Индекс Наг: раздача слонов

Вот и закончился второй этап сбора данных по нашему проекту "Индекс Наг". Анкеты собраны и в этой статье расскажем, кто оказался в числе тринадцати счастливчиков, и подсчитаем результаты. Сразу хочу оговориться, что данные до конца не обработаны и отчета еще нет. Но мы стараемся и очень стремимся закончить эту работу.

Объявленный в начале декабря года две тысячи тринадцатого очередной этап сбора данных по исследованию общественного мнения о качестве услуг интернет сервис-провайдеров закончен. Всего в процессе приняло участие более двух тысяч респондентов - этого достаточно с точки зрения социологии, чтобы заявить, что данные довольно репрезентативны реальному состоянию дел. Для тех, кто захочет немного поспорить о необходимости и достаточности респондентов в подобных исследованиях - милости прошу для начала изучить матчасть. Например, учебник социологии Инны Феликсовны Девятко, моего преподавателя социологии в далеком прошлом. Или пролистать ISO/TR13425:2003 "Guidelines for the selections of statistical methods in standardization and specifications", пусть бы и в русском переводе.

Могу только сказать, что ведущие российские социологические исследовательские организации считают необходимым и достаточным число респондентов в пределах 1500-1600 человек. Погрешность, разумеется, в данном исследовании имеется, и я расскажу, где она есть.

Итак, всего в ходе сбора данных было получено **2107 анкет**, которые первым делом были обработаны на заведомо сфальсифицированные и дублирующие. Фальсификация выявлялась довольно просто - как правило, доморощенные тролли пишут ерунду в открытых полях формы и пишут нереальные контактные данные. Например, телефон заведомо неверный (да-да, слава Е.164!) или электронную почту с отчетливо видным фейковым адресом.

Еще некоторые товарищи по ряду причин, которые мне не очень понятны, заполняли анкету несколько раз. Было несколько человек, которые не просто дублировали свои ответы, а пытались как-то их разнообразить. Очевидно, они ожидали увеличения вероятности выигрыша. Такие анкеты нещадно удалялись, оставалась только одна, первая заполненная по отметке времени.

В итоге чистки анкет осталось чуть меньше - **2069**, а данные из разных серий ответов сведены в единую таблицу. Итоговая мастер-таблица прошла еще раз проверку на дубли - некоторые респонденты могли попасть под "рекламный обстрел" несколько раз - и было удалено еще десяток продублированных ответов. Таким образом, в общей итоговой базе исследования осталось 2058 анкет, а коэффициент фальсификации составил всего 2,3% от принятых ответов. Я считаю, это хороший результат, свидетельствующий о том, что респонденты довольно ответственно подошли к заполнению откровенно сложной анкеты.

Подсчет количества заполненных открытых и необязательных полей подтвердил это, Например, описание "идеального оператора", открытое необязательное поле, заполнили 1237 человек, причем, совершенно неоднозначно, а некоторые довольно пространно. То есть, коэффициент значимых ответов в открытой форме составило 60% от общего числа респондентов и эти данные необходимо будет обработать.

Подсчет стандартных статистических параметров по результатам дал результат в ±2.16% максимального отклонения от полученных в данном исследовании от истинных данных генеральной совокупности. Исходные данные для подсчетов были взяты следующие:

Z = Z фактор - 1,95 для 95% доверительного интервала.

p = процент интересующих респондентов или ответов, в десятичной форме - 0,5 (то есть ответы да/нет, для ответов с большим количеством вариантов точность растет)

C = доверительный интервал, в десятичной форме 0,03 (±3%)

Практическая ценность полученной цифры заключается в том, что статистически значимыми можно принимать ответы, если их количество больше 2,16% от общего количества анкет, что составляет округленно (от 2058 респондентов) - 45 ответов. Попробую пояснить на следующем примере: имеет смысл считать количество ответов для определенной группы респондентов, если их общее число в выборке превышает 45 человек. Остальные не играют роли, ибо находятся за пределами статпогрешности. Вот, например, респонденты по отмеченному статусу, которые назвали себя:

1. Представителем организации поставщика услуг (Интернет-провайдера) себя назвали 84 респондентов (4% опрошенных).

