

## Секция

### «Антихрупкость репрезентативности»

**Модератор Рыбакова Ольга:** Дорогие коллеги, добрый вечер! Мы рады видеть вас здесь, в этой аудитории, мы начинаем нашу секцию, которая будет посвящена проблемам выборки. Мы назвали ее «Антихрупкость репрезентативности» не для того, чтобы завлечь как можно больше слушателей, участников, а потому, что на наш взгляд «толебовская антихрупкость» – это метафора, которая больше, чем что-либо другое, какое-то другое определение подходит к пониманию той ситуации, тех проблем, которые сложились сегодня в нашей отрасли, и связаны они с проблемой выборочного метода, его использования.

Не случайно в своем развитии выборочный метод, как вы знаете, прошел путь от абсолютного одобрянца, когда считалось, что любой опрос, который проведен на основе использования вероятностной выборки, он достоверен на 100%, ему можно верить. Сегодня многие и коллеги, и, что особенно напрягает, обычно, заказчики слова «репрезентативность», «ошибка», «погрешность» произносят, как мантру. Действительно, по мере того, как развивалась наша индустрия, вопросы и проблемы, связанные с сочетанием опросного метода и различных типов выборки множилось, множилось, множилось и требовали анализа, изучения, погружения. Поэтому, на наш взгляд, само название «антихрупкость репрезентативности» связано с тем, что очень ценным является выявление различного рода ошибок, проблем, трудностей, ограничений, с которыми мы сталкиваемся по мере развития и опросных технологий, и понимания, что такое выборочный метод, каковы границы использования разных типов выборки. Поэтому я надеюсь, что наша секция сегодня внесет определенный вклад, хороший, позитивный, конструктивный вклад в изучение способов усиления антихрупкости выборочного метода в практике социологических исследований.

Прежде, чем мы начнем нашу секцию, я приглашаю вас почтить память Асхата Кутлалиева, который один из самых крутых профессионалов, специалистов в области социологических исследований, выборочного метода, к сожалению, сегодня не с нами, а он в этих вопросах был одним из самых-самых крутых специалистов. Давайте почтим его память (*минута молчания*).

Спасибо большое. Мы объявляем первый доклад. Владимир Журавлев, Московский государственный лингвистический университет.

**Журавлев Владимир:** Всем добрый вечер. Мне выпала честь открывать нашу замечательную секцию, сегодня в своем докладе я хочу поднять и, наверное, частично предложить некие способы решения ряда проблем, с которыми сталкиваются исследователи в условиях современной реальности, когда доля традиционных методов постепенно сокращается, снижается, у нас появляются новые инструменты, новые способы работы и взаимодействия с респондентами, их опросов, в частности – всевозможные онлайн методы. Также последнее время очень сильно на слуху всевозможные вещи, касающиеся больших данных, работы с ними, их плюсов, минусов, ограничений.

В самом начале хочу сделать некое предупреждение и предостережение, что то, о чем я буду говорить сегодня, во многом экспериментальные вещи, которые требуют где-то дальнейшей проработки, часть гипотез может подтвердиться, часть гипотез может быть опровергнута. Но в этом и заключается, по моему мнению, работа социолога и работа исследователя – это тестировать, принимать, отвергать и двигаться дальше, чтобы приблизиться к некоей методологической и методической корректности.

Для начала хочу разобрать основные определения, дефиниции, с которыми сегодня в процессе подготовки доклада я плотно разбирался, работал.

Методология. Под «методологией» я понимаю некий общий подход, общую стратегию исследовательскую в рамках которых мы принимаем то, как мы будем работать с нашей объектно-предметной областью, каким образом ее исследовать, от каких и из каких теоретических и методологических предпосылок исходить. В данном случае я в качестве основ брал социологический конструктивизм, потому что как раз он нам позволяет изучать некое социальное явление в определенной конкретной среде, в частности в контексте моего доклада это социальные сети.

Я далее «социальные сети» понимают как некую среду, в которой существует человек, пользователь, в которой он осуществляет некое поведение, у этого поведения есть индикаторы проявления. Дальше мы с этим человеком можем взаимодействовать при помощи опроса, задавать ему вопросы, создавать информацию и дальше также получать некую пассивно собираемую информацию о том, на какие страницы, сайты он переходил, где он совершал клик, где он делал отметку «мне нравится», подписывался на что-то и т.д. Очень важный принцип дополнительности, я считаю, что это по аналогии, например, в естественных науках, когда есть некие сферы исследования, которые вносят вклад в различные науки, т.е. на стыке биологии, физики рождаются какие-то новые вещи. В данном случае на стыке информатики, программирования, на стыке социологии могут рождаться новые подходы, которые нужно тестировать, нужно пробовать и двигаться дальше.

Если говорить о второй части методики, о методической составляющей, то это уже конкретный набор операций, процедур, приемов, методов, которые мы используем для того, чтобы что-то изучить и для проведения конкретного социологического исследования. В частности важным моментом здесь является выборочное исследование, выборочный метод и механизмы, которыми мы осуществляем некий выборочный отбор. Это во главе угла. Потому что то, как и в каком виде, и насколько методически корректно мы соберем нашу выборку, это определяет дальнейший результат и наши возможности, и анализы, и работы с этими данными, их интерпретацией и т.д.

Первое. Очень много на разных конференциях и со стороны экспертов идет интересных дискуссий, полемики на тему столкновения традиционных онлайн-методов, в чем их плюсы, в чем их минусы, преимущества и недостатки. Я хочу обозначить несколько проблем, которые присутствуют в онлайн-методах и в частности в онлайн-опросах – это, например, проблема ненаблюдения. Когда мы говорим об онлайн-исследовании, мы, естественно, никаким образом не сможем охватить не пользователя интернета, если только осознанно не будем прилагать для этого дорогостоящие очень большие усилия. Это проблема ненаблюдения.

Проблема неответа. У нас отсутствует интервьюер, поэтому мы не можем приложить дополнительные усилия для того, чтобы убедить человека принять участие в нашем опросе. То, насколько хорошо мы соверши первый контакт, тут, по сути, происходит либо переход в опросы – человек соглашается, самостоятельно придумав себе некое объяснение и рационализацию, почему он это делает, либо человек просто игнорирует, как, например, вы можете вспомнить количество интернет-рекламы, с которой вы сталкиваетесь, и то количество, на которое вы все-таки кликаете, переходите, пытаетесь что-то изучить и посмотреть. Здесь работает ровно та же самая логика.

Следующий момент – проблема самоотбора. Когда мы говорим об онлайн-опросе, мы априори не можем контролировать генеральную совокупность. Первое – неизвестна вероятность попадания каждого из интернет-пользователей в нашу выборку, второе – для некоторых пользователей она равно нулю. Уже некий принцип вероятностного отбора и построения вероятностного отбора выборки нарушается в этом случае.

Наконец, еще одна важная проблема – мотивированный трафик. Это люди, которые совершают какое-то действие, например, прохождение опроса, исходя из набора корыстных мотивов. Это может быть получение вознаграждения за конкретный опрос денежное или в денежном эквиваленте, получение какого-то промо-кода, скидки и т.д. Все это я отношу к категории «мотивированный трафик», т.е. он мотивированный в плохом смысле. Он мотивированный не потому, что человек, респондент хочет внести вклад в изучение какого-то явления, а потому, что он не фокусируется на тематике опроса, дает не очень качественные ответы. Как правило, средняя скорость прохождения подобной категории опроса кратно меньше, статистически значимо отличается от медианы и от средней продолжительности прохождения опроса нормальными респондентами и, в принципе, часто люди врут. Когда мы как исследователи допускаем ошибку и выносим, например, в приглашение опроса тематику и человек может догадаться, какой ему нужно дать ответ на скринирующие вопросы для того, чтобы пройти дальше, мы таким образом его подталкиваем сказать неправду и принять участие в этом опросе.

Дальше есть важный момент – то, как мы уже разработанный инструментарий и разработанную сначала программу, потом анкету через операционализацию понятий, через работу инструментария, как мы ее реализуем, в какой форме, в каком виде.

Видов может быть несколько.

Первый – доступные бесплатные и платные так называемые «конструкторы опросов». Здесь представлены примеры наиболее распространенных глобальных конструкторов – это Серви Гизма, Серви Манки, Тайп Форм, Гугл Формы, есть еще большой-большой список, который можно перечислять. У этих конструкторов опросов есть определенные ограничения, связанные с бесплатным использованием, например, либо объем вопросов, которые вы можете создать в бесплатном аккаунте, либо количество респондентов, которые вы можете опросить. Как правило, они ограничивают либо одним, либо вторым параметром, т.е. либо не больше 10 вопросов на вашу анкету, на ваше исследование, либо не больше 100 человек вы можете опросить, это тоже является неким ограничением.

Любой внешний опрос, то, к чему я веду, это некая ссылка. Для человека в зависимости от того, где вы его пытаетесь опросить, она является неким, будем

называть это «внешним переходом». Например, вам эта ссылка приходит в социальной сети ВКонтакте или в Одноклассниках, вы на нее кликаете и переходите на какой-то внешний сайт. Соответственно, на этом этапе уже происходит отказ от дальнейшего участия, люди думают, что что-то странное, мошенничество или еще что-то, сейчас у нас очень развита паранойя – много мошеннических схем, поэтому люди иногда бросают и не идут дальше. Это неудобно, это нарушает то, что человек планировал делать. Например, он пришел в сеть пообщаться, посмотреть видео, картинки, полистать ленту и тут надо куда-то переходить, это нарушает его некий сценарий привычных действий, это нарушает его пользовательскую сессию на конкретном сайте.

В противовес этому есть инструмент или способ, который я специально вынес в кавычки – так называемый «нативный опрос», потому что при буквальном переводе с английского на русский смысл немножко теряется. Суть та же – это родной опрос для того места, где вы его проводите. Он не отличается с точки зрения дизайна внешнего вида интерфейса и человек в данном случае остается внутри социальной сети ВКонтакте, не совершает никуда переход, его левое меню доступно, он может вернуться к своим действиям, может вернуться обратно в опрос и т.д., что существенно влияет на повышение так называемого «респонс рейта», т.е. отклика на опрос и более качественного прохождения.

Если сравнивать внешний и нативный опросы, то, конечно, нативные опросы требуют гораздо больше ресурса и определенного набора знаний, навыков и компетенций с точки зрения их разработки, программирования и создания.

Переходим к основной вещи, проблематику которой я хотел сегодня поднять. Так что же все-таки делать? Возможен ли репрезентативный опрос в интернете? Если возможен, то на каких условиях, с какими ограничениями, если невозможен, то что же нам с этим делать?

