

Исследование возможности использования online-опросов для репрезентации больших социальных групп

Настоящая статья представляет собой сравнительный анализ результатов двух исследований, проведенных среди жителей 12 российских городов и трех регионов – Татарстана, Красноярского края и Ямало-ненецкого автономного округа. На основе сравнения массивов рассматриваются возможности использования онлайн-опросов больших социальных групп.

Ключевые слова: методология, сравнительный анализ, online, регионы, социально-демографические характеристики

Как показывает практика, предпочтение он-лайн опросам отдается в случае, когда необходимо охватить либо очень узкую, либо труднодостижимую целевую группу. Однако, все чаще исследователи задумываются над тем, насколько эффективными он-лайн опросы могут оказаться тогда, когда требуется охват самых широких социальных групп и даже шире – репрезентирующих население в целом – опросов. Для этой цели необходимо сопоставить данные он- и офф-лайн опросов, проведенных с помощью сходного инструментария.

В ноябре 2008 года Ассоциации российских социологических центров по собственной инициативе провела первую волну исследования «Межрегиональный индекс кризисного сознания». Опросы во всех городах и регионах репрезентировали генеральную совокупность их населения. В городах и регионах было опрошено от 300 до 1500 человек, данные по объему выборки в каждом городе приведены в таблице 1.

¹ Звоновский Владимир Борисович - кандидат социологических наук, председатель правления Фонда социальных исследований. т.(846) 2296733 (многоканальный)
E-mail: Zvb@fond.sama.ru

В конце января 2009 года компания OMI-Russia по идентичному инструментарию провела опрос участников своей панели, живущих в этих же городах и регионах, а также в Москве, Уфе и Омске. Опрос охватил только тех участников панели, которые выбрали в шкале по доходу варианты: «Мы можем без труда приобретать вещи длительного пользования, однако для нас затруднительно приобретать действительно дорогие вещи» и «Мы можем позволить себе достаточно дорогостоящие вещи - квартиру, дачу и многое другое». Очевидно, что этот опрос не был репрезентативным и его можно условно охарактеризовать как опрос представителей людей среднего достатка. Количество опрошенных в каждом из этих городов также приведено в таблице 1.

Таблица 1.

		МИКС-1	МИКС-online
1	Москва	-	405
2	Архангельск	595	9
3	Волгоград	702	39
4	Воронеж	705	45
5	Кострома	410	4
6	Красноярский край	450	14
7	Нижний Новгород	300	39
8	Омск	-	35
9	Республика Татарстан	1500	34
10	Ростов-на-Дону	315	33
11	Самара	500	63
12	Санкт-Петербург	500	168
13	Саратов	387	27
14	Уфа	-	32
15	Чебоксары	505	9
16	Челябинск	400	40
17	Ямало-Ненецкий автономный округ	300	4
18	Ярославль	441	12
	ВСЕГО	8010	1012

Результаты обоих измерений были объединены в два массива данных: МИКС-1 и МИКС-online соответственно. Затем в методических целях были сравнены распределение ответов на вопросы социально-демографического блока в обоих массивах. Отличия между ними в основном были статистически значимыми. Очевидно, однако, что источник различий

находится не только среди социальных факторов, влияющих на изучаемые переменные, но и среди факторов, связанных со способом сбора данных.

Различия между двумя массивами в самом общем виде можно разделить на три большие группы.

1. Неустранимые. Экономический кризис, отношение к которому различных социальных групп было предметом исследования, в период сентября-февраля имел высокую динамику в России. А поскольку между замером МИКС-1 и замером МИКС-online прошло более двух месяцев, то следует допустить, что обнаруженные изменения могли быть связаны не с какими-либо социальными факторами, или даже способами сбора данных, а сформированы изменениями в социальной реальности. При этом, мы не можем оценить вклад реально имевшейся динамики и каким-либо образом корректно связать отличия между двумя измерениями с изменениями в экономической ситуации в стране и мире за прошедшие два месяца. Поэтому при анализе факторов, дифференцирующих два массива, мы не будем использовать отношение к кризису, которое очевидным образом могло измениться за время между исследованиями, а сфокусируемся на социально-демографических признаках.
2. Частично устранимые. Как видно из таблицы 1, в массиве МИКС-online - иная пропорция между регионами и городами исследования. Разумеется, различия можно устранить, перевзвесив массив, но весовые коэффициенты в большинстве случаев будут недопустимо большими. По этой причине мы будем исходить из того, что оба массива нерасчленимы на региональные подвыборки и выборка МИКС-1, и выборка МИКС-online, взятые в целом, репрезентируют население совокупности городов, где проходило исследование.
3. Устранимые. Как видно из таблицы 1, в массиве МИКС-online значительна доля жителей Москвы, в то время как в выборке МИКС-1 их не было вовсе. Очевидно, что значительная их доля в одной выборке и отсутствие в другой,

могут в существенной мере и сформировать отличия в результатах. Таким образом, для учета влияния московской подвыборки следует разделить массив МИКС-online на московскую и региональную подвыборки и сравнить их с распределениями тех же признаков в массиве МИКС-1.

