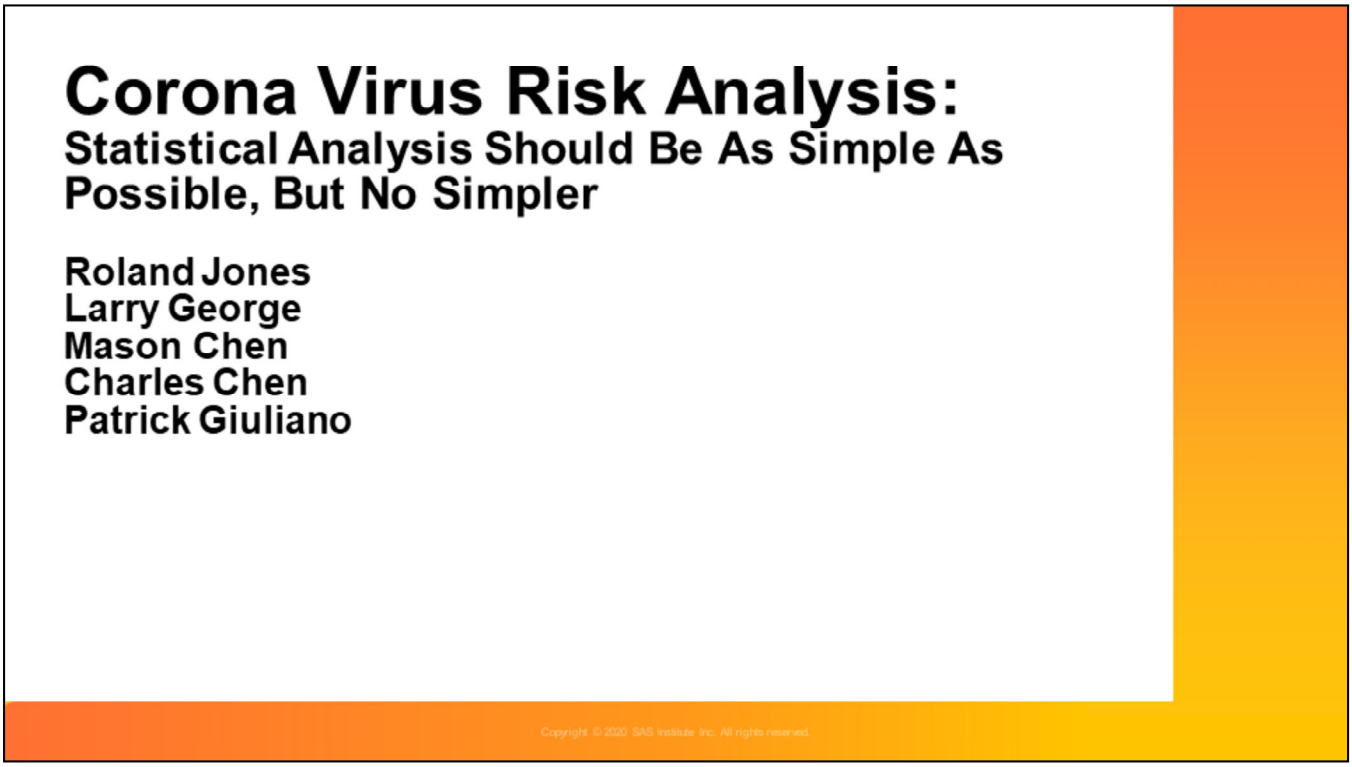
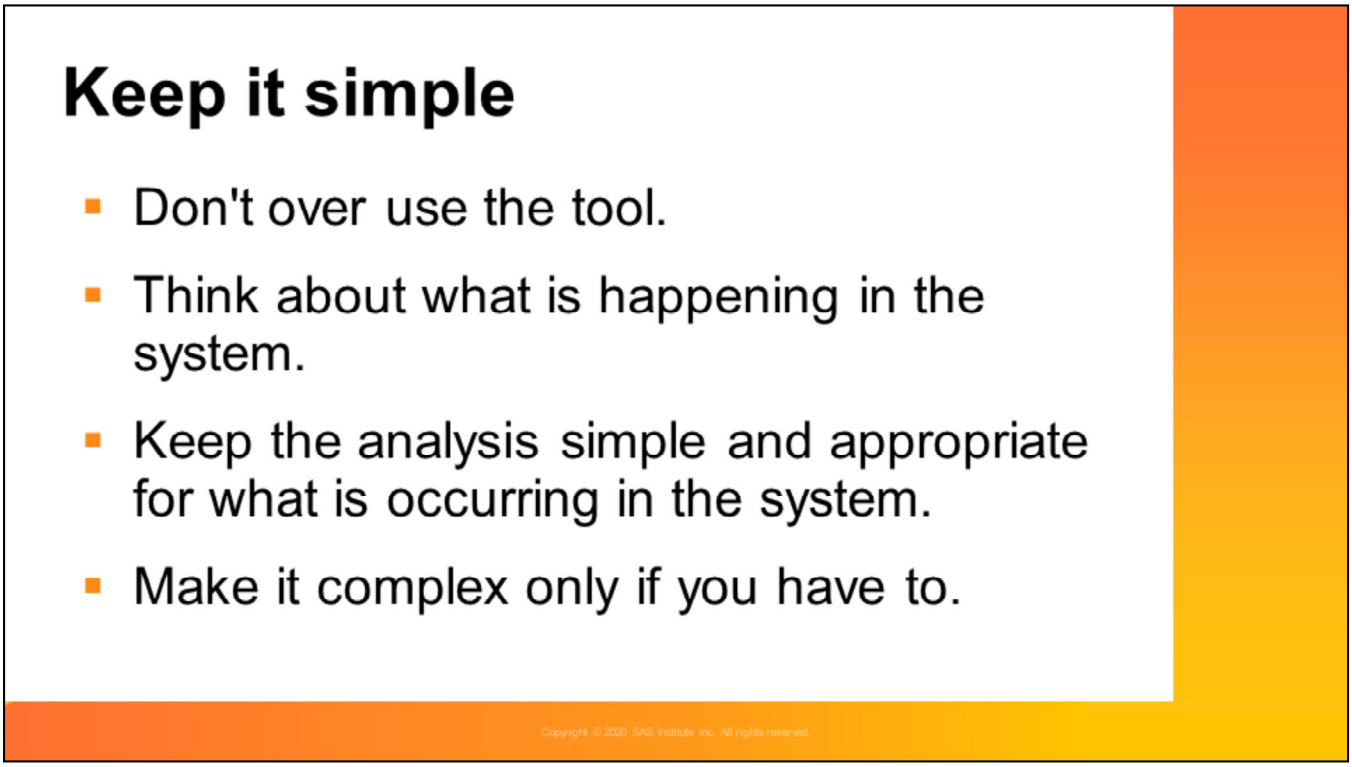


