

ABSTRACT

Generally social surveys and opinion polls are of the fixed sample size type, though based on probability sampling techniques. Wald's sequential procedure is rarely put to use. Peel and Skipworth (1970) present the results of a pioneering attempt of applying sequential methods to social surveys. If some preliminary data are available, the researcher can set up some null hypotheses regarding the phenomena to be studied. The Wald SPRT and the generalized SPRT provide an efficient hypothesis testing procedure for a prescribed (α, β) resulting in considerable saving of the resources. We illustrate the case of voting preference in the city of Edmonton, Canada using the sequential approach. Telephone calls are used to gather data and at each call to accept, or reject, the null hypothesis, or to continue the experiment is made.

1. INTRODUCTION

In the social sciences there appear to be very few application of sequential methods as described by Wald (1973), and Wetherill (1975). The one exception seems to be the work of Peel and Skipworth (1970) on marriage and family size in the English city of Hull. The present paper describes the results of the application of Wald's sequential probability ratio test (SPRT) to determine the preferences of Edmonton area voters prior to the Canadian federal election of May 22, 1979. We feel that public opinion could be gauged much more quickly and efficiently using the SPRT. The purpose of this research was to establish if sequential methods provide reasonably accurate measures of voters' preference, rather than study its correlates.

2. SEQUENTIAL TESTS

2.1 Binomial Sequential Probability Ratio Test

The SPRT used here is the special cases developed by Wald for testing the mean of a binomial distribution. The random variable x can only take the values 0 and 1. Two sets of hypotheses, one about the proportion of support for the Progressive Conservative party and another about the degree of support for its leader, Joe Clark are tested. Support for the Progressive Conservative party or the leader of the Conservative party in the latter case is assigned the value 1, while opposition to either assumes the value 0. If

p = the proportion of Conservatives (or Clark supporters) and

d_m = the number of Conservatives (or Clark supporters) in the first m units sampled,

then the SPRT is written as follows:

$$(1.1) \frac{P_{1m}}{P_{0m}} = \frac{p_1^{d_m} (1-p_1)^{m-d_m}}{p_0^{d_m} (1-p_0)^{m-d_m}}$$

where $p^{d_m} (1-p)^{m-d_m}$ represents the probability of obtaining the observed sample (Wald, 1973, p.90). Further, when $A = 1-\beta/\alpha$ and $B = \beta/1-\alpha$ the zones of acceptance, rejection and continuation may be expressed as follows:

ressed as follows:

i) reject H_0 if

$$(1.2) \log \frac{P_{1m}}{P_{0m}} \geq \frac{1-\beta}{\alpha};$$

ii) accept H_0 if

$$(1.3) \log \frac{P_{1m}}{P_{0m}} \leq \log \frac{\beta}{1-\alpha}$$

iii) continue sampling if

$$(1.4) \log \frac{\beta}{1-\alpha} < \log \frac{P_{1m}}{P_{0m}} < \log \frac{1-\beta}{\alpha}$$

$$\text{where } \log \frac{P_{1m}}{P_{0m}} = d_m \log \frac{p_1}{p_0} + (m-d_m) \log \frac{1-p_1}{1-p_0}$$

Wald (pp.92-3) has shown that a few simple algebraic steps will convert the boundaries of the three zones into what he terms the acceptance number, a_m , and the rejection number, r_m . These numbers can be evaluated for each value of m using

$$(2.1) a_m = \frac{\log \beta / (1-\alpha)}{\log \frac{p_1}{p_0} - \log \frac{(1-p_1)^m}{(1-p_0)^m}} + m \frac{\log \frac{(1-p_0)}{(1-p_1)}}{\log \frac{p_1}{p_0} - \log \frac{(1-p_1)}{(1-p_0)}}$$

$$(2.2) r_m = \frac{\log \frac{(1-\beta/\alpha)}{p_0} + m \log \frac{(1-p_0)}{(1-p_1)}}{\log \frac{p_1}{p_0} - \log \frac{(1-p_1)}{(1-p_0)}}$$

The equations (1.2) to (1.4) can be restated in the following manner

i) reject H_0 if $d_m \geq r_m$;

ii) accept H_0 if $d_m \leq a_m$;

iii) continue sampling of $a_m < d_m < r_m$.

When the values of the acceptance and rejection numbers have been determined the SPRT becomes the simple task of counting the number of conservative backers (or Clark supporters). Sampling ceases and a decision is made whenever either the acceptance or rejection number is at least equalled.

