Figure 3.10. SAMPLE SURVEY DESIGN/EXECUTION: *The Literary Digest* Poll

The article EM7601 reprinted below is an account of a famous polling debacle; the author is Maurice C. Bryson, Department of Statistics, Colorado State University, Fort Collins, Colorado, U.S.A. 80523.

EM7601: The American Statistician 30(#4), November, 1976, pages 184-185

The Literary Digest Poll: Making of a Statistical Myth

Pedagogically, there is no more useful or enjoyable device than the horrible example. Not only does it serve a worthwhile purpose, but it gives us an opportunity for an ego-boosting snicker at somebody else's dumb mistake. Thus, statistics books concerned even remotely with the principles of survey sampling rarely miss the opportunity to point out one of history's finest examples of how not to conduct a survey: *The Literary Digest's* presidential poll of 1936.

As most readers know, *The Literary Digest* was a popular magazine of the 1920s and 1930s, which had established a reputation for political prognostication by successfully predicting winners of presidential elections on the basis of "straw polls". In 1936, though, the history of successes came to a crashing halt when the *Digest* predicted a 3-to-2 victory for the Republican nominee, Governor Alf Landon, over the incumbent Roosevelt. Roosevelt, of course, not only won but pulled off one of the greatest landslides of history, winning 62% of the popular vote and carrying 46 of 48 states.

The explanation of *The Digest's* failure, by now a staple of statistical literature, is given at length by Robert Reichard in *The Figure Finaglers*: "The time was late 1935. The opinion ostensibly being measured: the voter's choice for the president – the incumbent, President Roosevelt, or the challenger, Senator Alfred M. Landon of Kansas. Everything was planned impeccably – with a statistically significant number of voters to be called up from all sections of the country. But the planners forgot one basic fact: the use of the phone itself was introducing a bias into the sampling. Remember, this was 1935, and the people who owned phones at that time did not represent a cross section of the American public. Quite the contrary. Telephones were a luxury then – and the people being sampled were the relatively affluent ones – and hence the ones more likely to vote for the Republican candidate" [5].

One of the most widely used sampling texts, by Mendenhall *et al*., repeats the same story, of how "the prediction was in error because more Republicans than Democrats had telephones" [4]. A more general text, by Weinberg and Schumacher, has the story just a bit different: "The *Digest* had made its error in choosing a sample of ten million persons originally selected from telephone listings and from the list of its own subscribers" [8]. Finally, a sort of ultimate word (and a slightly different story) comes from the dean of public opinion experts, George Gallup: "*The Digest's* sample of voters was drawn from lists of automobile and telephone owners" [1].

The stories differ in minor detail, but have one important feature in common: *they are all wrong*! The telephone-survey story is a *myth*; like all myths, it has a germ of truth in it, but like many myths, it misses the real point of what actually happened. Good statisticians should have caught the mistake, since the telephone-survey excuse is inherently implausible. Even in the depths of the depression, phones were not all that unusual, and it is readily estimated that there were about 12 million residential phones [7], representing roughly 40% of the households in the U.S. Since voter participation tends to be highest among the well-to-do, the telephone owners shouldn't have been all that bad as a sample of the *voting* population. Furthermore, consider quantitatively *The Digest's* prediction that Landon would get about 60% of the vote. If he had obtained 60% of the votes of all those with telephones, then – assuming two voters per phone – he would have had a block of over 14 million votes, out of the mere 16 million he actually got. That would leave Roosevelt carrying the non-phone voting population by an incredible 27-to-2 million vote margin. If this were not already implausible enough, one could note that Landon won heavily only in the (non-Southern) rural areas of the country, where telephones were relatively scarce. In such a well-to-do area as Westchester County, New York, presumably well-populated by phone owners, the Landon margin was a modest 51% [6].

