

DOI: 10.14515/monitoring.2015.3.03
УДК 303.62:004.738.5

Правильные ссылки на статью:

Федоровский А.М. Качество онлайн-опросов. Методы проверок // Мониторинг общественного мнения: экономические и социальные перемены. 2015. № 3. С. 28-35.

Fedorovskii A. M. Quality of Online Surveys. Methods of Control//Monitoring of Public Opinion: Economic and Social Changes. 2015. № 3. Pp. 28-35.

А.М. ФЕДОРОВСКИЙ КАЧЕСТВО ОНЛАЙН-ОПРОСОВ. МЕТОДЫ ПРОВЕРОК

КАЧЕСТВО ОНЛАЙН-ОПРОСОВ. МЕТОДЫ
ПРОВЕРОК

QUALITY OF ONLINE SURVEYS. METHODS OF
CONTROL

ФЕДОРОВСКИЙ Алексей Михайлович – директор по стратегическому планированию ООО «Онлайн Интервьюер». E-mail: alexey@onlineinterviewer.ru

FEDOROVSKII Aleksei Mikhailovich –Director of Strategic Planning, Online Interviewer Ltd. E-mail: alexey@onlineinterviewer.ru

Аннотация. Статья содержит актуальные рекомендации для проверки качества заполнения анкет, предназначенных для проведения онлайн-исследований. Предлагается система показателей для проверки массива данных, собранных методом CAWI. Формулируются перспективные направления работы по усилению качества онлайн-исследований как в маркетинговом, так и в научном сегменте.

Abstract. The paper presents recommendations concerning the quality control of the questionnaires in online surveys. A system of indicators to check data array collected using the CAWI method is developed. The scope of the future work to improve the quality of online surveys both for marketing and scientific purposes is proposed by the author.

Ключевые слова: онлайн-опрос, CAWI, проверки, качество, панели, онлайн-анкеты, самозаполнение

Keywords: online survey, CAWI, checks, quality, panels, online questionnaires, self-completion

Устойчивая тенденция последних лет в России – перевод максимального количества исследовательских проектов в цифровой формат. Основные игроки рынка для своих проектов выбирают планшеты или онлайн [Искакова, 2015], способствуя аналогичному переходу в цифровую эру и для своих подрядчиков [Ананьев, 2015].

Тема онлайн-исследований вписывается в глубокий методологический дискурс. Речь идет о самостоятельном подходе к сбору и анализу данных, специфика которого может влиять на результаты исследования, но остается пока недостаточно изученной [Мягков, 2008]. Дискуссии, посвященные сравнению данных, полученных с использованием методов личных интервью и телефонных опросов, представлены шире, чем вопросы, касающиеся метода сбора данных посредством самозаполнения анкет в интернете, также известном как CAWI

(computer-assisted web interviewing). В данной статье будет анализироваться именно этот метод.

Об активизации распространения CAWI в исследовательской сфере свидетельствует статистика Европейской организации маркетинговых исследований ESOMAR. С 2010 г. объем онлайн-анкетирования в России вырос в 3 раза – с 5% в 2010 г. до 14% в 2013 г. [Global Market Research, 2013].



Рисунок 1 – динамика методов сбора данных за 2010-2013 гг.

Увеличение объема сбора данных методом самозаполнения происходит на фоне снижения объема личных интервью (F2F). Подобная тенденция существует и в мировой практике, где доля CAWI вдвое превышает как личные интервью, так и телефонные опросы [Global Market Research, 2013].



Рисунок 2 – Объем методов сбора данных в мире в 2013 г.

На сегодняшний день исследовательские проекты, в которых сбор данных идет с онлайн-заполнением анкет, предполагают приглашение респондентов из специализированных онлайн-панелей. Именно о такого рода проектах пойдет речь ниже¹.

Проекты подобного типа делятся на две общие категории в зависимости от источника программирования электронной версии анкеты:

- с программированием исследовательским агентством-заказчиком;
- с программированием на стороне подрядчика.

