Sanders Leads Buttigieg, Biden in New Hampshire

Candidate	Percent
Bernie Sanders	24%
Pete Buttigieg	15%
Joe Biden	15%
Elizabeth Warren	10%
Amy Klobuchar	6%
Tulsi Gabbard	5%
Tom Steyer	5%
Andrew Yang	3%
Michael Bennet	0%
Deval Patrick	0%

Source: Suffolk University poll. Feb 3-4, 2020.
Margin of error for 500 likely voters is +/- 4.4%.

October 1948 Polling Results

Candidate	Crossley Poll	Gallup Poll	Roper Poll
Thomas Dewey	50%	50%	53%
Harry Truman	45%	44%	38%

- Impact of *quota sampling* misrepresented the strength of the Republican vote
- Operating under then-current view that voters decided by Labor Day, all major polling organizations ceased polling two weeks prior to Election Day.
- Harry Truman ran such an effective campaign that he won 75% of the late-deciding voters

Candidate	Actual Vote
Harry Truman	50%
Thomas Dewey	45%

What does this "margin of error" mean?

Candidate	Percent
Bernie Sanders	24%
Pete Buttigieg	15%
Joe Biden	15%
Elizabeth Warren	10%
Amy Klobuchar	6%
Tulsi Gabbard	5%
Tom Steyer	5%
Andrew Yang	3%
Michael Bennet	0%
Deval Patrick	0%

Source: Suffolk University poll. Feb 3-4, 2020.
Margin of error for 500 likely voters is +/- 4.4%.

A Short & Simple Polling Method

- Identify your target population (e.g., registered voters)
- Identify and quantify meaningful attributes about that population
  - Gender
  - Education
  - Location
  - Party affiliation
  - Income
  - …
- Define a meaningful sample size for this population and the targeted subjects who will represent this population effectively
- Carry out the poll
- Adjust/weight the results to match up with your target population

A Tasteful Sampling Exercise

- You are asked to poll a large bag of M&Ms for their views on light snacking.

- Your M&M population is diversified as follows:
  - 25% Blue
  - 20% Orange
  - 15% Green
  - 15% Yellow
  - 15% Red
  - 10% Brown

- Take a sample of 20 M&Ms and see how your sample matches up!

Processing your sample

- Do you think your sample size was large enough for your purposes? *(Correct answer: NO! I WANT MORE M&MS!)*

- Did you have representation from every subgroup? How closely did it match the target population?

- How would you need to adjust your polling results to make this sample reportable?

This is an example of how sampling error is introduced to a poll.

* The actual peanut M&M population in a bag is as follows: 23% Blue, 23% Orange, 15% Green, 15% Yellow, 12% Red, 12% Brown.

How Sampling Error is Calculated for Polls

$$Margin\ of\ Error = \pm 1.96 \left(\sqrt{\frac{pq}{(n-1)}} \right)$$

- p = the proportion (result) being considered
- q = 1 − p
- n = number in sample

Margin of Error is the possible variation up or down from the stated result.

Example: Potential Presidential Race

Candidate	Poll Result
Bernie Sanders	52%
Donald Trump	44%

$$Margin\ of\ Error = \ \pm1.96 \left(\sqrt{\frac{pq}{(n-1)}} \right)$$

- Sample size of 1051 registered voters
- p = .52
- q = .48
- n = 1051

$$Margin\ of\ Error = \ \pm1.96 \left(\sqrt{\frac{(.52)(.48)}{(1051-1)}} \right) = \pm1.96(0.015) = \pm0.030 = \pm3\%$$

Now, your turn… back to our initial example:

Candidate	Percent
Bernie Sanders	24%
Pete Buttigieg	15%
Joe Biden	15%
Elizabeth Warren	10%
Amy Klobuchar	6%
Tulsi Gabbard	5%
Tom Steyer	5%
Andrew Yang	3%
Michael Bennet	0%
Deval Patrick	0%

Source: Suffolk University poll. Feb 3-4, 2020.
Margin of error for 500 likely voters is +/- ▪.

What is the margin of error on Bernie Sanders' 24% result?

