

A Statistical Study of Buchanan’s 2000 Presidential Vote Total in Palm Beach, FL

Layth C. Alwan
Associate Professor of Business Statistics
University of Wisconsin-Milwaukee
PO Box 742
Milwaukee, WI 53201
alwan@uwm.edu

11/11/00, 8 am, CDT

Introduction

Democratic officials, and in turn media, have reported that one of the central issues of concern in the Florida vote count for the 2000 Presidential election is a potential “irregularity” in Palm Beach county. It is suggested that a number of people inadvertently voted for Pat Buchanan when they intended to vote for Al Gore due to the format of the ballot. An examination of the ballot (see Figure 1 below) suggests that the assertion of a “confusing” ballot is not without merit. In a recent note, Professor Fox notes that such a butterfly ballot format corresponds to what previous researchers have called an “incompatibility between input and output” which has been recognized to increase effort and error (Fitts and Seeger, 1953; Wickens, 1984; these references are found at Professor Fox’s site). Professor Fox provides a wonderful analogy of a “stove in which burners are aligned in a square but controls are aligned in a straight line from left to right. Most people would have difficulty selecting the intended control – and most people would make occasional errors.” If we were to modify his analogy to be closer to the ballot format, we might consider the controls aligned in a straight line up and down the middle splitting the burners.

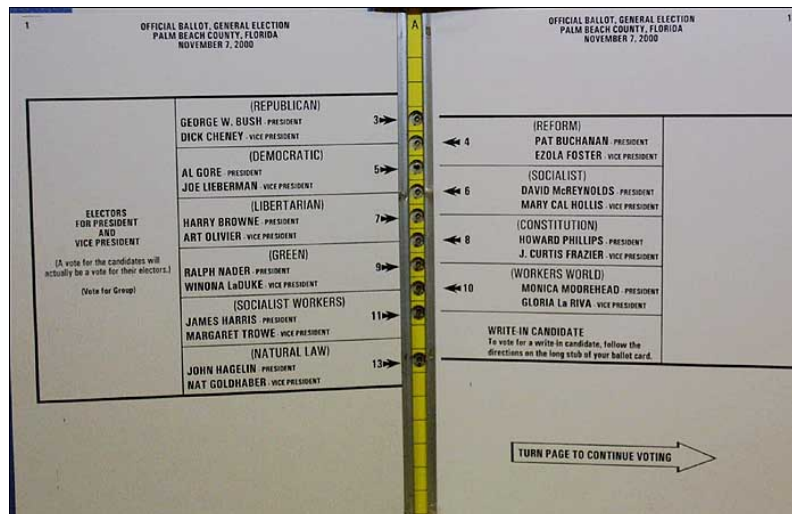


Figure 1 – Palm Beach Butterfly Ballot

There appears to be supporting evidence of the notion of confusion with the Palm Beach ballot given the large number of disqualified ballots (approx. 19,000) due to seemingly accidental punching of multiple candidates. Furthermore, it is curious to note that adjacent counties to Palm Beach were associated with substantially less Buchanan votes in percentage terms.

Isolated Analysis of the Buchanan Vote Totals

Our interest here is to statistically study the Buchanan vote total in Palm Beach and estimate its level of unusualness. We begin our analysis with the following chart provided by many media sources:

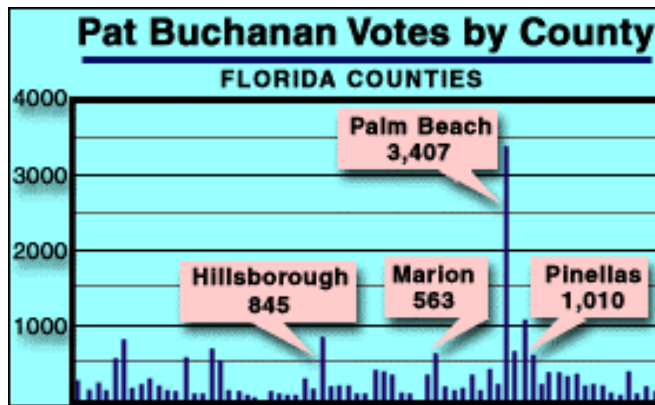


Figure 2 – County Vote Totals for Buchanan

From the above chart, a natural initial reaction is to believe that Buchanan’s vote count in Palm Beach is a gross outlier. Below is a distribution plot of Buchanan’s votes:

