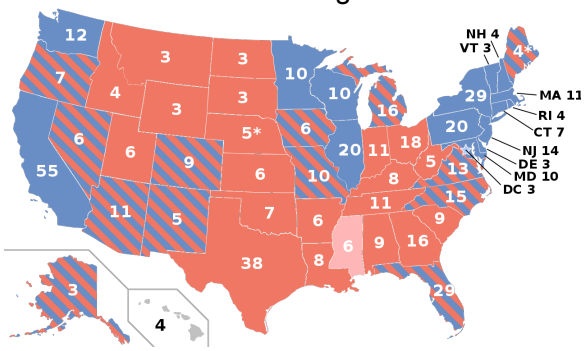
Do you think that grades in your school are inflated?

Which is best Snapchat or Instagram?

Polling



If you could interview anyone in the world, who would it be?

Which is your favorite sports team?

Have you ever thought that it would be nice to know what your classmates think about an issue? Have you ever considered conducting a poll? We wondered how hard it would be.

Before the Presidential election we heard a lot about polls and their results. The companies that perform polls try to represent how the actual voters in the United States (about 138,000,000 people) will cast their ballots from a sample of only 1,000 polled individuals.

Pollsters must be very careful about choosing the population of their sample voters.

1. If you were going to survey a sampling of the students in your school about who they support as a state Senator, what considerations might you want to take into account to be able to choose a representative sample?

2. If you would like to know what town restaurants are favored by your school's population and you polled only the students in your math class, how might your polling sample not be representative of your school's population?

3. What sort of considerations do you think the polling organization need to take into account when trying to pick a representative sample of the U.S. voter population?

4. When you figure out the percentages that you would like to gather from each of those descriptors of voters, what obstacles might you encounter in actually choosing those voting subjects?

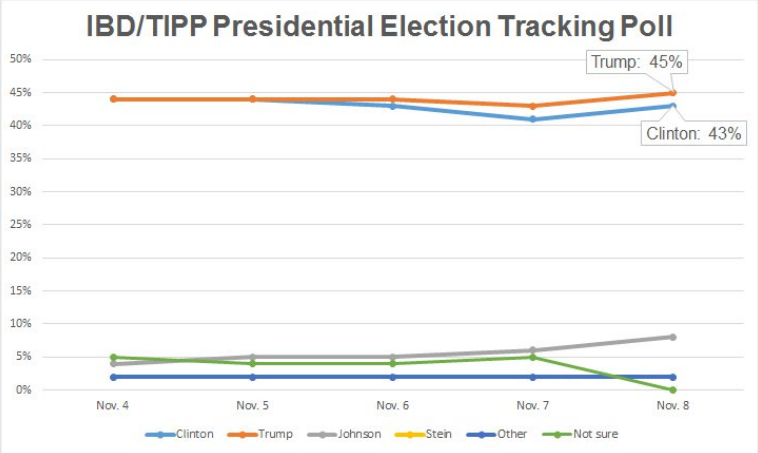
5. What do you think are some common errors in choosing a random, representative population of the soon to be voters?

As you can see, choosing a population-representative sample of 1,000 subjects is tricky. When the poll has been created and completed, I find even the understanding of the reporting to be tricky and often misrepresented. Here's one example of a poll taken by IBD (Investors Business Daily) just before the November 8th, 2016 Presidential election.

> Results based on survey of 1,107 likely voters. This poll was conducted from November 4, 2016 to November 6, 2016. The margin of error: +/- 3.1 percentage points and a 95% confidence rating. IBD's polling partner TechnoMetrica used "traditional" telephone methodology using live interviewers for data collection for its public opinion surveys. Roughly 65% of interviews come from a cell phone sample and 35% from a Random Digit Dial (RDD) landline sample.

