

ПРОГРАММНЫЕ ТРЕБОВАНИЯ К ВЫБОРКЕ

(Ядов В.А. Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности. М.: «Добросвет», 1999. С. 111-122)

В подавляющем большинстве случаев социолог использует тот или иной способ выделения из большой совокупности явлений и объектов изучения некоторую их часть в надежде, что на этой выборочной совокупности могут быть выявлены свойства объекта исследования в целом. (В отличие от идеализированного объекта исследования как определенной области социальной реальности, содержащей предмет изучения, здесь имеется в виду эмпирический объект: конкретные индивиды, группы, организации, регионы, локализованные во времени и пространстве).

Тип и способы выборки прямо зависят от целей исследования и его гипотез. Чем конкретнее цель и чем яснее сформулированы гипотезы, тем правильнее будет решен вопрос о выборке.

Наиболее строгие требования предъявляются к выборкам дескриптивных и аналитико-экспериментальных исследований, наименее строгие — к исследованиям по разведывательному плану. В последнем случае отбор «единиц наблюдения» на объекте подчиняется довольно простым правилам: следует выделять полярные группы по существенным для анализа критериям. Численность таких несистематических выборок строго не определяется. Все зависит от состояния получаемой информации.

Наблюдение или опрос в таком исследовании продолжаются до тех пор, пока не обнаружится, что получена информация, достаточно разнообразная для формулировки гипотез. Следовательно, состав и объем выборки заранее не фиксируются, а устанавливаются опытным путем по мере развития исследования.

В исследовании дескриптивного плана выборка, напротив, должна быть строго репрезентативной.

Требования репрезентативности выборки означают, что по выделенным параметрам (критериям) состав обследуемых должен приближаться к соответствующим пропорциям в генеральной совокупности. Между тем строго репрезентативную выборку по всем важным для проблематики исследования параметрам обеспечить невозможно, и поэтому следует гарантировать репрезентацию по главному направлению анализа данных.

Прежде всего, надо уяснить, какие из имеющихся сведений о характеристиках генеральной совокупности существенны для целей исследования. Во многих случаях это половозрастной, социально-профессиональный, имущественный состав обследуемых, их пространственная локализация. Половозрастная структура «замыкает» на себя многие показатели семейного состояния, уже известные по другим данным. Возраст содержит указания на жизненный опыт и, как правило, на рабочий или профессиональный стаж. Социально-профессиональные, социально-статусные характеристики — это

свидетельства о различиях в системе реального положения людей и их особых интересах, позиций. Пространственная локализация (по территории, подразделениям предприятий и учреждений, по другим административным и производственным «локалам») важна и с точки зрения особенностей условий этой деятельности (например, центр и периферия, основные и вспомогательные службы), и с точки зрения адресности итоговых выводов и рекомендаций, которые должны быть «привязаны» к административным или производственным ячейкам, имеющим четкие границы и часто самоуправляемым. В сочетании трех названных параметров — половозрастной структуры, социального состава, пространственной локализации — можно, как правило, быть уверенным, что выборка будет представительна для анализа многих социальных проблем. Понятно, что это правило имеет исключения в зависимости от конкретных условий и особых целей исследования (например, в этнически неоднородной среде существенно иметь в виду репрезентацию по критерию национальной принадлежности). Мера подобия выборочной модели структуре генеральной совокупности оценивается ошибкой выборки, а пределы допустимой ошибки опять-таки зависят от цели исследования.

Иногда требуется повышенная надежность, как это имеет место в экономических и демографических обследованиях, например при переписях населения. Здесь существенные ошибки оборачиваются миллионными потерями материальных ресурсов и просчетами планирования.

