



ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

Подлазов А.В.

Исследование
статистических методов
выявления выдуманных
результатов выборов: Часть
1. Круглые числа

Рекомендуемая форма библиографической ссылки: Подлазов А.В. Исследование статистических методов выявления выдуманных результатов выборов: Часть 1. Круглые числа // Препринты ИПМ им. М.В.Келдыша. 2019. № 147. 28 с. <http://doi.org/10.20948/prepr-2019-147>
URL: <http://library.keldysh.ru/preprint.asp?id=2019-147>

**Ордена Ленина
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Российской академии наук**

А.В. Подлазов

**Исследование
статистических методов выявления
выдуманных результатов выборов:
Часть 1. Круглые числа**

Москва — 2019

А.В. Подлазов

Исследование статистических методов выявления выдуманных результатов выборов: Часть 1. Круглые числа

Наиболее грубым способом фальсификации результатов выборов является простое их выдумывание, из-за которого возникает преобладание психологически притягательных значений электоральных характеристик. В работе конструируются, обосновываются и исследуются тесты для выявления таких фальсификаций. Все эти тесты строго формальны, основываются только на официальных данных о результатах выборов и не используют каких-либо предположений о поведении избирателей или эталонных результатов. Особый акцент сделан на максимально возможном упрощении как самих тестов, так и способов представления результатов их применения, что должно сделать их общедоступными для использования и проверки.

Для федеральных выборов 2000-2018 гг. выдуманные результаты в значимых количествах обнаруживаются для примерно 40% избирательных кейсов, из которых не менее 10% удастся уличить индивидуально. Массовыми фальсификациями данного типа затронуты примерно 20÷30 субъектов РФ.

Ключевые слова: электоральная статистика, электоральные фальсификации, психологически привлекательные значения, круглые числа, статистические гипотезы, метод максимального правдоподобия

A.V. Podlazov

A study of statistical methods for identifying fictitious election results:
Part 1. Round numbers

The crudest form of electoral frauds is simply to make up results. This gives rise to a predominance of psychologically attractive values of electoral characteristics. I develop, ground and explore tests to identify such frauds. All these tests are strictly formal, base only on official data on the election results and do not use any assumptions about voters' behavior or reference results. I pay special attention to the maximum possible simplification of both the tests themselves and the ways of presenting the results of their application. This should make them publicly available for use and verification.

Fictitious results present in significant quantities for approximately 40% of the electoral cases of 2000-2018 federal elections in Russia. At least 10% of them are caught individually. About 20÷30 subjects of Russian Federation are affected by mass frauds of this type.

Key words: electoral statistics, electoral frauds, psychologically attractive values, round numbers, statistical hypotheses, maximum likelihood method

Работа выполнена при поддержке РФФИ (проект 18-01-00619-а).

ПОСТАНОВКА ЗАДАЧИ

Бурное развитие гражданского наблюдения за выборами вообще и инструментария электоральной статистики в частности породило богатый спектр методов выявления фальсификаций результатов выборов. Наибольший интерес в этой связи представляют методы, обладающие самостоятельной доказательной силой, для обеспечения которой исследователь отказывается от опоры на:

- любые нерегулярные данные (результаты социологических опросов или голосования на участках, оборудованных электронными урнами);
- субъективные свидетельства (видеозаписи голосования и заявления наблюдателей), объективизация которых требует участия уполномоченных инстанций;
- предположения о размещении и поведении избирателей (связь электоральных предпочтений избирателей с их социальным составом или результатов голосования с явкой избирателей);
- эталонные данные (избирательные участки/выборы, где/когда итоги голосования были, предположительно, подведены честно).

Иначе говоря, желателен такой подход, который основывается только на вскрытии внутренних несообразностей в официальных результатах выборов. Несмотря на столь жесткие ограничения, оказывается возможным, действуя формальным образом, обнаружить самый наглый тип фальсификации результатов, связанный с их выдумыванием. Если числа в избирательных протоколах не находятся в какой-либо связи с содержимым урн для голосования, то массив этих чисел приобретают свойства, нетипичные для истинных результатов. Это возможно продемонстрировать с помощью инструментария проверки статистических гипотез, чему и посвящена настоящая работа.

Следует особо отметить, что описываемые методы не позволяют выявлять фальсификации, основанные на разного рода физических манипуляциях с избирательными бюллетенями: их умышленно неверном прочтении, вбросах, каруселях и т.п. Такие действия являются пусть и мошенническими, но процедурами, в ходе выполнения которых получают числа, обусловленные логикой этих процедур. А здесь речь идет исключительно об обнаружении чисел, выдуманных непосредственно фальсификаторами.

Предметом анализа являются результаты 9 выборов федерального уровня, проводившихся в нашей стране в XXI в., – президентских 2000, 2004, 2008, 2012 и 2018 гг. и парламентских 2003, 2007, 2011 и 2016 гг., для которых ЦИК РФ обнародовал официальные результаты с детализацией до избирательных участков (за исключением выборов 2000 г. по Якутии).

Анализ осуществляется на уровне субъектов федерации по причине существенной разницы между ними как в масштабах фальсификаций, так и в предпочитаемых способах их осуществления. Всего, если исключить из рассмотрения результаты голосования на участках, сформированных за рубежом, имеются *N = 769 избирательных кейсов* – голосований в данном субъекте федерации в данном году, для которых необходимо проверить достоверность результатов.

Базовая идея

При выдумывании результатов в них преобладают *психологически привлекательные числа*. Для целочисленных электоральных характеристик такими являются круглые числа, а для процентных, обычно приводимых с одной цифрой после запятой, – значения без десятых долей (т.е. круглым оказывается количество промилле).

У целочисленной случайной величины с разбросом во многие десятки и даже сотни единиц последняя цифра должна практически равновероятно принимать все возможные значения. Если же встречаемость цифры «0» в конце числа превышает 10%, можно предполагать наличие выдуманных результатов. Проверка этого предположения сводится к проверке статистической гипотезы о естественном возникновении превышения. И если ее *значимость* α – вероятность совершить ошибку I рода, отвернув данную гипотезу, когда она верна, – окажется мала, то наличие фальсификаций следует считать подтвержденным.

Регистрация целочисленной электоральной характеристики на конкретном избирательном участке представляет собой испытание Бернулли с вероятностями *успеха* (круглое число) $p = 1/10$ и *неудачи* (некруглое число) $q = 9/10$. Количество успехов описывается биномиальным распределением с вероятностью регистрации i успехов в n испытаниях $P_n^i = C_n^i p^i q^j$, где $j = n - i$. Таким образом, не менее k успехов регистрируются с вероятностью

$$\alpha = \alpha_n^k = \sum_{i=k}^n P_n^i . \quad (1)$$

Это и есть вероятность того, что, хотя встречаемость цифры «0», равная k/n , и может казаться аномально высокой, ее подъем над уровнем p есть лишь результат случайного стечения обстоятельств, принять которое за признак фальсификации было бы ошибкой.

В случае масштабных фальсификаций величина α оказывается чрезвычайно мала. Поэтому далее для удобства вместо нее всюду приводится ее *десятичный показатель* $\text{ра} = -\lg \alpha$, увеличение которого на единицу соответствует уменьшению на порядок вероятности отсутствия фальсификаций.

Если проверяемая гипотеза верна, то $\text{Prob}\{\alpha \leq z\} \leq z$, где равенство достигается только для z , являющихся значимостями возможных исходов испытания. А значит, при $n \rightarrow \infty$ распределение значимостей для всех возможных исходов испытания приближается к равномерному на $[0;1]$.

Чтобы упростить восприятие показателей значимости, проинтерпретируем их следующим образом. Выборка объемом N должна содержать в среднем не более μ кейсов с $\alpha \leq \mu/N$ (для бесконечного числа испытаний – ровно μ). Если же таких кейсов оказывается больше, можно полагать их фальсифицированными. Параметр μ определяет количество допускаемых при этом ошибок. Далее используются интуитивно понятные значения $\mu = 1, 1/10$ и $1/100$, которым соответствует интерпретация кейсов как *подозрительных*, *исключительных* и *неве-*

роятных при $pa \geq \lg N \approx 2,89$, $pa \geq \lg 10 \cdot N \approx 3,89$ и $pa \geq \lg 100 \cdot N \approx 4,89$ (для полной выборки из $N = 769$ кейсов).

В пределе бесконечного числа испытаний для любой выборки кейсов следует ожидать ошибочной интерпретации 1 кейса как подозрительного в каждом тесте, как исключительного – в 1 тесте из 10 и как невероятного – в 1 тесте из 100. Разумеется, это – средние показатели. Например, приблизительно для $1/e \approx 36,8\%$ тестов подозрительных кейсов не будет вовсе, в равной доле случаев такой кейс будет ровно 1, вдвое реже – 2 и лишь в $8,0\%$ случаев – 3 и более.

Использование больших μ , повышающее интерпретационные пороги для α (понижающее для pa) нецелесообразно, т.к. ведет к неприемлемому числу ложноположительных срабатываний теста, затрудняющих выделение сигнала из-под шума. Вместе с тем значительная, если не основная, доля фальсифицированных кейсов имеет как раз относительно высокие значимости. И хотя без риска многочисленных ошибок эти кейсы не удастся назвать поименно, их вполне можно пересчитать.

Математическое ожидание значимости гипотезы для кейса, содержащего n испытаний, составляет

$$\bar{\alpha} = \sum_{k=0}^n P_n^k \cdot \alpha_n^k = \sum_{k=0}^n P_n^k \sum_{i=k}^n P_n^i.$$

Отношение $\alpha/\bar{\alpha}$ служит мерой достоверности кейса. В самом деле, если рассмотреть все возможные исходы испытания, заменяя для них истинные значения α нулями с вероятностью ε , то среднее значение этого отношения, очевидно, составит $1 - \varepsilon$. Таким образом, из N кейсов достоверны $T = \sum_{l=1}^N \alpha_l / \bar{\alpha}_l$, где индекс l нумерует кейсы, а *всего (интегрально)* фальсифицированы $F = N - T$ кейсов.

Исходя из асимптотически равномерного распределения значимости, можно оценить стандартное отклонение величины F как $\sqrt{N/3} \approx 16$, или примерно 2% от полного числа кейсов. Следует отменить, что эта погрешность описывает не сложности с выявлением фальсификаций, а естественный разброс величины F в их отсутствие. При наличии фальсификаций ее погрешность, естественно, будет ниже.

Для всех выполняемых тестов как интегральный объем выявляемых ими фальсификаций, так и количество кейсов, проинтерпретированных как подозрительные, исключительные или невероятные, сведены в итоговую табл. 8 на стр. 26. Столбцы этой таблицы следует читать через «в том числе»: всего фальсифицировано, в т.ч. подозрительные, в т.ч. исключительные, в т.ч. невероятные.

Две введенных меры выявляемых фальсификаций – интегральная и интерпретационная – обладают принципиально разными свойствами. Первая является аддитивной, т.е. при разбиении выборки на части сумма найденных для них объемов фальсификаций совпадает с тем, сколько всего фальсифицированных кейсов получается на полной выборке. Результаты интерпретации показав-

телей, напротив, существенно зависят от размера выборки. Скажем, если рассматривать только кейсы конкретного года, она значительно сократится, что приведет к соответствующему уменьшению интерпретационных порогов для $ра$ (примерно на $0,94 \div 0,97$ в зависимости от числа субъектов федерации в разные годы). Те значения показателя, которые лишь при этом становятся подозрительными, в таблицах, приводимых далее, даны черным шрифтом, а те, которые всё равно остаются под порогом, – серым. Показатели же, являющиеся и для полной выборки подозрительными (исключительными, невероятными), выделяются жирным синим (фиолетовым, красными) шрифтом.

Две введенные меры фальсификаций кроме различий в свойствах имеют еще и разные объекты контроля. Если общий объем фальсификаций в пределах погрешности больше нуля, это говорит о неспособности избирательных комиссий организовывать выборы и честно подводить их итоги. Если же имеются подозрительные, а тем более исключительные или невероятные кейсы, которым своевременно не дана должная юридическая оценка с указанием конкретных лиц, виновных в фальсификации результатов, это говорит о неспособности правоохранительной системы пресекать электоральные преступления.