6. Where do you think the 95% part comes from? Any thoughts?

7. What's with the +/-3 percentage point's margin of error?



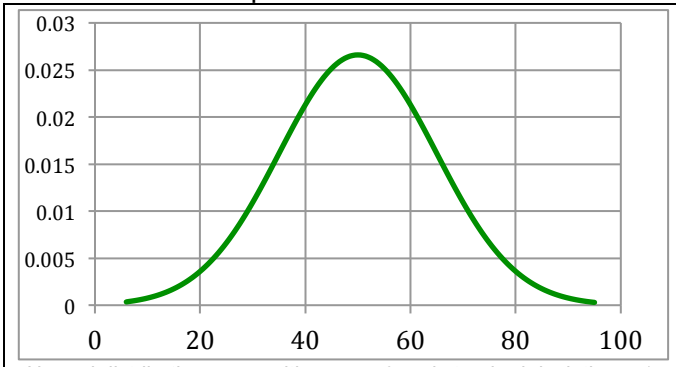
8. On November 7th, the poll results showed that 45% of those polled were in favor of Mr. Trump. If you saw a list of each of the 1,000 (actually 1,107 people) voters who took the poll, would you expect that exactly 450 of them were voting for Mr. Trump?

9. What do you expect the range would be of the number of people out of 1,000 who actually said that they were voting for Mr. Trump?

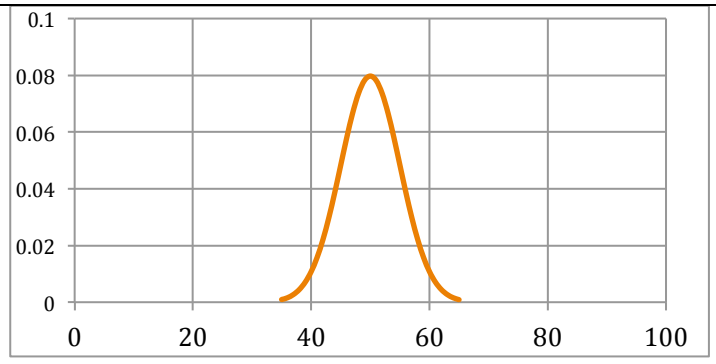
10. Since we can't survey everyone in the U.S., how many voting individuals in the U.S. would be represented by each of those about 1,000 dots on the graph (1,000 polled people)?

We found that in polling, it is assumed that the data frequency will generally follow a normal distribution ... a bell curve of results when graphed. Most of the respondents are represented by the high point of the graph.

Here are two examples of bell curves.



Normal distribution curve with mean 50 and standard deviation = 15.



Normal distribution curve with mean 50 and standard deviation = 5.

There is a measurement called the standard deviation that allows you to know that the data hangs pretty close to the mean or that the data is spread out over a large range.

11. When you look at the two bell curves pictured above, you see that they both have a mean value = 50 but they have two different standard deviations. How would you describe how those different standard deviations have affected the graphs?

The standard deviation is pretty much the average *distance* of all of your data from the mean of the data.

If $\bar{X}$ = the mean of all of your data,

"n" is the number of data points in your collection,

Σ means "the sum of"

Then $\sum(\bar{X} - X)^2$ = sum of all of the squared differences between your data and the mean,

and $\sqrt{\frac{\sum(\bar{X}-X)^2}{n}}$ or $\sqrt{\frac{\sum(\bar{X}-X)^2}{n-1}}$ is the standard deviation of the data.

$\sqrt{\frac{\sum(\bar{X}-X)^2}{n}}$ Use this formula for standard deviation if you know all of the data in your population.

$\sqrt{\frac{\sum(\bar{X}-X)^2}{n-1}}$ Use this formula for standard deviation when you have only a sample of your population data.

12. Why do you think I defined standard deviation as the *distance* from the mean instead of the *difference* from the mean?

Let's use an example to find some statistics.

In our middle school, all of the 8th grade algebra students are given the same final exam. Lets see what we can predict and how much accuracy we can expect in our predictions about all 200 of the scores if we only use 25 grades as our representative sample.

Suppose that these are the scores of the 25 students in your math class for that final exam.