Здесь я очень много проводил различных экспериментов, в таблицу пытался вынести различные типы. Во-первых, два параметра – наличие списка исходного. Т.е. есть ли у вас понятие о генеральной совокупности в целом либо хотя бы о выборочной совокупности, с которой вы планируете работать, есть ли этот список? Отлична ли вероятность ВКонтакте с каждым из пользователей, респондентов в этом списке от нуля, известна ли она? Сможем ли мы на основании этого построить вероятностную или хотя бы псевдо-вероятностную выборку? Второе – наличие непосредственно вероятностного механизма отбора. На пересечении у нас получается несколько различных категорий. Остановлюсь на них подробнее. Давайте пойдём с правого нижнего угла.

Первое – опросы без ограничений. Это любой опрос, который выложен или представлен на любой странице в интернете. Мы ничего не контролируем, есть какой-то трафик или аудитория, или пользователи, которые посещают эту страницу, любой из тех, кто проявил к этому интерес и прошел опрос, попадает в нашу выборку. Т.е. это полный самоотбор, нет вероятностного механизма отбора, соответственно, мы никаким образом не контролируем ни выборочную, ни генеральную совокупность.

Дальше – перехватывающие опросы. Данный сценарий возможен в том случае, если вы являетесь владельцем сайта. Если вы являетесь владельцем сайта, у вас есть возможность технически показывать опрос не постоянно всем, а, например, показывать его с каким-то шагом. Понятно, что это может приводить к некой систематической

ошибке, если это нужно проверять и тестировать, но, условно говоря, вы можете, например, добавлять туда элемент рандомизации и показывать какому-то эн случайному пользователю, пришедшему на ваш сайт, распределив сбор данных или показ опроса по времени, по частоте и т.д. Здесь у нас появляется некий механизм вероятностного отбора, мы можем говорить о том, что мы немножко как бы улучшили ситуацию и собираем чуть менее искаженную, смещенную выборку.

Следующий вариант – когда у вас есть, например, база или массив данных, например, вся аудитория вашего сайта, и вы знаете, что это все активные пользователи, с каждым из них вы имеете техническую возможность сконтактировать и пригласить для участия в опросе. Т.о. вы можете уже построить нормальный полноценный план выборки, сконтактировать с этими людьми, собрать эти данные, обработать их, проанализировать и опять же приблизиться к некой репрезентативности. Но опять же не забываем, что в данном случае мы генеральную совокупность искусственно сужаем до аудитории вашего сайта или ресурса.

Онлайн-панели. Онлайн-панели, например, такие как UMI или Tiburon, или какие-то другие панели предлагают тоже работу с генеральной совокупностью. Опять же мы говорим не о всей аудитории Рунета, не о всех пользователях, а о конкретном конечном наборе пользователей, которые выразили свое согласие участвовать в панельных исследованиях и, соответственно, принимать участие в опросах. Здесь опять же мы можем частично приблизиться. С отсутствием вероятностного механизма отбора – это тоже, по сути, зеркально опросы, в которых мы просто план выборки не строим и не контролируем это.

Что мы можем как исследователи, как люди, приложившие усилия для изучения существующих методических и методологических подходов, что мы можем здесь сделать?

Первая возможность, которая становится очевидной, - это попытаться нивелировать систематическую ошибку, максимально широко рекрутируя респондентов в опрос, т.е. по времени и по источникам. Максимальное количество источников, где мы можем это сделать, вывесить это приглашение, максимально длительно по времени, но опять же с учетом сроков, которые на нас накладывает решение либо научно-исследовательской, либо какой-то бизнес-задачи.

Дальше – добавление неких элементов квотирования, т.е. контроль профиля собираемой выборки, но здесь мы опять же должны иметь некий изначальный источник данных, которым мы доверяем. Это могут быть данные «Росстата», это могут быть данные «Медиаскопа», веб индекса об распределении по социально-демографическим параметрам, по географии пользователей интернета, это могут быть еще какие-то данные, но мы должны им доверять, потому что таким образом м свою выборку начинаем контролировать с точки зрения конкретных групп, которые мы туда набираем.

Добавление элементов рандомизации.

Далее – искусственное ограничение генсовокупности, это то, о чем я уже сказал, когда часто нам приходится немножко умирять свои исследовательские амбиции и говорит, хорошо, Рунет мы не сможем нормально репрезентировать исследовать, давайте будем сужать целевую аудиторию, исследуемую нами, с этом случае мы, по

крайней мере, более качественно сможем соблюсти все методы и процедуры, и приблизиться к репрезентативной выборке.

Дальше – есть метод игнорирования на этапе сбора данных всего этого. Дальше уже через некий набор аналитических и статистических процедур взвешивать данные, восстанавливать данные, пытаться их моделировать, но здесь опять же без того самого волшебного источника, про который я сказал, без исходных установочных данных мы не сможем это делать т.е. нужно понимать, подо что мы корректируем нашу выборку.

Наконец, если говорить о проблеме неохвата или непокрытия, например, те самые непользователи интернета, конечно же, мы можем дополнительно сделать отдельную ветку, провести с ними фейс-ту-фейс интервью, дойти до них вживую, это может быть поквартирник или еще что-то. Но это будет дорого, это может существенно удлинить сроки и, например, в какой-то ситуации это может быть неадекватным решением нашей исследовательской задачи.

Инструменты, которые могут использоваться.

Во-первых, это рекламные инструменты для рекрутирования в наш опрос. Это пример рекламного кабинета ВКонтакте, в котором мы можем заводить рекламное объявление, создавать определенный креатив, текст, картинку для привлечения человека в опрос, возможно говорить о том вознаграждении, которое предполагается или нет. Дальше – выбирать тот набор таргетингов, который доступен, – это пол, возраст, география, это могут быть интересы и различные технические параметры вплоть до браузера, через который человек выходит в интернет и, соответственно, операционная система.

Мы можем работать с так называемым «органическим трафиком», создавая всевозможные сообщества вокруг опросов, работать с теми самыми подписчиками. Для работы с подписчиками у нас есть по сообществу тоже вся аналитика доступная, по полу, возрасту и другим параметрам, которыми мы можем управлять, и географией в том числе.

Наконец, к этим данным мы можем при помощи открытых инструментов через так называемый API ВКонтакте получать данные о всевозможных дополнительных параметрах и обогащать данные опрошенных людей либо той выборочной основы, с которой мы планируем работать. Делается это посредством либо простых URL запросов в адресной строке браузера и мы получаем ответ – информацию о пользователе, здесь в качестве примера фамилия и имя у того профиля, который открыт и в котором человек не закрыл свой профиль от использования и анализа, либо мы можем использовать специальное программное обеспечение такое, как Пайчар и в нем писать простые скрипты, которые позволяют нам эту информацию тоже регулярно получать.

Дальше процесс достаточно просто устроен. Это все мы можем выгрузить и сохранить в CSV, и открыть в нашем любимом SPSS для того, чтобы уже заниматься непосредственно анализом и статистической обработкой этих данных.

В качестве резюме хочу сказать, что я считаю, что одна из задач социологов, социологов-исследователей и всех практиков в области исследований на текущий момент – это как можно больше экспериментировать, но не бездумно, а с оглядкой на те методологические и методические основы, которые есть и существуют, т.е. отталкиваться

от тех подходов, которые есть, дальше пробовать и пытаться приблизиться к репрезентативности, пытаться делать опросы более методологически и методически корректными и ни в коем случае не отчаиваться и не думать, вдруг кто-то из здесь сидящих будет тем человеком, который придумает и предложит какой-то новый подход, который станет популярным и будем всеми нами эффективно использоваться.

На этом все, большое спасибо. Если есть вопросы, готов ответить.

**Попкова Светлана:** Добрый день, Светлана Попкова, компания «Онлайн Маркет». Я хотела уточнить по поводу опыта использования рекламного трафика. По вашему опыту, сколько времени нужно на запуск этой кампании от ее настройки до получения первых откликов?

**Журавлев Владимир:** При работе с рекламными компаниями есть контролируемый исследователем момент – это все подготовка, т.е. непосредственно заведение руками, создание этого объявления. Есть неконтролируемый момент – это так называемая «модерация», которая осуществляется на стороне рекламной площадки. Вот модерация чаще занимает времени больше, чем непосредственно заведение кампании. Т.е. ваша рекламная кампания может модерироваться от 15 минут до 2 часов, а заведение занимает, если это у вас уже некая налаженная процедура, понятные шаги, вы знаете, что вы делаете, 15-20 минут максимум.

**Березин Николай:** Николай Березин, компания «Онлайн интервьюер». Вы проводили исследования в рамках вашей работы в MAIL.RU и в Яндексе или в рамках текущих диссертационных работ тех примеров, которые вы иллюстрировали?

**Журавлев Владимир:** Хороший вопрос. Мне кажется, вы спрашиваете, насколько это возможно реализовать, не являясь частью этих компаний. Я в какой-то момент для себя определился, что для того, чтобы эффективно совмещать рабочую деятельность и научно-исследовательскую деятельность, это должно быть в каком-то едином предметном поле. Поэтому у меня нет четкой границы, я очень много экспериментов и исследований на практике запускал и в Яндекс, и в MAIL.RU, я запускаю те же самые инструменты, самостоятельно лично не пользуясь никакими преимуществами и привилегиями работника одной из этих компаний. На текущем этапе у меня больше пропорций именно в рамках моей научно-исследовательской деятельности, я этим занимаюсь. До этого, как я показывал в самом начале, до 2018 года я ежедневно этим занимался в качестве моей работы.

**Березин Николай:** Еще один вопрос. Вы показали все это на примере ВКонтакте. Понятно, что количество площадок должно быть максимальным, их должно быть много, они должны быть разнообразными. Скажите, пожалуйста, по вашему опыту, где еще, кроме ВКонтакте, где активнее, где получается сложнее?

**Журавлев Владимир:** Спасибо, тоже хороший вопрос. Я работал с рекламными площадками так называемыми РСЯ, это рекламная сеть Яндекса, она включает в себя огромное количество разных сайтов, площадок, порталов и т.д. Я запускал во ВКонтакте и в Одноклассниках, и запускал в Facebook. Если говорить о дополнении, то, например, с точки зрения социально-демографических характеристик комбинация ВКонтакте+Одноклассники дает очень хорошее покрытие всех возрастных групп. У ВКонтакте есть возрастной перекосяк в более младшую возрастную категорию – это школьники, студенты и меньше представлены средний возраст и старший. В

Одноклассниках обратная пропорция, т.е. категория людей 45+ там представлена очень хорошо. Facebook дает некую серединку. На самом деле, многие пользователи пересекаются, у них есть профили везде, но по своему стилю, типу поведения и поведенческим стратегиям социальная сеть определяет то, как ты себя ведешь в Facebook отличается от того, как ты себя ведешь во ВКонтакте и по типу действий, и по некой специфике, и т.д. На текущий момент из реальных наших, так скажем, отечественных производителей и площадок комбинация ВКонтакте+Одноклассники дает оптимальное соотношение по покрытию и по цене усилий, ресурсов, которые нужно приложить для того, чтобы реализовать там опрос, исследование.