Практические вычисления

Значимость α можно рассчитывать как непосредственным суммированием по формуле (1), так и с помощью функций, доступных в электронных таблицах (Microsoft Excel, Apache OpenOffice Calc, Google Spreadsheets). Однако при этом следует избегать напрашивающегося вызова $\alpha = \text{BinomDist}(n-k; n; q; 1)$, т.к. уже при двузначных $ра$ реализация этой функции оказывается ненадежной. Значительно лучше эквивалентное представление $\alpha = \text{BetaDist}(p; k; n-k+1)$, позволяющее вычислять значимость α , пока та представима нормализованным числом с плавающей точкой (т.е. при $ра < 308$). Единственным недостатком этой функции является то, что при $k = 0$ вместо $\alpha = 1$, возникающего как результат предельного перехода $k \rightarrow 0$, она возвращает ошибку, связанную с невозможностью непосредственного вычисления. Об этом нужно помнить и делать соответствующую проверку числа успехов при расчете значимости.

Чтобы перейти к представлению значимости через бета-распределение, сначала продифференцируем сумму (1) по вероятности успеха

$$\frac{d\alpha}{dp} = n \sum_{i=k}^n (C_{n-1}^{i-1} p^{i-1} q^{n-i} - C_{n-1}^i p^i q^{n-1-i}) = n C_{n-1}^{k-1} p^{k-1} q^{n-k}, \quad (2)$$

а затем проинтегрируем обратно последнее оставшееся слагаемое

$$\alpha = n C_{n-1}^{k-1} \int_0^p t^{k-1} (1-t)^{n-k} dt = I_p(k, n-k+1). \quad (3)$$

Здесь $I_p(x,y) = B_p(x,y) / B(x,y)$ – регуляризованная неполная бета-функция. Если читатель, как и автор, не склонен пользоваться электронными таблицами, а

предпочитает программировать самостоятельно, то простую вычислительную реализацию этой функции можно найти, например, в различных изданиях книги “Numerical recipes: The art of scientific computing” [Press W.H., Teukolsky S.A., Bethe H.A., Vetterling W.T. – Cambridge University Press].

ЦЕЛОЧИСЛЕННЫЕ ЭЛЕКТОРАЛЬНЫЕ ХАРАКТЕРИСТИКИ

Начнем анализ с двух простейших электоральных характеристик:

- *размер избирательного участка* – число избирателей, включенных в их списки на момент окончания голосования;
- *участие в выборах* – число избирательных бюллетеней, выданных избирателям, проголосовавшим на участке.

Хотя для фальсификаторов эти характеристики и не являются целевыми, они всё равно подтасовываются, что позволяет на них апробировать методику и подготовить почву для анализа более важных процентных характеристик.

В данном разделе из рассмотрения исключаются участки, где анализируемые характеристики не достигают 20 чел. Введение такого *порога отсечения* необходимо, чтобы избежать искажений, вносимых карликовыми спецучастками, где количество избирателей может определяться не естественными факторами, а задаваться штатными расписаниями определенных служб. Корректность выбранного порога проверяется далее.

Размер избирательного участка

Для целого ряда избирательных кейсов имеет место преобладание избирательных участков круглого размера. Десятичные показатели значимости гипотезы о естественном возникновении такого преобладания представлены в табл. 1.

Здесь и далее в подобных таблицах приводятся данные по всем субъектам федерации, хотя бы раз продемонстрировавшим подозрительный показатель в рассматриваемом тесте. При этом субъекты сортируются и раскрашиваются в соответствии с максимальным зафиксированным для них показателем.

В последнем столбце таблиц дается оценка общей доли фальсифицированных кейсов субъекта. Необходимость использовать именно доли, а не количества объясняется тем, что некоторые субъекты существовали не всё время, что не позволяет напрямую сравнивать количества. Относиться к этой оценке следует с осторожностью, т.к. ее погрешность может быть довольно велика из-за небольшого числа выборов, проводимых в одном субъекте федерации (погрешность измеряется десятками процентных пунктов в отсутствие фальсификаций, хотя при их наличии становится ниже). Аналогичным образом в последней строке таблиц дается оценка общей доли фальсифицированных кейсов (для всех субъектов, а не только для включенных в таблицы). Вновь используются доли, т.к. количество субъектов федерации варьируется по годам (от 83 до 89). Здесь данные становятся более надежны (в отсутствие фальсификаций погреш-

Таблица 1. Показатели значимости для круглого размера участка

| Субъект федерации \ Год | 2000 | 2003 | 2004 | 2007 | 2008 | 2011 | 2012 | 2016 | 2018 | Всего |
|------------------------------|------------|------------|-------------|-------------|-------------|------------|------------|------------|------------|------------|
| 1. Чечня | — | 3,3 | 13,6 | 10,0 | 24,7 | 6,5 | 6,1 | 1,4 | 2,1 | 99% |
| 2. Дагестан | 2,3 | 4,9 | 10,5 | 2,7 | 4,1 | 7,2 | 2,4 | 7,9 | 4,5 | 100% |
| 3. Кабардино-Балкария | 3,4 | 2,7 | 6,8 | 6,0 | 9,1 | 0,7 | 1,9 | 0,3 | 0,2 | 69% |
| 4. Краснодарский кр. | 0,4 | 2,9 | 2,6 | 2,4 | 8,8 | 1,6 | 3,1 | 2,6 | 7,0 | 91% |
| 5. Северная Осетия | 2,4 | 0,7 | 2,2 | 1,9 | 0,7 | 2,7 | 7,4 | 2,7 | 4,6 | 91% |
| 6. Башкортостан | 1,6 | 1,2 | 3,0 | 5,4 | 2,1 | 2,1 | 0,1 | 0,4 | 2,3 | 70% |
| 7. Ростовская обл. | 1,1 | 0,2 | 5,2 | 1,4 | 0,5 | 1,9 | 3,2 | 1,7 | 1,0 | 74% |
| 8. Татарстан | 1,4 | 1,4 | 1,9 | 4,6 | 1,9 | 1,2 | 1,4 | 0,9 | 0,3 | 83% |
| 9. Кемеровская обл. | 0,2 | 0,3 | 4,6 | 0,6 | 3,8 | 0,2 | 0,7 | 0,7 | 3,4 | 49% |
| 10. Корякский а/о | 4,2 | 0,2 | 0,0 | — | — | — | — | — | — | 14% |
| 11. Пермский кр. | — | — | — | 0,2 | 3,3 | 1,6 | 0,7 | 0,0 | 1,3 | 41% |
| 12. Самарская обл. | 1,7 | 1,3 | 0,3 | 0,6 | 3,3 | 0,5 | 0,4 | 1,5 | 0,6 | 62% |
| 13. Саратовская обл. | 0,7 | 1,4 | 2,8 | 0,5 | 0,4 | 3,2 | 0,6 | 2,6 | 0,0 | 53% |
| 14. Тыва | 0,2 | 0,2 | 0,5 | 1,4 | 0,8 | 0,2 | 0,1 | 2,9 | 1,2 | 32% |
| Интегрально (по всем) | 33% | 20% | 26% | 31% | 42% | 21% | 24% | 26% | 31% | 28% |

ность составляет около 6% пунктов). Наконец, в правом нижнем углу таблиц дается оценка доли фальсификаций по всему массиву кейсов, приводимая также и в сводной табл. 8.

Как можно видеть из табл. 1, во многих случаях гипотезу о естественном возникновении избытка круглых чисел можно с уверенностью отклонить как имеющую исключительно низкую значимость, т.е. уже на этом уровне выявляются массовые фальсификации. При этом цветные показатели встречаются неоднократно для «красных» субъектов (хотя бы раз продемонстрировавших невероятные показатели), но лишь по разу для «фиолетовых» и «синих» субъектов (где показатели были лишь исключительными или подозрительными), т.е. склонность к фальсификациям очевидным образом кластеризуется.

Обращает на себя внимание некоторая странность, связанная с тем, что указываемый в избирательных протоколах размер участка вообще не является прямым результатом голосования. В его ходе число зарегистрированных избирателей может лишь корректироваться, но определяться оно должно заранее – при составлении их списка.

Представляются реалистичными две версии. С одной стороны, фальсификаторы могут заполнить весь протокол выдуманными числами, неумышленно затрагивая тем самым и число избирателей, включенных в список. С другой – возможна попытка максимизировать явку, просто приравняв участие в выборах размеру участка, что требует выдумывания не двух разных чисел, а лишь одного, которое в силу своего неестественного происхождения и оказывается круглым.

Сопоставить эти версии можно, исключив из анализа участки с *абсолютной явкой*, т.е. те, где число избирателей, принявших участие в голосовании, в точности совпало с числом зарегистрированных избирателей. При этом, как показывают сводные результаты (см. табл. 8), выявляемые масштабы фальсифика-

ций уменьшаются незначительно. Таким образом, основной причиной, по которой возникает избыток участков круглого размера, оказывается именно тотальная фальсификация результатов. Фальсификаторы порой не дают себе труда выяснить даже то, сколько на самом деле у них на участке имеется избирателей.

Участие избирателей в выборах

Как можно видеть из табл. 2, число избирательных бюллетеней, выданных избирателям, оказывается круглым значительно чаще, чем число самих избирателей, и позволяет выявить существенно большее количество фальсифицированных кейсов в большем числе субъектов. Табл. 8 показывает, что из тестов, использующих лишь на одну электоральную характеристику, данный тест является самым сильным в определении интегрального объема фальсификаций.

Дело здесь в массовой распространенности весьма специфического типа фальсификаций, связанного с (реальным или виртуальным) использованием на участке всех имеющихся бюллетеней без остатка. А поскольку бюллетени участковым комиссиям зачастую отпускают десятками, это приводит к многочисленным случаям круглого участия.

Таблица 2. Показатели значимости для круглого участия в выборах

| Субъект федерации \ Год | 2000 | 2003 | 2004 | 2007 | 2008 | 2011 | 2012 | 2016 | 2018 | Всего |
|-------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|------------|-------------|
| 1. Дагестан | 15,8 | 33,8 | 25,4 | 19,1 | 40,1 | 95,5 | 26,4 | 25,6 | 7,1 | 100% |
| 2. Татарстан | 0,7 | 0,3 | 1,9 | 8,6 | 23,5 | 1,1 | 1,8 | 2,2 | 3,9 | 82% |
| 3. Чечня | — | 13,0 | 14,9 | 1,3 | 0,2 | 0,3 | 0,0 | 0,7 | 0,4 | 32% |
| 4. Башкортостан | 3,8 | 0,4 | 3,0 | 11,1 | 7,5 | 2,7 | 4,1 | 1,7 | 2,0 | 91% |
| 5. Северная Осетия | 0,9 | 1,1 | 5,0 | 0,8 | 4,0 | 1,8 | 3,1 | 8,5 | 7,4 | 92% |
| 6. Краснодарский кр. | 1,5 | 1,5 | 0,9 | 1,9 | 6,9 | 0,7 | 1,0 | 3,1 | 4,2 | 89% |
| 7. Кемеровская обл. | 0,9 | 0,1 | 2,4 | 3,8 | 6,7 | 0,4 | 1,8 | 5,1 | 2,7 | 69% |
| 8. Липецкая обл. | 1,3 | 0,3 | 0,3 | 0,3 | 5,8 | 0,9 | 1,3 | 1,3 | 1,5 | 61% |
| 9. Хабаровский кр. | 0,4 | 0,1 | 0,4 | 0,8 | 1,1 | 0,9 | 1,3 | 0,2 | 5,1 | 46% |
| 10. Крым | — | — | — | — | — | — | — | 0,4 | 5,0 | 63% |
| 11. Карачаево-Черкесия | 1,4 | 0,6 | 3,7 | 4,0 | 4,9 | 0,6 | 2,3 | 1,1 | 3,8 | 86% |
| 12. Пермская обл. | 0,5 | 4,5 | 2,8 | — | — | — | — | — | — | 80% |
| 13. Кабардино-Балкария | 2,3 | 4,4 | 1,7 | 0,8 | 2,8 | 2,0 | 2,4 | 2,5 | 0,5 | 88% |
| 14. Самарская обл. | 0,3 | 0,1 | 1,0 | 0,1 | 4,4 | 0,1 | 0,0 | 0,6 | 0,6 | 6% |
| 15. Тюменская обл. | 0,4 | 0,3 | 1,1 | 0,6 | 1,8 | 2,0 | 0,1 | 4,3 | 1,4 | 55% |
| 16. Пензенская обл. | 0,6 | 0,3 | 1,4 | 3,5 | 1,6 | 0,2 | 0,5 | 0,2 | 0,4 | 41% |
| 17. Ставропольский кр. | 0,1 | 0,3 | 1,2 | 0,4 | 0,2 | 1,0 | 2,9 | 0,9 | 3,5 | 43% |
| 18. Мордовия | 0,4 | 0,2 | 0,9 | 3,5 | 1,9 | 1,4 | 0,8 | 0,9 | 0,1 | 52% |
| 19. Ярославская обл. | 0,4 | 2,2 | 0,1 | 3,3 | 1,7 | 0,8 | 0,1 | 0,7 | 0,2 | 35% |
| 20. Сахалинская обл. | 0,9 | 0,7 | 1,1 | 3,1 | 0,9 | 0,6 | 3,2 | 0,2 | 0,8 | 67% |
| 21. Алтайский кр. | 0,0 | 0,1 | 0,2 | 0,2 | 1,5 | 0,6 | 0,4 | 1,8 | 3,1 | 21% |
| Интегрально (по всем) | 41% | 20% | 33% | 43% | 61% | 42% | 26% | 47% | 43% | 39% |

