Assignment: **#FlattenTheCurve**

Spring 2020 Advanced Data Analytics
Samford University Professor Joy Buchanan

This semester, we have been talking about predictive modeling. In March of 2020, college students throughout the United States are being told that they may not continue to physically mix with their college friends. This is a huge sacrifice being asked of young adults, who are not themselves likely to suffer from contracting the virus that originated in Wuhan in 2019.

Why? How can we predict what will happen with and without this drastic measure? We can do predictive modeling. The entire country is taking directives from epidemiologists and data analysts right now.

Using the small amount of data that we have to make decisions about what to do next is what analytics is good for. I write this from the United States. This week, all of the organizations that I subscribe to (church... kid soccer...) have decided to stop meeting to avoid spreading the disease.

Part of the reason that we are making these decisions in the United States is that we can look at data from other countries where the virus hit sooner. One of the best sources of data is the World Health Organization (WHO).

The purpose of event cancellations is enforcing "social distancing" to slow the spread of the disease. To understand why, consider what would happen if we do not change our behavior. As I write this, there are only a few confirmed cases in the United States. If you only look at the level of the numbers, this feels like no big deal.

To comprehend cancelling in-person classes and the highly-anticipated basketball games, you have to understand growth rates and exponential growth.

I'm going to use some simple numbers here. An excellent resource, if you are looking for the latest accurate statistics, is https://ourworldindata.org/coronavirus.

Consider this simple example (borrowing from Our World in Data): Say there were only 1000 confirmed cases today, and there were 500 cases three days ago. We would say it took three days for the number of confirmed cases to double. Now suppose that the doubling time is three days for a month straight. How many cases would there be at the end of a month?

Say a month has 10 doublings. We get 1000 in the first doubling. Then the second doubling takes us from 1000 to 2000, and so on.

1. How many cases are there at the end of one month?
2. How many cases occur halfway through the second month (assume another 5 doublings)?
3. If 5% of all cases requires medical attention, how many people are seeking medical attention after a month and a half?

The medical system is currently in use. People get sick and have babies. People need surgeries and hospital beds such that there is not a LOT of extra space to cope with a pandemic outbreak. So, it's important to predict ahead of time how many people will need medical attention.

Models can often be expressed with mathematical formulas. You can model exponential growth with a simple formula

$$y = ab^x$$

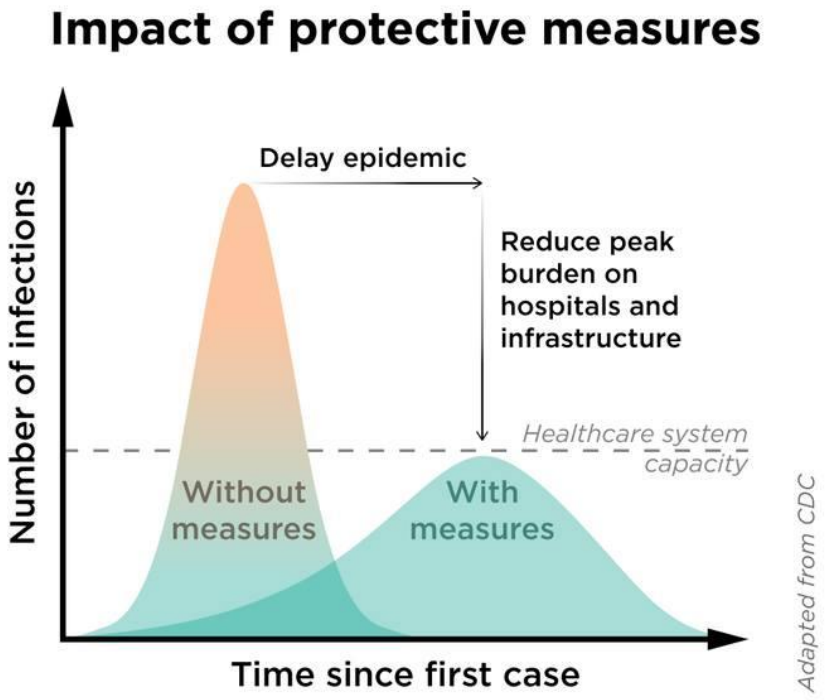
a = initial value (the amount before measuring growth or decay)
b = growth factor
x = number of time intervals that have passed

Recall our example that started with an initial value of 500. Try calculating the number of cases after one month again. The number of time intervals is 10, so you can use the formula:

$$500 * 2^{10} = 512000$$

Using the formula makes it easier, and now we can make changes quickly based on different parameters. Let's say that the growth factor is 3 instead of 2, because people are going to sports arenas and not washing their hands.