When we put together our abstract proposal at the beginning of May I was concerned that Covid-19 would be old news by October. At the time of recording, the 21st of August, this is far from the case. I really hope that by the time you watch this in October things will be well under control and life will be returning to normal, but I suspect that it won't.

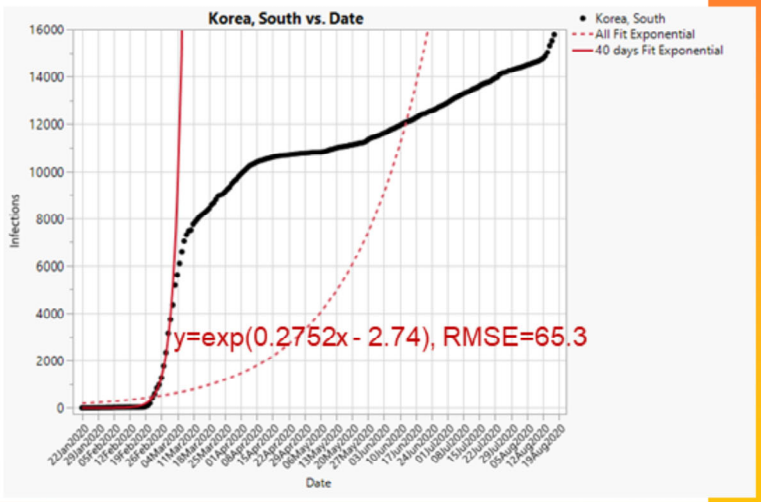


With all the power of JMP it is tempting to throw the data into the tool and see what comes out. The Covid-19 pandemic is an excellent case study of why this should not be done. The complications of incomplete and sometimes manipulated data, changing environments, changing behavior, and changing knowledge and information, these make it particularly dangerous to just throw the data into the tool and see what happens. Get to know what is going on in the underlying system. Once the system is understood, the effect of the factors I have listed can be taken into account, allowing the modeling and analysis to be appropriate for what is really happening in the system, avoiding analyzing or being distracted by the imperfections in the data.

It also makes the analysis simpler. The overriding theme of this presentation is keep things as simple as possible, but no simpler. There are some areas towards the end of the presentation that are far from simple, but even here, we still work to keep things as simple as possible.

Approach #1: Partitioning the data

- What function fits?
- Try Exponential
- Not a good fit
- Decent fit to just the first 40 days



We started by looking at the outbreak in South Korea

- It had a high early infection rate
- It is a trustworthy and transparent data source

All data in this presentation comes from the Johns Hopkins database as it stood on the 21st of August when this presentation was recorded.

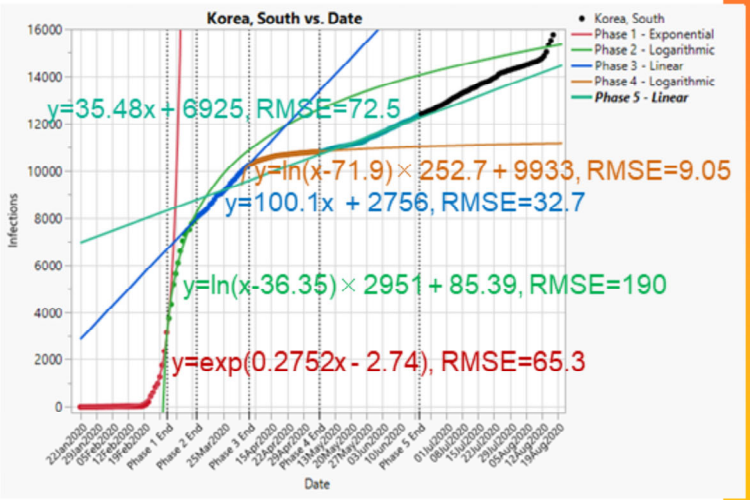
This is a difficult data set to fit a trend line to [It is actually worse that it looks from this graph, but I will get onto this in a moment]. We know that disease naturally grows exponentially, so try fitting an exponential.

This is not a good fit, and it is difficult to see how any function could fit the whole data set.

Something that looks like an exponential can be seen in the first 40 days, so lets fit to just that. There is a good exponential fit. We can partition the data into different phases and fit a function to each phase separately.

Partition and Fit separately

- 5 phases chosen
- Best fit function selected for each phase



5 partitions were chosen for the data as it stood on the 19th of June, resulting in 5 phases

Partitions were chosen where the curve seemed to transition to a different kind of behavior

Parameters in the fit functions were optimized using JMPs nonlinear fit tool (details of how to use this tool are in the appendix). Nonlinear also produced the Root Mean Square Error result (the sigma of the residuals).

Good fit for each phase (root mean square error is impressively low). However, as partition points were specifically chosen where the curve changed behavior, low RMSE is to be expected.

