The Error in the Florida Presidential Vote Count is 800 Votes

Robert Weiss
Department of Biostatistics
UCLA School of Public Health
Los Angeles CA 90095-1772 U.S.A.
Phone (310) 206-9626
Fax (310) 267-2113

November 18, 2000

Abstract

The Florida Presidential Contest count and recount is a natural experiment which provides us the ability to estimate the uncertainty in voting tabulations at the county and state level. I analyzed the changes in vote counts for Bush and for Gore from the original vote tally to the first vote recount. The standard deviation of the usual vote counting is approximately 800 votes. The probability that an accurate count would overturn a Bush lead of 300 is estimated at 35% and the probability of overcoming a lead of 930 is 12%.

The 2000 Presidential election is extremely close. The result will depend on the final Florida vote tally. The original vote total had George Bush leading Al Gore by 1784 votes. A recount by all Florida counties using their usual methods of vote tabulation then put Bush's lead at 327 votes. Subsequent adjustments which were not the same for all counties gave a lead of 300 for Bush. I analyze the differences between the original count and the first recount for Bush and Gore, not including subsequent adjustments. These two counts were similar except for the ordering in time.

The Associated Press reported changes from the first count to the first recount by candidate and county. This data is given in tables 1 and 2.

I analyzed[1] the changes in Gore's and Bush's vote tabulations from the first to the second recount (columns 4 and 5 in tables 1 and 2). While these are usually positive, there are some negative changes. Two predictors were used to predict both the mean and the log of the variance of these values. Predictors used are (1) GOREPREF; the fraction difference in preferences between Bush and Gore ignoring other candidates, (2) LOGTOTAL the log of the total number of votes cast for Bush and Gore.

Neither covariate was an important predictor of the change of votes. However, both are important predictors of the variance. As the size of the county increases, the variance of the change increases. The effect is the same size for both Bush's change and Gore's change. The greater GOREPREF, the greater the variability of the change in Gore's vote. The smaller GOREPREF, the greater the variability of the change in Bush's vote. The effects are the same size, except for the difference in sign.

I calculate[2] the standard deviation of the total statewide vote difference between Bush and Gore to be 800 votes. Barring a perfect recount, any final tally that gives the victor a lead which is not a substantial multiple of 800 votes will lead many to correctly conclude that an accurate counting of the ballots might well have given Florida's electoral ballots to the other candidate. I calculate the probability that an accurate count might overcome a lead of 300 votes as 35%. The probability that an accurate count might overcome a lead of 930 votes (the current tally including absentee ballots) is 12%.

This analysis estimates variability due to machine recounts and organizational difficulties such as finding or losing ballots from one recount to the next. It does not include bias due to any causes such as missed, lost or disqualified ballots. It does not address directly the possible variability of hand retabulations compared to the machine tabulations.

Footnotes

Footnote 1. The statistical model used is

$$y_{ij} \sim N(\mu_{ij}, \sigma_{ij}^2) \tag{1}$$

where y_{ij} is the county i vote change for $j = 1$ Bush, or for $j = 2$ Gore. The models for μ_{ij} and σ_{ij}^2 are

$$\begin{aligned} \mu_{ij} &= \beta_0 + (-1)^j * \beta_1 * \textsc{Gorepref}_i + \beta_2 * \textsc{logtotal}_i \\ \log(\sigma_{ij}^2) &= \gamma_0 + (-1)^j * \gamma_1 * \textsc{Gorepref}_i + \gamma_2 * \textsc{logtotal}_i \end{aligned}$$

The coefficient of Gorepref was of opposite sign for Bush recount and gore recounts, hence the factor $(-1)^j$ in each linear predictor. The complete model includes a $N(0, V_k)$ prior for each coefficient β_k and γ_k, with $V_0 = 1000$, and $V_1 = V_2 = 2$. Results did not change when prior variances were increased to $V_0 = 10000$ and $V_1 = V_2 = 4$. Posteriors are generally bell-curve shaped. Table 3 gives output. WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml) was used for computations, with an MCMC sample of size 160000 for results in table (3) and 400000 for the remainding computations after ignoring the first 4000 samples for warmup.

Footnote 2. This model was used to calculate the uncertainty in the tabulations. Suppose vote counts g_{im} (for Gore) and b_{im} (for Bush) in county i and count m=1, or recount m=2 are distributed around some true county and candidate specific value $g_{\mathrm{truth},i}$ or $b_{\mathrm{truth},i}$ with a county and candidate specific variance, $\tau^2_{g,i}$ or $\tau^2_{b,i}$

$$\begin{aligned} g_{im} &\sim N(g_{\mathrm{truth},i}, \tau^2_{g,i}) \\ b_{im} &\sim N(b_{\mathrm{truth},i}, \tau^2_{b,i}) \end{aligned}$$

Then the differences $y_{i1} = b_{i2} - b_{i1}$ and $y_{i2} = g_{i2} - g_{i1}$ have variances $\sigma^2_{i1} = 2 * \tau^2_{b,i}$ and $\sigma^2_{i2} = 2 * \tau^2_{g,i}$, and the predictive variances for the truth are therefore $\sigma^2_{ij}/2$. Using this model for the variance produces a standard error for the total difference between Bush's total and Gore's total. Sensitivity analysis suggested that the conclusion of an 800 vote standard error was fairly robust. Introducing quadratics for the mean and variance in the formulae for μ_{ij} and $\log(\sigma^2_{ij})$ increases the standard error to 1000 and decreases the 35% probability somewhat, but all other changes did not change the 800 number by even a few percent and did not change the 35% at all.

