A primer on polls

Technote

Introduction

Public opinion polls are pervasive. For the most part, published reports by pollsters fail to provide sufficient information to allow expert appraisal of the overall validity of the results. Some polls do report overall sample size and a theoretical measure of accuracy, but the crucial aspects of the poll are often obscure.

Overall sample size and accuracy

Usually, a poll will report the overall sample size and indicate that this provides, for example, an error of less than 4%, 19 times out of 20. Much is made of this scientific "benediction," but what does it mean?

To interpret the accuracy of the figure, consider the following example:

"If an election were held today, what party would you vote for?"

TABLE 1

Party	Percentage
Blues	35%
Reds	30%
Yellows	10%
Undecided	20%
Refused	5%

The statement "±4%, 19 times out of 20" simply means that if 20 identical surveys were conducted, we would expect 19 of them to provide responses within +4 percent of the unobserved "true" mean. Some surveys report that "any value for the Blues between 31 and 39 is statistically significant." This is a first approximation, but strictly speaking, it is inaccurate. Assuming that 35% is the true value (as determined by a complete census or 100% sample), then 19 out of the 20 surveys are expected to show a value in the range of 31 to 39.

Many polls provide results for a subset of the sample. For example, in a national poll, the results for an individual province may be reported. Thus, with 4% of the national population, a well-executed national sample of 1,000 (i.e., completely random over all households) may produce a Manitoba sample size of 40 or so. The error rate of 4%, 19 times out of 20 does not apply to these data. For such a small sample size, the theoretical error rate is over 15%, 19 times out of 20. So, if the Blue party leads the Red party by 8%—a healthy margin in a large sample (1,000)—in a small sample (40), this is meaningless; it is a difference of three people (8% times 40).

It is possible to increase the sample in the smaller regions, but even doubling the sample size (from 40 to 80) only reduces the error rate to 11%—hardly a worthwhile increase in precision. The practice of not reporting theoretical errors or even sample size subsets of the data can be misleading. As a rough rule, any subsample of fewer than 400 respondents should be viewed with caution.

Refusals

The refusal rate, which is the percentage of all rebuffed attempts to conduct an interview, must always be reported. Quite typically, as many as 70% or higher of all contacted individuals will refuse to participate in the interview. Sometimes people are busy, and other times, they object to the subject or the interruption. It is not safe to assume that those who refuse have the same opinions as those who participate.

Comparing basic information of refusers and participants is an important control for sample quality. On most polls, gender can usually be inferred from a respondent's voice. As well, the telephone prefix may indicate the location of the respondent. A typical pattern is shown below.

TABLE 2

	Percent of refusals
Winnipeg	76.4%
Non-Winnipeg	23.6%
Female	34.5%
Male	65.5%

It is important to note that the refusal pattern is not representative of the population. Based on census data, it may be known that 55–58% of the Manitoba population live in the city. The 76.4% of all refusers residing in the city is typical of many telephone polls. A number of validated procedures are available to "balance" the sample and make it representative of the population. Normally, random sampling will produce representation, but refusals are a non-random error. It is important to ensure that the final sample closely resembles the population in important attributes.

Research design

Sampling must be random. All units (household, individuals, etc.) ought to have a known probability of being included in the sample.

The simplest approach is to randomly draw telephone numbers, names from an electoral list, telephone book, etc. In this way, everyone on the list has an equal chance of being selected. Many other random sampling designs are also used.

There is no perfect list (sampling frame); for example, the telephone directory omits unlisted numbers and covers only about 95% of the population. Rather than looking up numbers in a telephone directory (which would not provide unlisted numbers), computer generation of the numbers is often used. Random digit dialling processes have proven to be efficient and accurate. The simplest approach is to select a number from the phone book and add one. If the selected number is 555-1234, the dialled number would be 555-1235. Current methodologies are more complex and use a combination of printed directories and random digit dialling.