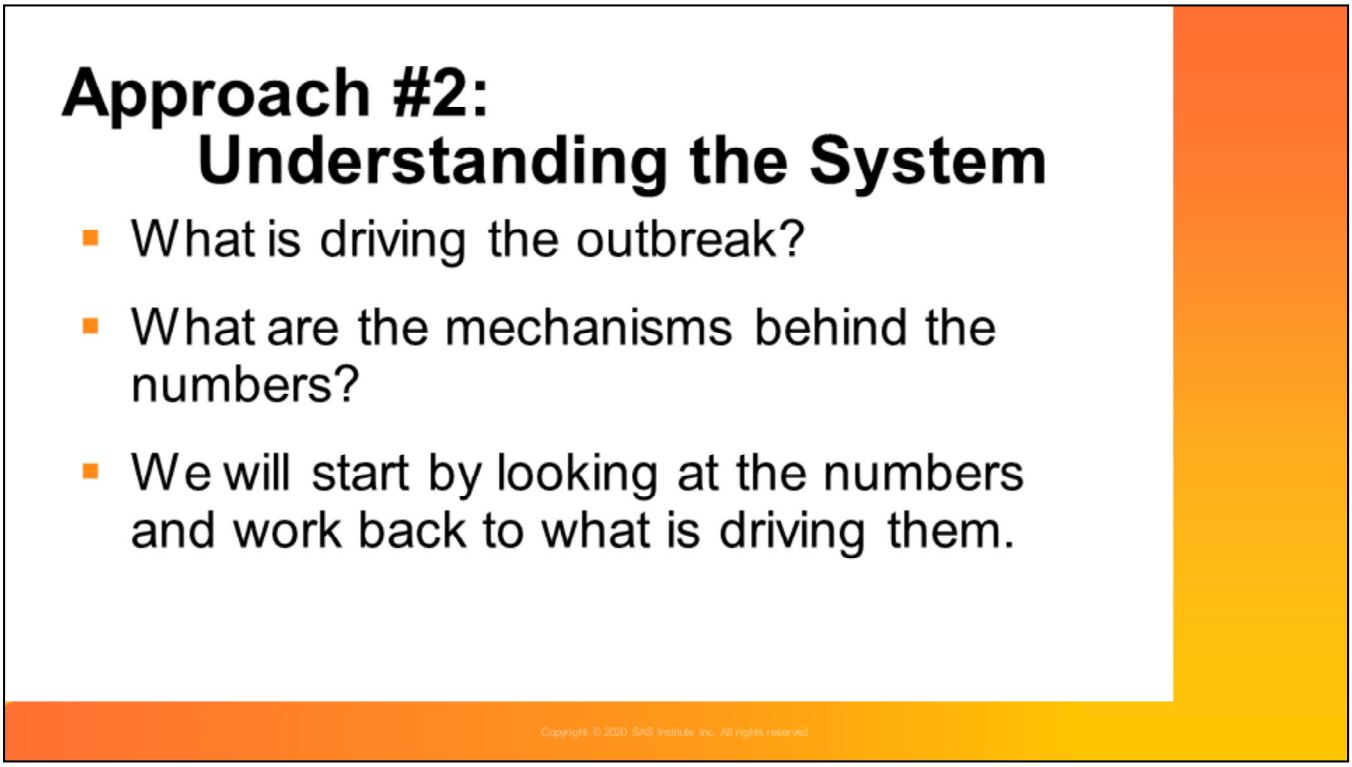
Trend lines have negligible predictive ability, because the partition points are chosen looking at the existing data. This can be seen in the data present since the analysis was performed on the 19th of June. With the extra data available, we could choose different partition points, and get better fit, but this would not help to predict beyond the new data.

Partition points do show where the outbreak behavior changes, but this could be

seen before all the analysis was performed.

No indication is given as to why the different phases have different fit functions.

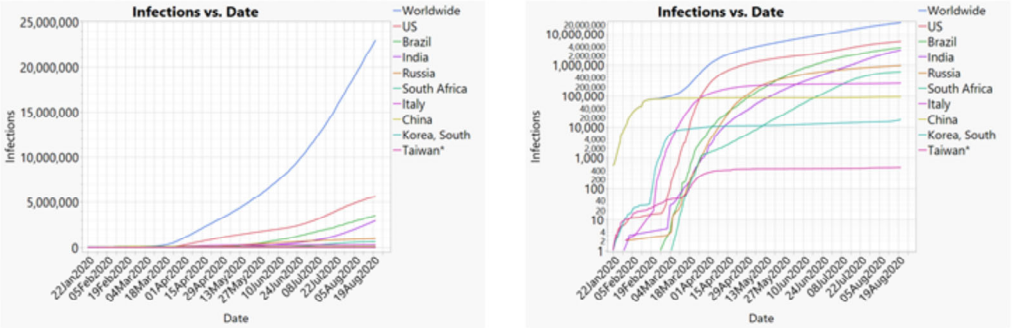
This exercise does illustrate the difficulty of modeling the outbreak, but does not give us much useful information on what is happening, or where the outbreak is heading. We need something simpler.



We are dealing with a system that contains self learning. As we, as a society, learn more about the disease, we modify our behavior to limit its spread, changing the outbreak trajectory. Lets look into the mechanics of what is driving the outbreak, starting with the numbers themselves, and working backwards to see what is driving them.

Data Representation: Graphs

- Graphs not numbers, but the right graphs



- Disease grows exponentially, so log graphs are appropriate, and show more detail

The news is full of Covid-19 numbers. USA hits 5 million infections and 150K deaths. California has higher infections than New York. Daily new infections in the US could top 100K. Individual numbers are not that helpful. Graphs help to put the numbers into context. The right graphs help us to see what is happening in the system.

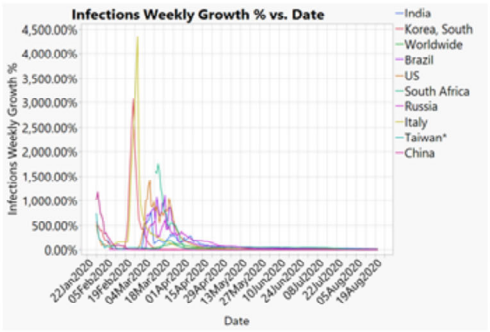
Disease grows exponentially. One person infects 2, who infect 4, who infect 8... Human eyes differentiate poorly between different kinds of curves, but differentiate well between curves and straight lines. Plotting on a log scale changes exponential growth and exponential decline into straight lines.