**Реплика из зала:** Стоимость рекрута одного респондента сколько примерно составляет?

**Журавлев Владимир:** Такого понятия, наверное, нет, потому что очень сильно зависит от конкретного опроса, от его длительности, от того, насколько вы качественно подбираете и создаете рекламное объявление. На самом деле стоимость человека, прошедшего опрос, может варьироваться от 10-20 рублей до бесконечности. Есть некоторые опросы, которые, например, нецелесообразно таким образом проводить. Например, это та точка, в которой онлайн-панели всегда будут давать фору подобному методу просто потому, что там паделлисты, например, готовы проходить опросы 20-30, в некоторых случаях 40 минут за какую-то мотивацию, а у обычного человека та доля внимания, на которую мы можем рассчитывать, максимум 10-15 минут, а еще лучше 5 минут, а еще лучше – три вопроса задать, но несколько раз его поймать и продолжать с ним некий диалог. Здесь очень условно, это надо тестировать. Даже то, с чем я столкнулся: если посадить двух человек заводить рекламные кампании и заниматься привлечением респондентов, у них будут разные результаты по стоимости и по количеству усилий. Хорошо работает понимание. Например, кто сам понимает, что ВКонтакте востребовано, как это все работает, какие интересы у людей, у сообщества, тогда это работает лучше и получается дешевле.

Спасибо большое за вопросы. На последнем слайде мои контакты, если кто-то не успел или хочет какую-то еще информацию получить, пожалуйста, я открыт и готов пообщаться в Telegram, в Facebook или на почту, если кому-то удобно.

**Модератор:** Спасибо, Владимир. Следующий доклад Мария Макушева, про онлайн-исследования продолжим.

**Макушева Мария:** Добрый вечер, коллеги. У меня сейчас будет достаточно сложная задача, потому что у меня по содержанию много пересечений с тем, о чем говорил Владимир, постараюсь не повторяться, это такой дисклеймер. Откуда эти данные? Работа во ВЦИОМ с нашими технологическими партнерами еще во ВЦИОМ, мы в течение двух лет проводили методические эксперименты с выборками типа интерсепт перехватывающими, ривер, с обогащением данных вместе с ДМП платформами, это некоторое такое обобщение будет разных экспериментов.

Там, где я буду ссылаться на данные «САТІ», это ВЦИОМ-СПУТНИК, вероятностная телефонная выборка, такой эталон методический в данном случае.

Была идея, как лучше построить доклад – взять один узкий частный вопрос и разобрать его более глубоко или общую схему, как устроена выборка, какие у нее есть особенности, что нужно учитывать при ее применении. Выбрала второй путь. Если будут



более частные вопросы, пожалуйста, я готова к общению, задавайте в аудитории или после.

Владимир об этом отчасти говорил, я буду говорить преимущественно о выборках типа ривер или интерсепт, при которых мы опрашиваем людей, предварительно не рекрутированных в панель, приглашая их к участию в опросе посредством различных каналов – баннеры, офферы, с помощью ДМП формируя сегменты, используя социальные сети. Здесь данные собраны преимущественно в двух или трех – либо ВКонтакте, Одноклассники, Facebook либо только ВКонтакте, Одноклассники. И рекламные сети – мы использовали Mail, Яндекс, Google в начале, потом отказались в силу особенностей, будет время, остановлюсь подробнее. Это всегда принципиально таргетированные по квотным параметрам опросы. Если это общие целевые группы, например, население определенных возрастных когорт, определенных территорий как в тех данных, на которые я буду ссылаться, то это по половозрастным и территориальным характеристикам. Если это опросы узких целевых групп, например, любители российского кинематографа, то это таргетные интересы. В общем, вся та информация, которая есть у рекламных сетей или у ДМП, на нее можно таргетировать.

О чем хотела сказать, построить доклад вокруг различных эффектов, с которыми мы сталкиваемся в интернет-опросах. Понятно, что ключевая проблема этой выборки – она не является вероятностной, а у любой невероятностной выборки мы не можем ее оценить сдвиги, мы не можем посчитать ошибку, она явно заведомо теоретически хуже, чем выборка вероятностная, но первая не всегда достижима, а вторая повсеместно используется. Поэтому, наверное, логичнее говорить не о недостатках, а об эффектах. Это более рабочий подход к ситуации, от которой мы все равно никуда не уйдем.

Эффекты. Не буду говорить очевидного об аудиторных эффектах, интернет-аудитория не равна населению, там представлены не все, т.е. потенциально доступны не все, а если говорить о реально доступных людях, то практически максимально трудно опросить людей 55+, максимально трудно опросить жителей населенных пунктов 100-, и это имеет свои пересечения. Например, Москва 55+ ОК, 18-24 город 100-, село – тоже нормально. Об этом подробно говорить не буду.

Начну с саморекрутирования, с эффекта, связанного с самоотбором. Желание, действие, когда человек принимает решение, встроено в некоторый смысловой алгоритм, мы его не можем не учитывать, когда анализируем эти данные. Здесь представлена некоторая воронка, она иллюстрирует ситуацию, в которой мы показываем приглашение несоизмеримо большему количеству людей, чем те, которые в итоге проходят по ссылке и чем те, которые в итоге оказываются в базе с комплитами.

Это достаточно серьезное влияние. В чем оно проявляется? Оно проявляется в первую очередь в возрасте и дальше идут связанные с этим эффекты. Есть потенциально доступные в интернете люди, есть те, кто имеет большую склонность попасть в такого типа выборку. Здесь у нас на первый план выходит возраст, гораздо легче рекрутировать. Если мы не квотируем возрастной интервал 18-34, например, внутри, то мы получаем вот такую картину. Оранжевые – это данные CATI, вероятностная выборка общероссийская, сопоставимая группа, интернет-пользователи, голубая линия – это ривер. Если мы квотируем 18-24, то все равно получаем большую долю 18-19 летних. У нас все равно выборка всегда будет сползать в сторону самых молодых интернет-

пользователей. Это влечет за собой то, что в выборке выше доля студентов, ниже доля людей с высшим образованием, ниже доля работающих, причем, доля работающих будет ниже в любом возрастном интервале.

Здесь показаны доверительные интервалы для CATI и насколько в них попадет или не попадет ривер. Это немножко странная игра с цифрами, потому что для ривер мы не имеем права рассчитывать доверительные интервалы, но, тем не менее. Хоть для какого-то сравнения я эти данные приведу вместе, так как выборка невероятная.

У нас есть сдвиги по возрасту, по материальному положению, эффект саморекрутирования это дает, у нас есть сдвиги по работающим, неработающим аудиторные, также у нас есть сдвиги по поведению и медиапотреблению. Понятно, что опрашивая людей в интернете, мы имеем гораздо больший шанс получить в выборке выше долю так называемых «хэви юзерс» - интенсивные пользователи интернета, те, кто пользуется ежедневно больше 4-х часов или больше 3-х часов по методике разных компаний. При опросе телефонном в выборку в категории 18-45 интернет-пользователей примерно 44% оказывается хэви юзерс, в сопоставимой категории в онлайн-опросе 67%, всегда разброс существенный, что, в общем, и понятно.

Есть большая разница в использовании социальных сетей. Те, кого мы опрашиваем в интернете гораздо интенсивнее ими пользуются в сопоставимой категории, чем те, кого мы опрашиваем по телефону, это для нас тоже свидетельствует об эффекте самоотбора, в результате которого наиболее интенсивные пользователи оказываются в нашей выборке.

Есть отличия в практиках. В силу того, что в этой выборке оказывается больше людей неработающих и больше молодежи, даже если мы все это перевзвешиваем, все практики, связанные с серьезными покупками, например, бытовой техники, с госуслугами оказываются значительно ниже представлены, чем в сопоставимой категории при опросе CATI. Гораздо больше при этом развлекательного контента потребляется, более представлены такого типа интересы. Т.е. отличия в практиках. Это вкратце про самоотбор.

Следующий эффект связан с площадкой. Интернет-пространство неоднородное, где ты опросишь, то и получишь. Недавно коллеги, которые экспериментировали с ривером и делали таргет на разного типа интересы, смогли опросить таким образом накануне выборов, что КПРФ набрало в три раза больше Единой России, в клик стрим были «Ведомости», «Коммерсант», разные деловые издания. В общем, здесь факт нахождения на определенной площадке и склонность именно на ней отреагировать на какое-то объявление, т.е. некоторое доверие ей не является нейтральным, это все приводит к сдвигам. Мы фиксировали для каждого опрошенного источник рекрута, потом анализировали их различия. Когда у нас накапливалось хотя бы тысячи 3, 4, 5, опрошенных через ривер, у нас уже были достаточно наполненные категории по разным источникам, чтобы мы могли их сравнить. Мы действительно видим довольно существенные различия между, во-первых, теми, кого мы рекрутировали из социальных сетей и теми, кого мы рекрутировали через рекламные сети. Это различие и в возрасте, и в населенных пунктах, стрелочками представлено «выше», «ниже», подробнее, наверное, пояснять времени нет, если что, потом на вопрос отвечу. Есть разница между

разными социальными сетями, как отмечал Владимир, ВКонтакте дает наиболее молодую аудиторию и т.д.

Есть разница, которую мы получили по сегментам, которые дают ДМП и которые можно взять у Гугл Аналитики. Например, при опросе в социальных сетях мы получаем большую долю аудитории с развлекательными интересами, с интересами к играм. Внизу представлены данные по категориям, по сегментам рекламных сетей, по которым есть значимая разница между выборками. Т.е. аудитории из разных каналов отличаются, чтобы получить что-то похожее на интернет-аудиторию, нужно комбинировать все возможные каналы, т.е. они корректируют сдвиги друг друга.

Далее – то, что, наверное, более интересно, чем предыдущие вещи, менее очевидно – сдвиги, связанные с самоотбором, сдвиги, связанные с эффектом площадки и, наконец, сдвиги, которые мы получаем, сравнивая, т.е. то, как мы объясняем для себя разницу в данных между вероятностной выборкой, CATI и опросами в интернете. Она не всегда связана с аудиторией. Мы проанализировали разницу в политических вопросах, в вопросах, связанных с поведением в интернете, в общем, в разных типах и обнаружили несколько эффектов.