Footnote 3. The author thanks Janet Sinsheimer and John Boscardin for helpful comments on earlier versions of this paper.

County	Gore Vote	Bush Vote	Gore change	Bush change
Alachua	47365	34124	65	62
Baker	2392	5610	0	0
Bay	18850	38637	0	0
Bradford	3075	5414	3	1
Brevard	97318	115185	0	0
Broward	386561	177323	43	44
Calhoun	2155	2873	0	0
Charlotte	29645	35426	4	7
Citrus	25525	29766	24	22
Clay	14632	41736	2	-9
Collier	29918	60433	13	7
Columbia	7047	10964	0	0
Dade	328764	289492	62	36
De Soto	3320	4256	-2	0
Dixie	1826	2697	1	-1
Duval	107864	152098	184	16
Escambia	40943	73017	-15	-12
Flagler	13897	12613	6	5
Franklin	2046	2454	4	6
Gadsden	9735	4767	170	17
Gilchrist	1910	3300	0	0
Glades	1442	1841	2	1
Gulf	2397	3550	8	4
Hamilton	1722	2146	4	-7
Hardee	2339	3765	-2	1
Hendry	3240	4747	1	4
Hernando	32644	30646	0	0
Highland	14167	20206	15	10
Hillsboro	169557	180760	28	47
Holmes	2177	5011	23	26
IndianRiver	19768	28635	-1	8
Jackson	6868	9138	0	0
Jefferson	3041	2478	3	-3

Table 1: Data as reported by the Associated Press. County, Gore vote at first recount, Bush vote from first recount, change in Gore vote from first count to first recount, change in Bush vote from first count to first recount. The difference between the candidates here is 327. The adjustments lead to a difference of 1782 at the first vote; an error of 2 votes is different from reported first totals of 1784. It is not known where the source of the error is, although it is not important in terms of the conclusions of this article.

County	Gore Vote	Bush Vote	Gore change	Bush change
Lafayette	789	1670	1	1
Lake	36571	50010	16	47
Lee	73560	106141	30	18
Leon	61425	39053	0	0
Levy	5398	6858	-5	-2
Liberty	1017	1317	6	1
Madison	3014	3038	3	0
Manatee	49177	57952	8	4
Marion	44665	55141	17	6
Martin	26620	33970	1	106
Monroe	16483	16059	0	0
Nassau	6879	16280	-73	-124
Okaloosa	16948	52093	24	50
Okeechobee	4588	5057	-1	0
Orange	140220	134517	105	41
Osceola	28181	26212	4	-4
Palm Beach	269696	152954	751	108
Pasco	69564	68582	14	1
Pinellas	200629	184823	417	-61
Polk	75196	90191	219	90
Putnam	12102	13447	11	8
Santa Rosa	12802	36274	7	26
Sarasota	72853	83100	-1	0
Seminole	59174	75677	286	384
St. Johns	19502	39546	20	49
St. Lucie	41559	34705	0	0
Sumter	9637	12127	3	1
Suwannee	4075	8006	-9	-8
Taylor	2649	4056	2	6
Union	1407	2332	8	6
Volusia	97063	82214	0	0
Wakulla	3838	4512	3	1
Walton	5642	12182	5	6
Washngton	2798	4994	2	11
Total	2909871	2910198	2519	1064

Table 2: Remainder of the data; all the same as in table 1.

μ/σ^2	parameter	mean	sd	2.5%	97.5%	$P(\beta > 0\|Y)$
μ	intercept	-2.93	8.91	-20.5	14.7	.37
	GOREPREF	.0757	1.37	-2.60	2.76	.52
	LOGTOTAL	.687	1.04	-1.37	2.74	.75
σ^2	intercept	-8.38	.967	-10.3	-6.46	0.0
	GOREPREF	2.47	.387	1.71	3.22	1.0
	LOGTOTAL	1.49	.0924	1.31	1.67	1.0

Table 3: Results from county level analysis. No coefficients are significant for the mean, but both the GOREPREF and LOGTOTAL are very significant for the variance. The sign of the GOREPREF coefficient enters positively for the Gore change data and negatively for the Bush change data. The 2.5%, 97.5% columns give left and right end points of a 95interval. The far right column is the probability that the coefficient, either β_k or γ_k is greater than zero. When the sign of the coefficient is well determined, we say that the parameter is statistically significant. Small numbers may be interpreted as a 1-sided pvalue, and large (near 1) can be subtracted from 1 to give a one-sided pvalue.