Примечание. Аккуратная репрезентативная территориальная выборка в современной России требует систематических коррекций. Это связано с тем, что территориальные административные границы (именно они являются основами официальной статистики населения) формировались для целей, не совпадающих с социально-исследовательскими и помимо того состав и структура населения в периоды реформации неустойчивы. Как показывает М. С. Косолапов [Косолапов М.С. Принципы построения многоступенчатой вероятностной выборки для субъектов РФ // Социологические исследования. — 1997. - №10.], достоверные расчеты общенациональной и региональных представительных выборок — сложная исследовательская задача. Экономная выборка предполагает также аккуратные расчеты маршрутов исследователей (интервьюеров и др.), что составляет особую проблему в данном регионе.

Гораздо чаще социологические обследования проводятся для уяснения общих тенденций, общей ориентировки в сфере социальной политики.

Весьма полезна следующая приблизительная оценка надежности результатов выборочного обследования [Экспертные оценки в социологических исследованиях / Ред. Крымский С.Б., Жилин Б.Б., Паниотто В.И. Киев: Наукова думка, 1990. С. 36].

Повышенная надежность допускает ошибку выборки до 3%, обыкновенная — до 3—10% (доверительный интервал распределений на уровне 0,03—0,1), приближенная — от 10 до 20%, ориентировочная — от 20 до 40%, а прикидочная — более 40%.

В аналитических и экспериментальных исследованиях проблема статистической репрезентативности выборки оказывается второстепенной в сравнении с необходимостью обеспечить качественное представительство изучаемых социальных объектов.

В.А. Ядов рассматривает следующий пример. В изучении образа жизни населения

некоторого города мы, следуя правилам дескриптивного обследования, хотим обеспечить представительство всех групп населения соответственно их пропорциям в составе генеральной совокупности с отклонением $\pm 5\%$ от истинного распределения. Такая выборка, представительная в качественном отношении, будет также и статистически репрезентативной, но следует решить, нужно ли это.

Напомним, что репрезентативные выборки необходимы лишь в том случае, если целью исследования является получение суммарных данных в отношении изучаемого объекта в целом. В нашем примере — это все население данного города. Тогда в выводах социолог имеет право сообщить, что в среднем горожане так-то оценивают различные условия жизни и деятельности, в среднем такая-то доля населения проявляет высокую активность в таких-то видах деятельности, а такая-то — низкую и т. п. Но с практической точки зрения, не говоря уже о теоретических задачах изучения образа жизни, нам гораздо важнее выявить специфику условий и образа жизни различных групп населения и в том числе тех, которые, будучи малочисленными, нуждаются в специальном внимании.

Допустим, что в составе населения города имеется 370 ветеранов Отечественной войны. Чтобы получить более или менее достоверную информацию об условиях их жизни и их проблемах, надо обеспечить должное численное представительство этой категории граждан в выборочной совокупности. Но поскольку выборка статистически репрезентативна, то при численности населения города, скажем, в 100 тыс. и численности выборочной совокупности в 2 тыс., т. е. при двухпроцентной выборке, доля ветеранов в выборочной совокупности составит 60 человек. Много это или мало? Возможно, этой численности достаточно для того, чтобы сделать статистически достоверные заключения о простейших частных показателях условий их жизни, например об уровне обеспеченности жилищем ветеранов войны, в сравнении со среднестатистическими показателями на всю выборку населения города. Но как только мы захотим углубить анализ, мы обнаружим, что численность подвыборки ветеранов явно мала. К примеру, важно установить, какова доля ветеранов войны, проживающих в отдельной квартире и без семьи, т. е. одиноких. В таком случае придется составить табличку размерностью 2×2 (две градации "проживают с семьей" и "одиночки" + две градации по критерию наличия своей комнаты или квартиры). В каждой клеточке этой таблицы может быть в пределе по 15 единиц наблюдения ($60:4=15$). Конечно, реальное распределение окажется иным. Так, ветеранов-одиночек, не имеющих собственной комнаты, не будет вовсе. Зато одиночек, проживающих в отдельной квартире, может оказаться, допустим, 5—10 человек. Вместе с тем именно эта категория ветеранов и составляет предмет особого внимания. Однако при численности подвыборки в 10 человек никакой дальнейший статистический анализ уже невозможен.