Во-первых, разница в поведенческих вопросах, в вопросах на знание, на них ответить относительно легко, она всегда меньше, чем разница между CATI и ривером, чем разница в ответах на оценочные, установочные вопросы. Всегда в онлайн выше доля затрудняющихся. Мы решили, что это эффект коммуникативной ситуации, сейчас поясню, почему. Общение с живым человеком, который тебя удерживает на линии, ждет от тебя ответа, это тебя немного торопит, понимает его содержательно, интерпретирует, ты понимаешь, что он понимает твой ответ содержательно – эти три эффекта играют решающую роль, возможно, даже не всегда выборка, а именно эти три эффекта.

Чувствительность ривера к различным сложностям инструмента дает ряд следствий. Например, мы заметили, что ривер не прощает ошибок в анкете, не прощает ее длины, чрезмерной сложности, неоднозначности, тупости, скучных вопросов. Мы сразу наблюдаем рост доли затрудняющихся, обрывов, снижения качества данных, которые анализируются автоматически по таким параметрам как прокликивание, стрейт лайнеры, зигзагеры, кто работает с онлайн-опросами, понимает, о чем речь.

Онлайн-опрос очень чувствителен к тому, что у респондента гораздо ниже ответственность. У него на том конце трубки нет кого-то, кто ждет ответа, кто реагирует, поэтому гораздо выше доля обрывов, снижения качества данных. Эта ситуация воспринимается как более анонимная, мы наблюдаем это в сензитивных вопросах. При сравнении с CATI и онлайн мы видим, что в онлайн гораздо выше доля неодобряемых ответов на разные вопросы.

У меня, к сожалению, нет времени, чтобы пояснить, чувствительность к длине – понятно, на одном примере попытаюсь чувствительность к сложности вопроса. Мы видим, что в вопросах, где мы просим совершить какое-то более серьезное усилие, что-то припомнить, еще и к этому свое отношение выразить, у нас гораздо больше получается доля затрудняющихся ответить на вопрос – разрыв между долями затрудняющихся в CATI и в ривере. Например, на вопрос, политики, которым я доверяю «Кому вы доверяете из современных российских политиков?» доля затрудняющихся очень высока и в телефонном опросе, и в онлайн опросе. Но разница между долями

гораздо больше в этом вопросе, нежели в вопросе о выборе партии, например. Т.е. в опросе о выборе партии и там, и там доля затрудняющихся сопоставима. Если мы берем более сложный тип вопроса, то сразу увеличивается разрыв между онлайн и CATI. Нормативность и анонимность - в онлайн опросе всегда выше доля неодобряемых ответов.

Пара выводов. Антихрупкость для меня – это постоянная рефлексия. Есть выборка, которая имеет набор сдвигов по ряду значимых переменных, это постоянная оценка того, насколько конкретно эти переменные, по которым сдвинута эта выборка, связаны с твоим предметом исследования, насколько они связаны с возмущающими для твоего предмета переменными. Т.е. это постоянный анализ того, является это ковариата или не является, в общем, постоянная рефлексия любого типа выборки, разные эффекты, анализ эффектов каждого из типа выборки на твой конкретно предмет, такой релятивизм методический, но, мне кажется, именно он наиболее рабочий. Набор рекомендаций по инструментарию, источникам трафика и типам задач, которые можно решать, нужно решать и категорически нельзя решать с помощью онлайн-опросов.

**Реплика из зала:** Исследовательский центр «Контекст», CATI. Вопрос, связанный с тем влиянием интервьюера, о котором вы говорите, на данные. Были ли у вас экспериментальные планы с небелковыми интервьюерами, как это недавно анализировалось в Facebook? Есть ли такая разница с робо CATI?

**Макушева Мария:** Это очень интересная тема. Лично я не занималась анализом данных робо CATI, я знаю, что такие эксперименты ставят. Может быть, в зале присутствуют бывшие мои коллеги из ВЦИОМ, которые делают непосредственно эти опросы. Знаю, что такие эксперименты проводились методические, очень хотела бы сама узнать, есть ли разница и какова она. Это очень интересно, я с вами согласна. Спасибо за вопрос, он важный, оставим без ответа пока.

**Модератор:** Следующий докладчик Александр Романович, тема онлайн-опросов.

**Романович Александр:** Здравствуйте. У меня специфический доклад об онлайн-опросах, сразу скажу, я очень скептически отношусь к онлайн-опросам, скажу больше, я негативно к ним отношусь. Расслабьтесь, кто этим занимается, я имею в виду те опросы, которые проводят СМИ в интернете на своих сайтах, каких-то информационных порталах, а теперь уже и группы в социальных сетях и вообще кто только не проводит, сейчас у нас все стали социологами.

Поднимите руки, кто не любит такие опросы так же, как их люблю я. Есть люди. Остальные, как я понимаю, либо представители СМИ, либо не знаю, кто еще.

Почему я их ненавижу? Я давно в профессии и наблюдаю, как развивался этот сегмент опросов, их еще называют даже не интернет-опросы, а «голосовалки», немножко уничижительно. Лет 10 назад они появились первый раз и проводили их достаточно серьезные СМИ, которые выходили в интернет. Они достаточно скрупулезно подходили к этому вопросу, зачастую даже звонили нам или другим социологам, чтобы узнать, как правильно надо задать вопрос, какие альтернативы ответов поставить. Тогда у меня с ними была одна проблема, я с ними бился, потому что, проведя опрос, они тут же публиковали результаты, допустим, воронежцы считают так-то или россияне считают так-то. Я всегда проводил разъяснительную беседу, ребята, какое имеет отношение ваш опрос на вашем сайте к мнению воронежцев и тем более россиян? Они начинали

мяться и говорить «вообще-то, наше издание читает широкий круг людей, так что можно, наверное, считать, что это отражает мнение жителей города, поскольку город наш достаточно типичный, то, наверное, и россияне думают точно так же». Долго с ними боролся, но оказалось, что это были только цветочки.

Потом пошла вторая волна, можно назвать ее «фейковые результаты опросов». Когда вдруг практически все интернет-ресурсы, информационные ресурсы взорвались какими-то рейтингами, какими-то результатами опросов. Только за прошлый год Воронеж поучаствовал. Был опубликован какой-то рейтинг самых пьющих городов, Воронеж на почетном месте находится, потом курящих, потом людей, которые пользуются обценной лексикой, тут мы тоже молодцы, в топ входим.

Я как социолог, мы проводим много инициативных исследований, я прекрасно знаю, сколько в Воронеже курят, сколько пьют и сколько поминают маму. Но, к сожалению, не могу сравнить с Саратовом, с Самарой, с другими городами, я не знаю, мы на 1 месте или на 52, этим, наверное, пользуются.

Года 2 назад, вот Максим сидит, он, наверное, тоже это на себе почувствовал, был опубликован рейтинг, даже назову, какая это компания «Зум Маркетинг», может быть, вы ее знаете, что я буду стеснительно не называть. Якобы они опросили 40 тысяч человек и Воронеж занял второе место по количеству проживающих мусульман, 32%. Ребята, мне за день позвонило около 30 представителей всевозможных СМИ в основном федеральных, но и местных тоже. За последние 2 часа много раз упоминаемый Сергей Романович Хайкин позвонил, говорит: «Саша, что у вас творится? Откуда у вас столько мусульман?», я говорю: «Сергей Романович, успокойтесь, меньше 1% - 0,8%, конечно, погрешность, все мы понимаем, если с гастарбайтерами, то 1,2%». Не скажу, что это были ягодки, ягодки были дальше.

Сейчас практически каждая группа ВКонтакте или желтые СМИ, которые расплодились, с большим удовольствием почувствовали себя социологами, наша профессия теперь десокрализирована. Эти псевдо-социологи, думали, мы сами можем еще лучше, они вопросы какие-то скучные задают, то ли дело мы сейчас зададим вопрос быстро, оперативно, результат будет супер! Не успевают выступить Жириновский, предположим, со своим предложением «а давайте коронуем Владимира Владимировича и сделаем его главой династии новой царской в России», тут же выходит опрос «Как вы считаете, надо позволить нашему президенту бессрочное президентство», т.е. отменить на два срока «Или же основать ему новую династию монархическую и ввести в монархию?», других вариантов ответа нет, даже затруднились ответить. Ребята не проиграют, какие бы результаты ни получились, хайп им обеспечен – цитируемость, разлетится по всей стране, что 56% россиян хотят, чтобы Путин стал новым монархом, предположим или еще что-нибудь.

Понятно, что те люди, которые посмотрели на это предложение проголосовать как на что-то страшное, у них нет вариантов «ни то, ни другое», «затрудняюсь ответить», они просто не будут голосовать. Но всегда найдутся те люди, которые поставят галочку, а это и надо, нужен хайп, он есть.

Я увлекся, надо переходить к слайдам.

Я в Facebook выложил пост и предложил воронежским СМИ или каким-то группам ВКонтакте, которые занимаются такими голосовалками поучаствовать в эксперименте.

Говорю, ребята, давайте мы проведем супер честное проведение «Квалитас», проведем опрос репрезентативный по городу Воронежу, возьмем вопросы из разных сфер – социально-экономические, культурологические, социально-политические, зададим мы, и вы в точно такой же формулировке зададите на своих интернет-ресурсах. Понимая, что они напрягутся, уж не хочу ли я их вывести на чистую воду. Я им объяснил, что не жду, что результаты должны совпасть, они и не должны совпасть, но зато мы с вами сможем сверить и посмотреть, так сказать, по гамбургскому счету, кто ваша аудитория, куда она смещается, это либералы или консерваторы, или люди с высокими культурными потребностями или со здравым смыслом в этом плане. Они расслабились и около 15 ресурсов изъявили желание, никого специально не просил.

В итоге получил я результаты по 7 СМИ и группам. У всех свои были причины – у кого-то по техническим вопросам, кому-то учредитель не разрешил, была пара ресурсов, где в одном никто не проголосовал, а во втором два или три человека, но, тем не менее, они пытались некие результаты представить.

Тут представлены местные издания, первые четыре – это СМИ, причем, газета «Мое!» онлайн - это самое популярное издание в Воронеже с совершенно сумасшедшим охватом. «Блокнот-Воронеж», «ТВ Губерния», «Берег», «Агентство Бизнес Информация». Как только я окунулся в эту проблему, хотя нам и так понятно, что во всех интернет-опросах невозможно отследить зависимости, мы не можем соцгентов проследить, да много каким проблем, вы знаете.

Я окунулся в следующее, помимо того, что я сказал, еще, оказывается, множество проблем у них. У многих движки настроены так, что они не могут одновременно ставить несколько вопросов для голосования, только один, а я им предоставил 9 вопросов. Невозможно ставить нужное количество альтернатив, они говорят « у нас только 5 и не больше», спрашиваю 2а как вы по выборам в Госдуму, где 7-10 могут быть?», а вот так вот. Невозможно давать множественные ответы, потому что когда они программировали, они не знали, что, оказывается, могут быть множественные ответы, они знали, что либо «да», либо «нет».

