How Do Statistical Polls Work?

Example #24 — A 2-foot × 2-foot × 2-foot box contains 150,000 black and white marbles. Suppose you had to estimate the proportion of marbles in box that are black. What would you do?

Well, you could reach into the box with both arms and mix up the marbles and draw a random sample of 500 marbles from all areas of the box. You count the marbles in your sample and 215 (43%) of the 500 marbles are black. Given this sample statistic, what can you say about the true proportion of black marbles in the box?

- Your best estimate of the true proportion of black marbles in the box is the sample proportion: 43%
- How accurate is your estimate? If you did a good job of mixing the marbles and drawing a random sample of 500 marbles from the box, then statistics can tell you how accurate your estimate is — expressed as a **Margin of Error** (or sampling error).

Sample Size	95% MOE	95% Confidence Interval around 43%	99% MOE	99% Confidence Interval around 43%
200	6.9%	36.1% – 49.9%	9.1%	33.9% – 52.1%
500	4.4%	38.6% – 47.4%	5.8%	37.2% – 48.8%
1000	3.1%	39.9% – 46.1%	4.1%	38.9% – 47.1%
1500	2.5%	40.5% – 45.5%	3.3%	39.7% – 46.3%

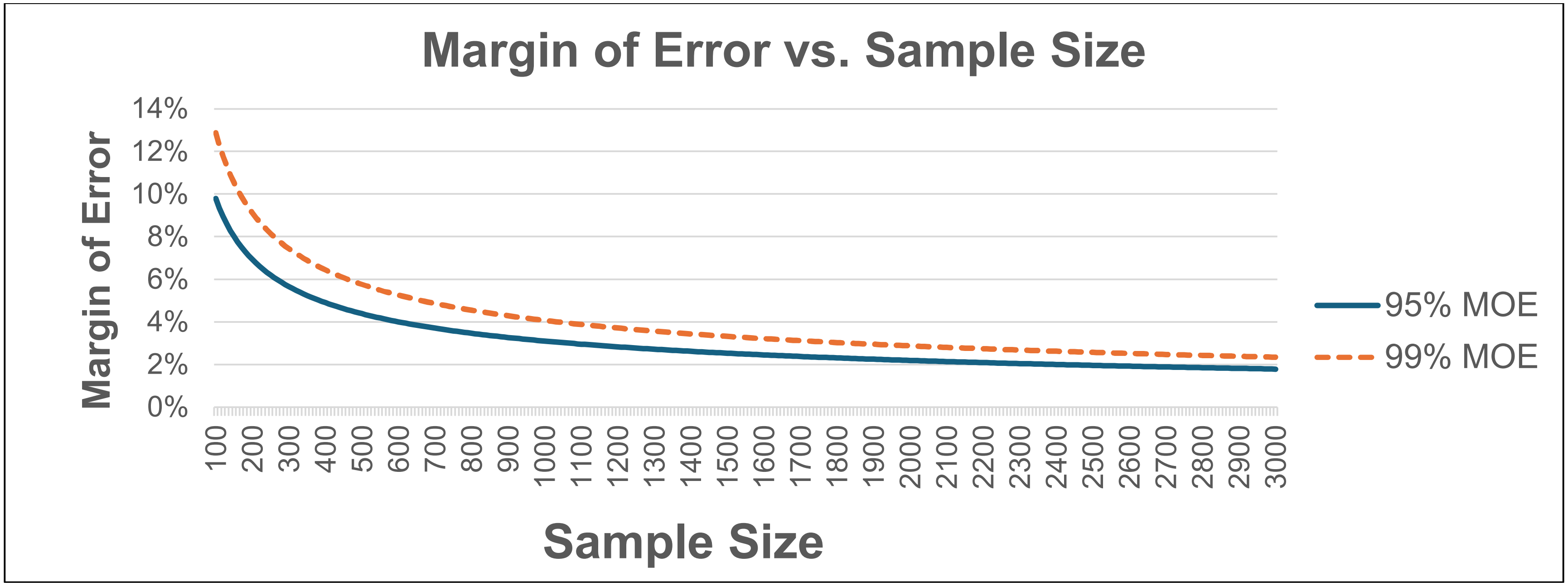
- 95% Margin of Error = $1.96 \times \frac{0.5}{\sqrt{N}} = 1.96 \times \frac{0.5}{\sqrt{500}} \approx 4.4\%$