Источник программирования – важная составляющая процесса поддержки качества данных. С одной стороны, самостоятельное программирование исследовательским агентством делает проект надежнее с точки зрения инструментов проверки качества заполнения анкет. С другой – практика показывает, что спектр возможностей и набор используемых для программирования инструментов ограничен. Это приводит к визуально и функционально не привлекательным анкетам, снижающим заинтересованность респондента в ее заполнении, приводящим к незавершенному заполнению. Для агентства это влечет риск не набрать нужный объем выборки и не уложиться в сроки.

По внутренней статистике компании «Онлайн Интервьюер», доля самостоятельного программирования анкет исследовательскими агентствами составляет около 60%. В 40% случаев агентство обращается к подрядчику. Конечно, оно может требовать определенного набора проверок качества заполнения анкет и от подрядчика, однако недобросовестный подрядчик может искусственно занижать возможности проверки качества для создания себе более комфортных условий сбора данных. При этом агентству трудно проконтролировать саму процедуру проверки и установить факт ее выполнения.

Методы проверки качества заполнения анкеты онлайн можно разделить на два больших блока:

- процедуры проверки во время заполнения;
- процедуры проверки при обработке полученных данных.

Процедурам проверки ответов респондентов во время заполнения анкет посвящено много исследований [Медведева, 2015; Puleston, Eggers, 2012; Phillips, 2012]. В них анализируется также влияние процедуры на результаты исследований. Особое внимание уделяется проблеме «удержания респондента» в опросе, т.е. получения от него завершенной анкеты.

Обобщим имеющиеся подходы к проверке ответов **во время заполнения анкеты**.

- 1 время заполнения анкеты (выбраковка анкет со слишком коротким временем заполнения);
- 2 время заполнения определенных блоков или отдельных вопросов анкеты (превышение установленного минимума для ответа на вопрос или блок вопросов);

¹ Помимо проектов с участием респондентов-панелистов реализуются проекты с приглашением респондентов, не зарегистрированных в специализированных панелях, являющихся пользователями тематических интернет-площадок (так называемые *river sampling*), и проекты с использованием клиентских контактных баз. У этих типов проектов также есть своя специфика.

- 3 соответствие ответов на логически связанные вопросы в разных частях анкеты;
- 4 исключение респондентов по категории исследования и/или количеству заполненных анкет за определенный период;
- 5 сопоставление половозрастных данных, указанных в анкете и профиле респондента в панели;
- 6 включение вопросов-ловушек (например, в табличном вопросе в одной из строк просьба отметить определенный вариант ответа);
- 7 логическая соответствие (например, при доходе в 10 000 руб. в месяц наличие автомобиля премиум-класса);
- 8 устранение возможности неоднократного заполнения анкеты одним и тем же респондентом;
- 9 проверочные вопросы с картинками, видео- или аудиоматериалами для контроля показа картинки или видео и передачи звука;
- 10 проверка качества ответов на открытые вопросы (формат текста, количество знаков, сравнение с тезаурусом).

Важно понимать, что наличие проверок качества заполнения онлайн-анкеты может вызвать неудобства для респондента. Здесь необходимо выдержать баланс. В онлайн-опросах нет интервьюера, который может уговорить закончить интервью. При клике на «красный крестик» в верхнем углу окна онлайн-респондент будет потерян, и скорее всего безвозвратно для данного проекта.

Проверки при *обработке полученных данных*.

Проверка времени заполнения анкет (при отсутствии проверки и отбраковывания во время заполнения).

На данном этапе отбраковываются анкеты, на заполнение которых респондент затратил неоправданно мало времени. Один из способов реализации такой проверки – удаление из базы тех интервью, которые выходят за пределы двух среднеквадратических отклонений от значения медианы. Такие респонденты, так называемые спидстеры (англ. speedsters), прокликают анкету, не вчитываясь в вопросы и задания. При реализации данной проверки важно помнить, что онлайн-анкета может допускать перерыв в заполнении. Респондент может вернуться к заполнению анкеты, например, на следующий день. Однако если в анкете демонстрировались видео- или другие материалы для оценки (например, тексты концепций), длительный перерыв (от нескольких часов до нескольких суток) не рекомендуется, так как может быть забыто содержание показанных материалов.

