
СТАТИСТИЧЕСКИЕ МЕТОДЫ И АНАЛИЗ ДАННЫХ

А.А. Гусаров, В.Б. Гухман
(Тверь)

К ОЦЕНКЕ СПОРНЫХ РЕЗУЛЬТАТОВ ВЫБОРОВ С ПОМОЩЬЮ СТАТИСТИЧЕСКОГО КРИТЕРИЯ СОГЛАСИЯ

В статье обосновывается использование статистических критериев согласия, в частности, критерия «хи-квадрат» для оценки легитимности результатов выборов. На конкретных примерах иллюстрируется логика принятия решений в спорных случаях.

Ключевые слова: выборы, равномерное распределение, статистическая гипотеза, критерий согласия, статистическая значимость, уровень надежности, уровень значимости.

Актуальность социально-политической проблемы легитимации демократически избранной власти существует всегда, особенно если выборы происходят по одной из мажоритарных систем с результирующим незначительным преимуществом одной из сторон – кандидатов во власть (лиц, партий). Именно несущественность разницы в количестве голосов, отданных за тех или иных кандидатов (доли единицы процентов от общего числа учтенных голосов), порождает сомнение. А не находится ли эта разница в пределах точности расчетов, которая, как всегда, оставляет желать

Андрей Александрович Гусаров – доцент Тверского государственного технического университета. E-mail: maedtver@yandex.ru.

Владимир Борисович Гухман – доктор философских наук, профессор Тверского государственного технического университета.

лучшего, или, что хуже, не сфальсифицированы ли результаты? Действительно, как только выборы с подобным исходом состоялись, проигравшие нередко апеллируют к высшим судебным инстанциям и/или к электорату, требуя «справедливого» пересмотра результатов, которые достаточно близки к *равномерному* (равновероятному) распределению голосов между кандидатами, тем более, если кандидатов всего два. При этом победители заявляют о безоговорочной победе, хотя о «безоговорочности», оказывается, можно спорить. Так, при системе «50%+1 голос» (прямые выборы или через выборщиков) в различного рода публикациях дискутировались (в указанном ключе) личностные дилеммы «Буш–Гор», «Буш–Керри», «Янукович–Ющенко» и др.¹ Дискуссии заканчивались лишь с вынесением юридического вердикта типа «таков закон». В определенной мере это свидетельствует о методических просчетах демократических избирательных технологий и соответствующих недостатках избирательного права, допускающего такие социально-политические коллизии.

Возникает вопрос, возможно ли предложить методику оценки результатов выборов, которая в некоторой мере компенсировала бы влияние объективных случайных факторов, связанных с (не)явкой избирателей, мелкими нарушениями и незначительными ошибками результирующих подсчетов? При положительном ответе на вопрос и принятии такой методики «на вооружение», потребуется, конечно, уточнение соответствующих положений избирательного законодательства, например, закона о ГАС РФ «Выборы» [1]. Известные нам публикации по проблемам выборов носят, в основном, публицистический характер (прогнозы, степень их реализации, возникающие социальные проблемы [2; 3; 4; 5]) и крайне редко посвящаются методическим вопросам.

¹ Подобная проблема характерна также для парламентской практики (например, при голосовании по законопроектам) и региональных выборов власти.

В этой связи нашей целью является внесение (в порядке дискуссии) конструктивного импульса в разработку такой методике, решение на основе математико-статистической теории проверки статистических гипотез, в частности, использования статистических критериев согласия для оценки результатов выборов. Впервые такая проблема была поставлена в статье В.Б. Гухмана [6]. В книге В.В. Платонова [7], близкой по проблематике, статистические критерии согласия используются для оценки так называемых *индексов представительности парламента* применительно к пропорциональной системе выборов в парламент по партийным спискам. Предлагаемая в статье (см. [6]) и детализируемая ниже методика ориентирована на другую систему голосования и на оценку *правдоподобия гипотез* о «победителях» и «проигравших» в случаях, сомнительных для избирательных комиссий.

Проверка статистических гипотез как механизм оценки результатов выборов

Рассмотрим приложение теории статистических гипотез к оценке результатов выборов в сомнительных случаях, полагая, что голоса избирателей представляют собой *выборку генеральной совокупности*, которая, в свою очередь, имела бы место, явись все избиратели на выборы¹.

