High Turnout Equals Democracy? A Statistical Analysis of Periodical Turkey Voting Data.

Robert Ke, Keming Zhang, Yuka Shimazaki

September 2, 2024

1 Background

Democratic societies are founded on the concept of conducting elections that are both free and fair, ensuring that every citizen's vote has an equal value. Government interventions that undermine or deprive free and fair voting behaviors should be questioned, examined, and censured if necessary.

Turkey is known for its high turnout rate and a great frequency of political voting events. However, Turkey has also remained controversial in terms of being a free country. "Turkey could be on the brink of dictatorship," posted by the *Economist*. *Freedom House* also questioned Turkish freedom under the rule of President Recep Tayyip Erdoğan and his Justice and Development Party (AKP). The press gave Turkey a low 32 overall democracy index, and lamented with, "deepening economic crisis ... government new incentives to suppress dissent and limit public discourse."

Therefore, it could be doubtful that the incumbent government intervened in the allegedly "democratic" elections. The aim of this paper is to analyze whether the voting results form any statistical irregularities.

2 Literatures and Methodolgies

Recent statisticians shed light on how to detect fraud with ballot data.

Benford's law underwent a revival as a tool for potentially detecting election fraud (1). In its basic and simplest form, Benford's Law notes that in numerous real-life processes, the first significant digit's logarithm tends to be uniformly distributed. Any discrepancies from this pattern could suggest the presence of alternate, potentially dishonest government interventions.

A different strategy for detecting fingerprints of election fraud is to examine the distribution of votes and turnout numbers. This paper refers to the method provided by this paper (2) as the "**turnout-vote nexus**". Authors also observed that ballot stuffing induces a high correlation between turnout and voting results – unusually high vote counts tend to co-occur with unusually high turnout numbers.

The third approach, referred to as "**suspicious number**", is based on the finding that in certain suspected fraudulent elections, a significant number of polling stations report a winning party's vote percentage that exactly divides by five – for instance, 50%, 65%, 75% (3)(4). Common sense suggests that in large voter populations, the chances of a party consistently securing such rounded-off vote percentages across numerous precincts are extremely slim without some kind of vote tampering.

It demonstrates that even perfectly legitimate election data can display seemingly anomalous patterns that might be wrongly interpreted as fraud, unless the process generating the data is thoroughly understood. Considering the increasing agreement within empirical studies that claims of election fraud frequently lead to violence after the election (5)(6), it's crucial for methods of detecting fraud to place a significant emphasis on minimizing false positive errors. Thus, to minimize the false positive errors, the turkey voting data has to undergo all three proposed methods with a strict statistical threshold, and the conclusion will be relaxed.

3 Data and Analytics

The data panel is integrated by Prof. Cantay Caliskan, a computational social science working at the Georgen Institute of Data Science, University of Rochester to understand the Turkey elections voting situation. The data panel records several noteworthy voting matrics across important Turkey elections from 2014 to 2023, consisting of 11 elections, namely 2014 prior-referendum presidential election (2014c), 2014 local election, 2015 June general election (J), 2015 Nov general election(N), 2018 General election (presidential part) (2018c), 2018 General election (Assembly part)(2018p), 2019 local elections, 2023 General election (Assembly part)(2023p), 2023 General election (first-round presidential part) (2023c1), 2023 General election (2nd-round presidential part) (2023c2), and finally the most significant 2017 Constitutional Referendum. The matrics or variables specifically are the number of eligible voters, the number of vote, the number of valid and invalid votes, and detailed vote counts for various parties, including AKP and other parties like SAADET, HDP, CHP, MHP, independent, and so on, so the number of variables inside the data is approximately 121. The data contains 973 rows representing district-level observations for each election and its