Проверка на заполнение табличных вопросов (так называемый straight lining или zig-zagging).

Стрейтлайнинг (straight lining) – это заполнение анкеты по определенному «маршруту», по одной вертикальной, горизонтальной линии или зигзагом. Респондент не вчитывается в вопрос или задание, прокликает анкету, стараясь скорее перейти к следующему вопросу. При проверке таблицы с такими ответами анализируются по смысловой нагрузке вопросов. В некоторых случаях одинаковые ответы на все пункты таблицы приемлемы. Проведение

проверки такого рода имеет смысл для оценочных таблиц с количеством альтернатив больше двух. Интервью считается плохим, если одинаково заполнено более 50% вопросов.

Проверка на идентичные анкеты.

Такие анкеты появляются в массиве по двум основным причинам. Первая – отсутствие технологического контроля на неоднократное заполнение анкеты одним и тем же человеком. Второй причиной обычно является недобросовестность подрядчика, проводящего сбор данных. В этом случае данные могут быть сдублированы или заполнены по шаблону для выравнивания недостающих квотных ячеек. В файле данных идентичные записи обнаруживаются путем визуального контроля – сортировкой ответов по нескольким вопросам.

Проверка ответов на открытые вопросы.

В открытых ответах обязательна проверка на смысловую адекватность, соответствие вопроса и ответа (например, неадекватным на вопрос «Опишите Ваши эмоции» будет ответ «Вторник»). Но большое количество открытых вопросов обычно приводит к тому, что респондент дает менее развернутый ответ на каждый следующий вопрос такого типа. При отсутствии интервьюера, способного мотивировать дать полный ответ, быстро теряется интерес к полным ответам и, как следствие, ответы становятся короче и однотипнее.

Проверка на нетипичное заполнение нетаргетируемых квот.

«Таргетируемость» (англ. targeting – нацеливание) предполагает возможности отбора респондентов для участия в опросе из панели. Существует 3 типа параметров отбора:

- 1 Хорошо таргетируемые параметры. К ним относятся пол, возраст, место проживания. Сформировать состав выборки по этим показателям можно с высокой степенью точности. Совокупно эти параметры дают около 1–2% ошибок отбора.
- 2 Средне таргетируемые параметры. Сюда относятся все указанные в профиле респондента данные (кроме пола, возраста и места проживания). Они могут изменяться, поэтому позволяют сделать только предварительный отбор потенциальных респондентов. Вероятность ошибки варьируется от 2 до 50%. К данным с большей ошибкой относятся доход, владение компьютерной техникой, пользование услугами сотового оператора и т.д. Более точные – наличие детей, крупной бытовой техники, домашних животных и т.д.
- 3 Нетаргетируемые параметры, т.е. все квотные ограничения, которые не учтены в профиле панелиста. В эту группу параметров входят потребляемые марки, частота потребления определенных продуктов, посещения магазинов и т.д., т.е. все то, что невозможно учесть при составлении профиля в онлайн-панели. Неожиданный процентный рост такого параметра при сборе данных, например, ближе к завершению полевого этапа, дает серьезный повод задуматься о качестве полученных данных. Увеличение данного показателя в отличие от параметров

первой и второй групп невозможно вызвать большей или специализированной рассылкой приглашений потенциальным респондентам.

Анализ частотных таблиц помогает обнаружить нетипичное заполнение нетаргетируемых параметров. Распределения оказываются скошены к определенному варианту ответа. Особого внимания требует резкий рост и доли квотных параметров по конкретным дням плевого этапа (особенно к концу полевых работ).

Проверка на количество ответов «Затрудняюсь ответить».

