Chapter 2

THE STATISTICAL SIGNIFICANCE OF PALM BEACH COUNTY

David Andrews
Andrey Feuerverger

Abstract This paper emphasizes certain issues and problems that arise when a statistical analysis must be undertaken on complex and evolving data, under tight constraints of time. In such circumstances, it typically is not possible to develop extensive or problem-specific methodology, yet an answer may be required almost immediately, and must be correct, defensible, understandable, and carry impact. It must also be able to withstand the test of comparison with analyses yet to come.

We illustrate these points by presenting the background to, and an analysis of, the State of Florida results in the 7 November, 2000 U.S. Presidential elections with emphasis on Palm Beach County. The analysis we discuss was carried out in the days immediately following that election. The statistical evidence strongly suggested that the use of the 'butterfly' ballot in Palm Beach County had resulted in a significant number of votes having been counted for presidential candidate Pat Buchanan which had not so been intended. The design of the 'butterfly' ballot suggests that many of these votes had likely been intended for the Democratic candidate Al Gore. This confusion was sufficient to affect the overall outcome of the 2000 U.S. Presidential election, conferring the office to George W. Bush, and this result is statistically significant.

1. Mise en scène

On the evening of Tuesday November 7, 2000, the United States of America, along with much of the world, found itself in a state of suspended animation as a consequence of an inconclusive outcome to the U.S. federal election. While history will record the remarkable circumstances of that day, and its subsequent consequences, it should be borne in mind that at the core of these events and ensuing controversies, no discipline played a more substantive role than Statistics.

By now, some three years later, and almost on the eve of the November 2004 elections, many articles have begun to appear in statistical journals providing substantive analyses of data related to the November 2000 elections. In some of these articles, their authors develop new methodologies and explore their value for the analysis of such data. However, important as such work is, these articles do not (and, of course, do not claim) to capture the situation as it was "on the ground" among statisticians who became involved in the analysis of this evolving data set within its actual context of (punitively severe) "real time". In fact, to have been of any value at all in that election's context, and its legal battles in particular, it is hardly any exaggeration to say that such analyses had necessarily to be completed within a mere matter of days if not hours, and using less than the most comprehensive or most appropriate data possible. In addition, these statistical controversies arose, literally, without any prior warning whatever, and fell upon statisticians who, in many cases, already held previous commitments and could thus make only a portion of their time available for carrying out the required analyses. In short, this was a perfectly typical problem, under entirely typical circumstances, in the real-world arena of statistical practice.

The authors of this article have been involved in forecasting elections, and particularly election night forecasting for Canadian television networks where primary tasks have included 'declaring' candidates of ridings to be 'elected' using statistical algorithms designed to allow such 'calls' to be made as quickly and accurately as possible after the counting of ballots commences, as well as 'declaring' the winning party as quickly as possible, both under appropriately stringent accuracy requirements. It was therefore natural for us to become involved in statistical analysis of the November 2000 elections. In particular, within what were the very real constraints of time (and of our schedules), we analyzed the results of the vote counts in the State of Florida with particular reference to Palm Beach County

It is widely accepted now, and it was clear enough even then, that more than any other among the many peculiarities of that election, it was the use of the 'butterfly ballot' in Palm Beach County that cost Al Gore and the Democratic Party the Presidency of the United States, and it was for that reason that we had focused all our statistical energies on Palm Beach County. As is well known, the butterfly ballot used (designed by Palm Beach's subsequently much beleaguered supervisor of elections, Theresa LePore) was found to be confusing by many voters, and it is claimed that this confusion led some voters intending to vote for Al Gore to cast their vote instead for the Reform Party candidate, Pat Buchanan, whose name appeared interveningly and adjacent to

Gore's on the ballot; see Figure 2.1. It is perhaps an understatement to say that the demographics of Palm Beach County were not favourable to a candidate such as Pat Buchanan (even by his own admission); and indeed every published statistical analysis to date has reaffirmed that the total of 3407 votes that were counted for Buchanan in Palm Beach is out of all proportion to what is predicted when reasonable statistical models are applied to patterns occurring in the remaining 66 (out of the total of 67) Florida counties. Furthermore, such models generally indicate that the extent to which the Buchanan vote exceeded the numbers predicted by such models is significantly greater than the final number of 537 votes (out of a total of some 6 million votes) by which the Republicans ultimately carried the State of Florida, thus gaining its 25 electoral seats, and hence — by way of a resulting 4 vote margin of 271 to 267 in the Electoral College — the Presidency of the United States. Thus it seems very highly probable, though of course not conclusively provable, that had a different voting mechanism been used in Palm Beach County, the outcome of the 2000 election would have been different. To resort to linguistic coincidence, what happened in Palm Beach County exemplifies the 'butterfly effect', a term coined by chaos theorists concerned about such matters as the unforecastability of climate over long time horizons due to the chaotic instabilities of the governing physical dynamics and associated equations.

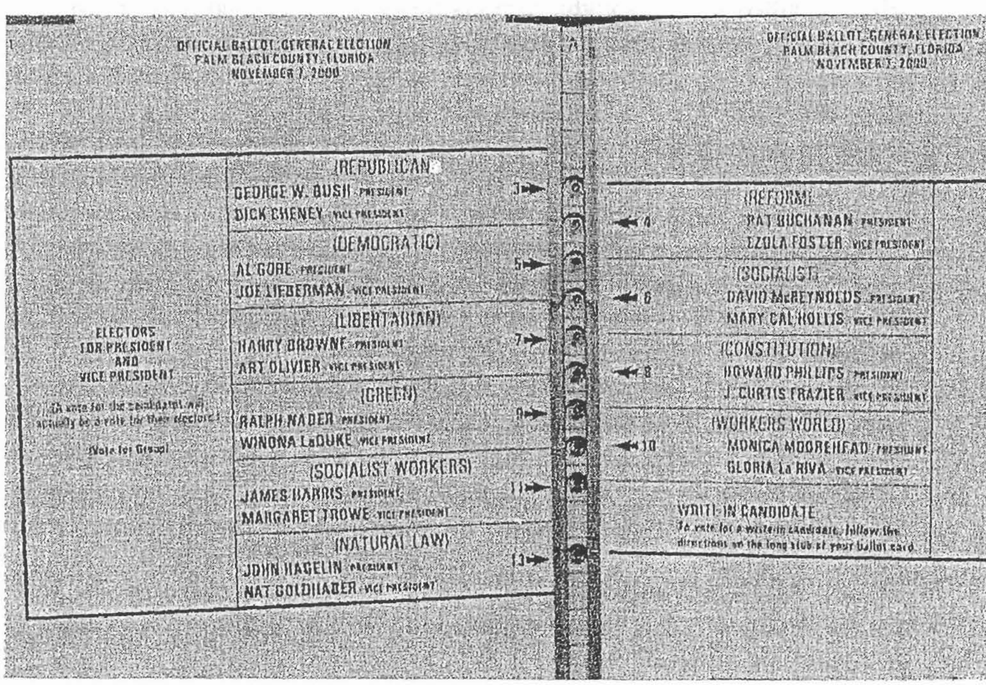


Figure 2.1. The Butterfly Ballot of Palm Beach County.

