55,000 residents desperately need your help!*

Andrew Gelman†

February 27, 2004

One day last summer, we received a fax, entitled HELP!, from a member of a residential organization:

> Last week we had an election for the Board of Directors. Many residents believe, as I do, that the election was rigged and what was supposed to be votes being cast by 5,553 of the 15,372 voting households is instead a fixed vote with fixed percentages being assigned to each and every candidate making it impossible to participate in an honest election.
>
> The unofficial election results I have faxed along with this letter represent the tallies. Tallies were given after 600 were counted. Then again at 1200, 2444, 3444, 4444, and final count at 5553.
>
> After close inspection we believe that there was nothing random about the count and tallies each time and that specific unnatural percentages or rigged percentages were being assigned to each and every candidate.
>
> Are we crazy? In a community this diverse and large, can candidates running on separate and opposite slates as well as independents receive similar vote percentage increases tally after tally, plus or minus three or four percent? Does this appear random to you? What do you think? HELP!

Figure 1 shows a subset of the data. These vote tallies were deemed suspicious because the proportion of the vote received by each candidate barely changed throughout the tallying. For example, Clotelia Smith's vote share never went below 34.6% or above 36.6%. How can we HELP these people and test their hypothesis?

We start by plotting the data: for each candidate, the proportion of vote received after 600, 1200, . . . votes; see Figure 2. These graphs are difficult to interpret, however, since the data points are not in any sense independent: the vote at any time point includes all the votes that came before. We handle this problem by subtraction to obtain the number of votes for each candidate in the intervals between the vote tallies: the first 600 votes, the next 600, the next 1244, then next 1000, then next 1000, and the final 1109, with the total representing all 5553 votes.

*To appear in *Chance*. We thank a referee for helpful comments, Jouni Kerman for help on the data analysis, and the National Science Foundation for financial support.

†Professor, Department of Statistics and Department of Political Science, Columbia University, New York, gelman@stat.columbia.edu. www.stat.columbia.edu/∼gelman/.

Figure 3 displays the results. Even after taking differences, these graphs are fairly stable—but how does this variation compare to what would be expected if votes were actually coming in at random? We formulate this as a hypothesis test and carry it out in five steps:

1. *The null hypothesis* is that the voters are coming to the polls at random. The fax writer believed the data contradicted the null hypothesis; this is what we want to check.

2. *The test statistic* is some summary of the data used to check the hypothesis. Since the concern was that the votes were unexpectedly stable as the count proceeded, we define a test statistic to summarize that variability. For each candidate i, we label $y_{i1}, \ldots, y_{i6}$ to be the numbers of votes received by the candidates in each of the six recorded stages of the count. (For example, from Figure 1, the values of $y_{i1}, y_{i2}, \ldots, y_{i6}$ for Earl Coppin are $55, 51, \ldots, 104$.) We then compute $p_{it} = y_{it}/n_t$ for $t = 1, \ldots, 6$, the proportion of the votes received by candidate i at each stage. The test statistic for candidate i is then the sample standard deviation of these six values $p_{i1}, \ldots, p_{i6}$, a measure of the variation in his or her votes:
$$T_i = \mathrm{sd}_{t=1}^{6} p_{it}.$$

3. *The theoretical distribution of the test statistic if the null hypothesis were true.* Under the null hypothesis, the six subsets of the election are simply six different random samples of the voters, with a proportion π_i who would vote for candidate i. From the binomial distribution, the proportion p_{it} then has a mean of π_i and a variance of $\pi_i(1-\pi_i)/n_t$. On average, the variance of the six p_{it}'s will equal the average of the six theoretical variances, and so the standard deviation of the p_{it}'s—the test statistic—should equal, on average, the theoretical value $\mathrm{avg}_{t=1}^{6} \pi_i(1-\pi_i)/n_t$. The probabilities π_i are not known, so we follow standard practice and insert the empirical probabilities, p_i, so that the expected value of the test statistic, for each candidate i, is
$$T_i^{\mathrm{theory}} = p_i(1-p_i)\mathrm{avg}_{t=1}^{6}(1/n_t).$$

4. *Comparing the test statistic to its theoretical distribution.* Figure 4 plots the observed and theoretical values of the test statistic for each of the 27 candidates, as a function of the total number of votes received by the candidate. The theoretical values follow a simple curve (which makes sense, since the total number of votes determines the empirical probabilities p_i, which determine T_i^{theory}), and the actual values appear to fit the theory fairly well, with some above and some below.

