

Ф. Н. ИЛЬЯСОВ

ИЛЬЯСОВ Фархад Назипович - кандидат философских наук.

Аннотация. В статье вводится понятие "репрезентативность результатов опроса", показывается специфика определения генеральной совокупности в маркетинге, обосновывается и иллюстрируется на конкретном примере эмпирический подход к определению репрезентативности результатов опроса и объёма выборки.

Ключевые слова: репрезентативность выборки, репрезентативность результатов опроса, реконструкция генеральной совокупности, распределение ответов

В социологии и маркетинге, как представляется, произошла, возможно невольная, мистификация понятия репрезентативности. В отчётах о проводимых опросах нередко даются некоторые характеристики репрезентативности. В одних случаях ссылаются на гипотезу о соответствии результатов опроса закону нормального распределения. В других случаях говорят о степени соответствия выборки социально-демографической структуре генеральной совокупности. В третьем варианте просто пишут примерно следующее: "погрешность данных исследования не превышает 3,3%" или "допустимая статистическая погрешность - 2,8%". Что имеется в виду в последних случаях - не расшифровывается, и не совсем понятно. Не исключено, это просто фигура речи, призванная создавать позитивный образ исследовательской фирмы в глазах потребителей результатов опросов. На самом деле выражение "погрешность не превышает", судя по всему, всего лишь метафора, а репрезентативность на практике достигается опытом построения "правильных" выборок.

М. С. Сваффорд, М. С. Косолапов, П. М. Козырева отмечали: "Многие организации, проводящие социологические обследования, просто ограничиваются заявлениями о том, что их выборка репрезентативна. Даже худшие из этих организаций настаивают на репрезентативности своей выборки. Но если задать им вопрос, что имеется в виду под этой "репрезентативностью", они ответят, что это коммерческая тайна или выдадут краткое описание, которое столь неполно, что специалист по выборке просто не сможет его оценить" [1]. Таким образом, можно сказать, что репрезентативность одна из самых таинственных и мистифицированных тем в социологии и маркетинге, хотя формально солидная литература по теории выборки существует - от статей и диссертаций, до серьёзных и признанных монографий (см., например: [2]).

Обычно в материалах, где обсуждается репрезентативность, речь идёт о репрезентативности выборки. В самом общем описании статистическая репрезентативность выборки¹ - это её свойство правильно отражать **исследуемые**² характеристики генеральной совокупности.

В оценке репрезентативности выборки можно условно выделить следующие формальные критерии: а) отбор респондентов производится из целевой группы (например, потребители определённого продукта), б) выборка соответствует социально-демографическим характеристикам генеральной совокупности (если они известны заранее).

¹ Иногда в литературе, даже научной, используется не совсем корректное выражение "репрезентативность исследования". Репрезентативность это характеристика эмпирической базы - выборки массового опроса, а не исследования. В исследовании может и не быть массового опроса, а может быть несколько массовых опросов.

² Требование отражать "все свойства" генеральной совокупности представляется избыточным и даже недостижимым.

Строго говоря, указанные критерии являются необходимыми, но недостаточными, так как прямо не связаны с характером распределения ответов на содержательные вопросы анкеты. Однако ситуация с этими критериями репрезентативности выборки представляется достаточно понятной. Другим критерием является размер выборки - количество респондентов должно быть достаточным для того, чтобы правильно отразить характер распределения признаков в генеральной совокупности.

В ряде случаев декларируется, что репрезентативность выборки определяется исходя из гипотезы о нормальном распределении ответов на вопрос, здесь используются построения и допущения, взятые из теории математической статистики [3]. Важно отметить, что к результатам конкретного массового опроса это не имеет прямого отношения, потому в социологии подобный подход нельзя признать в достаточной мере корректным. Статистическая теория не всегда приложима к социологии.

Значительное число вопросов, используемых в массовых опросах, имеет номинальный уровень измерения. Репрезентативность распределения ответов на подобные вопросы анкеты остаётся, как правило, неизвестной, даже гипотетически.

При анализе шкал с высоким уровнем измерения (интервальных и т.д.), использование гипотезы о нормальном распределении ответов для оценки репрезентативности ответов во многом остаётся под вопросом. Например, распределение по возрасту в популяции не подчиняется нормальному распределению - с увеличением возраста количество респондентов естественным образом уменьшается. Распределения по экономическим показателям - по доходу, зарплате, цене покупки и т.д., имеют правоскошенный характер, то есть, не симметричны по своей природе, а стало быть, измерение репрезентативности ответов на подобные вопросы на основании гипотезы нормального распределения имеет некоторую систематическую погрешность.

