

Forecast

In Presidential Elections :

What Is Wrong With

The Polls ?

Hanxiang Peng
Department of Mathematics
University of Mississippi
February 2006

The 2000 election is the CLOSEST modern contest by both measures: the popular and the electoral votes.

ELECTORAL (270 NEEDED)

George W. BUSH	271
AL GORE	267
Difference	4

POPULAR

AL GORE	50, 158, 094
GEORGE W. BUSH	49, 820, 518
Difference	201, 596

Source: The New York Times

FORECASTS BY FOUR MAJOR POLLS

ABC News Poll	BUSH	GORE
Oct. 29-31	49%	45%
Error +/- 3 pct. pts		

CBS News Poll	BUSH	GORE
Oct. 29-31	44%	43%
Error +/- 4 pct.pts		

CNN/USA Today/Gallup Poll	BUSH	GORE
Oct. 29-31	48%	43%
Error +/- 2 pct.pts		

Pew Research Center Poll	BUSH	GORE
Oct. 25-29	47%	43%
Error +/- 3 pct.pts		

Source: The New York Times

HOW THE MODERN POLL IS CONDUCTED

Here is how the New York Times / CBS News Poll conducted the latest poll.

- **Randomly** selected the sample of telephone exchanges by a *computer* from a **complete** list of more than 42, 000 residential exchanges across the country.
- With each exchange, **random** digits were added to form a **complete** telephone number, thus permitting access both to listed and unlisted numbers.
- With each household, one adult was designated by a **random** procedure to be the respondent of the survey.

- 1,279 adults throughout the US were interviewed by telephone Wednesday through Saturday.

In addition, the results have been **weighted** to take account of *household size* and *number of telephone lines* into the residence and to adjust for variations in the sample relating to *geographic region, sex, race, age and education*.

This is a typical example of how modern polls are conducted. The method is called **Probability Method**.

Source: The New York Times

However, they still failed to forecast! So, what is wrong with these polls?

POSSIBLE REASONS

A **representative** sample is essential to guarantee a good forecast. So, we may ask:

- The sample not representative. (How can we know it is or it is not?)
- The Probability Method not good. —The selection of sample is left to **CHANCE**, instead of Human's judgment. (Is there other better method? for example, hand-pick the sample that resembles the population?)
- The sample not large enough. (Can we improve the forecast by a larger sample?)
- Non-respondents. (Were the respondents and non-respondents different?)

- Large number of voters turned out in the last minutes otherwise wouldn't.
- Voters changed their minds (e.g., from for Democrat to for Republican).
-

1936 ELECTION: ROOSEVELT VS LANDON

The *Literary Digest* poll:

- based on 2.4 million responses—the largest number of people ever replying to a poll — out of 10 million questionnaires sent.
- backed by the enormous prestige of the *digest*—which had called the winner since 1916.

	Roosevelt's %
Election result	62%
Digest prediction	43%

(The *Literary Digest* went bankrupt soon after!)

Source: Freedman, Pisani and Purves. *Statistics. Third Edition.*

GALLUP'S PREDICTION

Just setting up his survey organization, George Gallup, by using a different method, was able to

- predict the *Digest* prediction using its sample, well in advance of its publication.
- correctly forecast the Roosevelt victory, based only on his 50, 000 samples.

	FDR's %
Election result	62%
Digest prediction	43%
Gallup's prediction of <i>Digest</i> prediction	44%
Gallup's prediction	56%

Source: the same as above.

WHAT WENT WRONG WITH THE *DIGEST*?

- Its 10 million addresses came from sources like telephone books, club membership lists.
- Remember, at the time, only one household in four had a telephone.
- ⊙ The poor were screened out.

So there was a strong **BIAS** against the poor in the *Digest's* sampling procedure. We call it **SELECTION BIAS**.

— Taking *large* sample does not help. This just repeats the same mistake on a larger scale.

— Then, *why did it predict correctly before 1936?*

Prior to 1936, this bias may not have affected the predictions very much, because the poor and the rich voted along a similar line. But in 1936 they did not: the poor voted for Roosevelt while the rich for Landon.

This is just its first blunder in sampling. There is a second blunder.

In the *digest* poll, only 2.4 million people bothered to reply, out of the 10 million who got questionnaires.

— *Were the non-respondents and the respondent different?*

— No, they were not! The non-respondents did NOT respond. But...