Despite their poor record, many surveys still use convenience samples, such as newspaper readers, those who log into an Internet survey, or those who call a toll-free line. Four major issues are associated with convenience samples.

- People who read any given newspaper may not be representative of the population.
- Readers who respond are not usually representative of all readers.
- Not everyone has access to the Internet.
- Repeat callers can load one side of the question.

Therefore, no claim can be made about how the population feels on the issues addressed, no valid error rate can be calculated, and sample quality cannot be assessed. Even modest quality cannot be assessed. Even modest basis for calculating the error rate. For example, a researcher may appear to increase the validity by doubling the sample size from 40 to 80, but if these extra respondents are simply added to the sample from a biased sample list, the error rate cannot be computed. There are appropriate ways to adjust the sample size, but many public opinion polls do not employ them because they make the analysis more difficult and slow the entire polling process.

Question wording

Question wording and positioning influence response patterns. Considerable research has been done on this in the past decade, but while some general principles have been established, much remains unclear. All reports must state the exact question phrasing. While many do this, some still report a condensed version of the full text.

The sensitivity of responses is exemplified in Trial A and Trial B, below. Note the shift in response when all other factors are controlled, except for a slight shift in wording. Also notice that the addition of the words "usually" and "probably" have opposite effects in different questions. The question wording dilemma is a key area of research in polling.

TRIAL A

When you take a problem to a lawyer you have less control over how the problem is resolved.
(n=637)

Agree	Neutral	Disagree
59.7%	2.7%	37.6%

When you take a problem to a lawyer you usually have less control over how the problem is resolved.
(n=657)

Agree	Neutral	Disagree
63.3%	3.4%	33.3%

TRIAL B

Stricter gun laws would decrease the number of murders.
(n=863)

Agree	Neutral	Disagree
62.5%	2.0%	35.5%

Stricter gun laws would probably decrease the number of murders.
(n=829)

Agree	Neutral	Disagree
58.5%	3.6%	37.9%

* Does not include don't know and refusals.

Consider a common question: *"What do you think is the most important problem facing Canada?"* Some polls report that anywhere from 15-20% will report the deficit as the most important problem, while others register under 5%. Why is there a difference?

In addition to question wording, format is important to the response pattern. Few researchers openly manipulate respondents through tricky wording: more common is an unconscious bias. In the question above, if a list that includes "deficit" as a possible response is read, then about 15-20% of respondents will identify it as most important. If respondents are not read a list of alternatives (i.e., the question is open), only 5% might volunteer "deficit" as the most important problem. It is important to be aware that for these type of questions, current news stories often influence respondents. Whose responses are being collected: the public's or those of the media?

"Open" questions without lists read will obtain general responses, many of which are compound concepts (e.g., *"economic growth and justice"*). These responses are difficult to unravel. The presentation of a list "solves" these issues—possibly at the expense of accuracy or possibly ignoring information that would have emerged had the respondent been allowed freedom of expression.

Reading lists also encourages the respondent to focus on the first or last item heard and confines the answers to those the researcher provides. Therefore, lists should be randomized for each respondent.

Method of survey

The telephone survey is common because it is efficient; the telephone directory has reasonable coverage (except for low-income transient population and special groups), and the development of computer-aided telephone interviewing (CATI) has made complex questionnaires easy to manage. CATI also allows interviewers to schedule call-backs and re-contact respondents who are unavailable.

However, telephone calls are intrusive, and many confuse a telephone survey with a sales solicitation. This is not surprising, since many marketing firms start their pitch as a phoney sales survey. As well, a telephone survey of more than a few minutes is onerous, and the interviewer risks losing the respondent or at least his or her focused attention on prolonged questions.

Face-to-face surveys may be preferred on the grounds that it is harder to refuse an interviewer at the door. However, the logistics of face-to-face interviewing on a national scale are formidable. Because of the low density of the population, rural areas may be under-sampled, producing significant bias.

