In case you’d like to see why the “best fit line” is also called a “least squares regression line”…here ya go!

Look back to exercise 9 (in the Larson book) and its data. Assume, for simplicity, that the tree is perfectly circular (although this argument holds even if it isn’t).

**E1.** Notice that your dependent variable is “circumference”. What are the units of circumference?

**E2.** Estimate this tree’s circumference in 2012.

**E3.** What I’d like to do is analyze the cross-sectional area of the tree (that is, the area of the circle that would be left at the top of the stump if you cut the tree down). Using the following formulas, show me why the area of a circle is equal to the square of its circumference divided by 4\(\pi\); that is, \(A_{\text{circle}} = \frac{C^2}{4\pi}\)

1. Formula for circumference of a circle: \(C_{\text{circle}} = \pi d\)

2. Formula for the area of a circle: \(A_{\text{circle}} = \frac{\pi d^2}{4}\)

**E4.** What are the units of area, if circumference is measured in inches?

“E” Answers.

**E1.** Inches.

**E2.** 26.27 inches, according to my best-fit line from problem 9. Don’t worry about calculating a margin of error right now – we’ll get into those in MTH 243 and MTH 244!

**E3.** Here are some hints: First, solve the circumference formula, \(C_{\text{circle}} = \frac{\pi d}{4}\), for \(d\). Next, substitute the result into the \(d\) in the area formula \(A_{\text{circle}} = \frac{\pi d^2}{4}\). Cancel common factors, and you’ve got it!

**E4.** Square inches (or in\(^2\)).
The data at left shows the number of words in Facebook’s “privacy policy” in recent years (pretty ironic for a site that prides itself on small snippets of the English language, eh?). What’s interesting, to me at least, is that these data are pretty linear!

For these data, use your Excel sheet to create the best–fit equation of a line. Let $x$ be years, and let $y$ be the number of words. If you don’t want to deal with scientific notation in the regression equation, let your $x$–values be 5, 6, 7, and so on.

1. **(2 points)** Interpret the slope of this model, in context.
2. **(2 points)** According to this model, how many words should have been in Facebook’s privacy policy in 2012?  
   *Note*: don’t worry about the MOE for this prediction.
3. **(2 points)** If the pattern continues, in which year will Facebook have more than 15,000 words in its privacy policy?
4. **(1 point)** Which of the questions above asked you to extrapolate?
5. **(3 points)** Name one reason that extrapolating with any data set can lead to inaccurate conclusions.
Scary stuff, eh? It seems that one in five Americans are worried about getting Ebola! Yikes!!!!!! Yes, I, too, remember the furor that Ebola caused in America. Too bad, however, I don’t fully believe that claim. 😊

1. **(2 points)** Read the first paragraph of the article and tell me what percentage of Americans “say they worry about getting the Ebola virus.”

2. **(2 points)** Find the margin of error (“MOE”) of the poll (it’ll be near the bottom of the page – scroll down).

3. **(6 points)** Apply that MOE to the number you got in 1 to more correctly state the headline. You might want to look at the “E” questions above if you get stuck.
In class, we looked at how to apply a “Margin Of Error” (MOE) to a predictive statistic. If you recall, we had to do this because any time a sample is drawn (even if done properly), the resulting summary statistic will vary a little bit from the true value you’re trying to estimate. The margin of error, in theory, takes care of that difference.

Here’s a whimsical example: do you like Reese’s Pieces? Yeah, me too. Have you ever thought “Hmmmm…I wonder what fraction of the candies in those bags are orange?” Of course you haven’t. But this is precisely the kind of question that researchers ask all of the time. They need to estimate the percentage of patients who get better on a certain medication, or the number of likely voters who will actually vote, or the percentage of invasive grasses in a certain wilderness zone. The catch, as you may have guessed, is that the researchers can’t ask everybody. If they could, however, they’d know exactly the answer they were looking for. But they can’t – in most cases, it’s too expensive and time consuming. So, they draw a sample, which is less than 100% certain. However, the MOE allows them to be more certain then they would with just a single sample figure.

OK, back to the whimsy…suppose a researcher needs to know what percentage of Reese’s Pieces (overall) were orange. So, since she can’t check every single one, she draws a random sample of 30 of them. Here’s what she got!

(In case you want to use the simulator that I got the image at left from, here’s the link: http://www.rossmanchance.com/applets/OneProp/OneProp.htm?candy=1. You’ll find, as you take my classes, that simulations are wonderful ways of exploring the word around you. Ask any airline pilot!).

You might notice that 10 out of the 30 candies were orange, which would be $\frac{1}{3}$, of them, or about 33%. So, should we say that 1 out of every 3 Reese’s Pieces are orange? You could. But it’s not quite right. Here’s why....

Suppose another researcher has the exact same idea as the first one – and he sets out to estimate the percentage of Reese’s pieces that are orange. His budget’s a little better, thought, and he’s able to check 100 candies. That’s

It’s a little harder to see, but this researcher got 53 out of 100 orange. So who’s right? Some of you might say that this researcher is “righter” since his sample is bigger. Assuming the sample is drawn randomly, that’s a good argument. But neither researcher is right – both of their results have a little bit of “offness” to them. Check it out...I’ll do three more sample, of varying numbers of candies checked, and you can see the results:

number orange: 109
number total: 200
% orange: 54.5%

number orange: 12
number total: 25
% orange: 48%

number orange: 521
number total: 1000
% orange: 52.1%

1 Don’t Google it! Yet, anyway. ☺
2 News media does. More on that in a minute.
So what do we say the answer is? One way to answer would be to combine all the results, or take a weighted average (more on that in MTH 243 and MTH 244). But, even if you did that, you’d have error, right?