Допустим, что два кандидата во власть набрали примерно одинаковое количество голосов с небольшой разницей в пользу одного из них, что дает повод считать его «победителем», даже если эта разница – результат стечения случайных обстоятельств. Оценим, достаточна ли она со статистической точки зрения для корректного принятия решения избирательной комиссией. Обра-

¹ При стопроцентной явке избирателей мы имеем дело с генеральной совокупностью и вопрос о выборке избирателей становится вообще беспредметным. Но проблема легитимности результатов выборов и в этом случае остается актуальной.

тимся к аппарату проверки статистических гипотез с целью определения статистической *незначимости* (статистическая гипотеза H_0) или *значимости* (альтернативная статистическая гипотеза H_1)¹ отклонения реального случайного (вероятностного) распределения голосов от теоретического равномерного распределения. Такая оценка требует применения *критериев согласия* теории с практикой.

Воспользуемся, например, популярным непараметрическим критерием согласия «хи-квадрат» К. Пирсона, одна из простейших дискретных форм которого имеет вид [8, с. 192; 9, с. 167]:

$$\chi^2 = n \sum_{i=1}^k \frac{(p_i^* - p_i)^2}{p_i}, \quad (1)$$

где n – количество наблюдений (общее число проголосовавших); k – число дискретных разрядов распределения (по числу кандидатов²); p_i^* и p_i – соответственно реальная (опытная) и теоретическая (математическая) вероятности i -го разряда распределения (здесь $p_i^* = m_i / n$); m_i – число голосов, отданных за i -го кандидата.

В (1) для теоретического равномерного распределения $p_1 = p_2 = \dots = p_k = 1/k$ [при $k = 2, p_1 = p_2 = 0,5$ (50%)]. При этом должно выполняться условие нормировки вероятностей:

$$\sum_{i=1}^k p_i^* = \sum_{i=1}^k p_i = 1 \text{ (100\%)}. \quad (2)$$

Из (1) следует, что при малых разностях $p_i^* - p_i$ величина χ^2 может оказаться меньше некоторого критического значения χ_{crit}^2 , которое при проверке статистических гипотез определяет, какую

¹ Понятие статистической (не)значимости равносильно понятию (не)существенности отклонения, которое на практике всегда имеет место (в большей или меньшей степени).

² Включая кандидата «против всех», если таковой «допускается» к выборам согласно действующему законодательству.

гипотезу (H_0 или H_1) принять в качестве правдоподобной: при $\chi^2 > \chi_{crit}^2$ отклонение реальных результатов выборов от теоретического равномерного распределения полагается статистически значимым, неслучайным (H_1); при $\chi^2 \leq \chi_{crit}^2$ – статистически незначимым, случайным (H_0). Критические значения χ_{crit}^2 обычно табулированы (см., например, [10, с. 334; 11, с. 499–502; 12]).

Из (1) также следует, что с увеличением объема выборки n (при тех же χ^2 и теоретическом распределении вероятностей) уменьшается допустимый разброс вероятностей в числителе под знаком суммы. Это значит, что для признания правдоподобия гипотезы H_0 (о равномерности эмпирического распределения) ошибка приближения должна уменьшаться, что естественно при росте числа наблюдений, т.е. активности избирателей.

Если представить $p_i^* = m_i / n$, то (1) преобразуется к виду, часто используемому в практике статистических расчетов [10, с. 241; 11, с. 308]:

$$\chi^2 = \sum_{i=1}^k \frac{(m_i - np_i)^2}{np_i}. \quad (3)$$

Полагаем, что избирательным комиссиям не составит слишком большого труда воспользоваться одной из дискретных формул критерия «хи-квадрат», представив проценты голосов, обычно используемые на практике, в безразмерной вероятностной форме. Что касается значений χ_{crit}^2 , то для их вычисления нет острой необходимости обращаться к таблицам – достаточно воспользоваться возможностями стандартных компьютерных программ, например MS Excel или SPSS.

Заметим, что критерий согласия χ^2 лишь позволяет оценить, насколько правдоподобна (или неправдоподобна) гипотеза о равномерности распределения голосов между кандидатами – и не более того. С помощью критерия χ^2 в общем случае нельзя определить «победителей» и «побежденных». Это можно сделать только в частном случае, когда кандидатов всего два ($k = 2$).

Поэтому данная методика предлагается не взамен, а в дополнение к существующим методам. Она представляется полезной, если результаты выборов сомнительны, в частности, на завершающем этапе, когда приходится выбирать между двумя оставшимися кандидатами.