These arguments constitute scanty proof, but should be enough to make one suspicious, and to warrant a check of original sources. When we go back to the 1936 *Digest* itself, we find the following:

"Hundreds of astute 'second-guessers' have assured us that the reasons for our error were 'obvious' .... The one most often heard runs something like this: '..... *The Digest*, polling names from telephone books and lists of automobile owners, simply did not reach the lower strata .....' [But] the 'have-nots' did not re-elect Mr. Roosevelt ..... As Dorothy Thompson remarked in the *New York Herald Tribune*, you could eliminate the straight labour vote, the relief vote, and the Negro vote, and *still* Mr. Roosevelt would have a majority ..... Besides – we *did* reach these so-called 'have not' strata. In the city of Chicago, for example, we polled *every third registered voter.* In the city of Scranton, Pennsylvania, we polled every *other* registered voter. And in Allentown, Pennsylvania, likewise other cities, we polled every registered voter .... The fact is that we were as badly off there as we were on the national total .... All this conjecture about our 'not reaching certain strata' simply will not hold water" [9].

So the fact is that the telephone-survey story was not only incapable of explaining the error, but not even correct to begin with. What then did account for the fiasco? The answer, very simply, was *The Digest's* reliance on *voluntary response*. Ten million sample ballots were mailed to prospective voters, but only 2.3 million were returned. As everyone ought to know, such samples are practically always biased. The respondents represent only that subset of the population with a relatively intense interest in the subject at hand, and as such constitute in no sense a random sample. In the 1936 election, it seems clear that the minority of anti-Roose- velt voters felt more strongly about the election than did the pro-Roosevelt majority.

Correction of the telephone-survey myth is important precisely because the myth does conceal the real culprit, voluntary response. Voluntary response to mailed questionnaires is perhaps the most common method of social-science data collection encountered by statisticians, and perhaps also the worst. Somewhat profound decisions are often based on this kind of highly fallacious data. For example, many congressmen (100% of them, based on a non-random sample of five districts where this author has lived) use mailed questionnaires to see how their constituents feel about various issues, and justify subsequent votes on the results. It might not change any votes, but at least knowledgeable people ought to be aware of the irrelevance of such justifications. More generally, one should realize that voluntary response is such a pervasive problem that it may be expected to introduce bias into *any* survey using it. Whether the subject is political preference or university parking policies, the intensely-interested subset is certain to differ from the more apathetic elements of the population.

It would be nice to conclude this discussion by identifying the original perpetrator of the myth but, like most myth-makers, he or she has been lost in the shrouds of history. A careful survey of the *Reader's Guide to Periodic Literature* reveals no hint of the

(*continued overleaf*)

telephone-survey story (other than *The Digest's* own report quoted above) until 1948, when a *Scientific American* article by Rensis Likert reported an approximate version of the truth: "First, the poll was restricted to *Literary Digest* and telephone subscribers. Second, it obtained a biased sample of those subscribers, *i.e.*, only those people who answer mailed questionnaires" [3]. But by 1954, the current version of the myth was being reported as fact by Darrell Huff in his popular *How to Lie With Statistics.*. Huff referred only to "the ten million telephone and *Digest* subscribers [who] came from the list that had accurately predicted the 1932 election" [2]. The true problem, that of the non-random selecting of 2.3 million respondents out of the 10 million, had been lost. In the interests of good statistical procedure as well as accurate reporting of history, it is well that it should be found again.

REFERENCES

[1] Gallup, George: "Opinion Polling in a Democracy, in J. Tanur *et al.*, *Statistics: A Guide to the Unknown*, San Francisco: Holden-Day Publishers, 1972, 146-152.

[2] Huff, Darrell: *How to Lie With Statistics*, New York: W. W. Norton and Co., 1954.

[3] Likert, Rensis: "Public Opinion Polls," in *Scientific American*. 179; December 1948, 7-11.

[4] Mendenhall, William, Lyman Ott and Richard L. Schaeffer: *Elementary Survey Sampling*, Belmont, California: Wadsworth Publishing Co., 1971.