Of course, what is self-evident to statisticians does not always accord with U.S. constitutional law. Thus it came as a surprise to us, as to many others in the statistical community, although less so to specialists in law, that David Boies and others of the legal team representing Gore and the Democratic Party chose not to pursue Palm Beach's butterfly ballot in the courts. Instead, they pursued the matter of 'overcounts' and 'undercounts' (also known as the issue of 'completely unattached' versus 'hanging' or 'dimpled' chads) and fought to have hand recounts conducted in several counties — a legal battle which pitted statisticians Nicolas Hengartner for Gore against Laurentius Marias for Bush in the Leon County Circuit Court of Judge N. Sanders Sauls. Statistical analyses of state-wide counting issues lead to results which are less categorical and less dramatic; depending upon the exact standards set for counting, they mostly support Gore's position, but in certain instances they support the position of Bush. To quote from the Supreme Court's ultimate decision, "only in very close elections do such matters make a difference. ... upon due consideration ... it is obvious that the recount cannot be conducted in compliance with the requirement of equal protection and due process without substantial additional work"; see Supreme Court (2000). At the end, this matter travelled expeditiously from the Florida Eleventh Circuit Court to the Florida Supreme Court, and finally to the Supreme Court of the United States which — accepting what it called an "unsought responsibility" — rendered a controversial and deeply divided (five justices in favour, four against) verdict on December 12, 2000, bringing "this agonizingly long election process to a definitive conclusion". The mind-numbing ambiguity which gripped the world for more than a month was thus finally settled under strict protocols of American constitutional law, and George W. Bush was declared elected — the 43rd President of the United States.

Our purpose in writing this article is to emphasize certain issues that arise when a statistical analysis must be undertaken on a set of complex (and rapidly evolving) data under tight constraints of time. No political motive is intended by the authors, and we take no such positions here, other perhaps than supporting notions of reasonable democracy. Of course, it is well known that the *popular* vote went to Al Gore over George Bush by a margin exceeding 500,000 votes nation-wide, but that due to vagaries of the Electoral College system, Bush was the winning candidate. Such discrepancies from 'true' democracy are by no means uncommon. For example, in Canada, there have been many instances (both federally and provincially) in which the party which took leadership by virtue of winning the most ridings was not the party that obtained the highest popular vote. In fact, the allocation of campaign resources, as well as campaigning strategy, are typically decided by candidates after taking explicit account of election regulations. When viewed in the framework of the Electoral College system, the year 2000 U.S. elections were, in essence, a 'statistical tie' and, however one might define that term, such ties are not unduly rare. Even *exact* ties can occur in many types of elections and must then be broken using one of an imaginative variety of prescribed devices — such as tossing a coin, playing out a hand of poker by tying candidates, or by casting a tie-breaking vote by a prescribed official, to name three. In fact, in 'evenly contested' elections, the occurrence of ties is disproportionately high relative to the inverse of the number of voters. For if n is even and a fair coin is tossed n times, the probability of an exact tie is $2^{-n}\binom{n}{n/2} \approx \sqrt{2/\pi n}$ which scales as $1/\sqrt{n}$.

Returning to our purpose in writing this article, the first point we wish to make is that in typical problems of statistical consultation involving non-trivial data there frequently is not time to develop new and/or problem-specific methodology. An answer may be required immediately, and this generally necessitates using off-the-shelf solutions — more specifically, trusted and familiar statistical routines in familiar software packages. This is not to say there is no role for developing novel methods, but when time is the enemy, they cannot generally be relied upon. (Perhaps there is a modest role to play — in graduate courses on 'statistical consulting' — by the statistical equivalent of 'speed-chess'!)

Our second point is that while time may be of essence, and implementation of a 'demonstrably optimal' statistical procedure becomes an unattainable goal, it is nevertheless important for the resulting analysis to be essentially correct. By this we do not mean that statisticians (who after all are human) never make errors. In fact, even the term 'correct' here is ambiguous. The data here were not collected under controlled randomized conditions which lend precise structure to inference. There is no 'true model' underlying data of this type. The most that can be hoped for is to find models having a satisfactory degree of explanatory validity. In this process, experience, some understanding of the phenomena being investigated, as well as availability of an adequate number of experimental units are all invaluable commodities. Here being correct also entails not focusing categorically on only one plausible model, but in fact pursuing all, or as many reasonable models as is feasible. It also entails that a full account be rendered of the true uncertainties of conclusions which stem from uncertainties about the (non-existent) correct model. In this respect, selection of a 'best' model via an 'information' criterion does not provide full information to 'consumers' of the analysis. We posit that it is better, in such circumstances, to present as wide a variety of reasonably conceived analyses as possible, and allow variations among their results to speak for the true uncertainties inherent in the problem overall.

Indeed, such a collection of analyses contributes to the court of public knowledge and opinion without trivializing or simplifying to the point of what, strictly speaking, becomes untruth. They represent an admission of the real uncertainty which underlies the problem — an admission that there is no single right answer. Such an approach also recognizes that additional analyses will later be carried out by others, using ever more extensive data and more sophisticated models. Hence the analysis will be subject to the scrutiny of future light; there will be a sequence of tests to pass, and any conclusions of the analysis will be subjected to consistency checks against all analyses yet to come.

Our third point is that statistical analyses must sometimes be as straightforward and easy to understand as possible. Statistics is not easy for non-statisticians, and except in instances all too rare, lawyers and judges are not statisticians. Nevertheless they are amongst the decision-making professionals to whom the statistical ideas, analyses, and results need to be explained; and it is often they who, on the basis of their understanding, must make substantive decisions. That some level of complexity is necessarily inherent in an analysis of this type cannot be avoided, but one can avoid going farther than what is required for capturing the essential conclusions, at least insofar as materials 'presented' to such professionals is concerned. In this respect, understandable graphical displays that carry impact are of utmost importance.

This paper contains three additional sections. Section 2 discusses the data set whose analysis is present here. The actual analysis is carried out in Section 3. It is a matter of principle that the analysis we present in Section 3 is exactly as carried out by us in the days immediately after November 7, 2000. We have meticulously avoided embellishing or improving upon our original results or adding any follow-up to that work. It thus appears here 'warts and all' in its entirety. The only exceptions are that the explanatory text has been edited for clarity and context, and that fewer displays are included here in order to keep the length of this paper reasonable and avoid the repetitive display of graphs all having very similar appearance. A significant aspect of the work reported in Section 3 revolved around designing effective, convincing, and statistically accurate graphical displays. Finally, in Section 4, we cite some references to later work that others have done using more substantive data and more complex statistical models. We compare our result to theirs and thereby apply the 'test of time' to the work we had done. Some concluding remarks are also given there.

2. Description of the data

Immediately following recognition of the crucial importance of the vote in Florida to the outcome of the U.S. federal elections of 7th November, 2000, the ABC Television Network began posting, on its website, a county by county current tabulation of the votes for many of the candidates in that state. Although, in developing our analyses, we first worked with earlier posted versions of these counts, our final analyses were ultimately 'frozen' and based on the data set as posted on the ABC website at 9 am on Sunday, November 12th. This data reflects the votes as tabulated (after the first "recount") of the Florida State ballots. Of course, the nature of the statistical methods we used are such that our results were not expected to vary in any important way under the small subsequent changes to this data that were anticipated to occur under the repetitive recounting of the votes.