2. Представитель СМИ – 13 респондентов.
3. И как “частное лицо” себя назвали - 1961 человек (95,2%).

Видно, что принципиально имеет смысл сравнивать мнение представителей интернет-провайдеров и "частных лиц", но совершенно бессмысленно будет сравнивать мнение журналистов с кем бы то ни было, ибо их количество не достаточно, чтобы выйти за предел статистической погрешности - данных банально не хватает.

Магия статистики представлена на следующей табличке:

Таблица 1. Распределение ответов по сериям анкет

№ серии	Всего анкет	Принятых анкет	Вопрос I (ответов)	Доля	Отклоне-ние	Вопрос II (ответов)	Доля	Отклоне-ние
1	2	3	4	5	6	7	8	9
1	61	60	15	25,00%	1,27%	24	40,00%	3,12%
2	361	354	86	24,29%	0,56%	128	36,16%	-0,72%
3	121	121	37	30,58%	6,85%	44	36,36%	-0,51%
4	10	9	5	55,56%	-	5	55,56%	
5	269	268	70	26,12%	2,39%	89	33,21%	-3,67%
6	1	1	1	100,00%	-	0	0,00%	
7	107	107	15	14,02%	-9,71%	29	27,10%	-9,77%
8	23	22	4	18,18%		3	13,64%	
9	68	68	19	27,94%	4,21%	24	35,29%	-1,58%
10	40	37	10	27,03%		23	62,16%	
11	53	51	11	21,57%	-2,16%	20	39,22%	2,34%
12	723	703	151	21,48%	-2,25%	293	41,68%	4,80%
13	47	46	13	28,26%		19	41,30%	
14	223	222	54	24,32%	0,59%	62	27,93%	-8,95%
ИТОГО:	2107	2069	491	23,73%	0,00%	763	36,88%	0,00%

Это распределение ответов по сериям анкет. В смысле все анкеты не складывались в одну табличку, а были распределены по разным с целью разделить собираемые данные для дальнейшего анализа. Например, серия 1 - это форма, которая была размещена на сайте Наг.ру в статье про объявление сбора данных. Серия 2 - это была форма на специальной страничке, куда привлекали рекламой тоже с Нага, но со станиц справочника провайдеров. Серия 3 - это... Впрочем, тут долго рассказывать и вообще, это мое ноу-хау, которое я раскрывать не очень хочу.

Для каждой серии я рассчитал количество ответов, данных на один вопрос. Вопрос I звучал как "Готовы ли вы пользоваться бесплатным доступом в Интернет при условии просмотра рекламы и значительном ограничении скорости доступа? ", а подсчитывалось количество конкретных ответов "Считаю, что бесплатный доступ за рекламу приемлем только в публичных местах по технологии Wi-Fi без обязательств со стороны оператора".

Вопрос II - это количество ответов "Скорость ниже заявленной, но можно считать ее удовлетворительной” на вопрос "Соответствует ли реальная скорость подключения заявленной? "

Взяты были именно эти ответы на именно эти вопросы по причине их нейтральности и фактической независимости от субъективных факторов, таких как регион или конкретный оператор.

Я "подсветил красненьким" ответы, которые выходят за пределы расчетных ±2.16% для наглядности. Нет расчетов для серий, где количество ответов слишком мало (меньше 45), чтобы войти в доверительный интервал.

И по табличке четко видно, что из расчетного отклонения по обоим параметрам сильно выбились только две серии - 7 и 12. Причем, серия 12 по вопросу I отклонилась не так уж и много - в пределах буквально сотых долей процента. По серии 7 отклонение довольное больше. Причины этого отклонения я пока выясняю, но склонен отнести это к специфике источника - анкета этой серии была опубликована на сайте партнера - довольно крупного оператора и сбор данных по ней происходил несколько нетипично относительно других серий.