Следовательно, если мы хотим изучить в статистических показателях особенности условий и образа жизни каких-то определенных групп населения, репрезентативная выборка должна быть заменена целевой, в которой численность каждой интересующей нас группы будет достаточна для более основательного анализа. Такая выборка, будучи качественно представительной в отношении целей исследования, не является статистически репрезентативной в отношении генеральной совокупности.

Во многих случаях необходимы именно целевые выборки. (Иногда целевую выборку называют "социологической", в ней обеспечивается представительство по признакам, выявленным в предыдущих социологических исследованиях, а для реализации таких

выборки могут использоваться таксономические процедуры [см. Жабский М.И. Обоснование Репрезентативности социологического исследования // Социологические исследования. – 1982. - №2.; Кочанов Л. Алгоритм построения представительной территориальной выборки при наличии ограничений на стоимость измерений. М., 1992. Т.1 №1.].

Особенно это важно в исследованиях экспери–ментального плана. Скажем, проверяется эффективность введения новой формы организации труда. Ясно, что для этого следует отобрать подразделения, где введена новая организация, и для сравнения — аналогичные, где работа идет по старому. Следует гарантировать в выборке равную численность экспериментальных подразделений или организаций и "контрольных", работающих по прежней системе. При этом важно так подобрать эти подразделения, чтобы они были аналогичны по всем существенным характеристикам, кроме факта наличия или отсутствия новой формы организации труда. Формы собственности, профессиональный и квалифицированный состав работников, их половозрастная структура и, возможно, другие показатели должны быть сопоставимы. Решающее значение имеет здесь отнюдь не пропорциональность выборочной доли экспериментальных подразделений фирмы или предприятия в отношении к их доле в генеральной совокупности, но именно качественное представительство экспериментальных и контрольных объектов соответственно цели исследования. Численность (объем) выборки зависит от уровня однородности или разнородности изучаемых объектов. Чем более они однородны, тем меньшая численность может обеспечить статистически достоверные выводы. Но степень однородности социального объекта зависит, в сущности, от того, насколько детально мы намерены его исследовать. Практически любой, самый "элементарный" объект оказывается чрезвычайно сложным. Лишь в анализе мы представляем его как относительно простой, выделяя те или иные его свойства. Чем более основательным и детальным будет анализ, чем больше свойств данного объекта мы намерены принять во внимание в их сочетании, а не изолированно, тем больше должен быть объем выборки.

Для решения такого рода задач как раз и необходимы целевые аналитические выборки. В них учитывается не только структура изучаемой совокупности, но и ограничения, накладываемые на объем выборки целями исследования, глубиной анализа проблем. Используя статистический критерий Стьюдента, можно рассчитать объем выборок в зависимости от заданного уровня доверительного интервала ошибки вывода [Руковишников В.О., Паниотто В.И., Чурилов Н.Н. Опросы населения. М.: Финансы и статистика, 1984. С. 19—21], Чем меньше объем сравниваемых подвыборок (пусть это будут ветераны-одиночки и семейные), тем больше должно быть различие каждой пары сопоставляемых статистик (например, процентные различия оценок условий быта теми и другими). Если численность сравниваемых подвыборок неодинакова, за базу определения допустимой ошибки следует брать наименьшую подвыборку. В зависимости от объема подвыборки существен–ность процентных различий определяется таблицей:

Объем подвыборок по их численности	Значимая разность	Объем подвыборок по их численности
Значимая разность		

Программные требования к выборке

Добавил(а) Социология
17.01.11 12:00 -

в % при ошибке

в % при ошибке

не более 5 %

не более 5%

50 8	20	300
100 6,3	14	500
150 4,5	11,5	1000
200 2	10	5000

Допустим, что удовлетворительно оценивают условия быта 85% ветеранов-женщин и 79% мужчин, проживающих с семьями, и соответственно 32% женщин и 38% мужчин-одиночек. Разности в процентах составляют здесь: $85 - 79 = 6$ и $42 - 38,4 = 3,6\%$. При численности подвыборки до 150 человек и при 5-процентном уровне ошибки эти различия нельзя признать существенными, так как они должны перекрывать 11,5%. Но различия между соответствующими оценками одиночек и семейных будут существенны. Они составят для женщин $85 - 32 = 53\%$ и $79 - 38 = 41\%$ для мужчин. Такие различия значимы уже при выборках около 50 человек. Достоверный вывод звучит так: решающей является ситуация проживания ветеранов с семьей или одиноким. В какой мере эти обстоятельства больше переживаются мужчинами или женщинами, сказать трудно; наших данных для этого недостаточно.