Dotplot: Buchanan

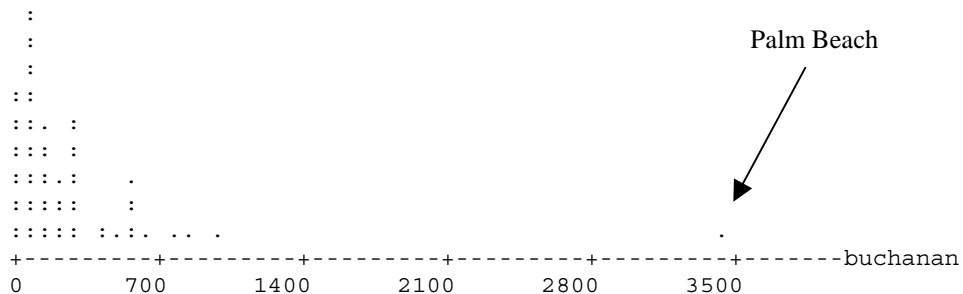


Figure 3 – Distribution Plot of County Vote Totals for Buchanan

One way to estimate to the outlier effect is to create an indicator variable for Palm Beach to serve as an independent variable in a regression which follows:

Regression Analysis: Buchanan versus Palm Beach

The regression equation is
 Buchanan = 211 + 3196 Palm Beach

Predictor	Coef	SE Coef	T	P
Constant	210.76	27.57	7.64	0.000
Palm Beach	3196.2	225.7	14.16	0.000

Figure 4 – Estimation of Outlier Effect of Palm Beach Buchanan Vote Total

If the null hypothesis that Palm Beach county is not unusual relative to other county vote counts is true, this model estimates that the 3196 votes for Buchanan over and above the mean level is associated with a test statistic of 14.16. The *p*-value for this statistic is approximately equal to 1.5E-21 or 1.5 in thousand billion squared. Such an astronomically small *p*-value, which also serves as an estimate of the outlier effect, will then be viewed as absolutely undeniable evidence that Buchanan’s vote total is an enormous outlier. The above analysis is the implicit analysis conducted when charts such as Figure 1 are provided as a support for an argument. The problem is that model diagnostics indicate clear violation of standard model assumptions. In particular, there is evidence that the random error component is not normally distributed as suggested by the following normal probability plot of the standardized residuals:

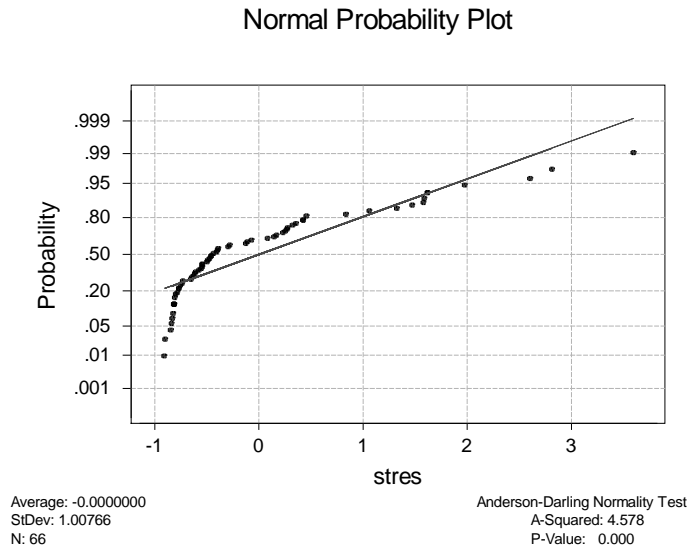


Figure 5 – Normality Probability Plot of Indicator Regression on Original Votes

Given the gross nonnormality, we have little confidence in the estimated measure of unusualness of the Buchanan's vote total in Palm Beach county. A viable remedy is to normalize the data. We have found that the logarithmic transformation does a satisfactory job. Below is a distributional plot of the Buchanan votes in log scale:

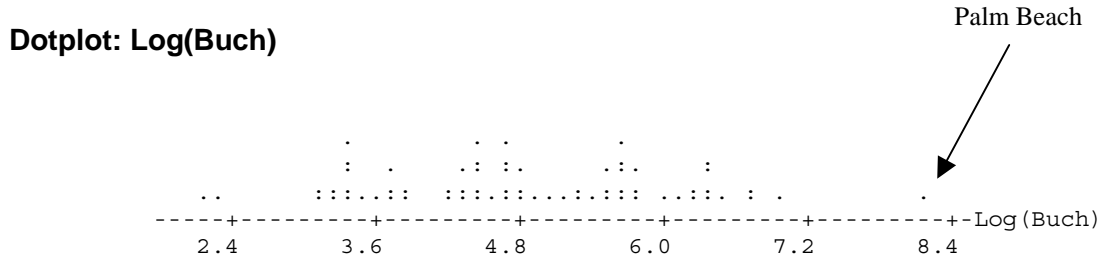


Figure 6 – Distribution Plot of County Vote Totals for Buchanan

Notice that the perceived unusualness of the Buchanan's Palm Beach vote total is dramatically less. Below is the corresponding regression analysis for the logged data:

Regression Analysis: Log(Buch) versus Palm Beach

The regression equation is
 $\text{Log(Buch)} = 4.78 + 3.35 \text{ Palm Beach}$

Predictor	Coef	SE Coef	T	P
Constant	4.7843	0.1400	34.18	0.000
Palm Beach	3.349	1.146	2.92	0.005

Figure 7 – Estimation of Outlier Effect of Palm Beach Buchanan Vote Total Based on Logged Data

Now, the *p*-value for statistic associated with the Palm Beach variable is 0.005 or 5 in 1000. The statistic is indeed significant at commonly accepted significance levels of 0.05 or 0.01 and thus it is reasonable to view the Buchanan's vote total in Palm Beach as a potential outlier deserving investigation into sources of causality such as a confusing ballot format. However, it is worthy to note that the level of unusualness (5 in 1000) is substantially less than the naïve analysis of the Buchanan vote total in original units. With respect to normality, below is a normality of the standardized residuals from the latest regression:

Normal Probability Plot

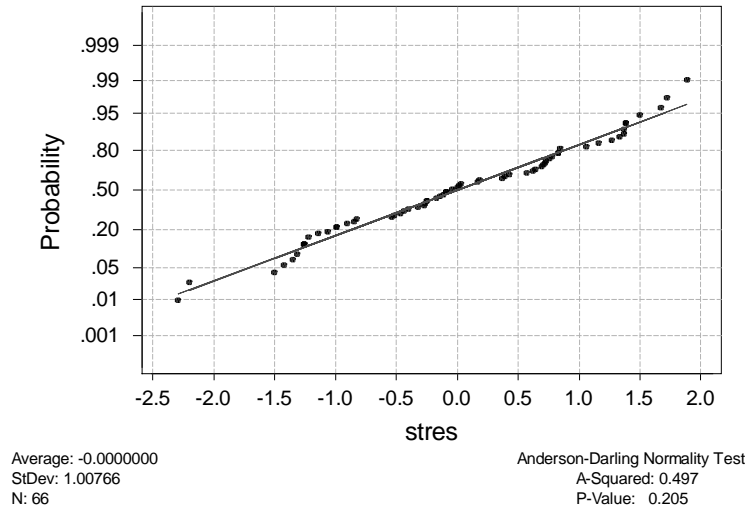


Figure 8 – Normality Probability Plot of Indicator Regression on Logged Votes

As opposed to the initial analysis, there is no evidence of substantial nonnormality. Hence, the analysis of the logged values provides a more appropriate perspective of the univariate analysis of the Buchanan vote totals.