Самое печальное, что у многих голосовалки заточены так, что есть вариант «другое», когда его нажимаешь, тебе говорят «а что другое?», свой вариант ответа сразу подверстывается к вопросу. Сам видел в Воронеже, было «За кого вы проголосовали на президентских выборах?», вариант «другое» кто-то написал «обезьяна шимпанзе», естественно, это пошло, я бы это не стал говорить, но Владимир Владимирович занял все-таки первое место, но с небольшим отрывом от этой самой обезьяны. Теперь на втором месте у нас обезьяна, Грудинин только на третьем оказался в этом опросе.

Мы еще обязательно вставили пол. Только два интернет-ресурса смогли дать все вопросы в том числе и пол, и возраст. Это по Облстату распределение Воронежа – мужчины и женщины, все понятно.

Давайте посмотрим при голосовании «Блокнот-Воронеж», тут совершенно другие цифры, мужчин 65%, «Воронеж – любимый город» - 56%, т.е. с точностью да наоборот мужчины и женщины. На удивление оказалось, что голосуют больше в голосовалках мужчины, а не женщины. Уже интересно.

Распределение по возрасту. Опять-таки Облстат для Воронежа, это мы видим, понятно. Голубенький «Квалитас» – тут нашей заслуги нет, мы работали по квотной

выборке, она полностью совпадает с Облстатом, вот, что мы видим: резкий всплеск мужчин 30-49 лет, мы делаем вывод, что больше мужчин среднего возраста, именно они и голосуют.

Было несколько вопросов оценить уходящий год, следующий вопрос – что прогнозируют на год грядущий. Как видим, все оценивают год лучше, кроме «Воронеж-Блокнот». «Блокнот» во многих регионах есть, это достаточно скептически настроенное издание к действующей власти, основывалось как орган «Справедливой России», сейчас немного покусывает не федеральную, но местную власть точно. Только у них, кто голосовал, выбились из общего тренда. Вопрос «Как вы видите 2019 год?», пользователи «Воронеж-Блокнот» более пессимистично оценили.

Мы много лет задаем вопрос «Кому принадлежит реальная власть в городе?». Как видим сейчас, первое – это «Квалитас», т.е. репрезентативный опрос, 32% говорят, областному правительству, городской администрации – будет второй слайд, мы сейчас увидим. Были разные времена. Очень много лет при других губернаторах у нас на первом месте было «В городе царит безвластие» или «Принадлежит власть теневой экономике». Посмотрите, как онлайн-пользователи говорят дельцам теневой экономики – при репрезентативном опросе 14%, при интернет-опросах 55%, 48% и т.д. Кстати, ближе всех к репрезентативному вопросу читатели «Агентства Бизнес Информации», потому что у них публика специфическая – подготовленная, грамотная публика, бизнесмены, топ-менеджеры и т.д., у них ответ «Областному правительству», они знают реальный расклад дел, что теневая экономика у нас давно в загоне, а областное правительство на ее место встало, красота.

Вот политический «За кого вы проголосуете на ближайших выборах?», я взял первые 4 места, не стал всех брать. «Единая Россия» 32%, что бы ни говорили злопыхатели, мы с вами прекрасно знаем, что такой рейтинг у нее и есть, не меньше, по крайней мере, что на выборах она побеждает – это реалии.

Теперь читатели «Воронеж-Блокнот» и «Воронеж – любимый город». Если у нас 32%, то у них 11% и 5% «Единой России», зато коммунисты красавцы, как видите.

Последний вопрос решили задать совершенно нейтральный, это было в декабре накануне Нового Года. Мы спросили «Где вы собираетесь Новый Год провести – дома или в гостях?», тут «Квалитас» результатов больше «Отметить дома», оно и понятно, потому что мы опрашивали всех, в том числе глубоких бабушек, а в интернет-опросе больше молодежи, которая чаще ходит в гости.

Теперь я должен сказать: те, кто верил этим голосовалкам, продолжаете верить или нет?

Спасибо за внимание.

**Реплика из зала:** Скажите, пожалуйста, какое было количество опрошенных онлайн и сколько опрос длился? С каждого ресурса минимальное количество...

**Романович Александр:** Забыл сказать. Когда я окунулся в эту проблему, я вдруг понял, почему, как правило, когда публикуют результаты своих голосований, они не указывают, сколько проголосовало. Тут я на них надавил, оказалось, что у большинства от 30 до 60 человек. Исключения – газета «Мое!», там около 1600 и «Блокнот-Воронеж», там около 1000 человек, остальные стыдливо не показывают.

**Реплика из зала:** Сколько опрос шел по времени?

**Романович Александр:** У «Квалитас» он шел два дня, у ребят где-то неделю. Мы опросили 600 человек методом телефонного интервью, а у всех по-разному – от 50 до 1500 человек.

**Модератор:** Александр, спасибо большое. Мария Гургуцкая, фонд «Социум», Екатеринбург.

**Бурлуцкая Мария:** Здравствуйте! Наверное, хорошо, что мы поменялись местами с Александром, потому что был онлайн, онлайн, мы сейчас пойдем из онлайн прямо на улицу, в самый что ни на есть оффлайн. В ходе своего доклада расскажу наш опыт изучения одного мега-события, но предваряя доклад, рассказ об опыте, хочу сказать, почему мне кажется, что этот вопрос сейчас очень актуален. Мы по опыту собственных исследований видим, с одной стороны, что сейчас растет, по крайней мере, в больших городах досуговая активность и интерес к разного рода ивентам. Они собирают аудиторию, все более разнообразная программа. Лето начинается, наверное, вы сами видите, почти каждые выходные что-то происходит. Я представляю Екатеринбург, фонд «Социум», вы тоже, наверное, видите по своей практике, есть запрос, у заказчиков есть запрос понимания аудитории этих событий, потому что они должны как-то монетизироваться, привлекать партнеров-рекламодателей, понимать, кто к ним приходит, что это за аудитория, насколько она платежеспособна, насколько она интересна.

Об одном из таких исследований я хочу вам рассказать, это URAL MUSIC NIGHT, сейчас будет минутка гордости своим городом и своим регионом. Кто знает про URAL MUSIC NIGHT? Мы в конце июня все должны к нам приехать, это действительно отличное событие. В двух словах, что это такое. Это одна ночь, когда приезжают музыканты, она начинается на каких-то площадках с 4-5, но самая активность начинается в 6 вечера и продолжается до 4 утра в разных местах города, разные жанры, разные музыканты.

Это сайты URAL MUSIC NIGHT, статистика несколько лет, это традиционный летний фестиваль, его придумали в нашем городе, он набирает популярность. Тут 2015 год, была тысяча исполнителей, 43 площадки, 70 тысяч посетителей. В 2018 году, когда у нас проходило исследование, тут уже результаты представлены, но мы в начале пути не знали этого, сколько людей придет, было известно, что 103 площадки и 2000 исполнителей. Это не всегда открытые площадки, это в том числе кафе, библиотеки, кинотеатры, крыши, под мостом, на мосту, на площади, в разных местах, на балконе оперного театра у нас очень популярная площадка. Рекомендую, будете у нас, обязательно приходите.

Это пятница, вечер, один день, когда весь город встает на уши, приезжают люди специально, потому что в том числе топовые музыканты приезжают, это действительно очень интересное событие. Вопрос заказчика и вопрос организаторов фестиваля: кто все эти люди? Количественно они более или менее пытаются сами оценить, но тоже была задача оценить количество, но главное – понять, кто это.

Как формировать выборку в таких условиях? Проблема та еще. На берегу было понятно, что это очень разнообразная публика. Почему? Потому что тут несколько измерений. Во-первых, музыка любая, она очень разножанровая. Федук приезжал, Абакара приезжала, приезжал Кустурица в прошлом году, т.е. понимаете, насколько это



ставит на уши разные сегменты аудитории. Плюс были КСП-шники, в общем, все жанры, т.е. там нет жанровой определенности.

С другой стороны, площадки все в центре более или менее локализованы, но все равно разбросаны, т.е. это пространственная задача. Плюс временная задача, потому что расписание формируется, на одних и тех же площадках меняются исполнители, меняются жанры, это все по времени, естественно. Публика меняется, потому что кто-то идет после работы сразу, кто-то приезжает специально, естественно, что до 4 утра выживают не все и т.д.

Задача была – решить трехмерно. Мы проводили опрос плюс наблюдение, но главное, о чем я буду говорить – о том, как мы реализовывали выборку для опроса. Мы отобрали из 103 площадок, мы очень подробно работали с пространством заранее и со списком тех площадок, которые есть, мы отобрали 66 площадок. С учетом того, что некоторые из них топовые, т.е. мы классифицировали и по жанрам, и по тому, как мы могли спрогнозировать вместе с организаторами, насколько там будет поток. Плюс мы распределили опрос по времени и подкрепили его наблюдением, т.е. отдельно ходили наблюдатели и отдельно ходили интервьюеры по площадкам. Соответственно, мы нашу выборку распределили с учетом, у них это называлось «ключевые площадки» - больше ключевые, остальные меньше, т.е. она была пропорционально разбита по времени, по типам площадок. Конечно, это был типичный уличный опрос, но таким образом мы попытались охватить это событие со всех сторон.

Понятно, что вопрос какой-то репрезентативности тут непросто. Мы попытались придумать процедуру параллельного контроля качества выборки, которую мы реализуем. Для этого было несколько инструментов.

Во-первых, мы контролировали то, насколько хорошо работают интервьюеры, насколько они выдерживают этот график. Сами интервьюеры обязаны были фотографировать эту ситуацию, во-вторых, наблюдатели, которые наблюдали, контролировали интервьюеров. У нас был анализ недостижимости, т.е. отказы они тоже фиксировали по полу, по возрасту, кто отказывается, это как в уличных опросах по месту продаж и т.д. Плюс у нас наблюдение контролировало выборку, потому что наблюдатели тоже фиксировали количественный состав и качественный состав публики по наблюдаемым показателям. Мы потом сверяли данные, полученные в опросе данные, полученные в наблюдении.

Чтобы еще охват, мы дополнительно обратились к сервису Яндекс аудитория, вы, наверное, знаете, есть такая история, он для рекламы придуман, конечно, чтобы локализовать рекламу, но он дает некоторые параметры аудитории – охват, распределение по полу, по возрасту на определенное время в определенном месте, в точке пространства. Пятница вечер обычного дня и пятница вечер, у нас сидел человек и это все настраивал и снимал в точках, где шел фестиваль, мы еще здесь сравнили. Т.о. мы смогли проконтролировать качество выборки. Я не скажу слово «репрезентативность», я на нем споткнулась, вы поняли паузу, качество нашей выборки. Это просто слайд, что мы достигли своих целей и показали заказчику, кто же пришел к нему на фестиваль, в том числе с анализом по площадкам, по ключевым во всяком случае.