The downloaded file was processed to make it suitable for reading into the statistical analysis package S-Plus which was used for our analyses. (See Becker, Chambers and Wilks, 1988.) This processed version of the data file is exhibited as Table 2.1. That table gives a breakdown by county, for all 67 Florida counties, of the number of votes counted (after first "recount") for the four presidential candidates George W. Bush, Al Gore, (Reform Party candidate) Pat Buchanan, and (Green Party candidate) Ralph Nader. The file also gives the number of votes cast for three Senatorial candidates, Bill Nelson, Bill McCollum, and Willie Logan. These Senatorial candidates are respectively Democrat, Republican, and an Independent with central leanings. Several other candidates also ran in both the presidential and the senatorial races but received extremely small vote counts; these were not reported on the ABC website and were not included in our analysis.

3. The statistical analysis

The analysis we present was based on the seven data variables appearing in Table 2.1, namely the raw vote counts for seven candidates — four presidential, and three senatorial. In addition to these seven variables, two variables can be constructed which are each proxies for 'county size', namely the total number of votes counted for these four presidential candidates in each county, and the total number of votes counted for the three senatorial candidates. Of course, these two county size proxy variables are in nearly perfect correlation. In our analyses, we used the first of these as our actual proxy for 'county size'. We also defined additional variables representing the proportions of votes cast for each of the seven candidates. Presidential candidate proportions were

Table 2.1. Florida Election Results, Raw Data, by County; First Recount.

This Data From ABC Website as at 9am Sunday November 12, 2000

	PRESIDENT				SENATE		
COUNTY	GORE	BUSH	BUCHANAN	NADER	NELSON	MCCOLLUM	LOGAN
ALACHUA	47365	34124	262	3215	49005	31003	1733
BAKER	2392	5610	73	53	3104	4578	50
BAY	18850	38637	248	828	22914	33901	358
BRADFORD	3075	5414	65	84	4118	4697	92
BREVARD	97318	115185	570	4470	112255	98813	2304
BROWARD	386561	177323	789	7099	377081	174980	5974
CALHOUN	2155	2873	90	39	2809	2055	31
CHARLOTTE	29645	35426	182	1462	28947	37026	746
CITRUS	25525	29766	270	1379	27566	27025	947
CLAY	14632	41736	186	562	16094	39054	561
COLLIER	29918	60433	122	1400	28207	60508	822
COLUMBIA	7047	10964	89	258	8942	9031	255
DADE	328764	289492	560	5352	304878	264801	11796
DE.SOTO	3320	4256	36	157	3593	3736	127
DIXIE	1826	2697	29	75	2450	2007	36
DUVAL	107864	152098	652	2757	121805	145930	3214
ESCAMBIA	40943	73017	502	1727	45907	67607	771
FLAGLER	13897	12613	83	435	13980	11988	293
FRANKLIN	2046	2454	33	85	2498	2018	31
GADSDEN.	9735	4767	39	139	10838	4306	553
GILCHRIST	1910	3300	29	97	2558	2561	46
GLADES	1442	1841	9	56	1649	1620	67
GULF	2397	3550	71	86	3393	2739	42
HAMILTON	1722	2146	23	37	2172	1722	44
HARDEE	2339	3765	30	75	2972	3051	88
HENDRY	3240	4747	22	103	3760	4513	116
HERNANDO	32644	30646	242	1501	32916	29099	1106
HIGHLANDS	14167	20206	127	545	11630	1434	296
HILLSBOROUGH	169557	180760	847	7490	176667	15919	5928
HOLMES	2177	5011	76	91	3201	3252	31
INDIAN.RIVER	19768	28635	105	950	21050	27223	545
JACKSON	6868	9138	102	138	8648	7529	168
JEFFERSON	3041	2478	29	76	3513	2101	139
LAFAYETTE	789	1670	10	26	1299	1175	28
LAKE	36571	50010	289	1459	40726	47333	800
LEE	73560	106141	305	3587	69308	107824	2610
LEON	61425	39053	282	1932	61728	35468	2914
LEVY	5398	6858	67	284	6654	5796	142
LIBERTY	1017	1317	39	19	1375	948	33
MADISON	3014	3038	29	54	3528	2492	69
MANATEE	49177	57952	272	2489	51400	54104	1608
MARION	44665	55141	563	1809	48947	50896	962
MARTIN	26620	33970	108	1075	21826	28065	456
MONROE	16483	16059	47	1090	16588	14778	455
NASSAU	6879	16280	90	253	8489	14413	198
OKALOOSA	16948	52093	267	985	18667	48897	421
OKEECHOBEE	4588	5057	43	131	5320	4410	244
ORANGE	140220	134517	446	3881	140897	119652	3014
OSCEOLA	28181	26212	145	732	29722	23856	624
PALM.BEACH	269696	152954	3407	5564	269835	154528	4385
PASCO	69564	68582	570	3392	73338	63081	2172
PINELLAS	200629	184823	1010	9986	207974	168508	7333

Table 2.1. (cont.) Florida Election Results, Raw Data, by County; First Recount.

This Data From ABC Website as at 9am Sunday November 12, 2000

	PRESIDENT				SENATE		
COUNTY	GORE	BUSH	BUCHANAN	NADER	NELSON	MCCOLLUM	LOGAN
POLK	75197	90196	538	2060	81159	79837	2990
PUTNAM	12102	13447	148	377	13124	11876	304
SANTA.ROSA	12802	36274	311	724	14504	33650	222
SARASOTA	72853	83100	305	4069	71434	80641	2094
SEMINOLE	59174	75677	194	1940	63037	69865	1377
ST.JOHNS	19502	39546	229	1217	20558	37275	930
ST.LUCIE	41559	34705	124	1368	41082	33050	1065
SUMTER	9637	12127	114	306	10271	11384	231
SUWANNEE	4075	8006	108	180	5544	6797	130
TAYLOR	2649	4056	27	59	3431	3386	105
UNION	1407	2332	29	32	1560	11	23
VOLUSIA	97063	82214	396	2436	99135	76033	2554
WAKULLA	3838	4512	46	149	4496	3778	131
WALTON	5642	12182	120	265	6842	10295	115
WASHINGTON	2798	4994	88	93	4065	3661	60

obtained by dividing their vote counts by the first of the vote totals mentioned, while senatorial candidate proportions were obtained by dividing their vote counts by the second of the totals. Finally, we also considered logarithmic transformations in each of these variables in order to obtain variables that more nearly satisfy the assumptions and requirements of the statistical methods we used. Such requirements typically include approximate linearity, constancy of variance, and normality (or at least approximate symmetry, together with tails that are not too heavy) of the model-generated error terms. For example, the logarithms of proportions (especially when small) often have more uniform variability over their range of values than do raw proportions; likewise, logarithms of raw vote counts often give more normally distributed residuals than untransformed values. Some might argue for logits of proportions, but here these are almost identical to logarithms which are more widely understood. Furthermore, positive variables are often described by additive models based on their logarithms. Others might argue for square roots of proportions because this tends to stabilize variance. However, on balance, we felt that linearity was more important here. Discussions on such matters may be found, for example, in Belsley (1980), Cook and Weisberg (1999), Draper and Smith (1998), Mosteller and Tukey (1977), or Myers (1990).

