



Assignment 2: On Being a Data Skeptic

Student ID: 261032796

Name: Ahmed Ibrahim

Email: ahmed.ibrahim6@mail.mcgill.ca

On December 7, 2020, a bill of complaint was made against Georgia, Michigan, Wisconsin, and Pennsylvania to challenge their administration of the 2020 US presidential election. This document was selected as the basis of the herein short study. In this document's nature of action section, two actions are relevant to this study being actions 10 and 11. Action 10 stipulates that Biden's probability of winning the popular vote in the four defendant states independently, given Trump's lead in the polls as of 3 am on November 4, 2020, is less than 10^{-13} %. The odds of winning all 4 states were tabulated to be less than 10^{-52} % [1]. The last relevant action compared Biden's performance and Hillary Clinton's performance in the 2016 general election regarding those same four states; they concluded both democratic leaders had the same degree of the unlikelihood of winning the mentioned states. The appeal for this case was led by Ken Paxton, the Texas Attorney General. In contrast, the statistical analysis to support these claims was made by Charles J. Cicchetti, an American economist, a deputy director of the Energy and Environmental Policy Center at Harvard University's John F. Kennedy School of Government.

Professor Cicchetti's assignment was to analyze the validity and credibility of the 2020 presidential election in battleground states. His analysis can be segregated into three main parts. Firstly, he investigates the differences in county votes for both Biden and Clinton. He also compared and tested the significance of the total ballots' change tabulated in early reporting compared to the subsequent tabulations. Secondly, he also compared the rejection ballots from Georgia for both 2016 and 2020. Thirdly, he investigated the validity of the absentee ballots for Wayne County, Michigan, and how they met the tabulating and reporting ballots' requirements.

In the first part, he tested the hypothesis that both Clinton and Biden's performances were statistically similar. He determined the Z-score by comparing the total number of votes received by both democratic candidates. He estimated the variance necessary to compute the Z-score by computing the product of the mean times the probability of the candidates not getting a vote (see column 1 of table 2 in appendix). He also computed a second set of Z-scores with an additional test to remove the effects of Biden's increased number of votes in retrospect to Clinton's (see column 2 of table 2 in appendix). Given the astronomically large Z-scores obtained, he rejected the hypothesis in both scenarios that both candidates' performance was similar. As for the second half of the first part, he investigated the numbers behind Biden's turnaround in the four states after the late ballots' addition. He computed a Z-score for each state to test if the early ballots' votes and ballots, including the late ballots, were statistically similar (see column 3 of table 2 in appendix). Yet again, the Z-scores computed were incredibly large, where the smallest value was from Michigan with 586 despite the early ballots accounted for at the minimum of 90 % of the final ballots. Thus, with a greater degree of confidence in this tabulation, he rejects the hypothesis that the early morning vs. after election tabulation were drawn from the same voter population.



In the second part, the mail-in absentee ballots in Georgia were investigated, given the upset occurred with a difference of only 12,670 votes. In 2016, the mail-in volume was 213,033, with a rejection rate of 6.42 %. However, in 2020, the volume skyrocketed to 1,316,943 with a drastically lower rejection rate of 0.36 % (see Tables 3 and 4 of the appendix). In this section, he argues that if the same rejection rate from 2016 occurred in 2020, this would have yielded a decrease of votes by 28,965 for Trump and 54,552 for Biden. This adjustment would yield a win by 12,917 for Trump.

In the last part, he analyzed absentee ballot data for Wayne County, Michigan, solely at the precinct level. He found that 174,384 out of these 566,694 were counted without a registration number for precincts in Detroit. It is worth noting that Biden won Wayne county with 68.4% of tabulated ballots. By assuming the ballots from the unregistered votes from Wayne County and the registered counterparts had a similar distribution of republican/democratic votes, and then by removing these unregistered ballots, Biden would lose 119,300 votes, and Trump would lose 52800 votes. This hypothetical scenario is not enough to swing a win in Michigan for Trump. Still, this notion was used to argue that Michigan's overall election administration was inaccurate and warranted a thorough audit [2].

