MATH E-3: Lecture 6

Quantitative Reasoning: Practical Math

Math E-3

March 1, 2016

https://farm8.staticflickr.com/7144/6810031577_b63fdb120f.jpg

Homework

- Assignment 5 is due Saturday, 3/5.
- Assignment 6 will be posted tomorrow.

Homework Assignment Submission
“Homework Help Center”

CHECKLIST

- ☑ single PDF file
- ☑ 4mb maximum
- ☑ file name (example: albrigo.assign1)
- ☑ work must be neat and legible (flag your final answer) and scanned to the appropriate drop box (pages in order and no upside down or sideways pages)
- ☑ review your assignment after submitting
- ☑ upload by Saturday, 11:59 a.m. (ET) deadline

It’s advisable not to wait until the last minute to upload your assignment, as you may experience much “upload traffic” and become locked out.

Review carefully the Homework Policies section outlined in the syllabus.

Quiz #1
(LECTURES 1-5)

Date: Tuesday, March 8

Time: 7:40 pm through March 9, 7:40 pm
(Eastern time)

Location: Online, MATH E-3 Canvas course site
(no proctor needed)
No class meeting

Quiz #1 is open book. You will have 75 minutes to complete the quiz.

Quiz #1 Review

- **Review Section**: Friday, 3/4, beginning at 5:30 pm (ET), live, online via Canvas Conference; recorded session and slides will be posted shortly thereafter to the Quiz #1 module on the course site home page.

- **Practice Test** has been posted. See Quiz #1 module)

Quiz #1 TIPS

- **Review** readings, homework assignments, lecture slides
- **Watch** the review section video and review section slides
- **Attend** online or on campus help section if needed
- **Complete** the Math E-3 Practice Test

Help/Review Sections

- **March 2:** online via Conferences, beginning at 7:30 pm ET, with Jessica

- **March 4:** online Quiz 1 review section, beginning at 5:30 pm ET, with Sue and Jessica

©Toons4Biz * illustrationsOf.com/7803

But first, show me the money . . .

If you're at the casino, or the bank, or withdrawing a large amount of money today -- you may notice something different with your money.

The U.S. Treasury Department has printed new $100 bills and the Federal Reserve begins distributing it to banks. The new, more secure bills have been in development for a decade and even had to overcome production problems over the past few years. But the bills are ready now.

The bills feature a number of new security features to thwart counterfeiters, including a 3-D security ribbon with images that move in the opposite direction from the way the bill is being tilted, and a disappearing Liberty Bell in an ink well.

http://www.marketplace.org/topics/economy/its-all-about-new-benjamins-new-100

On a related "note" . . .

How much cash is there in the US (and the world)?

HOW MANY BILLS IN YOUR WALLET?

The U.S. Treasury has released a new $100 bill. As of 2012 there were more than $863 billion* worth of $100 bills in circulation. Let's take a look at what we know about the cash that's out there and where it is.

The amount of cash per capita in circulation in the U.S. keeps going up. If you took all the bills in circulation and divvied them up between us, how much would each of us have?

*Source: federalreserve.gov-/paymentsystems/coin_curr-circvalue.htm

Where is the currency? A country by country breakdown of the Top 10 foreign locations for all U.S. bills*

* Source: mpra.ub.uni-muenchen.de

**Feige and Dean, 2004, E.L. Feige and J.W. Dean, Dollarization and Euroization in transition countries: Currency substitution, asset substitution, network externalities and irreversibility, in V. Alexander, J. Melitz and G.M. von Furstenberg (ed.), Monetary Unions and hard pegs: Effects on trade, Financial development, and stability, Oxford University Press, (2004) pp. 303-319.

How has the amount of $100 bills in circulation changed over time?

Currency in Circulation - Value

- In 1990, $100 bills made up roughly 52% of the total value of US circulation worldwide. By 2012, $100 bills accounted for nearly 77% of total value worldwide*

Currency in Circulation - Volume

- In 1990, $100 bills made up roughly 10% of the total volume of US circulation worldwide. By 2012, $100 bills accounted for 26% of total volume worldwide**

*Source:www.federalreserve.gov/paymentsystems/coin_currcircvalue.htm

**Source: www.federalreserve.gov/paymentsystems/coin_currcircvolume.htm

MARKETPLACE

A little Excel . . .

		bill	# per person	products
9		**bill**	**# per person**	**products**
10	Average bills per person in the US	$100	28	$2,800
11		$50	5	$250
12		$20	23	$460
13		$10	5	$50
14		$5	8	$40
15		$2	3	$6
16		$1	33	$33
17				
18			Total	**$3,639**

A bit more Excel

21	Average cash per person in the US	$3,639
22		
23	Total population of the US	315,000,000
24		
25	Total cash in the US (billions US$)	$1,146

Math in the news . . .

WGBH . . . 89.7 or ??

Italian earthquake from 2009 . . .

http://www.npr.org/blogs/thetwo-way/2012/10/22/163400917/italy-finds-scientists-guilty-of-manslaughter-for-2009-earthquake-forecast

2009 Italian earthquake

Six Italian scientists have been sentenced to six years in prison for what a judge said was a faulty forecast of the 2009 earthquake in L'Aquila.

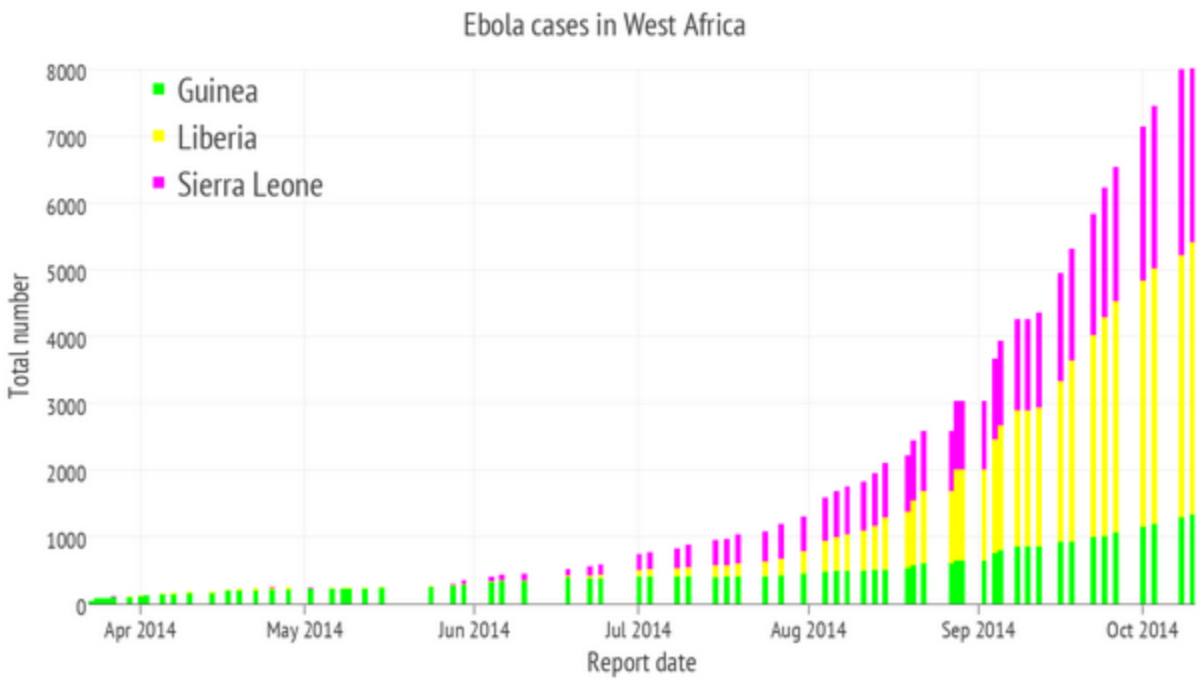
The BBC reports that prosecutors said the scientists, who work for the National Commission for the Forecast and Prevention of Major Risks, "gave a falsely reassuring statement before the quake, while the defense maintained there was no way to predict major quakes."

According to NBC News, what happened is that L'Aquila had been feeling tremors in late March. One local man, who was not a scientist, made the prediction that a big one was on its way. Responding to the man on March 31, the group of scientists concluded it was **"improbable"** that the area would experience a major earthquake, "although they stopped short of entirely excluding the possibility."

On April 6, a magnitude 6.3 earthquake killed more than 300 people.

Andrew Revkin, over at *The New York Times*' Dot Earth blog, wrote about the case last year. He called the trial "a medieval-style attack on science."

Ebola cases in West Africa



Ebola cases in West Africa (Data: WHO / Chart CC BY 4.0: JV Chamary / Source: http://onforb.es/1sCVxE1)

http://www.forbes.com/sites/jvchamary/2014/10/13/ebola-trends/

According to the same story . . .

The total number of cases was rising at an exponential rate. As of 14 September 2014, the doubling time was 16 days in Guinea, 24 days in Liberia and 30 days in Sierra Leone.

So, focusing just on Guinea, suppose from the graph that there were 1300 cases on October 13, 2014. How many cases would there be in 3 months if this rate of increase were to continue? How about in 1 year?