Давайте опубликую еще немного демографических данных, чтобы был понятен еще один недостаток метода CAWI - невозможность устанавливать квотирование по определенным демографическим группам.

Гендерно-возрастная составляющая респондентов получилась вот такая:

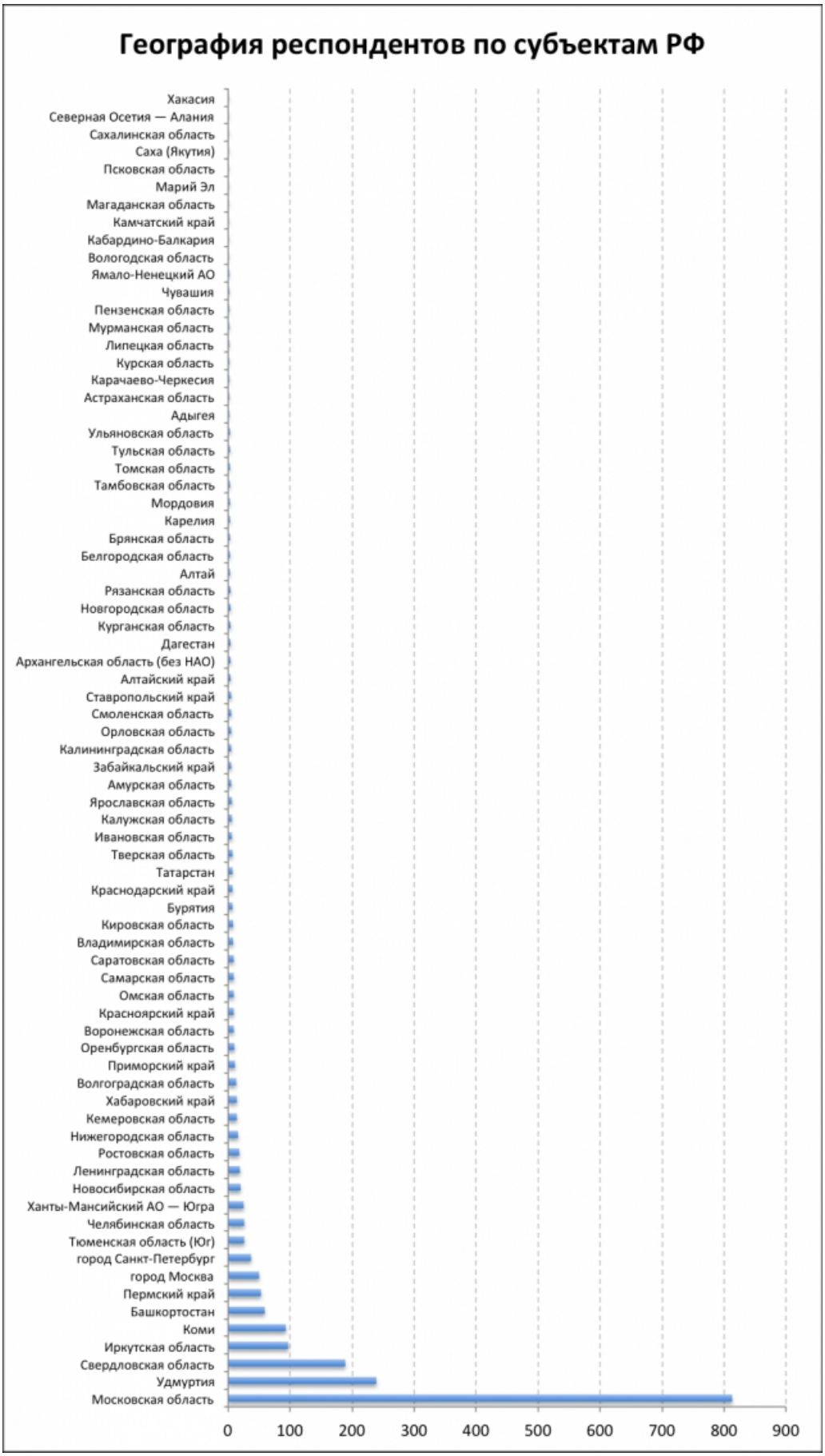
	25-34 года	35-44 года	45-54 года	До 24 лет	Старше 55	Grand Total
Женский	177	89	30	67	18	381
Мужской	820	378	137	280	62	1677