Columns	Description	Type	Possible Value
Province	All province in Turkey	String	name of Turkey province
District	All district in Turkey	String	name of Turkey district
Eligible_2014_c	The eligible voter in 2014 prior-referendum presidential election (2014c)	Numeric	limit to the total number of people in each district
Voter_2014_c	The actual voter in 2014 prior-referendum presidential election (2014c)	Numeric	less than the number of eligible voters
Valid_2014_c	The valid votes in 2014 prior-referendum presidential election (2014c)	Numeric	less than the number of actual voters
Invalid_2014_c	The invalid votes in 2014 prior-referendum presidential election (2014c)	Numeric	=total vote-valid vote
AKP_2014_c	The votes for AKP in 2014 prior-referendum presidential election (2014c)	Numeric	votes for AKP
SAADET_2014_c	The votes for SAADET in 2014 prior-referendum presidential election (2014c)	Numeric	votes for SAADET
HDP_2014_c	The votes for HDP in 2014 prior-referendum presidential election (2014c)	Numeric	votes for HDP
CHP_2014_c	The votes for CHP in 2014 prior-referendum presidential election (2014c)	Numeric	votes for CHP
MHP_2014_c	The votes for MHP in 2014 prior-referendum presidential election (2014c)	Numeric	votes for MHP
Independent_2014_c	The votes for independent in 2014 prior-referendum presidential election (2014c)	Numeric	votes for independent
Other_2014_c	The votes for other parties in 2014 prior-referendum presidential election (2014c)	Numeric	votes for Others

Figure 1: data dictionary

metrics.

Here is a sample data dictionary showing the main variables used in analysis for one of the Turkey elections, 2014 prior-referendum presidential election (2014c). The data dictionary for the other elections will follow a similar description, data type, and possible value.

The analytical methods we will use on the dataset is listed below.

1. Benford Law examination: The frequent distribution of the first digit in the voting dataset will be analyzed. According to the Benford Law, the explanatory variable here is all possible first digits from 1 to 9, and the response variable here is the amount of each first digit. A logarithmic function will be fitted onto the data. The null hypothesis will be the data is "realistic" and forms a log-decrease function in terms of first digits. [Rejecting the null hypothesis under great confidence indicates the election results may be spurious.] The same process will be applied to each voting district.

2. Turnout-vote nexus examination: The distribution of turnout rate and voting result (vote for AKP or Erdogan) across districts will be side-by-side compared. All the elections included in the dataset will be analyzed. The correlation coefficient will be calculated. The unusually high correlation will be examined with linear regression, calculating the beta's robustness using student's t-test. The student t-test is testing the null hypothesis that the coefficient of the OLS is not zero. A spurious result, as pointed out in (1), will have a robust and over 0.5 beta. The rejection of the null may indicate a spurious election.

3. Suspicious number examination: A descriptive analysis of the votes will be performed. This paper pinpoints on whether there is a binomiality and uneven distribution across districts in terms of voting results. The percentage numbers that have multiples of 5 will be marked.

References

[1] Mebane, W.R. (2006). Election Forensics: Vote Counts and Benford's Law. Political Methodology Society, University of California, Davis, CA.

[2] Mebane, W.R., & Sekhon, J.S. (2004). Robust estimation and outlier detection for overdispersed multinomial models of count data. *American Journal of Political Science*, 48, 392–411.

[3] Klimek, P., Yegorov, Y., Hanel, R., & Thurner, S. (2012). Statistical detection of systematic election irregularities. *Proceedings of the National Academy of Sciences*, 109(41), 16469–16473. https://doi.org/10.1073/pnas.1210722109

[4] Gehlbach, S. (2012, January 27). Electoral Fraud in Russia: Report from the Russian Blogosphere. *The Monkey Cage*.

[5] Tucker, J. (2007). Enough! Electoral fraud, collective action problems, and post-communist Colored Revolutions. *Perspectives on Politics*, 5(3), 535–551.

[6] Collier, P. (2009). *Wars, Guns, and Votes: Democracy in Dangerous Places*. New York: Harper.