SAMPLING ISSUES IN THE DESIGN AND IMPLEMENATION OF HISPANICS EXIT POLLS

Robert L. Santos, Lorraine E. Porcellini, Institute for Survey Research, Temple University
and Robert Brischetto, Southwest Voter Research Institute
Robert L. Santos, Temple University, Rm 505 USB, 1601 N. Broad St., Phila., PA 19122

KEY WORDS: Exit Poll, noncoverage, Hispanics, oversampling, multi-stage

I. Introduction

An exit poll is a sample survey in which voters are selected to be respondents as they emerge from their voting location. For the past twenty years, exit polls have been employed to predict and analyze election day voting in the United States. The goal of this paper is to develop sampling issues relevant to the design and implementation of exit polls of Hispanic voters. The 1988 Texas Primary Exit Poll, conducted by the Southwest Voter Research Institute, will serve to illustrate one approach to the design of an Hispanic voter exit poll.

A review of the literature failed to identify published methodological work on Hispanic exit polls. However, a substantial literature exists on the topic of rare element sampling and surveys of minority populations. For example, Kish (1965a) and Kalton and Anderson (1984) discuss techniques amenable to rare element sampling. Ericksen (1976), Tourangeau and Smith (1985) and Santos (1985) illustrate area probability sample designs for household surveys of minority populations. Levy (1983) discusses general methodology of election day polls of the general population.

II. Survey Goals and Sampling Issues

Revelant sampling issues in an Hispanic exit poll are readily delineated through the use of a concrete example. To this end, consider an exit poll of Hispanics in the state of Texas for the "Super Tuesday" Democratic and Republican party primaries in March 1988.

The survey goals called for a sample of about 2,000 Hispanic voters from Texas. (An "Hispanic" is defined using a self-identification question in the polling instrument which is similar to the standard 1980 Census Spanish Origin query.) A maximum sampling error of 2.0 to 2.5 percentage points for percentage estimates was desired. Because the distribution of Hispanic voters in Texas is markedly different than that of the general population, intentional noncoverage was not to exceed 5 percent.

Several design constraints were imposed on the Hispanic exit poll. First, an exit poll methodology was employed. This called for a two-stage cluster sample design, in which primary sampling units (PSUs) consisted of precincts, and secondary sampling units were voters within precincts. Voters would be selected for this poll as they exit their voting location. Secondly, due to cost considerations, no more than 50 PSUs were used. Third, in order to facilitate the analysis of the survey data, a design close to equal probability sampling (epsem) was desired. Finally, the utility of exit poll data was to be considered: Timeliness was of paramount importance. Consequently, field procedures (e.g., selection procedures, the questionnaire itself, reporting procedures) were to be as simple as possible. Also the estimation task, including the calculation of sampling errors which reflect the complex nature of the sample design, was to be straightforward.

The first important design issue for this survey was the choice of sampling frame for the selection of PSUs. (A sampling frame is simply the list of elements from which the sample is drawn.) Fortunately, a convenient frame was available from the office of the Secretary of the State of Texas. Through this office, a precinct level data tape was obtained which contained precinct identifiers, geographic data (i.e., county), counts of total registered voters (TRV) and counts of Spanish surnamed registered voters (SSRV). (The data were updated through 1987.) No data were provided on voter turnout from a previous election. Ideally PSUs should be selected with probabilities proportional to the number of Hispanic registered voters who actually vote. In absence of such information, PSU probabilities were based on SSRV.

Rare element sampling plays an important role in the remaining design issues. To begin, consider the population of Spanish surnamed registered voters in Texas. In 1987, approximately one million SSRV resided in Texas. This represented about 14 percent of all registered Texans. Table 1 presents the percentagewise distribution of the SSRV population by precinct density of SSRV. About 60 percent of SSRV reside in high density SSRV precincts (40% or more SSRV concentration). However, about one sixth of the SSRV population resides in precincts with less than 10% concentrations of SSRV. Just over one quarter of the SSRV population resides in precincts with less than 20% SSRV concentrations.

Table 1: Percentagewise Distribution of Spanish Surnamed Registered Voters (SSRV) and Total Registered Voters (TRV) by Precinct Density of SSRV

Precinct Density of SSRV	Percentagewise Distribution of the: SSRV Population	TRV Population
0 - 0.9%	1.3%	24.3%
1 - 4.9%	8.1	35.9
5 - 9.9%	7.2	12.9
10 - 19.9%	9.6	8.9
20 - 39.9%	13.3	6.3
40 - 59.9%	13.4	3.6
60 - 79.9%	21.0	4.0
80 - 100.0%	26.0	4.1
Total	100.0%	100.0%
(N)	(1,021,088)	(7,482,687)

Thus, a significant portion of the population resides in areas which, if sampled, require modest to major screening efforts. This gives rise to the next design issue -- the identification of eligibles in an exit poll.