Известны и другие критерии согласия (непараметрические критерии Колмогорова, омега-квадрат, параметрические критерии [11, с. 303–308, 317–320]), но они представляются нам более трудоемкими в реализации, чем критерий «хи-квадрат».

Предлагаемая методика, конечно, может «удлинить» избирательную кампанию (с соответствующими дополнительными затратами), но с не меньшими проблемами (рисками) приходится сталкиваться при существующих методах принятия решения. Кроме этого, риск ошибки случайного выбора, а вот такого риска как раз и можно избежать (с контролируемой достаточно высокой доверительной вероятностью¹), используя предлагаемую методику.

Приложения критерия согласия «хи-квадрат» к оценке результатов выборов

Пример 1. Региональные выборы губернатора в РФ

В табл. 1 и на рис. 1 в отсортированном виде представлены опубликованные в Интернете итоги первого тура выборов губернатора одной из областей РФ [13]. Выборы состоялись в декабре 1999 г. Число кандидатов 11 и «кандидат» *Против всех* – итого 12 кандидатов. Система голосования мажоритарная, двухступенчатая; во второй тур выходят два кандидата, набравшие наибольшее абсолютное (и относительное) число голосов в первом туре.

¹ При доверительной вероятности γ (уровень надежности оценки) риск принять неверную гипотезу из H_0 и H_1 равен $\alpha = 1 - \gamma$ (уровень значимости ошибки). Так, при $\gamma = 0,95$ (95%) $\alpha = 0,05$ (5%), при $\gamma = 0,99$ (99%) $\alpha = 0,01$ (1%). Значение α (γ) задается пользователем при определении χ^2_{crit} (по математическим таблицам либо с помощью компьютерных средств).

Таблица 1

РАСПРЕДЕЛЕНИЕ ГОЛОСОВ В 1-м ТУРЕ ВЫБОРОВ ГУБЕРНАТОРА
(ПРИМЕР)

№	Кандидаты	Число голосов	Процент голосов
1	К1	257483	32,38
2	К2	181413	22,81
3	К3	98911	12,44
4	К4	47115	5,93
5	К5	43494	5,47
6	К6	29490	3,71
7	К7	29350	3,69
8	К8	21882	2,75
9	К9	12331	1,55
10	К10	10166	1,28
11	К11	2461	0,31
12	Против всех кандидатов	48930	6,15
13	Количество недействительных бюллетеней	12142	1,53
	Всего (из них учтенных голосов – без недействительных бюллетеней)	$n = 795168$ (783026)	100 (контроль)



Кандидаты
Рис. 1. Распределение голосов по результатам первого тура выборов (пример 1)

Даже без расчетов видно, что распределение голосов существенно неравномерное. Расчеты по приведенной выше методике подтверждают этот вывод:

$\chi^2 = 1135881,3$ (расчет произведен при $k = 13$, так как учитывается процент недействительных бюллетеней);

$\chi^2_{crit} = 21,03$ (расчет проведен в MS Excel (статистическая функция ХИ2ОБР) при уровне надежности (доверительной вероятности) 95% (0,95) и числе степеней свободы $k - 1 = 12$):

$$\chi^2 \gg \chi^2_{crit}. \quad (4)$$

Если не учитывать процент недействительных бюллетеней, то при сумме голосов $n = 783026$ и $k = 12$ результаты расчетов будут следующими.

Таблица 2
 РАСПРЕДЕЛЕНИЕ ГОЛОСОВ В 1-м ТУРЕ ВЫБОРОВ
 (БЕЗ НЕДЕЙСТВИТЕЛЬНЫХ БЮЛЛЕТЕНЕЙ)

Кандидаты	Процент голосов
К1	32,88

Окончание табл. 2

Кандидаты	Процент голосов
К2	23,17
К3	12,63
К4	6,02
К5	5,55
К6	3,77
К7	3,75
К8	2,79
К9	1,57
К10	1,30
К11	0,31
<i>Против всех</i>	6,25
Контрольная сумма	100

$$\chi^2 = 1024862,1; \chi_{crit}^2 = 19,675; \chi^2 \gg \chi_{crit}^2$$

Таким образом, усиленное неравенство (4) сохраняется – распределение голосов существенно неравномерное.

Согласно действовавшему (в тот период) законодательству во второй тур выборов прошли кандидаты К1 и К2. В табл. 3 и на рис. 2 приведены итоги второго тура выборов, включая «кандидата» *Против всех*.