Да, женских ответов опять меньше, чем хотелось бы - всего 19% от общего количества респондентов. Но, как уже отмечалось в предыдущей исследовании - это не баг, это фича. Данный показатель всего лишь отражает тот факт, что выбор интернет-провайдера и взаимодействие с ним дело преимущественно мужское. С этим нужно просто смириться и пользоваться. Управленческое решение же из этого факта можно принять, например, в области сегментирования рекламных материалов - на мой взгляд, более правильным будет ориентировать рекламу на привлечение именно мужской аудитории.

География респондентов в разрезе федеральных округов получилась следующая:

С прискорбием вынужден сообщить, что в этот раз порог статпогрешности не прошли не только совершенно ожидаемо Северо-Кавказский федеральный округ, но еще и Дальний восток и Юг. Это плохо, друзья. Но тут трудно что-либо поделать - проникновение интернет-доступа в этих регионах действительно сильно ниже, чем в других федеральных округа Российской Федерации. Можно попытаться объединить ЮФО и СКФО в один, но даже в этом случае (54 анкеты) не сыграют значительной роли. Слишком мало данных.

По субъектам федерации получилась следующая картина:



Так же получился перегиб с количеством респондентов в отдельных субъектах, хотя, почти во всех регионах России был хотя бы один респондент, за исключением Чечни, где, к сожалению, никто не отметился. То есть охват получился действительно всероссийский, и не побоюсь этого слова - федеральный.

Регионов же, прошедших порог статпогрешности немного:

Московская область	813
Удмуртия	239
Свердловская область	189
Иркутская область	97
Коми	93
Башкортостан	59
Пермский край	53
город Москва	50

По этим регионам имеется возможность построить сравнительные графики - как отличается потребление от общей картины, и это будет сделано в итоговом отчете. Остальные субъекты, увы, до сравнительных характеристик не дотянули. Даже Санкт-Петербург не сдюжил, набрав всего 37 респондентов. А "странный" на первый взгляд, набор регионов, порог прошедших, объясняется наличием там партнеров, которые приняли деятельное участие в сборе данных. Посему, в третьем исследовании, которое запланировано на лето текущего года, приглашается как можно больше партнеров-операторов - если вы хотите узнать, как относятся клиенты к вашей компании, к вашим конкурентам, и чего ожидает клиент в конкретной местности, то выход только один: спросить клиента. А спрашивать по уже готовой методике с поддержкой и конечным результатом гораздо выгоднее, чем пытаться изобрести свой велосипед. Кроме того, ценность данной методики заключается в возможности сравнительного анализа.

Относительно типа населенного пункта, где проживают респонденты в общем, получилась следующая картина:

[Индекс Наг: раздача слонов]

Как видно из круговой диаграммы, все типы населенных пунктов представлены в количестве достаточном для сравнения между собой - даже где-то примерно поровну.

Очень приблизительно, наиболее встречаемые операторы по бренду выглядят так:

Инфолинк	659
Марк-ИТТ	213
Ростелеком	168
Нет в списке	162
Билайн (Вымпелком)	90
ТТК (ТрансТелеКом)	76
Усинск-сеть	62
IRKNET TELEKOM	62
Дом.ru (ЭР-Телеком)	60
Башинформсвязь	49
Урал-ТТК	48
MTS	55

Приблизительно, потому что из-за довольно большого списка действующих операторов довольно большое количество респондентов поленились или не смогли найти своего оператора и ответили "нет в списке". Теперь предстоит в ручную расставить 162 оператора по списку, так что расклад вполне может и измениться.