Таблица 3. Показатели значимости для круглого участия в выборах без учета участков с абсолютной явкой или жадным голосованием

| Субъект федерации \ Год | 2000 | 2003 | 2004 | 2007 | 2008 | 2011 | 2012 | 2016 | 2018 | Всего |
|-------------------------------|------------|-------------|-------------|------------|------------|------------|------------|------------|------------|------------|
| 1. Дагестан | 2,8 | 11,5 | 13,6 | 4,8 | 6,6 | 1,0 | 5,7 | 6,9 | 2,5 | 98% |
| 2. Пермская обл. | 0,4 | 4,7 | 1,9 | — | — | — | — | — | — | 74% |
| 3. Хабаровский кр. | 0,4 | 0,1 | 0,2 | 0,3 | 0,7 | 0,5 | 0,9 | 0,2 | 4,5 | 24% |
| 4. Тюменская обл. | 0,3 | 0,3 | 1,0 | 0,3 | 1,8 | 2,3 | 0,2 | 4,4 | 1,1 | 50% |
| 5. Кабардино-Балкария | 1,7 | 4,0 | 0,6 | 0,1 | 2,4 | 1,2 | 2,5 | 2,3 | 0,4 | 68% |
| 6. Северная Осетия | 0,5 | 1,3 | 2,1 | 0,7 | 2,7 | 1,2 | 1,2 | 3,6 | 1,2 | 84% |
| 7. Чечня | — | 3,5 | 1,1 | 0,2 | 0,2 | 0,1 | 0,0 | 0,5 | 0,2 | 5% |
| 8. Ярославская обл. | 0,4 | 2,3 | 0,1 | 3,3 | 1,5 | 0,6 | 0,1 | 0,7 | 0,1 | 33% |
| 9. Карачаево-Черкесия | 1,5 | 0,7 | 3,0 | 0,3 | 0,9 | 0,4 | 1,9 | 1,2 | 1,9 | 70% |
| 10. Ставропольский кр. | 0,1 | 0,2 | 1,1 | 0,3 | 0,1 | 0,8 | 2,7 | 1,0 | 3,0 | 34% |
| Интегрально (по всем) | 31% | 15% | 22% | 18% | 34% | 26% | 18% | 38% | 29% | 26% |

Определим как *жадные* те участки, где количество избирателей, принявших участие в выборах, в точности совпадает с числом бюллетеней, полученных участковой избирательной комиссией, но при этом оказывается меньше числа зарегистрированных избирателей. Последнее ограничение необходимо для исключения «герметичных» участков, находящихся в труднодоступных местах, где и впрямь могут проголосовать все зарегистрированные избиратели, что требует получения такого же количества бюллетеней. Кроме участков с абсолютной явкой жадными не считаются и участки, оборудованные комплексами электронного голосования, при использовании которых избирательные протоколы зачастую заполняют неправильно, указывая в качестве числа полученных бюллетеней не число одноразовых карточек для голосования, а просто число проголосовавших избирателей, что ошибочно может восприниматься как жадность.

Табл. 8 показывает, что тест на круглое участие в выборах еще менее чувствителен к исключению из рассмотрения участков с абсолютной явкой, чем тест на круглый размер участка, тогда как исключение из рассмотрения жадных участков заметно сокращает улов. Таким образом, именно жадность фальсификаторов, стремящихся выжать максимум голосов, является основной причиной избытка участков с круглым числом избирателей, принявших участие в выборах.

Если исключить из рассмотрения оба упомянутых типа участков, как сделано в табл. 3, то количество кейсов, уличенных в фальсификациях индивидуально, станет очень небольшим, хотя ее выявленный интегральный объем останется высоким (см. табл. 8). Фактически только в Дагестане фальсификаторы испытывают систематическую и немотивированную склонность к массовому выдумыванию круглого участия. В других субъектах федерации это либо происходило лишь эпизодически, либо было обусловлено стремлением максимизировать явку (с фальсификацией размера участка или без оной).

Отрицательный контроль

Проанализировав встречаемость цифры «0» в конце электоральных характеристик, уместно сделать то же самое и для всех прочих цифр. Поскольку нет весомых оснований предполагать массовые фальсификации в пользу цифр «1», «2», ... «9», гипотеза о естественной распространенности любой из них должна иметь высокую значимость. Для каждого кейса ограничимся самой часто встречающейся последней цифрой, отличной от «0», дающей наименьшую значимость (ей соответствует наибольший показатель).

Рассмотрение такой гипотезы требует коррекции интерпретационных порогов. В предположении о независимости встречаемости разных цифр пороги минимальной значимости установятся на уровне $\alpha(\mu) = 1 - (1 - \mu/N)^{1/d} \approx \mu/Nd$, где μ – ожидаемое число ложноположительных срабатываний теста на выборке (1, 1/10 и 1/100 для рассматриваемых интерпретаций), а d – число тестов, для которых берется минимальная значимость. Его определение нетривиально. С одной стороны, хотя для отрицательного контроля служат только 9 цифр, здесь следовало бы считать, что их всё же 10, коль скоро гипотеза о круглых числах уже была рассмотрена как выделенная априорно. Но, с другой стороны, встречаемости разных цифр очевидно взаимосвязаны, что требует уменьшения d как минимум на 1. В итоге принято значение $d = 9$ как верхняя оценка числа независимых значимостей. В результате увеличения порогов $\rho\alpha$ на $\lg d \approx 0,95$ подозрительными становятся кейсы с $\rho\alpha \geq 3,84$, исключительными – с $\rho\alpha \geq 4,84$ и невероятными – с $\rho\alpha \geq 5,84$.

В случае размера участка обнаруживается только 2 подозрительных кейса, что не выходит за пределы вероятного: $\rho\alpha = 4,5$ для выборов 2004 г. в Новгородской обл. и $\rho\alpha = 4,0$ для выборов 2012 г. в Архангельской обл. Однако для обоих этих кейсов доминирующей последней цифрой оказалась «5», что дает основание подозревать фальсификации к полукруглым числам – вещь экзотическую, но не невообразимую. После исключения из рассмотрения участков с абсолютной явкой лишь новгородский кейс остается подозрительным, что позволяет допустить его естественное происхождение, тогда как выбытие архангельского кейса как раз свидетельствует в пользу подгонки размеров участков к полукруглым числам.

В случае участия избирателей в выборах обнаруживается только 1 подозрительный кейс: $\rho\alpha = 4,5$ для выборов 2016 г. в Тамбовской обл. Доминирующая цифра – «2». Исключение участков с абсолютной явкой картины не меняет.

Обнаружение лишь единичных подозрительных кейсов для распространенности непривлекательных цифр при отсутствии исключительных и, тем более, невероятных подтверждает достаточность выбранного порога отсека в 20 избирателей. Если бы карликовые участки при анализе были сохранены, то тест на избыточность последних цифр, отличных от «0», выявил бы для размера участка (участия избирателей) 7 (5) подозрительных кейсов, в т.ч. 3 (4) исключительных, в т.ч. 2 (2) невероятных.

Распределение минимума значимости нескольких не независимых гипотез неизвестно, поэтому не удается посчитать, сколько всего кейсов было фальсифицировано к какой-либо из цифр «1», «2», ... «9» в конце рассмотренных электоральных характеристик. Однако для каждой из цифр по отдельности такой расчет можно выполнить. Его результаты показаны на рис. 1. Как можно видеть, кроме чудовищного избытка цифры «0» имеется пусть и небольшой, но существенный избыток цифры «5» (2,5 и 4,2 стандартных отклонения для размера участка и участия в выборах соответственно). А из прочих цифр особенно сильно дискриминирована цифра «9». Желание округлить ее вверх, наверно, может возникнуть даже у тех жуликов, которые изначально не планировали фальсифицировать рассматриваемые характеристики именно таким способом.

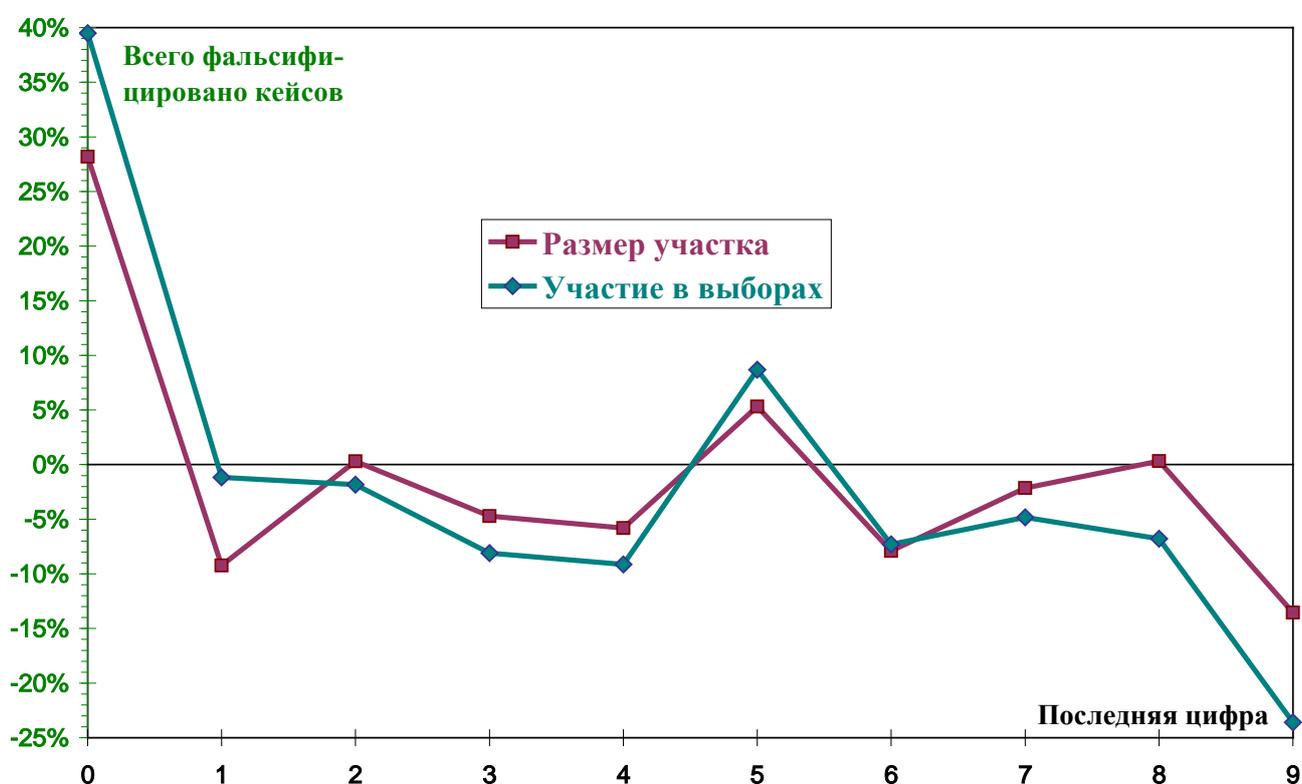


Рис. 1. Интегральный объем фальсификаций в пользу каждой из последних цифр электоральной характеристики

Отрицательные значения означают, что фальсификации осуществлялись за счет данной цифры, а не в ее пользу.