Проблемы, конечно, есть. Назову две, с которыми мы столкнулись. Одна носит, скорее, организационный характер, потому что реализация подобного рода исследований потребовала вывода на улицы огромного количества сотрудников – не только интервьюеров, но и все сотрудники пошли. Плюс в том, что все получили удовольствие от самого мероприятия.

Методическая проблема. Тут то, что мы попытались решить, но пока не очень понятно. Ясно, что требование одинаковая вероятность попадания в выборку, она не реализуется, потому что мы просто встречаем каких-то людей, которые активнее всего и дольше всего ходят, у них вероятность попадания в нашу выборку больше с разнообразными вкусами. Но это пока не решается, наверное. У нас этот прогнозируемый сдвиг в сторону самых активных посетителей фестиваля. Но, наверное, с точки зрения заказчика это и не так плохо, потому что они контактируют активнее с теми материалами, которые предоставляют партнеры и рекламодатели. На этом все.

**Реплика из зала:** Скажите, откуда вы, какое агентство представляете.

**Бурлуцкая Мария:** Фонд «Социум», Екатеринбург.

**Реплика из зала:** Что мешало заказчику во время продажи билетов или это бесплатное мероприятие?

**Бурлуцкая Мария:** Бесплатное мероприятие. Мы в этом году будем повторять, у них некоторые топовые площадки при регистрации, мы будем еще этот канал использовать. У нас были всякие гуляния на Чемпионат мира, тоже там опросы велись, это совершенно свободная история – фестиваль, когда люди приходят, уходят, там не только екатеринбуржцы, около 10% потока было приезжих, которых мы тоже никаким пост-опросом не можем выявить, а количество туристов нарастает. Это только опрос здесь и сейчас, в этом и проблема.

**Реплика из зала:** Вы публиковали, открывали эти данные?

**Бурлуцкая Мария:** Они открыты, они опубликованы, насколько понимаю, на нашем сайте, на сайте URAL MUSIC NIGHT, они это не скрывают, потому что им важно партнеров привлекать, наоборот.

**Реплика из зала:** Мария, хотела уточнить пример повторности, помимо сложности вообще расчетов и планирования выборки, как решался вопрос свободного хождения, вопрос учета наблюдателями людей?

**Бурлуцкая Мария:** Повторности попадания. При расчете количества аудитории мы это учитывали в каком смысле – у нас был, естественно, вопрос в интервью, сколько площадок посетили, у нас там сложная была система, мы и это учитывали. Мы понимали, сколько разные люди попадались на глаза наблюдателям, плюс я сейчас не воспроизведу, но мы придумали систему подсчета при наблюдении – количество людей, сколько уходит за 5 минут, сколько приходит и вот три раза на площадке. А в смысл опроса, что они в большей вероятности попадают, я уже сказала, тут решения особого не существует. У нас была мысль перевзвешивать в сторону тех, кто меньше... вот этих количество площадок посещенных, но как-то это нам показалось некорректным потом. Понятно, о чем я? Долю тех, кто много площадок посетил, они ответили, что много, активная аудитория и увеличили долю тех в выборке, в ответах.

**Реплика из зала:** Т.е. ограничение, которое было обозначено для заказчика, что мы будем считать так и это надо учитывать.

**Бурлуцкая Мария:** Да, да. Мы учитываем, что все равно у нас сдвигка в сторону активной аудитории какая-то идет, и можно предположить, что дело не только в выборке, что они еще и активнее соглашаются отвечать, самые активные, но это предположение уже. А выборка – математически понятно, если человек в 18:00 начал и в 04:00 закончил и много площадок прошел, он много раз нам попался. Мы в поиске, тут решения нет.

**Казаков Илья:** Илья Казаков, Москва. Спасибо, первый раз вижу настолько технологически проработанную выборку. У меня вопрос по количеству вопросов в анкете, приблизительно опишите их характер и решалась ли как-то проблема шумов.

**Бурлуцкая Мария:** Что был шумно?

**Казаков Илья:** Нет, шумы в социологическом смысле. Мало того, что шум в качестве звука и шум респондента навеселе.

**Бурлуцкая Мария:** Вы знаете, это удивительная история – там вообще почти не пьют! Это правда! Во-первых, запрещена продажа, но и с собой особо не приносят. Я вам еще раз могу предложить, вы приезжайте, таких хороших лиц на редкость, понятно. Это лирика, что называется. Мы это, конечно, обсуждали, но тут у меня ответа, кроме того, что какая-то безопасность интервьюера должна быть и он к таким не подойдет, хотя у них было правило систематического отбора. Они стояли и смотрели, кто как бы отходит от площадки, третьего ил пятого считали и к нему обращались с предложением принять участие в опросе. Анкета была не очень большая, но и не совсем маленькая. Я количество вопросов вам не скажу, в общем, она была на трех листах, если я правильно помню, там был блок социально-демографический, блок по расходам и доходам довольно-таки большой, в том числе расходов на самом фестивале, потому что это коммерческая история все-таки, был блок с кем пришел, описание, практика посещения фестиваля, плюс был вопрос на какие еще они ходят подобного рода мероприятия в городе, но это уже внутренняя история, чтобы это еще с какими-то событиями сопоставить.

**Реплика из зала:** Почему на бумаге опрос?

**Бурлуцкая Мария:** Почему не на планшете? Тут, во-первых, количество и потом мы не могли рисковать. Знаете, что-то сбойнет, а событие это не повторить. Удобнее – второе, ночь – третье, в общем, и безопасность. Мы 66 площадок опрашивали, у нас, конечно, был маршрут, но было несколько рядом стоящих интервьюеров, в общем, там человек 30 работало без наблюдателей. Не могли мы рисковать. Мы обсуждали, конечно, планшетный опрос.

**Реплика из зала:** В села посылают с планшетами.

**Бурлуцкая Мария:** Я понимаю, но тут все равно толпы. По организационным соображениям, скажем так.

**Реплика из зала:** Последний вопрос. У вас на одном слайде было указано, что вы изучали недостижимость, неответы красненьким как-то так. Можете рассказать?

**Бурлуцкая Мария:** Это типичная история, поэтому не стала комментировать, как и в торговом центре когда ставим. Наверху стоит человек, вот он третьего спросил «давайте ответим», говорит «нет, я не буду отвечать», дальше он отмечает пол, возраст примерно на глаз отказавшегося и, по-моему, у нас в этом опросе было, он с компанией пришел или с кем. Дальше это все отмечает, отмечает, ага, согласился, т.е. к каждой анкете. Потом мы это пересматриваем, перепроверяем выборку, например, насколько у нас мужчины в отказ пошли по сравнению со структурой. Т.е. структуру аудитории мы пересчитываем на отказавшихся по наблюдаемым параметрам.

**Реплика из зала:** А потом и сами это компенсируете?

**Бурлуцкая Мария:** В перевзвешивании данных иногда да, иногда нет. В этом мы не компенсировали, мы специально делали такой слайд, но тут и не было большого. Мы же тоже, когда отбирали событие, про которое будем рассказывать, мы это учитывали. Это успешное в этом смысле, потому что опять же аудитория очень лояльна к фестивалю, тут больших перекосов не было, даже особо много отказов не было. Поэтому нам и тут было удобно работать, и пьяных не было, и вообще все было очень хорошо с этой точки зрения. Можно было с планшетами, наверное, идти.

**Реплика из зала:** И еще немного фонда «Социум», Екатеринбург. Я буду очень краткой, хочу познакомить вас с одним интересным кейсом из нашей практики. Мы с 2005 года проводим омнибусные опросы ежемесячно, выборка 600 и на протяжении всех этих лет мы проделали несколько кульбитов в погоне за репрезентативностью, вот об этом я и хотела рассказать.

Начинали мы в 2005 году, выборка была 600, был дедовский телефонный опрос. Т.е. наши интервьюеры работали по проводным телефонам, мы генерировали базу данных по Екатеринбургу, базу телефонов, генерировали номера, квотировали отбор по полу и возрасту в домохозяйствах и таким образом мы работали до 2012 года. В 2012 году мы перешли на уличный опрос по причине, о которой я скажу позже. Выборка та же, стартовые точки в 16-ти микрорайонах города, при этом каждый микрорайон был разбит на квадранты, и осуществлялась каждый месяц ротация этих квадрантов. Т.е. мы постепенно в течение нескольких месяцев закрывали практически все маленькие планировочные микрорайоны города. Обязательно в уличниках квотирование по полу и возрасту. В 2016 году мы снова переходим на телефонные опросы, но уже через конденсатор, выборка та же, генерация номеров мобильных телефонов плюс проводные не более 25%. С 2016 года мы работаем без квотирования на заключительном этапе, используем исключительно, как нам кажется, вероятностную выборку.

Что происходило, почему мы осуществили эти два кульбита? Как мы работали с заказчиками? Одно дело, когда вы понимаете, что по каким-то причинам выборка, к которой вы привыкли и с которой вы работаете на протяжении нескольких лет, а это была телефонная по проводным номерам выборка, она перестала работать. Мы понимаем, что мы даем какие-то странные данные, которые не совпадают со статистикой заказчиков, в частности это крупные банки, это девелоперы, которые говорят «ребята, у нас другие цифры, что-то у вас происходит не то». Хотя социологи не любят больших перепадов, наши цифры достаточно хорошо себя чувствовали, но на той выборке, на которой мы работали.

Мы озадачились вопросом, что же происходит с нашей выборкой, появилось ощущение, что что-то некачественно, с какими-то смещениями мы сталкиваемся. Мы сформулировали гипотезу: в телефонном интервью мы теряем статусное активное население, т.е. горожан, которые практически не пользуются проводными телефонами, они просто к ним не подходят. Естественно, просто так перейти на другой формат выборки, опыт показывает и все это знают ресечеры, что при изменении методики скажут показатели, скажут результаты, всегда они дают значительное изменение показателей, и заказчики на такое изменение показателей будут реагировать негативно. Мы это знали заранее, поэтому мы обратились к формату методических экспериментов для того, чтобы не просто сказать заказчикам «так, ребята, все, переключаемся», а объяснить им с цифрами в руках, почему это необходимо делать.

Мы очень осторожно на протяжении нескольких замеров, т.е. несколько месяцев начали вводить уличные опросы и каждый раз сопоставляли. Т.е. мы попытались проверить гипотезу, контролируя при этом пол и возраст и в телефоннике, и в уличнике, посмотрели, сопоставили несколько статусных показателей, таких как уровень образования, уровень дохода, возрастные когорты. А также взяли несколько поведенческих показателей из вопросов наших постоянных заказчиков, постоянных подписчиком омнибуса с тем, чтобы проверить, как работает уличник и как работает телефонник.