[5] Reichard, Robert S.: *The Figure Finaglers*. New York: McGraw-Hill, 1974.

[6] Scammon, Richard M.: *America at the Polls*, Pittsburgh: Pittsburgh University Press, 1965.

[7] United States Census Bureau: *Statistical Abstract of the United States*, Washington: U.S. Government Printing Office, 1942.

[8] Weinberg, George H. and John A. Schumacher: *Statistics: An Intuitive Approach*, Belmont, California: Wadsworth Publishing Company, 1962.

[9] "What went wrong with the polls?" *Literary Digest* 122, November 14, 1936, pp. 7-8.

1 The word *error* epitomizes the discussion of the article reprinted overleaf on page 3.19 and above; as discussed in detail in Statistical Highlights #6 and #18 for example, *we* use this word *only* with a well-defined technical meaning and we usually distinguish *four* components of the *overall* error in an investigation like a sample survey to answer a Question with a descriptive aspect. The article EM7601 uses *error* four times:

∗ ..... the prediction was in *error* because ..... [last paragraph of Column 1];
∗ *The Digest* had made its *error* ..... [last paragraph of Column 1];
∗ ..... the reasons for our *error* were ‘obvious’ ..... [last paragraph of Column 2];
∗ ..... incapable of explaining the *error*, ... [second paragraph of Column 3];

- Explain briefly the meaning of *error* in each of these statements.
- From a statistical perspective, what is undesirable about such statements? Explain briefly.

2 In the third and fourth paragraphs overleaf on page 3.19, explanations are given from four sources (Reichard, Mendenhall *et* al., Weinberg and Schumacher, and Gallup) for *The Literary Digest's* substantially wrong estimate of a population proportion; characterize each of these explanations in terms of the component of overall error that is said to be involved.

- Why are such characterizations useful? Explain briefly.
- Which component of overall error does the *author* of the article give as the explanation? Explain briefly.
  - What evidence does the author cite with regard to *study* error as the explanation? Explain briefly.
- Who is mentioned in the article as giving an explanation in terms of *two* components of error? Explain briefly.

3 In the paragraph at the top of the third column overleaf on page 3.19, *The Literary Digest* cites their high sampling fractions in a number of cities and comments that ..... *we were as badly off there as we were on the national total*. Comment on the implications of this information for the answers to the preceding Question 2.

4 The phrase *voluntary response* (as a synonym for non-response) is used four times in the article EM7601:

∗ ..... the *Digest's* reliance on *voluntary response*..... [second paragraph of Column 3];
∗ ..... the real culprit, *voluntary response* ..... [third paragraph of Column 3];
∗ *Voluntary response* to mailed questionnaires ..... [third paragraph of Column 3];
∗ ..... one should realize that *voluntary response* ..... [third paragraph of Column 3];

- Comment critically on the use of this phrase in such statements.

5 Comment critically in context on the statement in the third paragraph of the third column overleaf on page 3.19: ..... *many congressmen (100% of them, based on a non-random sample of five districts where this author has lived) use* .....

6 Comment critically on the statement near the bottom of the third column overleaf on page 3.19: *More generally, one should realize that voluntary response is such a pervasive problem that it may be expected to introduce bias into any survey using it.*

7 The last sentence of the third paragraph near the bottom of Column 3 overleaf on page 3.19 is: *Whether the subject is political preference or university parking policies, the intensely-interested subset is certain to differ from the more apathetic elements of the population.* Comment critically on using this statement to argue that those who feel strongly about an issue form the population whose attribute we want to estimate and so non-response is *not* of appreciable concern.

- Where in the article EM7601 does the author use an argument somewhat like this? Explain briefly.

The article EM7601 reprinted overleaf on page 3.19 and above is also used in Figure 8.7b of the STAT 220 Course Materials, in Figure 3.4b of the STAT 332 Course Materials and in Statistical Highlight #17.