ПРОЦЕНТНЫЕ ЭЛЕКТОРАЛЬНЫЕ ХАРАКТЕРИСТИКИ

Наиболее интересны для фальсификаторов следующие электоральные характеристики, изменяемые процентами:

- *общая явка избирателей* – доля зарегистрированных избирателей, принявших участие в выборах (получивших бюллетень);
- *результат власти* – доля избирателей, поддержавших партию/кандидата власти, измеряемая от числа проголосовавших (опустивших бюллетень в урну).

Как уже было сказано, психологически привлекательными для процентных характеристик являются числа с нулем десятых, т.е. при фальсификациях круглыми оказываются промилле явки избирателей и результата власти. Сразу оговоримся, что для явки из рассмотрения исключаются участки с ее абсолютным значением, которое может возникать из-за отсутствия заранее составленных списков избирателей, что делает ее фиктивной величиной.

При анализе рассматриваемые величины превращаются в целые числа путем умножения на 1 000 и округления до ближайшего целого. Может показаться, что исследование так преобразованных процентных характеристик ничем не отличается от исследования изначально целочисленных. Однако процедуру существенно усложняет то обстоятельство, что процентные характеристики изначально являются не десятичными, а простыми дробями.

Получение целых процентов

В избирательных протоколах не указываются никакие проценты – только целые числа. И фальсификатор, выдумывающий значение дробной электоральной характеристики, вынужден как-то получать числитель и знаменатель дроби. При этом он более свободен в выборе числителя (число избирателей, принявших участие в выборах или поддержавших власть), нежели знаменателя (число зарегистрированных или проголосовавших избирателей). Поэтому числитель обычно получается умножением желательного целого процента на уже известный знаменатель и округлением результата до ближайшего целого. Если после этого снова рассчитать процент, то из-за ошибок округления он может уже и не оказаться целым. Как видно из врезки на рис. 2, при знаменателях, меньших 1 000, фальсификатор, выбравший психологически привлекательный целый процент, выдает себя лишь с некоторой вероятностью. Поэтому для повышения чувствительности теста может быть целесообразно исключить из рассмотрения небольшие участки, для которых знаменатель слишком мал.

Однако, с другой стороны, подтасовки чаще осуществляются именно на небольших участках, где хуже организован контроль голосования и подведения его итогов. Так что исключение таких участков, скорее всего, уменьшает общую долю фальсифицированных результатов, затрудняя их выявление. Но даже будь фальсификации равновероятны для участков любого размера, сохранение доли успехов k/n при уменьшении числа испытаний n повышает значимость гипотезы, тем самым снижая чувствительность теста. Поэтому здесь необходим некоторый баланс между борьбой с влиянием ошибок округления и сохранением представительности выборки.

Еще одно важное обстоятельство связано с тем, что дробные величины имеют склонность концентрироваться на значениях с малыми знаменателями ($\frac{1}{2}$, $\frac{1}{3}$, $\frac{2}{3}$, $\frac{1}{4}$, $\frac{3}{4}$ и т.п.). Чем меньше знаменатель, тем большим числом способов могут возникнуть такие значения. В результате одни последние цифры промилле могут оказаться более вероятны, чем другие. И чтобы приблизить их распределение к равномерному, опять-таки необходимо исключение небольших

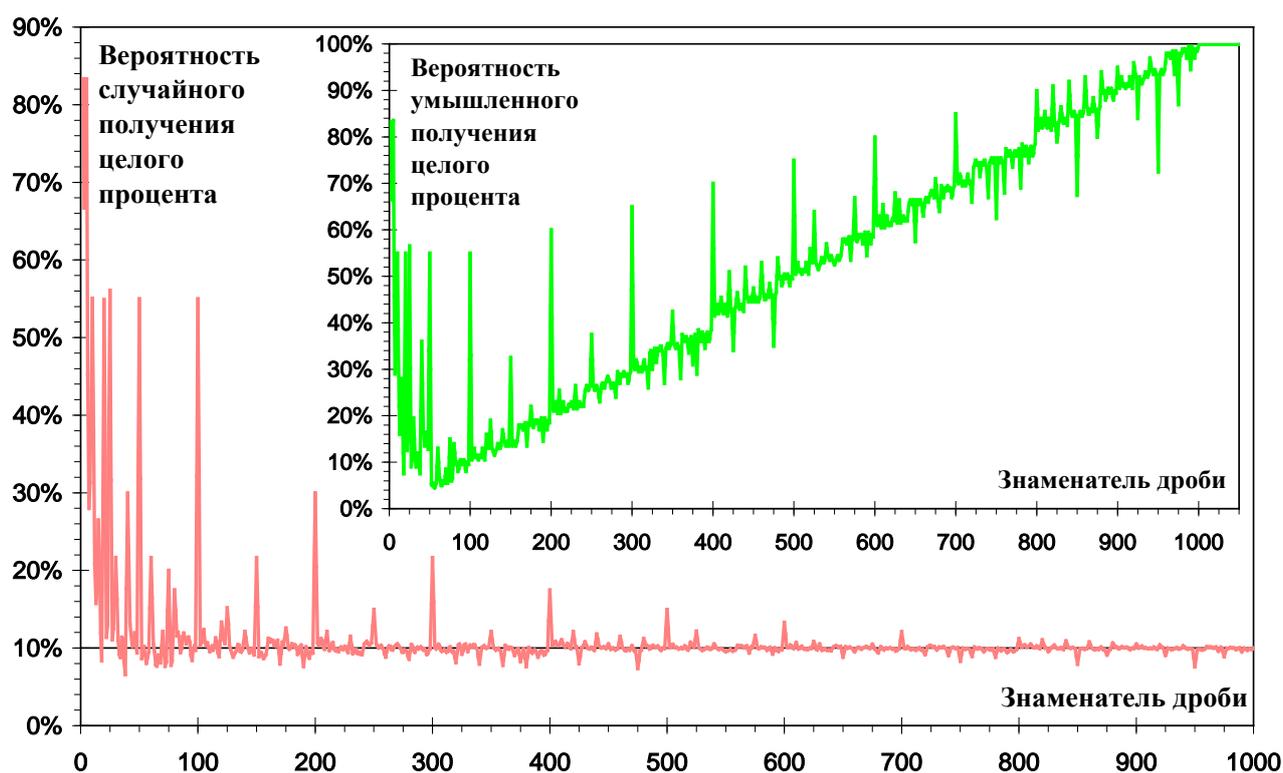


Рис. 2. Вероятность получения целого процента в зависимости от знаменателя дроби

Если числитель выбирается случайно, то по мере увеличения знаменателя вероятность приближается к 10%. Если же числитель вычисляется исходя из целого процента, то вероятность успеха возрастает примерно линейно, достигая 100% для четырехзначных знаменателей.

Оба графика сглажены биномиальный фильтром 2-го порядка.

участков. Но в этом случае оно обеспечивает уже не оптимальность теста, а его корректность. Как показывает рис. 2, вероятность получения целого процента при случайном числителе приближается к 10% для знаменателей, измеряемых сотнями.

Определение порогов отсечения

Приведенные выше соображения подсказывают установить порог отсечения по величине знаменателя в диапазоне от нескольких сотен до тысячи. К сожалению, для дробных электоральных характеристик невозможно, выбрав порог из каких-либо общих соображений, убедиться в адекватности его величины с помощью отрицательного контроля, как это было сделано для целочисленных характеристик.

Отрицательный контроль здесь в принципе неосуществим, поскольку у промилле явки и результата нарушение равновероятности последних цифр может возникать не только при независимом выдумывании разных круглых чисел, но и при организованном выдумывании одного и того же числа, не обязательно являющегося круглым. В последнем случае возникают сгустки – резкие выбросы на гистограмме распределения участков по величине рассматриваемой ха-

рактеристики. Методам выявления сгустков будет посвящена отдельная работа, но сама возможность их расположения на некруглых промилле не позволяет проверить, какой порог отсечения достаточен для обеспечения равновероятности последних цифр.

Поэтому пороги отсечения выбирались так, чтобы приблизить эмпирическую вероятность возникновения целых процентов к теоретической. На материале всех кейсов рассматривались участки со знаменателем (размером или голосованием) выше некоторого порога. Для них вычислялась вероятность получения целого процента (явки или результата) при случайном числителе. Полученная зависимость представлена на рис. 3.

В качестве порогов отсечения были выбраны минимальные значения знаменателя, при которых эмпирическая вероятность опускается до 10%. Далее для анализа явки учитываются участки с не менее чем 601 зарегистрированным избирателем, а для анализа результата – с не менее чем 523 проголосовавшими.

Оценим неточность определения показателя значимости ra , обусловленную возможной неточностью вероятности успеха p при испытании Бернулли. Эластичность значимости по вероятности может быть ограничена сверху с помощью формулы (2):

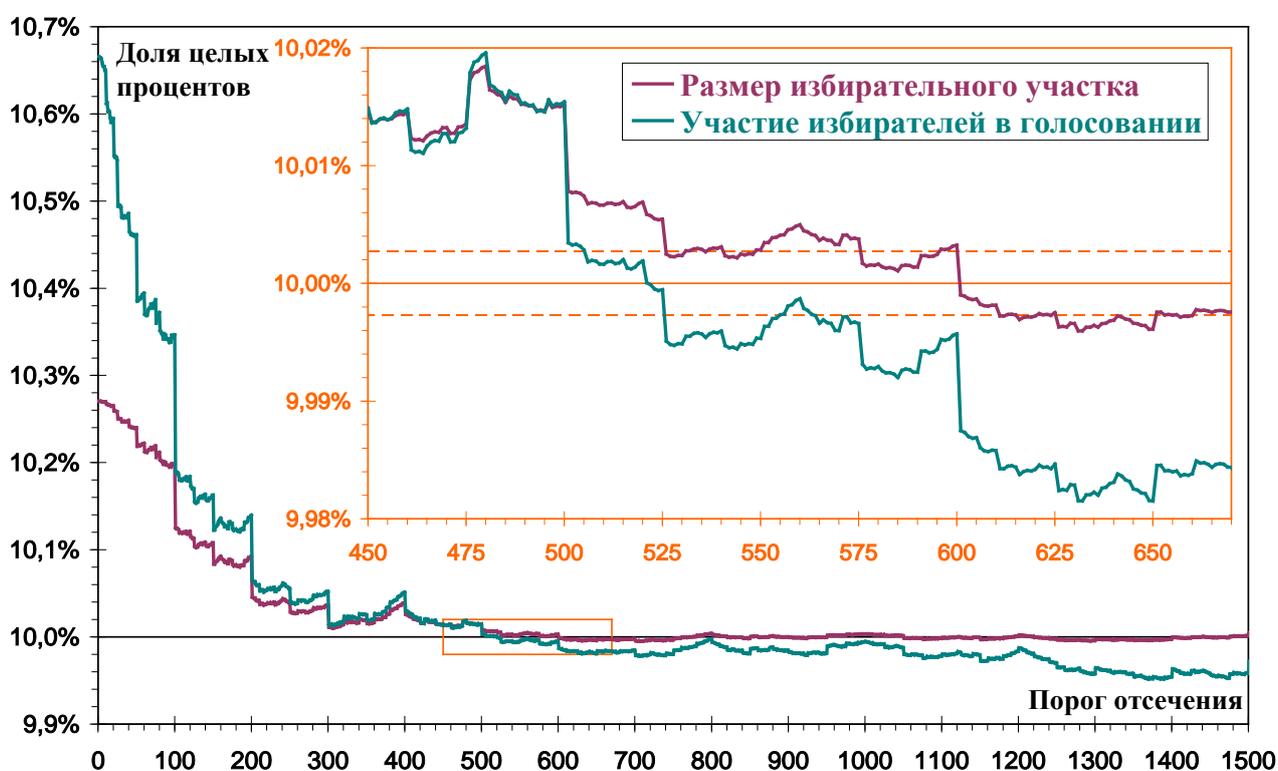


Рис. 3. Эмпирическая вероятность получения целых процентов

На врезке увеличен участок графика, где вероятность пересекает уровень в 10%. Пунктиром обозначены допуски, гарантирующие точность ra до десятых.

$$\frac{d \ln \alpha}{d \ln p} = p \frac{n C_{n-1}^{k-1} p^{k-1} q^{n-k}}{\sum_{i=k}^n C_n^i p^i q^{n-i}} = \frac{k}{1 + \frac{p}{q} \frac{n-k}{k+1} \left(1 + \frac{p}{q} \frac{n-k-1}{k+2} (1 + \dots) \right)} < k. \quad (4)$$

Отсюда получается ограничение и на погрешность показателя $\delta p \alpha < \lg e \cdot k/p \cdot \delta p$. При выбранных порогах отсечения регистрируемое число успехов $k < 430$ для всех кейсов. Если при этом потребовать допуска $\delta p < 0,0027\%$, то можно гарантировать $\delta p \alpha < 0,05$, т.е. вычисление показателя значимости с 1 точной цифрой после запятой. Реальная погрешность его определения будет несколько ниже – особенно при $k \approx pn$, когда первые множители перед скобками в формуле (4) еще близки к 1.

