Classroom Simulations Using R: Margin of Error in a Public Opinion Poll

Bruce E. Trumbo
Department of Statistics, CSU Hayward
www.sci.csuhayward.edu/~btrumbo
“Margin of Error” in a Public Opinion Poll

Many polls quote a margin of sampling error — depending on the number $n$ of people sampled:

- $\pm 2\%$ for $n = 2500$
- $\pm 3\%$ for $n = 1100$

What is the basis for these numbers?
These Margins of Error
Ignore Subjective Issues

- Was the correct population sampled?
- How to interpret “refusals”?
- How to allocate “undecideds”?
- Questions understood?
- Truthful answers given?
- Many other crucial issues.
Margins of Error Are Based on Principles of Probability

- Assume a random sample from the population used.
- When subjective factors enter: They can only make the margin of error larger — never smaller.
Simulation Illustrates Probability Rules

Requires faith that a computer can take a random sample from a specified population.

In practice, simulation of the scope shown in some examples here requires very high-quality statistical software. (Apologies to Bill Gates: Excel® — without first rate statistical add-ins — won’t do.)
Our Simulations Use R

- Based on S, developed at Bell Labs
- Many built-in probability/statistics functions.
- Best random number generators known.
- Runs very fast (on today’s PCs: at the “supercomputer speed” of 1980s)
- Widely used in research / industry / biotech
- Best of all in a budget crisis: Free software.
Two Preliminary Examples of “Sampling” With R

1. A *poker hand* (cards numbered 1 to 52):

   \[
   \text{sample}(1:52, 5, \text{rep}=F)
   \]

Example of result:

\[
> \text{sample}(1:52, 5, \text{rep}=F)
\]

[1] 16  3 33 17 51

Interpret this as: 3♦, 3♥, 7♣, 4♦, Q♠
2. *Toss a fair coin* 10 times:

```r
sample(0:1, 10, rep=T)
```

Example of result:

```r
> sample(0:1, 10, rep=T)
[1] 1 0 0 1 0 1 0 0 0 0 1
```

Interpret this as

H T T H T H T T T H

4 Heads in 10 tosses, so the proportion of Heads is 4/10 or 0.4 or 40%.
Simulating an Election Poll

To simplify, suppose:

- No undecided voters
- 53% of voters favor Candidate “A”
- 47% of voters favor Candidate “B”
- Randomly sample 25 voters.

Can we **detect from data** for 25 subjects that Candidate “A” is in the lead?
R code:

```r
sample(0:1, 25, rep=T, prob=c(.47, .53))
```