As statistical data analysis can be a powerful tool, its utility depends on the assumptions underpinning them [3]. In the first part, extreme Z-scores were obtained when comparing the total votes obtained by Biden and Clinton, which lead to the odds used by Paxton throughout the appeal. However, the analyst made flawed and minimal assumptions; for instance: he did not account for the fact that the two democratic candidates are inherently different individuals. For some voters, Hillary reminded them of Bill Clinton's 8 years in office. Thus, in some ways, Biden was considered more acceptable than Clinton in ways large and small, personal, and political, sexist, and not [4]. This section's tests also did not account for demographic and geographical differences in Biden and Clinton voters. It is worth noting that Biden bested Clinton's performance in every age-related voting in 2020, only allowing Trump to gain ground among the population between the ages of 30 to 44 years of age. From a geographical standpoint, Biden emerged victorious in the suburbs with 5 % more points than Clinton, where she fell to Trump in this category. As for small cities and rural communities, Biden gained 8 % more points, while Trump dropped from 61 % to 57 % compared to 2016 [5]. Another factor that could provide insight into the massive gap in voting between Biden and Clinton is that the 2020 election had the highest voter turnout seen in 120 years with a predicted turnout of 66.9 % [6]. The analyst attempted to correct the large gap of votes between the two candidates but did not provide details about his assumptions. There was no mention of the pandemic's effects on the voting turnout and distribution in these battleground states. The pandemic had colossal socio-economical effects on the country, yet this was not accounted for in any of the analyst's assumptions [7]. Many more factors could be enumerated as worthy candidates to illustrate how Biden and Clinton's campaigns lead to different outcomes.

Furthermore, one can wonder what the point of testing the null hypothesis for these two democratic candidates to have similar outcomes since it would be expected that the Z-scores would be at extremes given the many possible sources of distinctions. Besides the assumptions, the choice of the statistical tool was inherently wrong here. Binomial distribution was used for these tests; however, this would require that voters vote independently of each other and as if they were



flipping coins to decide their votes. These requirements are not met as voters can influence others' votes, and most placed their ballots while pre-meditatively knowing their decision. Thus, a binomial model was not the right choice; it might have been worthwhile using a hypergeometric distribution instead.

As for the Z-scores computed by comparing the early ballots and later tabulated ballots, the analyst failed to mention that in those four states, all absentee ballots cannot be counted before Election Day [8]. Furthermore, there was no mention of the pandemic being a possible reason for the more significant number of absentee ballots in 2020 than in 2016. The analyst also did not provide the voters' demographic or geographical data tabulated later during the Election day. These ballots could have come from many diverse or few regions. Nevertheless, it was also reported that Democrats were much more likely than Republicans to vote as absentees [9]. Also, there was no confirmation that all the remaining ballots post 3:10 am on November 4 were predominantly absentee ballots. They were also reporting a bottleneck of ballots to be counted from the most populous counties on Election day, which are, of course, centered around heavily Democratic cities [10].

As for the second part of his analysis, the analyst pointed out that the low-rejection rate favored Biden. The 2020 election saw a nation-wide decrease in rejection rates, where only Mississippi saw an increase from 1.6 % to 2.3 % and Illinois from 1.6 % to 1.7%. The rest of the states saw decreases such that Massachusetts saw a remarkable decrease from 3.3 % to 0.6 %. Based on election-administration experts, several factors have led to this decline. An essential factor is that many absentee voters submitted early, and many states reported a steep drop-off in the share of absentee votes received post-deadline. They are other factors that they alluded to, such as social-media reminders, improved mailing services, and changed policies that contributed to fewer rejected absentee votes [11]. Thus, the analyst was simply hopeful in assuming that the rejection rate should have remained at 6.42 % and not 0.36 %.

For the last part, concerning the unregistered 174,384 absentee ballots, these ballots came from eligible voters. Chris Thomas, who served as Michigan's director of elections for 36 years, stated in sworn testimony that those ballot verifications took place before the tabulation [12]. As a result, a judge refused to stop Detroit-area votes' certification based on the premise of this case [13].

In conclusion, despite a statistical analysis being done by a notable figure in economics, it remained worthwhile to investigate his assumptions, choice of model, and dig deeper into his claims. Nonetheless, many aspects of his analysis were found dubious and incorrect.



Appendix

Table 1. Early Ballots and Percent Increases Between 2016 and 2020.