Рассчитанный уровень допуска δp обозначен на врезке рис. 3. Как можно видеть, для знаменателей, находящихся в широкой окрестности выбранных порогов отсечения, отклонение эмпирической вероятности от теоретического значения составляет как раз порядка δp . Поэтому возможные отклонения порогов, которые были бы оптимальны для отдельных кейсов, относительно значений, определенных для всего массива кейсов, не должны оказывать существенного влияния на точность результатов.

Обращает на себя внимание то, что эмпирическая вероятность получения целого процента существенно сильнее флуктуирует вокруг своего теоретического значения при использовании в качестве знаменателя участия в голосовании, чем размера участка. Это имеет две причины. С одной стороны, проголосовавших избирателей, очевидно, не больше чем зарегистрированных. Поэтому при одном и том же значении абсциссы ордината на рис. 3 вычисляется по выборке меньшего объема, что увеличивает флуктуации. А с другой стороны, как установлено ранее, фальсификация к круглым числам намного более распространена для участия избирателей по сравнению с размером участка. Ну а участие в голосовании отличается от участия в выборах лишь на число бюллетеней, которые избиратели забрали с собой. Поскольку выдуманные избиратели не склонны к такому поведению, то и голосование оказывается круглым намного чаще, чем размер. Вероятность получения целого процента при круглых знаменателях сложным образом зависит от того диапазона, в котором они находятся. Расчеты показывают, что она смещена как раз вверх и вниз для значений, относящихся соответственно к левым и правым частям приведенных графиков, что и вызывает такое их поведение.

Явка избирателей и результат власти

Значения показателей значимости гипотезы о естественном возникновении круглых промилле явки избирателей и результата власти представлены в табл. 4 и 5.

Таблица 4. Показатели значимости для круглого промилле явки избирателей

| Субъект федерации \ Год | 2000 | 2003 | 2004 | 2007 | 2008 | 2011 | 2012 | 2016 | 2018 | Всего |
|-------------------------|------|------|------|------|------|------|------|------|------|-------|
| 1. Татарстан | 0,3 | 0,6 | 1,0 | 2,5 | 10,7 | 18,4 | 22,6 | 67,7 | 33,1 | 80% |
| 2. Башкортостан | 1,1 | 0,3 | 29,1 | 13,2 | 30,3 | 17,3 | 17,6 | 36,0 | 20,3 | 88% |
| 3. Краснодарский кр. | 0,1 | 0,6 | 0,8 | 1,5 | 4,6 | 1,5 | 2,6 | 1,5 | 32,2 | 73% |
| 4. Дагестан | 1,4 | 1,4 | 23,6 | 7,8 | 13,7 | 1,0 | 7,9 | 22,8 | 27,7 | 96% |
| 5. Кемеровская обл. | 0,7 | 0,1 | 2,3 | 3,0 | 7,0 | 1,1 | 4,0 | 16,4 | 21,1 | 78% |
| 6. Кабардино-Балкария | 2,6 | 3,9 | 5,1 | 4,3 | 20,1 | 0,2 | 10,6 | 1,7 | 0,2 | 74% |
| 7. Северная Осетия | 0,4 | 0,8 | 3,9 | 5,8 | 9,6 | 0,1 | 0,5 | 5,9 | 1,7 | 64% |
| 8. Мордовия | 0,1 | 1,7 | 4,4 | 6,5 | 4,7 | 6,7 | 2,4 | 9,4 | 1,8 | 83% |
| 9. Карачаево-Черкесия | 0,7 | 0,0 | 0,3 | 1,5 | 8,8 | 3,7 | 1,8 | 0,7 | 4,3 | 61% |
| 10. Ямало-Ненецкий а/о | 0,0 | 0,0 | 0,4 | 0,2 | 2,1 | 0,3 | 7,6 | 2,6 | 0,5 | 25% |
| 11. Саратовская обл. | 0,2 | 0,4 | 0,5 | 1,4 | 2,4 | 0,7 | 1,8 | 5,6 | 6,7 | 66% |
| 12. Москва | 0,5 | 0,4 | 1,9 | 0,5 | 5,6 | 6,2 | 0,6 | 1,3 | 0,3 | 59% |
| 13. Ставропольский кр. | 0,2 | 0,4 | 0,1 | 1,1 | 0,3 | 0,1 | 0,1 | 0,4 | 5,6 | 6% |
| 14. Ростовская обл. | 0,1 | 0,1 | 1,2 | 1,7 | 2,1 | 0,1 | 1,8 | 4,6 | 2,6 | 45% |
| 15. Чечня | — | 3,5 | 4,5 | 0,0 | 0,4 | 0,0 | 0,0 | 0,1 | 0,2 | –18% |
| 16. Московская обл. | 0,3 | 0,4 | 0,6 | 2,8 | 3,9 | 0,6 | 1,3 | 0,6 | 1,9 | 64% |
| 17. Ингушетия | 0,0 | 0,4 | 0,1 | 1,9 | 1,0 | 0,4 | 3,8 | 0,7 | 0,1 | 26% |
| 18. Санкт-Петербург | 0,4 | 0,5 | 0,7 | 0,5 | 3,0 | 0,1 | 0,9 | 0,6 | 0,4 | 40% |
| 19. Тыва | 0,8 | 2,9 | 0,4 | 0,1 | 0,3 | 0,5 | 1,1 | 1,0 | 0,3 | 42% |
| Интегрально (по всем) | 2% | 4% | 19% | 12% | 28% | 4% | 13% | 31% | 33% | 16% |

Таблица 5. Показатели значимости для круглого промилле результата власти

| Субъект федерации \ Год | 2000 | 2003 | 2004 | 2007 | 2008 | 2011 | 2012 | 2016 | 2018 | Всего |
|-------------------------|------|------|------|------|------|------|------|------|------|-------|
| 1. Татарстан | 0,5 | 1,3 | 3,9 | 9,1 | 52,2 | 27,6 | 27,3 | 43,9 | 16,0 | 91% |
| 2. Башкортостан | 0,2 | 0,8 | 39,8 | 16,0 | 39,9 | 12,5 | 11,0 | 40,3 | 11,1 | 84% |
| 3. Дагестан | 3,6 | 6,6 | 17,0 | 5,0 | 10,5 | 5,5 | 11,4 | 29,2 | 16,2 | 100% |
| 4. Кабардино-Балкария | 2,6 | 2,7 | 3,3 | 6,6 | 24,7 | 5,0 | 1,7 | 1,8 | 0,1 | 82% |
| 5. Кемеровская обл. | 0,7 | 0,1 | 1,1 | 2,8 | 15,2 | 3,2 | 5,6 | 2,1 | 7,4 | 75% |
| 6. Краснодарский кр. | 0,4 | 0,1 | 0,5 | 0,1 | 14,1 | 2,6 | 3,3 | 1,5 | 5,2 | 46% |
| 7. Мордовия | 0,4 | 1,4 | 6,5 | 4,8 | 9,0 | 2,9 | 3,0 | 13,3 | 2,5 | 91% |
| 8. Ростовская обл. | 0,6 | 0,4 | 2,7 | 4,9 | 10,3 | 1,6 | 1,0 | 3,3 | 0,7 | 78% |
| 9. Москва | 0,0 | 0,0 | 0,4 | 0,2 | 1,1 | 9,5 | 0,4 | 1,4 | 0,3 | 16% |
| 10. Карачаево-Черкесия | 0,2 | 1,3 | 0,7 | 9,0 | 9,3 | 0,7 | 0,0 | 0,0 | 2,5 | 39% |
| 11. Саратовская обл. | 1,7 | 0,0 | 5,1 | 0,3 | 2,4 | 4,0 | 6,6 | 1,9 | 2,1 | 68% |
| 12. Чечня | — | 1,5 | 6,3 | 0,0 | 0,7 | 0,0 | 1,3 | 0,0 | 0,7 | 18% |
| 13. Ингушетия | 0,3 | 0,7 | 0,8 | 5,8 | 0,1 | 0,4 | 0,4 | 0,7 | 0,4 | 34% |
| 14. Тюменская обл. | 0,0 | 1,1 | 2,5 | 1,7 | 3,6 | 4,8 | 1,6 | 2,1 | 2,3 | 76% |
| 15. Северная Осетия | 0,3 | 0,0 | 1,5 | 3,1 | 2,5 | 0,5 | 0,4 | 4,8 | 0,2 | 42% |
| 16. Ставропольский кр. | 0,2 | 0,1 | 0,9 | 0,9 | 0,0 | 2,3 | 1,7 | 1,3 | 4,0 | 43% |
| 17. Брянская обл. | 0,1 | 0,1 | 0,3 | 0,4 | 0,2 | 0,4 | 0,2 | 3,0 | 1,2 | 10% |
| 18. Чукотский а/о | 0,5 | 0,8 | 0,0 | 0,5 | 2,9 | 0,7 | 0,0 | 0,2 | 1,3 | 35% |
| Интегрально (по всем) | 10% | 3% | 19% | 11% | 26% | 24% | 19% | 25% | 31% | 18% |

Для фальсификаторов эти электоральные характеристики являются целевыми, потому их психологически привлекательные значения позволяют проин-

терпретировать как исключительные, подозрительные или невероятные большее число избирательных кейсов, нежели размер участка или участие в выборах (см. табл. 8). С другой стороны, вероятностный характер получения круглого промилле в сочетании с необходимостью исключать из рассмотрения значительное число небольших участков приводят к тому, что процентные характеристики по сравнению с целочисленными дают существенно меньшую оценку общего объема фальсификаций.

Иначе говоря, явка и результат более чувствительны к массовым фальсификациям, но плохо видят эпизодические, а размер и участие – наоборот. Данный вывод косвенно подтверждается доминированием в табл. 4 и 5 субъектов, продемонстрировавших невероятные показатели, тогда как в табл. 1 и 2 таких субъектов оказывается не более половины. Сводка данных по подверженности субъектов фальсификациям приведена в табл. 9 на стр. 26.

Как уже было сказано, отрицательный контроль не работает для процентных электоральных характеристик. Из данных, приведенных в табл. 8, можно видеть, что имеются многочисленные случаи значимого преобладания в конце промилле как явки, так и результата цифр, отличных от «0», связанные с директивным определением этих характеристик при фальсификации (как и ранее для отрицательных контролей здесь используются интерпретационные пороги значимости, уменьшенные в 9 раз). Более того, многие выявленные случаи доминирования цифры «0» на самом деле тоже обусловлены тем, что на разных участках фальсификаторами выдумывался один и тот же процент, но просто он был выбран целым.

