The 2008 Presidential Election: Statistical Analysis of Voting Trends

Allison Deal
Rhyian Johnson
Hillary Superak
Sarah Wong

ESE 405 – Reliability & Quality Control
December 8, 2008
Contents

I. Introduction ........................................................................................................................................... 2

II. Methodology & Assumptions ............................................................................................................. 3

III. Calculations, Graphs, and Analysis ................................................................................................ 5

   A. Gender .............................................................................................................................................. 5
   B. Age .................................................................................................................................................. 6
   C. Race ................................................................................................................................................. 7
   D. Income Level ................................................................................................................................. 8
   E. Race and Gender ............................................................................................................................ 9

V. Conclusions ......................................................................................................................................... 12

References ............................................................................................................................................. 14

Appendix .............................................................................................................................................. 15
I. Introduction

The 2008 presidential election, and the associated campaign events, comprised a period of United States political activity that will go down in history as one of the most consequential and momentous occasions ever. Voter turnout reached its highest level of all time, with over 128 million votes cast. Moreover, fundraising for both major parties reached record high amounts, leading to abundant advertising and media coverage. The election, which was tracked around the globe, truly had an international impact beyond the borders of the United States.

One of the primary reasons for the historical nature of the 2008 presidential election was the backgrounds of the candidates. The Democratic ticket consisted of Barack Obama, a senator from Illinois, and his running-mate, Joe Biden, a senator from Delaware; the Republican ticket consisted of John McCain, a senator from Arizona, and Sarah Palin, the governor of Alaska. Obama was the first African-American candidate for a major political party; Palin, if elected, would have been the first ever female vice-president.

The purpose of this report was to analyze the trends of voters in the 2008 presidential election. Specifically, the following analyses investigate the voting patterns of various demographic groups. The principal question of interest was:

*Which demographic factors had the most impact on the candidate selected by a voter?*

Another topic of analysis was the accuracy of pre-election polls. Numerous media sources published weekly polls of the opinions of likely voters in the weeks leading up to Election Day. Clearly, the proportions of voters leaning towards Obama or McCain fluctuated on a weekly basis. For this analysis, the question of interest was:

*Do pre-election polls provide a more accurate prediction of the actual election outcome as Election Day draws nearer?*
II. Methodology & Assumptions

The voting trends of the 2008 election were investigated by comparing the overall proportion of the population that voted for the winning candidate, Barack Obama, to the proportion of voters in different demographic groups who favored Obama at the polls. A one-proportion z-test was utilized in the calculation of a p-value for each test:

\[
z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}
\]

The test statistic, \(z\), was calculated based on \(p_0\), the overall proportion of the population that voted for Obama, and \(\hat{p}\), the proportion for the given demographic group that voted for Obama. The number of individuals in each group, \(n\), was calculated by multiplying the percentage of total respondents in the given demographic group by the total number of respondents. A positive z-score indicated that the members of the certain demographic voted more strongly for Obama than the overall American population voted. The following hypotheses were tested for each demographic parameter:

**Null Hypothesis**

\[H_0: \text{The proportion of surveyed voters in the category of interest who voted for Obama is the same as the overall population proportion of voters who favored Obama.} \]

\[\hat{p} = p_0\]

**Alternate Hypothesis**

\[H_a: \text{The proportion of surveyed voters in the category of interest who voted for Obama is different from the overall population proportion of voters who favored Obama.} \]

\[\hat{p} \neq p_0\]

For the pre-election polling data, the sample proportion from each point in time was compared to \(p_0\) based on the overall election data. The z-scores were computed and used in the comparisons.
Negative z-scores indicated that fewer voters were leaning towards Obama in the pre-election poll than actually voted for Obama; positive z-scores indicated that more voters were leaning towards Obama.

A standard significance level of $\alpha = 0.05$ was used in all analyses. Once the test statistic was calculated, it was used to generate the p-value for the test. If the p-value was less than 0.05, then the variable being examined was deemed to be a significant factor in voters’ decisions.

In order to use a one-proportion z-test, we must work under the assumption that the samples are simple random samples. All data were obtained from CNN News. The sampling methodology employed by CNN does legitimately constitute simple random sampling. The data are based on exit polls from randomly selected precincts across the nation. In other words, every precinct has an equal chance of being included in the sample; larger or more important precincts are not given extra weight.

Interviewers stand outside the selected polling locations and question every third or fifth voter, depending on the size of the precinct. Voters are interviewed throughout the entire duration of the precinct’s hours. This aspect of the sampling methodology helps avoid bias that could potentially arise if, for example, only morning or evening voters were surveyed. Overall, the methodology used to obtain the data for this study is creditable with creating a simple random sample, so this assumption is met.