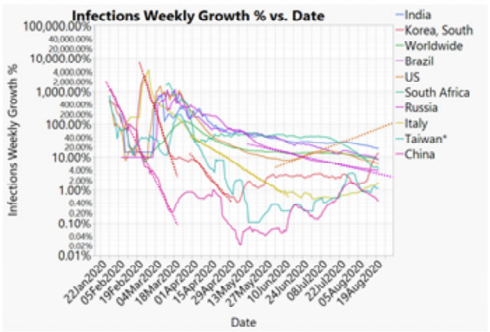
Also, on a log scale, early data is now visible. Many countries show 1 sometimes 2 plateaus which are not visible on a linear scale [Remember on the Korea graph, I said that it was more difficult to fit a function than the graph was showing]. How can we model this kind of behavior? Lets keep on digging.

The slope of the log infections graph is the percentage growth. Plotting percentage growth gives us more good information.

Data Representation: % Growth

- Again, a log scale helps to show details





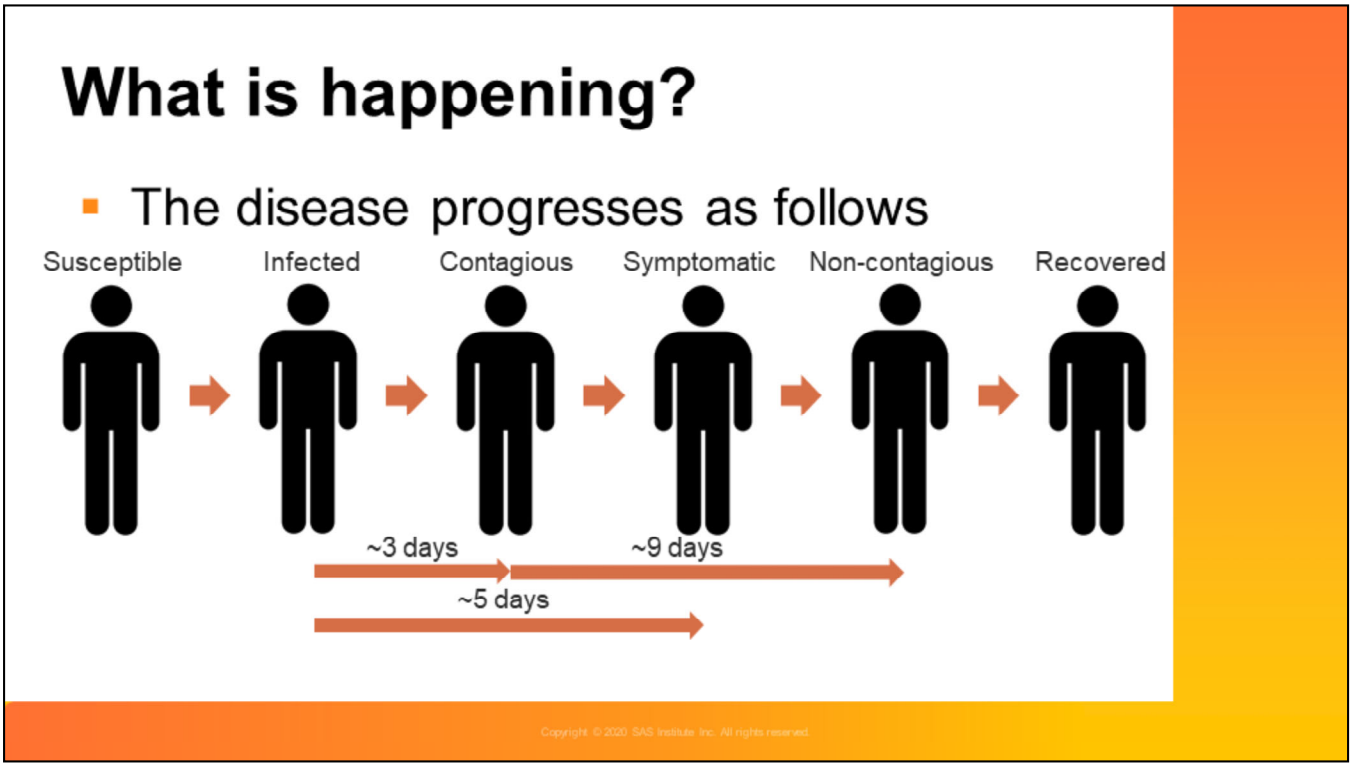
- Why does % growth decline exponentially?

Parentage growth helps to highlight where things changed. The decline in growth in the US can be seen to be slackening off from Mid April, and finally reversing a little after the 10th of June. This is visible, but is not as clear in the infections graphs. It is much clearer in the % growth graphs.

Interesting observation. Many countries show linear decline in % growth when plotted on a log scale. Italy is a particularly fine example of this, but it can also be seen in China, South Korea, and Russia. Why is this happening? Intuitively, I expected that when behavior changes, growth would drop down to a lower % and stay there, not exponentially decline toward zero.

I started plotting graphs of Covid-19 back in late February, not to predict the outbreak, but because I was frustrated with the graphs that were being published. After seeing this linear decline in % growth I started paying an interest in prediction. Extrapolating the % growth line through linear regression works pretty well as a predictor, but only works when the growth is declining. It does not work at all well when the growth is increasing. If we extrapolate the US growth line from the 17th of June to the 1st of July it predicts that we will be at 30% weekly growth by the 22nd of July, and 100% weekly growth by the 26th of August, and keep on growing beyond this. Cleary this model does not match reality.

I will come back to this exponential decline in percentage growth later. For now, lets keep looking at what is physically going on as the disease spreads.