Если вариант «Затрудняюсь ответить» выбирается в половине ответов, анкета помечается как подозрительная. Это свидетельствует о невнимательном отношении респондента к ней (или о некорректном составлении вопросов).

Проверка на заполнение анкет по дате и времени.

Результаты анализа 30 онлайн-проектов, проведенных 2015 г., показывают, что максимальное количество анкет заполняется респондентами в будние дни и в рабочее время.



Рисунок 3 – Количество заполненных анкет по дням недели в 2015 г.

В выходные и в праздничные дни активность снижается в 2–3 раза, а в хорошую погоду – в 4 раза. Этот фактор необходимо учитывать в сроках проекта.

Второй фактор, который также необходимо учитывать, – время заполнения анкет с учетом часовых поясов. Нетипичные активности респондентов в ночное или утреннее время по местному времени дают повод усомниться в источнике данных. Например, если большая часть анкет для Владивостока заполнена после полуночи по местному времени, когда в Москве еще не закончен рабочий день, то заполнение с большой долей вероятности происходит не во Владивостоке.

Проверка на идентичность IP-адресов.

При программировании опроса на стороне исследовательского агентства целесообразно включение регистрации IP-адреса, с которого респондент заходит на анкету. Это позволит осуществить дополнительную косвенную проверку места жительства. Однако возможна ситуация, при которой IP-адрес не будет совпадать с отмечаемым местом проживания, в случаях, когда респондент находится на отдыхе или в командировке. IP-адреса заполняемых анкет с мобильных устройств могут существенно отличаться от фактических, так как многие сотовые операторы для выхода в интернет используют общий ограниченный пул адресов.

В заключение акцентируем внимание на том, что применять проверки для определения качества анкет необходимо в совокупности. Важно помнить, что респондент, заполняя анкету на своем компьютере или другом устройстве, может быстро потерять интерес в силу разных причин (персональных, не связанных с особенностями анкеты или напрямую вызванных ею – слишком длинная, скучная, сложная) и прервать заполнение. Одной из причин прекращения участия в опросе может быть большое количество проверок, вызывающих у респондентов негативную реакцию, они начинают давать заведомо неверные ответы и вообще прекращают участие в исследовательских проектах.

Общие рекомендации можно сформулировать следующим образом:

- 1 Запрашивать осуществляемые процедуры проверок у подрядчика, выполняющего программирование электронной версии анкеты и сбор данных.
- 2 Уменьшить количество проверок во время заполнения анкеты, увеличив проверки на этапе обработки данных, избегать переизбытка проверочных вопросов для респондента.
- 3 Обеспечить наличие набора из 3–4 проверок для каждого планируемого проекта.
- 4 Учитывая характер заполняемости анкет в праздничные и выходные дни планировать сроки полевого этапа.

Литература

- 1 Исакова Е.Б. Планшетизация – первые практические уроки. Опыт общероссийского омнибуса ВЦИОМ / Секция «От бумаги к планшетами» // Материалы V Грушинской конференции 2015. Москва, 2015.
- 2 Ананьев А.Я. Практика опроса на планшетных компьютерах / Секция «От бумаги к планшетами»// Материалы V Грушинской конференции 2015. Москва, 2015.
- 3 Мягков А.Ю. Всегда ли респонденты говорят правду // Социологические исследования. 2008. № 9. С. 20–31.
- 4 Global Market Research 2013 // ESOMAR. ISBN 92-831-0270-3.
- 5 Медведева Е. Как фабрикуются анкеты в автоматизированных опросах / Секция «От бумаги к планшетами»// Материалы V Грушинской конференции 2015 г. Москва.
- 6 Puleston J., Eggers M. Dimensions of online survey data quality. What really matters? // Congress 2012 – Accelerating Excellence – Celebrating 65 Years And Beyond. Atlanta, 2012.

- 7 Phillips K. Data Use: An evaluation of quality-control questions// Quirk's Marketing Research Review. December. 2013. URL: <http://www.quirks.com/articles/2013/20131205.aspx>.