$$Margin\ of\ Error = \pm1.96\left(\sqrt{\frac{pq}{(n-1)}}\right)$$

$$Margin\ of\ Error = \pm1.96\left(\sqrt{\frac{(.24)(.76)}{(500-1)}}\right)$$

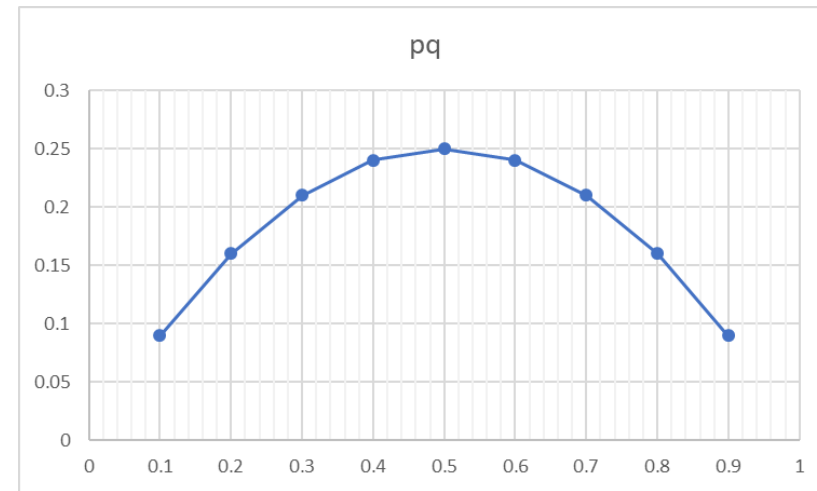$$Margin\ of\ Error = \pm1.96(0.019) = \pm0.037 = \pm3.7\%$$

Wait a minute, why was our calculation different?

Candidate	Percent
Bernie Sanders	24%
Pete Buttigieg	15%
Joe Biden	15%
Elizabeth Warren	10%
Amy Klobuchar	6%
Tulsi Gabbard	5%
Tom Steyer	5%
Andrew Yang	3%
Michael Bennet	0%
Deval Patrick	0%

Source: Suffolk University poll. Feb 3-4, 2020.
Margin of error for 500 likely voters is +/- 4.4%.

Margin of error is typically calculated for publication based upon an even split between two candidates, which essentially maximizes the error.

$$Margin\ of\ Error = \pm 1.96\left(\sqrt{\frac{(.5)(.5)}{(n-1)}}\right) = \pm 1.96\left(\sqrt{\frac{.25}{(n-1)}}\right) = \pm\frac{1.96(.5)}{\sqrt{(n-1)}} = \pm\frac{0.98}{\sqrt{(n-1)}} = 4.4\%$$

A final twist: Is the difference between two candidates (say, Sanders & Buttigieg) significant?

Candidate	Percent
Bernie Sanders	24%
Pete Buttigieg	15%
Joe Biden	15%
Elizabeth Warren	10%
Amy Klobuchar	6%
Tulsi Gabbard	5%
Tom Steyer	5%
Andrew Yang	3%
Michael Bennet	0%
Deval Patrick	0%

If there are more than two possible choices, it is not just taking the poll margin of error, doubling it, and determining if the difference is within the range or not.

$$Margin\ of\ Error = 1.96 \left(\sqrt{\frac{(p_1 + p_2) - (p_1 - p_2)^2}{(n-1)}} \right)$$

Plugging and chugging…

$$Margin\ of\ Error = 1.96 \left(\sqrt{\frac{(.24 + .15) - (.24 - .15)^2}{(500-1)}} \right)$$

Margin of Error $= 5.4\%$, not 8.8% or even 7.4%!

So yes, this difference is legitimately significant

And a final thought...

Candidate	Percent
Bernie Sanders	**24%**
Pete Buttigieg	**15%**
Joe Biden	15%
Elizabeth Warren	10%
Amy Klobuchar	6%
Tulsi Gabbard	5%
Tom Steyer	5%
Andrew Yang	3%
Michael Bennet	0%
Deval Patrick	0%

Source: Suffolk University poll. Feb 3-4, 2020.
Margin of error for 500 likely voters is +/- 4.4%.
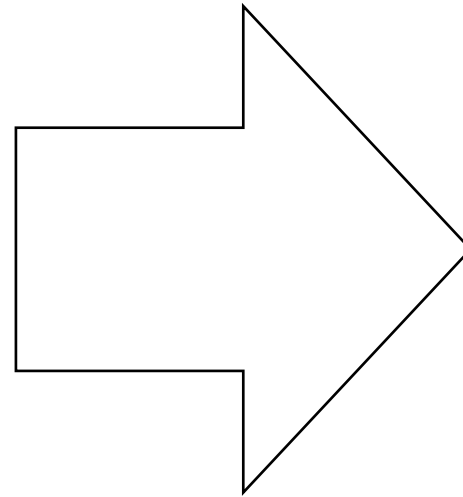
Candidate	Percent
Bernie Sanders	**24%**
Pete Buttigieg	**23%**
Elizabeth Warren	13%
Joe Biden	11%
Amy Klobuchar	6%
Tulsi Gabbard	4%
Tom Steyer	3%
Andrew Yang	3%
Michael Bennet	1%
Deval Patrick	1%

Source: Suffolk University poll. Feb 5-6, 2020.
Margin of error for 500 likely voters is +/- 4.4%.