How high is this building?

Math 2606 Group Project #1

Due: March 10, 2017 by 5 pm

Statistics in the field

While the mathematical theory of statistics is incredibly useful in guiding our understanding of scientific practice, there is a lot to be gained just by considering our simple models of data collection very carefully. For this project, you're going to consider the implications of additive errors on estimates. We've already seen additive error a couple of times. It just means that we assume that the error is added to a true value; in the context of the mean:

$$x_i = \mu + \epsilon_i,$$

where x_i is a single data point. Later on in the course we'll learn how to think about distributions related to ϵ_i in terms of probability distributions. For the moment, we're just going to consider the implications of the additive part.

Finding the height of Searles

The first step in this project is to estimate the height of the Searles Science Center. There are any number of ways to do this so I'm going to be specific and ask you to use the mirror method using similar triangles. This whole method is pretty much summed up in this diagram:

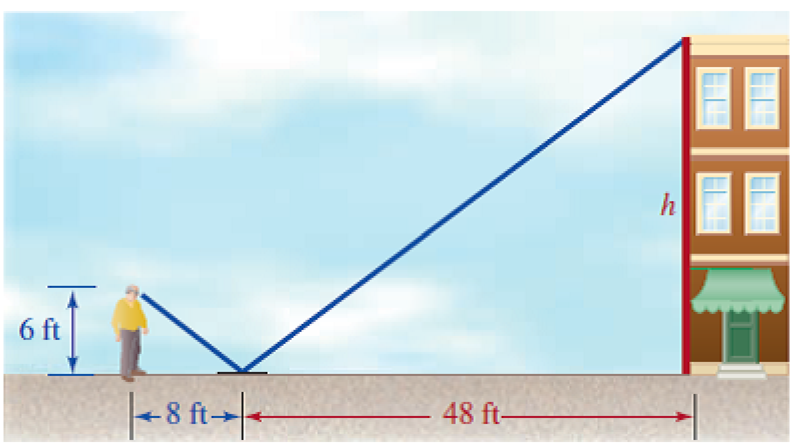

Figure 1: Diagram of the mirror method.

For mathematical convenience, let's label the two heights h_1 and h_2 and the two lengths d_1 and d_2, with the smaller ones getting the 1 subscripts and the larger ones getting the 2 subscripts. It's pretty easy to estimate h_2 given the other three values using similar triangles:

$$\frac{h_2}{d_2} = \frac{h_1}{d_1}.$$

Project outline

This project is about figuring out how to come up with the best estimate of h_2 and will lead you through an analysis of understanding the consequences of additive errors, and testing whether or not that is a good model.

Note: While the diagram above is in imperial units, it is absolutely essential that you use the metric system for your measurements below.

1. If needed, come by my office to pick up a mirror. Fine to use your own as well.
2. With your group, find d_1, h_1, and d_2 at least ten times for each quantity. For each distance, take a set of measurements together (i.e. don't do d_1 ten times then h_1 ten times; do one set of d_1, h_1, d_2 and then do another.) Also: don't just assume that people know their heights: they definitely don't know their heights to their irises. (It would be wise to use a single person.)
3. With these measurements, estimate the height of Searles.
   - For each measurement set, find an estimate of the height. Average the set of estimates.
   - Average all measurements and
4. Assume each of these estimates has additive error. Then:

   $$h_2 + \epsilon_{h_2} = \frac{(h_1 + \epsilon_{h_1}) \cdot (d_1 + \epsilon_{d_1})}{d_2 + \epsilon_{d_2}}.$$

   Of course, if we had perfect knowledge of the measured quantities, we'd know that $h_2 = \frac{h_1 \cdot d_1}{d_2}$. Use these expression to find a formula for the error in ϵ_{h_2}.
5. For each of your directly measured quantites, estimate the standard deviation up to an order of magnitude (e.g. 1 cm, 1 mm, 1 dm).
6. Use your formula from above for the error in h_2 to find which measurement contributes to the error the most.
7. To understand this a bit more richly, use `R` to plot the various measurements against the inferred values of h_2.
8. What does this suggest about the model of additive errors?

To pass in...

- A typed, carefully written summary of your work, including tables of the data, the formulas you derived, and plots produced. You may do this in whatever way you'd like though I'd suggest that `Rmarkdown` is the easiest.
- This doesn't have to be that long: together with the plots and tables, three pages is likely enough.
- In terms of style, imagine a thorough laboratory report: describe what you did and the results you found.
- That last bullet point above is the one exception: I'd like you to think about what does this tell you about the model of additive errors. Think a little bit about this; perhaps drop by my office for a chat. Additive error was assumed, then used through the calculation. What do the final plots tell you about this assumption? Why? (No need to expound at great length; a thoughtful paragraph or two is enough.)