Таблица 3

РАСПРЕДЕЛЕНИЕ ГОЛОСОВ ВО 2-м ТУРЕ ВЫБОРОВ ГУБЕРНАТОРА

№	Кандидаты	Число голосов	Процент голосов
1	К1	299 869	46,48
2	К2	296 458	45,95

Окончание табл. 3

№	Кандидаты	Число голосов	Процент голосов
3	Против всех кандидатов	42 242	6,55
4	Количество недействительных бюллетеней	6600	1,02
	Контрольная сумма	645 169	100
	Общее число проголосовавших (из них учтенных голосов)	$n = 645169$ (638569)	

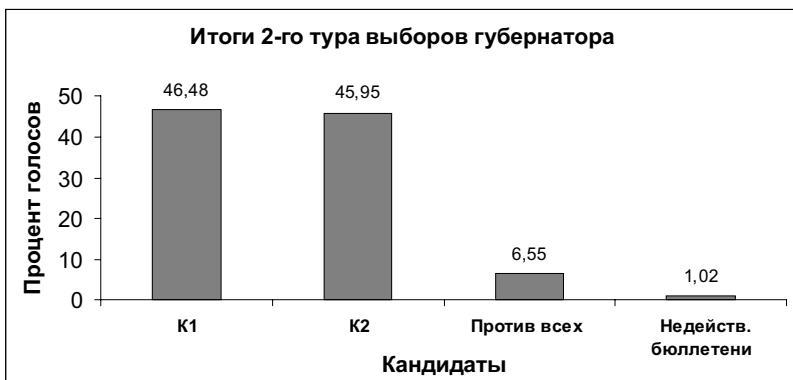


Рис. 2. Распределение голосов по результатам второго тура выборов

Распределение голосов опять неравномерное (в основном, из-за голосов, отданных за «кандидата» *Против всех*, и недействительных бюллетеней), на что недвусмысленно указывают и результаты расчетов по критерию «хи-квадрат» (с учетом недействительных бюллетеней):

$$\chi^2 = 468565,3; \chi_{crit}^2 = 7,815; \chi^2 \gg \chi_{crit}^2.$$

По простому большинству голосов кандидат K1 был признан победителем во втором туре выборов и стал легитимным губернатором области.

Если исключить недействительные бюллетени и голоса, отданные против всех, то при соответствующем перерасчете процента голосов, отданных за двух кандидатов ($k = 2$), данные второго тура выборов приобретают вид (табл. 4).

Таблица 4

РАСПРЕДЕЛЕНИЕ ГОЛОСОВ ВО 2-м ТУРЕ ВЫБОРОВ
ГУБЕРНАТОРА (ДВА КАНДИДАТА)

Кандидаты	Число голосов	Процент голосов
K1	299869	50,29
K2	296458	49,71
Сумма голосов	596327	Контроль: 100

$$\chi^2 = 19,51; \chi_{crit}^2 = 3,84; \chi^2 > \chi_{crit}^2$$

Обратим внимание, что неравенство (4) уже неусиленное, но все же имеет место. Следовательно, правдоподобна гипотеза $H1$, т.е. распределение голосов между двумя кандидатами статистически значимо отличается от равномерного распределения, и *губернатор, действительно, легитимен* (с уровнем надежности 95%, т.е. мы можем ошибаться с уровнем значимости 5%).

При уровне надежности наших оценок 99% $\chi_{crit}^2 = 6,635$ и, как и прежде, $\chi^2 > \chi_{crit}^2$.

На весьма однозначных результатах последних расчетов (в пользу легитимности избранной власти) сказала прежде всего активность избирателей – относительно большое значение n в (1), (3). Поэтому нам представляется, что для снижения остроты обсуждаемой проблемы желательно, чтобы как можно больше избирателей участвовало в голосовании. Более того, явку на выборы каждый избиратель, если он желает достойной власти, должен считать не только своим правом, но и гражданским долгом¹.

¹ Во многих странах (Бельгия, Пакистан и др.) неявка избирателя на выборы считается противоправным деянием с соответствующими последствиями.

Из проведенного анализа также следует, что корректное использование предлагаемой методики в сомнительных случаях, когда легитимность избранной власти под вопросом, возможно, если избирательным комиссиям исключить из рассмотрения недействительные бюллетени и голоса, отданные за «кандидата» *Против всех*. При двухальтернативном выборе целесообразно убрать графу *Против всех* из бюллетеней.