  95% probability that the true proportion is 43% ± 4.4%

- 99% Margin of Error = $2.576 \times \frac{0.5}{\sqrt{N}} = 2.576 \times \frac{0.5}{\sqrt{500}} \approx 5.8\%$

  99% sure that the true proportion is 43% ± 5.8%

This example of estimating the proportion of black marbles in a box is analogous to how a statistical poll works!



- Most statistical polls have a 95% Margin of Error between 2–5%. For this reason, sample size for most polls is about 500 – 1500.

<u>Statistical Polls in the News</u>

- A Reuters/Ipsos poll (April 2025) of 1,029 U.S. adults nationwide showed an approval rating for President Trump of 41%. The poll had a margin of error of about 3 percentage points.

- In a CBS News/YouGov poll (March 2025) of 2351 U.S. adults, 61% of respondents said artificial intelligence will have a bigger impact on society than Internet. The margin of error is ±2.5%

- A Washington Post poll (June 2025) of 1000 people finds Americans opposing U.S. airstrikes against Iran by a 20-point margin — 45 percent to 25 percent — with 30 percent unsure. Overall margin of error is ± 3.6 percentage points.

Sampling Error (Margin of Error) and Survey Bias

There are two main sources of error in a survey statistic: sampling error (margin of error) and survey bias. A well-designed statistical survey should:

- Have a large enough sample size to achieve the desired sampling error — typically 2 – 5 percentage points. Sampling error **can** be reduced by increasing the sample size.

- Minimize survey bias, which is tendency of a sample statistic to over- or under-estimate a population parameter. Increasing the sample size **cannot** correct for survey bias. Sources include:

    - Under-coverage — when some segments of population are under-represented in the sample — e.g. a "convenience sample," such as sampling marbles from only the top of box; the Literary Digest poll for the 1936 presidential election
    - Non-response bias — when certain individuals chosen for the sample are unwilling or unable to participate — e.g. full-time workers who are too busy to participate
    - Voluntary response (self-selection) bias — when sample members are self-selected volunteers — e.g. a radio call-in show that solicits audience participation in a survey
    - Response bias — when survey participants do not tell the truth for some reason — e.g. when the wording of the question favors one response over another, or responses are biased toward what people believe is socially accepted

What Does "Statistically Significant" Mean?

- A difference is "statistically significant" if it is unlikely due to random chance — i.e. the difference is real

- 1000 people with a disease are randomly divided into treatment and control groups to test a new drug vs. a standard drug. 74% of the treatment group recovers from disease vs. 66% of control group. Difference is statistically significant — i.e. very unlikely due to random chance (less than 1% chance), meaning that the new drug really is more effective than the standard drug.
- AP-NORC Poll of 1200 adults: 40% in June 2025 said we were on "right track" vs. 25% in September 2025. This difference is statistically significant — i.e. not likely due to chance (less than 1% chance), meaning a real change in the opinion of adults.
- To test two ad campaigns, a sample of 200 users is exposed to Campaign A and 200 users to Campaign B. Conversion rates are 17% and 21%. The difference is not statistically significant — could have plausibly occurred due to random chance.

<u>Example #25.1</u> — You are told a coin is fair with Pr (*Heads*) = ½. To test whether the coin is fair, you flip the coin 500 times and get 500 × 43% = 215 *Heads*. Do you still believe the coin is fair?

Key Question: How likely is it that only 43% of 500 coin tosses would be *Heads* if the coin were fair?