Репрезентативность выборки, определяемая на основании гипотезы о нормальном распределении, строго говоря, является всего лишь гипотезой. Как отмечал А. А. Давыдов: "...встречающиеся утверждения, что выборка в целом репрезентативна, вообще говоря, неправомерны. ...репрезентативность нельзя "вычислить" чисто логическими средствами. Репрезентативность - это всегда гипотеза, которая может быть лучше или хуже обоснована, но истинность которой может быть установлена только в результате эмпирического исследования" [4].

В маркетинге большей частью генеральная совокупность вообще неизвестна, ее характеристики - тем более. Кроме того, именно в рамках маркетинговых исследований нередко возникает задача выявления размеров, характера и структуры генеральной совокупности. О. Б. Божков называет подобную реконструкцию генеральной совокупности "обратная задача выборки" [5].

В коммерческой социологии и маркетинге произошёл отход от использования строго вероятностных выборок, хотя этот факт особо и не афишируется. "В реальных исследованиях, - указывает А. Чуриков, - применяется даже больше неслучайных выборок, чем случайных" [6]. Поскольку указанное явление носит явно массовый и долговременный характер, его нельзя назвать немотивированным. Помимо того, что, как указывалось ранее, теория математической статистики имеет относительное отношение к социологии, надо сказать, она сложно применима к реальной практике построения выборок.

О. Б. Божков отмечает: "репрезентативность не всегда обеспечивается методами построения выборки, разработанными в математической статистике" [5]. В маркетинге, в том числе, как показывают консультации, в полевых отделах крупных опросных центров, похоже, преобладает практика не случайных выборок. Можно предположить, что один из наиболее типичных алгоритмов выборки в маркетинге состоит из следующих ступеней: (1) не случайный ("интуитивный") отбор: типичный регион -> типичный населённый пункт -> типичный район в населённом пункте; (2) метод маршрутного опроса в типичном районе населённого пункта; (3) не случайный отбор респондента в домохозяйстве по заданным поведенческим критериям и социально-демографическим квотам.

Практика подтверждает обоснованность такого подхода, например, по критериям повторяемости результатов при повторных опросах с одним дизайном выборки, более того, схожесть результатов разных опросных фирм, правильные прогнозы, например, результатов выборов. А "полевики" сохраняют по этому поводу "коммерческую тайну", должно быть, ввиду того простого соображения, что у них нет "теоретического обоснования", которое можно было бы "продать". То есть признать, что они, создавая вы-

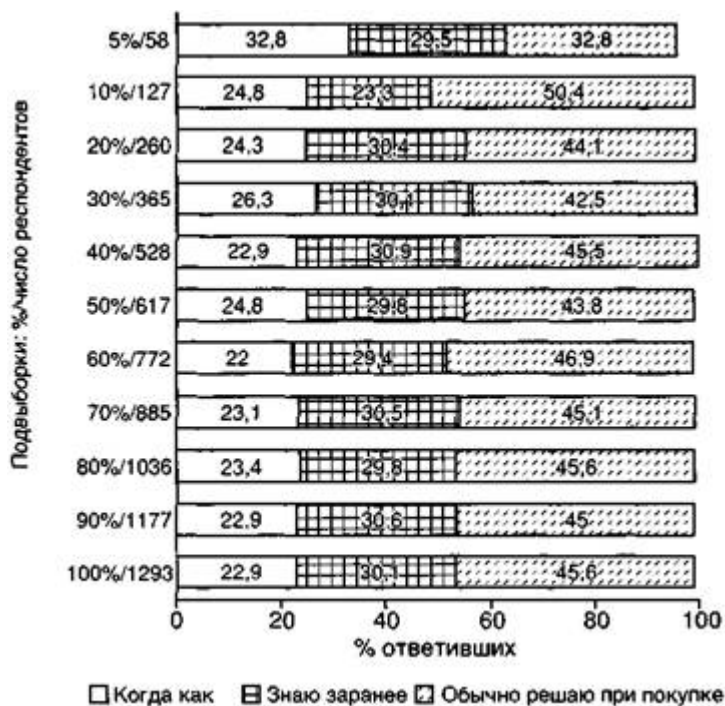


Рис. 1. Характер распределения ответов по подвыборкам разного объема

борки, "действуют на ощупь, без методологии, и не знают, как измерить репрезентативность результатов" - это уронить себя в глазах заказчика. Эмпирия, видимо, здесь опережает теорию - "практически делают, а объяснить не могут". Возникает парадоксальная ситуация, можно сказать; репрезентативность результатов опроса есть, а теоретического обоснования репрезентативности нет. То есть репрезентативность здесь выступает в форме своеобразного ноумена, не имеющего "эмпирического объяснения".