Случай абсолютной явки

При анализе явки не учитывались участки с ее абсолютным значением. Однако не все они возникают из-за невозможности заранее составить списки избирателей. Иногда 100%-я явка, которая тоже является психологически привлекательной, всё-таки выдумывается. Чтобы различить эти ситуации, необходимо иметь исчерпывающий перечень *специальных участков*, где избиратели голосуют не по месту жительства, а по месту временного пребывания, что зачастую не позволяет заранее составить списки. В общем случае получить такой перечень можно, только соотнося адреса участков с расположенными по этим адресам объектами, что крайне трудоемко. Однако представить, какой эффект на значимость гипотезы оказывает появление нескольких новых успехов в наборе испытаний, можно на примере Москвы, где для специальных участков используются выделенные ряды номеров: 3000+ в 1999-2004, 3100+ в 2007-08, 3200+ в 2011-12 и 3600+ в 2016-18 гг. Если для выборов в Москве выполнить тест на круглое промилле явки, исключая из рассмотрения ее абсолютные значения только для участков с номерами из этих рядов, то для 2008 г. ρ возрастет с 5,6 до 6,2, а для 2011 г. – с 6,2 до 8,8, что довольно существенно, хотя уже и не меняет интерпретации этих кейсов как невероятных. Увеличение показателя для выборов других лет не поднимает его над интерпретационными порогами и кажется не-

принципиальным. Однако за счет этого увеличения общий объем выявленных фальсификаций явки в Москве возрастает почти на 1,5 кейса (16% пунктов) – с 5,3 (59%) до 6,8 (76%).

Выполненный анализ требует дополнительного контроля на случай, если приведенные ряды содержат не все номера спецучастков (хотя об этом и не имеется никаких данных). Для их дополнительного поиска воспользуемся тем, что нарезка участков слабо менялась между парламентскими и следовавшими за ними президентскими выборами. И если участок с некоторым номером хотя бы для одних выборов из такой пары имел явку в 100%, а для других – не менее 90%, нельзя исключать того, что он тоже специализированный если не де-юре, то де-факто. Ни для одних выборов таких участков в Москве не было больше 5. Единственный заметный результат, к которому приводит их исключение при абсолютной явке состоит в непринципиальном уменьшении ρ в 2011 г. до 8,7.

Таким образом, имеются значительные резервы для повышения эффективности теста на круглое промилле явки, требующие, однако, учета типа избирательных участков. Это вряд ли выполнимо для всего набора кейсов, но в отдельных представляющих интерес случаях – вполне.

ОБЪЕДИНЕННЫЕ ТЕСТЫ

Выдумывание результатов может осуществляться по-разному в разных субъектах федерации, из-за чего тесты, основанные на определенных электро-ральных характеристиках, различаются по эффективности, вскрывая лишь свою часть массива фальсификаций. Поэтому разумно попытаться объединить результаты всех рассмотренных тестов. Сделать это можно двумя принципиально разными способами – вполне строгим, но довольно сложным и совсем простым, но приближительным и имеющим ограничения.

Многомерный тест

Рассмотрим сразу d независимых электро-ральных характеристик. Пусть для характеристики c в n_c испытаниях наблюдаются k_c успехов при фиксированных вероятностях успеха и неудачи p и $q = 1 - p$. Вероятность регистрации $\mathbf{i} = \{i_c\}$ успехов в $\mathbf{n} = \{n_c\}$ испытаниях

$$P_{\mathbf{n}}^{\mathbf{i}} = \prod_{c=1}^d \frac{n_c!}{i_c! j_c!} p^{i_c} q^{j_c}, \text{ где } \mathbf{j} = \mathbf{n} - \mathbf{i}.$$

Для определения значимости гипотезы о естественном возникновении наблюдаемого исхода $\mathbf{k} = \{k_c\}$ необходимо просуммировать эти вероятности по *критической области*, к которой принадлежат все те исходы, которые «не лучше» исхода \mathbf{k} с точки зрения проверяемой гипотезы.

Для сравнения векторных исходов воспользуемся методом максимального правдоподобия, опирающегося на определение таких гипотетических вероятностей успеха p_c , при которых вероятность именно рассматриваемого исхода была бы максимальной. С точностью до постоянного коэффициента функция правдоподобия имеет вид

$$L = \prod_{c=1}^d p_c^{i_c} q_c^{j_c}, \text{ где } q_c = 1 - p_c.$$

Пространство вероятностей $\{p_c\}$ распадается на области нулевой гипотезы $H_0: \forall c: p_c \leq p$ и альтернативной гипотезы $H_1: \exists c: p_c > p$. Соответственно, обозначим $L_0 = \sup_{H_0} L$ и $L_1 = \sup_{H_1} L$ максимальные правдоподобия этих гипотез. Скалярное отношение правдоподобия $\lambda = L_0/L_1$ и позволяет сравнивать векторные исходы между собой.

Положение глобального максимума G функции $L: \tilde{p}_c = i_c/n_c$ ($\tilde{q}_c = j_c/n_c$) находится как решение уравнений

$$\frac{\partial \ln L}{\partial p_c} = \frac{i_c}{p_c} - \frac{j_c}{q_c} = 0.$$

Точка $G \in H_0$, если для всех c без исключения выполнены условия $\tilde{p}_c \leq p$, и $G \in H_1$, если хотя бы одно из них нарушено. Для области, в которую точка G не попала, функция правдоподобия достигает своей точной верхней грани где-то на границе областей, перпендикулярной одной из осей p_c .

Введем обозначения $u_c = p^{i_c} q^{j_c}$, $\tilde{u}_c = \tilde{p}_c^{i_c} \tilde{q}_c^{j_c}$ и $v_c = u_c/\tilde{u}_c = (p/\tilde{p}_c)^{i_c} (q/\tilde{q}_c)^{j_c}$.

Если $G \in H_0$, т.е. $\forall c: i_c \leq pn_c$, то $L_0 = \prod_{c=1}^d \tilde{u}_c$ и $L_1 = \max_c \{v_c\} \cdot \prod_{c=1}^d \tilde{u}_c$, т.о. $\lambda^{-1} = \max_c \{v_c\}$.

Если $G \in H_1$, т.е. $\exists c: i_c > pn_c$, то $L_0 = \prod_{c: i_c > pn_c} u_c \cdot \prod_{c: i_c \leq pn_c} \tilde{u}_c$ и $L_1 = \prod_{c=1}^d \tilde{u}_c$, т.о. $\lambda = \prod_{c: i_c > pn_c} v_c$.

В частности, для одномерного случая, рассмотренного ранее, $\lambda = 1/v$ при $i \leq pn$ и $\lambda = v$ при $i > pn$, где $v = (np/i)^i (nq/j)^j$ достигает максимума при $i = pn$, что означает монотонную зависимость $\lambda(i)$, делающую критическую область простым лучом $i \geq k$.

В общем случае в критическую область входят все потенциальные исходы \mathbf{i} , такие что $\lambda(\mathbf{i}) < \lambda(\mathbf{k})$ или же $\lambda(\mathbf{i}) = \lambda(\mathbf{k})$, но при этом $P_n^{\mathbf{i}} \geq P_n^{\mathbf{k}}$. Выбор здесь бóльших, а не меньших вероятностей отражает то обстоятельство, что вырож-

дение отношения λ происходит по тем координатам, для которых $i_c \leq pn_c$. При этом вероятность исхода еще является возрастающей функцией числа успехов i_c , а значит, именно увеличение P соответствует сдвигу в сторону критической области.

Вхождение в отношение правдоподобия данных не от всех характеристик является основной, но не единственно возможной причиной вырождения. Иногда оно происходит и случайно. А его необходимо снять полностью, чтобы быть уверенным в том, что при справедливости нулевой гипотезы вероятность получения ее значимости, не превосходящей наблюдаемой величины, равна этой величине.

Снять вырождение можно двумя способами. Первый – учитывать вырожденные исходы с весом, обратно пропорциональным кратности вырождения, второй – принудительно упорядочить их. Первый способ, будучи однозначным, чрезвычайно сложен в вычислительном отношении, поэтому в силу крайней редкости вырождения предпочтительным здесь сочтен второй способ, пусть и предполагающий некоторый произвол. Характеристики сортируются по возрастанию их *важности*, т.е. в том порядке, в котором они вводятся в настоящей работе, и при $\lambda(\mathbf{i}) = \lambda(\mathbf{k})$ и $P_n^i = P_n^k$ исход \mathbf{i} не включается в критическую область, только если первая по важности несовпадающая координата c векторов \mathbf{i} и \mathbf{k} такова, что $i_c < k_c$.

Суммирование вероятностей по критической области дает значимость α . Ее десятичные показатели для 4-мерного теста, использующего все описанные выше электоральные характеристики, приведены в табл. 6. А в табл. 8 и 9 представлены итоговые данные не только для него, но и для тестов меньшей размерности, использующих все возможные пары и тройки характеристик, чтобы была понятна ценность каждой из них при таком подходе. Вполне ожидаемо, чем выше размерность теста, тем он, как правило, оказывается эффективнее. В отношении числа проинтерпретированных кейсов (табл. 8) это правило не знает исключений. В отношении числа уличенных в фальсификациях субъектов (табл. 9) тесты большей размерности обычно выигрывают у тестов меньшей размерности, использующих те же характеристики, а если и проигрывают кому-то из них, то не более 1 субъекта.

Таблица 6. Показатели значимости для комбинации круглых размера участка, участия в выборах, промилле явки избирателей и промилле результата власти

| Субъект федерации \ Год | 2000 | 2003 | 2004 | 2007 | 2008 | 2011 | 2012 | 2016 | 2018 | Всего |
|-------------------------|------|------|------|------|------|-------|------|-------|------|-------|
| 1. Татарстан | 0,9 | 1,3 | 5,2 | 20,0 | 82,6 | 43,5 | 48,1 | 109,3 | 48,5 | 96% |
| 2. Дагестан | 18,8 | 41,8 | 70,5 | 29,7 | 62,7 | 103,7 | 42,9 | 79,4 | 50,1 | 100% |
| 3. Башкортостан | 3,9 | 0,7 | 69,4 | 40,1 | 74,2 | 29,7 | 28,6 | 73,4 | 30,7 | 96% |
| 4. Кабардино-Балкария | 7,2 | 9,8 | 12,9 | 13,8 | 51,5 | 5,2 | 12,7 | 3,4 | 0,1 | 82% |
| 5. Краснодарский кр. | 0,7 | 2,7 | 2,1 | 3,2 | 29,2 | 3,3 | 6,4 | 5,1 | 43,2 | 96% |
| 6. Чечня | — | 17,2 | 34,4 | 9,1 | 23,0 | 5,2 | 5,4 | 0,8 | 1,4 | 95% |
| 7. Кемеровская обл. | 0,6 | 0,0 | 6,7 | 6,6 | 27,6 | 2,6 | 8,4 | 20,0 | 29,7 | 74% |
| 8. Мордовия | 0,1 | 2,3 | 9,6 | 11,6 | 12,7 | 8,0 | 3,7 | 20,1 | 2,6 | 83% |
| 9. Карачаево-Черкесия | 1,0 | 0,8 | 4,6 | 11,6 | 19,9 | 3,5 | 2,7 | 0,8 | 7,7 | 90% |
| 10. Северная Осетия | 1,8 | 1,0 | 9,0 | 8,1 | 13,0 | 2,8 | 8,3 | 17,5 | 10,5 | 97% |
| 11. Москва | 0,2 | 0,2 | 1,7 | 1,6 | 6,8 | 13,3 | 1,1 | 1,4 | 0,0 | 49% |
| 12. Ростовская обл. | 0,6 | 0,2 | 6,2 | 6,5 | 10,7 | 1,8 | 3,4 | 6,7 | 2,1 | 80% |
| 13. Ставропольский кр. | 0,5 | 0,2 | 1,2 | 0,8 | 0,7 | 1,7 | 3,6 | 1,4 | 9,8 | 68% |
| 14. Ямало-Ненецкий а/о | 0,9 | 0,6 | 0,8 | 1,8 | 3,8 | 2,0 | 9,1 | 1,9 | 0,7 | 82% |
| 15. Саратовская обл. | 2,1 | 1,1 | 6,1 | 0,7 | 2,9 | 5,9 | 6,5 | 9,0 | 7,5 | 93% |
| 16. Ингушетия | 1,7 | 0,3 | 0,3 | 6,0 | 0,6 | 0,4 | 3,2 | 1,4 | 1,0 | 59% |
| 17. Тюменская обл. | 0,2 | 0,5 | 2,2 | 2,1 | 5,6 | 5,2 | 0,9 | 4,5 | 2,9 | 76% |
| 18. Самарская обл. | 0,9 | 0,6 | 0,4 | 0,1 | 5,5 | 0,2 | 0,6 | 0,8 | 1,2 | 45% |
| 19. Липецкая обл. | 1,0 | 0,0 | 0,2 | 0,1 | 5,5 | 0,5 | 1,3 | 1,1 | 1,7 | 35% |
| 20. Московская обл. | 2,0 | 0,3 | 1,6 | 3,3 | 5,2 | 0,5 | 1,1 | 0,4 | 3,5 | 71% |
| 21. Сахалинская обл. | 0,8 | 0,7 | 2,3 | 4,5 | 2,1 | 1,3 | 2,9 | 1,4 | 0,9 | 88% |
| 22. Крым | — | — | — | — | — | — | — | 0,1 | 4,5 | 24% |
| 23. Хабаровский кр. | 0,3 | 1,4 | 0,3 | 0,3 | 0,5 | 0,8 | 0,7 | 0,2 | 4,0 | 34% |
| 24. Санкт-Петербург | 1,4 | 0,6 | 0,2 | 0,8 | 3,7 | 1,0 | 1,1 | 0,4 | 0,4 | 55% |
| 25. Корякский а/о | 3,6 | 0,1 | 2,0 | — | — | — | — | — | — | 49% |
| 26. Коми | 1,4 | 0,5 | 0,6 | 0,7 | 3,3 | 0,0 | 0,5 | 1,4 | 0,2 | 37% |
| 27. Пермская обл. | 0,7 | 3,3 | 2,4 | — | — | — | — | — | — | 85% |
| 28. Ярославская обл. | 0,1 | 1,7 | 0,3 | 3,2 | 1,7 | 0,3 | 0,1 | 0,2 | 0,7 | 21% |
| 29. Брянская обл. | 0,4 | 1,2 | 1,0 | 0,1 | 1,4 | 0,1 | 0,1 | 3,0 | 2,9 | 36% |
| Интегрально (по всем) | 31% | 19% | 37% | 40% | 60% | 39% | 36% | 46% | 56% | 40% |