Авторы приведенных расчетов отмечают, что выборки на уровне 500 человек позволяют анализировать таблицы сопряженности с 4 признаками из трех градаций каждый, а выборки в 1000 единиц расширяют возможности уверенного анализа до таблиц с 6 признаками из пяти градаций. Все это при условии обеспечения доверительного интервала, не превышающего 5% статистически значимой ошибки.

Общее правило таково: объем выборки при заданном уровне доверительного интервала должен быть не менее чем nK единиц наблюдения, где n — объем подвыборки по столбцу, а K — число столбцов.

Объем выборки зависит также от уровня доверительного интервала допустимой ошибки, каковая, как уже говорилось, задается целесообразной точностью итоговых обобщений: от повышенной до ориентировочной. Однако здесь имеются в виду так называемые случайные ошибки, связанные с природой любых статистических погрешностей. Именно они и вычисляются как ошибки репрезентативности

В. И. Паниотто приводит следующие расчеты репрезентативной выборки с допущением 5-процентной ошибки [Паниотто В.И., Максименко В.С. Количественные методы в социологических исследованиях. Киев: Наукова думка, 1982. С. 81].

Объем выборки 222 286

333 350 360 370 385 398

Для совокупности более 100000 выборка составляет 400 единиц. Если же иметь в виду генеральные совокупности численностью от 5 тыс. и больше, то, по расчетам того же автора, можно указать величины фактической ошибки выборки в зависимости от ее объема [Паниотто В.И. Качество социологической информации. Киев: Наукова думка, 1986. С. 82], что для нас весьма важно, памятуя, что величина допустимой ошибки зависит от цели исследования и необязательно должна приближаться к 5-процентному уровню.

Фактическая ошибка при данном объеме выборки, % 20 15 10 9 8 7 5 4

Наряду со случайными возможны ошибки систематического характера. Они зависят от организации выборочного обследования. Это разнообразные смещения выборки в сторону одного из полюсов выборочного параметра.

Объем выборки определяется аналитическими задачами исследования, а ее репрезентативность — целевой установкой программы. Именно программа задает образ необходимой генеральной совокупности для проведения выборки. Будет ли это все население или особые его структурные образования, все элементы изучаемого объекта или только выделяемые по заданным программой критериям.

Генеральную совокупность составляют все единицы определенного в программе объекта. Теперь следует обеспечить равную их вероятность попадания в выборочную совокупность.

При небольших по численности генеральных совокупностях применяют случайную бесповторную выборку, где обеспечивают равную вероятность попадания в исследование всех ее единиц по полному их списку из генеральной совокупности. Имея полный список работников предприятия (например, 2000 человек) и определив объем выборочной совокупности (например, в 200 человек), устанавливаем шаг выборки делением первого на второе ($2000:200$) и получаем шаг отбора — каждый 10-й из списка. Здесь важно не допустить систематической ошибки из-за отсутствия в списке, скажем, какого-то подразделения, например работающих в филиале предприятия.

При больших генеральных совокупностях, как это имеет место в опросах населения, используют многоступенчатый отбор по районам, т. е. крупным структурным составляющим генеральной совокупности: регионы, типы поселений, кварталы города.