State	2016	2020	Percentage Change (%)
Georgia	2,438,644	3,957,889	162.30
Michigan	1,277,405	3,111,414	243.60
Pennsylvania	288,996	2,504,518	866.60
Wisconsin	825,620	1,924,838	233.10

Table 2. Z scores of Battleground States.

State	Biden Votes	& Clinton Percentage	Early to Later
Georgia	396.3	108.7	1891
Michigan	290.4	90.7	736
Pennsylvania	198.5	77	1271
Wisconsin	333.1	107.4	586

Table 3. 2016 Mail-in Absentee Ballots in Georgia

2016 Mail-in Volume	213,033
2016 Mail-in Ballots Rejected	13,677
2016 Mail-in Rejection Rate	6.42 %

Table 4. 2020 Mail-in Absentee Ballots in Georgia

2020 Mail-in Volume	1,316,943
2020 Mail-in Ballots Rejected	4,786
2020 Mail-in Rejection Rate	0.36 %



References

1. Coleman, W. T. (2020). The Supreme Court of the United States: *The John F. Sonnett Memorial Lectures at Fordham University School of Law*, 12548(Mc 059), 200–233.
<https://doi.org/10.2307/j.ctv19x569.18>.
2. For, M., Consideration, E., The, O. F., For, M., To, L., Bill, F. A., Complaint, O. F., Expedition, F. O. R., Any, O. F., Consideration, P., Motion, F., Interim, F. O. R., & Is, R. (2020). *In the Supreme Court of the United States*. 20.
3. Wickham, H. (2015). *Done wrong*. http://www.amazon.de/Advanced-Chapman-Hall-Hadley-Wickham/dp/1466586966/ref=sr_1_1?s=books-intl-de&ie=UTF8&qid=1459782184&sr=1-1&keywords=Advanced+R
4. Lerer, L., Epstein R. J. (2021). *Why These Voters Rejected Hillary Clinton but Are Backing Joe Biden*. <https://www.nytimes.com/2020/10/18/us/politics/biden-clinton.html>.
5. Hal, M., Gal, S. (2020). *How the 2020 election results compare to 2016, in 9 maps and charts*. <https://www.businessinsider.com/2016-2020-electoral-maps-exit-polls-compared-2020-11>.
6. Grzeszczack, J. (2020). *Election 2020 Voter Turnout 'at 67 Percent'*, Highest in 120 Years. <https://www.newsweek.com/election-2020-voter-turnout-67-percent-highest-120-years-1544552>.
7. Chaggari, S. (2020). *How the COVID-19 pandemic has changed the US election*. <https://www.aljazeera.com/news/2020/10/29/how-the-pandemic-has-changed-the-u-s-election>.
8. NCSL (2020). *VOPP Table 16: When Absentee/Mail Ballot Processing and Counting Can Begin*. <https://www.ncsl.org/research/elections-and-campaigns/vopp-table-16-when-absentee-mail-ballot-processing-and-counting-can-begin.aspx>.
9. Litke, E. (2020). *Split in voting methods puts COVID, mail center stage in Wisconsin*. <https://www.politifact.com/article/2020/oct/10/voting-covid-mail-wisconsin-republican-democrat/>.
10. Litke, E. (2020). *Trump again flat wrong with claims about Wisconsin voter fraud*. <https://www.politifact.com/factchecks/2020/nov/20/donald-trump/trump-again-flat-wrong-claims-about-wisconsin-vote/>.
11. Rakich, N., *Why So Few Absentee Ballots Were Rejected In 2020*. <https://fivethirtyeight.com/features/why-so-few-absentee-ballots-were-rejected-in-2020/>.
12. Hendrickson, C. (2020). *Texas AG's Supreme Court lawsuit makes false claim about absentee ballots in Detroit*. <https://www.freep.com/story/news/local/michigan/detroit/2020/12/11/supreme-court-election-case-makes-false-claim-detroit-ballots/6509541002/>.
13. White, E. (2020). *Judge refuses to stop certification of Detroit-area votes*. <https://apnews.com/article/joe-biden-donald-trump-michigan-elections-detroit-1eb5f7ea8cbfb0031769e300f4d529ce>.