Отдельной проблемой для многомерных тестов является оценка интегрального объема фальсификаций, поскольку при увеличении размерности стремительно возрастает вычислительная сложность расчета средней значимости $\bar{\alpha}$. И если в 2-мерных тестах ее удалось вычислить для всех кейсов, то в 3-мерных пришлось ограничиться кейсами, включающими не более 1 000 избирательных участков, а в 4-мерном это ограничение ужесточилось до 200 участков. Однако, как было отмечено выше, по мере роста числа испытаний распределение значимостей приближается к равномерному. При этом $\bar{\alpha} \rightarrow \frac{1}{2}$ сверху, и с увеличением размерности этот процесс ускоряется. Поэтому в тех случаях, когда среднюю значимость не удастся определить непосредственно, ее можно с высокой точностью оценить. В качестве нижней оценки используется значение $\frac{1}{2}$, а в качестве верхней – минимальное значение, определенное для

входящих тестов меньшей размерности (например, для 3-мерного теста, использующего любые 3 характеристики, вычислялся минимум средней значимости для всех 2-мерных тестов, использующих эти характеристики попарно). Разница между результатами, получаемыми для этих двух оценок, определяет дополнительную погрешность интегрального объема фальсификаций. Ею можно пренебречь, поскольку она оказывается на 2 порядка ниже систематической погрешности, рассчитанной ранее. Впрочем, для еще большего снижения погрешности используются средние арифметические обеих оценок.

На фоне успехов многомерных тестов в интерпретации субъектов неожиданным оказывается то, что максимальная оценка интегрального объема фальсификаций существенно не увеличивается при повышении размерности, замерев на уровне чуть более 300 кейсов (примерно 40%), даваемом в одиночку тестом на круглое участие в выборах. Можно предположить, что этот предел обусловлен не ограничениями предложенных методик, а распространенностью фальсификаций, связанных с выдумыванием электоральных характеристик.

На уровне 2-мерных тестов имеется отчетливый антагонизм в эффективности выявления фальсификаций целевых и нецелевых электоральных характеристик. Тест, использующий размер участка и участие избирателей, оказывается среди всех 2-мерных тестов лучшим по общему объему выявляемых фальсификаций, но одновременно и худшим – по числу проинтерпретированных кейсов. Иначе говоря, целочисленные электоральные характеристики, не являющиеся целевыми для фальсификаторов, выдумываются хоть и много где, но обычно не очень массово. А вот 2-мерный тест, использующий явку избирателей и результат власти, будучи одним из лучших по числу проинтерпретированных кейсов, оказывается худшим по общему объему выявляемых фальсификаций. Иначе говоря, целевые для фальсификаторов процентные электоральные характеристики если уж выдумываются, то массово, что компенсирует их неудобство для тестов (вероятностный характер видимости фальсификации и высокие пороги отсечения).

Отмеченный антагонизм исчезает уже на уровне 3-мерных тестов. Из них одновременно и самым лучшим по общему объему выявляемых фальсификаций, и одним из лучших по числу проинтерпретированных кейсов оказывается тест, использующий размер участка, участие в выборах и результат власти (см. табл. 8). Иначе говоря, если сосредоточиться на важнейшей из целевых характеристик, то побочные эффекты фальсификаций вполне успешно можно детектировать с помощью обеих нецелевых.

Множественный тест

Многомерный тест сложен и в теоретическом обосновании, и в практической реализации. Существует иной подход к объединению тестов, при котором ценой незначительного снижения строгости можно добиться примерно тех же результатов в интерпретации фальсифицированных кейсов.

Этот подход связан с описанием кейса вероятностью ошибочно отклонить одновременно несколько гипотез о естественном возникновении избытка

психологически привлекательных значений для d различных электоральных характеристик. В предположении их независимости эта вероятность ξ и ее показатель $p\xi$ даются соответственно произведением и суммой значимостей и их показателей для индивидуальных гипотез: $\xi = \prod_{c=1}^d \alpha_c$ и $p\xi = \sum_{c=1}^d p\alpha_c$.

Интерпретационные пороги для произведения ξ даются квантилями ее распределения уровней μ/N . К сожалению, определить это распределение удастся только приближенно, если считать значимости α_c равномерно распределенными на $[0;1]$, что, как уже было сказано, справедливо лишь в пределе бесконечного числа испытаний.

Исходя из равномерного распределения сомножителей $\text{Prob}\{\alpha \leq z\} = z$, индукционным путем легко находим и распределение их произведения $\text{Prob}\{\xi \leq z\} = z \cdot h(-\ln z)$, где $h(t) = \sum_{c=0}^{d-1} t^c / c!$ – отрезок разложения экспоненты.

Десятичные показатели порогов p легко определяют с помощью стремительно сходящегося итерационного процесса $p := \lg h(p/\lg e)N/\mu$. Их значения, рассчитанные для $N = 769$ и различных d , приведены в табл. 7.

Таблица 7. Пороги для интерпретации $p\xi$

| Интерпретация \ d | 1 | 2 | 3 | 4 |
|---------------------|------|------|------|------|
| Подозрительно | 2,89 | 3,88 | 4,74 | 5,53 |
| Исключительно | 3,89 | 4,98 | 5,92 | 6,77 |
| Невероятно | 4,89 | 6,06 | 7,06 | 7,97 |

Индивидуальные результаты кейсов для множественного теста здесь не приводятся. Табл. 8 позволяет убедиться, что при использовании как различных пар или троек электоральных характеристик, так и их полной четверки, общая эффективность множественного теста оказывается почти такой же (хотя и чуть худшей), как и у описываемого далее многомерного теста. То же самое можно сказать и про интерпретацию конкретных кейсов.

Таким образом, преимущество множественного теста состоит лишь в его вычислительной простоте, покупаемой приближенностью интерпретационных порогов. Кроме того, в отличие от многомерного теста, для множественного не удастся определить интегральный объем фальсификаций, что является ценой «лобового» подхода к объединению разных электоральных характеристик. Поэтому основные результаты здесь приводятся и анализируются именно для многомерного теста.

Динамика фальсификаций

Возможность отслеживать подверженность фальсификациям как значений отдельных электоральных характеристик, так и результатов выборов в целом, дает материал для рассмотрения эволюции фальсификационного дела. Рис. 4 демонстрирует, как менялась от выборов к выборам доля субъектов, затронутых фальсификациями. Рекорд здесь, вне всякого сомнения, устанавли-

вают выборы 2008 г., а за второе место конкурируют выборы 2007 и 2018 гг. с небольшим преимуществом последних.

В целом графики для явки избирателей и результата власти ведут себя примерно одинаково за исключением заметно большей распространенности фальсификаций результата на парламентских выборах 2007 и особенно 2011 гг. Это было время, когда партия власти шла к конституционному большинству в парламенте. После достижения этой цели уже фальсификации явки стали чуть более распространенными, чем фальсификации результата, что может быть попыткой скрыть падение интереса избирателей к выборам, которые более ни на что не влияют.

По мере ужесточения требований (переход от общей доли к долям всё более категорично интерпретированных фальсификаций) относительная мощь тестов, использующих целочисленные характеристики, ослабевает. Для общей доли тест на круглое участие почти так же эффективен как 4-мерный тест. Для выявления подозрительных кейсов все электоральные характеристики поодиночке примерно одинаково эффективны. Для исключительных и невероятных кейсов больше фальсификаций позволяют выявить процентные характеристики.

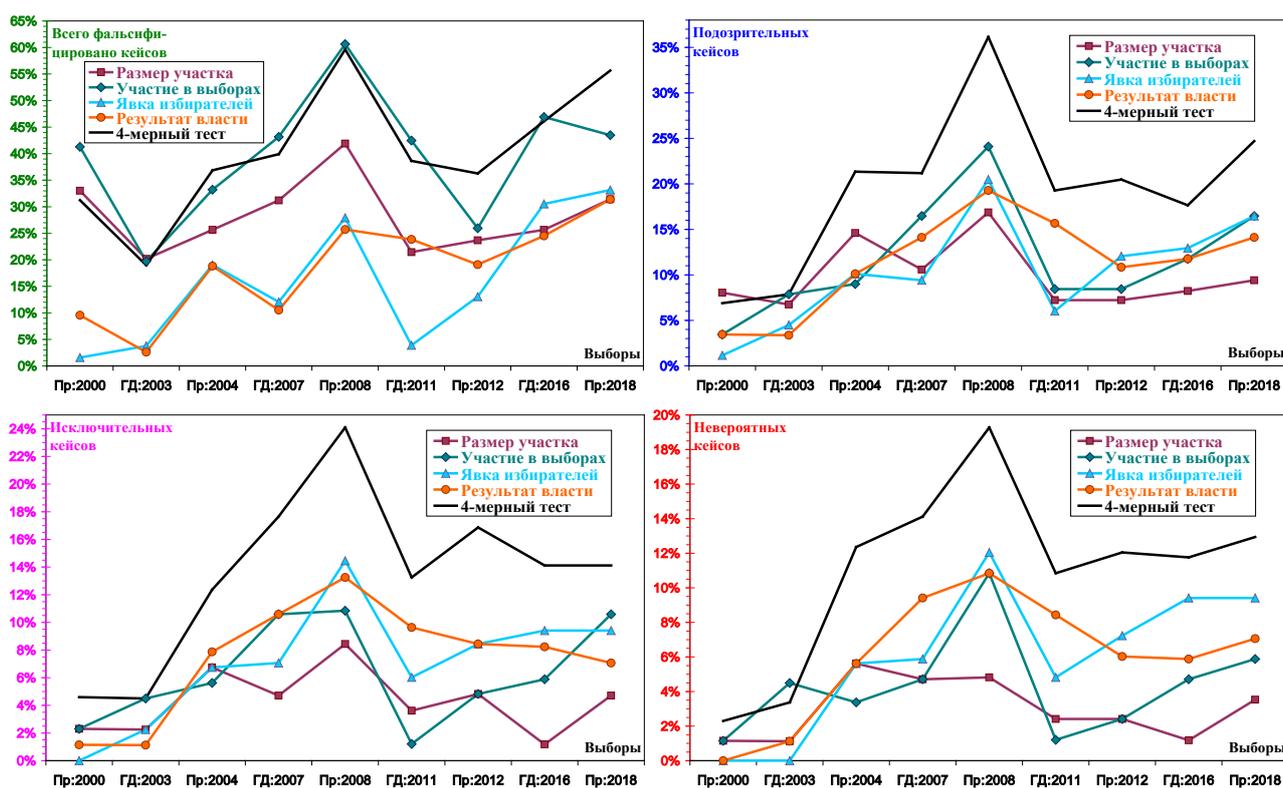


Рис. 4. Динамика долей интегральных и интерпретированных фальсификаций к психологически привлекательным числам

Интегральные графики (левый верхний рисунок) воспроизводят последние строчки табл. 1, 2, 4, 5 и 6. Данные интерпретационных графиков можно получить, подсчитывая по столбцам этих таблиц раскрашенные показатели (включая фиолетовый в синий, а красный – и в тот, и в другой), т.е. доли, показанные на последовательных рисунках, включают в себя друг друга так же, как и значения в столбцах в табл. 8.

