

А.А. Вейхер

Национальный  
исследовательский  
университет «Высшая школа  
экономики»  
(Санкт-Петербург)

# РАЗВИТИЕ МЕТОДОВ ОЦЕНКИ ВАЛИДНОСТИ ДАННЫХ КОЛИЧЕСТВЕННЫХ СОЦИАЛЬНО- ЭКОНОМИЧЕСКИХ ОПРОСОВ

---

## Проблемные аспекты валидности социально-экономических опросов

Социально-экономические опросы продолжают расширять свою нишу в сфере информационного обеспечения. За 65 лет (с появления Consumer Sentiment Index, а затем Consumer Confidence Indicator (Michigan, USA)) результаты множества опросов населения, потребителей, экспертов различного типа, менеджеров и руководителей предприятий и организаций стали высокочисленными для принятия социальных и экономических решений. Они находятся в одном ряду с индикаторами, получаемыми статистикой из документов (акты гражданского состояния (демография), регистры различных экономико-правовых действий (индексы бирж, таможенные обороты, регистрируемые сделки), отчеты и декларации граждан и хозяйствующих субъектов (налоги) и т.п.). Ныне в аналитических отчетах, в сводных индексах (например, в Leading Indicator Composite Index [OECD, 2012]) опросные и документарные индикаторы-показатели применяются на равных.

Между тем они отличаются принципами оценивания валидности [Вейхер, 2008]. Данные из регистров и иных документарных источников — результаты объективированной фиксации фактов действий или состояния вещей. Их валидность сути общественных явлений или процессов, подлежащих измерению, определяется четкостью критериев их квалификации (для включения в круг наблюдаемых фактов) и соблюдением процедур фиксации (статистического наблюдения), которые могут нарушаться при наличии интересантов их фальсификации (неполный учет всего круга фактов, искажения их отнесения к той или иной градации шкалы измерения и т.п.). Довольно часто эти факты подвергаются сплошному наблюдению, получаемые

значения показателей измеряются метрическими шкалами, погрешности сбора данных признаются несущественными, и во всех расчетах используются абсолютные значения показателей.

Данные социально-экономических опросов практически всегда собираются при помощи управляемой выборочной процедуры. Валидность получаемых показателей имеет вероятностный характер. (Это в полной мере относится и к данным из документарных источников, когда они обследуются выборочно, но для опросных данных это свойственно всегда.) Такой тип переменных ставит проблему их включения в расчеты любых индексов: индекс как сводная величина совокупности индикаторов тоже должен быть переменной с указанием доверительного интервала ее точности при определенной надежности.

В данном случае нет возможности рассмотреть математическую сторону дела, связанную с формированием вероятностных параметров сводной величины как результата цепных и аддитивных арифметических операций с совокупностью переменных, имеющих вероятностный характер. Отметим только, что в имеющейся практике есть ряд противоречий, без понимания которых интерпретация получаемых результатов может быть неверной, дающей пищу для обоснованной критики этих результатов.

Так, в публикациях мониторинговых замеров CSI в США (и в других странах) значения этих индексов приводятся в процентах с точностью до десятичного знака, и в комментариях интерпретируются изменения уровня индекса, даже если они составляют несколько десятых процента [Short, 2012]. В тех же публикациях часто указывается, что для рядового трендового опроса выборка составляет 500 респондентов. Такому объему выборки, согласно объяснениям методики сбора данных [Curtin, 1999], соответствует доверительный интервал  $\pm 5\%$  (для значения доли 50%). Совсем иной оказывается интерпретация тех же результатов по графику значений, сглаженному скользящей средней по четырем точкам [Ibid.]. Нам не удалось пока обнаружить разработок по этим вопросам для методологии получения и использования опросной информации социально-экономической направленности.

Охарактеризованные проблемы относятся к внутренней валидности методик рассматриваемого типа опросов. По нашему мнению, еще более острыми являются проблемы внешней валидации их результатов.

## **Место внешней валидности в опросной практике**

За 80 лет после введения Терстоуном понятия валидности (валидации) [Thurstone, 1931] были предложены более десятка вариантов этого тер-

мина и несколько методов оценивания валидности, так как в широком смысле характеристика валидности присуща любому исследованию [Winter, 2000; Никитина, 2009]. Для предмета нашей работы — валидности представительного опросного обследования — традиционно признаются очевидными два типа валидности: внутренняя и внешняя. Целью методов оценки внутренней валидности являются улучшение качества инструментов сбора данных и координирование всего процесса их создания и использования. Два способа оценивания внешней валидности (конкурентный и прогностический) фиксируют адекватность результатов генерализации целевой популяции по результатам обследования с другими (чаще более масштабными) выборками и генеральными совокупностями.