- Per Example #23: for sample size = 500, 99% Confidence Interval for the true proportion of *Heads* is 37.2% – 48.8%
- Note that: 99% Confidence Interval does **<u>not</u>** include 50%
- It is very unlikely (less than 1% chance) that 43% of 500 coin tosses would be *Heads* by random chance if the coin were fair
- Hence, **statistically significant** at the 1% level of significance
- Decision: The coin is loaded (not fair) — i.e. we reject the hypothesis that the true probability of *Heads* = 0.5
- But we could be wrong: there is a very small chance (≈ 0.18%) that 500 flips of fair coin could get 43% *Heads* (or worse)

<u>Example #25.2</u> — Now suppose you flip the coin 500 times and get 500 × 47% = 235 *Heads*. What do you decide now?

- For sample size = 500, the 95% Confidence Interval for the true proportion of *Heads* is 47% ± 4.4% (42.6% – 51.4%)
- Note that: 95% Confidence Interval **<u>does</u>** include 50%
- Hence, **<u>not statistically significant</u>** at 5% level of significance
- Decision: The coin is fair (not loaded). By convention, if a difference is not statistically significant at 5% level, then it is generally considered to be **<u>not</u>** statistically significant.
- Note that we are **<u>not</u>** saying that the coin is definitely fair. Rather we are saying that a 47% sample proportion (on 500 tosses) could have plausibly occurred due to random chance (with probability ≈ 18%), even if coin were fair. We don't have enough statistical evidence to say that the coin is loaded.

<u>Example #26.1</u> — Box #1 and Box #2 each contain around 150,000 black and white marbles. Suppose you had to answer the following question: Is the proportion of black marbles in Box #1 greater than that in Box #2? What would you do?

You could draw a random sample of 500 marbles from each box. Suppose that 74% of the 500 marbles sampled from Box #1 are black and 66% of 500 marbles sampled from Box #2 are black.

- Using a statistical test: it is highly unlikely (less than 1% chance) that a difference as great as 74% vs. 66% could occur by chance if the proportion of black marbles in Box #1 were **less than** or equal to Box #2
- Hence, the difference between the sample proportions is statistically significant at the 1% level of significance
- Decision: The overall proportion of black marbles in Box #1 is **greater than** that in Box #2
  (But there is a small chance ≈ 0.3% that we could be wrong)

Example #26.2 — Suppose that 70% of the 500 marbles from Box #1 are black and 66% of 500 marbles from Box #2 are black

- Not statistically significant at 5% level: a difference of 70% vs. 66% could plausibly occur by chance (≈ 8.7%), even if overall proportion of black marbles in Box #1 were **less than** Box #2
- Decision: Even though the sample proportion of black marbles in Box #1 is greater than that in Box #2, we do **not** have sufficient statistical evidence to conclude that the overall proportion of black marbles in Box #1 is greater than that in Box #2

Example #26.3 — 1000 people with disease are randomly divided into treatment and control groups to test a drug. 74% of the treatment group recover from disease vs. 66% of control group. Is there statistical evidence that the drug helps to cure the disease?

- This is analogous to Example 26.1. Yes, there is statistical evidence that the drug does help to cure the disease.

<u>Example #27</u> — Just because a result is statistically significant does not necessarily mean it is "practically significant" or important

- A coin is tossed a million times, resulting in 50.1% *Heads*. Because of the extremely large sample size, this is statistically significant at the 5% level of significance. But this difference is not meaningful. You would still say that the coin is fair.

- In a large clinical trial of a drug with 10,000 people, 61% of the treatment (drug) group recovered from a certain disease vs. 59% for the control (placebo) group. This difference is statistically significant at 5% level, but not "clinically significant."

- In a large clinical trial of a drug to test for side effects, the drug doubled the risk of cancer: 0.2% of the treatment group developed a rare form of cancer vs. 0.1% for the control group. Even if this difference were statistically significant, it may be of little importance because the disease is so rare in either case.