Большое количество респондентов, отметивших как "свой оператор" небольших региональных провайдеров, объясняется тем, что именно эти операторы приняли деятельное участие в сборе данных. То есть это ни в коей мере не рейтинг - просто так получилось. Для рейтинга известности брендов было отдельное открытое поле, данные из которого еще предстоит обработать.

Хочу высказать свою благодарность тем операторам, которые приняли деятельное участие в исследовании: Инфолинк (Московская область), Марк-ИТТ (Ижевск), Усинск-Сеть (Усинск, республика Коми), ИркНет Телеком (Иркутск) и Уралтранстелеком. Спасибо вам, коллеги!

Ну, и в заключение этой статьи, но вовсе не серии в целом, обещанный розыгрыш слонов призов. В прошлый раз мы делали розыгрыш на публичном мероприятии, а именно на КРОСе - из аквариума бумажки доставали участники конференции. В этот раз подгадать большое публичное мероприятие не удалось. Посему мы просто и без затей уселись в офисе Наг.ру и разыграли обещанные призы и даже чуточку больше. У нас нет и не было цели, чтобы призы достались каким-то конкретным людям

- посему, можете поверить, никаких передергиваний или подыгрываний не было и быть не могло. Репутация честного человека, поверьте, дороже, чем стоимость призов в сумме чай, не автомобиль разыгрывали.

Видео сего действа получилось вот такое:

А список победителей такой.

Маршрутизаторы SNR CPE W4N-3G достаются следующим счастливым респондентам:

1. Станислав Сергеевич, Тюменская область, телефон заканчивается на *2529
2. Виталий, Московская область, телефон *5470
3. Алёна, Новосибирская обл., *3137
4. Стасов Валерий Алексеевич, Московская область, телефон *9877
5. Сергей, Тюменская область, телефон *5655
6. Антон, Московская область, телефон *1884
7. Степан, Свердловская область, телефон *0333
8. Галина, Удмуртия, телефон *4880
9. Александр, Удмуртия, телефон *7854
10. Сергей, Омская область, телефон *5788

Имя здесь указано ровно так же, как автор указал в анкете, а телефон, по понятным причинам, мы не публикуем целиком, но указанных цифр вполне хватает для идентификации счастливчика.

Специальный приз, предоставленный компанией Мегафон, "умная ручка" Echo SmartPen достается Сергею из Иркутской области с последними цифрами телефона *0097. Надеюсь, он разберется что это за гаджет, ибо вещь достаточно необычная.

Ну, и два главных приза, опять таки "умные" часы Sony SmartWatch 2, достаются везучим людям:

1. Данил, Московская область, телефон *9523
2. Дмитрий, Иркутская область, телефон *5327

С часами произошла несколько неприятная вещь - мы планировали купить в официальном интернет-магазине Sony, где было обещано разместить их "в феврале 2014 года". Однако, в феврале в продаже они не появились, посему, в записи мы этот девайс показать не смогли. К счастью, на момент написания статьи, часы все же появились в продаже и тут же заказаны.

Всем победителям высланы электронные письма с целью уточнения адреса для доставки. Если в течение недели на них не будет ответа - мы позвоним. Если же и на звонок не получим ответа - что ж. Приз будет оставлен до следующего розыгрыша, который планируется провести летом с примерной датой май-июнь сего года.

В качестве анонса следующей статьи серии "Индекс Нага" скажу, что это будет анализ влияния факторов скорость/цена/качество и других параметров оказания услуг на итоговую оценку NPS операторов, по принципу "что важнее". В комментариях принимаю ставки на наиболее значимый параметр.

Собственно отчет планируется написать до конца марта сего года, и он будет доступен по смешной цене всего в 29 900 рублей. Предзаказ уже можно отправлять по адресу klimarev@gmail.com. Скидки в 20% от стоимости будут предоставлены тем компаниям, которые уже приобрели отчет и партнерам-участникам исследования.