

А.В. Подлазов

**Формальное выявление выдуманных
результатов выборов**

Рекомендуемая форма библиографической ссылки

Подлазов А.В. Формальное выявление выдуманных результатов выборов // Проектирование будущего. Проблемы цифровой реальности: труды 3-й Международной конференции (6-7 февраля 2020 г., Москва). — М.: ИПМ им. М.В.Келдыша, 2020. — С. 176-190. — <https://keldysh.ru/future/2020/15.pdf> <https://doi.org/10.20948/future-2020-15>

Размещено также видео выступления

Формальное выявление выдуманных результатов выборов

А.В. Подлазов

Институт прикладной математики им. М.В. Келдыша РАН

Аннотация. Для выдуманных электоральных результатов типично преобладание круглых значений электоральных характеристик. Оно может быть выявлено методами математической статистики. Для федеральных выборов 2000-2018 гг. выдуманные результаты в значимых количествах обнаруживаются для примерно 40% избирательных кейсов, из которых не менее 10% удается 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

Formal identification of fictitious election results

A.V. Podlazov

RAS Keldysh Institute of Applied Mathematics

Abstract. A predominance of round values of electoral characteristics is typical for fictitious electoral results. It can be detected by methods of mathematical statistics. 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.

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

Постановка задачи

Среди методов выявления электоральных фальсификаций особое место занимают методы, обладающие самостоятельной доказательной силой. Они основаны на поиске с помощью инструментария математической статистики внутренних несообразностей в официальных результатах выборов. Такие несообразности возникают, в частности, при *выдумывании* результатов, т.е. внесении в избирательные протоколы

чисел, не находящихся в какой-либо связи с содержимым урн для голосования. Массив таких чисел приобретает свойства, нетипичные для истинных результатов. Их анализ здесь осуществляется на уровне субъектов федерации по причине существенной разницы между ними в масштабах и способах фальсификаций. Результаты президентских и парламентских выборов, проводившихся в нашей стране в XXI в., образуют 769 избирательных кейсов – голосований в данном субъекте в данном году.

В выдуманных результатах преобладают *психологически привлекательные числа*. Для целочисленных электоральных характеристик таковыми 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 в десятки и сотни единиц последняя цифра практически равновероятно принимает все возможные значения. Когда встречаемость цифры «0» в конце числа превышает 10%, можно предполагать наличие выдуманных результатов. При низкой *значимости* α гипотезы о естественном возникновении превышения 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) $q = 9/10$. Вероятность регистрации i успехов в n испытаниях $P_n^i = C_n^i p^i q^{n-i}$. Не менее k успехов регистрируются с вероятностью $\alpha = \alpha_n^k = \sum_{i=k}^n P_n^i$.

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 чрезвычайно мала. Поэтому далее для удобства вместо нее всюду приводится ее *десятичный показатель* $\rho\alpha = -\lg \alpha$, 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 $\alpha \leq \mu/N$. Если же таких кейсов больше, можно полагать их фальсифицированными. Параметр μ определяет количество допускаемых при этом ошибок. Далее

используются интуитивно понятные значения $\mu = 1, 1/10$ и $1/100$. Им соответствует интерпретация кейсов как *подозрительных, исключительных* и *невероятных* при $\rho\alpha \geq \lg N \approx 2,89$, $\rho\alpha \geq \lg 10 \cdot N \approx 3,89$ и $\rho\alpha \geq \lg 100 \cdot N \approx 4,89$ для 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 бóльших μ нецелесообразно, т.к. ведет к неприемлемому числу ложноположительных срабатываний теста. Вместе с тем значительная, если не основная, доля фальсифицированных кейсов имеет как раз сравнительно высокие α . И хотя без риска многочисленных ошибок эти кейсы не удастся назвать поименно, их вполне можно пересчитать.

Мерой достоверности кейса служит отношение $\alpha/\bar{\alpha}$, где математическое ожидание значимости $\bar{\alpha} = \sum_{k=0}^n P_n^k \alpha_n^k = \sum_{k=0}^n P_n^k \sum_{i=k}^n P_n^i$. В самом деле, если для возможных исходов испытания с вероятностью ε заменять истинные значения α нулями, то среднее значение этого отношения, очевидно, составит $1 - \varepsilon$. Таким образом, из N кейсов достоверны $T = \sum_{l=1}^N \alpha_l / \bar{\alpha}_l$, 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 при их отсутствии. При наличии фальсификаций ее погрешность, естественно, будет ниже.