ИТОГИ И ОБСУЖДЕНИЕ

Таблица 8. Количество и доля фальсифицированных избирательных кейсов

| Тест | | Всего | | Подозри- тельные | | Исклю- чительные | | Неверо- ятные | |
|--|---|-------|-----|---------------------|------|---------------------|------|------------------|------|
| Круглый размер участка (табл. 1) | | 217 | 28% | 34 | 4,4% | 23 | 3,0% | 17 | 2,2% |
| То же, но без участков с абсолютной явкой | | 185 | 24% | 27 | 3,5% | 19 | 2,5% | 13 | 1,7% |
| Круглое участие в выборах (табл. 2) | | 304 | 39% | 48 | 6,2% | 34 | 4,4% | 24 | 3,1% |
| То же, но без участков с абсолютной явкой | | 277 | 36% | 42 | 5,5% | 28 | 3,6% | 21 | 2,7% |
| То же, но без жадных участков | | 234 | 30% | 28 | 3,6% | 14 | 1,8% | 8 | 1,0% |
| То же, но без участков обоих типов (табл. 3) | | 197 | 26% | 15 | 2,0% | 10 | 1,3% | 5 | 0,7% |
| Круглое промилле явки (табл. 4) | | 123 | 16% | 55 | 7,2% | 48 | 6,2% | 38 | 4,9% |
| Некруглое промилле явки | | — | — | 7 | 0,9% | 5 | 0,7% | 4 | 0,5% |
| Круглое промилле результата власти (табл. 5) | | 141 | 18% | 57 | 7,4% | 47 | 6,1% | 40 | 5,2% |
| Некруглое промилле результата власти | | — | — | 12 | 1,6% | 8 | 1,0% | 5 | 0,7% |
| 2-мерный тест | размер участка и участие в выборах | 320 | 42% | 74 | 9,6% | 49 | 6,4% | 35 | 4,6% |
| 2 множителя | | — | — | 73 | 9,5% | 48 | 6,2% | 34 | 4,4% |
| 2-мерный тест | размер участка и явка избирателей | 219 | 28% | 76 | 9,9% | 64 | 8,3% | 57 | 7,4% |
| 2 множителя | | — | — | 74 | 9,6% | 63 | 8,2% | 57 | 7,4% |
| 2-мерный тест | размер участка и результат власти | 249 | 32% | 74 | 9,6% | 62 | 8,1% | 50 | 6,5% |
| 2 множителя | | — | — | 73 | 9,5% | 61 | 7,9% | 50 | 6,5% |
| 2-мерный тест | участие в выборах и явка избирателей | 276 | 36% | 79 | 10% | 62 | 8,1% | 53 | 6,9% |
| 2 множителя | | — | — | 77 | 10% | 59 | 7,7% | 52 | 6,8% |
| 2-мерный тест | участие в выборах и результат власти | 294 | 38% | 80 | 10% | 64 | 8,3% | 56 | 7,3% |
| 2 множителя | | — | — | 77 | 10% | 62 | 8,1% | 56 | 7,3% |
| 2-мерный тест | явка избирателей и результат власти | 160 | 21% | 79 | 10% | 65 | 8,5% | 55 | 7,2% |
| 2 множителя | | — | — | 81 | 11% | 65 | 8,5% | 55 | 7,2% |
| 3-мерный тест | участие, явка и результат (все кроме размера) | 273 | 36% | 91 | 12% | 74 | 9,6% | 65 | 8,5% |
| 3 множителя | | — | — | 88 | 11% | 75 | 9,8% | 65 | 8,5% |
| 3-мерный тест | размер, явка и результат (все кроме участия) | 237 | 31% | 92 | 12% | 77 | 10% | 69 | 9,0% |
| 3 множителя | | — | — | 90 | 12% | 78 | 10% | 70 | 9,1% |
| 3-мерный тест | размер, участие и резуль- тат (все кроме явки) | 324 | 42% | 92 | 12% | 78 | 10% | 71 | 9,2% |
| 3 множителя | | — | — | 92 | 12% | 78 | 10% | 69 | 9,0% |
| 3-мерный тест | размер, участи и явка (все кроме результата) | 306 | 40% | 97 | 13% | 75 | 9,8% | 61 | 7,9% |
| 3 множителя | | — | — | 97 | 13% | 75 | 9,8% | 61 | 7,8% |
| 4-мерный тест (табл. 6) | размер, участие, явка и результат | 309 | 40% | 104 | 14% | 84 | 11% | 79 | 10% |
| 4 множителя | | — | — | 103 | 13% | 83 | 11% | 76 | 9,9% |

Таблица 9. Количество и доля субъектов федерации, для которых хотя бы раз были зафиксированы фальсификации

| Тест | Подозри- тельные | | Исклю- чительные | | Неверо- ятные | |
|--|---------------------|-----|---------------------|-----|------------------|-----|
| Круглый размер участка (табл. 1) | 14 | 16% | 10 | 12% | 7 | 8% |
| Круглое участие в выборах (табл. 2) | 21 | 25% | 15 | 18% | 10 | 12% |
| Круглое промилле явки избирателей (табл. 4) | 19 | 22% | 15 | 18% | 13 | 15% |
| Круглое промилле результата власти (табл. 5) | 18 | 21% | 16 | 19% | 13 | 15% |

| Тест | | Подозрительные | | Исключительные | | Невероятные | |
|-------------------------|---|----------------|-----|----------------|-----|-------------|-----|
| 2-мерный тест | размер участка и участие в выборах | 26 | 30% | 19 | 22% | 11 | 13% |
| 2 множителя | | 25 | 29% | 18 | 21% | 10 | 12% |
| 2-мерный тест | размер участка и явка избирателей | 24 | 28% | 17 | 20% | 16 | 19% |
| 2 множителя | | 21 | 25% | 17 | 20% | 16 | 19% |
| 2-мерный тест | размер участка и результат власти | 22 | 26% | 16 | 19% | 14 | 16% |
| 2 множителя | | 21 | 25% | 15 | 18% | 14 | 16% |
| 2-мерный тест | участие в выборах и явка избирателей | 27 | 32% | 21 | 25% | 16 | 19% |
| 2 множителя | | 25 | 29% | 18 | 21% | 15 | 18% |
| 2-мерный тест | участие в выборах и результат власти | 25 | 29% | 22 | 26% | 17 | 20% |
| 2 множителя | | 23 | 27% | 20 | 23% | 17 | 20% |
| 2-мерный тест | явка избирателей и результат власти | 20 | 23% | 17 | 20% | 16 | 19% |
| 2 множителя | | 21 | 25% | 17 | 20% | 16 | 19% |
| 3-мерный тест | участие, явка и результат (все кроме размера) | 26 | 30% | 21 | 25% | 18 | 21% |
| 3 множителя | | 26 | 30% | 20 | 23% | 18 | 21% |
| 3-мерный тест | размер, явка и результат (все кроме участия) | 24 | 28% | 18 | 21% | 17 | 20% |
| 3 множителя | | 23 | 27% | 18 | 21% | 18 | 21% |
| 3-мерный тест | размер, участие и результат (все кроме явки) | 28 | 33% | 22 | 26% | 19 | 22% |
| 3 множителя | | 28 | 33% | 21 | 25% | 18 | 21% |
| 3-мерный тест | размер, участи и явка (все кроме результата) | 30 | 35% | 22 | 26% | 16 | 19% |
| 3 множителя | | 31 | 36% | 21 | 25% | 16 | 19% |
| 4-мерный тест (табл. б) | размер, участие, явка и результат | 29 | 34% | 23 | 27% | 20 | 23% |
| 4 множителя | | 29 | 34% | 22 | 26% | 19 | 22% |

Сводные результаты применения разных тестов, представленные в табл. 8 и 9, подводят исчерпывающий итог предложенным методам, давая представление о распространенности выдумывания результатов и позволяя сравнить различные тестов по способности к его выявлению.

Современные информационно-компьютерные технологии, обеспечивая прозрачность (доступность информации о происходящих событиях всем желающим ее получить), сами по себе не гарантируют публичности (превращения доступной информации в осознанный факт, требующий принятия управленческих решений). В результате возникает драматический разрыв в уровне развития между обществом и его институтами, особо острый в нашей стране. Предложенный инструментарий призван отчасти устранить его в сфере наблюдения за чистотой выборов. За исключением многомерного теста все описанные методы выявления выдуманных результатов чрезвычайно просты. Проверка их теоретического обоснования доступна студенту технического вуза, а практическое применение – старшекласснику с компьютером.

Проведенный анализ, будучи заточенным под весьма специфический тип фальсификаций, оптимален для общей оценки не столько достоверности результатов выборов, сколько дееспособности избирательной и судебной ветвей власти в стране. Фальсификации, при которых содержимое избирательных протоколов не соответствует содержимому урн для голосования, элементарно подтверждаются независимым пересчетом бюллетеней. И то, что такой пересчет не выполняется

для тех случаев, когда есть основания подозревать выдуманность результатов, говорит о принципиальной непригодности существующей системы избирательных комиссий для поведения итогов голосования. А то, что даже математически строго доказанный факт фальсификации не получает автоматически надлежащего юридического оформления, говорит о принципиальной непригодности существующей судебной системы для разрешения конфликтов в электоральной сфере.

Особо следует отметить, что не вся территория страны массово подвержена выдумыванию результатов. Из 94 субъектов федерации, существовавших в разные годы, 55 ни разу ни в одном тесте не оказались хотя бы подозрительными. Еще 13 субъектов никогда не имели показателей хуже чем подозрительные, для которых еще высок риск случайного возникновения. Наконец, результаты выборов еще в 3 субъектах никогда не были невероятными.

Вместе с тем не следует полагать, что несклонность или малая склонность избирательных комиссий этих субъектов к выдумыванию результатов означает отсутствие в них фальсификаций вообще. В условиях полной безнаказанности фальсификаторов предполагаемая электоральная чистота субъекта, к сожалению, остается не заслугой его властей, а недоработкой исследователей. Поэтому необходимо распространение методов электоральной статистики как на другие типы фальсификаций, так и на выборы регионального уровня.

Автор считает приятным долгом поблагодарить С.А. Махова за неоценимую помощь с освоением теоретических основ проверки гипотез.

Содержание

| | |
|--|----|
| Постановка задачи..... | 3 |
| Базовая идея..... | 4 |
| Практические вычисления | 6 |
| Целочисленные электоральные характеристики | 7 |
| Размер избирательного участка..... | 7 |
| Участие избирателей в выборах | 9 |
| Отрицательный контроль | 11 |
| Процентные электоральные характеристики | 12 |
| Получение целых процентов..... | 13 |
| Определение порогов отсечения | 14 |
| Явка избирателей и результат власти | 16 |
| Случай абсолютной явки..... | 18 |
| Объединенные тесты..... | 19 |
| Многомерный тест | 19 |
| Множественный тест | 23 |
| Динамика фальсификаций..... | 24 |
| Итоги и обсуждение | 26 |