Wald (p.105) gives a rule for reaching a decision if the sequential test is terminated at a specified n_0 but no decision has yet been made, i.e.,

$$a_m < d_m < r_m : \text{ i) if } \frac{a_{n_0} + r_{n_0}}{2} \geq d_{n_0} \text{ accept } H_0,$$

and

$$\text{if ii) if } \frac{a_{n_0} + r_{n_0}}{2} \leq d_{n_0} \text{ reject } H_0.$$

He maintains that the use of this rule does not involve any serious error. A second suggestion Wald makes is to simplify the process by sampling in groups, rather than singly. Wald (pp.103-3) argues that for practical purposes, the effects of grouping are small and may actually decrease the values of α and β to ones smaller than those employed in the test.

2.2 The Average Sample Number (ASN) Function

The ASN function was evaluated to obtain an approximation of what the sample size might be for a given α, β, p_0 and p_1 values. Equation (3.1) below gives the value of the expected sample size for a given p :

$$(3.1) E_p(n) = \frac{L(p) \log \beta + (1-L(p)) \log A}{p \log p_1/p_0 + (1-p) \log \frac{1-p_1}{1-p_0}}$$

where $A = \frac{\beta}{1-\alpha}$ and $B = \frac{1-\beta}{\alpha}$ and $L(P)$ represents the value of the operating characteristics function at p . It can be shown that $L(p)=\beta$ at $p = p_1$ and $L(p) = \alpha$ at $p = p_0$. This allows the comparison of the expected and actual sample size.

3. SAMPLING PROCEDURE

3.1 Sampling Frame

The 1979 Edmonton and Vicinity Phone Book which lists residential phone numbers for the city of Edmonton and the nearby communities was the sampling frame. The phone book probably better reflects the residents of this region as of November 1978 rather than May 1979. This source was used because it was easily accessible and the sample could be drawn quickly and inexpensively. This sampling frame suffers from the fact that transients, recent migrants into the area, residents with unlisted numbers or without phones are excluded. Another difficulty is that the target population is eligible voters and not households with phones.

3.2 Sampling Design

Two independent random samples were selected, one of Edmonton residents and another of residents of the nearby communities and towns.**The samples were drawn in a two stage process. First, a simple random sample of twenty pages of Edmonton and seventeen pages for the surrounding region were selected from the phone book. Next, a systematic random sample of ten phone numbers was selected from each page using a different random start in each case. Non-residential numbers were removed and if a page had more than four such numbers it was replaced (two pages in the Edmonton sample were replaced). The phone numbers were then listed in their order of selection.

The interviewing was done by phoning respondents during the period of May 15 to May 18. The table gives a breakdown showing types of incomplete calls and the number of refusals in the completed calls. For the Edmonton sample 28 of 137 eligible voters contacted refused to participate, while in the sample of outlying areas 15 of 97 eligible voters contacted, refused to participate. This gives response rates of approximately 80 per cent in both cases when unanswered calls, phones no

TABLE (1). Breakdown of the Telephone Calls made by Type

	Total calls	Phones not in service	Phones not answered	Total completed calls
Edmonton	191	14	21	137
Outlying area	149	11	23	97

TABLE(1) contd. Breakdown of the Telephone Calls made by type

	Ineligible to vote	Eligible to vote Refused	Responded
Edmonton	19	28	109
Outlying area	18	15	82

longer in service and ineligible voters are not counted.

4. PARTY PREFERENCE AND LEADERSHIP QUESTIONS

Respondents in each region answered two questions. The first question asked was whether they favored the candidate of the Progressive Conservative party and the second asked was, which of the three leaders of the major political parties in Canada, they felt, would make the best Prime Minister. These are standard questions in election polls and their wording was obtained from that used in the Gallup Poll. The exact wording of the questions is given in Appendix A.

5. GALLUP POLLS

The Canadian Institute of Public Opinion releases monthly surveys of voter preferences in its Gallup Report. The survey results are based on a national sample of at least a 1000 respondents and are reported at a significance level of $\alpha = 0.05$. It was decided to stop sampling if the samples reached a size of 100, because the region being sampled is relatively homogenous, and because of resource and time limitations. For our sequential sample it was decided to let $\alpha = 0.05$ as in the Gallup Polls. Since β could be set to any desired level when employing a sequential test, which is not possible with a method based on fixed sample size, β was chosen equal to 0.05.

6. HYPOTHESES AND CRITICAL REGIONS

6.1 Hypotheses

There were four statistical hypotheses for testing two concerned the proportion of Conservative vote and two the degree of preference for the leader of the Conservative party. The hypotheses are summarized in Table 2. H_0 is a simple hypothesis in each case, while the alternate hypothesis H_1 is composite and the one in parenthesis is the simple hypothesis to which the composite is equivalent.