Две меры фальсификаций – интегральная и интерпретационная – обладают принципиально разными свойствами. Первая – аддитивна, т.е. при разбиении выборки на части сумма найденных для них объемов фальсификаций совпадает с их объемом на полной выборке. Результаты интерпретации показателей, напротив, существенно зависят от размера выборки. Скажем, если рассматривать кейсы только конкретного года, она значительно сократится, что приведет к соответствующему уменьшению порогов интерпретации $\rho\alpha$. Те значения показателя, которые лишь при этом становятся подозрительными, в приводимых далее таблицах даны черным шрифтом, а те, которые всё равно остаются под порогом, – серым. А показатели подозрительные (исключительные, невероятные) и для

полной выборки выделяются жирным синим (фиолетовым, красными) шрифтом.

Кроме различий в свойствах введенные меры 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 виновных, это говорит о неспособности правоохранительной системы пресекать электоральные преступления.

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 не являются целевыми, они всё равно выдумываются, что позволяет на них апробировать методику и подготовиться к анализу более важных процентных характеристик. В данном разделе из рассмотрения исключаются участки, где размер и участие не достигают 20 чел. Введение такого порога отсечения позволяет избежать искажений, вносимых карликовыми спецучастками, размер которых задается штатными расписаниями определенных служб.

Для целого ряда 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 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.

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%

Субъект федерации \ Год	2000	2003	2004	2007	2008	2011	2012	2016	2018	Всего
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%

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%

В последнем столбце таблиц дается оценка общей доли фальсифицированных кейсов субъекта. 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). Здесь данные уже более надежны (в отсутствие фальсификаций погрешность составляет около 6% пунктов). Наконец, в правом нижнем углу таблиц дается оценка доли фальсификаций по всему массиву кейсов.

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.

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 определяется оно заранее. Однако если весь протокол заполняется выдуманными числами, можно и не выяснять, сколько на участке имеется избирателей.

Табл. 2 демонстрирует, что число избирательных бюллетеней, выданных избирателям, ожидаемо оказывается круглым значительно чаще, чем число самих избирателей, и позволяет выявить существенно большее количество фальсифицированных кейсов в большем числе субъектов. Из тестов, использующих лишь одну электоральную характеристику, этот – самый сильный 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.

Процентные электоральные характеристики

Для фальсификаторов наиболее интересны следующие электоральные характеристики, изменяемые процентами:

– *общая явка избирателей* – доля зарегистрированных избирателей, принявших участие в выборах (из рассмотрения исключаются участки со 100% явкой, которая может возникать из-за отсутствия заранее составленных списков избирателей, что делает ее фиктивной величиной);

– *результат власти* – доля избирателей, поддержавших партию/кандидата власти, измеряемая от числа проголосовавших.

Как уже было сказано, психологически привлекательными здесь являются круглые промилле. Поэтому может показаться, что исследование процентных характеристик, если умножить их на 1 000 и округлить до ближайшего целого, ничем не отличается от исследования целочисленных. Однако в избирательных протоколах не указываются никакие проценты – только целые числа. Иначе говоря, процентные характеристики изначально являются не десятичными, а простыми дробями, которые склонны 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, необходимо исключение небольших участков.

Пороги отсекаения выбирались так, чтобы приблизить эмпирическую вероятность возникновения целых процентов к теоретической. На материале всех кейсов рассматривались участки со знаменателем (числом зарегистрированных или проголосовавших избирателей) выше некоторого порога. Для них вычислялась вероятность получения целого процента (явки или результата) при случайном числителе (числе избирателей, принявших участие в выборах или поддержавших власть). Полученная зависимость представлена на рис. 1. Для явки эмпирическая вероятность опускается до 10% при учете участков с не менее чем 601 зарегистрированным избирателем, а для результата – с не менее чем 523 проголосовавшими.