Opportunity for Statistical Adjustment

The initial analysis of the Buchanan vote totals was conducted within a univariate framework, that is, an analysis of Buchanan votes in isolation. An interesting question is whether or not we can refine our analysis of Buchanan's Palm Beach vote total given information on voting totals for other candidates? Consider, for instance, a plot of Buchanan's vote totals versus Bush's vote totals:

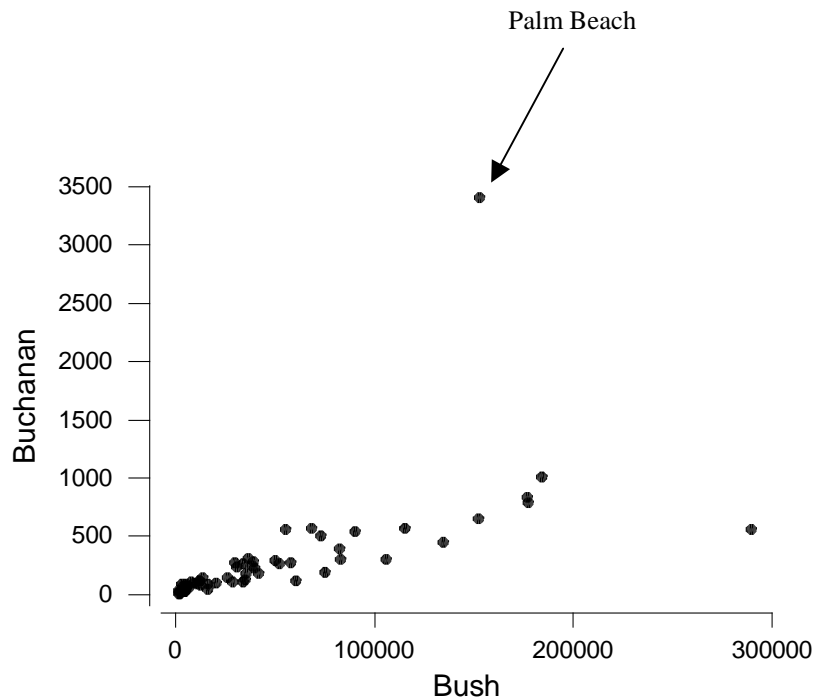


Figure 9 – Scatter Plot of Buchanan versus Bush Vote Totals

As with the univariate analysis in original units, Buchanan's vote total for Palm Beach stands out as a potential enormous outlier. We also see an apparent relationship underlying the variables in question, in particular, there appears to be a positively correlated relationship. Given the apparent relationship, one would be tempted to fit a regression line to the data. However, the scatter plot indicates a potential pitfall. Careful examination of the scatter plot shows nonconstant variance, that is, heteroscedasticity. The likely reason for heteroscedasticity in this application is due to varying population sizes of the counties. We expect smaller variability in smaller counties versus larger counties. Such heteroscedasticity has the potential of undermining the interpretation of the regression output and statements of prediction derived from the estimated equation. Fortunately, the nature of this heteroscedasticity can be remedied with a transformation which again turns out to be the logarithmic transformation.