Our choice for dependent variable was governed by the purpose of the analysis, namely, to determine the number of votes for Buchanan attributable to the nature of the ballot in Palm Beach. This is equivalent to estimating the proportion of such votes. However the number of votes for any candidate in a county is roughly proportional to the number of electors in the county. Because counties vary considerably in the number of electors, the variation in votes is considerably higher than the variation in proportions. Following the general principle of removing known sources of variation, it is more effective for estimating the overvote to use proportions rather than votes.

Voter preferences are determined by many personal factors associated with age, education, income, party affiliation, etc., and by many community factors associated with employment, density, and so on. While data on these were not readily available to us, we do have the effect of voter preferences as shown by votes cast for the other candidates. The proportions of votes for these other candidates act as surrogates for underlying determining variables. For this reason explanatory models were constructed from the proportions of votes for the other candidates. In addition, community size may be associated with other important factors. (Small communities may be more closely knit, and may inspire greater voter participation, for example.) To voter proportions, we added total vote as a surrogate for such variables. Using these variables, we found that the proportion of votes for Buchanan in a county could be reliably predicted from the pattern of votes for other candidates in the counties.

On the basis of these variables, we analyzed a large number of regression models fitted to this data. Our rationale for examining many models was based firstly on our expectation that such data would be subjected to widespread analysis by many others. We therefore chose to report the results of many reasonably selected analyses in the hope of assisting those who would seek to form their own conclusions. Secondly, fitting explanatory variables explains variation leaving less variation in the estimation of error. Hence methods of variable selection can result in error estimates that are biased downward and may thus affect the validity of prediction intervals. (Indeed, this leads to questions of research caliber.) Partly for this reason, we considered instead sets of possible explanatory variables, fitting a sequence of increasingly complex models. We contend that the resulting collection of error estimates provides more information than would the results from variable selection.

In our reporting, all regression analyses were based on a single 'dependent' variable, namely the logarithm of the proportion of votes (county by county) cast for Pat Buchanan. Regression models were fitted for this dependent variable against many plausible combinations of predictor variables. Given a regression model which fits such data in a statistically appropriate way, an estimated vote count for Buchanan can be obtained from a model-based estimate of the logarithm of the Buchanan proportion, by just exponentiating (to obtain the estimated proportion) and then multiplying by the county vote total (to obtain the estimated Buchanan vote count). Our regression analyses were all based on first removing Palm Beach County from the statistical fitting. Since Florida has 67 counties, our regression fits are all based on 66 (multivariate) observations. For each model we fitted in this way, we obtained the model-based prediction of the number of votes for Buchanan in Palm Beach County. This estimate was then compared to the actual number of 3407 votes counted for Buchanan in Palm Beach. It is the difference between these two numbers that is of essence and the focus of our analysis. To assess the statistical significance of this difference, we obtained (for each model) a prediction interval for the log Buchanan proportion, and transformed this (in the manner indicated above) into a prediction interval for the Buchanan vote count in Palm Beach. These regressions were all based on unweighted, ordinary least squares, and classical statistical prediction intervals having confidence levels of 95%. Appropriate checks were conducted, during the course of this work, to assure that the key requirements for regression analysis were being adequately met.

Three series of regression analyses were carried out. The 'series A' regressions were each based on a single predictor variable, while the 'series B' and 'series C' regressions were each based on two and three predictor variables, respectively. There are a total of 7, 6, and 3 regressions in the 'A', 'B', and 'C' series and they are summarized in Tables 2.2, 2.3, and 2.4 (discussed further below.) The key findings from these analyses were then transformed into a carefully crafted series of figures which we correspondingly labelled as Graphs A1–A7, B1–B6, and C1–C3; these figures are all very similar in overall impact and appearance, and to save space only some of them are reproduced here. However it should be borne in mind that the repetitively similar nature of all 16 figures in itself carries useful information. A substantial amount of our efforts was spent on the design of these displays since we had anticipated that they might be used by non-statisticians, and thus knew that they had to accurately, effectively, and unbiasedly carry the statistically substantive information. The particular displays reproduced here include Graphs A2, A6, B2, B4, C2 and C3. See Figures 2.2–2.4. (The criteria for their inclusion here is explained further below.)

Each of these displays plots the actual Buchanan vote counts on the vertical axis, against the predicted number (i.e. the model-based 'fitted' values) for the Buchanan vote counts on the horizontal axis, for all 67 counties. (The regression fits themselves, of course, excluded Palm Beach.) Note, however, that a logarithmic scale was used for the horizontal axis. The 67 bullet points on each graph give the actual versus predicted Buchanan vote counts in the counties, and the smooth solid curve represents the regression fit to the data, i.e. the model-based predicted values; it is just the identity line. (Except for the effect of the logarithmic horizontal scale, this curve would be a straight line with height equal to the horizontal axis coordinate at each point.) The 67 vertical line segments represent the model-based *prediction* intervals, each having confidence level 95%, for the Buchanan vote count for each of the counties. As may be discerned, the tops and bottoms of these intervals do not lie along any particular patterns; this reflects the effects of the predictor variables in the models which are not otherwise visible in these graphs. Nevertheless the lengths of the prediction intervals do increase in a reasonably smooth manner with the fitted vote counts for Buchanan. Note that a high proportion of the actual votes for Buchanan (other, of course, than Palm Beach county) fall inside their respective prediction bands. Other indications of the quality of these fits may be obtained by examining typical residual plots; as a general rule, the quality of these fits are all very good. We also experimented with weighted regressions using weights related to the county vote totals (e.g., inverse total votes, or inverse square root of total votes). These results all corroborate the findings presented, but are not included here.

On each of these plots, two arrows are used to help identify the bullet point for Palm Beach county. As is evident from examining the plots, the data point for Palm Beach County is a statistically highly significant outlier in every one of these models. It is possible to determine an approximate two-sided p-value for the Palm Beach County outlier through the following simple device: the prediction interval for Palm Beach County can be enlarged by increasing its significance level until the upper end of the interval just touches the Palm Beach County bullet. The difference between unity and the confidence level of that interval gives the two-sided p-value. (If one were to adopt the viewpoint that the nature of the butterfly ballot could not lead to an *undercount* for the Buchanan vote, then a one-sided p-value could be obtained by taking one half of the two-sided p-value.) Significance levels so computed do not account for any selection effect that might have led to the study of Palm Beach County. However, allowing for such selection bias at most multiplies the significance levels by a factor of 67, and this would still leave our results statistically significant. In this respect, it is also worth noting that Florida had been selected by others for focus because its results were so close, while the selection of counties as the natural unit of aggregation was also made by others.

Table 2.2. Series "A" Regression Results — One Predictor Variable.