Роль критериев внешней валидности данных обследований отличается в различных сферах исследований и изменяется во времени. Известные ранние репрезентативные опросы избирателей (Дж. Гэллуп, А. Кросли, А. Кэнтрил, Э. Роупер) получили признание общественности, потому что они валидизировались простым и понятным индикатором прогностической валидности — результатом выборов.

В настоящее время методы внешней валидности редко используются вне электоральных опросов. Они очень скупо упоминаются в учебниках (см. один из типичных учебников [De Leeve, Нох, Dillman, 2008, р. 53], где есть только один параграф 3.8.4 «Comparison with External Data» в 18 строк). Внешняя валидность даже не обсуждается в общей статье девятнадцати ведущих специалистов по теме «Оценка погрешностей из-за отказов отвечать» [Groves et al., 2008].

Основная причина вышеописанной ситуации с внешней валидностью проста: если мы знаем величину или распределение конкретного индикатора в общей популяции, то в выборочном обследовании часто нет необходимости.

Однако внешняя валидизация сохранила свою роль в особой форме распространения различных методов «взвешивания» (weighting) полученных данных.

Традиционно в публикациях опросных обследований присутствует такая фраза: «Результаты репрезентативны по возрасту и полу», так как совпадение распределения случайной выборки по показателям возраста и пола с распределением по этим же показателям в генеральной совокупности рассматривается как подтверждение хорошего качества выборки. Однако добиться такого совпадения в большинстве опросов довольно несложно. Распространение же ситуаций отклонения пропорций распределений респондентов в выборочных совокупностях от известных из внешних источников пропорций в генеральных совокупностях по более широкому кругу параметров привело к появлению математически весьма серьезных методов «взвешивания». Эти методы широко и успешно применялись до тех пор, пока причины невы-

полнения правила равной вероятности попадания в выборку респондентов можно было считать случайными. Данная посылка переставала соответствовать реальности по мере того, как первостепенным становился фактор отказа потенциальных респондентов участвовать в опросе. В обзоре-введении специального тематического выпуска журнала «Public Opinion Quarterly» [Singer, 2006] прямо указывается: если отказ участвовать в опросе становится главной причиной невыполнения заданной выборки (ранее такой причиной была невозможность установить контакт с индивидом, попавшим в выборку, из-за его отсутствия), это разрушает основу представительности выборочного обследования. В результате затрудняется статистическая оценка погрешностей получаемых данных, их валидности описываемой реальности.

## **Причины сокращения возможностей применения методов коррекции выборки**

Растущие уровни отказов от участия в опросах могут быть объяснены, по моему мнению, последствием проникновения опросов в массовую культуру. По большому счету расширение применения разнообразных современных опросных технологий существенно повлияло, наряду с новыми технологиями распространения и предъявления информации, на восприятие большинством общественных субъектов — от рядовых граждан до лидеров общества — общественной жизни в различных ее проявлениях: от узнавания популярности разных сортов пива и звезд эстрады до оценок скрытно назревающих экономических катаклизмов и вероятного поведения разных групп населения на предстоящих выборах. По сути, произошедшие и происходящие изменения можно без чрезмерного пафоса назвать *качественно новым состоянием рефлексии обществом общественной жизни*.

Такая популярность опросов стала причиной устойчиво растущих трудностей их проведения, особенно за последние 20 лет. Как видно из данных по США, более 70% взрослого населения уже были респондентами в опросах, в том числе 45–50% — в течение последнего года [Kim, 2011, p. 181]. Около 40% американцев сказали, что в продолжение года отказывались от предложений принять участие в опросах [Ibid., p. 182], треть из них не верит в их результаты [Ibid., p. 178]. Доля респондентов, которые видят что-то интересное в участии в опросах, сократилась за период с 1995 по 2003–2006 г. с двух третей до одной трети [Ibid., p. 185].

Так же развивается ситуация в других странах. Более трети граждан России в 2004 г. ответили, что им приходилось участвовать в опросах [Исследовательская группа ЦИРКОН, 2005]. 40% жителей России рассматривают опросы

как чисто коммерческую деятельность, а 28% согласны, что результаты опросов часто фабрикуются и публикуются для оказания давления на граждан; в больших городах мнение последних широко распространено [Исследовательская группа ЦИРКОН, 2005]. По нашим данным, в ходе избирательных кампаний в Санкт-Петербурге один из троих взрослых получает предложение участвовать в опросе. Опросы по некоторым темам (например, электоральные, маркетинговые) стали частью субкультур различных социальных групп. Накапливание опыта участия в опросах и знакомство с их результатами, публикуемыми в СМИ, заметно влияют на готовность принимать в них участие. Поэтому в настоящее время сокращение числа респондентов является причиной систематических погрешностей в выборочных распределениях ряда социальных групп.