3 months = roughly 96 days

day	# cases
0	1,300
16	2,600
32	5,200
48	10,400
64	20,800
80	41,600
96	83,200

1 year = approximately 368 days

day	# cases
0	1,300
16	2,600
32	5,200
48	10,400
64	20,800
80	41,600
96	83,200
112	166,400
128	332,800
144	665,600
160	1,331,200

176	2,662,400
192	5,324,800
208	10,649,600
224	21,299,200
240	42,598,400
256	85,196,800
272	170,393,600
288	340,787,200
304	681,574,400
320	1,363,148,800
336	2,726,297,600
352	5,452,595,200
368	**10,905,190,400**

Is this possible?

10,905,190,400 Ebola cases in a year?!

That's more than the population of the world!

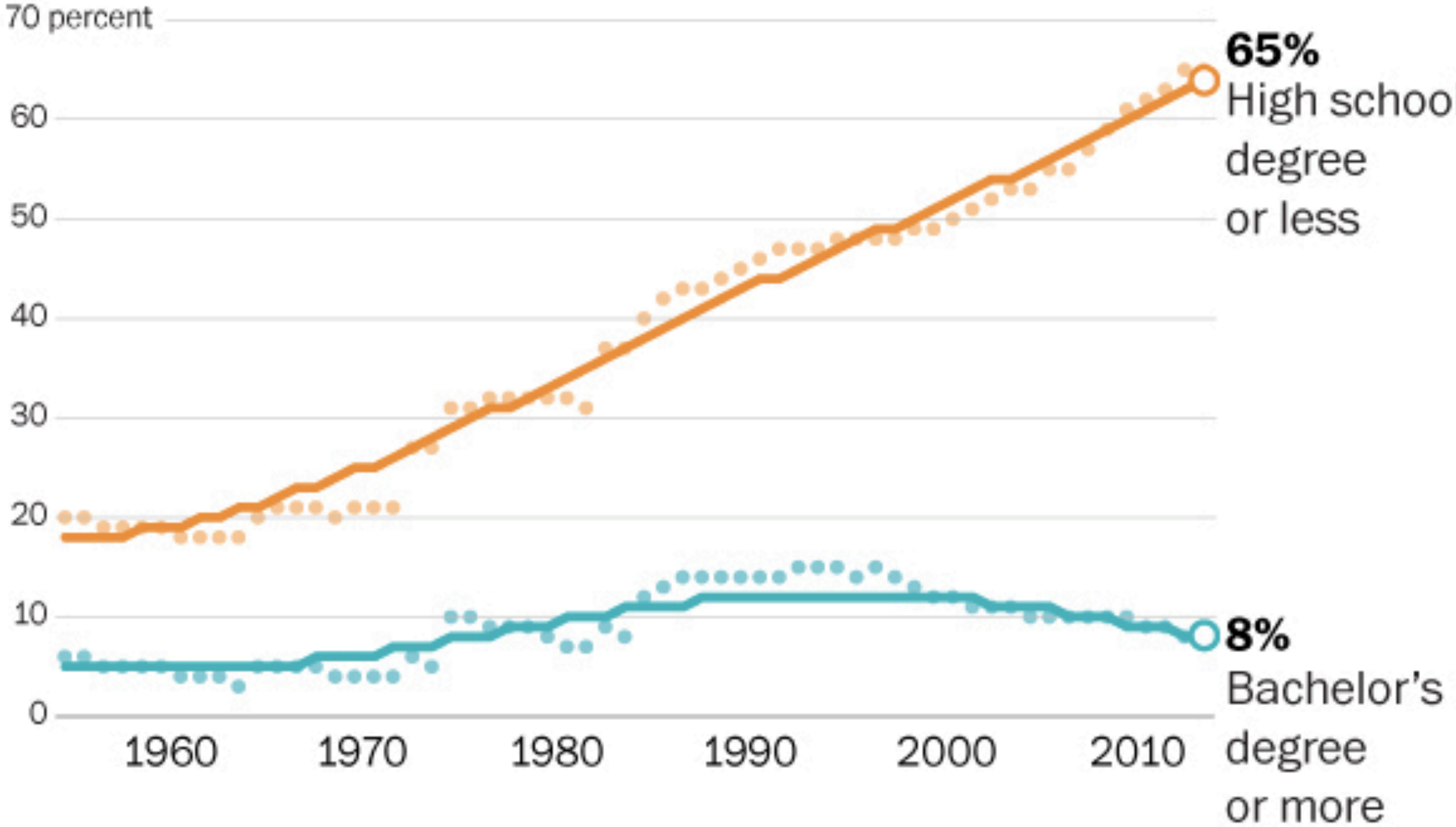
Fortunately, the hard work done by medical people and ordinary citizens helped to bring an end to the disease.

This is an example of "extrapolating too far."

There's nothing wrong with the math, just the way it is being used.

Children living in a single-parent home

In 2012, 65 percent of children whose mothers never made it past high school spent at least part of their early childhood in a single-parent household, up from 20 percent in 1953.



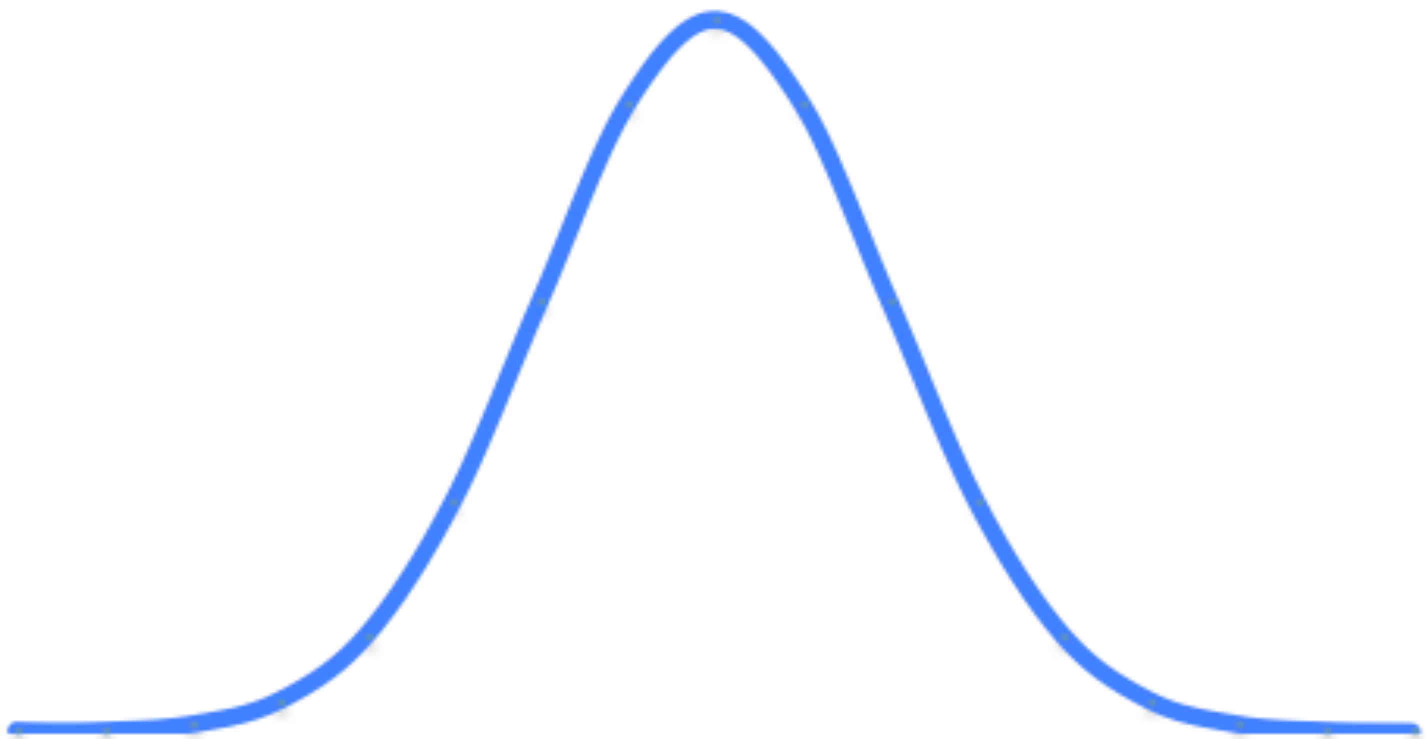
Sources: Robert Putnam and U.S. Census Bureau

THE WASHINGTON POST

More on probability . . .

The “Bell Curve”

A preview of Probability type II



Formula for the Normal Curve

(not on the first test…)

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

http://en.wikipedia.org/wiki/Normal_distribution

“Rough and Ready rules” for the Normal Distribution

- 68% of the data will lie within one (1) standard deviation of the mean (so 34% on each side).
- 95% of the data will lie within two (2) standard deviations of the mean (so 47.5% on each side).
- [99.7% of the data will lie within three (3) standard deviations of the mean.]