Web-based surveys are becoming common because many people believe (mistakenly) that they cost less. (See Technote on Mail and Web Surveys.)

Checklist for reporting a survey

The following is a minimum set of requirements for reporting a survey:

- **Precise question phrasing must be provided in the report.**
- **The sample size and overall <u>theoretical</u> accuracy must be stated.**
- **Sample size and theoretical accuracy for <u>all</u> analyzed subsets should be indicated.**
- **Assessment of overall sample representativeness is important.** How does the sample compare with secondary sources such as census and taxation data? If there are biases (and there usually are, no matter how careful the design and execution), have the results been weighted to correct for this?
- **The overall refusal rate (percentage of all those who were contacted who refused to participate) should be reported.**
- **The interview dates and survey method must be clearly presented.** Polls that are undertaken over a long period of time or that do not employ a random sample should be disregarded as "unscientific."
- **The authorship and sponsorship of the report must be clearly stated.**

Calculating the approximate error rate for a survey sample

The usual procedure for computing error rate at the 5% (19 times out of 20) level employs the following formula:

$$B = +1.96(1-n/N)\ [p(1-p)/n]^{1/2}$$

This assumes a normal distribution, where "n" is the sample size, "N" is the population, "p" is the response split (yes/no, or male/female) to a two-way question, and "B" is the error rate. Assuming the question has only two possible responses and that the "typical" pattern for response can be predicted, this formula can be used to calculate the theoretical error rate. Typically, we assume p=.5 (50%) implying that we expect the proportion of "yes" to be 50%.

Note that for a large N (Canada or Manitoba), the value of n/N is very close to zero, so the term 1 - n/N becomes very close to 1. If P is expected to be close to .50, then the term $[p(1-p)]^{1/2}$ equals .5, which when multiplied by 1.96 is close to 1. A simple approximation for the theoretical error rate is given by:

$$B = [1/n]^{1/2}$$

A sample of 400 has an approximate error rate 5% 19 times out of 20, a sample of 1,000 yields an error of 3.2%, and a sample of 5,000 produces a rate of under 1.5%. The marginal gain in accuracy from increasing sample size falls as the sample rises with sample size.

The theoretical error rate drops as p diverges from 50%. For example, if p has a typical value of .1 (1-p = .9), the error rate for a sample size of 800 becomes 2.02%.

Note that these are theoretical error rates based upon prior knowledge of the response pattern (p). Researchers usually use a 50/50 split because that gives the highest error rate and, in the face of uncertainty, produces the most conservative estimate.

Each cell in Table 3 shows the error rate (percentage, 19 times out of 20) for various combinations of sample size and population. Notice how little population size (N) influences this error. Recall that these pertain only to a simple yes/no question, where the expected response is 50/50 (i.e., complete uncertainty). For any breakdown (e.g., sex by region), larger samples are required. For regional, gender, or cultural analysis at the national level in Canada, samples of 5,000 are required.

TABLE 3

Error rate (19 times out of 20)				
Population (N)	**Sample size (n)**			
	100	**500**	**1,000**	**5,000**
10,000	9.70%	4.16%	2.79%	0.69%
50,000	9.78%	4.34%	3.04%	1.25%
100,000	9.79%	4.36%	3.07%	1.32%
500,000	9.80%	4.38%	3.09%	1.37%
1,000,000	9.80%	4.38%	3.10%	1.38%
25,000,000	9.80%	4.38%	3.10%	1.39%

Additional readings

Blankenship, A.B., & Breen, George Edward. (1993). *State of the art marketing research.* Chicago: American Marketing Association.

Cochran, William G. (1977). *Sampling techniques.* Toronto: John Wiley & Sons.

Dillman, Don A. (1978). *Mail and telephone surveys: The total design method.* Toronto: John Wiley & Sons.

Peterson, Robert A. (2000). *Constructing effective questionnaires.* Thousand Oaks: Sage.

For additional information, please contact admin@pra.ca