При «уровне отклика» ниже 70% воздействие погрешностей из-за отказа отвечать делается неустранимым [Groves, Couper, 1998]. В опросах, проводимых в 2005–2010 гг., при числе отказов отвечать, превышающих 25% плановой выборки, любая попытка корректировать выборку перевзвешиванием по полу и возрасту приводила к занижению среднего числа детей у женщин в возрасте 30–45 лет. Соответственно информация о социальных проблемах, связанных, например, с детьми, неизбежно искажалась.

Расширяется круг исследований социальных групп, о которых нет информации как о генеральной совокупности. К таковым относятся фрилансеры; люди, которые придерживаются новых политических взглядов; малые предприниматели, не регистрирующие свой бизнес. Численность этих групп невелика, и выявление их путем представительного опроса населения дает мало информации. В лучшем случае удается узнать, что они составляют 1–2% населения. Только из очень большого числа общей выборки можно получить данные для расчетов состава таких малых подгрупп. Высокие затраты на получение столь малой информации делают представительное исследование таких групп практически нереальным.

## **Практика использования критериев внешней валидности**

Критерий внешней валидности в явной форме, а не только в форме «взвешивания» по известным стратам, на деле используется много лет.

В качестве примера можно назвать уже упоминавшийся известнейший проект Consumer Sentiment Index (индекс потребительских настроений (ИПН), начат в 1946 г.). Этот индекс, по сути, является латентной интегрирующей переменной, отражающей экономические настроения граждан США. Его использование основывается на предположении, что их чувства будут

проявляться в массовом поведении (покупки, нахождение и потеря работы и т.д.). Существует хорошая взаимосвязь между значениями индекса как в целом, так и по его составным частям с объективными индикаторами экономики США. В этом случае прогностическая валидность используется как чисто статистическая, без совпадения содержательного значения, отражаемого переменными. Налицо модель «черного ящика», и связи компонент в формуле расчета индекса интерпретируются как реально слабые и в целом не имеют индикатора во внешних источниках.

Исследователи, работающие на этом проекте, тщательно изучают воздействие роста отказов от участия в опросах на получаемые результаты [Curtin, 2000]. Долговременная практика отслеживания динамики сильной корреляции ИПН и внешних индикаторов экономического развития позволит акторам утверждать, что прогнозная ценность индекса слабо зависит от колебаний факторов, нарушающих выборку. С 1999 г. размещенная на веб-сайте ИПН информация о методологии проверки валидности динамики индекса гласит: «иные факторы, кроме собственно выборочных, также влияют на точность этих (и других) результатов. Эти факторы могут включать эффекты от формулирования вопросов, способности респондентов формулировать свои ответы и мнения, отказов от участия в опросах и неполноты покрытия плановой выборкой генеральной совокупности. Стандартное измерение этих эффектов отсутствует, но их влияние должно проявляться при использовании данных этих и других опросов. Поскольку измеряемые эффекты представлены во всех опросах, отмеченные преимущества наблюдений серии опросов во времени проявляются в том, что невыборочные влияния остаются относительно постоянными во всех выборках, а потому их влияние на оценки во времени минимизируется» [Curtin, 1999].

## **Развитие методики внешней валидности — результаты применения**

Мы предлагаем применять подход «внешней валидности» в варианте «метод перекрестной валидации» как статистической модели с очевидными связями между параметрами и внешним увязанным с опросным индикатором. Кроме «целевых» (связанных с целью опроса) автор предлагает включать в опросник несколько других вопросов. Они не представляют существенного интереса, но могут быть использованы в расчетах индикаторов, сравнимых со статистическими данными из независимых источников.

Приведем пример. Качество выборки в опросах о теневой зарплате среди работающих в Санкт-Петербурге (Россия) оценивалось с помощью внеш-

него индикатора «Зарплата до вычета налогов (номинальная начисленная зарплата)» из данных официальной статистики по сравнению с опросным индикатором. Было проведено три опроса в 2004–2007 гг. Мы применяли трехступенчатую выборку: 1) стратифицированная по численности жителей 50 территорий города; 2) вероятностная для отбора домов на каждой из территорий; 3) квотная для отбора жителей в отобранных домах (по полу и семи возрастным группам). Численность опрошенных составляла от 500 до 800 человек.