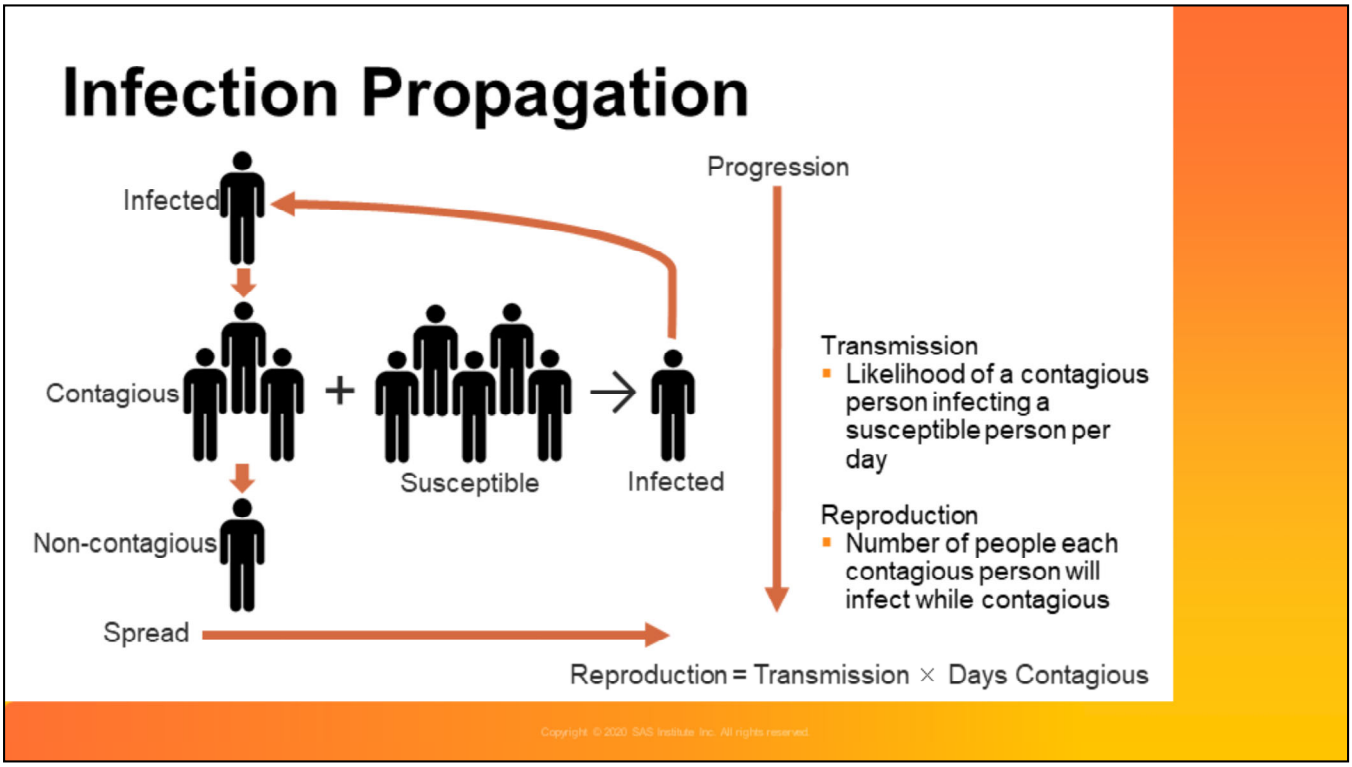
People progress from Susceptible → Infected → Contagious → Symptomatic → Non-contagious → Recovered

This is the Markov SIR model. SIR stands for Susceptible, Infected, Recovered. The three extra stages of Contagious, Symptomatic, and Non-contagious help us to model the disease spread and relate it to what we can measure.

Note the difference between infected and contagious. Infected means that you have the disease. Contagious means that you can spread it on to others. It is easy to confuse the two, but they are different and will be used in different ways further into this analysis.

The timings shown are best estimates and can vary greatly. Infected to symptomatic can be from 3 to 14 days, and some infected people are never symptomatic.

The only data that we have access to is confirmed infections, which usually come from test results which usually follow from symptomatic. Even if testing is performed on non-symptomatic people, there is an ~5 day delay between infection and positive test results, so we are always looking at old data. We can never directly observe the true number of people infected.



The disease progresses from top to bottom.

We have a pool of contagious people, fed by infected people becoming contagious, and drained by contagious people becoming non-contagious.

The disease spreads from left to right.

New infections are created when susceptible people come into contact with contagious people and becoming infected. Infected people join the queue waiting to become contagions, and the cycle continues.

This cycle is controlled by transmission, how likely a contagious person is to infect a susceptible person each day.

If we know how many days a person remains contagious, we can derive reproduction, how many people a contagious person is likely infect while they are contagious.

The whole cycle revolves around the number of people contagious and the

transmission/reproduction. The time individuals stay in the contagious should be constant, unless Covid-19 starts to mutate. The transmission can vary dramatically depending on social behavior and the size of the susceptible population [can be reduced through herd immunity or vaccination].

Some Math

Days Contagious (DC) ≈ 9

People contagious ≈ last 9 days infections

Percentage Growth in Contagious (PGc)

- Derived from people contagious trend

$$\text{Transmission} = \frac{PGc \times (PGc+1)^{DC}}{(PGc+1)^{DC}-1}$$

Reproduction = Transmission × DC

Our best estimate is that days contagious averages out at about 9.

We can estimated people contagious as the number of people confirmed infected in the last 9 days. In some respects this is an under estimate because it does not include people that are infected and not yet symptomatic, or that are asymptomatic, or that do not yet have positive test results. In other respects it is an over estimate because it includes people who were infected a long time ago but are only now tested positive. It is an estimate.

From the estimate of people contagious we can derive the percentage growth in contagious. It does not matter if people contagious is an over estimate or an under estimate. As long as the percentage error in the estimation remains constant, the percentage growth in contagious will be accurate.

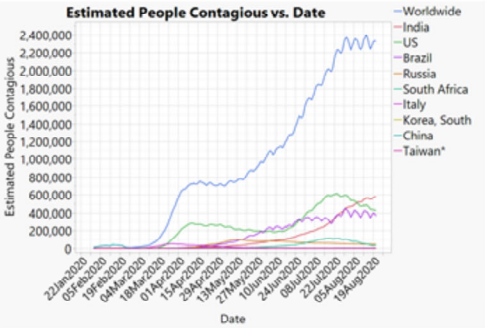
Percentage grown in contagious is important because we can use it to derive transmission. The derivation of the equation relating the two can be found in the appendix. Note that this equation allows you to derive Transmission and then Reproduction from Percentage Growth in Contagious, but can not tell you the Percentage Growth in Contagious for a given transmission. [needed if you want to see how the outbreak is expected to progress for a given Reproduction]. It can only be found by solving numerically. I have outlined how to do this using JMPs Fit Model

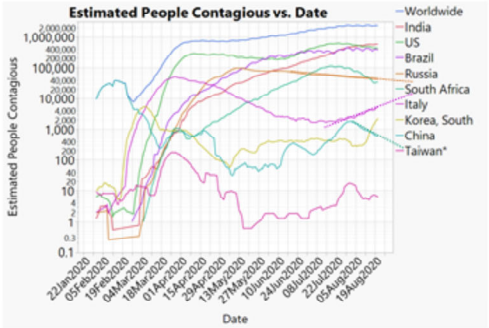
tool in the Appendix.