34	61	64	68	74
55	61	65	68	77
57	63	65	70	78
60	63	67	71	82
60	64	67	72	84

13. Find the mean of this data.

14. Calculate the absolute deviations from the mean and the squares of those deviations in the chart below.

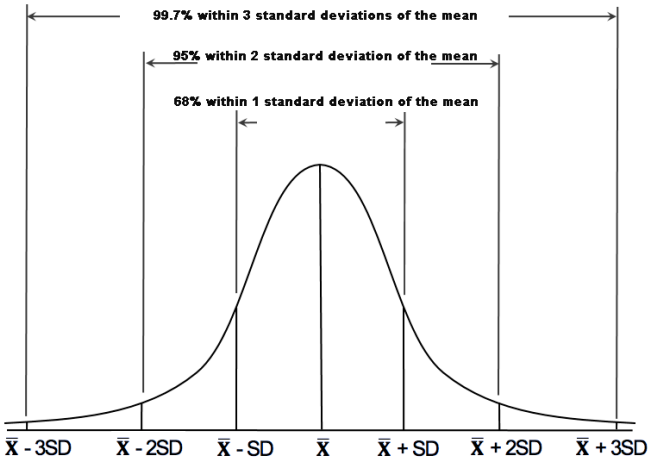
25 Student's scores	Mean absolute deviation	Mean deviation squared
34		
55		
57		
60		
60		
61		
61		

63		
63		
64		
64		
65		
65		
67		
67		
68		
68		
70		
71		
72		
74		
77		
78		
82		
84		

15. Now calculate the standard deviation of this sample.

16. What might you infer from this mean and this standard deviation?

We found that to getting a 95% confidence interval has to do with the standard deviation of the data. Almost all of the data that follows a normal distribution will fall within 2 standard deviations of the mean or 95 percent of the data will be found within 2 standard deviations of the mean.



68% of the population will be within 1 standard deviation of the mean.

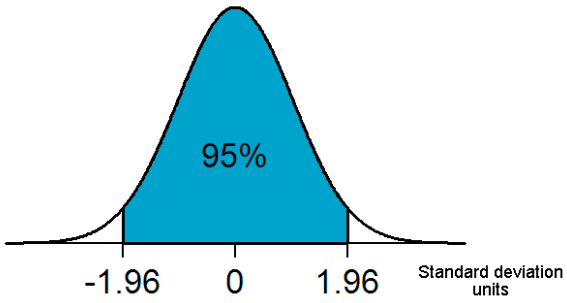
95% of the population will be within 2 standard deviations of the mean.

99.7% of the population will be within 3 standard deviations of the mean.

17. So, in the example that you just worked out about our Algebra final, what would be the range of grades that you would expect represents 95% of the grades in all of the 200 8th grade students?

Now lets figure out the +/- 3.1percentage points error calculation. The margin of error (MOE) of a survey is related to its sample size and its desired confidence value.

If we want people to be able to reproduce our poll and get similar results 95% of the time then we have a confidence interval of ∓ 1.96 standard deviations from our mean score.



Notice that this bell shaped curve has been translated so that its mean is at 0 and the positions of 2 standard deviations from the mean are now at ± 1.96.

One formula for Margin of Error simply involves your desire of a confidence level and your sample size.

- For a 90% confidence level, MOE $\approx \frac{0.82}{\sqrt{n}}$
- For a 95% confidence level, MOE $\approx \frac{0.98}{\sqrt{n}}$
- For a 99% confidence level, MOE $\approx \frac{1.29}{\sqrt{n}}$

18. Do the math. Was the margin of error for a 95% confidence level correct? Show your work.

19. In our Algebra final exam example our population was 200 8th grade students and our sample size was 25 students. What would be the margin of error in our Algebra final exam?

20. What might you like to create a poll for?

21. When you chose your random sample, what considerations might you have to include to make sure that you get a representative sample of your population?

Sources: http://www.quickanddirtytips.com/education/math/how-to-use-statistics-to-understand-poll-results
http://www.investors.com/politics/ibd-tipp-presidential-election-poll/
http://www.suffolk.edu/academics/10741.php
http://stattrek.com/statistics/dictionary.aspx?definition=z score
http://www.statisticshowto.com/what-is-standard-deviation/