Еще одним устранимым отличием является наличие в массиве МИКС-1 респондентов, не имеющих доступ в Интернет, и, наоборот, отсутствие неинтернетной публики в МИКС-online. Доступность к оперативной информации вообще, и в Интернете, в частности, является фактором, влияющим на отношение к кризису. Во-первых, потому, что информированность о предмете или явлении, скорее всего, влияет на отношение к нему. Во-вторых, потому что экономически наиболее активные социальные группы, чаще других обращаются к Интернету, а отношение этих групп к кризису, скорее всего, должно отличаться от отношения других социальных групп и населения в целом.

Третьим отличием массивов было отсутствие в МИКС-online трех нижних сегментов респондентов по их самооценке потребительского статуса. Исследование МИКС-1 показало наличие тесной взаимосвязи между потребительским статусом и восприятием экономического кризиса. Поэтому, очевидно, что отсутствие наименее доходных групп в массиве должно значимо влиять на отношение к кризису всей выборочной совокупности.

Гипотеза исследования:

Устранение обозначенных отличий - выравнивание – массивов, хотя и нивелирует их отличия по социально-демографическому составу, но не сглаживает различий в особенностях поведения.

С целью проверки этой гипотезы будем постепенно устранять отличия двух массивов и пошагово сравнивать их.

Выравнивание по потребительскому статусу.

Прежде всего, удалим из общего массива МИКС-1 все анкеты, не удовлетворяющие критерию «потребительского статуса», принятого в МИКС-online и оставим только анкеты с вариантами 4 и 5 ответа на вопрос об уровне доходов. Таких анкет оказалось 2109. Назовем этот массив МИКС-45.

Если сравнить доли вариантов 4 и 5 в каждом из массивов, как показано в таблице 2, видно, что в оффлайновом массиве присутствуют значимо большая доля респондентов с наивысшим потребительском статусом. Иначе говоря, онлайн-опрос не охватывает наиболее доходные группы населения. Это может происходить в силу того, что регулярная деятельность в Сети не является для этой группы повседневной практикой в силу их занятости и, возможно, более старшего возраста.

Таблица 2.

Соотношение различных групп потребительскому статусу в массивах				
	МИКС-45	МИКС-45-сетевики	МИКС-Онлайн	МИКС-Онлайн-регионы
Мы можем без труда приобретать вещи длительного пользования. Однако для нас затруднительно приобретать действительно дорогие вещи	86%	92%	97%	97%
Мы можем позволить себе достаточно дорогостоящие вещи - квартиру, дачу и многое другое.	14%	8%	3%	3%

Последнее можно легко подтвердить, сравнивая распределение по социально-демографическим характеристикам каждый из массивов. Главное отличие – существенно иной средний возраст и дисперсия этого показателя. В онлайн-массиве средний возраст оказался равным 29 лет, а дисперсия +/-9, в офф-лайн-массиве – 37 и +/-14 лет соответственно. Как видим, сетевые респонденты и существенно моложе (на 8 лет) и существенно однороднее

своих офф-лайновых соотечественников. Это заметно смещает выборку в сторону средних значений материального достатка.

Кроме этого, сетевая публика работает в иных отраслях экономики. Как видно из таблицы 3, в online-массиве более, чем вдвое выше доля занятых в отраслях связи и ИТ (15% против 6% в массиве МИКС-45), что вполне объяснимо способом сбора данных. Также выше доля занятых в науке, образовании, здравоохранении и юриспруденции (17% и 14% соответственно). При этом снижена доля работающих в промышленности, строительстве и на транспорте (11% и 12% против 18% и 18% соответственно).

Таблица 3.

Соотношение групп, занятых в различных отраслях экономики в массивах				
	МИКС-4 5	МИКС-4 5- сетевики	МИКС- Онлайн	МИКС- Онлайн- регионы
сельское или лесное хозяйство	2	1	0,4	0,6
промышленное производство	18	13	11	14
строительство и транспорт	18	14	12	13
связь и информационные технологии	6	12	15	12
государственное и муниципальное управление	3	3	2	2
наука, образование, здравоохранение, юриспруденция	14	12	17	17
финансы и инвестиции	4	7	9	7
услуги, в т.ч. торговля	30	35	30	29
коммунальное хозяйство	2	2	2	1,5
военная служба, милиция, силовые органы, в т.ч. частные	3	2	3	3