Reproduction and Transmission are very closely linked, but reproduction has the advantage of ease of understanding. If it is >1 the outbreak is expanding out of control, infections will continue to grow and there is no end in sight. If it is less than 1, the outbreak is contracting coming under control. There are still new infections, but their number will gradually decline until they hit zero. The end is in sight though it may be a long way off.

Back To The Graphs

- People contagious drives the outbreak





The number of people contagious is the underlying engine that drives the outbreak.

People contagious grows and declines exponentially. We can predict the path of the outbreak by extrapolating this growth or decline in the people contagions.

Remember the interesting observation that infections percentage growth declines exponentially. Here is why. If reproduction is less than 1 and constant, people contagious will decline exponentially toward zero. People contagious drives the outbreak. The percentage growth in infections is proportional to the number of people contagious. So if people contagious declines exponentially, the percentage growth in infections will also decline exponentially. Mystery solved.

The slope of people contagious plotted on log scale gives us contagious percentage growth, which then gives us transmission and reproduction through the equations on the last slide.

Notice the weekly cycle in the data, particularly clear for Brazil, but also visible in other countries. This could be due to numbers getting reported differently over the weekend, or people being more likely to get infected at the weekend. Either way, we will have to take this seasonality into account when using people contagious to

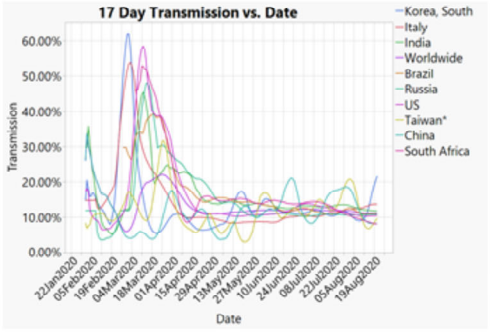
predict the outbreak.

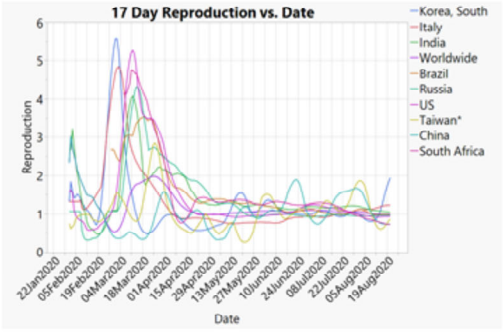
Because social behavior is constantly changing, transmission and reproduction changes as well, so we can not use the whole distribution to generate reproduction. We chose 17 days as the period over which to estimate reproduction. We found that 1 week was a little to short to filter out all the noise. 2 weeks gave better results. 2.5 weeks was even better. Having the extra half week evened out the seasonality in the data.

There is a Time Series Forcast tool in JMP that will do this for us, including the seasonality, but because we are performing the regression on small sections of the data, we did not find the tool helpful.

Transmission & Reproduction

- These tell us where the outbreak is heading





- Reproduction > 1 means people contagious is growing, the outbreak is expanding

Here are the derived transmission and reproduction numbers. You can see that they can change quickly.

It is easy to get confused by these numbers. South Korea is showing a significant blip in reproduction, but is doing well. The USA, Brazil, India, and South Africa are all doing badly, but seem to have reproduction close to or less than 1.

Aid to Understanding the Numbers: A Little Bit of Calculus

- We are all familiar with distance, speed and acceleration

Parameter	Description	Pandemic Equivalent
Distance	How far traveled	Number of infections or deaths
Speed	Rate of change of distance	New infections or deaths
		People contagious
Acceleration	Rate of change of Speed	Transmission
		Reproduction

To help reduce the confusion around reproduction, a little calculus.

Driving a car, the gas pedal controls acceleration. To know where you are going to be, you need to know where you are, how fast you are going, and how much you are accelerating or decelerating.

To know where the pandemic is going to be, we need to know;

- how many infections there are (the equivalent of distance traveled)
- how fast the infections are expanding, or how many people are contagious (the equivalent of speed)
- how fast the people contagious is growing, that is the transmission or reproduction (the equivalent of acceleration).

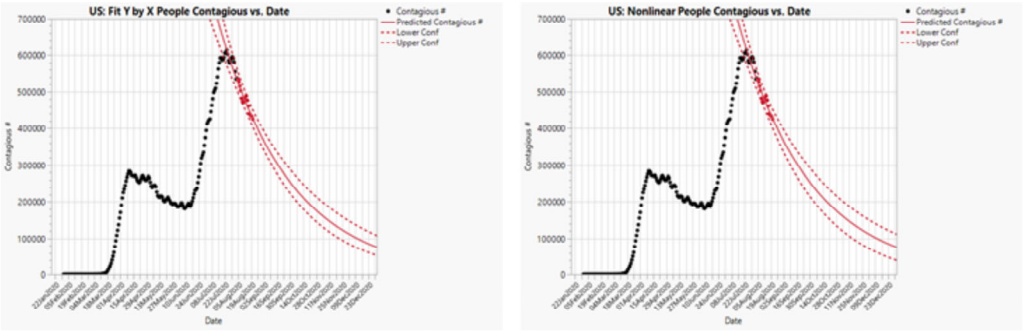
There is a slight difference. Distance grows linearly with speed, and speed grows linearly with acceleration. Infections grow linearly with people contagious, but people contagious grows exponentially with reproduction. There is a slight difference, but the principals are the same.

