

Survey Sampling (Presidential Polls)

Sources: D. Freedman, R. Pisani, R. Purves. *Statistics*. 3rd ed. New York, Norton, 1998; <http://poll.gallup.com>; <http://www.presidenceselect.org/>

Literary Digest correctly predicted winner of presidential election in 1916, 1920, 1924, 1928, and 1932 by mailing questionnaires.

1936 Literary Digest Poll

Questionnaires mailed to 10 million people, 2.4 million responses.

Literary Digest Prediction: Landon 57%, Roosevelt 43%.

Actual Results: Roosevelt 60.6%, Landon 36.8%.

Problem: The sample was not representative. Names to whom surveys were sent came from phone books, club membership lists. Poor were undersampled.

1936 Gallup Poll

Gallup Prediction: Roosevelt 55.7%, Landon 44.3%.

Actual Results: Roosevelt 60.6%, Landon 36.8%.

Quota sampling: Interviewers were assigned to survey a specific number of people in different categories, based on race, gender, age, and income. This gave better results than the Literary Digest survey.

1948 Gallup Poll

Gallup Prediction: Dewey 49.5%, Truman 44.5%.

Actual: Truman 49.8%, Dewey 45.1%.

Problem: Convenience sampling. Interviewers surveyed assigned number of people in certain categories, but within categories could choose arbitrarily.

1952-2004 Gallup Poll

Year	Winner	Gallup	Actual	Error
1952	Eisenhower	51.0	54.9	3.9
1956	Eisenhower	59.5	57.6	1.9
1960	Kennedy	49.0	49.7	0.7
1964	Johnson	64.0	60.6	3.4
1968	Nixon	43.0	43.4	0.4
1972	Nixon	62.0	60.3	1.7
1976	Carter	49.0	50.1	1.1
1980	Reagan	47.0	50.8	3.8
1984	Reagan	59.0	59.2	0.2
1988	Bush	56.0	53.4	2.6
1992	Clinton	49.0	43.0	6.0
1996	Clinton	52.0	49.2	2.8
2000	Bush	48.0	47.9	0.1
2004	Bush	49.0	50.7	1.7

In 1952, the Gallup poll began choosing people at random, and since then has predicted the winner correctly in all 14 presidential elections. In 2004, the sample size was only 2014.

Random sampling

- 1) **Simple random sample.** Choose n people, every sample of size n equally likely to be chosen.

- 2) **Stratified random sample.** Divide the population into groups called strata, then do simple random sampling in each stratum. (Example: sample 500 men and 500 women rather than any 1000 people.) This can reduce variability.

- 3) **Cluster sample.** Divide the population into clusters. Select a few clusters at random and sample only from those selected. (Example: exit polling is only done at a few polling stations.) This can reduce cost.

Nonresponse bias

In 2004, exit polls predicted Kerry to win in Ohio, Florida. Instead, Bush won both states and consequently the presidency. Republicans more likely to decline to participate in exit polling.

Voluntary response sampling

Voluntary response sample: many people invited to respond. All who respond are counted.

Surveys done through the internet or radio talk shows suffer from voluntary response bias, have no scientific value.

Wording of the question

Time magazine poll on 4/5/1993 reported favorable ratings for:

Hillary Clinton	57%
Hillary Rodham Clinton	49%

Wording of question can substantially alter results.