Regression Model	Predicted Buchanan	Estimated Overcount	95% Pred. Interval	Minimum Overcount
A1: $\mathcal{L}$Tot	760	2647	(250, 1680)	1727
A2: $\mathcal{L}$Gore	686	2721	(229, 1640)	1767
A3: $\mathcal{L}$Bush	867	2540	(282, 2182)	1225
A4: $\mathcal{L}$Nader	852	2555	(290, 2064)	1343
A5: $\mathcal{L}$Nelson	690	2717	(221, 1695)	1712
A6: $\mathcal{L}$McColl	935	2472	(296, 2423)	984
A7: $\mathcal{L}$Logan	797	2610	(292, 1812)	1595

The regression series A1–A7, B1–B6, and C1–C3 are also summarized in Tables 2.2, 2.3 and 2.4. Table 2.2 gives the results of seven regressions for the dependent variable '$\mathcal{L}$Buchanan$\mathcal{P}$' (i.e. the logarithm of the proportion of votes cast for Buchanan — see below) each against a single predictor variable, with an intercept term included. In model A1, the predictor variable is the logarithm of the total number of votes cast for all presidential candidates, while models A2–A7 are based on the same 'independent' variable, but modelled, respectively, against the logarithm of the total number of votes for each of the six other candidates (presidential as well as senatorial) appearing in the Table 2.2 (and in the order given in that table): Gore, Bush, Nader, Nelson, McCollum, and Logan. In Table 2.2 (as in Tables 2.3 and 2.4), the first column lists the predictor variables defining the regression. The notations 'Tot', 'Gore', 'Bush', 'Buchanan', 'Nader', 'Nelson', 'McColl', and 'Logan' represent the total (presidential) vote count, and the 7 individual candidate vote counts. When a vote proportion instead of a vote total is used, we indicate this by a suffix $\mathcal{P}$, and when a logarithm is taken we indicate this by a prefix $\mathcal{L}$. Tables 2.3 and 2.4 give the results, respectively, of our regression analyses each using two and three predictor variables, but always the same dependent variable, namely, $\mathcal{L}$buchanan$\mathcal{P}$. Thus, for example, the row labelled as 'C2' in Table 2.4 gives the results for the regression of $\mathcal{L}$buchanan$\mathcal{P}$ against the predictor variables $\mathcal{L}$Tot, Gore$\mathcal{P}$, and Nader$\mathcal{P}$, i.e., against the logarithm of the total vote counts, and the vote proportions for Gore and for Nader. The remaining columns in Tables 2.2, 2.3, and 2.4 respectively give the model-based vote projections for Buchanan in Palm Beach County, the estimated overcount of the vote for Buchanan (obtained by taking the difference between the projection in column 2 and Buchanan's actual count of 3407 votes), 95% prediction intervals for the Buchanan vote, and in the last column, the minimum overcount for Buchanan consistent with these prediction intervals. It may be seen that none of our point predictions for Buchanan's vote exceeds 1000, with most lying well below this value, and correspondingly, the estimated overcounts all exceed 2400. Further, our *minimum* overcounts are all well above 1000, except for one instance which occurs in one of the single-predictor regressions. At final tally, the official plurality by which the State of Florida was won by Bush amounted to 537 votes. As may be seen, the minimum overcount exceeds this value by a substantial (and significant) margin in every one of these regression models.

Table 2.3. Series "B" Regression Results — Two Predictor Variables.

Regression Model	Predicted Buchanan	Estimated Overcount	95% Pred. Interval	Minimum Overcount
B1: $\mathcal{L}$Gore+$\mathcal{L}$Bush	622	2785	(195, 1503)	1904
B2: $\mathcal{L}$Tot+$\mathcal{L}$Logan$\mathcal{P}$	850	2557	(303, 1939)	1468
B3: $\mathcal{L}$Tot+$\mathcal{L}$Logan	839	2568	(298, 1920)	1487
B4: $\mathcal{L}$Tot+Bush$\mathcal{P}$	616	2791	(192, 1496)	1911
B5: $\mathcal{L}$Nelson+$\mathcal{L}$McColl	697	2710	(215, 1738)	1669
B6: $\mathcal{L}$Logan+$\mathcal{L}$Nader	798	2609	(291, 1824)	1583

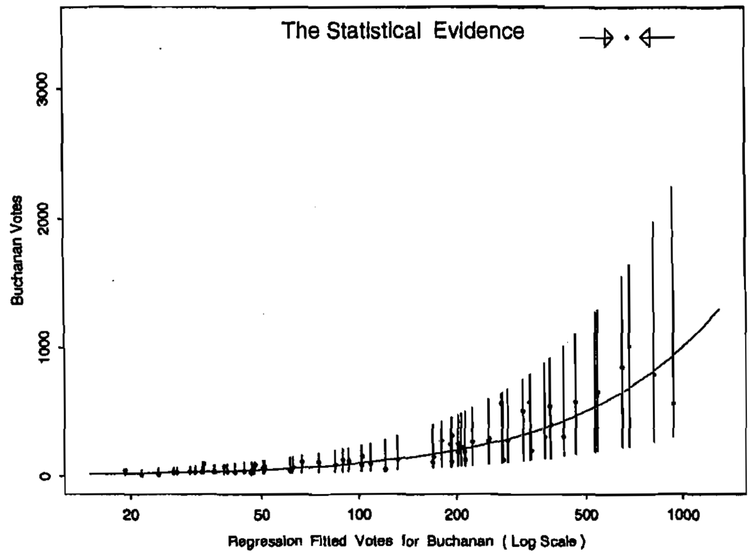
Table 2.4. Series "C" Regression Results — Three Predictor Variables.

Regression Model	Predicted Buchanan	Estimated Overcount	95% Pred. Interval	Minimum Overcount
C1: $\mathcal{L}$Tot+$\mathcal{L}$Gore$\mathcal{P}$+$\mathcal{L}$Nader$\mathcal{P}$	663	2744	(201, 1621)	1786
C2: $\mathcal{L}$Tot+Gore$\mathcal{P}$+Nader$\mathcal{P}$	631	2776	(185, 1555)	1852
C3: $\mathcal{L}$Tot+Logan$\mathcal{P}$+Nader$\mathcal{P}$	860	2547	(285, 2036)	1371

Finally, we indicate the criteria by which the Graphs A2, A6, B2, B4, C2 and C3 were selected for inclusion here. The extent to which a figure is visually and statistically 'persuasive' is governed by the distance of the Palm Beach County bullet point from its model predicted value, as measured in units of the width of its prediction band. Thus, within each of the series A, B and C graphs, we selected their most and least persuasive representatives for inclusion here. (Note that the same graphs would have been selected had the criterion been based instead on the minimum overcount values.)

To check that the use of linear models had not hidden important information, tree regressions were fitted to both vote counts and vote proportions. In both cases, the known dependence of variance on size was approximated by weighting these tree regressions by the inverse of the total vote counts. Tree regression fits step functions based on explanatory variables and is not restricted to any assumed linearity. In effect, it groups counties with the same fitted values. See Breiman et al. (1984) for details concerning this procedure.