TABLE 2: The Hypotheses

Region	Question	
	Party Preference (proportion favor- ing the Conserva- tives)	Preference for Prime Minister (proportion in favor of the Conservative leader)
Edmonton	$H_0: p_0 = 0.53$ $H_1: p_1 \geq 0.57$ ($H_1: p_1 = 0.57$)	$H_0: p_0 = 0.30$ $H_1: p_1 \geq 0.34$ ($H_1: p_1 = 0.34$)
Outlying	$H_0: p_0 = 0.44$ $H_1: p_1 \geq 0.48$ ($H_1: p_1 = 0.48$)	$H_0: p_0 = 0.30$ $H_1: p_1 \geq 0.34$ ($H_1: p_1 = 0.34$)

The null hypotheses (H_0) were obtained in the following manner. The value p_0 for the party preference question represent the Conservative popular vote in the last 1974 federal general election for the electoral districts correspond- in to Edmonton's geographic boundaries and that for the district containing the surrounding towns (Report of the Chief Electoral Officer, 1975)**. The value for p_0 in the leadership question was taken from the Gallup Report of December 16, 1978. The figure used was the support for the Conservative leader outside Quebec. The two sets of hypotheses on leadership preference are the same because no separate information was available for each region. The p_1 values were

obtained by adding 0.04 to each p_0 value. The value 0.04 was added because four per cent is the standard margin of error reported in most opinion polls, especially those done by the Canadian Institute of Public Opinion. The difference between the null and the alternate hypothesis should be at least the margin of error associated with these polls.

There are two reasons that all the alternate hypotheses are of the form $p \geq p_1$. First the popular vote in the province of Alberta in the 1974 federal election was 61 per cent for the Progressive Conservative party and second, voters sentiment in the country seems to be in favour of the Conservatives or at least opposed to the governing Liberal party. These pieces of information suggest that the Conservatives may increase their popular vote in the Edmonton area. In addition since the leader of the Conservative party, Joe Clark is a native Albertan and Alberta a very Conservative province, it is likely that he may be more favorably evaluated in Alberta.

6.2 Critical Regions

The values of the acceptance number a_m and the rejection number r_m were determined using (2.1) and (2.2) for each m of the four hypotheses with $\alpha = \beta = 0.05$.

7. DATA AND FINDINGS

7.1 Sample Size

The expected sample size was found by employing equation (3.1). The expected sample size along with the actual sample size for each of the

hypotheses is displayed in Table 3. Only one of the four tests terminated prior to the sample size N reaching 100. The test for the leadership question in the outlying areas was stopped when the question on party preference reached a conclusion at $N = 71$. The difference in N in the two questions is due to the fact that undecided voters are excluded when calculating the proportion in favor of a particular party.

TABLE 3: Expected and Observed Sample Sizes

Region	Question	
	Party Preference	Preference for Prime Minister
Edmonton	$E_{p_0}(n) = 817$ $E_{p_1}(n) = 821$ $N = 100$	$E_{p_0}(n) = 728$ $E_{p_1}(n) = 712$ $N = 100$
Outlying	$E_{p_0}(n) = 824$ $E_{p_1}(n) = 821$ $N = 71$	$E_{p_0}(n) = 728$ $E_{p_1}(n) = 712$ $N = 81$

The results suggest that the tests should have been permitted to terminate at a larger N , this would have increased the likelihood of reaching a decision prior to stopping the tests. The expected sample sizes $E(n)$ are larger for the following reasons: i) the smaller the difference between the values of p_0 and p_1 the larger the sample required detect the difference; ii) the values of p_0 and p_1 used are very close to the maximum values for the ASN curve (Wald, p.99); iii) as the value of α and β decrease the sample size increases, but smaller values of these parameters have the desirable quality of increasing the strength of the test.

7.2 Sex Composition

The sex composition of each sample is given in Table 4. Sex may be of some relevance because studies of political behavior suggest that women vote more conservatively (Lipset, 1970, p.221). Undecided voters are excluded from the proportions calculated for the party preference question. It would have been preferable if the samples had about ten per cent more males as this would have been more representative of the Alberta population.

TABLE 4: Sex Composition (-number of males)

Region	Question	
	Party Preference	Preference for Prime Minister
Edmonton	$37/100 = 0.37$	$38/100 = 0.38$
Outlying	$27/71 = 0.38$	$28/81 = 0.35$

7.3 Findings

Tables 5 and 6 summarize the results of the survey for the two questions and the two areas. In Edmonton and the surrounding towns 58 per cent and 71.8

TABLE 5: Party Preference

Region	Party		N	Undecided
	Conserv- atives	Non-Cons- ervatives		
Edmonton	58 (0.58)	42 (0.42)	100	9
Outlying	51 (0.718)	20 (0.282)	71	11

per cent of the respondents preferred the Conservative candidate. The support for Clark, the

TABLE 6: Preference for Prime Minister

Region	Leader					N
	Clark	Trudeau	Broad- bent	Do not None Know		
Edmonton	31 (0.31)	33 (0.33)	24 (0.24)	5 (0.05)	7 (0.07)	100
Outlying	30 (0.37)	26 (0.32)	11 (0.14)	6 (0.07)	8 (0.10)	81

Conservative leader, was 31 per cent in Edmonton and 37 per cent in the surrounding communities.