The exit poll methodology calls for the selection of voters as they emerge from their voting locations. For Hispanic exit polls, we ideally would subselect among Hispanic voters. Unfortunately, there is no way of knowing the ethnicity of a voter without a screening procedure. One of two alternative design strategies must then be adopted. A screening instrument could be administered to voters (prior to the actual interview) to determine eligibility. Alternatively, Hispanic voters could be permitted to fall into the sample (as a subclass) as a consequence of sampling among all votes. In order to simplify data collection, the latter strategy was adopted.

Accepting Hispanics as a subgroup has one major drawback. To obtain an efficient sample of Hispanics, they should be drawn in approximately equal numbers per PSU. However this gives rise to disparate PSU workloads. A precinct with 10 percent SSRV density requires eight times as many interviews (to get the same number of Hispanics) as one with 80 percent SSRV density. Unequal PSU workloads can cause staffing problems and may lead to unequal interviewer workloads. In turn, this increases survey costs and the potential for nonsampling error (e.g., nonresponse).

Intentional noncoverage was employed to eliminate those areas which, if selected, would have proved too costly to survey. All precincts which contained less than 100 registered voters or which exhibited SSRV concentrations of 1 percent or less were deleted from the sampling frame. This resulted in a loss of only 1.5 percent of the SSRV population in Texas. Roughly 42 percent (3,300) of all precincts were dropped, representing about one quarter of total registered voters. The resultant high Hispanic coverage rate, coupled with the significant deletions of total registered voters and precincts are highly desirable in rare element sample designs.

Oversampling precincts with higher density SSRV was also considered. Disproportionate sampling can be an effective tool for reducing the high costs of screening, and under certain circumstances may produce samples with minimum variance. However, disproportionate sampling was not employed because a principal survey goal called for an epsem sample. This was unfortunate, since Table 1 shows that almost 74 percent of the SSRV population resides in precincts with at least 20 percent SSRV concentration. For instance, two strata could have been created: a 20+% SSRV concentration stratum, and a less than 20% SSRV concentration stratum. The ratios of their respective average SSRV densities is roughly 9.5 to 1. Waksberg (1973) suggests that such ratios may lead to significant gains in precision via disproportionate sampling.

III. The Actual Design

The sample design utilizes a two stage paired selections model. Twenty-five strata of approximately equal numbers of SSRV were assembled by collapsing neighboring cells (whenever necessary) of precincts in a cross tabulation of a sevenfold categorization of SSRV density and a sixfold categorization of total numbers of registered voters. Minimum sized PSUs were then created by combining precincts containing few SSRV with those containing larger numbers of SSRV. Next, two PSUs were independently selected from each stratum without replacement and with probabilities proportional to SSRV. The paired strata were designed to facilitate sampling error computations in the analysis stage. A total of 50 PSUs were drawn.

In the second stage of sampling, voters were selected systematically as they emerged from their precinct voting locations. Generally, sampling rates within precincts were inversely proportional to the first stage selection probability. To facilitate the voter selection process, they were rounded to a convenient fraction. Within precinct rates varied substantially, ranging from 1 in 1 to 1 in 11. The second stage sampling rate, f_2, was set to yield an overall rate of 1 in 70:

$$f = 2\ (\text{SSRV})\ /(\text{Stratum Total}) * f_2 = 1/70$$
$$f_2 = (1/140)(\text{Stratum Total} / \text{SSRV}),$$

where Stratum Total denotes the total SSRV count in a given stratum and SSRV denotes the number of Spanish surnamed registered voters in a given PSU.

IV. Survey Results

The Texas Primary Exit Poll gathered 4,443 interviews. Of these, 2,114 were Hispanic voters. Over 200 Hispanic interviews were obtained from the Republican Primary, while over 1,800 resulted from that of the Democrats. Although there is no comparison available to the true population totals, the Hispanic sample displayed heterogeneity across such demographics as sex, age, education and household income. Table 2 provides these percentage distributions.

Table 2: Distributions of the Hispanic Sample* by Demographics

Sex	%
Male	48
Female	52

Age	18-25	26-35	36-45	46-55	56-65	66+
%Vote	18	24	23	13	14	7

Education	Not HS	HS	Some College	College Grad
%Vote	28	28	27	16

HH Income ($1,000)	10 & less	10-19.9	20-29.9	30-39.9	40-49.9	50+
%Vote	28	25	19	13	7	8

* The distributions are based on a sample of N = 2,114 from the 1988 Texas Primary Exit Poll conducted by Southwest Voter Research Institute.