Graph A2: Palm Beach County Vote is Out of Line



Graph A6: Palm Beach County Vote is Out of Line

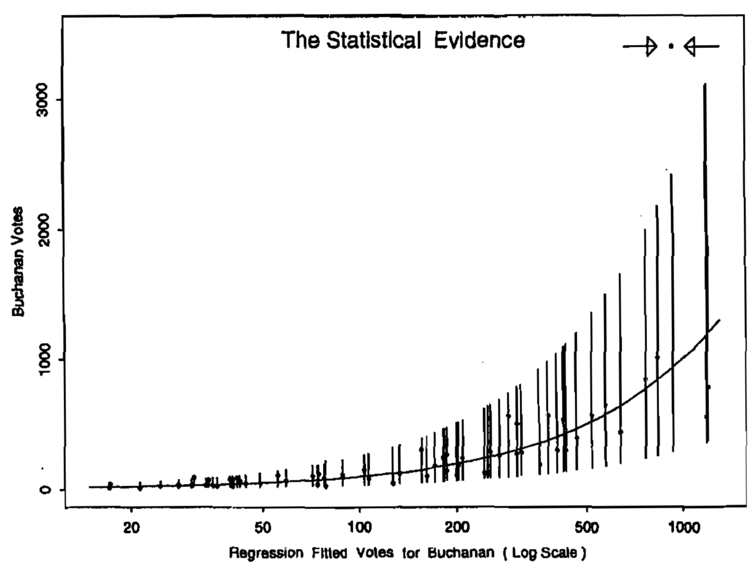
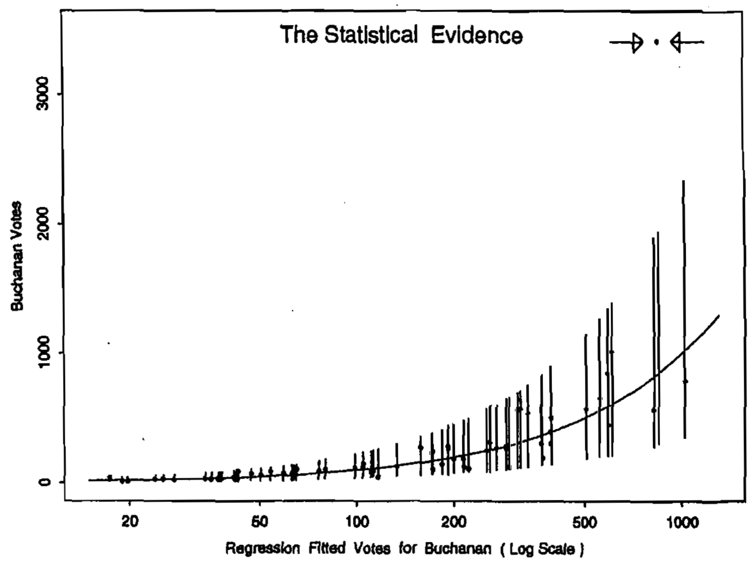


Figure 2.2. Two results of regression analysis (with Palm Beach omitted) for log Buchanan vote percentages, each against a single predictor variable (see text). Intervals shown are level 95% prediction bands for Buchanan's vote in all 67 counties.

Graph B2: Palm Beach County Vote is Out of Line



Graph B4: Palm Beach County Vote is Out of Line

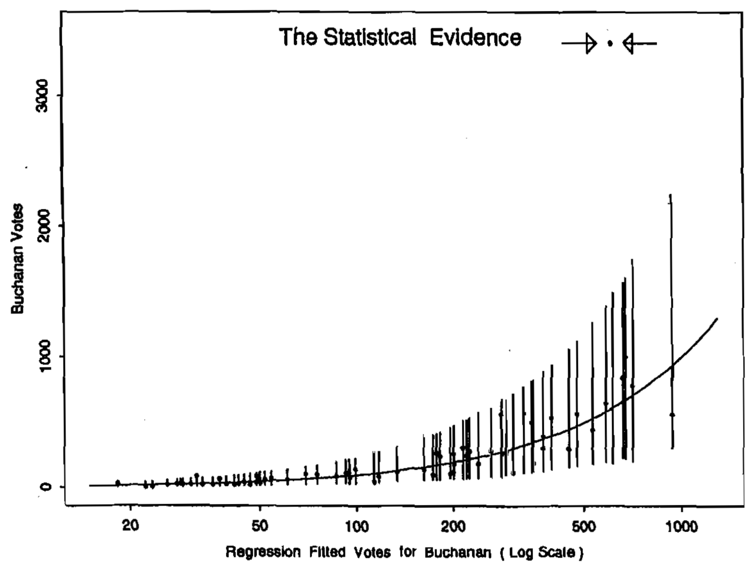
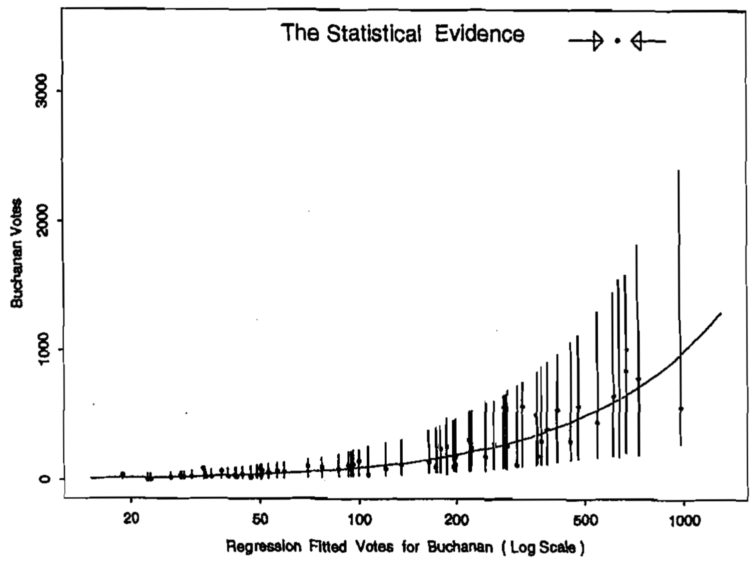


Figure 2.3. Two results of regression analysis (with Palm Beach omitted) for log Buchanan vote percentages, each against two predictor variables (see text). Intervals shown are level 95% prediction bands for Buchanan's vote in all 67 counties.

Graph C2: Palm Beach County Vote is Out of Line



Graph C3: Palm Beach County Vote is Out of Line

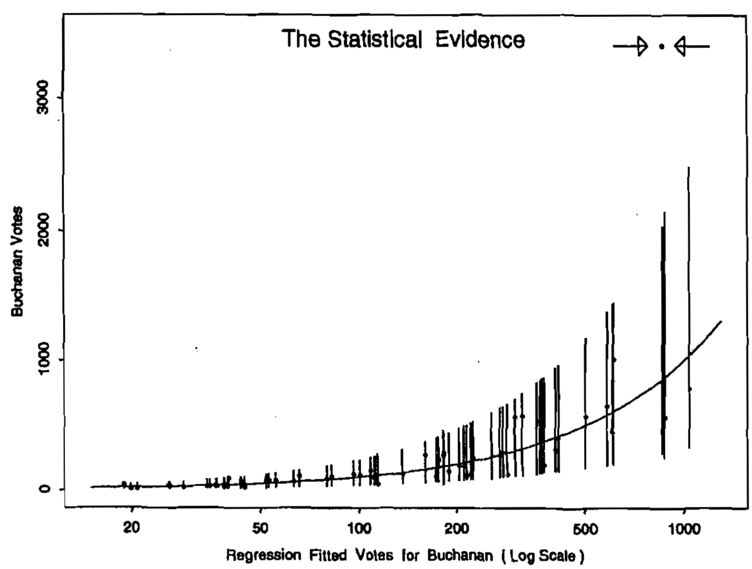


Figure 2.4. Two results of regression analysis (with Palm Beach omitted) for log Buchanan vote percentages, each against three predictor variables (see text). Intervals shown are level 95% prediction bands for Buchanan's vote in all 67 counties.