Below is a scatter plot of Buchanan versus Bush vote totals in transformed units:

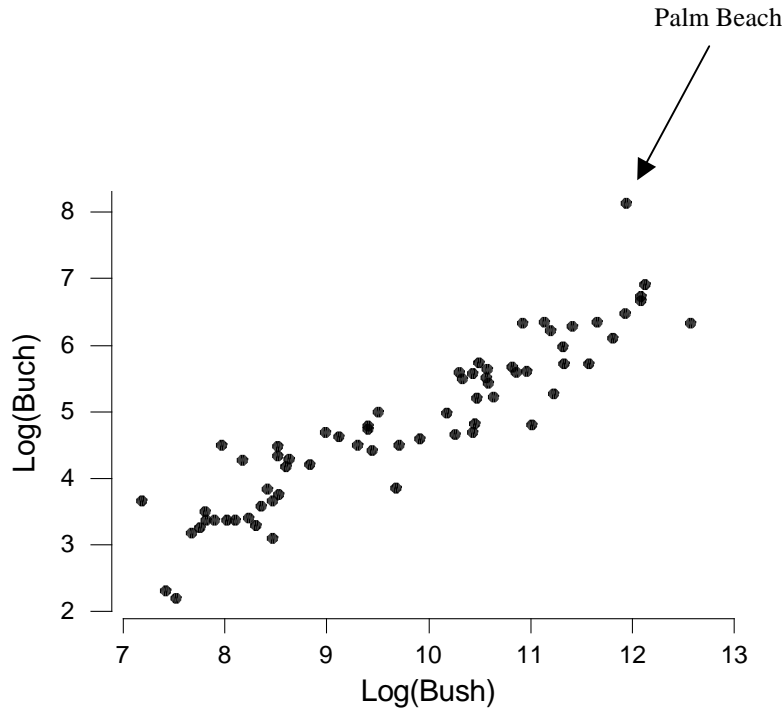


Figure 10 – Scatter Plot of Buchanan versus Bush Vote Totals in Logarithmic Scale

The impact of the transformation is obvious on a few fronts. First, the nonconstant variance is no longer a concern. Second, the relationship between the two variables is more clearly seen; in particular, in the form of a linear positive relationship. Third, the level of unusualness of the Buchanan vote total in Palm Beach is far less than the naïve perspective.

Implicitly, we are viewing “Log(Buch)” as the dependent variable in a subsequent regression analysis. With Figure 9, we are looking at only one potential predictor variable. For completeness, we should consider the vote totals for the other candidates: Gore, Nadar, Browne, Hagelin, Harris, McReynolds, Moorehead, and Phillips. Since some of the other candidate vote totals were 0 in some counties, we added 1 to the data for such candidates prior to taking the logarithmic transformation. Below are the backwards stepwise regression results:

Stepwise Regression: Log(Buch) versus Log(Gore), Log(Bush), ...

Backward elimination. Alpha-to-Remove: 0.1

Response is Log(Buch on 10 predictors, with N = 67

Step	1	2	3	4	5	6
Constant	-0.8153	-0.8125	-0.5881	-0.9811	-1.1126	-0.5863
Log(Gore)	0.10	0.09				
T-Value	0.51	0.52				
P-Value	0.611	0.602				
Log(Bush)	0.470	0.472	0.510	0.557	0.584	0.447
T-Value	2.68	2.80	3.39	3.83	4.00	5.14
P-Value	0.010	0.007	0.001	0.000	0.000	0.000
Log(Nader)	-0.29	-0.29	-0.25	-0.22	-0.14	
T-Value	-1.80	-1.82	-1.84	-1.69	-1.17	
P-Value	0.077	0.074	0.070	0.096	0.248	
Log(Browne)	0.18	0.18	0.17	0.17		
T-Value	1.56	1.58	1.54	1.52		
P-Value	0.123	0.120	0.130	0.133		
Log(Hagelin)	0.097	0.096	0.098			
T-Value	1.11	1.13	1.16			
P-Value	0.271	0.262	0.251			
Log(Harris)	-0.194	-0.194	-0.180	-0.174	-0.158	-0.161
T-Value	-3.04	-3.06	-3.15	-3.05	-2.78	-2.85
P-Value	0.004	0.003	0.003	0.003	0.007	0.006
Log(Mc)	-0.003					
T-Value	-0.04					
P-Value	0.965					
Log(Moore)	0.269	0.269	0.282	0.305	0.332	0.299
T-Value	2.58	2.61	2.84	3.13	3.43	3.22
P-Value	0.012	0.011	0.006	0.003	0.001	0.002
Log(Phillips)	0.195	0.195	0.187	0.179	0.201	0.210
T-Value	2.51	2.54	2.50	2.39	2.71	2.84
P-Value	0.015	0.014	0.015	0.020	0.009	0.006
Palm Beach	1.48	1.47	1.53	1.58	1.51	1.51
T-Value	3.36	3.59	3.91	4.04	3.84	3.82
P-Value	0.001	0.001	0.000	0.000	0.000	0.000
S	0.377	0.374	0.371	0.372	0.376	0.377
R-Sq	91.64	91.64	91.60	91.40	91.07	90.86
R-Sq(adj)	90.15	90.32	90.44	90.38	90.17	90.11
C-p	11.0	9.0	7.3	6.6	6.8	6.2