Наша гипотеза оказалась верной, мы ее подтвердили, потому что эти и ряд других показателей свидетельствовали о том, что да, статусное население в уличнике представлено лучше формата телефонника, который мы проводили все эти годы.

С этими цифрами мы провели серию переговоров с заказчиками и смогли им доказать, что такой переход необходим, что смещение, которое мы получаем, работая в формате прежнего телефонника, существенно, и мы даем искаженные данные за счет всех этих эффектов. Заказчики пошли нам навстречу, мы обнулили омнибус, а вы понимаете, что омнибусные исследования – это постоянный мониторинг, т.е. они вынуждены были поставить точку под трендами и начать с начала уже с другими показателями. Это был первый этап.

Второй этап. Мы проработали до 2016 года в формате уличника и постепенно наша удовлетворенность этим переходом снижалась, мы понимали, что сталкиваемся с еще одним эффектом, с еще одной проблемой, которая нарастала. В 2016 году мы поняли, что результаты, которые мы выдаем нашим заказчикам, нам кажутся сомнительными, т.е. есть какие-то внутренние эффекты, которые снова заставили нас задуматься и перейти к практике методических экспериментов. А именно мы сформулировали гипотезу, что на улицах мы теряем мобильную и активно работающую часть населения, те социальные группы, которые пользуются личным автотранспортом, редко бывают на улицах, плюс все-таки уличник не репрезентативен, мы понимали, что получаем латентные смещения, эффекты, связанные прежде всего с человеческим фактором, т.е. работа интервьюера в формате уличника.

В 2016 году мы сделали еще один поворот, мы, сохраняя часть выборки уличникам, перевели значительную долю, т.е. сделали дополнительный опрос при помощи КТ, у нас к тому времени как раз появился КТ центр, мы сделали уличный опрос. Посмотрите на этот интересный график, как я уже сказала, гипотеза была связана с тем,

что мы теряем социальные слои, которые по улице не ходят, которые ездят на машинах, если вопрос о том, есть ли у вас машина в домохозяйстве практически не показал существенных различий между результатами уличника и телефонника, то вот количество машин в домохозяйстве оказалось весьма хорошо работающим индикатором. Мы обнаружили, что участвующие в телефоннике горожане обладают гораздо большим количеством машин в домохозяйстве, на основании этого мы сделали вывод, что наша гипотеза подтверждается. С цифрами в руках, вы можете посмотреть раскладку по полу, в уличнике у нас провисала молодая возрастная когорта, в уличнике у нас гораздо больше была представлена когорта старшего поколения.

Посмотрите на интересный вопрос «Намерение улучшить жилищные условия», уличник давал 14%, телефонник дал 27,8%, т.е. очевидно, что телефонник как раз восполняет эту часть аудитории, которую мы теряли в уличнике. Таким образом мы после серии переговоров с заказчиками совершили еще один переход – переход на КТ, сейчас мы и работаем в формате телефонного КТ омнибуса, пока довольны, но проблемы есть.

Когда заказчики произносят слово «репрезентативность», я все равно внутренне вздрагиваю, потому что в том числе и КТ опросы дают те эффекты, в частности о них Мария говорила, это и неодинаковая доступность разных социальных групп для интервьюеров, это предпочтения интервьюеров в выборе времени проведения опроса, правда, это уличника касается, это нарушения отборов респондентов домохозяйств, смещение из-за отказов от участия в опросе, т.е. эффекты можно множить.

Поэтому я слово «репрезентативность» оставляю под большим вопросом, и мы продолжаем анализировать, продолжаем проводить методические эксперименты, я считаю, что методические эксперименты сегодня – это очень надежный способ как раз оценки различного рода эффектов, связанных с выборкой, с ее изучением сначала для себя, а потом обязательно должно это все обсуждаться с заказчиками, потому что я считаю, что исследовательский процесс должен быть прозрачным. Спасибо.

**Модератор:** Никита Анатольевич, пожалуйста. Без презентации.

**Тюриков Никита:** Добрый день, коллеги, Тюриков Никита Анатольевич, «Центр общественно-политических проектов и коммуникации». Я выступаю без презентации, у меня своеобразное выступление непрофессионала о непрофессионалах для профессионалов, потому что здесь все социологи, как я понимаю, я не социолог, но часто сталкиваться приходится. До меня выступал коллега, который закончил свою речь фразой «Вы верите этим опросам?», мне кажется, он просто не туда ее адресовал. Здесь таким странным сайтовым опросам мало, кто верит, но, тем не менее, тенденция формирования общественного мнения и использования соцопросов во всех формах – дорога с двусторонним движением. Наряду с профессиональными социологами в ней очень активно принимают участие практически все представители гуманитарных профессий. Поэтому хотим мы этого, не хотим, все вот эти сайтовые, телефонные, случайные и прочие опросы, сделанные непрофессионалами, существовать будут. Вопрос возникает: что с ними делать, как с ними работать?

У нас возник кейс, мы работали с картами нарушения на выборах. В стране у нас существует две базовых карты: карта нарушений – это карта нарушений орг и карта сообщений – это портал на 24.ру. Вроде бы, показалось на первый взгляд, что здесь

можно пойти чисто статистическим подходом, предположить, что это случайная выборка однородная по одному показателю, т.е. сообщение гражданина о нарушении на выборах, и попробовать с этим что-то сделать.

Здесь начинается моя любимая история – это не история успеха, наоборот, история неуспеха, потому что она более показательна в точки зрения того, а как же с этим быть. Брали Единый День голосования 2018 года, 4800 избирательных кампаний и 930 сообщений, на первый взгляд, цифра, позволяющая с этим массивом работать. Но сразу сработал у нас первый ограничитель – а действительно ли эти сообщения можно оценить с точки зрения их однородности? Первое, что проходило, это верификация. Тут оба портала помогли, потому что карта нарушений размещает эти данные без всякой верификации, второй – карта сообщений интерполирует эти данные и верифицирует одновременно. Плюс были парамельные работы. По меньшей мере, треть этих сообщений таковыми не являются, т.е. их надо из случайной выборки удалять, потому что они нам не помогают.

Получается, что фактически весь этот массив, хорошо, он есть физически, если вы хотите его использовать для какой-то экстраполяции данных, интерпретации данных, чего угодно – уже приходится перебирать руками.

Второй момент, который, вроде бы, для непрофессионала облегчает жизнь, для социолога, возможно, наоборот, усложняет – это наличие внутренних фальсификаторов. Когда начинаем проверять эти внутренние классификаторы, выясняется, что они или настолько размыты, что туда попадает все, что можно, либо они, наоборот, настолько сужены, что приходится их перевзвешивать, с внутренними классификаторами работа тоже практически невозможна в изначальном виде, как это сделано.

В качестве примера 18 марта 2018 года выборы президента. 337 сообщений, нарушение прав наблюдателей, членом комиссий, представителей СМИ. Страшная цифра, можно пересчитать из выборки проценты и сказать, что по всей стране идет кошмар и ужас. Самый вопиющий момент – 26 случаев давления и удаления наблюдателей. Когда начинаешь каждый поштучно разбирать, выясняется, что в результате у нас остается 15 эпизодов. Тут же мы смотрим данные ЦИК, которые заявляла Элла Памфилова, 12 эпизодов. Получается, что не то, что статистика или социология, тут здравый смысл подсказываем, что можно взять официальные данные, проверяем все, но приходим к тому, что официальные данные не врут, а зачем нам тогда выборку с этой классификацией использовать. Немного странная история, не получается в этом ничего делать.

Последний момент – то, что в мультфильме про одну Кунг-фу Панду говорила умная Черепаха, что случайности случайные, а потом народ шутил, что случайности не случайны. Здесь много приводилось примеров разных вариантов активизма, в том числе и в вопросах, разные внутренние мотивации людей, которые принимают участие в телефонных опросах, в интернет-опросах, в том числе подготовленных профессиональными социологами. Когда речь идет не об опросе, а о некой формирующейся совокупности данных, которые не отработаны профессиональными социологами для социологических задач, а отработаны для других задач, то фактор активизма становится абсолютно доминирующим.

Например, срез по 2018 году показывает, что у нас в тройке лидеров по нарушению на выборах, первое и второе места, может, вы догадаетесь, а знаете, кто был на третьем месте? Никогда не догадаетесь - Ивановская область. Москва – понятно, Самара – тоже понятно, Ивановская область почему? Начинаешь разбираться и выясняется, что там просто неравнодушные люди. И сразу картинка меняется.

Мы в результате приходим к выводу, что эта случайная выборка не случайная, а мы получаем конкретную совокупность. Все, что мы дальше с этой совокупностью делаем, это точно также, как представлен нам воронежский опрос «а почему у нас воронежская неделя 60% против?», мы можем анализировать только то, что у нас есть в качестве числовых показателей. Но проблема-то в другом совершенно. Проблема заключается в том, что на основании таких данных в том числе выстраиваются экстраполяции на... и не только вот это касается нашего кейса с нарушениями, а выстраиваются экстраполяции на страну в целом, на молодежь в целом, все сталкивались с исследованиями, когда на основании пары тройки ВУЗов в паре субъектов пытаются экстраполировать.

Поэтому для меня как не для профессионала, но как для человека, который часто сталкивается с разными данными - и социологическими и квази-социологическими хотелось бы, чтобы социологи более активно предлагали и доступные методы какой-то репрезентации и какой-то проверки выборок и в том числе сами это активно продвигали. Потому что верю или не верю я опросу – это неправильный подход абсолютно, люди будут верить в то, что им хочется верить, а вот большинству других непрофессионалов сделать правильные выводы и правильные исследования из имеющихся данных и четко понимать, почему какие-то массивы можно использовать, а какие-то надо сразу выбросить, здесь очень большая проблема именно для несоциологов.

С понимаю, что у вас на кухне все плохо, но мы как профессионалы на кухню и не ходим, мы в зал обеденный садимся и ждем, пока принесут, если пойдем на кухню, то точно застрелимся. Но, тем не менее, такого инструментария популяризаторского и объяснительного, к сожалению, крайне мало, а хотелось бы. Спасибо.

