CBS NEWS -CAMPAIGN 68 WARREN J. MITOFKSY - CBS NEWS

CBS News will report the results of the 1968 elections for each state-wide presidential contest and all gubernatorial and senatorial contests. We shall also analyze the source of the candidates' votes and compare them with past elections at state, regional, and national levels. The source of a candidate's vote will be estimated by geographic area; by size of city and rural areas; for areas classified by income, occupation, past voting behaviour, and various other characteristics by which geographic areas can be classified. We shall also attempt some measurement of the issues influencing people to vote as they do. We shall try to separate, in our analysis, the salience of issues as they influence change in voting behaviour from the past. It may be somewhat obvious, but let me point out the source of the various analyses. The estimated vote for various classifications of geographic areas will be based on samples of precincts. Actual returns on election night will provide the source of these data. Any estimate related to issues will, of necessity, be the result of preelection sample surveys. In addition to statewide elections, we will estimate, by geographic region, the party composition of the House of Representatives. If the election for president is not to be decided in the Electoral College, where an absolute majority of electoral votes is necessary, the election will be decided by the House of Representatives, where each state has one vote. If this contingency seems likely, we shall report the majority party for each state in the House of Representatives.

Our computer hardware, supplied by IBM, consists of two 360's, model 65, each with one million bytes of core. All programs, past data and election night data will be retained in core and it will not be necessary to access disk or tape storage for any calculations. On election night, all input will be by use of IBM 2260's, which will eliminate any use of punch card equipment. For those of you unfamiliar with this device, the 2260 is a display screen with the ability to enter data or recall data from a computer. Another device, the IBM 2250, will be used for graphic display and display of all estimates of candidates' percentages on air. The programming is being done by Informatics, a company which has extensive real-time programming experience on such projects as the airline reservation system and the Western Union message switching system.

In 1966, for the first time at CBS, election estimates were based on a probability design. This year we hope we have made some significant improvements both in procedures and use of available data. This includes use of both NES and returns from samples of precincts.

The estimation of the outcome of an election is treated in many ways as a traditional sample design problem. A single stage sample of precincts is selected with probability proportionate to size from a stratified frame which includes data from a past election for all precincts in a state. The stratification is based on past vote for precincts as well as county characteristics. Various estimators are available that will make maximum use of the usually high correlations of total vote or party vote with past elections, such that variance can be minimized. In fact, precise estimation of election results would not be conceptually difficult if it were not for the requirement imposed by the networks for calling the winner at the earliest possible time.

Before enumerating these difficulties, let me clarify a few points that are usually raised about early calls. First, the results of an election within a state are not broadcast without some minimum amount of actual election returns being available from the state. This means that either some or all of the polls are closed within a state before estimates are possible. Second, if all polls within a state are not closed at the time a winner is announced for that state, the local stations in the state are cued and have the option of not broadcasting the result.

Third, I know of no study that has adequately estimated the number of potential voters who have not voted at the time results are announced for the east. And, of this group of potential voters, those who have been aware of the announced results, and who also have been influenced either not to vote or to vote for a particular candidate. The last part of such a study should estimate the impact of an election.

The first special statistical problems related to the early call of winners has to do with missing reports for sample precincts. The usual assumption for most statistical models is that accurate measurement exists for all elementary units in the sample. If measurement is missing for elementary units, then this component of error must be included in the determination of the mean square error. If elections are to be called as soon as possible after the polls close, with a predetermined risk of calling the wrong winner, it is necessary to measure the error due to missing returns from precincts, as more than half of the precincts may be missing at the time of a call.

If the correlations between a past and present election are high, it is possible to impute results to missing precincts. If correlations are low, it may be reasonable to assume that those precincts reporting are approximately a random subset of the initial sample. Of course, the subset would have to be relatively large. It is also possible, for defined geographic areas within a state where the reporting of precinct results is known to be slow, to make election day estimates of actual voters. These estimates may be superior to imputation based only on actual returns. In any case, it is necessary to select the better imputation procedure and estimate the resulting contribution to the mean square error.

The second major statistical problem relates to quality control. On election night it is almost impossible to determine whether results are reported correctly for a particular precinct. I might clarify what I mean by incorrect results. This happens when unknown to us, a precinct's boundaries have been changed or when precinct names have been changed. In either case, we can receive correct results for a geographic area different than the one selected in the sample. This can yield misleading results as the various estimators usually depend on the correlation of total vote, and in some cases of party vote between present election and a past election. Some control of input data can be accomplished by having various criteria for acceptance available to the input operator on recall from the computer through the 2260. Alternatively, the report may be correct but the results may act like an outlier. The recognition of "outliers" versus trends in data is a significant problem. Consequently, the acceptance or rejection of data must be flexible enough so as not to distort the estimate at any given point in time. The rejection of data should be done so that the additional bias will be less than the reduction in sampling error. An error in quality control procedure occurred for the Maryland gubernatorial election of 1966 where CBS called George Mahoney a "probable winner". The call resulted from a failure to distinguish between a trend and an outlier. Consequently, data was completely and incorrectly rejected when it should have been accepted.

The third problem has to do with the constantly changing precinct boundaries. If sample precincts are selected from either 1964 or 1966 sampling frames, it is quite likely that a substantial number of boundaries have been significantly modified. It is possible through a great deal of tedium to resolve boundary changes for individual precincts within slightly larger geographic areas that remain common over the years. An unbiased measure of size for either 1964 or 1966 total vote can be obtained for a precinct as it is geographically defined in 1968. Reasonable approximations to the party vote also can be made. Failure to recognize boundary changes during the time of the field work, results in much of the incorrect data described earlier.

Last, but not least of the problems, is the necessity of establishing criteria to exercise the proper options of quality control, imputation, weighting, estimation, and estimation of the mean square error. For example, while a regression estimate theoretically might yield the smallest variance, the problem of estimating the regression coefficient reliably might make a ratio estimate more desirable. Of course, "if the Wallace third party is at all effective ..., analyses based largely on past experience might go wrong and 'all bets would be off'," as our chairman prophetically said in his book about elections 20 years ago. In the case of base correlation of party vote with a past election, estimate will be based on current data only. Also, imputation must proceed differently when either the correlation with party vote is small, or when a third party candidate has a reasonable share of the vote.

Other criteria, related to the decision as to the winner are under the final review of a statistical team. We believe it is possible to program most of the decision process, but not all. People are still most able to recognize patterns or irregularities. The recognition of the winner of a race is the focus of the decision. The actual percentages for candidates is a by-product not subject to the same rigor as the winner decision. Elections are the only sample survey problem where three groups are doing the identical survey and parameters are available (in most states) shortly after the estimates. If the proper options are exercised incorrectly, the failure will soon be known coast-to-coast.

I might conclude by pointing out that almost any statistical design is adequate for a landslide election. Furthermore, no design will be reliable for an extremely close contest. It is the middle ground, where elections are won by 2 to 10 percent margins, that the networks' election estimates have their greatest efficacy for reporting results in a timely and informative way.

IMPROVING FEDERAL STATISTICS ON CRIME AND CRIMINALS

٧I

٠.

Chairman, DANA M. BARBOUR, U. S. Office of Statistical Standards

<pre>Improving Criminal Statistics - LESLIE T. WILKINS, University of California Uses of Surveys for Estimating Crime Incidence - ALBERT D. BIDERMAN, Bureau of Social Science Research, Inc</pre>	Page
	102
	107
Truth and Consequences in Criminal Statistics - JOHN P. CONRAD, Bureau of Prisons	112