Errors in Sample Surveys

Questions to ask before you believe in a poll

What the margin of error doesn’t say
1.0 Errors in Sample Surveys

■ The sole measure of the worth of a sample survey is the reliability with which estimates based on it reproduce the truth about the population. Implicit in this evaluation is the idea of repeatedly sampling.

■ Discrepancies of $\hat{p}$ from $p$ could be due to:
  ▶ variability
  ▶ bias

■ Variability in $\hat{p}$ is caused purely by the act of taking a sample.

■ Bias in $\hat{p}$ is caused by under-coverage of the population and/or problems during data collection, such as non-response.
1.1 Example

Identify whether repetition of the following scenarios will result in variability and/or bias in \( \hat{p} \).

- A librarian randomly selects 100 titles from the library data base to estimate the proportion that start with the letter “A”.

- An opinion poll to assess the proportion of Seattle-ites who support same-sex marriage uses the telephone directory as the sampling frame. This misses all those people who are not listed in the phone book.

- On a survey to estimate the proportion of teens who are sexually active, embarrassed students deliberately give the wrong answer.
2.0 Questions to Ask Before You Believe a Sample Survey

- Who carried out the survey?
- What was the population of interest and what was the sampling frame?
- How was the sample selected?
- How large was the sample?
- What was the response rate? (number of individuals who responded/ number of individuals who were asked)
- How were the interviews conducted?
- When was the survey conducted?
- What were the exact questions asked?
2.1 Example

Ann Landers once asked readers of her advice column: if you had to do it all over again, would you have children? She received over 10,000 responses with 70% saying a “no”. Would you trust this result as an estimate for the proportion of Ann Landers’ readers who regret their decision to bear a child?
2.2 Example

A 1993 Roper poll found that 1 out of 3 Americans are open to doubt there was a holocaust. The poll surveyed 992 adults at random and asked them the question:

*Does it seem possible, or does it seem impossible to you that the Nazi extermination of the Jews never happened?*

Would you trust this result as an estimate for the proportion of Americans who doubt there was a Holocaust? Why or why not?
A survey is carried out by the Finance Department to determine average household size in a certain city. They draw a S.R.S. 1,000 households. After several visits, the interviewers only find people at home in 653 of the sample households. Rather than face such a high non-response rate, the department draws a second batch of households and uses the first 347 completed interviews in the second batch to bring the sample back to its planned size. Using this data, they estimate average household size in the city to be 3.1 persons. Would you trust this result? Why or why not?
3.0 What the Margin of Error Doesn’t Say

Caution

The reported margin of error for a sample survey that uses probability methods only gives the magnitude of random sampling variability. Under-coverage, non-response error and other practical difficulties can cause a large bias in our result that is not covered by the margin of error.
3.1 Example

- When the Current Population Survey asked the adults in its sample of 50,000 households if they voted in the 2000 Presidential election, 55% said they had. The margin of error was less than 1%. Interpret the margin of error.

- Post election, it was reported that only 51% of the adult population voted in that election. Why do you think the C.P.S. result missed by more than the margin of error?