибочные прогнозы сделали все знаменитые институты опросов общественного мнения. При этом некоторые из них избрали в качестве «типичного» случая население штата Мэн, так как прежде жители этого штата всегда «угадывали» будущего президента. В описываемом случае «нетипично» (т.е. за проигравшего выборы Дьюи) проголосовали только два штата — Мэн и Вермонт. Поэтому поговорку «Как голосует Мэн, голосует вся Америка» пришлось перефразировать: «Как голосует Мэн, так голосует Вермонт».

Метод «снежного кома» — это еще один (наряду с выборкой доступных случаев) интересный подход к отбору из «редких» совокупностей. Его идея такова: первоначально идентифицированная небольшая группа членов интересующей социолога совокупности служит источником сведений о других членах этой совокупности, так что выборка постепенно разрастается вширь подобно снежному кому, катящемуся с горы. Этот метод использовал, например, П. Лазарсфельд с коллегами в исследовании «влиятельных людей» и неформальных связей. Помимо властвующих элит данный метод применяют в изучении других групп, также избегающих широкой известности, — например, наркоманов или коллекционеров антиквариата. Для этого метода существуют определенные приемы оценки систематической ошибки, однако они слишком сложны, чтобы обсуждаться здесь.

К выборкам, не основанным на случайном отборе, относится и квотная выборка, когда-то чрезвычайно популярная даже среди профессиональных статистиков и практически не используемая сейчас. Идея квотной выборки проста: изучаемая совокупность разбивается на такие социально-демографические группы, которые исследователь почему-либо считает важными. Обычно критериями разбивки становятся пол, возраст, национальная принадлежность, место жительства и т. п. Далее, основываясь на уже известных (обычно из официальной статистики) пропорциях этих групп в генеральной совокупности, социолог составляет полевые задания для интервьюеров, указывая, сколько женщин, мужчин, лиц с высшим образованием и т. п. нужно опросить. Например, интервьюер получает задание опросить десять женщин старше 50 лет, восемь мужчин 35 — 45 лет и трех восемнадцатилетних девушек, проживающих в г. Санкт-Петербурге. В результате должна получиться выборка, представляющая все заданные пропорции групп в генеральной совокупности.

Основная проблема квотного отбора заключается в том, что он носит неслучайный характер и осуществляется лично интервьюером. Последний выбирает респондентов, в конечном счете, по собственному усмотрению. Хотя число мужчин или женщин, рабочих или пенсионеров, которых следует опросить в данном районе или местности, задано заранее, интервьюер решает, в какую квартиру ему удобнее позвонить, с кем из членов семьи провести интервью, куда вернуться вторично, если на звонок никто не ответил, и т. п. Это неизбежно ведет к систематическим смещениям в процессе отбора, причем не существует никаких методов для оценки величины возникающей систематической ошибки.

Еще один очевидный недостаток квотного отбора связан с тем, что обычно невозможно даже приблизительно оценить количество отказов от участия в опросе. Если интервьюер сталкивается с человеком, не желающим отвечать на вопросы, или просто недоброжелательным, или вызывающим у него антипатию, интервьюер всегда волен попрощаться и попытать счастья в соседней квартире.

По указанным причинам квотные выборки «вышли из моды» среди социологов, несмотря на свою относительную дешевизну.

Оценивая полезность и применимость вышеописанных «неслучайных» методов отбора в исследовательской практике, следует, прежде всего, сказать, что в определенных обстоятельствах никакой другой альтернативы просто не существует. В ситуации нехватки денег, персонала, времени либо первичной информации о генеральной совокупности социологи использовали и будут использовать впредь выборки доступных случаев, метод «снежного кома» и даже (к сожалению) квотную выборку. При этом профессиональный долг социолога заключается в том, чтобы оценить, пусть даже очень приблизительно, величину и источники возникающей выборочной ошибки.

Безусловно, разумно использовать целевые выборки в пилотажных исследованиях, в экспериментах, в том числе методических (т. е. нацеленных на проверку и отработку анкет, опросников, шкал и т. п.).

Однако всегда следует помнить о том, что возможность обобщения любых оценок, полученных на целевой выборке, для генеральной совокупности в целом, т. е. внешняя валидность результатов исследования, чаще всего оказывается сомнительна.

**Библиографический список**

1. Кокрен У. Методы выборочного обследования. М.: Статистика, 2010.

2. Петренко Е. С., Ярошенко Т. М. Социально-демографические показатели в социологических исследованиях. М.: Статистика, 2009.

3. Территориальная выборка в социологических исследованиях. М.: Наука, 2010.

4. Чурилов Н. Н. Проектирование выборочного социального исследования. Киев: Наукова думка, 2008.

5. Маслова О. М. Качественная и количественная социология: методология и методы (по материалам круглого стола) // Социология: 4М. 2009. № 5—6.