4. If the growth factor is 3, then how many cases do we get after 1 month (10 doublings)?
5. If 5% of those people are seeking medical attention, then what do you expect to happen to the medical system? Assume that doctors get sick at the same rate as others, so at least 5% of them cannot work due to symptoms.



The solution our society is choosing is to #FlattenTheCurve

A good picture has been circulating throughout this discussion. This powerful example of visual analytics has driven huge changes in the American way of life. Something to keep in mind as you consume news, especially if you do so through social media, is the role of visual analytics. People are using slick graphs to drive policy. Visual analytics is more than descriptive statistics; visual analytics can communicate a trend.

We can't stop time. We can't reduce the number of active cases. We may be unwilling to completely prevent the spread of the virus. The variable in the exponential growth model that we have control over is the growth factor. If a sick person only infects 1 person in 3 days instead of 3 people in 3 days, then there is a big difference.

Are you surprised that 1 versus 3 can make a difference? It can be hard to grasp exponential growth.

I'm going to quote extensively from a recent article by Tyler Cowen called "Bill Gates Is Really Worried About the Coronavirus. Here's Why" I suggest you read in full at https://www.bloomberg.com/opinion/articles/2020-03-03/how-fast-will-the-new-coronavirus-spread-two-sides-of-the-debate

> Just how bad will the new coronavirus be? I can't answer that question, but I have observed the debate splitting into two broad camps: Call them the "growthers" and the "base-raters."
>
> The term growthers refers to the notion of exponential growth, and indeed the number of Covid-19 cases appears (by some accounts) to be following an exponential pattern. Some scientists have estimated that the number of cases doubles about every seven days. If you play that logic out, it is easy enough to see how people might be complacent at first, then in a few months there is a public health crisis.
>
> Of course, that process of doubling won't go on forever. At some point, the number of people who have already been exposed to Covid-19 would become so large that their immunity could lower the subsequent rate of spread. Furthermore, society would adjust by having fewer large gatherings — many conferences already are being canceled — and by taking other precautions.
>
> Still, the growthers find it easy to imagine that the number of cases might overwhelm the capacity of the U.S. health care system. Even if you think a speedy American (or more likely Singaporean) response argues against this scenario, it is harder to be equally sanguine about all the world's nations, most of which are much poorer and have lower-quality public health systems than the U.S.
>
> The growther approach seems most common among people trained in mathematics, finance, and those who work in technology. Finance is centered on the idea of exponentially compounding returns, where small initial gains turn into something quite large. So financial professionals understand the growther perspective.

That is why something that seemed like a problem for other countries in February of 2020 and then seemed like a very minor problem in the United States is suddenly impacting our lives in a huge way.

The good news for college students is that the disease has a very low death rate for young adults. The thing that young people can do is to help stretch out the length of time the disease takes to make its way through the population. That allows medical professionals to give care to those who need it.

One thing that scientists are working on right now is how to parametrize the model correctly, so that they can make the most accurate predictions about how the disease will progress. We do not know precisely how long it will take for the next doubling to occur in the United States, but we do know how to make it take longer through social distancing.

Comprehension questions

If we flatten the curve, we will achieve:

a. Completely stop the spread of the virus
b. Slow the spread of the virus so that sick people can get treatment
c. Get everyone sick as fast as possible
d. Increase the doubling time window

The #FlattenTheCurve visual communicates:

a. The current number of confirmed cases
b. The number of predicted cases
c. The way the trend of the virus affects the medical system
d. The number of hospital beds in use over time