Техническое задание

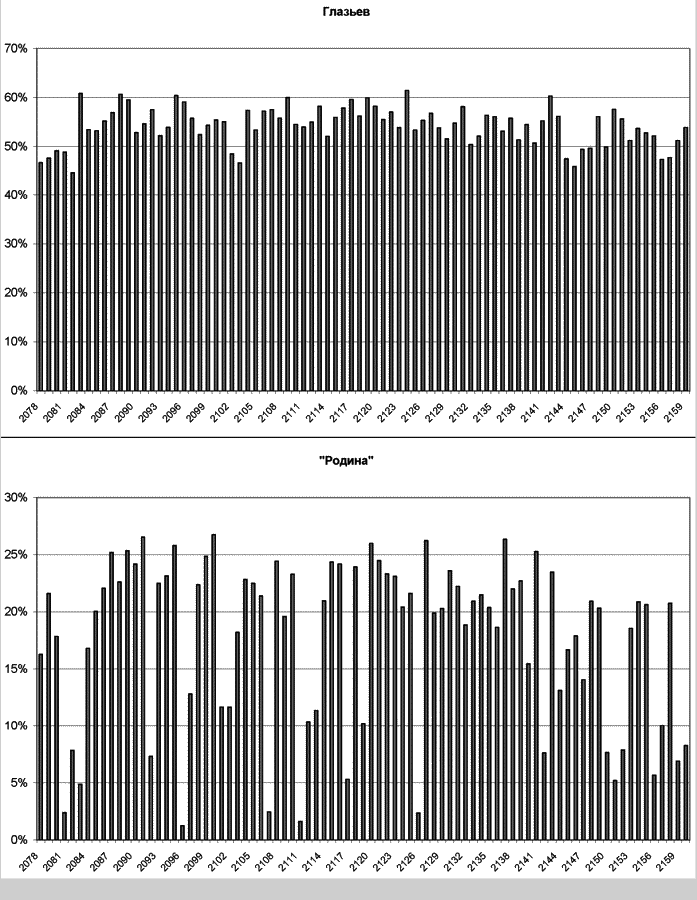
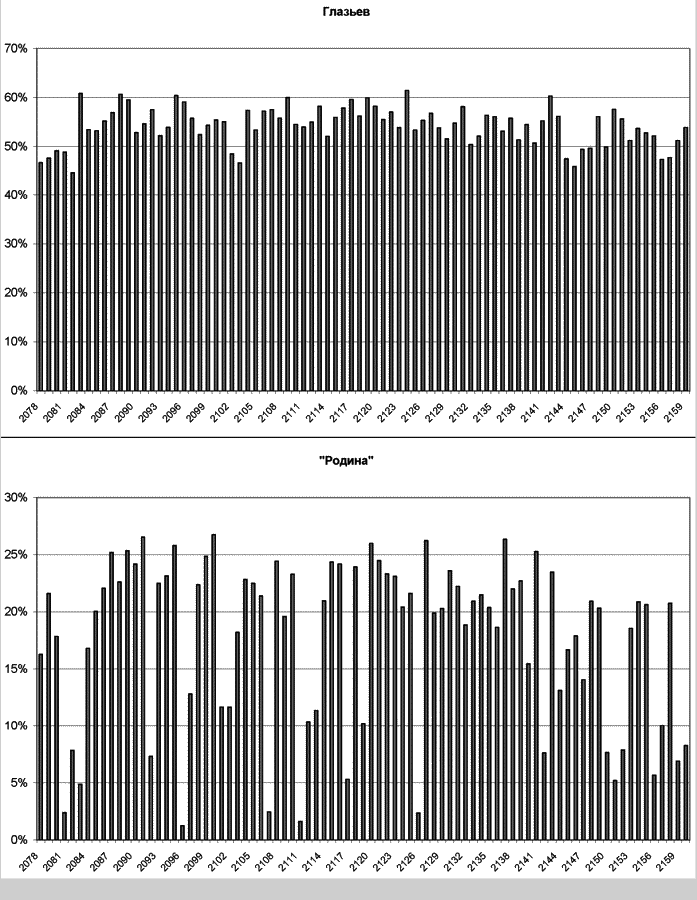
«Модель для определения вбросов на участках по статистическим данным других участков»

# Методы выявления фальсификаций

## Дисперсионный анализ

Одним из наиболее простых и понятных методов выявления фальсификаций является сравнительный (дисперсионный анализ) – сравнение итогов голосования на близких по составу электората избирательных участках или территориях. Если электорат достаточно однородный, то при честном проведении голосования и подсчета голосов разброс не должен быть велик. Фальсификации же обычно осуществляются не по всем избирательным участкам, и это приводит к заметному увеличению разброса. Так на рисунке 1 из первой диограммы мы можем сделать вывод о достаточной однородности городского электората, и в сравнении с ней вторая гистограмма ясно указывает на фальсификацию итогов голосования.

Рисунок 1



## Метод корреляции с явкой Собянина – Суховольского

Метод основан на предположении, что в отсутствии фальсификаций результаты кандидатов или партий (в процентах от числа проголосовавших избирателей или действительных голосов) не зависят от активности избирателей. Если же результаты кандидатов или партий на отдельных территориях (избирательный участок, район и т. п.) выражать в процентах от списочного числа избирателей, то зависимость таких результатов от показателя явки должна выражаться прямой, исходящей из начала координат, тангенс угла наклона которой равен среднему проценту голосов от числа проголосовавших избирателей.

Такие «нормальные» графики действительно часто наблюдаются на выборах (см. левый график на рисунке 2). Однако встречаются и графики, которые с точки зрения данной гипотезы следует считать «аномальными»: в них точки, характеризующие итоги голосования на отдельных территориях, располагаются вдоль регрессионной прямой, пересекающей ось ординат заметно выше или ниже начала координат (см. правый график на рисунке 2).

