Supplemental Report on 2004 Washington Gubernatorial Election

Jonathan N. Katz

California Institute of Technology

28 April 2005

Since completing my report of April 10, 2005, I have received updated data on invalid ballots cast in the Gubernatorial election. In the new dataset all of the invalid ballots can now be allocated to voting precincts. This greatly improves and simplifies the analysis.

None of my substantive findings have changed from my previous report. However, the new analysis of the distribution of invalid votes here supersedes my analysis presented in Section 4.2 of my earlier report.

The next section of the report clarifies how the estimation was conducted and the final section presents the updated analysis of the invalid votes

1 Estimating the Distribution of Invalid Ballots

The thought experiment that underlies the estimation is to think about drawing the 1183 invalid ballots from the pool of all of the ballots cast in the election. How many of these ballots would we expect to be votes for Gregoire? Given that we use secret ballot, there is no way to distinguish invalid from valid ballots and thus a randomly selected ballot from a given precinct, or any other geographic unit, must have the same chance of being a vote for the Democratic candidate given by the known fraction of Democratic vote in the precinct.

This is a very well studied problem in statistics (see, for example, Larson 1982:155–164). Drawing a random sample from a known population where units are of two possible types, in our case Democratic and Republican votes, is governed by the Binomial distribution.[1]

---

[1] Although the more standard case is where we want to estimate the true fraction of a type in the population given a random sample from it. Here we know the true fraction of votes for Gregoire, but we want to know is a possibly drawn sample what is the number of Gregoire votes. This leads to some differences in the formula.

Let x_i be the number of Democratic votes from a sample of n_i in a given precinct i. Then the expected value — i.e., our best estimate — of x_i is:

$$E[x_i] = n_i p_i$$

Where p_i is the fraction of Democratic ballots in the precinct and n_i are the number of invalid ballots in the precinct.

In order to calculate the confidence interval of our estimate, we will also need to know the variance of x_i, which summarizes the spread of x_i we would see on any given draw. The variance is

$$V[x_i] = n_i p_i (1 - p_i)$$

The quantity of interest is the total number of invalid Democratic votes in the state, so we need to sum over the precincts. The expected number of total invalid Democratic votes is

$$E\left[\sum_i x_i\right] = \sum_i n_i p_i$$

The variance of this sum is just the sum of the variances (since the estimates are independent across precincts):

$$V\left[\sum_i x_i\right] = \sum_i n_i p_i (1 - p_i)$$

The 95% confidence interval[2] can be constructed by using the standard Normal approximation:

$$\sum_i n_i p_i \pm 1.96 \times V\left[\sum_i x_i\right]^{\frac{1}{2}}$$

2 Impact of Invalid Votes on Election Outcomes

Since completing my report of April 10, 2005 we have revised numbers of invalid ballots. A new summary of the invalid ballots by category can be found in Table 1. This table is slightly different from my previous report in that any invalid ballot that has been challenged has been removed and thus these totals are really a conservative estimate of the number of invalid ballots in the election. The other central difference between the two tables is that the

---

[2]The 95% confidence interval represents our uncertainty since we are estimating the effect from only a sample of data. The 95% confidence interval is constructed so that if we were to collect a new sample of data and re-estimate the quantity of interest, 95 times out of 100 this new estimate would be in this interval.

Table 1: Updated Invalid Votes in the 2004 Washington Gubernatorial Election Excluding All Challenged Cases

Type	King County	Other County	Total
Felons	724	219	943
Deceased	35	14	49
Dual Multi-State	2	1	3
Dual In-State	10	2	12
Non-Citizens	2	0	2
Not Registered Provisional	174	0	174
Total	947	236	1183

total number of invalids has risen to 1183. All of the other patterns discussed in my previous report still hold.

Given that we can now allocate all of the invalid ballots to precincts it is possible to do a single analysis as outlined above. The estimated distribution of the invalid ballots to the candidates in the gubernatorial race can be found in Table 2. The key results are found the third row, labeled All, that allocates all of the invalid votes. However, for consistency with my earlier report I have also separately allocated the invalid ballots from felons, by far the most numerous of the invalid ballots, and non-felon.

The second column in Table 2 gives the total number of invalid ballots of that type — for example, there were 240 non-Felon invalid ballots. The next two columns of the table give the estimated number of those ballots that voted for Gregoire and Rossi respectively. It should be noted that only the fraction voting for Gregoire was estimated, since the number who voted for Rossi must be the total number minus the number who voted for Gregoire. The numbers in parenthesis are the 95% confidence interval for the estimate. For example, the lowest number of felon invalid votes for Gregoire that is consistent with the data is 556.30 and highest is 612.00.

Consider just the final row of the table that allocates all of the invalid ballots. We see that the fewest number of invalid votes for for Gregoire that is consistent with the data is 691.49. This implies that 491.51 voted for Rossi if this case held. Gregoire's lead would be $129 - 691.49 + 491.51$ or -70.98. In other words, if the invalid votes were excluded, Gregoire would have lost by just under 71 votes instead of winning by 129 as in the final manual recount.

Table 2: Estimated Distribution of Invalid Votes

Type	Total	Gregoire	Rossi
Felons	943	584.30	358.70
		(556.60, 612.00)	(331.00, 386.40)
Non-Felons	240	138.35	101.65
		(124.09, 152.60)	(87.40, 115.91)
All	1183	722.64	460.36
		(691.49, 753.80)	(429.20, 491.51)

In fact, given the results in Table 2, we can construct the 95% confidence interval for Gregoire's margin of victory if the invalid votes were removed. It is -70.98 to -195.60. Thus, under every likely scenario that is consistent with the observed data, Rossi would win if the invalid votes were excluded.

4

3   References

Larson, Harold J. 1982. *Introduction to Probability Theory and Statistical Inference*, 3rd Edition. New York: John Wiley &Sons.