(*) It’s time to avoid the unavoidable! Let’s attack the idea of a margin of error! Look back to the first set of results: 10 out of 30 orange, or about 0.33 orange. The MOE for an experiment with that many trials (30) is about 17%...so you’d have to say that “I’m statistically confident that there are about 33% ± 17% orange Reese’s Pieces in a bag.” Said a different way, “I’m statistically confident that there are between 16% (33% - 17%) and 50% (33% + 17%) orange Reese’s Pieces in a bag.”

(**) Remember, you have to do this because there is uncertainty! Any time you can’t sample an entire population (that is, check every Reese’s Piece), you must express the uncertainty. Let’s do it again!

In the second set of data, we got 53 out of 100 orange (53%). The MOE for an experiment with that many trials is about 10%, so you’d have to say that “I’m statistically confident that there are about 53% ± 10% orange Reese’s Pieces in a bag”, or, more understandably, “I’m statistically confident that there are between 43% and 63% orange Reese’s Pieces in a bag.”

Two things to note:

1. (2 points) As the sample size goes up, the MOE shrinks. Why do you think that is?

2. (2 points) Notice that the percentages that we got originally (ignoring the MOE) are different (33% and 53%). Why would they be? Seems they should be the same, no?

OK – you ready to try your own? You bet!

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Percentage Orange</th>
<th>MOE</th>
<th>Conclusion?</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>54.5%</td>
<td>7%</td>
<td>“There are likely between 47.5% and 61.5% orange Reese’s Pieces in a bag.”</td>
</tr>
<tr>
<td>25</td>
<td>48%</td>
<td>19.5%</td>
<td>3. (2 points) You answer this one!</td>
</tr>
<tr>
<td>1000</td>
<td>52.1%</td>
<td>2%</td>
<td>4. (2 points) You answer this one, too!</td>
</tr>
</tbody>
</table>

What you just learned about are called confidence intervals – a range of values that you feel “confident” will contain the number you’re looking for. In this one, you were interested in finding out what percentage of Reese’s Pieces were orange. A quick Google Search tells me that 50% should be. Notice that 50% isn’t the actual percentage we ever actually got with one of our experiments (we got 33%, 53%, 54.5%, 48%, and 52.1%). However...

(2 points) In how many of the confidence intervals that we constructed did the number 50% fall? Use all 5 from up there: the two we did in the paragraphs marked “*” and“**”, as well as the three in the table.

That’s the power! And I know that analyzing candy’s kinda silly – but this mathematics is the same used in way more important analyses (for example, this one...or this one). Welcome to the awesome world of statistics!

---

3 You’ll calculate WHY in MTH 244!
So, in class we did a couple of cost analyses (light bulbs, water heaters). I’d like to do one more!

A few years ago, my family decided it was time to put PV (photovoltaic) panels on our house. It was a big decision for us, since it required us to save for years before we actually did it. But, the time was right, the ETO, the state, and the feds gave us some good incentives, and we decided, “Let’s do this!”

So, with all of the incentives accounted for, we had to pony up around $5000 to cover the startup cost of the panels’ purchase and installation (technically, we had to provide even more, but with the tax breaks, it washed out to an overall out-of-pocket cost of $5000).

Now, electricity in Bend carries an average rate of about 10.78 cents (that’s $0.1078) per kilowatt hour (a kilowatt hour, or kWh, is defined to be “using 1000 watts of power for the duration of 1 hour”). Based on my own analysis of our power bills, we save (on average) around 15% on our electric bills by using the panels.

1. **(2 points)** What’s the Rule family average electric cost per kWh?

Looking at our power bill, we use about 11,000 kWh of power each year. How much does that power cost us? Do it two ways:

2. **(2 points)** As if we didn’t have the panels (that is, use the $0.1078 per kWh)

3. **(2 points)** Using the panels (that is, use the rate you found in part 1).

You now have two slopes and two y – intercepts (technically, one’s 0) to use in a cost analysis! Use your Excel sheet (the “Exact” tab) to approximate how many years it will take until our investment in solar will start saving us money.

4. **(2 points)** We installed the panels in 2010. In approximately which year will having the panels cost us less than not having them?

5. **(2 points)** Take a screen capture of the graph you used and include it with this quiz! Here’s a [video to help you take a screenshot](https://www.youtube.com/watch?v=Guj1zHEhSEg) if you need it! (link: [https://www.youtube.com/watch?v=Guj1zHEhSEg](https://www.youtube.com/watch?v=Guj1zHEhSEg))
Quiz 5.

I remember learning, years ago, about Morse Code. I wondered why the letters “E” and “T” has such small representations, while others, like Z and J, had such long ones. I eventually learned a secondary topic:

Letter frequency

My assumption

(points) is it right?