Оценим погрешность определения показателя значимости α , обусловленную возможной неточностью вероятности успеха p при испытании Бернулли. Продифференцировав значимость по вероятности успеха

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

можно ограничить сверху эластичность значимости по вероятности:

$$\frac{d \ln \alpha}{d \ln p} = p \frac{n C_{n-1}^{k-1} p^{k-1} q^{n-k}}{\sum_{i=k}^n C_n^i p^i q^{n-i}} < k.$$

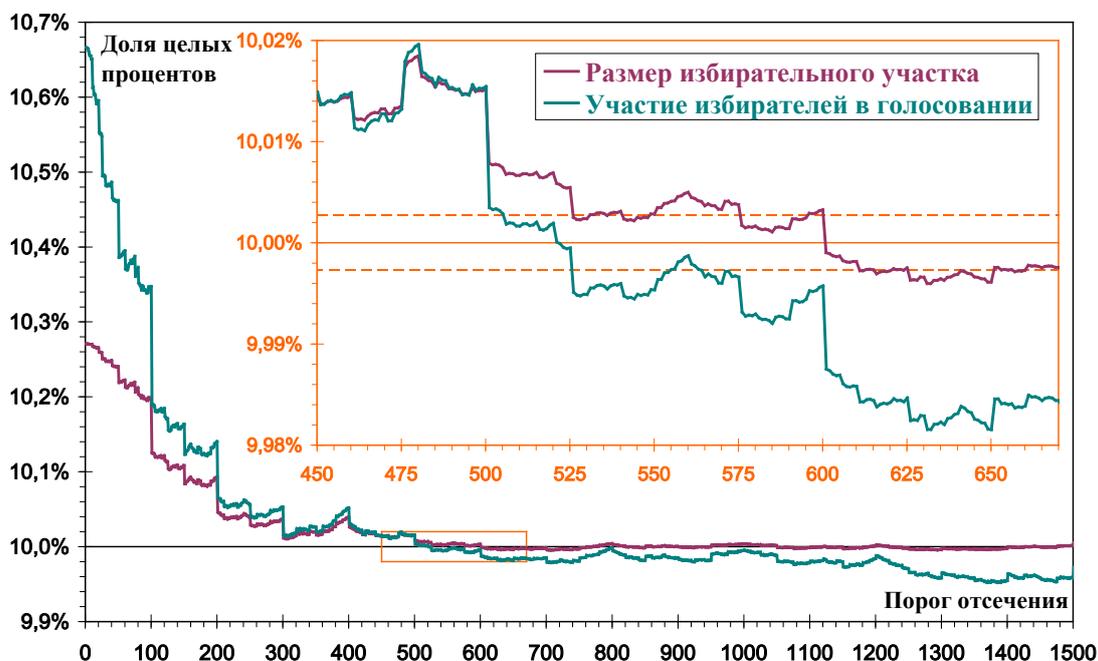


Рис. 1. Эмпирическая вероятность получения целых процентов. На врезке увеличен участок графика, где вероятность пересекает уровень в 10%. Пунктиром обозначены допуски, гарантирующие точность α до десятых.

Отсюда получается ограничение и на погрешность показателя $\delta p \alpha < k \cdot \lg e \cdot \delta p / p$. 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 одной точной цифрой после запятой.

Рассчитанный уровень допуска δp обозначен на врезке рис. 1. 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.

Показатели значимости гипотезы о естественном возникновении круглых промилле явки избирателей и результата власти, рассчитанные на основе определенных порогов отсечения, представлены в табл. 3 и 4.

Для фальсификаторов эти электоральные характеристики являются целевыми, потому их психологически привлекательные значения позволяют проинтерпретировать как исключительные, подозрительные или невероятные большее число избирательных кейсов, нежели размер участка или участие в выборах. 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. 3 и 4 субъектов, продемонстрировавших невероятные показатели, тогда как в табл. 1 и 2 таких субъектов оказывается не более половины.

Таблица 3. 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%

Субъект федерации \ Год	2000	2003	2004	2007	2008	2011	2012	2016	2018	Всего
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 4. 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%

Многомерный тест

Результаты выполненных выше тестов, вскрывающих разные части массива фальсификаций, можно объединить.