The USA, Brazil, India, and South Africa have all traveled a long distance (high

infections) and are traveling at high speed (high contagious). Even a little bit of acceleration has a very big effect on the number of infections.

South Korea is not going fast (low contagious), so has headroom to respond to acceleration and get things back under control without covering much distance. Also, when the number of people contagious is low, adding in a small number of new people produces significant acceleration. Countries that have things under control are prone to blips in reproduction. You have to take all 3 factors into account (number of infections, people contagious, and reproduction) to decide if a country is doing well or doing poorly.

Regression Comparison



- Prediction lines are not that different
- Nonlinear confidence lines are much better

Within JMP there are a couple of ways to perform the regression to get percentage growth in contagions, the Fit Y by X tool, and the Nonlinear tool.

I have details on how to use both these tools in the Appendix, but lets compare the results they produce.

These graphs compare of the results from both tools. The 17 data points used to make the prediction are shown in red.

The prediction lines from the two methods are just about identical, though there are some noticeable differences in the confidence lines.

The confidence lines from the nonlinear tool are much better. The Fit Y by X tool transposes the data into a linear space before finding the best fit straight line. This results in the lower confidence line pulling closer to the prediction line after transposing back. Confidence lines are not that useful when the parameters that define the outbreak are constantly changing. Best case, they will help you to see when the parameters have definitely changed.

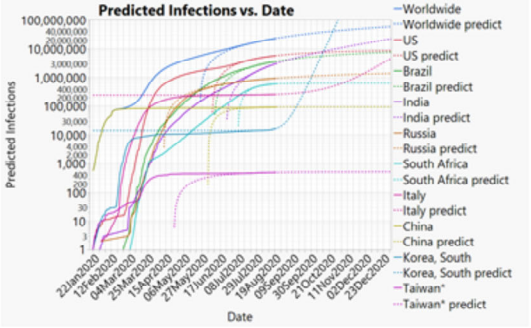
In my scripts, I use linear regression calculated in column formulas because it is

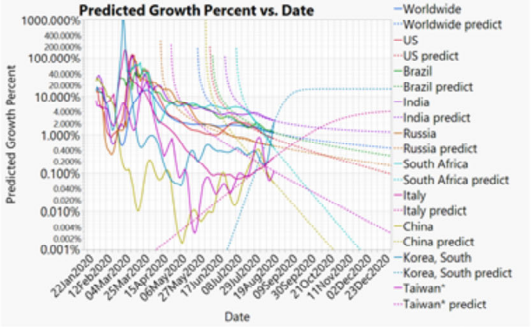
easy to adjust with variables. This allows the analysis to be adjusted on the fly without having to pull up a tool in JMP. I do not currently use confidence lines in my analysis, though I am working to integrate them into the column formulas.

Linear regression is simpler and produces almost identical results. Keep it simple.

A little more math

New Infections = People Contagious × Transmission



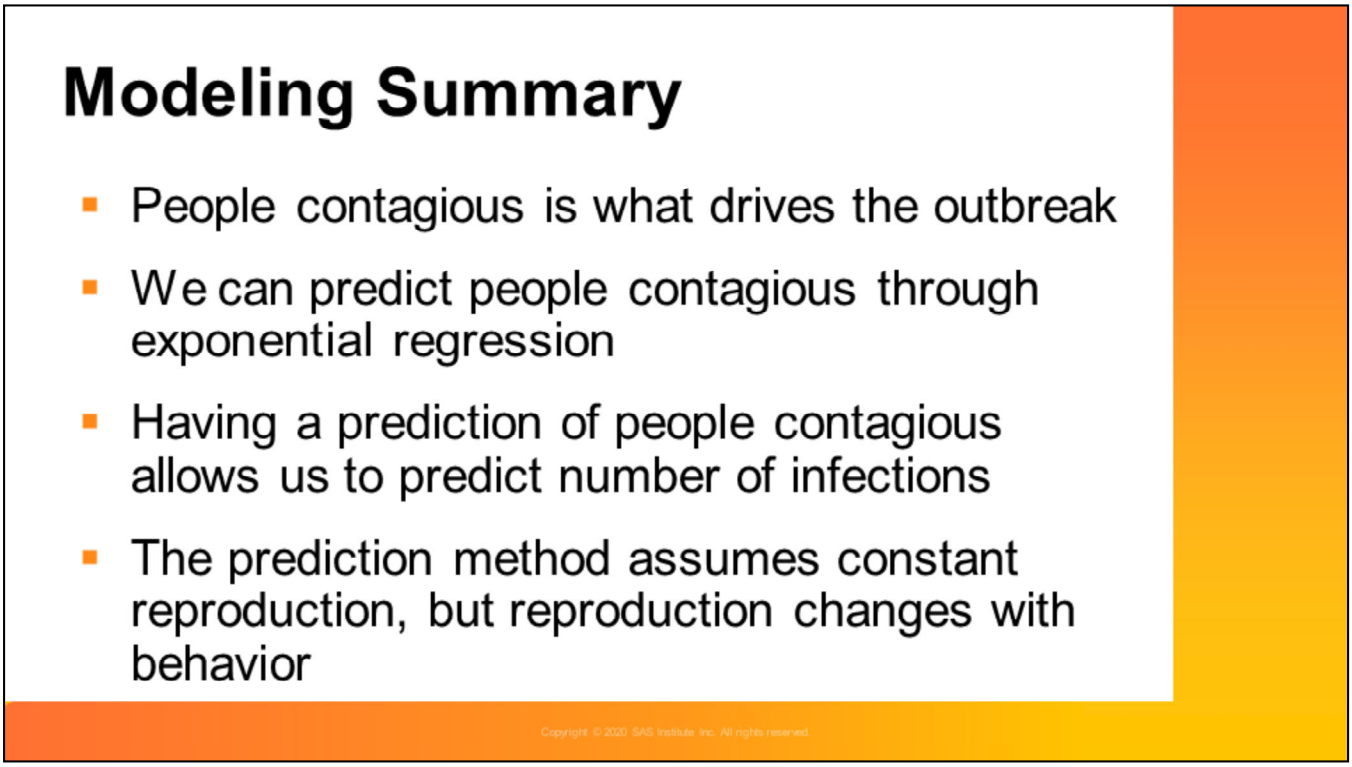


We have seen how fitting an exponential to the number of people contagious can be used to predict where the people contagious will be in the future, and also to derive transmission.