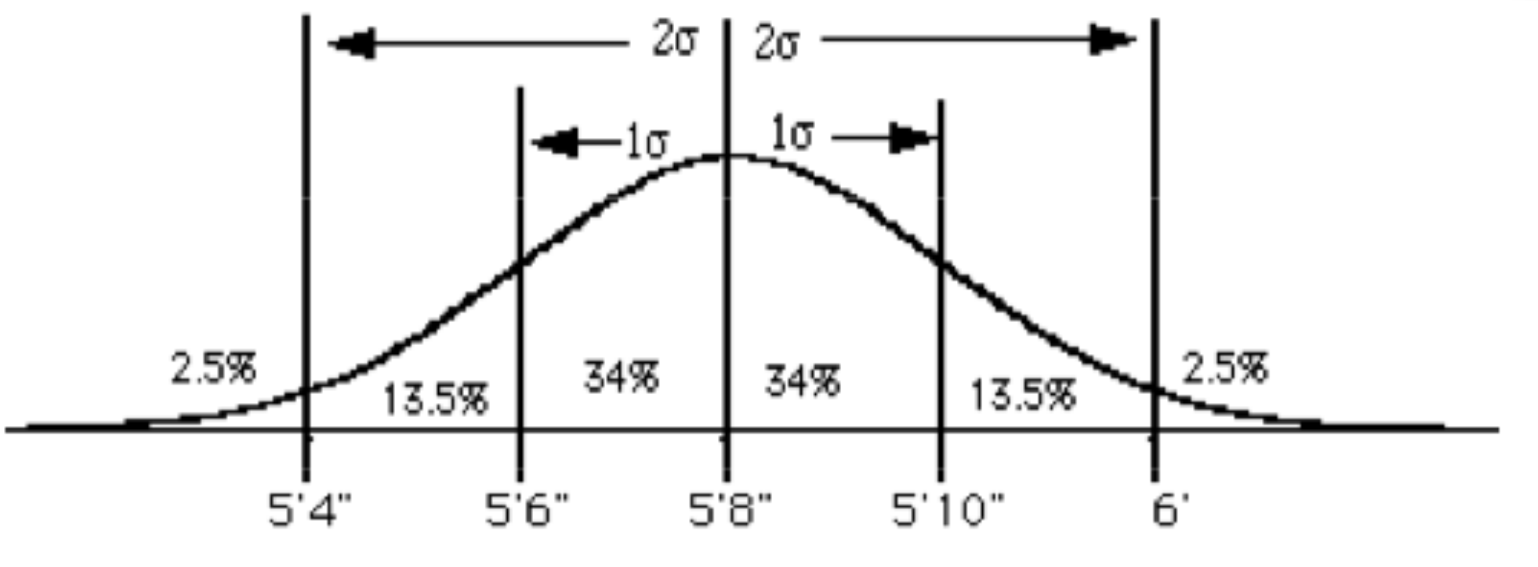
We’ll chiefly use the first two of these.

What does this mean?

Example: Heights of men (over 25) in the USA

Suppose that researchers have determined that the mean height of men in the USA is **5'8"**, and that the standard deviation is **2** inches; in addition, you are told that the distribution of men's heights is ***normal*** (as many sets of data like this are).

Normal distribution of heights of US adult males

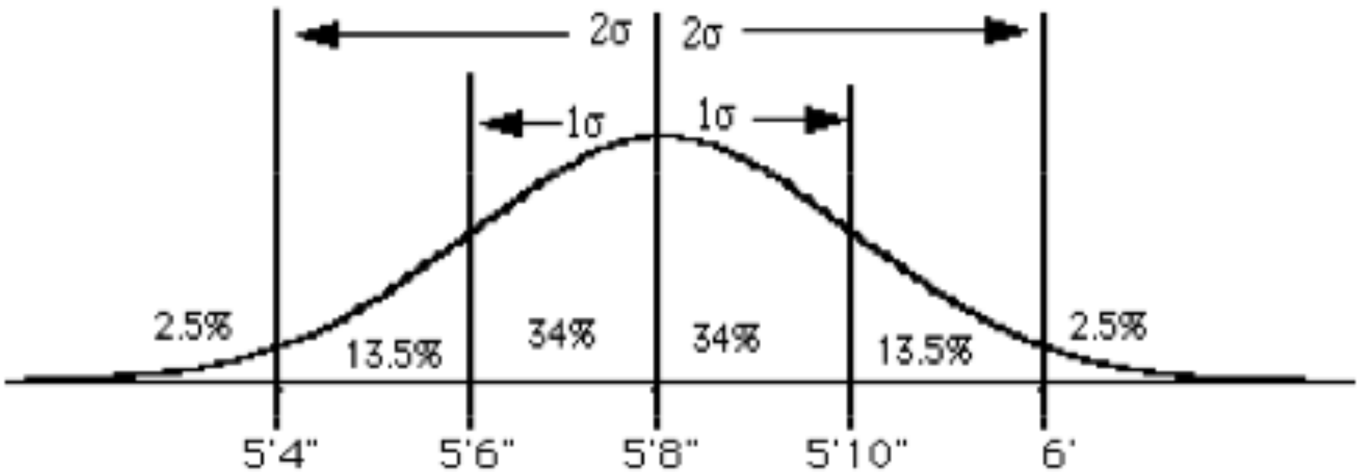


How to use this?

- Based on the previous information, what percentage of US males over 25 have a height:

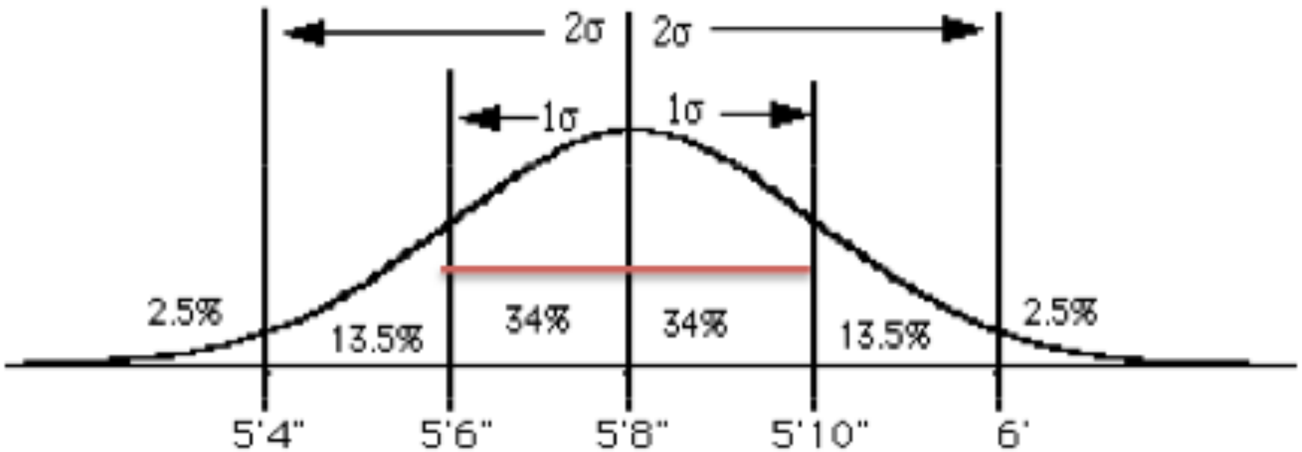
- a) between 5’6” and 5’10”?
- b) greater than 5’10”?
- c) less than 5’4”?
- d) between 5’4” and 6’?
- e) greater than 6’?

What percent have heights:



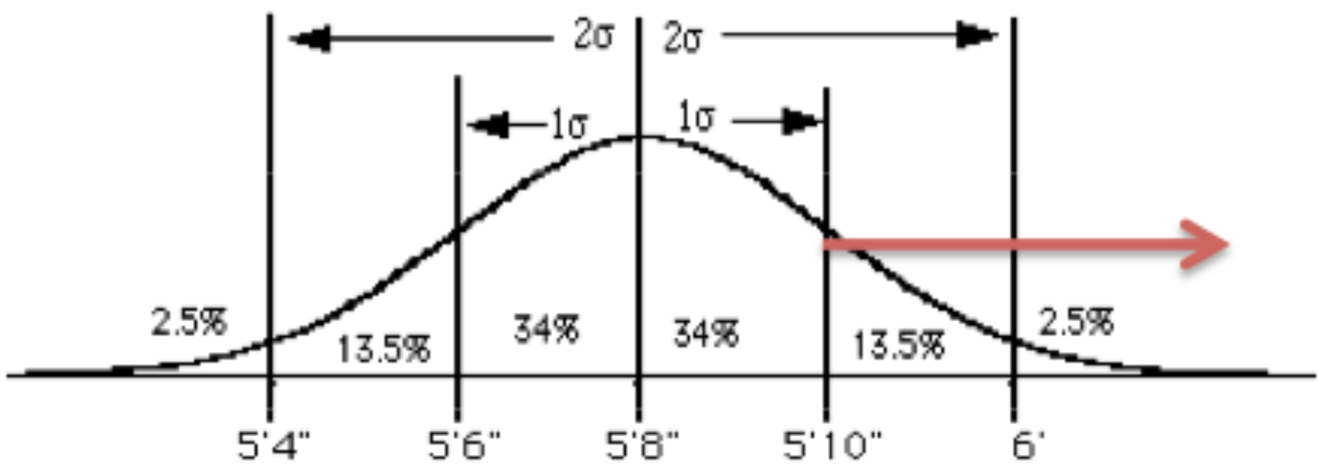
- a) between 5’6” and 5’10”?
- b) greater than 5’10”?
- c) less than 5’4”?
- d) between 5’4” and 6’?
- e) greater than 6’?

What percent have heights:



- a) between 5'6" and 5'10"? 34% + 34% = **68%**
- b) greater than 5'10"?
- c) less than 5'4"?
- d) between 5'4" and 6'?
- e) greater than 6'?