5. *Numerical summaries using χ^2 tests.* We can also express the hypothesis tests numerically. Under the null hypothesis, the probability of a candidate receiving votes is independent of the time of each vote, and thus the 2×6 table of votes excluding or excluding each candidate would be consistent with the model of independence. (See Figure 4 for an example.) We can then compute for each candidate a χ^2 statistic, $\sum_{j=1}^{2} \sum_{t=1}^{6} (\mathrm{observed}_{jt} - \mathrm{expected}_{jt})^2/\mathrm{expected}_{jt}$, and compare to a χ^2 distribution with $(6-1) \times (2-1) = 5$ degrees of freedom.

Unlike the usual application of χ^2 testing, in this case we are looking for unexpectedly *low* values of the χ^2 statistic (and thus p-values close to 1), which would indicate vote propoportions that have suspiciously little variation over time. In fact, however, the χ^2 tests for the 27 candidates show no suspicious patterns: the p-values range from 0 to 1, with about 10% below 0.1, about 10% above 0.9, and no extreme p-values at either end.

Another approach would be to perform a χ^2 test on the entire 27×6 table of votes over time (that is, the table whose first row is the top row of the left table on Figure 5, then continues with the data from Earl Coppin, Clarissa Montes, and so forth). This test is somewhat suspect since it ignores that the votes come in batches (each voter can choose up to 6 candidates) but is a convenient summary test. The value of the χ^2 statistic is 114.7, which, when compared to a χ^2 distribution with $(27-1) \times (6-1) = 130$ degrees of freedom, has a p-value of 0.83—indicating slightly less variation than expected, but not statistically significant. That is, if the null hypothesis were true, we would not be particularly surprised to see a χ^2 statistic of 114.7.

Thus, our graphical and numerical summaries both conclude that the intermediate vote tallies are consistent with random voting. As we explained to the writer of the fax, opinion polls of 1000 people are typically accurate to within 2%, and so, if voters really are arriving at random, it makes sense that batches of 1000 votes are highly stable. This does not rule out the possibility of fraud, of course—it merely shows that this aspect of the voting is consistent with the null hypothesis.

In any case, the fax writer's vision of overturning a fraudulent election—and our vision of exposing injustice (and maybe getting on TV) were frustrated, leaving us in the end with nothing more than a good story and a rare example of non-overdispersed binomial data.

Clotelia Smith	208	416	867	1259	1610	2020
Earl Coppin	55	106	215	313	401	505
Clarissa Montes	133	250	505	716	902	1129
...	...	...	...	...	...	...

Figure 1: Subset of results from the cooperative board election, with cumulative votes for each candidate (names altered for anonymity) tallied after 600, 1200, 2444, 3444, 4444, and 5553 votes. These data were viewed as suspicious because the proportion of votes for each candidate barely changed as the vote counting went on. (There were 27 candidates in total, and each voter was allowed to choose 6 candidates.)

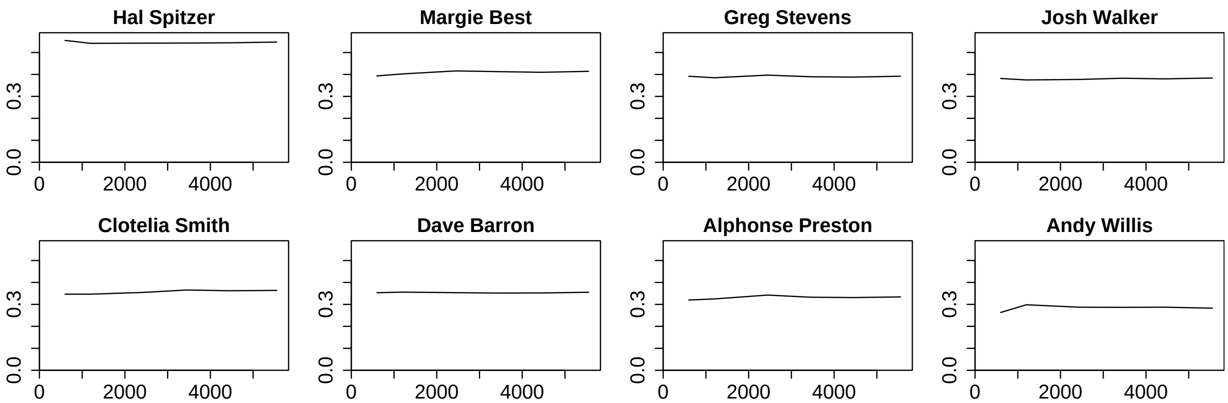

Figure 2: Proportion of votes received by each candidate in the cooperative board election, after each stage of counting: 600, 1200, 2444, . . . , 5553 votes. There were 27 candidates in total; for brevity we display just the leading 8 vote-getters here. The vote proportions appear to be extremely stable over time; this might be misleading, however, since the vote at any time point includes all the previous vote tallies. See Figure 3.