The overall response rate for this survey was 59 percent. It is calculated as the ratio of completed interviews to total voters selected. Table 3A suggests that there was substantial differential response by precinct density of SSRV. PSUs with 1 to 20 percent SSRV concentrations achieved on average a response rate about 14 points lower (54.5 vs 68.6) than PSUs with 60 percent or more SSRV density. There are two possible explanations for this phenomenum. First, PSUs with lower SSRV densities tended to have higher sampling rates within PSUs. With rates as high as 1 in 1, it may not be surprising that many voters failed to respond, since the logistics of such "censuses" would be difficult to operationalize. Secondly, many of the field staff were Hispanic, and there may have been a component of nonresponse (in non-Hispanic PSUs) attributable to the disparate ethnicities of the interviewer and the selected subject.

Another interesting result involves a comparison of voting trends across SSRV densities. Table 3B suggests that the percentage vote among Hispanic democrats for Dukakis varied considerably by precinct SSRV concentration. This suggests that future samples should ensure that the low density SSRV precincts are represented, since Hispanic voters in such areas may vote differently than their counterparts in higher SSRV density precincts. Although not included in this paper, it would be interesting to check for differential voter turnout by SSRV density, as well.

Table 3A: Response Rates by Precinct Density of SSRV

	1-20%	21-60%	60+%	Total
Response Rate	54.5%	55.2%	68.6%	59.1%

Table 3B: Percent Vote for Dukakis Among Hispanic Voters by Precinct Density of SSRV

	1-20%	21-40%	41-60%	61-80%	81+%
%Vote	40.2	50.2	58.3	48.3	54.1

Design effects of percentage estimates were calculated for selected items in the questionnaire. The percentage was treated as a ratio mean, and the computation followed a Taylor expansion approximation adapted to a paired selections model (found in such sampling texts such as Kish [1965b]). Design effects (DEFF) are defined as the ratio of the actual variance of a statistic to that obtained under a simple random sample of the same size:

$$DEFF = \text{Actual Var}(p) \; / \; \text{SRS Var}(p).$$

Design effects averaged 2.10 and ranged from 1.41 to 2.73. We conclude that the precision requirements for the survey were obtained.

Although the sample design was epsem, the sample was not self-weighting for two principal reasons. First, the field staff failed to adhere to the desired PSU sampling rates. This was mostly due to interruptions of the field staff during data collection, and due to late arrivals of the staff (i.e., arriving after the polls had opened). Of course, a second factor is nonresponse. There were no substitutions or adjustments permitted in the sampling process. Thus, the sampling process continued irrespective of the subjects' cooperation.

The weighting procedure utilized in this study involved the calculation of an "effective" within PSU sampling rate. It was defined as the ratio of the total votes cast (as reported at the close of the polls) to the number of completed interviews. Within each PSU, this implicitly assumes (1) completely random nonresponse, and (2) equal turnout rates among Hispanics and non-Hispanics. To the extent that these assumptions are violated, estimates based on the weighted survey data may be biased.

V. Considerations for November

On November 7, 1988, Southwest Voter Research Institute will conduct Hispanic exit polls in Texas, California and perhaps New Mexico. In light of our experience with the Texas Primary Exit-Poll, there are two design issues presently under consideration. The first deals with the issue of oversampling. We have seen that epsem designs will require weighting to adjust for nonresponse and departures from prespecified selection rates within PSUs. Since weighting will be required regardless of the design, and since disproportionate sampling could likely produce noticeable gains in precision, an optimal allocation design should be seriously considered.

Secondly, a more realistic weighting procedure should be investigated. The assumption of random nonresponse within a PSU is tenuous at best. Perhaps field staff could record detailed (to the extent possible) demographic characteristics on nonresponders. Such items might include sex, age (in gross categories) and "guessed" ethnicity. These data could then be employed in a more realistic nonresponse adjustment. It might be difficult, however, to design an adjustment scheme which could be adopted on election night.

References

Ericksen, E.P. (1976)
Sampling a rare population: a case study. Journal of the American Statistical Association, Vol.71, pp. 816-822.

Kalton, G., and Anderson D.W. (1984)
Sampling rare populations for health surveys. Proceedings of the Section on Survey Research Methods, American Statistical Association.

Kish, L. (1965a)
Selection techniques for rare traits. In Genetics and Epidemiology of Chronic Diseases, Public Health Service Publication, No.1163.

Kish, L. (1965b)
Survey Sampling. Wiley, New York.

Levy, M.R. (1983)
The methodology and performance of election day polls. Public Opinion Quarterly, vol.47, pp.54-67.

Santos, R. L. (1985)
One approach to oversampling Blacks and Hispanics: the national alcohol survey. Proceedings of the Section on Survey Research Methods, American Statistical Association.

Tourangeau, R., and Smith, W. (1985)
Finding subgroups for surveys. Public Opinion Quarterly, Vol.49, pp.351-365.

Waksberg, J. (1973)
The effect of stratification with differential sampling rates on attributes of subsets of the population. Proceedings of the Social Statistics Section, American Statistical Association, pp.429-434.