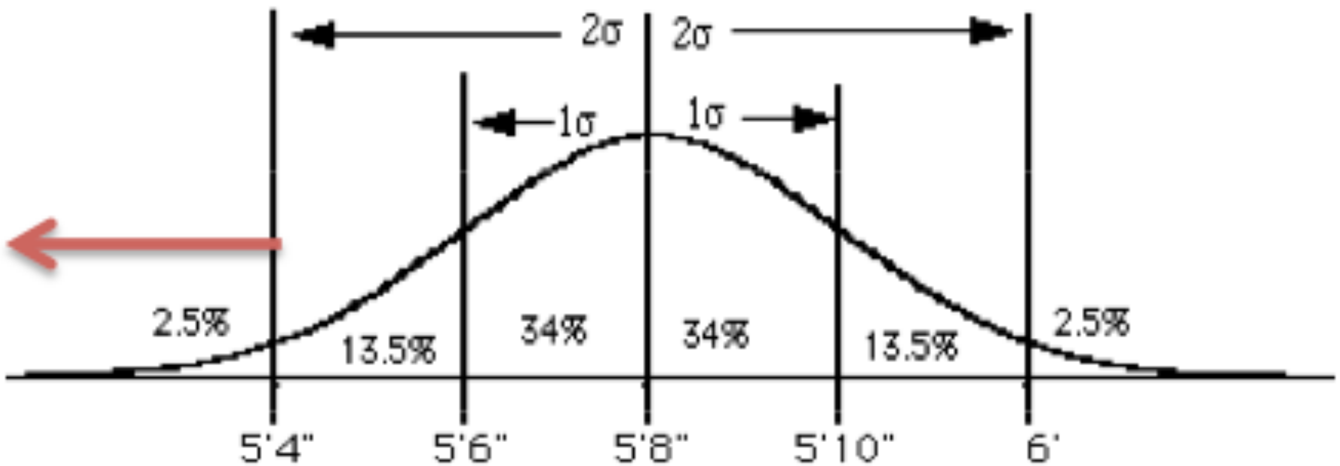
What percent have heights:



- a) between 5'6" and 5'10"?
- b) greater than 5'10"?
- c) less than 5'4"?
- d) between 5'4" and 6'?
- e) greater than 6'?

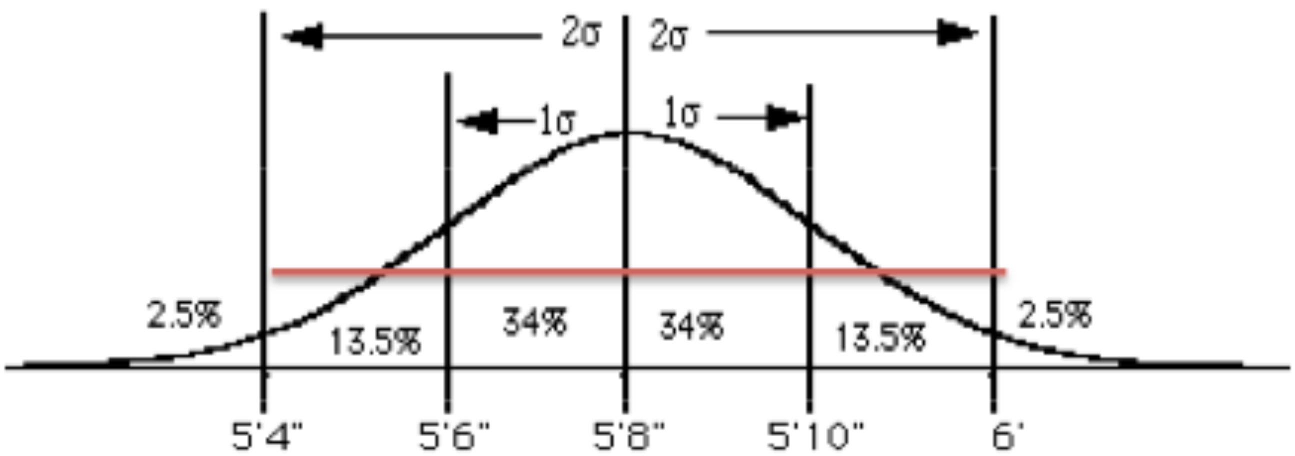
13.5% + 2.5% = **16%**

What percent have heights:



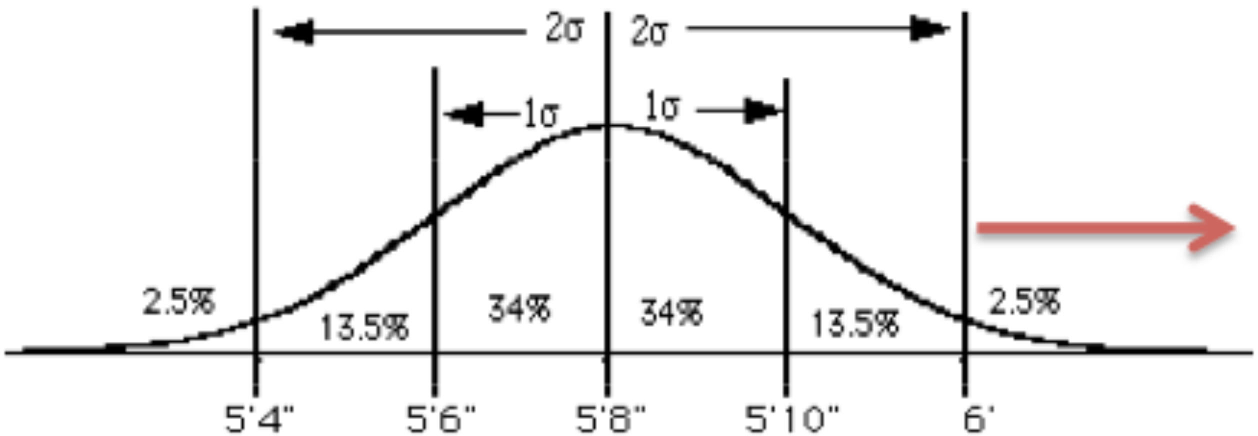
- a) between 5’6” and 5’10”?
- b) greater than 5’10”?
- c) less than 5’4”? **2.5%**
- d) between 5’4” and 6’?
- e) greater than 6’?

What percent have heights:



- a) between 5'6" and 5'10"?
- b) greater than 5'10"?
- c) less than 5'4"?
- d) between 5'4" and 6'? 13.5% + 34% + 34% + 13.5% = **95%**
- e) greater than 6'?

What percent have heights:



- a) between 5'6" and 5'10"?
- b) greater than 5'10"?
- c) less than 5'4"?
- d) between 5'4" and 6'?
- e) greater than 6'? **2.5%**

Notice what we cannot calculate using this method:

- What percent have a height of exactly 5’8”? Etc.
- *We can only calculate ranges of heights*
- What percent have heights between 5’8” and 5’9”?
- *We can only calculate ranges that are multiples of the standard deviation (at least for now . . .)*

A “leap” from statistics to probability (type II)

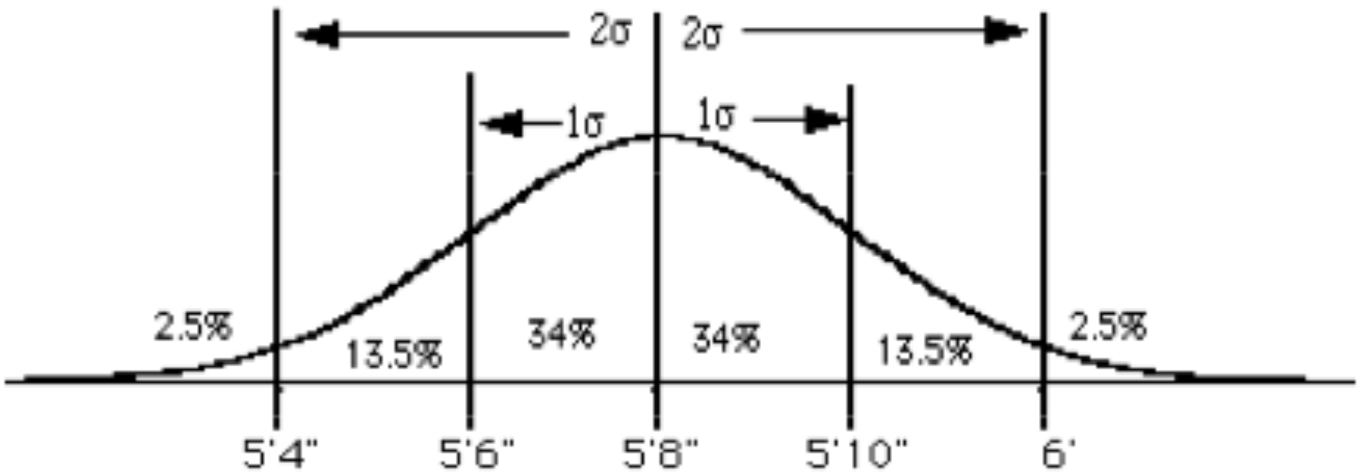
- We have already calculated that, based on these data, 68% of US males (age 25 or more) have heights between 5’6” and 5’10”.

- It is a short step (so to speak) to saying that, if you were to choose a US adult male at random, there would be a 68% chance that he would be between 5’6” and 5’10”.