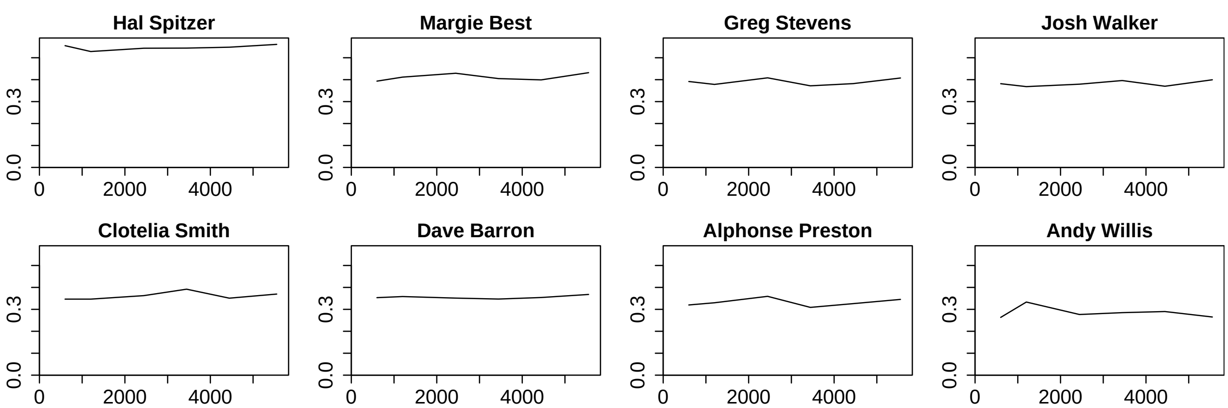

Figure 3: Proportion of votes received by each of the 8 leading candidates in the cooperative board election, at each disjoint stage of voting: the first 600 votes, the next 600, the next 1244, then next 1000, then next 1000, and the final 1109, with the total representing all 5553 votes.

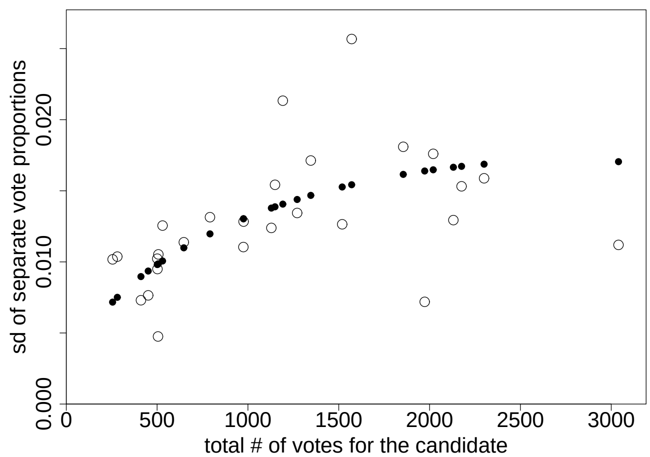

Figure 4: The open circles show, for each of the 27 candidates in the cooperative board election, the standard deviation of the proportions of the vote received by the candidate in the first 600, next 600, next 1244, ..., and the final 1109 votes, plotted vs. the total number of votes received by the candidate. The solid dots show the expected standard deviation of the separate vote proportions for each candidate, based on the binomial model that would be appropriate if voters were coming to the polls at random. The actual standard deviations appear consistent with the theoretical model.

Counts within each time interval

	1	2	3	4	5	6	Total
Yes	208	208	451	392	351	410	2020
No	392	392	793	608	649	699	3533
Total	600	600	1244	1000	1000	1109	5553

Expected counts

218.3	218.3	452.5	363.8	363.8	403.4
381.7	381.7	791.5	636.2	636.2	705.6

Figure 5: Left table: votes including and excluding Clotelia Smith (see top row of Figure 1) within each of the six time intervals tallied in the election. Right table: expected counts under the model of independence ($600 \times 2020/5553 = 218.3$, and so forth). The resulting χ^2 statistic is $\sum_{j=1}^{2}\sum_{t=1}^{6}(\text{observed}_{jt} - \text{expected}_{jt})^2/\text{expected}_{jt} = 5.8$, which when compared to a χ^2_5 distribution indicates no problems, with a p-value of 0.32. The χ^2 tests for the other candidates similarly show no problem with the null hypothesis.