Figure 11 – Stepwise Regression Results

Based on the stepwise regression output, we estimate the following regression model:

Regression Analysis: Log(Buch) versus Log(Bush), Log(Harris), ...

The regression equation is

$$\text{Log(Buch)} = -0.586 + 0.447 \text{ Log(Bush)} - 0.161 \text{ Log(Harris)} + 0.299 \text{ Log(Moorehead)} + 0.210 \text{ Log(Phillips)} + 1.51 \text{ Palm Beach}$$

Predictor	Coef	SE Coef	T	P
Constant	-0.5863	0.6177	-0.95	0.346
Log(Bush)	0.44740	0.08702	5.14	0.000
Log(Harris)	-0.16131	0.05669	-2.85	0.006
Log(Moore)	0.29886	0.09285	3.22	0.002
Log(Phillips)	0.21033	0.07400	2.84	0.006
Palm Beach	1.5061	0.3942	3.82	0.000

S = 0.3775 R-Sq = 90.9% R-Sq(adj) = 90.1%

Figure 12 - Regression of Buchanan versus Selected Predictor Variables in Logarithmic Scale

Before proceeding with model interpretation, we conduct some model diagnostics:

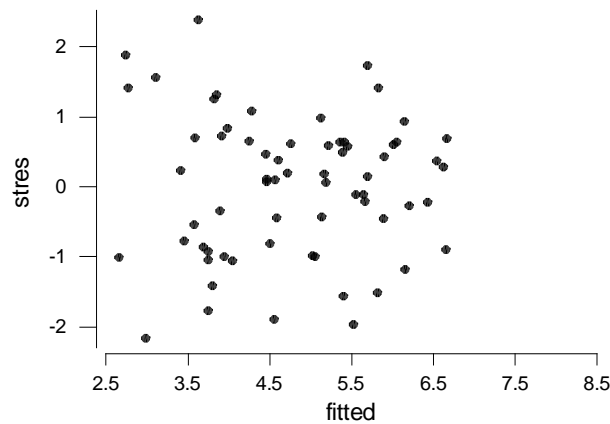


Figure 13 - Residuals vs. Fitted Values Plot

Normal Probability Plot

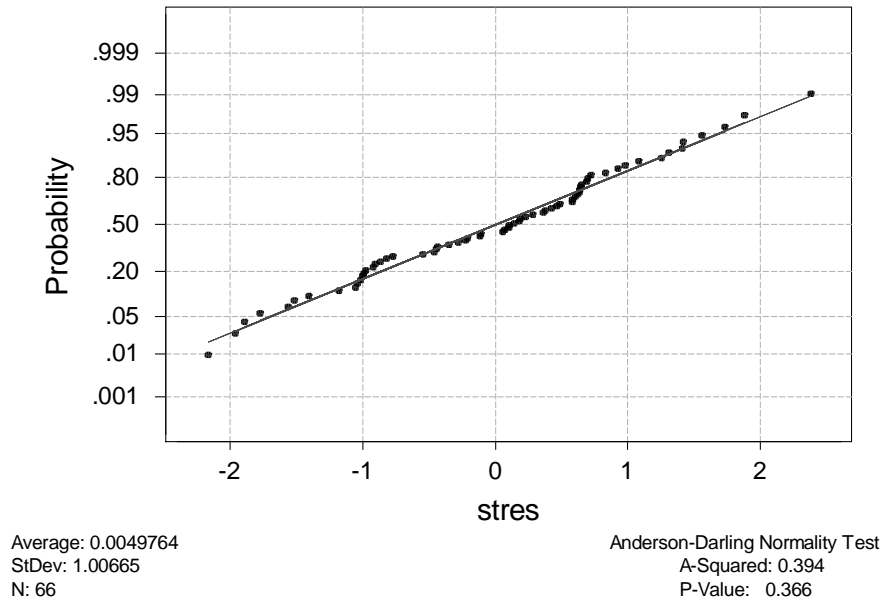


Figure 14 – Normality Probability Plot of Indicator Regression on Logged Votes

As we can see, both the randomness and normality checks of the residuals are satisfactory. All other standard diagnostics (not shown) are all satisfactory.

Note that the test statistics associated with Palm Beach is 3.82. This test statistic is associated with a p -value of 0.0003 or 3 in 10,000. Thus, the univariate-derived estimate of 5 in 1000 has been statistically adjusted to 3 in 10,000. Hence, the chances that Buchanan’s vote total in Palm Beach is due solely to chance is approximately 3 in 10,000.

Statistical Prediction of Buchanan Vote Total

As we can use the statistical model to assess the unusualness of the Buchanan vote total in Palm Beach, we can also use the model to predict his likely vote total in the county in absence of some underlying effects that render the observed total of 3407 votes unrepresentative of the general regularity of the data.

After removing the outlier effect from estimation, below we find a 99% prediction interval for the Buchanan’s vote total for Palm Beach:

Predicted Values for New Observations

New Obs	Fit	SE Fit	99.0% CI	99.0% PI
1	6.6274	0.1138	(6.3249, 6.9300)	(5.5792, 7.6757)

Figure 15 – 99% Prediction Interval for Logged Buchanan Vote Total

Translated into original units, a 99% prediction interval for Buchanan’s vote total in Palm Beach is (265, 2155). In this application, the lower prediction limit and even a point prediction are less useful for argumentative purposes. Since the focus of the alleged irregularity is that it resulted in excessive votes for Buchanan, such an argument is made stronger when we give benefit to the doubt by looking at the upper bound of the prediction interval. Thus, given Buchanan received 3,407 votes, an overly conservative estimate of the unexpected votes going in the direction of Buchanan is 1,252 (= 3,407 – 2,155). If we assume that all of these excess votes would have gone to Gore due to the ballot format, then it conservatively estimated that Gore was “shortchanged” by at least 1,252 votes.

One additional point is worthy of consideration. Buchanan’s vote total in Palm Beach for the 1996 primary election was close to expectation (actually slightly below) based on the similar modeling fitting procedures, that is, it was not an outlier. In an attempt to suggest other causal factors other than the ballot structure, Bush strategist Karl Rove noted that the Reform Party registration increased by 110% from 1996 to 2000. Even though such a “grassroots” factor might explain for some of the nearly 4 standard deviation outlier in the 2000 election, it is this author’s opinion that such a factor is unlikely to account for all the unexpected excessive votes.

For related information, see:

Professor Fox’s analysis: <http://faculty.fuqua.duke.edu/~cfox/Bio/election2000note.pdf>

Professor Adams’ and Fastnow’s analysis: <http://madison.hss.cmu.edu/>

Professor Brady’s analysis: <http://faculty.fuqua.duke.edu/~cfox/Bio/election2000note.pdf>

Professor Monroe's analysis: <http://www.indiana.edu/~playpol/pbmodel.pdf>