The other major assumption that goes into a one-proportion z-test is that the samples are sufficiently large. This assumption is required for the normal approximation to be valid. Specifically, the sample must meet the condition that $np_0 > 10$ and $n(1-p_0) > 10$. The value of $p_0$, the proportion of the total population of voters that selected Obama, was $0.527$, or $52.7\%$. The total sample size of all voters surveyed in the exit poll was $17,836$. Therefore, without going through any calculations, it is obvious that for any demographic subgroup of this sample, both $np_0$ and $n(1-p_0)$ will greatly exceed 10 since the overall sample is so large.
III. Calculations, Graphs, and Analysis

The following analyses investigate several questions concerning the habits of American voters. Although all data were drawn from the 2008 Presidential election, conclusions can be extrapolated to other local and national elections as well.

A. Gender

Is the proportion of males who voted for Obama significantly different from the overall population proportion of both genders who voted for Obama?

Among the 8,383 male voters (47% of the total population), 49% voted for Obama and 48% voted for McCain (Fig. 1)

\[
\hat{p} = \frac{4900}{8383} = 0.583 \\
\hat{q} = 1 - \hat{p} = 0.417 \\
\sigma = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.583(0.417)}{8383}} = 0.0068 \\
z = \frac{\hat{p} - p_0}{\sigma} = \frac{0.49 - 0.527}{0.0068} = -6.785 \\
p-value \approx 0
\]

Since the p-value for this test is approximately 0, there is strong evidence that the proportion of males who voted for Obama is significantly different from the overall population proportion who voted for Obama. Specifically, since the z-score is negative, a smaller proportion of males voted for Obama relative to the entire population; male voters tended to favor McCain more strongly than the overall population of both genders.
B. Age

Is the proportion of young voters (age 18-44) favoring Obama significantly different from the overall population proportion of voters of all ages who voted for Obama?

Among the 8,383 voters between the ages of 18 and 44 who were surveyed, 57.4% reported voting for Obama; only 41.8% voted for McCain (Fig. 2).

![Age (18-44) Pie Chart](image)

**Fig. 2.** Percentages of young voters for each candidate.

\[
z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0.574 - 0.527}{\sqrt{\frac{0.527(0.473)}{8,383}}} = 8.619 \quad \text{p-value} \approx 0
\]

The sample size, \( n \), was calculated by pooling two groups of voters from the sample data: voters age 18-29 (18% of the sample) and voters age 30-44 (29% of the sample). By taking a weighted average of the proportions in these two groups who voted for Obama, the overall “young voter” statistic was obtained.

The p-value for this test is approximately 0, indicating that the proportion of young voters who voted for Obama is significantly different from the overall population proportion who voted for Obama. Specifically, since the z-score is positive, a larger proportion of young voters favored Obama relative to the proportion of the entire voting population who favored Obama. This is an interesting finding since a major part of Obama’s campaign strategy was to target first-time voters, who tend to fall into the 18-44 age group.
C. Race

Is the proportion of African American voters favoring Obama significantly different from the overall population proportion of voters of all ethnicities who voted for Obama?

Among the 2,319 African American voters (13% of the total population), 95% voted for Obama and 4% voted for McCain (Fig. 3)

\[
\hat{p} = \frac{2319 \times 95}{2319} = 0.95
\]

\[
p_0 = 0.527
\]

\[
z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0.95 - 0.527}{\sqrt{\frac{0.527(0.473)}{2319}}} = 40.800
\]

p-value ≈ 0

Since the p-value for this test is approximately 0, there is strong evidence that the proportion of African Americans who voted for Obama is significantly different from the overall population proportion who voted for Obama. Specifically, since the z-score is positive and very large, a much larger proportion of African Americans voted for Obama relative to the entire population. It can be concluded that African American voters tended to favor Obama more strongly than the overall population of all ethnicities.

Fig. 3. Percentages of African American voters for each candidate.
D. Income Level

Is the proportion of voters who make less than $100,000 a year and voted for Obama significantly different from the overall population proportion of all incomes who voted for Obama?

Among the 13,199 lower-income voters (74% of the total surveyed population), 55% voted for Obama and 43% voted for McCain (Fig. 4).

![Income Level (<$100,000)](image)

**Fig. 4.** Percentages of lower-income voters for each candidate.

\[
\hat{p} - p_0 = \frac{0.55 - 0.527}{\sqrt{\frac{0.527(0.473)}{13,199}}} = 5.293 \quad \text{p-value} \approx 0
\]

The p-value for this test is again approximately 0, suggesting there is strong evidence that the proportion of voters who earn less than $100,000 per year and who voted for Obama is significantly different from the overall population proportion who voted for Obama. Specifically, since the z-score is positive, a larger proportion of this lower-income demographic voted for Obama relative to the entire population; lower-income voters tended to favor Obama more strongly than the overall population of all income levels.
E. Race and Gender

Is the proportion of African American male voters who voted for Obama significantly different from the overall population proportion of all voters who voted for Obama?