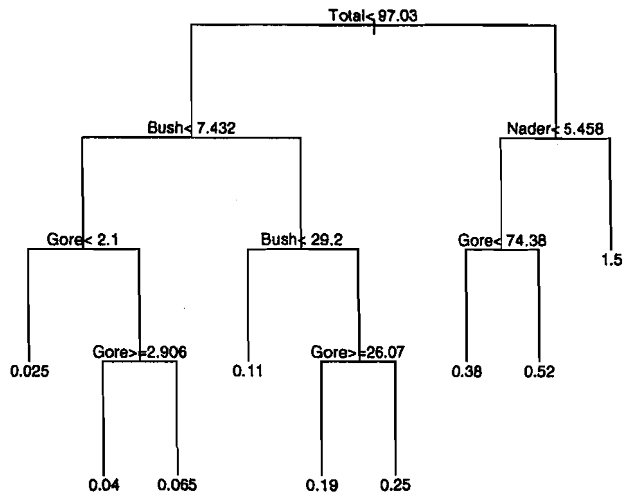
Graphs D1–D6 in Figures 2.5–2.7 show tree regressions using the data from all 67 counties, and predictor variables as indicated in the graph headings. Each plot gives either the average votes or the average proportions for Buchanan for the counties within the identified groups. The trees are displayed together with the splitting values of the explanatory variables. The groups of counties are positioned in the display so that average votes per county increases as one goes from left to right. Examination of the groups to which Palm Beach is assigned shows that even after extensive fitting, Palm Beach stands out with far higher values than other counties in those groups. For example, the plot in Graph D1 identifies one group at the right. This group is formed by those counties with total vote exceeding 97030, and Nader vote exceeding 5458. There are, in fact, 4 such counties. One is Palm Beach which 'gave' 3407 votes to Buchanan, and 3 others whose average Buchanan vote is less than a third of this value. The Buchanan vote in Palm Beach is thus seen to be extreme relative to the 3 other counties having similar voting patters. This unusual behaviour occurs in all of the 6 tree models fitted. The tree regressions show that the apparent unusual nature of Palm Beach County is not an artifact of the nature of a linear model.

While we knew then that further analysis on this (and expanded) data sets should (and no doubt would) be carried out, it was clear — even at that early stage for such analyses — that the votes as counted in Palm Beach County, with very high likelihood, did not reflect the intended votes of those voters. Every reasonable statistical model we had examined indicated a statistically significant overcount for Buchanan in Palm Beach County. It was our conclusion at the time, based on these analyses, that the explanation of this singular 'outlier' cannot rest on factors to which all counties are subject to a greater or lesser degree, but rather must rest on some aspect of the voting peculiar to Palm Beach County. It is widely believed, and we would now concur, that many of the overcounted Buchanan votes were in fact intended for Gore, and were likely misdirected owing to voter confusion with the butterfly ballot that was used in Palm Beach County.

4. Concluding remarks

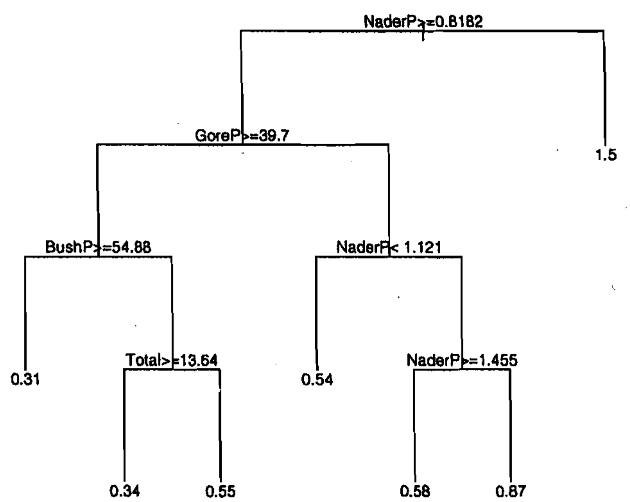
As indicated in the introduction, the analyses as presented in the previous section were carried out in the days immediately following the November 2000 elections. Since that time, many other analyses have, as expected, appeared. As examples, we refer the reader to Agresti and Presnell (2001, 2002), Hansen (2003), Smith (2002), Wand et al. (2001), and the references therein.

Graph D1: Classification and Regression Tree of Buchanan Vote Fitted to Total Vote and Votes for Gore, Bush and Nader Weighted by Inverse Total Votes (in 1000's).



End nodes give average Buchanan votes for counties.

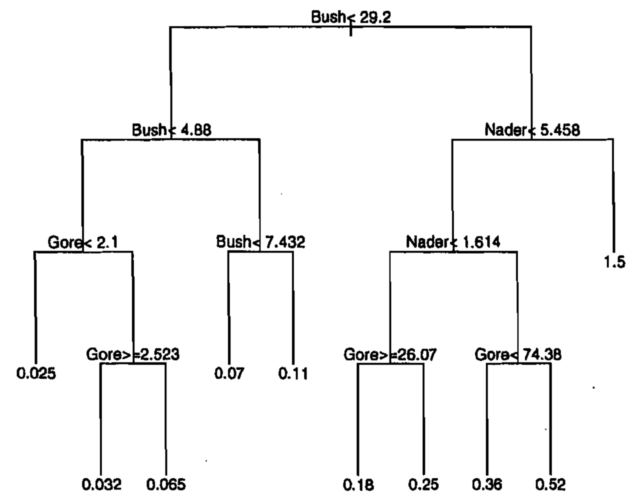
Graph D2: Classification and Regression Tree of Buchanan Percentage Fitted to Total Vote and Vote Percentages for Gore, Bush, and Nader Weighted by Inverse Total Votes (in 1000's).



End nodes give average Buchanan percentage for counties.

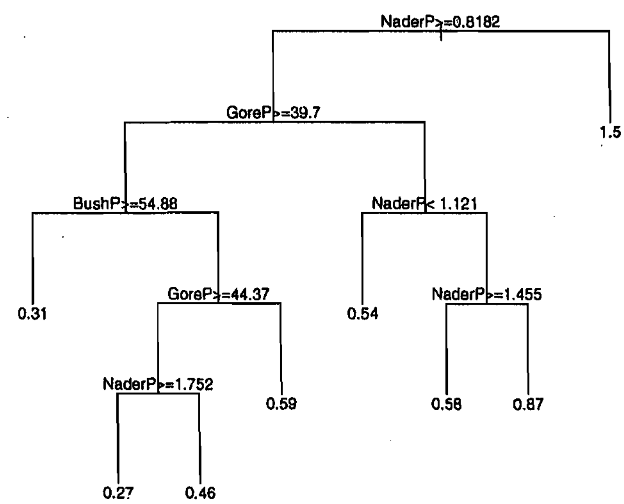
Figure 2.5. Classification and regression trees.

Graph D3: Classification and Regression Tree of Buchanan Vote Fitted to Votes for Gore, Bush, and Nader Weighted by Inverse Total Votes (in 1000's).



End nodes give average Buchanan votes for counties.