Важно ещё раз отметить, исследователя интересует, строго говоря, не репрезентативность выборки, а *репрезентативность результатов опроса*. В этом смысле - репрезентативность это не один (единственный) показатель, отражающий свойство результатов всего опроса в целом, а характер распределения ответов на вопросы, включенные в инструментарий исследования. Это ситуация, когда совпадает распределение ответов на отдельный вопрос в выборочной и генеральной совокупностях. Степень совпадения и является уровнем *репрезентативности ответов на конкретный вопрос*. В одном и том же опросе она может быть различной относительно ответов на разные вопросы. В этой связи имеет смысл под *уровнем репрезентативности результатов опроса* понимать диапазон колебания репрезентативности ответов на отдельные вопросы. Если при анализе результатов опроса стоит задача удостовериться не в репрезентативности выборки, а в действительной репрезентативности распределения ответов на определённый вопрос, то для этого можно использовать анализ изменения распределений ответов, при изменении объема подвыборок, извлекаемых из совокупности методом "извлечение случайной подвыборки". Тогда мы имеем дело не с гипотезой, а с эмпирически установленным фактом. Понятно, идея подобного подхода не нова, но нынче её легче реализовывать с использованием современных программ статистической обработки и анализа.

Предлагаемый подход иллюстрируется ниже на примере результатов случайно отобранного нами массового опроса³. На рис. 1 и далее приведены распределения

³ Опрос проводился методом очного стандартизированного интервью в марте 2006 г. в г. Москве ($n = 1293$). Генеральная совокупность - покупатели котлет-полуфабрикатов г. Моск вы. Использовалась двухступенчатая выборка, на первом шаге отбирались 9 типичных районов, на втором - случайный отбор респондентов "на улице" (с использованием определённого шага выборки - например, опрос каждого седьмого прохожего), при отсеке "не покупающих" и квотировании ("добору") по возрасту.