Сравнимый индикатор был рассчитан по трем параметрам из опросных данных: полная сумма всех заработков респондента, доля тех, кто получал зарплату без надлежащего оформления, величина зарплаты без оформления. В опроснике было два вопроса: «Вы лично фактически на всех местах, где работали, независимо от оформления, сколько за последний месяц заработали (примерно)?» и «В наше время немало людей получают часть заработной платы, иногда и всю, без оформления. Вам за этот же месяц приходилось получать заработную плату таким образом?» (Отказы отвечать на оба вопроса не превышали 15%.) Для каждого респондента рассчитывалась «законная» зарплата как «полная зарплата — теневая зарплата». Полученные величины пересчитывались в номинальную (до уплаты налога) и в среднюю номинальную зарплату для всех работников.

Валидность оценивалась степенью различия средних величин номинальной зарплаты по данным опроса и официальной статистики. Во всех трех опросах погрешность (разность между статистической и «опросной» средними величинами номинальной зарплаты по Санкт-Петербургу) не превышала 3,7%. Такая близость полученных величин возможна, если и в выборочной, и в генеральной совокупностях распределения по трем расчетным параметрам достаточно близки. Тем самым близость величин по номинальной зарплате становится индикатором близости величин теневой зарплаты, полученной в результате исследования и имеющейся в генеральной совокупности.

## **Выводы**

В исследованиях, где необходимо определять количественные параметры популяции, все чаще приходится сталкиваться с трудностями в обеспечении репрезентативности выборки. Основной причиной этого является отказ большого числа респондентов (более 30%) от участия в опросе. В таких случаях репрезентативность по разным переменным может существенно отличаться.

Предлагается определять представительность результатов опросов для различных целевых индикаторов. В случае расчета таких индикаторов из

определенных переменных на опросных данных это может быть несложным вариантом структурных уравнений.

Представляется перспективной дальнейшая разработка методов внешней валидности получаемых опросных данных в направлении построения рядов индикаторов представительности выборки, отдельных для различных предметов изучения.

## Литература

*Вейхер А.* Опросные технологии получения экономической информации — методологические проблемы развития // Известия СПбУЭФ. 2008. № 2. С. 34–35.

Исследовательская группа ЦИРКОН, ВЦИОМ. Группа 7/89. 2005. Социологические исследования в России: осведомленность и отношение населения. ЦИРКОН, Россия, retrieved 12 April 2012 <http://www.789.ru/new/projects/050315.pdf>

*Никитина Т.* Теоретическая и эмпирическая интерпретация и операционализация понятия «валидность» // Социология: теория, методы, маркетинг. Киев, 2009. № 2. С. 113–129.

*Curtin R.* Survey of Consumers — Data File Documentation, The University of Michigan, 1999, US, retrieved 16 April 2012 URL: <http://www.sca.isr.umich.edu/documents.php?c=i>

*Curtin R., Presser S., Singer E.* The Effects of Response Rate Changes on the Index of Consumer Sentiment // The Public Opinion Quarterly. 2000. Vol. 64. No. 4. P. 413–428.

*De Leeve E., Hox J., Dillman D.* International Handbook of Survey Methodology, Lawrence Erlbaum Associates. New York, 2008.

*Groves R., Couper M.* Nonresponse in household interview surveys. New York: John Wiley, 1998.

*Groves R. et al.* Issues Facing the Field: Alternative Practical Measures of Representativeness of Survey Respondent Pools // Survey Practice. 2008. Vol. 10. Retrieved April 15, 2012. URL: <http://surveypractice.wordpress.com/2008/10/30/issues-facing-the-field/>

*Kim J., Gershenson C., Glaser P., Smith T.W.* Trends in surveys on surveys // Public Opinion Quarterly. 2011. Vol. 75. No. 1. P. 165–191.

*Singer E.* Nonresponse bias in household survey // Public Opinion Quarterly. Special Issue. 2006. Vol. 70. No. 5. P. 637–645.

*Short D.* Michigan Consumer Sentiment at 12-Month High, February 24, 2012, retrieved March, 15, 2012. URL: <http://www.advisorperspectives.com/dshort/updates/Michigan-Consumer-Sentiment-Index.php>

*Thurstone L.* The Reliability and Validity of the Test, Ann Arbor, MI, 1931.

OECD-Library, retrieved May 14, 2012. URL: [http://stats.oecd.org/Index.aspx?DatasetCode=MEI\\_CLI#](http://stats.oecd.org/Index.aspx?DatasetCode=MEI_CLI#)

*Winter G.* A Comparative Discussion of the Notion of ‘Validity’ in Qualitative and Quantitative Research // The Qualitative Report [On-line serial]. 2000. Vol. 4 (3/4). Retrieved April 1, 2012. URL: <http://www.nova.edu/ssss/QR/QR4-3/winter.html>