Among the 892 African American males (5% of the total surveyed population), 95% voted for Obama and 5% voted for McCain (Fig. 5)

![Race & Sex (African American Male)](image)

**Fig. 5.** Percentages of African American male voters for each candidate.

\[
z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0.95 - 0.527}{\sqrt{\frac{0.527(0.473)}{892}}} = 25.301 \quad \text{p-value} \approx 0
\]

The p-value for this test – again approximately 0 – indicates strong evidence that the proportion of voters who are African American males and who voted for Obama is significantly different from the overall population proportion who voted for Obama. Specifically, since the z-score is positive and very large, a much larger proportion of this demographic voted for Obama relative to the entire population; African American male voters tended to favor Obama more strongly than the overall population of all backgrounds. This is a significant and expected trend because Obama himself is the first African American male to run (and to be elected) for office. It is interesting to note that almost all members of this demographic seemed to vote for someone of the same race and gender as themselves.
After considering the voting trends of several demographics, a new question was raised: *Do pre-election polls provide a more accurate prediction of the actual election outcome as Election Day draws nearer?* New data from incremental polling periods were gathered and the following analyses were performed.

Table 1 shows the percentage of voters who favored each candidate at various points in time leading up to the election. Table 1 includes data from approximately one poll per month for the five months prior to the election.

<table>
<thead>
<tr>
<th>DATE</th>
<th>OBAMA</th>
<th>z-score</th>
<th>p-value</th>
<th>MCCAIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>6/15/2008</td>
<td>0.470</td>
<td>-6.000</td>
<td>0</td>
<td>0.480</td>
</tr>
<tr>
<td>7/13/2008</td>
<td>0.490</td>
<td>-3.895</td>
<td>0</td>
<td>0.460</td>
</tr>
<tr>
<td>8/22/2008</td>
<td>0.490</td>
<td>-3.895</td>
<td>0</td>
<td>0.450</td>
</tr>
<tr>
<td>9/22/2008</td>
<td>0.520</td>
<td>-0.737</td>
<td>0.461</td>
<td>0.430</td>
</tr>
<tr>
<td>10/22/2008</td>
<td>0.540</td>
<td>1.368</td>
<td>0.171</td>
<td>0.430</td>
</tr>
<tr>
<td>11/2/2008</td>
<td>0.530</td>
<td>0.316</td>
<td>0.975</td>
<td>0.440</td>
</tr>
<tr>
<td><strong>11/4/2004</strong></td>
<td><strong>0.527</strong></td>
<td></td>
<td></td>
<td><strong>0.459</strong></td>
</tr>
</tbody>
</table>

*Table 1. Percentage of voters favoring each candidate, based on pre-election polls.*

As shown, the p-values were non-zero as the election drew nearer. In the polls furthest in time from the election, the sampled proportions were significantly different from the true population proportions on Election Day. Closer to Election Day, the proportions were not significantly different. This trend indicates that the pre-election polls became more accurate to the true election outcome over time. Notably, the highest p-value was from the poll closest to Election Day. The results of this analysis show that voters became more decisive about their choice of candidates as the election drew nearer.
Figure 6 on the following page shows the trends of American voters, as indicated by monthly pre-election polls. It should be noted that initially, McCain appeared to lead Obama in the national polls. However, over time, Obama’s lead grew as the election drew nearer. This trend may be a result of events leading up to the election. Many political analysts declared Obama victorious in the three debates that took place weekly in late September and early October. Figure 6 shows that during this period of time, Obama’s lead over McCain grew.
V. Conclusions

Barack Obama’s victory in the presidential election was a groundbreaking event that will go down in history for its political implications. Analysis of demographic factors provides a great deal of insight into the trends and patterns of Americans’ voting habits in general. While there are seemingly limitless demographic categories that can be analyzed, this particular study examined voter trends based on gender, age, income level, race, and leanings over time. Performance of a two-sided one-proportion hypothesis test enabled acceptance or rejection of the null hypothesis that each category of interest would vote in the same manner as the overall voting population.

The results of the hypothesis tests revealed that more important than acceptance or rejection of the null hypothesis is the strength of this decision. This analysis suggests that the demographics studied (gender, age, income level, and race) do in fact have significance in voter leanings. However, some categories have more strength than others. For example, the strongest correlation was found in African-Americans’ votes for Obama, and particularly, his overwhelming win of the African-American male vote. While analyses like these are highly simplified, it is interesting to see which demographic groups had the most impact on Obama’s win. This study suggests that race was the most influential factor followed by gender, age, and lastly income level. From June through September, there were significant shifts in Obama’s popularity; it was clear before Election Day that the voting population was settling on their preference for Obama. One trend in the pre-election polls that deserves notice is that the z-scores tended to get closer to 0 as Election Day drew nearer. This trend indicates that the polls became more accurate to the true Election Day proportions over time, meaning that they provided a better indication of the true results of the presidential election.

This analysis has fascinating implications for the campaigning policies of politicians in future elections. There is statistical evidence that certain demographic groups tend to vote in certain manners;
this information could provide a candidate with a valuable source of insight concerning which groups should be targeted at campaign events leading up to the election. The historical trends in the weekly polls show that voters are more likely to be swayed at an earlier point in time. As Election Day draws nearer, the statistics show that the reliability of polls of likely voters tends to increase as the estimated proportions draw closer to the true voting patterns of Americans on Election Day.
References

<http://abcnews.go.com/politics>