Рассмотрим сразу d независимых электоральных 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 α 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. Определяющая ее *функция правдоподобия* имеет вид $L \sim \prod_{c=1}^d p_c^{i_c} q_c^{j_c}$, 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 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 L достигает своей точной верхней грани где-то на границе областей, перпендикулярной одной из осей p_c . Примеры расположения соответствующих точек для $d = 2$ при различных соотношениях между \tilde{p}_c и p показаны на рис. 2.

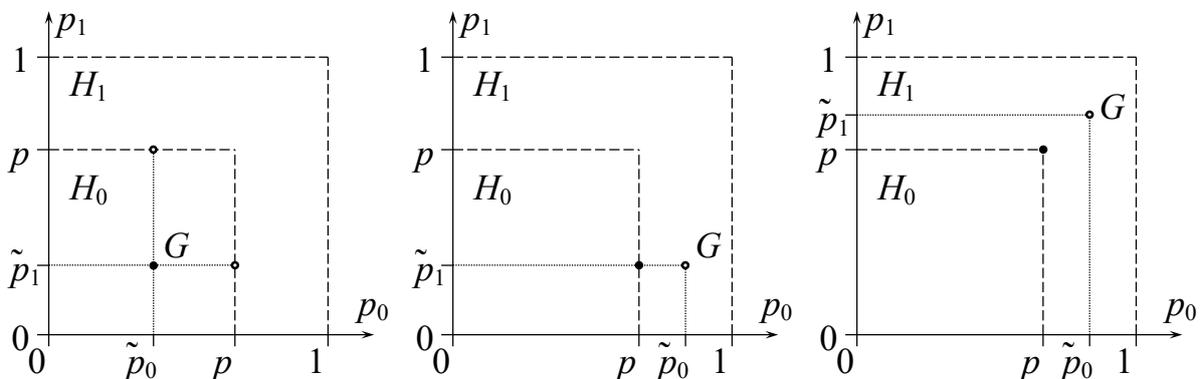


Рис. 2. Положения максимума функции правдоподобия в области нулевой (сплошная точка) и альтернативной (наколотая точка) гипотез

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

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. $\lambda^{-1} = \max_c \{v_c\}$.

А 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 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 вырождение 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 λ данных не от всех характеристик является основной, но не единственной возможной причиной вырождения. Иногда оно происходит и случайно. Для его окончательного устранения здесь использована сортировка электоральных характеристик 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 $P_n^{\mathbf{i}} = P_n^{\mathbf{k}}$ исход \mathbf{i} не включается в критическую область, только если первая по важности несовпадающая координата c векторов \mathbf{i} и \mathbf{k} такова, что $i_c < k_c$.

Чем выше размерность теста, тем он, как правило, оказывается эффективнее. В отношении числа проинтерпретированных кейсов это правило не знает исключений. В отношении числа уличенных в фальсификациях субъектов 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. Поэтому здесь рассмотрение ограничено 4-мерным тестом, учитывающим все электоральные характеристики. Для него показатели значимости приведены в табл. 5.

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 по мере роста числа испытаний распределение значимостей приближается к равномерному. При этом $\bar{\alpha} \rightarrow 1/2$ сверху, и с увеличением размерности этот процесс

ускоряется. Поэтому в тех случаях, когда среднюю значимость не удается 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. 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.

Таблица 5. Показатели значимости для комбинации круглых значений всех 4 электоральных характеристик

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%

Субъект федерации \ Год	2000	2003	2004	2007	2008	2011	2012	2016	2018	Всего
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%

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. 3 демонстрирует, как менялась от выборов к выборам доля субъектов, затронутых фальсификациями (для их интерпретации в качестве величины N берется число субъектов, в которых проводились выборы в данном году). Рекорд здесь, вне всякого сомнения, устанавливают 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 данной 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.