Изображение выглядит как линия, диаграмма, График, Параллельный

Автоматически созданное описаниеТангенс угла наклона кривой на графиках Собянина – Суховольского назвается «поддержкой дополнительных избирателей» (ПДИ) и в качестве критерия отклонения от «нормы» может использоваться разность между ПДИ и результатом кандидата (партии) относительно числа проголосовавших избирателей.

Рисунок 2

Второй параметр регрессионного уравнения (отрезок, отсекаемый регрессионной прямой на оси ординат) назвали смещением регрессионной линии (СРЛ). В идеальном случае значение СРЛ должно быть нулевым (то есть регрессионная линия должна попадать в начало координат), «нормальными» можно считать итоги голосования, при которых значение СРЛ по модулю не превышает 0,1 (или 10 %). Превышение этого порога свидетельствует об аномалиях.

При способе Собянина – Суховольского в «нормальном» случае коэффициент корреляции явки и результатов будет близок к единице.

## Метод исследования распределения явки

Метод заключается в построении графика, где по оси абсцисс откладывается явка, а по оси ординат – число избирательных участков, на которых явка укладывается в соответствующий однопроцентный интервал. Метод может быть использован для территорий, содержащих значительное число участков (две-три тысячи и более), для меньших территорий приходится использовать двух– или даже пятипроцентные интервалы, но точность и наглядность в этом случае значительно ниже.

Изображение выглядит как текст, снимок экрана, черно-белый, монохромный

Автоматически созданное описаниеПо мнению исследователей, использующих данный метод, на относительно однородных территориях при отсутствии административного воздействия получающиеся кривые распределения должны быть близки к гауссовым, то есть иметь один «горб» и быть достаточно симметричными.

Рисунок 3

На рисунке 3 показаны кривые распределения явки с 2000 по 2014 год. Кривая для выборов 2000 года почти идеальна, и на выборах 2012 года получилась точно такая же кривая, лишь смещенная в сторону более низких значений явки. Также близка к идеальной кривая для выборов 2014 года. Кривые для выборов 2005 года и 2007 года также имеют один основной «горб», но на них уже видны искажения. А вот кривые для выборов 2009 года и 2011 года вообще не имеют выраженного основного пика; они «размазаны» в широком диапазоне значений явки, но при этом имеют небольшие локальные пики – в основном на «круглых» значениях явки (40, 50, 55, 60, 65, 75 %). Очевидно, поведение электората не может так резко меняться в течение короткого времени, а активность однородного городского электората не может описываться пилообразными кривыми распределения.

# Функциональное назначение

Разрабатываемая «Модель для определения вбросов на участках по статистическим данным других участков» является инструментом для выявления вбросов по вышеуказанным методам. Предоставляя необходимые графики и статистики, она позволяет исследователям оценить вероятность фальсификации результатов путем вбросов.

# Требования к функциональным характеристикам

## Требования к составу выполняемых функций

Программа должна предоставлять пользователю возможность:

* Рассчитывать дисперсию голосов за кандидата X по избирательным участкам (дисперсионный анализ);
* Строить столбчатую диаграмму количества голосов за кандидата X по избирательным участкам (дисперсионный анализ, рисунок 1);
* Рассчитывать ПДИ (метод корреляции с явкой Собянина – Суховольского);
* Рассчитывать СРЛ (метод корреляции с явкой Собянина – Суховольского);
* Рассчитывать корреляцию явки и результата кандидата, выраженного в процентах в процентах от общего числа избирателей. (метод корреляции с явкой Собянина – Суховольского);
* Строить точечную диаграмму явки и результата кандидата, выраженного в процентах в процентах от общего числа избирателей (метод корреляции с явкой Собянина – Суховольского, рисунок 2);
* Строить гистограмму, где по оси абсцисс откладывается явка, а по оси ординат – число избирательных участков, на которых явка укладывается в соответствующий однопроцентный интервал (метод исследования распределения явки, рисунок 3);

## Организация входных данных

Данные по избирательным участкам предоставляются в файле формата .csv , где

* Первая строка – названия столбцов, причем третье и последующие имена столбцов задают имена кандидатов
* Все следующие строки содержат следующую информацию об участке:
  + Общее количество избирателей, прикрепленных к этому участку
  + Количество проголосовавших избирателей (для упрощения считаем, что количество проголосовавших = количество действительных бюллетеней)
  + Количество голосов отданных за кандидата , где

Входными данными также является взаимодействие пользователя с графическим интерфейсом (GUI) программы.

## Организация выходных данных

Выходные данные будут отображаться при помощи графического пользовательского интерфейса (GUI).

# Терминология

**Явка –** отношение (выраженное, как правило, в процентах) числа избирателей, принявших участие в выборах, к общему количеству граждан, имеющих право на участие в этих выборах.

**Вброс** – способ фальсификации, при котором число голосов за одного кандидата (одну партию) увеличивается, а за других остается неизменным; в этом случае на число «вброшенных голосов» увеличивается показатель явки, то есть результат кандидата (партии) увеличивается формально за счет снижения числа не проголосовавших избирателей.

# Использованные источники

1. Избирательные системы: российский и мировой опыт [Текст] / А. Е. Любарев ; Региональная общественная орг. "Либеральная миссия". - Москва : Либеральная миссия : Новое Литературное Обозрение, 2016. - 631 с. : ил., табл.; 21 см.; ISBN 978-5-4448-0569-5 c. 543 - 564
2. Явка избирателей – Википедия – URL: <https://ru.wikipedia.org/wiki/Явка_избирателей> (режим доступа: свободный, дата обращения 11.04.2024)