Пример 2. Президентские выборы в США (2000-й, 2004 гг.)¹

Выборы двухступенчатые: 1) прямое голосование граждан США, пользующихся избирательным правом; 2) избрание президента выборщиками от штатов с учетом результатов прямого голосования. Каждый из 538 выборщиков голосует за одного из двух кандидатов, как правило, набравших наибольшее число голосов в штате выборщика². Избранным на пост президента считается кандидат, набравший простое большинство голосов выборщиков.

Распределение голосов выборщиков на выборах 2000-го и 2004 гг. между двумя наиболее популярными кандидатами приведено в табл. 5.

Таблица 5

РАСПРЕДЕЛЕНИЕ ГОЛОСОВ ВЫБОРЩИКОВ

Выборы 2000 г.			Выборы 2004 г.	
Дж. Буш	А. Гор	Воздержался	Дж. Буш	Д. Керри
271	266	1	286	252

Примечание 1. В статистике 2000 г. учтем 537 выборщиков (без воздержавшегося).

Примечание 2. В отличие от примера 1 здесь имеет место не выборка, а генеральная совокупность избирателей, ибо *все без исключения* выборщики участвуют в избрании президента.

¹ По материалам [14].

² Имеются в виду 50 штатов и один Федеральный округ. При этом густонаселенные штаты представлены большим числом выборщиков, чем малонаселенные штаты.

Расчеты по предлагаемой методике дали следующие результаты:
Выборы 2000 г.: $\chi^2 = 0,046555$; выборы 2004 г.: $\chi^2 = 2,1487$.

При $\chi_{crit}^2 = 3,84$ (с уровнем надежности 95%) и $\chi_{crit}^2 = 6,635$ (99%) $\chi^2 < \chi_{crit}^2$.

Следовательно, если воспользоваться критерием согласия «хи-квадрат», то правдоподобна гипотеза H_0 о близости распределения голосов выборщиков к равномерному в обеих избирательных кампаниях. И это вполне объяснимо, так как волеизъявление граждан не отдало *явного предпочтения* одному из кандидатов. Так, в кампании 2000 г. голоса граждан США распределились даже обратно распределению голосов выборщиков (с небольшим преимуществом А. Гора): Буш – 49819600, Гор – 50156783 [14].

Причины подобного противоречия, видимо, хорошо известны в США, если судить по кричащим заголовкам изданий [3; 4; 5]. Но, как бы там ни было, таков закон!

В качестве одной из существенных причин мы можем предположить неадекватность представительства института выборщиков. И здесь, возможно, уместно было бы обратиться к методикам расчета индексов представительности, описанным в [5], что выходит за рамки настоящей статьи.

ЛИТЕРАТУРА

1. Федеральный закон о государственной автоматизированной системе РФ «Выборы». 2003. № 20-ФЗ.
2. Паниотто В.И. Реабилитация социологов на парламентских выборах // <http://www.pravda.com.ua/news/2006/4/7/40732.htm>.
3. 36 Days: The Complete Chronicle of the 2000 Presidential Election Crisis. N.Y.: Times Books, 2000.
4. Gillman H. The Votes that Counted: How the Court Decided the 2000 Presidential Election. Chicago: University of Chicago Press, 2001.
5. Miller M.C. Fooled Again: How the Right Stole the 2004 Election. N.Y.: Basic Books, 2005.
6. Гухман В.Б. Избирательные технологии – наука или шарлатанство? // Компьютерра. 2003. № 46 (521). С. 53–55.

7. *Платонов В.В.* Применение критериев согласия для оценки представительности парламента: Препринт WP 7/2004/03. М.: ГУ–ВШЭ, 2004.

8. *Айвазян С.А., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика (основы моделирования и первичная обработка данных). М., 1983.

9. *Гухман В.Б.* Введение в компьютерную обработку социологических данных. Тверь, 2004.

10. *Калинина В.Н., Панкин В.Ф.* Математическая статистика. М., 2001.

11. *Тюрин Ю.Н., Макаров А.А.* Анализ данных на компьютере. 3-е изд. М., 2003.

12. *Большев Л.Н., Смирнов Н.В.* Таблицы математической статистики. М., 1983.

13. Выборы, которые проходили в Тверской области за период деятельности Избирательной комиссии Тверской области // http://tverik.cikrf.ru/inf_d/elect.htm.

14. United States Presidential Election, 2000 // http://en.wikipedia.org/United_States_presidential_election%2C_2000; United States Presidential Election, 2004 // http://en.wikipedia.org/United_States_presidential_election%2C_2004.