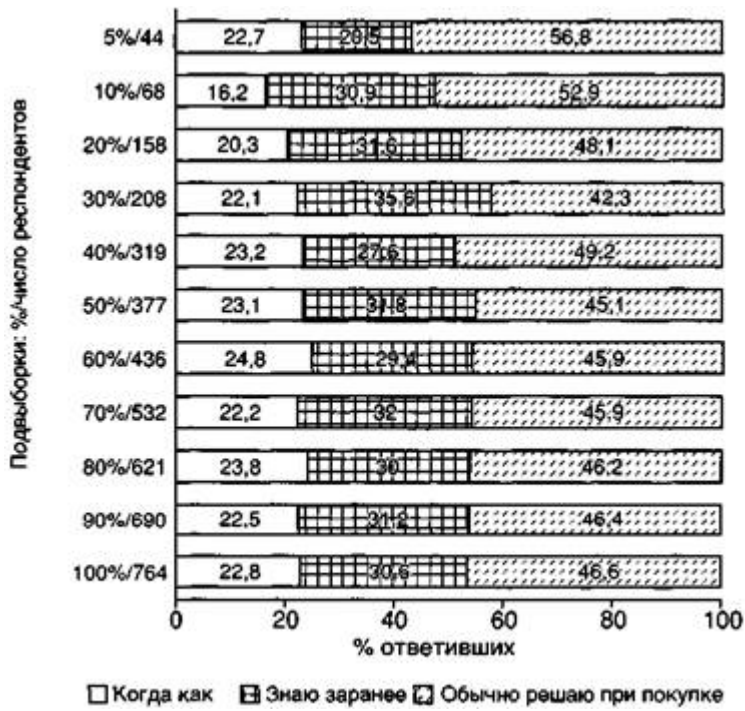


Рис. 2. Распределение ответов среди женщин для разных по объему подвыборок

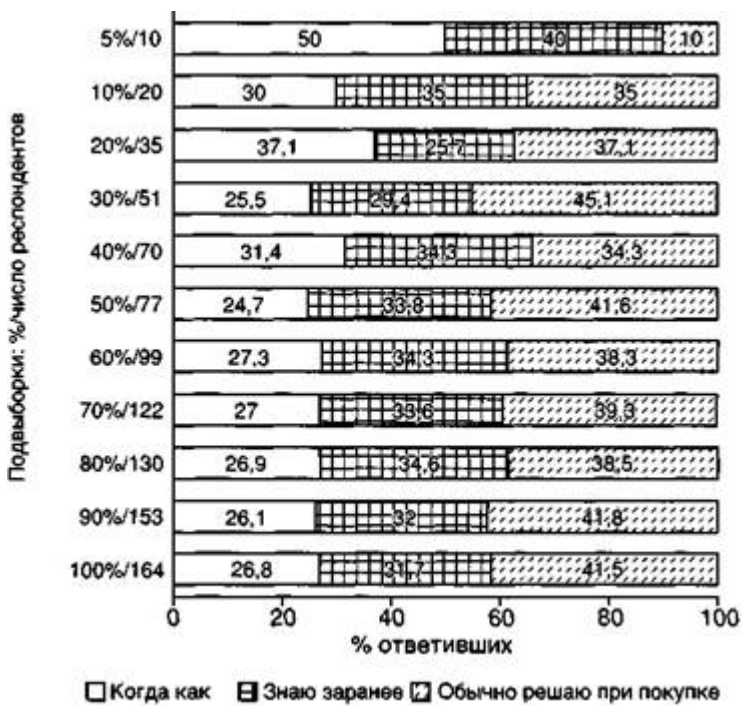


Рис. 3. Распределение ответов среди женщин 31 - 30 лет для разных по объему подвыборок

ответов на вопрос: "Когда Вы решаете купить котлеты-полуфабрикаты, вы всегда заранее знаете, какую марку купите, или чаще решаете в момент покупки?" (с тремя вариантами ответов). Как видно из приведённых данных, при размере выборки 772 респондента (подвыборка 60%) происходит стабилизация распределения ответов, - оно существенно не меняется по мере увеличения объема подвыборки. Следовательно, можно полагать, что для этого вопроса, при использовании того же алгоритма отбора респондентов, для анализа одномерного распределения, репрезентативной является выборка в 772 респондента. Как известно для изучения более гомогенной генеральной совокупности требуются выборки меньшего размера. Рассмотрим, как может уменьшаться размер репрезентативной выборки при увеличении гомогенности группы. Рис. 2 иллюстрирует, что при увеличении объема подвыборки в группе "женщины" до 377 респондентов (подвыборка 50%), происходит стабилизация распределения ответов. Следовательно, для анализируемого нами вопроса, при использовании того же дизайна выборки, репрезентативной является выборка в 377 женщин.

Для группы "женщины 31 - 40 лет", как это видно на рис. 3, стабилизация распределения ответов наступает при величине выборке в 77 респондентов (подвыборка 50%). Следовательно, для данного вопроса, при использовании того же дизайна выборки, для группы "женщин 31 - 40 лет", репрезентативной является выборка в 77 респондентов.

Описанный выше алгоритм анализа может использоваться для определения объёма репрезентативной выборки при проведении последующих опросов на указанную тему, при применении того же дизайна выборки и использовании того же вопроса. Корректно измерить в % ошибку репрезентативности распределения ответов на конкретный вопрос, основываясь на результатах опроса, без знания распределения в генеральной совокупности, невозможно (а если исследователю известно распределение в генеральной совокупности, то проведение опроса теряет смысл). Если же исходить из описанного выше эмпирического критерия репрезентативности, то можно определять, репрезентативны или не репрезентативны выборки, на основании самих результатов опроса, без знания распределения в генеральной совокупности. Например, характеристика репрезентативности рассматриваемого выше примера может описываться так: "Выборка для вопроса "х" репрезентативна для одномерного распределения, при размере случайной подвыборки равной 60%". Если указанная величина подвыборки, при сравнении с распределениями ответов в случайных подвыборках на другие вопросы анкеты, максимальна для данного опроса, то она может быть характеристикой всего опроса.

СПИСОК ЛИТЕРАТУРЫ

1. *Сваффорд М. С., Косолапов М. С., Козырева П. М.* Международные стандарты оценки качества социологических обследований // Мир России. 1999. N 1 - 2. С. 285 - 286.
2. *Йейтс Ф.* Выборочный метод в переписях и обследованиях. М.: Статистика, 1976. *Якушева А. А.* Формирование выборки: поиск новых возможностей // Социол. исслед. 2007. N8. *Тулбаева А. А.* Репрезентативность выборки в исследовании социальных объектов. Автореф. дис. ... канд. социол. наук. Екатеринбург, 2010.
3. Рабочая книга социолога. М.: Наука, 1977. С. 257 - 297.
4. *Давыдов А. А.* Репрезентативность выборки // Социол. исслед. 1990. N 1. С. 117.
5. *Божков О. В.* Эта неуловимая генеральная совокупность // Социол. исслед. 1987. N 3.
6. *Чуриков А.* Случайные и неслучайные выборки в социологических исследованиях // Социальная реальность. 2007. N 4. С. 89.