Сетевая выборка включает в себя более квалифицированных работников, занимающих в своих предприятиях более статусные позиции (таблица 4). Так, доля руководящих работников в онлайн-массиве почти вдвое выше (17% против 10%), почти в полтора раза больше численность работников высшей профессиональной квалификации (36% и 25% соответственно). А вот

доля рабочих и служащих в массиве МИКС-online ниже, чем в МИКС-45. Можно предположить, что в сетевом опросе в существенно меньшей степени участвовали представители групп рабочих и служащих, занятые по преимуществу в промышленности и строительстве. И, напротив, те, кому, скорее всего, по своим должностным обязанностям приходится регулярно использовать доступ в Интернет, чаще тех, у кого такой необходимости нет, участвовали в опросе МИКС-online.

Таблица 4.

Соотношение различных профессиональных групп в массивах				
	Массив МИКС-45	Массив МИКС-45 -сетевики	Массив МИКС-Онлайн	Массив МИКС-Онлайн-регионы
неквалифицированные рабочие, крестьяне	3	1	0	0
рабочие ср. и высокой квалификации, фермеры	18	11	7	8
служащие без квалификации	5	4	3	2
служащие средней и высокой квалификации	19	15	9	8
студенты и учащиеся	6	8	6	7
работники высшей профессиональной квалификации	25	30	36	38
руководящие работники	10	18	17	15
военнослужащие и милиция	2	1	3	2
прочие должности	9	11	19	19

Однако, наиболее существенные отличия мы находим в образовательных группах (таблица 5). Судя по всему, сетевое сообщество имеет радикально отличающийся от остального населения, даже равного ему по потребительскому статусу, уровень формального образования. В значительной степени это закреплено имеющимся должностным статусом работника. Таким образом, изучаемые нами выборки существенно отличаются по социально-демографическим характеристикам.

Таблица 5.

Соотношение групп с различным уровнем образования в массивах				
	МИКС-	МИКС-45	МИКС-	МИКС-

	45	-сетевики	Онлайн	Онлайн-регионы
Образование				
Неполное среднее	2	1	0,2	0,2
Среднее	23	8	4	3
Среднее специальное	23	18	11	11
Высшее	52	73	85	86

Выравнивание по представительству регионов.

На следующем шаге удаляем из массива МИКС-online анкеты, полученные от жителей Москвы. После этого в массиве, который назовем МИКС-online-регионы, осталось 607 анкет. Из таблицы 5 можно увидеть, что уровень формального образования, принадлежность к тому или иному потребительскому статусу и квалификация массива в целом и региональной его части практически не отличаются (см. табл. 2 и 4). Иначе говоря, респонденты online-исследования, т.е. те, кто регулярно работает или проводит свободное время в Сети, вне зависимости от того, живут ли в Москве или в российских регионах, практически не отличаются по своим ключевым социально-демографическим характеристикам. Следовательно, можно предположить, что они сходны и по своему отношению к актуальным вопросам и сюжетам общественной жизни, например, к экономическому кризису.

Тем не менее, различия имеются. Они касаются сфер экономики, в которых работают респонденты (см. табл.3). Взятые в отдельности обнаруженные изменения незначимы. Однако, все они дают смещение в сторону значений массива МИКС-45, т.е. региональная часть массива, собранного в Интернете, оказывается ближе к массиву, собранному традиционным образом в регионах, чем МИКС-online в целом. Отчасти, это может быть следствием отличий генеральной совокупности: ведь, очевидно, что в Москве выше доля занятых в области связи и информационных технологий, а также финансовом секторе. Таким образом, полученная в Интернете выборка в целом отражает

некоторые региональные отличия, характерные для генеральной совокупности.

Выравнивание по представительству в Сети.

Теперь сформируем в массиве МИКС-45 подвыборку имеющих более или менее регулярный доступ в Интернет. Опросник, который, напомним, был полностью идентичен для обоих массивов, содержал вопрос о различных потребительских практиках, в том числе о работе «с Интернетом, отправке и получении электронной почты». Таких в выборке МИКС-45 665 записей. Назовем их «сетевым сообществом», или «сетевиками», а массив – МИКС-45-сетевики.

Средний возраст респондентов этого массива 32 года, а дисперсия - 11 лет. Напомним, что средний возраст массива МИКС-45 был 37 лет, МИКС-online – 29 лет, а дисперсия – 14 и 9 лет соответственно. Как видим, средний возраст в МИКС-45-сетевики гораздо ближе к среднему возрасту массива МИКС-online, чем к массиву МИКС-45. Иначе говоря, возраст пользователей Интернета, охваченных традиционными опросными технологиями, очень близок аналогичному параметру респондентов, опрошенных по онлайн-технологии.