Thus I can re-word my previous questions:

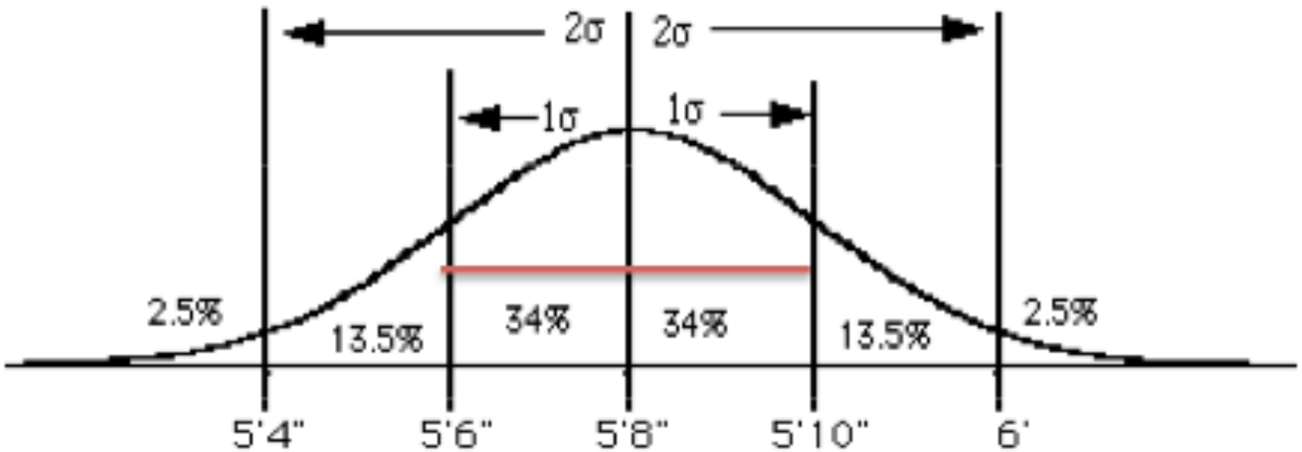
- Instead of asking: “What percent of US adult males have a height of . . .”
- I can instead ask: “If I were to choose a US adult male at random, what would be the probability that his height would be . . .”?
- So here again are the previous questions and answers, but from a <u>probability</u> perspective:

If I were to choose a US adult male at random, what is the probability that his height would be:



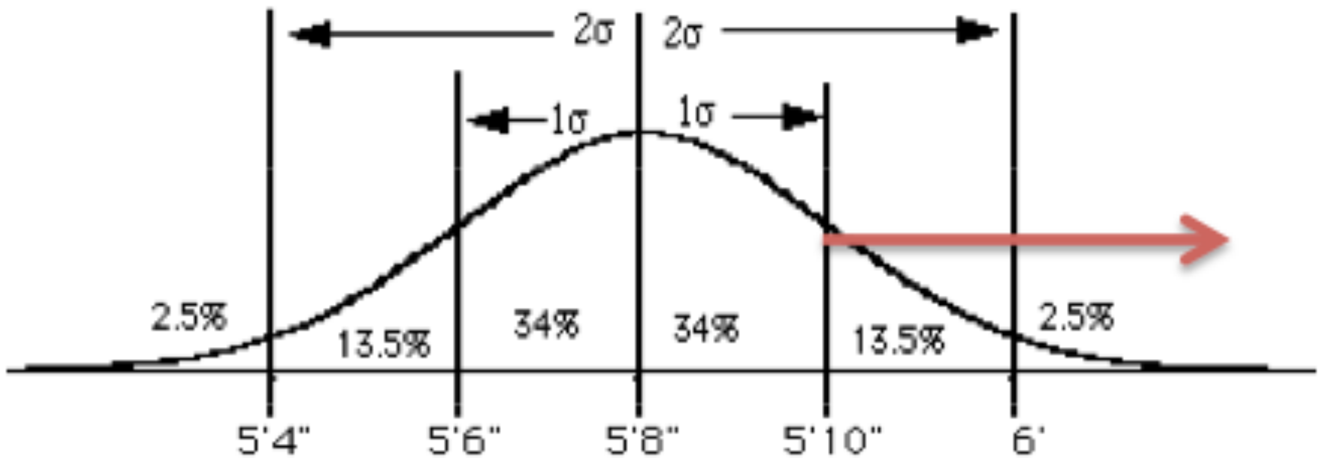
- a) between 5'6" and 5'10"?
- b) greater than 5'10"?
- c) less than 5'4"?
- d) between 5'4" and 6'?
- e) greater than 6'?

If I were to choose a US adult male at random, what is the probability that his height would be:



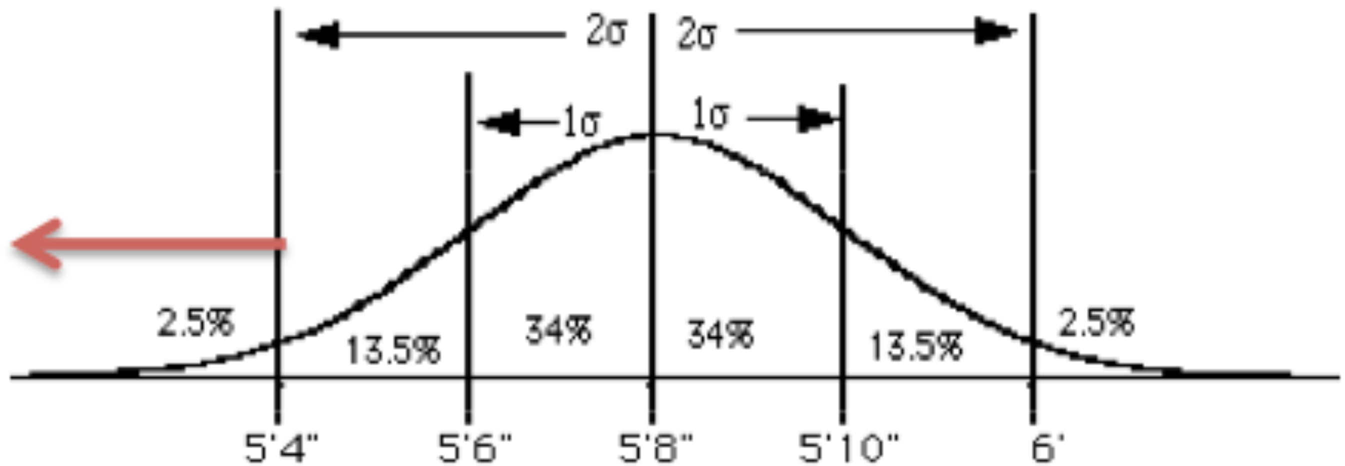
- a) between 5'6" and 5'10"? 34% + 34% = **68%**
- b) greater than 5'10"?
- c) less than 5'4"?
- d) between 5'4" and 6'?
- e) greater than 6'?

If I were to choose a US adult male at random, what is the probability that his height would be:



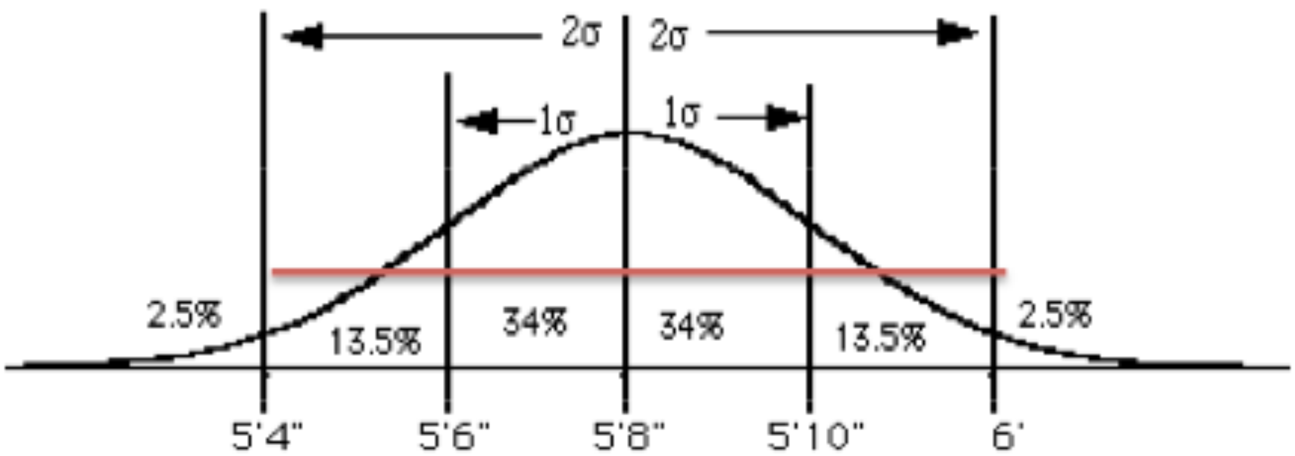
- a) between 5'6" and 5'10"?
- b) greater than 5'10"? 13.5% + 2.5% = **16%**
- c) less than 5'4"?
- d) between 5'4" and 6'?
- e) greater than 6'?