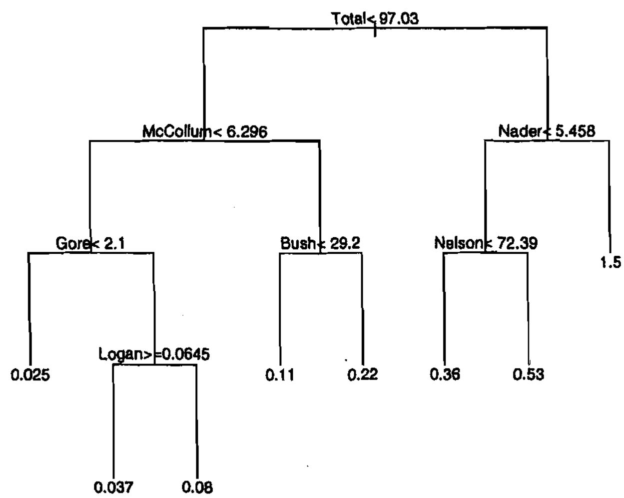
Graph D4: Classification and Regression Tree of Buchanan Percentage Fitted to Vote Percentages for Gore, Bush and Nader Weighted by Inverse Total Votes.



End nodes give average Buchanan percentage for counties.

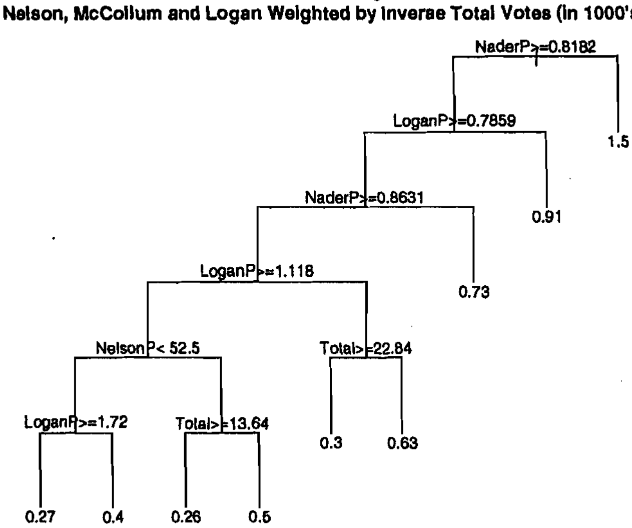
Figure 2.6. Classification and regression trees.

Graph D5: Classification and Regression Tree of Buchanan Vote Fitted to Total Vote and Votes for Gore, Bush, Nader, Nelson, McCollum and Logan Weighted by Inverse Total Votes (In 1000's).



End nodes give average Buchanan votes for counties.

Graph D6: Classification and Regression Tree of Buchanan Percentage Fitted to Total Vote and Vote Percentages for Gore, Bush, Nader, Nelson, McCollum and Logan Weighted by Inverse Total Votes (In 1000's).



End nodes give average Buchanan percentage for counties.

Figure 2.7. Classification and regression trees.

In particular, Smith (2002) carries out a carefully thought through series of regression analyses based on data which also include the vote counts for Libertarian presidential candidate Browne, but not the votes for 5 other fringe presidential candidates who each received less than 0.1% of the vote (although some of these latter may perhaps carry some degree of explanatory power for the Buchanan vote). Smith does not use data from any of the senatorial races, but he obtains and uses demographic data related to ethnicity, age, education and income for the counties. The predictor variables in his models were selected using either Mallow's C_p or backward selection. After detailed analysis, Smith selects three linear regression models for in depth consideration. In these models, the three resulting 95% prediction intervals range from a lower end value of 180 votes, to an upper end value of 758 votes for the Buchanan vote in Palm Beach. Further analyses, based on 3 logistic regressions, ranged similarly from 237 at the low end, to 606 votes at the high end.

Smith's prediction intervals and our own are consistent, and hence confirmatory for us. Of course, Smith's intervals are narrower, as one would expect, based as they are on a more comprehensive set of explanatory variables. In fact, Smith's intervals are remarkably tight, especially at the high end, and benefit heavily from inclusion of the demographic variables (although in his article Smith maintains that his analyses are not meant to be definitive but only a positive example of regression methods). The higher right endpoints for our own prediction intervals also suggest that the analyses which we carried out at the time were, in fact, fairly conservative.

By way of concluding remarks, it seems fair to say that statistical methodology is characterized by the development of increasingly sophisticated models. While developing theory leads to increasingly accurate assessment of significance, this significance is grounded in the correctness of the assumed model. While such models are generally accepted in designed experiments, this is not the case for observational studies where confounding covariates potentially lead to controversy. In fact there is, in such cases, no single correct answer, and even the term 'correct' as used here is ambiguous. In place of Herculean efforts to identify a single correct model, it is sometimes best to pursue and present many reasonable models. Within legal courts, it has been considered a useful practice to bring before jurors more than one case — typically the strongest case for, and the strongest case against. But in the court of public opinion there are many cases to be presented and it is the public who must then determine where the preponderance of evidence lies. The role of the statistician can be to present this full range of evidence using arguments and displays that must be as clear and as understandable as possible to be effective. Here we have illustrated this approach in the context of the singular events at Palm Beach County. Of course, there is no one 'right' answer. But by now, the court of public opinion has rendered a verdict: Bush would not have won this election with a clearer ballot in Palm Beach.

As an aside, we allow ourselves here to remark that in elections, it is not uncommon for a party to garner a majority of 'seats' while another party holds a larger share of the popular vote. Here the issues are not statistical, of course, but rather lie at the core of arguments for proportional representation. Such arguments sometimes confuse party support with local representation. The voting schemes of the U.S., Canada, and the U.K., for instance, are based on a principle that communities, and not parties, should be represented in the legislative bodies. So far, this has led to comparatively stable governments in those countries. There is no compelling evidence of which we are aware that proportional representation leads to better or more effective government.

Another point we have tried to illustrate here is that many statistical challenges can be usefully addressed via methodologies that have been carefully developed in the past and the problem considered here is a case in point. What is then new in any analysis is the application of the methodologies; this is the art of the statistician. Here the challenge is typical. It is not to test a hypothesis or to make inferential statements about a parameter. It is to measure something on the basis of available data, namely, the number of votes intended for Gore that went to Buchanan. Like all measures this one has bias and error and the challenge is to develop the measure and to estimate its bias and error. Furthermore, sometimes the effectiveness of an analysis depends critically on whether the statistician can deliver, in timely fashion, credible, understandable, and defensible results, under circumstances in which simplicity and presentation are highly influential.

Finally we might remark here that within statistics, outliers are always of essence and occur for many reasons. Some are no more than 'secretarial' errors, some are influential but nevertheless valid observations, some are symptomatic of model deficiencies, an occasional few are worth patenting, and a very few — like the one in Palm Beach County (for which clear and convincing explanations exist as to how it occurred) — change the course of history.

Acknowledgments. The authors thank Rob Tibshirani for helpful discussions and a motivating example, Barbara Thomson for assistance in computing the regression trees, and David Quance of the CTV Television Network for information concerning Canadian elections.

STATISTICAL MODELING AND ANALYSIS FOR COMPLEX DATA PROBLEMS

Edited by
PIERRE DUCHESNE
Université de Montréal and GERAD

BRUNO RÉMILLARD
HEC Montréal and GERAD

Springer

2005