На каждой ступени отбора следует обеспечить требования представительности населения, т. е. обоснованно отобрать регионы так, чтобы не было смещения по какому-то важному параметру (например, по этнонациональному). То же самое и на последующих ступенях отбора. В конечном счете отбор производится опять-таки систематически с установленным шагом отбора по списку граждан (из списков избирателей или иных), списку хозяйств на селе, путем посещения каждой, скажем, 20-й квартиры в списке квартир каждого 50-го дома выделенного квартала города. Многие обстоятельства усложняют проблему расчета ошибки и нередко могут привести к тому, что формально-статистически репрезентативная выборка окажется качественно непредставительной.

Итак, качество выборки зависит от трех условий: (а) от меры однородности социальных объектов по наиболее существенным для исследования характеристикам; (б) от степени дробности группировок анализа, планируемых по задачам исследования; (в) от целесообразного уровня надежности выводов из предпринимаемого исследования.

Очень часто малоопытный социолог не улавливает разницы между проблемой ошибки репрезентативности выборки и ошибки вывода из данного конкретного распределения в рамках выборочной совокупности.

Пусть выборка, достаточно репрезентативна и ошибка по тому или иному параметру выборки незначительна. Оценка уровня достоверности вывода по каждому конкретному распределению остается при этом проблемой самостоятельного анализа. (Приемы расчета разнообразных ошибок вывода рассматривает К. В. Кемниц, который подчеркивает, что формально-статистические методы расчета ошибок вывода должны предваряться «инженерным» (т. е. содержательным.) изучением распределения [Кемниц Ю.В. Математическая обработка зависимых результатов измерений. М., 1970. С. 4].

Несколько заключительных замечаний. Из сказанного выше может показаться, что обеспечить представительство данных в выборочном обследовании если и удастся, то ценой непомерных усилий, разумность затрат которых часто сомнительна.

Рекомендуется, во-первых, не отчаиваться и, во-вторых, рассуждать здраво, имея в виду программные цели исследования.

Если перед нами стоит задача выполнить дескриптивное обследование большой общественной значимости, в итоге которого должны быть сделаны заключения относительно генеральной совокупности в целом, следует, конечно, максимально реализовать все требования репрезентативной выборочной процедуры. Затраченные усилия будут не только оправданы, они просто необходимы, так как ошибки в выводах такого исследования недопустимы. Здесь ложная информация опаснее ее отсутствия (достаточно сослаться на ошибки прогнозов исхода выборов вследствие ошибок выборки опросов электората или ошибки в исследованиях рынка).

Если же задачи исследования более скромные, уровень надежности планируемых выводов с точки зрения их статистической точности можно смело понизить, но надо принять все меры к качественному представительству выборочной совокупности.

Преувеличенное внимание к формально-статистическим критериям достоверности выводов (и тем более их абсолютизация) за счет качества исходной информации и качества анализа — свидетельство профессиональной неопытности социолога.

Подчеркивая статистическую надежность данных, он вводит в заблуждение и себя, и,

хуже того, тех, кто привык верить в убедительность математических расчетов. Нельзя забывать о реальной природе того, что кроется за цифрами и математическими формулами. Ведь сами исходные характеристики, получаемые исследователем путем опросов или другими способами, лишь условно переводятся в количественные показатели. Часто эти количественные сведения весьма приблизительно отражают существо социальных процессов. Поэтому усилия, направленные на строгость статистического обоснования результатов, приобретают смысл только при условии серьезного качественного анализа проблемы, содержательного ее изучения. Бывает и так, что непредставительные в статистическом смысле данные, многократно повторяемые на разных подвыборках, как раз свидетельствуют об определенной социальной тенденции лучше, чем статистически достоверный вывод, сделанный на одной единственной выборке.

Следует постоянно помнить, что социолог призван сосредоточить внимание именно на существе социальных проблем, активно привлекать к постановке задач исследования других специалистов, практиков и теоретиков, внимательно следить за литературой по широкому кругу вопросов, относящихся к предмету исследования. Наконец, для решения собственно статистических задач, касающихся типа и объема выборки, он прежде всего обязан максимально четко сформулировать конкретные вопросы, подлежащие решению, и уже после этого обращаться к соответствующим расчетам разнообразных статистик.