**Реплика из зала:** Зачем вы хотели проанализировать эти карты?

**Тюриков Никита:** У нас, исходя из этой карты, некоторые исследовательские центры делают выводы о политическом климате, уровне демократизации, отношении к власти, к демократии, к выборам и сразу экстраполируют это на страну в целом. Люди делают это по разным причинам, предположим, они добросовестные исследователи. Задача была – проверить, можно ли на основании этого массива делать добросовестные исследования. Если на основании этого массива нельзя делать добросовестные исследования, т.е. цифра высчитывается, а дальше идет простая экстраполяция - 60% избирательных участков по России в целом... В России 97 тысяч избирательных участков, а мы 13 конкретных случаев начинаем... И самим больно, и глупость трещит.

**Реплика из зала:** А поправить эти данные нельзя вручную или их там слишком много?

**Тюриков Никита:** Нет, поправить эти данные вручную можно, но возникает вопрос «зачем?» Я вас на примере с удалением наблюдателей показал, что мы поправили данные вручную и получили официальные данные ЦИКа. Мы поправим еще кучу данных



вручную, получим близкое. Получается, что вот эту штуку надо всю вручную перебрать, чтобы доказать то, что все совпадет.

**Реплика из зала:** Вы же не знаете, что получится.

**Тюриков Никита:** Я один кейс в пример привел, по нескольким кейсам перебирали, получается, что власти, конечно, немного врут, но не настолько, насколько им приписывают. Эти карты по сути своей свою главную задачу выполняют, они обозначают болевые точки, дальше с болевыми точками разбираются, если есть нарушение, его надо пресекать, устранять, если нарушения нет, то ничего не надо. Тот, кто пытается делать из этого глобальные выводы, тот садится в лужу. Для чего все это делалось? Дать сигнал, чтобы разобраться, чтобы устранить конкретное нарушение, не для глобальных выводов. Кто пытается делать на этом глобальные выводы, садится в лужу в конкретном случае, причем, очень глубоко, но, тем не менее, народ не останавливается.

**Модератор:** Немного про то, что у социологов все нормально, загляните в папку, что вам дали, там лежит буклет без наименования компании, которая проводит опросы по онлайн панели, запустили омнибус, результаты через 24 часа, запуск каждый вторник и четверг, профильные данные, медиапотребление, репрезентативная выборка. Ну, зачем?

**Реплика из зала:** Сейчас уже немодно поквартирники проводить?

**Модератор:** К сожалению, доклад про поквартирные опросы мы не успели, мы старались вдвоем с Ольгой стимулировать докладчиков, чтобы они укладывались, но не получилось. Приблизился дедлайн нашей секции, но много людей осталось. Нас просили зачитать информацию, что у «Российской исследовательской недели» есть Telegram-канал, подписывайтесь, ставьте лайки, там самые интересные тезисы докладчиков, социологические данные, маркетинговые инсайты, эксклюзивные слайды спикеров, подписывайтесь.

Давайте проголосуем, нас пока никто не гонит отсюда, но мы выбиваемся из регламента. Вы хотите послушать Надежду? Надежда перед началом попросила меня вынесли ее за регламент. Я пытался, честно.

Доклад про поквартирный опрос и почему там все так плохо. Вчера сидели в кафе, представители крупной исследовательской компании без названия этой компании сказали, что они сами ходили в поле, они не понимают, как можно по методикам, которые прописаны, качественно провести поквартирный опрос.

**Галиева Надежда:** Мы поквартирники редко проводим, специализируемся на телефоне, но у нас был в прошлом году один поквартирник, расскажу про него. Это было исследование в двух средних городах России – Великие Луки и Мичуринск, выборка по 60 человек в каждом городе, население мы опрашивали во всех УИКах этих городов, соответственно, число людей опрашивали пропорционально количеству избирателей, в основном в пределах 10-13 человек на одном пике было опрошено. Мы не задавали маршрут или начальную точку, мы случайным образом выбрали две точки в каждом УИКе, и сказали интервьюерам, что эти две точки они обязаны посетить – зайти, спросить, поставить результат посещения. В пределах УИКа можете опрашивать где угодно, на придомовой территории в том числе, единственное, условия, чтобы человек

сказал адрес проживания, чтобы верифицировать, что он относится к этому УИКу. Соответственно, один респондент от домохозяйства и возраст.

Когда мы получили массив, я на различных сайтах нашла пару переменных, которые характеризуют дома, в которых опрашивали людей. У меня получились переменные: тип постройки – многоэтажная, частная, этажность и только для многоэтажного жилья – год постройки дома, материал несущих стен – панель, кирпич, дерево и комфортность, некомфортность жилья. К некомфортному жилью я отнесла аварийные дома, малосемейки, общежития и все года, старше 1960 года постройки. Эта вся информация открыта, ее можно найти на сайтах ЖКХ или реформе ЖКХ, я привела пару сайтов, которыми пользовалась.

Что получилось? Типичный дом опроса или типичный дом респондента. В Великих Луках получилась кирпичная пятиэтажка 80-ых годов постройки, это фото из самих городов, мы туда ездили с методическим аудитом. В Мичуринске – тоже кирпичная пятиэтажка и панельная, они в разной степени представлены, год постройки 60-ые, 70-ые, 80-ые года. Ни для кого не секрет, что интервьюеры не любят частный сектор, опрашивать в частном секторе ресурсозатратно и при этом малоэффективно.

Наша вся выборка смещается к многоэтажной застройке. Все бы ничего, но в средних, малых тем более города в России много частного сектора. Если Великие Луки, мы же там были, визуальное состоит из пятиэтажек, там не так много частного сектора, то Мичуринск практически одноэтажный, там очень-очень много частного сектора, наши люди, которые там проводили опрос, нам все время очень жаловались, как им много приходится ходить. В чем забавный факт, назову это так, в частном секторе опрашивали больше мужчин. Думаю, все знают, что мужчины – сложная категория для опроса, там вот именно среди всех опрошенных в частном секторе в Мичуринске 73 было мужчины, в Великих Луках 60% было мужчин, соответственно, в многоэтажной застройке 40-45%. Это немного спекуляция, потому что объем выборки небольшой, для Мичуринска человек 100, а для Великих Лук 50 что ли, т.е. здесь возможна ошибка довольно большая.

Дальше я построила кучу кросс-таблиц с этими новыми переменными, характеризующими тип постройки, всем соцдемом, всеми содержательными вопросами анкеты. Что я нашла? Ничего я не нашла, т.е. никаких зависимостей я не выявила. Оговорюсь, что я не использовала сложные методы, это все построено на АФИ квадрате.

Теперь про попытку сравнения запроектованной выборки и той выборки, которая была реализована интервьюерами. Как я говорила, у интервьюеров было две обязательные к посещению точки, к сожалению или к счастью, некоторые интервьюеры эти обязательные точки игнорировали, точнее, фальсифицировали – они подходили к дому и не заходы отмечали «никого нет дома», «отказ», «невозможно попасть на территорию» и т.д. Это случалось тогда, когда обязательная точка попадала именно в частный сектор, я уже говорила, что не любят туда ходить, если точка в частный сектор выпадала, они туда даже не заходили. Если обязательная точка выпадала на многоэтажную постройку, то, как минимум, одно интервью там было взято.

Чтобы сравнить запроектованную и реализованную выборки, анкетам был присвоен код соответствия запроектованной выборке. Код построен на сравнении по основным параметрам – тип постройки, материал, год постройки домов обязательных

точке и домов, в которые интервьюеры зашли. Если в УИКе дома, в которых были взяты интервью, соответствовали домам, вот эти обязательным точкам, т.е. обязательные точки – два дома в частном секторе и у нас интервью тоже взяты в частном секторе, таким анкетам присваивался код «соответствует выборке». Если не соответствуют дома, обязательные точки у нас – частный сектор, а все интервью взяты в многоэтажках, это «не соответствует выборке», иногда бывало «частично соответствует» - и там, и там взято.

Как вы можете видеть, половина анкет в Великих Луках была взята только на УИКах соответствующих запроектованной выборке, а в Мичуринске только 38% анкет была взята в УИКах, соответствующих запроектованной выборке. Думаю, это связано с тем, что Мичуринск более одноэтажный, и интервьюеры не ходили по частному сектору, избегали его, как могли.

Потом я точно также, как и в первом случае, построили кросс-табы с этой переменной и со всеми остальными переменными, несколько переменных привела – это уровни значимости. Как видите. Нет зависимости, только с полом в Мичуринске зависимость есть сильная, там мужчин опросили в УИКах, соответствующих выборке, а женщин в УИКах, не соответствующих выборке. Похоже, надо, чтобы интервьюеры все-таки следовали нашим инструкциям, когда они будут сложно достижимые группы опрашивать как-то быстрее и эффективнее.

Некоторые выводы. Получается, что респонденты, опрошенные в разных типах жилья, т.е. в многоэтажной застройке, в частном секторе, в комфортном или некомфортном жилье, в домах разных годов постройки, они в целом не отличаются друг от друга ни по соцдему, ни по содержательным вопросам анкеты. Респонденты, опрошенные в УИКах, соответствующих и несоответствующих запроектованной выборке тоже не отличаются.

Что это значит? У меня несколько предположений и вопросов. Возможно, это означает, что адрес проживания не является дифференцирующей переменной. Мы активно строим наши выборки по этим УИКах, чтобы равномерное распределение было по городу, а, может быть, достаточно опросить в двух-трех домах и вот тебе репрезентация города. Либо в опросах соглашались участвовать только определенные люди, это как бы понятно, но, может быть, это не только определенные люди по каким-то своим отличительным особенностям характера, т.е. нравится им поговорить с людьми, может быть, они – определенные люди с точки зрения соцдема или, может быть, какое-то мнение у них одинаковое по всем вопросам или по многим вопросам, либо сами интервьюеры выбирают какую-то определенную категорию граждан. Грубо говоря, даже в самом элитном районе, если ему выпал такой УИК, он найдет самого вредного и самого подходящего под нас, под среднестатистического человека. У меня все. Спасибо.

**Реплика из зала:** Сколько там жителей всего?

**Галиева Надежда:** Это средние города, там от 50 до 100 тысяч, по-моему, около 100.

**Реплика из зала:** Т.е. можно предположить что для них факт территориальный не играет никакой роль.

**Галиева Надежда:** Я так предположила, да.

**Реплика из зала:** А если бы там было, предположим, 400 тысяч, то была бы другая гипотеза?

**Галиева Надежда:** Не знаю.

**Реплика из зала:** Как вы аудио записали?

**Галиева Надежда:** Это опрос на планшетах.

**Реплика из зала:** Анкетами закрыли?

**Галиева Надежда:** Вы имеете в виду, что я вам показала. Да, это интервьюеры, они, к нашему счастью, зачастую забывают, что идет аудио или что оно начинается с первого опроса, поэтому они запустили анкету, пока едут куда-то.

**Модератор:** Коллеги, еще вопросы? Мне очень совестно, что мы вас задержали так, спасибо вам, что досидели до конца. Хорошего вечера, до свидания!