If I were to choose a US adult male at random, what is the probability that his height would be:



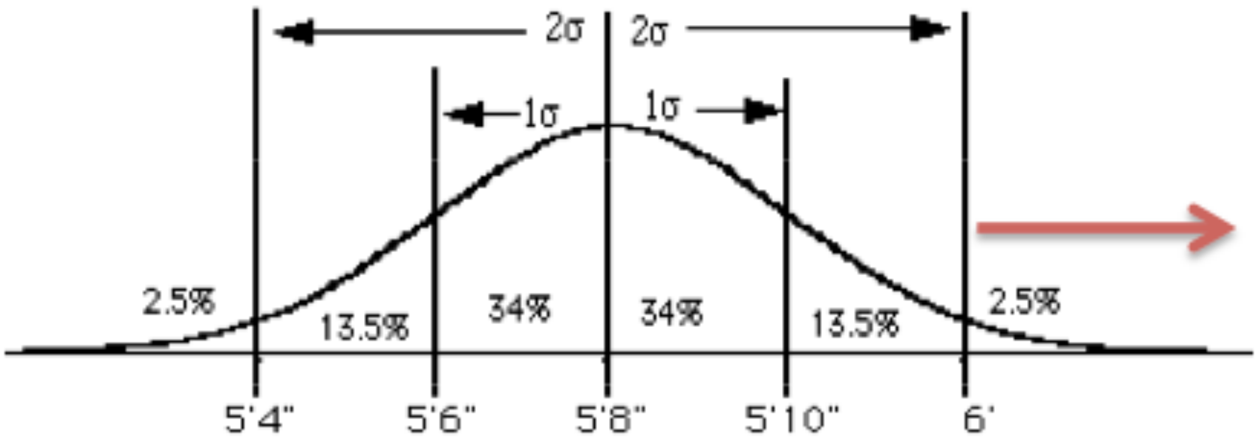
- a) between 5'6" and 5'10"?
- b) greater than 5'10"?
- c) less than 5'4"? **2.5%**
- d) between 5'4" and 6'?
- e) greater than 6'?

If I were to choose a US adult male at random, what is the probability that his height would be:



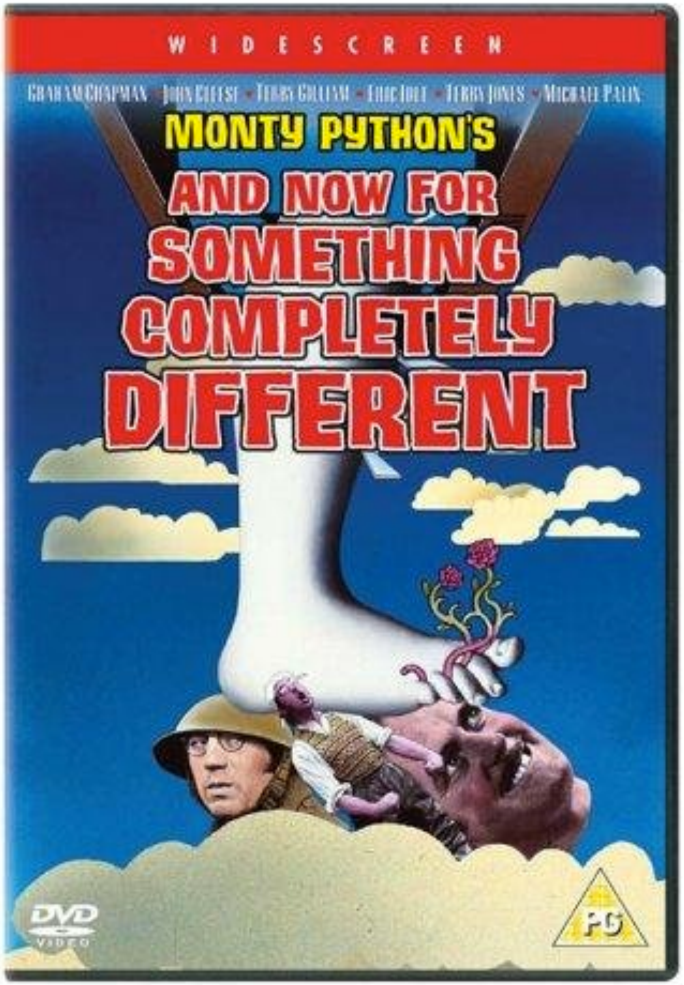
- a) between 5'6" and 5'10"?
- b) greater than 5'10"?
- c) less than 5'4"?
- d) between 5'4" and 6'? 13.5% + 34% + 34% + 13.5% = **95%**
- e) greater than 6'?

If I were to choose a US adult male at random, what is the probability that his height would be:



- a) between 5'6" and 5'10"?
- b) greater than 5'10"?
- c) less than 5'4"?
- d) between 5'4" and 6'?
- e) greater than 6'? **2.5%**

And now, for something completely different . . . well, at least partially different



http://www.imdb.com/media/rm4216226816/tt0066765

A new situation and a new formula . . .

- If you are dealing with "binary" data – i.e. data where there are only two options (think of them as "yes" and "no"), and you are taking many samples of the data, there is a formula for calculating the standard deviation of these samples, provided you already know the mean and the sample size.

For example . . .

(since today is "Super Tuesday . . .")

If, hypothetically, Hillary Clinton has 47% popularity among registered Democratic voters in Massachusetts (of course you can't know this in advance), and you and several other groups take lots of samples, each of 250 Massachusetts voters, these sample polls will not all have the same results. You can, however, calculate the standard deviation of these sample poll results, using the following fairly simple formula:

Standard deviation for samples

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

Using the new formula . . .

- Where p = the mean (or proportion) of the whole population, and n = the sample size (and of course σ = the st. dev.)

- So in this case, p = 0.47, and n = 250.

- And the standard deviation is $\sqrt{(0.47*(1-0.47)/250)}$ (I'll also write this on the board for clarity; note square root is done <u>last</u>)

- You should calculate this and get:
- Convert to a percent and round to 1 d.p. =

Using the new formula . . .

- Where p = the mean (or proportion) of the whole population, and n = the sample size (and of course σ = the st. dev.)

- So in this case, p = 0.47, and n = 250.

- And the standard deviation is $\sqrt{(0.47*(1-0.47)/250)}$ (I'll also write this on the board for clarity; note square root is done <u>last</u>)

- You should calculate this and get: 0.031565804
- Convert to a percent and round to 1 d.p. = **3.2%.**

Still to come . . .

We’ll be using this formula and the rough and ready rules quite a bit: for sampling and confidence intervals (today), and for hypothesis testing (the following class).

Sampling
and
Confidence Intervals

Samples . . .

- What are samples?
- What are they for?
- How should they be chosen?
- What should they be like?
- How large should they be?

Example: Martha Coakley's chances . . .

(from before last fall's election)

- Now that Martha Coakley is the Democratic gubernatorial candidate in the November election

- What are her chances of winning in the November election?

- How could we find out before November?

- Stand inside Harvard and ask people?

- Stand in downtown Boston?

- Call people? Which people? How many?

Choose a sample

- Let’s begin by asking say, 100 registered voters randomly chosen from around the country.
- Then those 100 people are what we call a *random sample*.
- But how do you choose people randomly?

- Should we only ask one group of 100 people?
- What if we sent out 10 grad students and had each ask 100 people? (why grad students?)
- Then we would have 10 samples each of size 100.

Sample fractions from 10 Coakley surveys

Question from Grad Student to a Random passerby: Will you vote for Martha Coakley in the gubernatorial election?
Answers are listed below from each of the 100 people the grad student asked.
Notes: We will assume for the purposes of this exercise that every person asked did give an answer. Hence, the result labeled 'Other.' (e.g. if a respondent answered, "maybe," it was not counted as a yes.)

Grad Student Name:	**Results** Yes	Other	Percent Yes
Janet	52	48	52%
Jeff	46	54	46%
Steve	56	44	56%
Jane	35	65	35%
Ann	52	48	52%
Dick	47	53	47%
Albert	40	60	40%
Pat	32	68	32%
Joseph	64	36	64%
Mike	46	54	46%

So, as you can see, this very last column is what we call the **"Sample Fractions."**

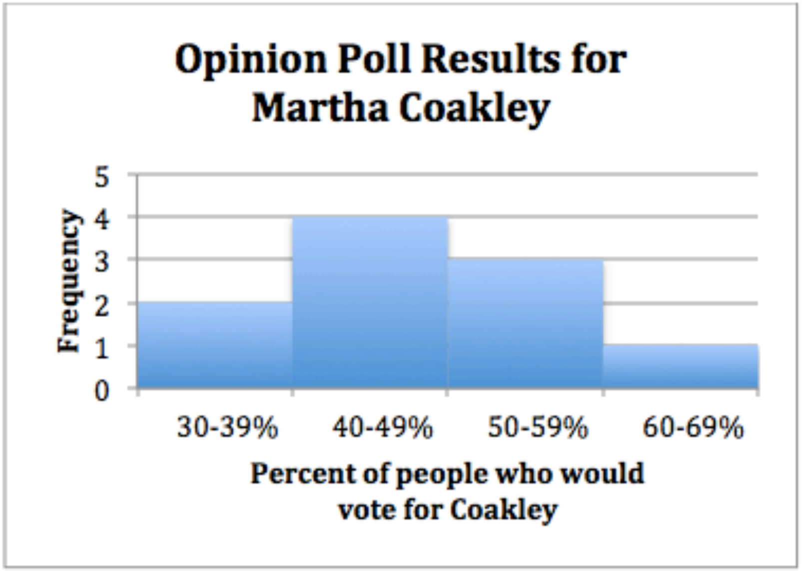
Coakley poll data . . .