The *Digest* made a special survey in Chicago. About 20% responded, and of those who responded over half favored Landon. But it turned out in the election Chicago went for Roosevelt, by a two-to-one margin. So

—Non-respondents can be very different from respondents. This is **non-respondent bias**.

The *Digest* poll was spoiled both by **selection bias** and **no-respondent bias**.

—To find out whether a sample is any good, ask how it was chosen. Any **selection bias**? any **no-respondent bias**? We may not be able to answer these questions just by looking at the data.

HOW DID GALLUP FORECAST THE DIGEST PREDICTION?

He just chose 3, 000 people **at random** from the same lists the *Digest* was going to use, and mailed them all a postcard asking how they planned to vote. He knew that a random sample was likely to be quite representative. We will discuss it in a while.

THE YEAR THE POLLS ELECTED DEWEY

Three major polls had declared Dewey the winner, with a lead of 5% points. However, the statistics didn't convince the American public.

The Election of 1948

	Truman	Dewey
Results	50%	45%
Crossly	45%	50%
Gallup	44%	50%
Roper	38%	53%

source: as above What went wrong for the polls? We need to know who the samples were picked.

—All three major polls used the same method: hand-picked the sample that resemble the population with respect to some key characteristics.

The method is called **Quota Sampling**.

QUOTA SAMPLING

How did it go?

Each interviewer was assigned a fixed quota of subjects to interview; the number falling into certain categories (like residence, sex, age, economic status, etc.) were also fixed. In other respect, the interviewers were free to select anybody they like.

For example, A Gallup Poll interviewer in St. Louis was required to interview 13 subjects, of whom

- exactly 6 were to live in the suburbs, and 7 in the central city.
- exactly 7 were to be men, and 6 to be women.

Of the 7 men (similar for women):

- * exactly 3 were under forty years old, and 4 over forty.
- * exactly 1 was to be black, and 6 white.
- *

The method looks good. However, the 1948 experience shows it worked very badly.
Let us see why.

- No quotas can be set to faithfully represent the nation's political opinions.

Many factors influence the voting behavior besides the ones the survey organizations control for.

- Unintentional bias on the part of the interviewers.

With the assigned quotas, the interviewers were **free** to choose. However, human choice is always subject to bias. Interviewers preferred Republicans in every presidential election. So too many Republicans were chosen.

PROBABILITY METHOD

What is a probability method?

—Say, a small town has 1, 000 eligible voters. Write the name of each voter on a ticket and put it in a box, and draw 100 from the box (shake the box after each draw, a ticket drawn out will not be returned to the box).

This method is called **simple random sampling**. It is the basic. All other methods can be quite complicated. But all of them have two important features:

- the interviewers have not discretion at all to whom they interview.
- there is a definite procedure to select the sample. It involves the planned use of chance.

HOW WELL DO THE PROBABILITY METHODS WORK?

From 1948 to 1992, the Gallup and many other major polls have used the probability methods to choose their samples. They have been able to predict the elections. Three points to notice

- the sample size has gone down sharply.
- no longer any consistent trend favoring either Democrats or Republicans.
- the accuracy has gone up appreciably.

WHY DID PROBABILITY METHODS WORK SO WELL?

- At first glance, it may seem that judgment is needed to choose sample.
- For instance, quota sampling gives that the percentage of men in the sample will be equal to that of men in the population.
- With probability sampling, certainty is reduced to likelihood.
- But judgment and choice usually show bias, while chance is impartial. This is why probability methods work better than judgment.

TELEPHONE SURVEYS

- Modern polls are now conducted by telephone. The savings in costs are dramatic, and the results were good from 1948 to 1992 (not including the 2000).
- In theory, 19 cases out of 20 the results based on such samples will differ by no more than three percentage points in either direction from what would have been obtained by seeking out all American adults. So unfortunately, this 2000 could have been fallen into this 3%!?

So the possible reasons for the false forecast of the 2000 election could be the last two reasons. (I can only dig out these reasons).

DIFFICULTIES IN REALITY WITH THE SIMPLE RANDOM SAMPLING

- Impossible to make a list of a large number of people, say 200 millions.
- Can't make those draws (Remember each ticket is equally likely to be drawn).
- Prohibitively expensive to send interviewers to all the country.

MULTISTAGE CLUSTER SAMPLING

So most survey organizations use a probability method called **multistage cluster sampling**. For example, the modern polls is conducted with the multistage method.

- The complete list of 42, 000 residential exchanges across the country is the first stage.
- A household is an intermediate stage.
- The randomly designed respondent in a household is the last the stage.