One result:

```r
> sample(0:1, 25, rep=T, prob=c(.47, .53))
[1] 1 0 1 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

Interpret 0s & 1s as candidate preferences

A B A A B B A B B A B B B
B A B A B A A A B A A B

B A B A B A A A B A A B

A B
What can we conclude from this survey of 25 subjects?

Poll result after 25: “A” has 12 in favor.  
After 25 subjects: Proportion = 48% for “A”.
Contradicts what we know about population. 
Intuitively, we know 25 is not a large enough number. 
How to look at this more objectively?
Find Proportions After Each Subject 1-25

Table Below Used to Find 1\textsuperscript{st} Through 5\textsuperscript{th} Proportions

<table>
<thead>
<tr>
<th>Subj. so Far</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opinion</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>A’s so Far</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Prop. so Far</td>
<td>1/1 = 1.00</td>
<td>1/2 = 0.50</td>
<td>2/3 = 0.67</td>
<td>3/4 = 0.75</td>
<td>3/5 = 0.60</td>
</tr>
</tbody>
</table>
Now plot the remaining 20 subjects.
Endpoint is 48%. Next, look at additional polls.
Unstable Trace of a 25-Subject Poll

Note: Here and below, vertical axis from 25% to 75% to help show detail.
Unstable Trace of a 25-Subject Poll

Proportion Favoring 'A'

Number of Subjects
Unstable Trace of a 25-Subject Poll
Unstable Trace of a 25-Subject Poll
Results Differ Widely

As we repeat the 25-subject poll over and over again, a collection of quite different results appeared.

Next: Superimpose on one graph the traces of 20 polls, each with 25 subjects. We see several different endpoints, most with values above 50%, but some below.
Focus on Endpoints — How Variable Are They?

Simulate 1000 polls with 25 subjects each. Results range from about 25% to 85%; 377 have misleading endpoints below 50%. A 25-subject poll cannot reliably distinguish between 53% and 50% for Candidate “A”. Here is a histogram of the 1000 endpoints.
Order from Randomness: Pretty Good Fit to Bell Curve

Density

0 1 2 3 4

0.2 0.3 0.4 0.5 0.6 0.7 0.8

Proportion Favoring 'A'
Polls with $n = 2500$ Subjects

Major polling organizations report such polls as having a $\pm 2\%$ margin of sampling error.

**Intended Meaning:**

If the true proportion for “A” is 53%, then most 2500-subject polls (95%) will have end points between 51% and 55%, thus detecting that “A” is the favorite.
Stability and Reproducibility

The following simulated traces show that 2500-subject polls are:

- **Stable:** By $n = 2500$ the trace of the proportion for “A” has settled to a nearly constant value.

- **Reproducible:** Repeated simulations give traces with endpoints near 53%.
Ultimately Stable Trace of a 2500-Subject Poll

Proportion Favoring 'A'

Number of Subjects
Ultimately Stable Trace of a 2500-Subject Poll

Proportion Favoring 'A'

0.3 0.4 0.5 0.6 0.7

Number of Subjects

0 500 1000 1500 2000 2500
Ultimately Stable Trace of a 2500-Subject Poll
Ultimately Stable Trace of a 2500-Subject Poll

Proportion Favoring 'A'

Number of Subjects
Red marks at right indicate 53% ±2%. Green line at 50%.
An “average” run: 19 out of 20 traces (95%) within margin of error.
Endpoints of 1000 Polls, n=2500: With Bell Curve

95% of endpoints within margin of error; only 3 in 1000 below 50%.
In interval (.51, .55): Black curve has 95% probability, Blue curve only about 12%
Law of Large Numbers

The Law of Large Numbers is a mathematical theorem.

Applied to polls:
By choosing a big enough $n$, the margin of error for estimating the Proportion Favoring “A” can be as small as we like.
In practice we can quantify this:

The margin of error for a “close race” is

\[ E = n^{-1/2} = \frac{1}{\sqrt{n}}. \]

For example, as we have seen,

\[ E = 2\% \text{ for } n = 2500 \]

The Fine Print:

- This margin of error applies to 95% of polls, based on bell curves (normal distributions).
- Don’t use the formula for \( n < 100 \).
- \( E \) for differences (“A’s lead over B”) and subgroups (“women”) are always larger.
Other Kinds of Random Processes

Stability: Some random processes never settle down to a target value, no matter how large the sample size.

Example: Here we model the position of a particle subject to a certain kind of random bombardment. Small dots show its position (vertical axis) at each point in time.
From an example in a forthcoming book by E. Suess and B. Trumbo (Springer).
Reproducibility: Some random processes have *more than one* target value.

*Example*: The endpoint of Brother-Sister Mating (relax, it’s a botanical example) can be either of two “homozygous crosses,” which we label 1 and 6.

Starting from $3 = (Aa \times Aa)$: endpoints $1 = (AA \times AA)$ and $6 = (aa \times aa)$ are equally likely.
From an example in a forthcoming book by E. Suess and B. Trumbo (Springer).
Limiting Distributions

If a process does not settle to a point, then it may (or may not) settle to a distribution.

Sometimes, the limiting distributions can be surprisingly complex. (Not always a bell curve!)

Here are two examples: The 1\(^{st}\) of practical importance, the 2\(^{nd}\) mainly recreational.