$$\textbf{Mean} = \frac{52+46+56+35+52+47+40+32+64+46}{10} = \textbf{47\%}$$

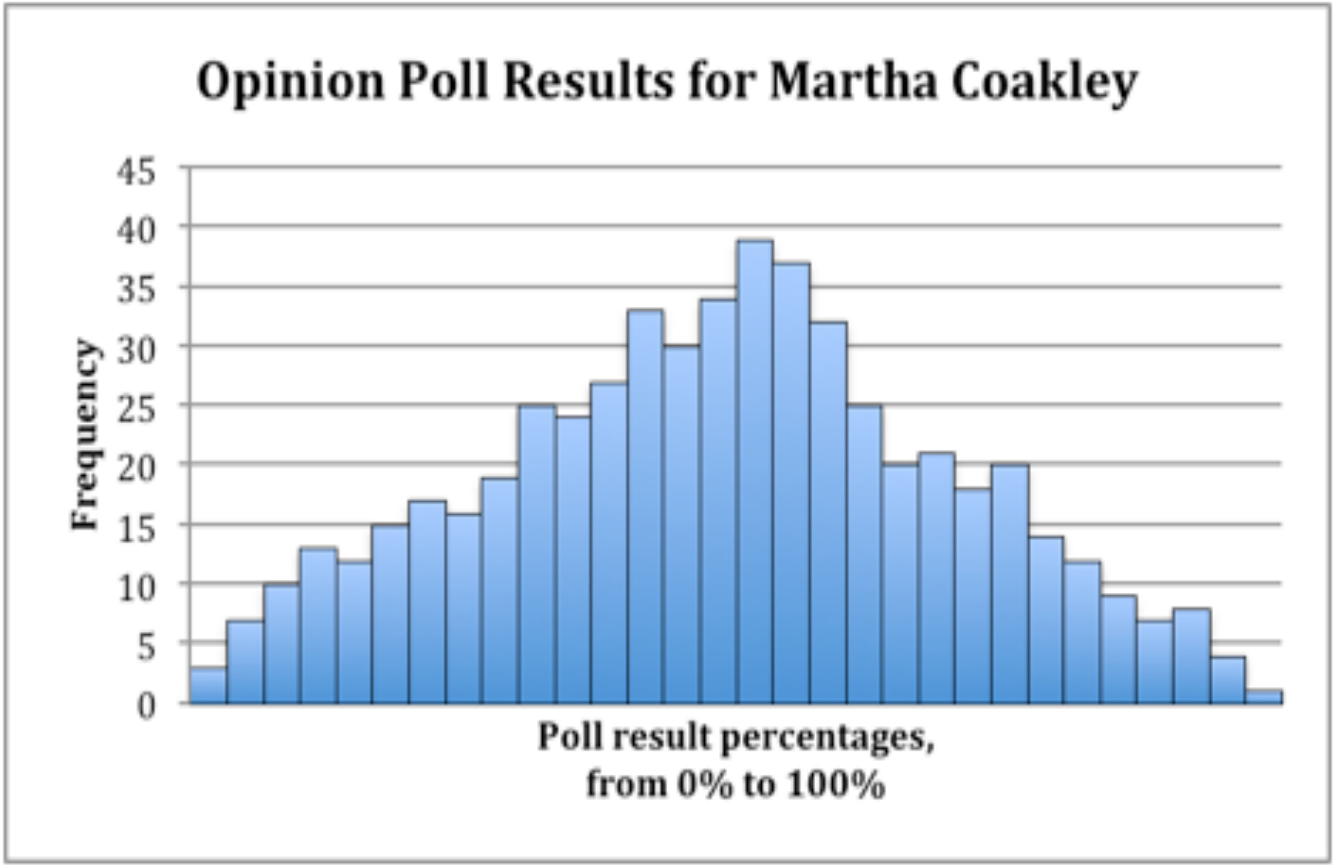
Raw Data

52%
46%
56%
35%
52%
47%
40%
32%
64%
46%

Percent Data	#Times Appeared Frequency
30-39%	2
40-49%	4
50-59%	3
60-69%	1



Histogram of "many" polls . . .



From earlier slide . . .

Recall the New Formula for the Standard Deviation

Now that we have a NORMAL DISTRIBUTION, we can use our rough and ready rules to find percentages and probabilities. But first, we need to calculate the **Standard Deviation**. How would we do that with 'thousands' of sample fractions? Instead of doing it the old way, we now have a simpler formula. This formula can only be used if we are working **with proportions (percents)** rather than with raw data.

The New Formula for Standard Deviation

The formula for calculating the standard deviation of *sample fractions*:

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

p is the mean of the population. (or in this case *hypothetical* or *sample* mean.)

n is the **size** of the sample we are looking at.

Standard deviation for Coakley data

Let's calculate the Standard Deviation for the Coakley data: I have labeled it as Example 1.

Ex 1: p = 47% and our sample size is 100. i.e. **n = 100**

Formula: $$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

First turn 47% into a decimal. 47% = 0.47 (Keep at least three decimal places.)

$$\sigma = \sqrt{\frac{.47(1-.47)}{100}}$$

Put (1 – .47) in the calculator first and get (1 – .47) = .53 So.

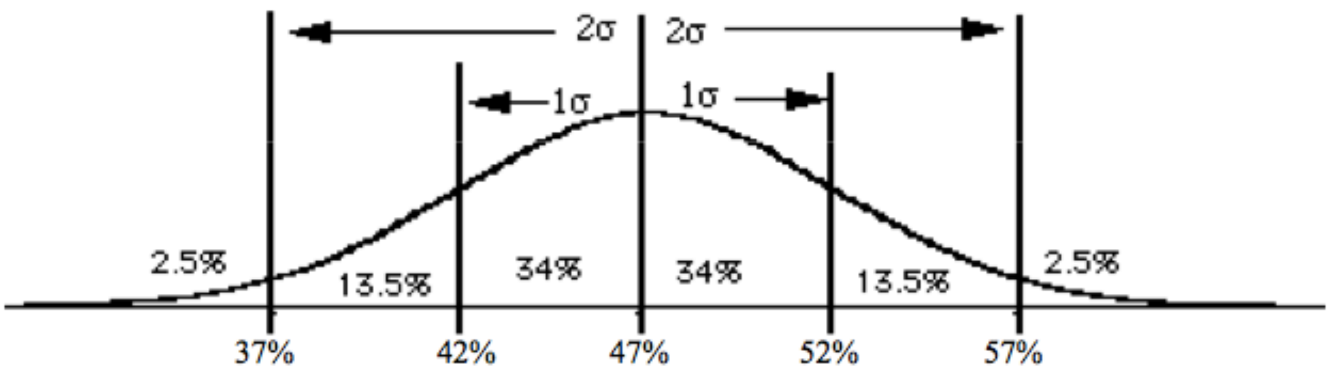
$$\sigma = \sqrt{\frac{.47(.53)}{100}}$$

$$\sigma = \sqrt{\frac{.2491}{100}} = \sqrt{.002491} = 0.0499099... = \mathbf{4.99099...\ \%}$$

Round: $\sigma = \mathbf{5\%}$

Histogram of Coakley data

Let's continue with our hypothetical example from the handout on Sampling and Statistical Inference. We looked at those who would vote for Martha Coakley. Our mean was 47% and the standard deviation was 5%. Let's draw the normal distribution curve for these figures. Put 47% as the mean in the center on the horizontal. Add 5% in one direction and subtract in the other, etc.



Predictive questions . . .

- In this case, we may ask the following questions:
-
- What is the probability that Coakley will get between 42% and 57% of the vote?
-
-
- What is the chance that Coakley will get less than 42% of the vote?
-
-

Predictive questions . . .

- In this case, we may ask the following questions:
-
- What is the probability that Coakley will get between 42% and 57% of the vote?
-
- **Answer:** Add the inside percentages that are associated with those on the horizontal.
-
- 34% + 34% + 13.5% = 81.5% (I think Coakley would be happy with those results.)
-
- What is the chance that Coakley will get less than 42% of the vote?
-
-

Predictive questions . . .

- In this case, we may ask the following questions:
-
- What is the probability that Coakley will get between 42% and 57% of the vote?
-
- **Answer:** Add the inside percentages that are associated with those on the horizontal.
-
- 34% + 34% + 13.5% = 81.5% (I think Coakley would be happy with those results.)
-
- What is the chance that Coakley will get less than 42% of the vote?
-
- **Answer:** 13.5% + 2.5% = 16%
-

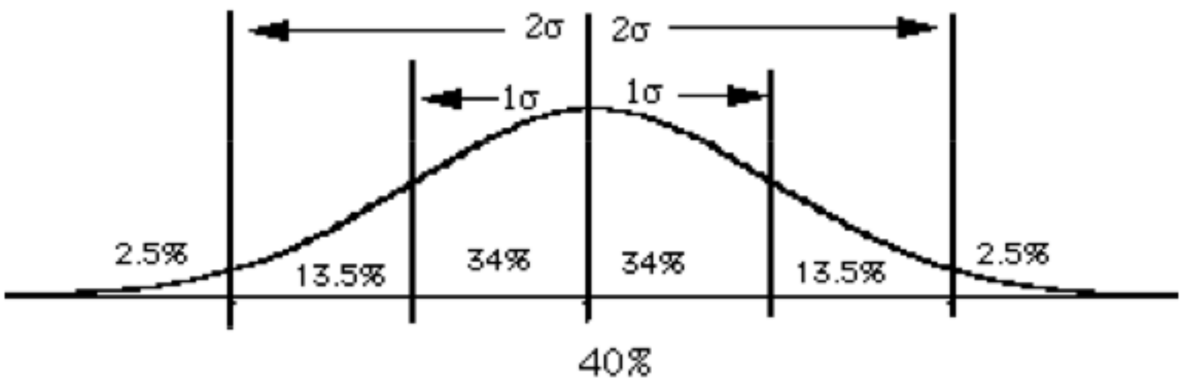
Confidence Intervals
(from when Deval Patrick was governor…)

Suppose Governor Deval Patrick asked you to find out his popularity in Massachusetts, and he gave you a small budget to take a survey of about 1000 people. Based on what you know about sampling, you take an <u>unbiased</u> poll of 1000 people and find out that 400 of them say they like Governor Patrick and would vote for him if the election were held right now (and if he were running again, which he isn't). Now should you go back and tell him that he has a 40% popularity rating??? (400/1000 = 40%)

Estimating Governor Patrick's popularity

- a) You understand that sampling does not necessarily yield the ***exact*** percentage of what the entire population of voters might say. Even if your poll is truly random, there is still a lot of **variation** in sampling results.
-
- b) If you could possibly (which you obviously cannot) take many of these polls of 1000 voters, you would build a histogram of sample fractions. (see handout on Sampling). The true fraction of voters who like your candidate would end up being the mean of all these trials. But this is not feasible!
-
- c) You know from our study of the normal distribution, that out of all these samples, 95% of them would be within two standard deviations of the true percent. We can turn this around a bit and say that *95% of the time the true percent should be no more than two standard deviations away from our one single sample mean*.
-
- d) However, we do not have very many samples - we only have one. So based on this one and only one sample, we assume it is a *pretty good estimate* of the true fraction of all voters. This is our mean, **p**, that we use in our estimate for the standard deviation.