Становятся также ближе к этим значениям характеристики респондентов, опрошенных по традиционным методикам, но указавшим на свой регулярный доступ в Интернет. Начнем с того, что распределение по группам потребительского статуса приближается к тому, что было зафиксировано в сетевом опросе (см. табл. 2). Это подтверждает высказанную выше гипотезу о том, что повседневные сетевые практики, включающие в себя онлайн-опросы, не характерны для верхних доходных групп. Отчасти это происходит по причине того, что наиболее доходные группы в большой степени составляют люди, принадлежащие к поколениям, для которых доступ в Сеть еще не стал ежедневной потребностью, а общение там – регулярным.

Распределение по сферам занятости у сетевиков, охваченных традиционными технологиями опроса, становится практически идентичным аналогичному распределению среди участников онлайн-опроса (см. табл. 3). Значимые отличия выявлены лишь по долям занятых в сфере услуг и в сфере науки, образования и здравоохранения. Возможно, что причина этих расхождений в том, что для занятых в сферах, требующих большей квалификации и интеллектуального труда, деятельность в Сети является более интенсивной и разносторонней. В то время, как для занятых в сфере услуг, она по преимуществу фрагментарна и однообразна. По этой причине, среди участников онлайн-опроса доля «продвинутых» пользователей выше, а среди сетевиков, опрошенных традиционным образом, выше доля пользователей, которых можно было бы назвать поверхностными.

Отчасти это подтверждается и распределением по группам квалификации (см. табл. 4). Среди участников онлайн-панели выше, относительно участников традиционного опроса, доля работников высшей профессиональной квалификации (инженеров, врачей, юристов, учителей и пр.). Напротив, опрос Ассоциации охватил в большей степени квалифицированных служащих, которые чаще всего заняты в разнообразных ведомствах, а также торговли, где взаимодействие с компьютером и мировой сетью носит довольно однообразный характер.

Это лучше всего увидеть в группах формального образования (см. табл. 5). Если среди участников панели имеющих высшее и неоконченное высшее образование составляет 86%, то среди имеющих доступ к Интернету их всего лишь 73%. Как видим, онлайн-опросы охватывают очень образованную часть Интернет-сообщества, ведущую в Сети активную и разнообразную деятельность.

Убедимся в этом, используя признак участия респондентов в различных современных практиках. Как показано в таблице 6, между массивами МИКС-45 и МИКС-онлайн-регионы есть существенные различия по их

повседневным практикам. А вот между массивами МИКС-1 «сетевики» и МИКС-онлайн значимая разница есть лишь в шести из пятнадцати видов практик. Как видим, и по набору повседневных практик сетевики, охваченные традиционным опросом, оказываются более схожими с теми, кто был привлечен через Интернет. При этом отличия сохраняются для преимущественно потребительских практик, таких как кредитование, оплата за товары и услуги карточкой и доставка товаров на дом.

Таблица 6.

Охват различными повседневными практиками в массивах			
	МИКС-45	МИКС-45-сетевики	МИКС-онлайн-регионы
брать кредит в банке	18	39	44
покупать товары в кредит	18	40	52
водить автомобиль	24	59	57
ездить за границу	15	41	39
расплачиваться за товары и / или услуги при помощи пластиковой карточки	18	46	68
иметь дело с иностранной валютой	21	54	62
пользоваться услугами косметических салонов	17	45	49
пользоваться услугами домработницы, помощницы по хозяйству, няни, сиделки	3	8	6
летать самолетами	16	42	39
приобретать спортивные товары и / или туристическое снаряжение	22	57	53
заниматься в фитнес-центре, спортивном клубе	13	37	45
вкладывать деньги в акции, ценные бумаги	4	11	14
пользоваться услугой доставки товаров на дом	14	35	45
получать дополнительное образование, повышать квалификацию	17	46	48
пользоваться мобильным телефоном	44	98	97

Проведенный сравнительный анализ показал, что

1. Региональная часть онлайн-панели по другим своим социально-демографическим характеристикам практически не отличается от части, сформированной жителями Москвы. Напротив, респонденты, опрошенные по обычной методике, существенно отличаются по своим характеристикам в зависимости от того, имеют ли они доступ в Интернет. Образно говоря, «московская прописка» в меньшей степени влияет на социально-демографические характеристики, чем «прописка в Сети».
2. Аудитории, охваченные онлайн-опросами по сравнению с теми, кто, хотя и имеет доступ в Сеть, но опрошен был традиционными способами, отличаются более высоким уровнем формального образования, более высокой квалификацией повседневной деятельности и более разнообразным потребительским поведением. При этом описанные аудитории сходны по возрасту.

Таким образом, проводя онлайн-опросы, направленные на получение репрезентативной выборки, следует учитывать, что получаемые массивы отражают мнение очень активной в своем социальном поведении группы населения.