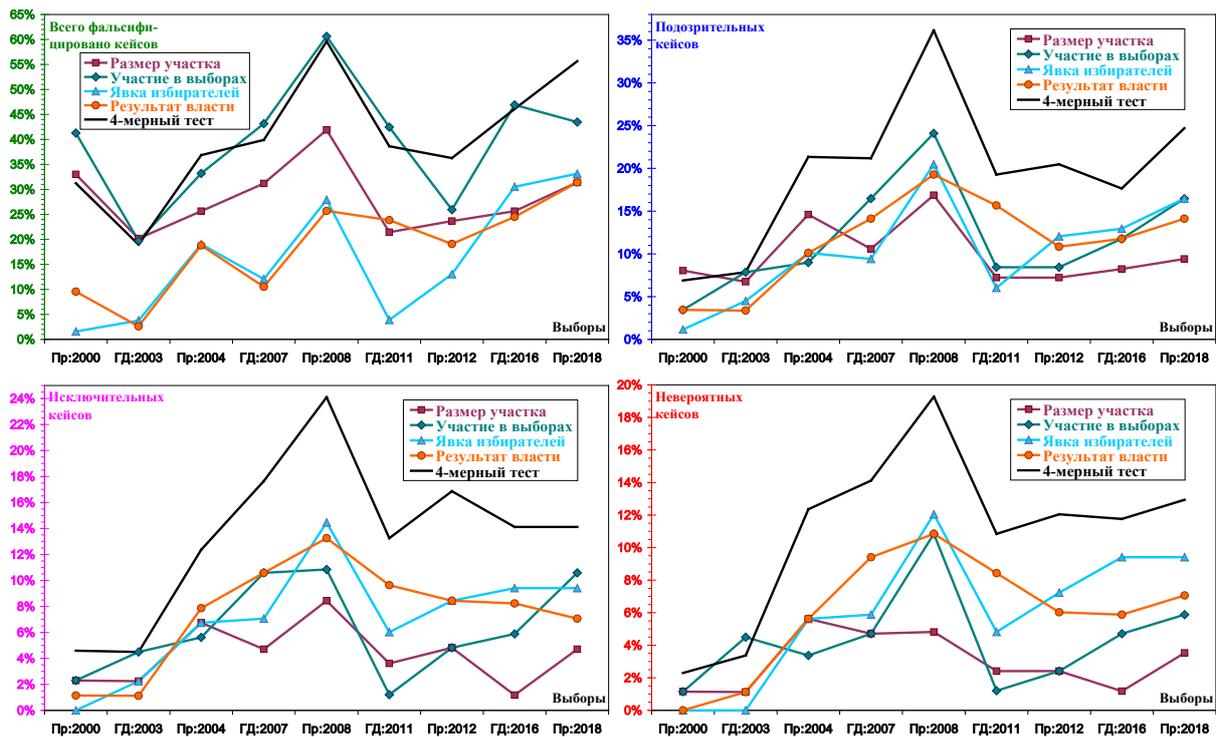


Рис. 3. Динамика долей интегральных и интерпретированных фальсификаций к психологически привлекательным числам

Итоги и обсуждение

Сводные результаты применения разных тестов, представленные в табл. 6 и 7, подводят исчерпывающий итог предложенным методам, давая представление о распространенности выдумывания результатов и позволяя сравнить различные тесты по способности к его выявлению. Проведенный анализ, ориентированный на весьма специфический тип фальсификаций, оптимален для общей оценки не столько достоверности результатов выборов, сколько дееспособности избирательной и судебной ветвей власти.

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. Причем все выявляемые фальсификации могут быть элементарно подтверждены 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.

Таблица 6. Количество и доля фальсифицированных избирательных кейсов

Тест	Всего		Подозри- тельные		Исклю- чительные		Неверо- ятные	
	Число	Доля	Число	Доля	Число	Доля	Число	Доля
Круглый размер участка (табл. 1)	217	28%	34	4,4%	23	3,0%	17	2,2%
Круглое участие в выборах (табл. 2)	304	39%	48	6,2%	34	4,4%	24	3,1%
Круглое промилле явки (табл. 3)	123	16%	55	7,2%	48	6,2%	38	4,9%
Круглое промилле результата власти (табл. 4)	141	18%	57	7,4%	47	6,1%	40	5,2%
4-мерный тест (табл. 5)	309	40%	104	14%	84	11%	79	10%

Таблица 7. Количество и доля субъектов федерации, для которых хотя бы раз были зафиксированы фальсификации

Тест	Подозри- тельные		Исклю- чительные		Неверо- ятные	
	Число	Доля	Число	Доля	Число	Доля
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. 3)	19	22%	15	18%	13	15%
Круглое промилле результата власти (табл. 4)	18	21%	16	19%	13	15%
4-мерный тест (табл. 5)	29	34%	23	27%	20	23%

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

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