Estimating Governor Patrick's popularity



Our estimate of 40% is the mean and the true proportion is somewhere within two standard deviations of this mean. What do you tell Governor Patrick? Let's actually do the steps and develop our wording.

Estimating Governor Patrick's popularity

Example 1: You take an unbiased poll of 1000 people to find out the popularity of a prospective candidate for whom you work. (Say Deval Patrick before he became Governor.) You find that out of the 1000 people polled, 400 of them say they like your Deval Patrick and would vote for him. ***Find a 95% Confidence Interval for the true proportion of people who approve of your candidate and would vote for him.***

Step 1: Obtain a value for **p** based on our own sample. You may need to calculate 'p' as it could be given in numbers such as in this case. We have 400 out of 1000 (our sample size was 1000), thus

$\mathbf{p} = \frac{400}{1000} = \mathbf{40\%}$ (the 40% came from your survey, i.e. you asked 1000 people and 400 gave a positive answer thus 40%.)

$\mathbf{n = 1000}$

Remember our “new” formula for standard deviation for samples

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

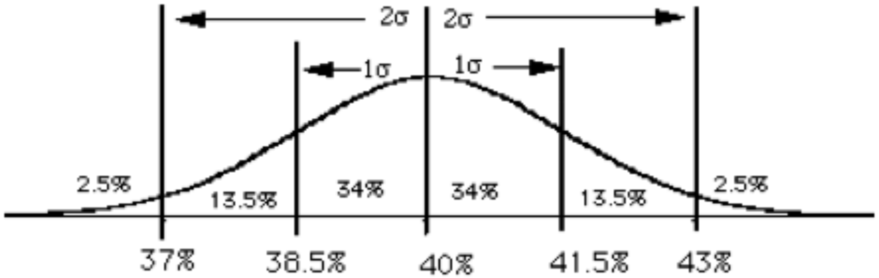
Estimating Governor Patrick's popularity

Step 2: Calculate the standard deviation based on this **observed** percentage. Use the new formula.

$$\sigma = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.40(.60)}{1000}} = 0.01549... = 1.5\%$$

NOTE: 2 times the standard deviation **IS** the Margin of Error. So **Margin of Error = 2(1.5%) = 3%**

Step 3: Calculate $\mathbf{p \pm 2\sigma}$ Drawing a picture helps here.



$\mathbf{p \pm 2\sigma}$ = 40% ± 2(1.5%) = **40% ± 3%** or from **37% to 43%**

Important note:

The Margin of Error is generally considered to be equal to

"Two times the Standard Deviation," i.e., 2σ

Estimating Governor Patrick’s popularity

Step 4: Construct the 95% confidence interval:

$p - 2\sigma$ <------> $p + 2\sigma$

Answer (1) The 95% confidence interval for the true proportion of voters who would vote for Deval Patrick is between 37% and 43. Write this in the notation below:

37% <-------------------------> 43%

This means that there is a 95% chance that the real true percent of voters who like Deval was somewhere between these limits. You can never be 100% sure of this but you can be 95% sure!

Answer (2) (another way of stating a Confidence Interval) The percent of voters who like your candidate is

40% plus or minus a *Margin of Error* of 3%.

Back to Coakley . . .

Example 2: Now we'll find the 95% Confidence Interval for the true proportion of people who will vote for Martha Coakley in the Massachusetts gubernatorial election in November.

Step 1: Obtain a value for **p** based on your own sample. You may need to calculate **'p'** as it could be given in numbers. In this case we had p=47%. (See the handout on Sampling and Statistical Inference.)

p = **47%**

n = 100 This time we had only asked 100 people, a much smaller sample than the 1000.

Back to Coakley . . .

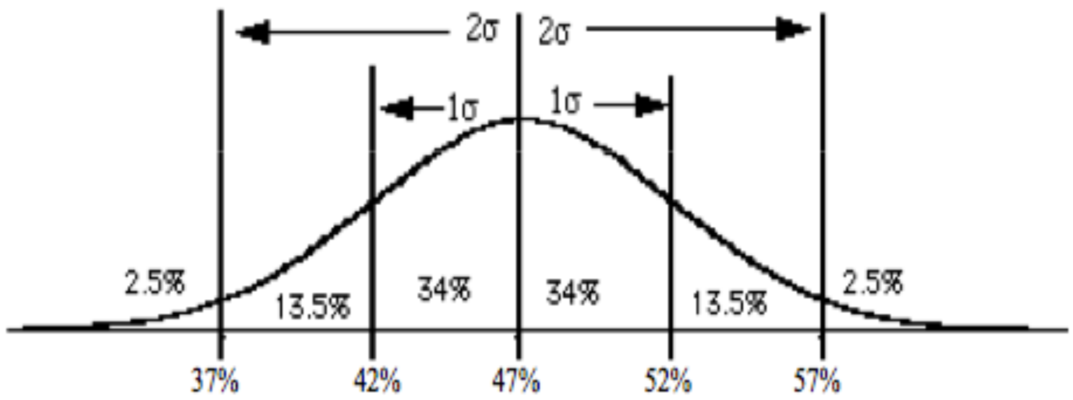
Step 2: Calculate the standard deviation based on this **observed** percentage. We already calculated the standard deviation as 5% (Page 2).

. σ **= 5%**

NOTE: 2 times the standard deviation **IS** the Margin of Error. So **Margin of Error = 2(5%) = 10%**

Back to Coakley . . .

Step 3: Calculate **$p \pm 2\sigma$** We show the curve below.



$\mathbf{p \pm 2\sigma}$ = 47% ± 2(5%) = **47% ± 10%** or from **37% to 57%**

Back to Coakley . . .

Step 4: Construct the 95% confidence interval:

$$p - 2\sigma \longleftarrow\!\!\!-\!\!-\!\!-\!\!-\!\!\longrightarrow p + 2\sigma$$

Use the wording as in ANSWER (1) or ANSWER (2) below.

Answer (1) The 95% confidence interval for the <u>true proportion</u> of people who say they will vote for Martha Coakley is:

37% <-------------> 57%

Note: This means that there is a 95% chance that the real true percent of these voters is somewhere between these limits. You can never be 100% sure of this but you can be 95% sure!

Answer (2) The percent of voters who say they intend to vote for Coakley is

47% plus or minus a *Margin of Error* of 10%.

NB: The Margin of Error is 10%. In example 1, our margin of error was only 3%. We had a much larger sample in example 1. (1000 versus 100 people) The margin of error shrinks as the sample size increases. This is because the standard deviation decreases as the sample size increases. And notice the reverse is true. Since the sample size was much smaller, our margin of error was larger.

Things can go wrong . . .

Landon in a Landslide: The Poll That Changed Polling

The 1936 presidential election proved a decisive battle, not only in shaping the nation's political future but for the future of opinion polling. The *Literary Digest,* the venerable magazine founded in 1890, had correctly predicted the outcomes of the 1916, 1920, 1924, 1928, and 1932 elections by conducting polls. These polls were a lucrative venture for the magazine: readers liked them; newspapers played them up; and each "ballot" included a subscription blank. The 1936 postal card poll claimed to have asked one fourth of the nation's voters which candidate they intended to vote for. In *Literary Digest*'s October 31 issue, based on more than 2,000,000 returned post cards, it issued its prediction: Republican presidential candidate Alfred Landon would win 57 percent of the popular vote and 370 electoral votes.

http://historymatters.gmu.edu/d/5168

http://en.wikipedia.org/wiki/United_States_presidential_election,_1936

Things can go wrong . . . Nationally:

	predicted		actual	
	poll numbers	percents	election numbers	percents
Landon	1,293,669	57.1%	16,681,862	37.5%
Roosevelt	972,897	42.9%	27,752,648	62.5%

Even in California . . .

	California			
	predicted		actual	
Landon	89,516	92.2%	836,431	32.1%
Roosevelt	7,608	7.8%	1,766,836	67.9%

What went wrong?

Sample bias

Self-selection bias

Others?

For more, see:

http://www.scribd.com/doc/259298/Why-the-1936-Literary-Digest-Poll-Failed