Table 7 shows which of the null hypotheses were rejected and the values of the test statistic d_m and the value of rejection number.

TABLE 7: Decisions Reached by Each Test

	Party Preference	Preference for Prime Minister
	Edmonton	reject H_0 $d_{n_0} \geq \frac{r_{n_0} + a_{n_0}}{2}$ 58 > 55
Outlying	reject H_0 $d_m = r_m = 51$	reject H_0 $d_{n_0} \geq \frac{r_{n_0} + a_{n_0}}{2}$ 30 > 25.5

Only the test of voter preference in the outlying areas terminated without being stopped. At $m=71$ the rejection number was equalled leading to the acceptance of the alternate hypothesis that $p = p_1 = 0.48$. The other three hypotheses were either accepted or rejected using Wald's rule. The test statistic d_{n_0} in all three cases is very close to

the critical value; therefore the tests were not clearly decisive in any of these three situations.

8. DISCUSSION

A few points merit brief mention. Since the election results are available, though incomplete with a few polls having not yet reported, some comparisons of the popular vote in the two areas may be made to the sample results reported above. The actual popular vote for the Conservatives in Edmonton was 58.6 per cent compared to the 58 per cent suggested by the sample, while for the surrounding region the relevant figures are 76.6 and 71.8 per cent (Edmonton Journal, May 23, 1979, p. 3-18). The actual vote for the outlying towns is

a crude estimate since most the associated electoral districts are beyond the area sampled and only when poll by poll results are available will a more accurate comparison be possible.

The leader of the Conservative party was more highly evaluated in the outlying ridings than in the city electoral districts. This could have been due to the fact that he is a candidate for his party in one of the ridings, but a check of the seventeen respondents from his riding showed that six selected him while the leaders of the other two parties were selected seven times. He actually did slightly worse in his own riding than in the other two outlying ridings.

The application of the SPRT did not work quite as well as was hoped. Only one of the four tests terminated without being stopped arbitrarily, but in this case there was a large saving in observations and effort that would have been required had the fixed sample size method been employed. The time devoted to interviewing the 234 respondents was approximately twenty hours and about another twenty hours were spent in preparation (selecting the sample and evaluating the acceptance and rejection numbers).

The availability of more time or interviewers would have permitted a more conclusive testing of the hypotheses for which sampling did not terminate.

FOOTNOTES:

**Only those suburban communities which could be phoned toll-free were included in the sample.

*** In 1974 the year of the last election the federal riding of Pembina contained all the surrounding towns, since then redistribution has placed these towns into the three ridings of Pembina, Wetaskiwin and Yellowhead (Report of the Electoral Boundaries Commission for the Province of Alberta 1976).

REFERENCES

- _____, Edmonton Journal, May 23, 1979.
- _____, "Trudeau Holds Overall Lead as 'Best Minister'", The Gallup Report, December 16, 1978.
- Lipset, Seymour Martin, Political Man: The Social Politics. Garden City: Doubleday & Company, 1960.
- Peel, J. and G. E. Skipworth "Sequential Sampling as a Technique in Sociological Surveys," Applied Statistics, XIX (1970), p. 27-33.
- Report of the Chief Electoral Officer Thirtieth General Election 1974. Ottawa: Information Canada, 1975.
- Report of the Electoral Boundaries Commission for the Province of Alberta 1973, Ottawa Information Canada, 1973.
- Report of the Electoral Boundaries Commission for the Province of Alberta 1976, Ottawa: Information Canada, 1976.

Wald, Abraham Sequential Analysis. New York: Dover Publications, 1973.

Wetherill, G. Barrie Sequential Methods in Statistics. London: Chapman and Hall, 1975.

APPENDIX A

The three questions asked of respondents are listed below.

1. Are you eligible to vote in the May twenty-second federal election?
2. If a federal election were held today, do you think you would favor the candidate of the Progressive Conservative party?
3. Regardless of any political feeling you may have and thinking just of the individuals, which man do you think would make the best Prime Minister for Canada - Pierre Trudeau, Joe Clark or Ed Broadbent?

Question 3 asks which of the leaders of the three major political parties in Canada would make the best Prime Minister. Trudeau was the Prime Minister and leader of the Liberal party, while Joe Clark was the leader of the Progressive Conservative party which formed the official opposition prior to the election. Broadbent still is the leader of the third party - the New Democratic Party (NDP).

The sex of respondents was noted as well.

FOOTNOTES - continued

* This research was supported by a grant from the General Research Fund of the University of Alberta.