Now that we have a prediction line for people contagious, we need to convert that back into infections. New infections = people contagious × transmission. Remember, transmission is the probability that a contagious person will infect a susceptible person each day.

In the predicted infections graph that result from this calculation, note that South Korea and Italy have low infections growth. However, they have high reproduction extrapolated from the last 17 days data. The model assumes that reproduction will not change, and the high reproduction results in high growth in 2 to 8 weeks time. For South Korea this is unlikely to happen because they are moving slowly and have the headroom to get things back under control. South Korea has had several of these blips as it opens up, and always manages to get things back under control.

In the predicted growth percent graph on the right, note how the increasing percentage growth in South Korea and Italy do not carry on increasing indefinitely, but plateau out after a while. Percentage growth is still seen to decline exponentially, but it does not grow exponentially, it grows and then plateaus out.



So, to summarize

The number of people contagious is what drives the outbreak. This metric is not normally reported, but is close to the number of new infections over a fixed time period. New infections in the past week is the closest regularly reported proxy to the number of people contagious. This is what we should be focusing on, not the number of infections, or daily new infections.

Exponential regression of people contagious will predict where the contagious numbers are likely to be in the future.

The percentage growth in contagious gives the transmission [likelihood of a contagious person infecting a susceptible person per day] and reproduction [the number of people a contagious person is expected to infect while contagious]. The contagious number and transmission number can be combined to predict the number of new infections in the future.

The prediction method assumes that transmission and reproduction are constant, which they are not. They change with behavior, but the predictions are still useful to show what will happen if behavior does not change, or how much behavior has to

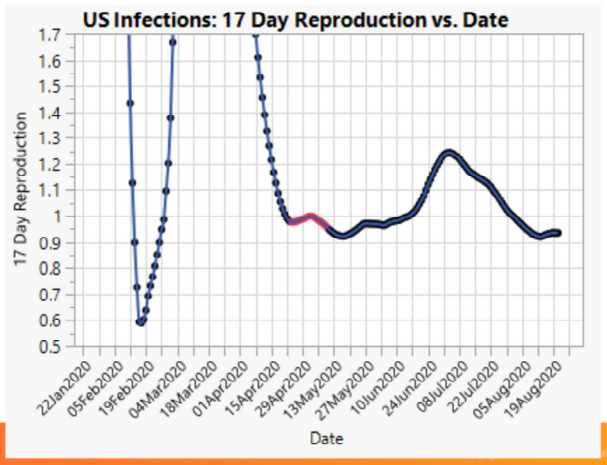
change to avoid certain milestones.

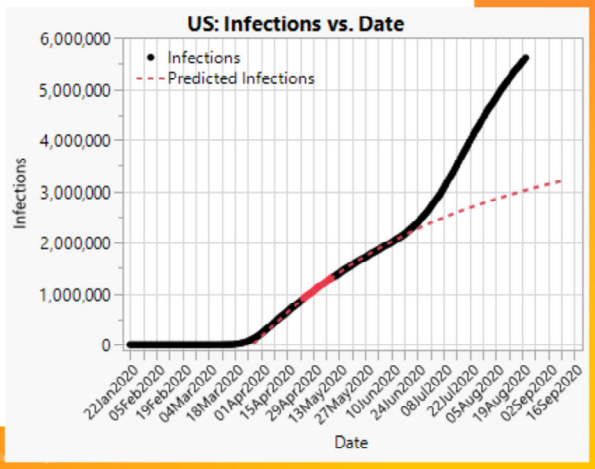
The only way to close this gap is to come up with a way to mathematically model human behavior. If any of you know how to do this, please get in touch, we can make a lot of money, though only for a short time.

That is the modeling. Lets check how accurate it is by looking at historical data from the US.

Prediction accuracy

- Works well when reproduction is constant, but not when reproduction is changing



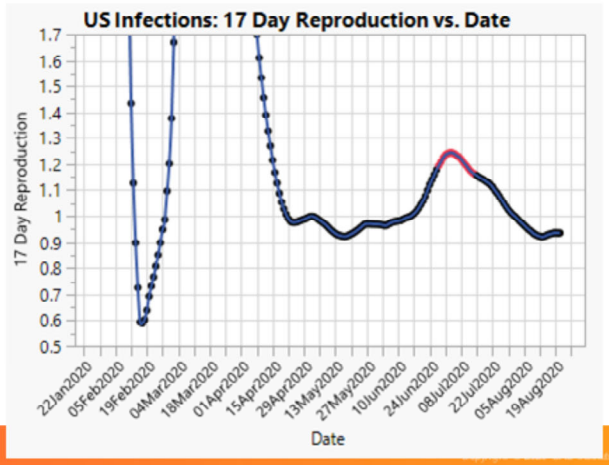


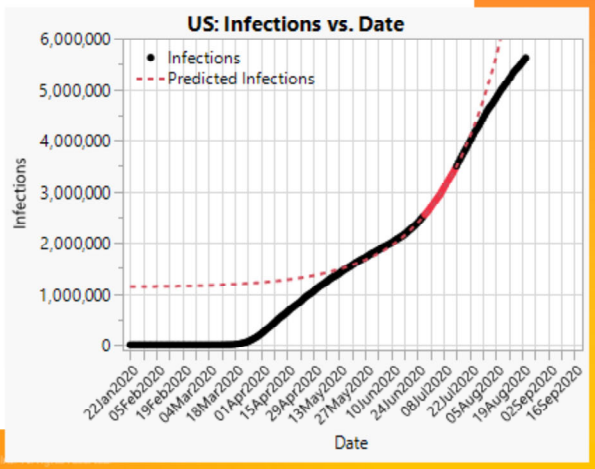
The prediction works well when reproduction is constant, but not when it is chancing.

The US prediction based on data from late April to early May is accurate as long as reproduction stays at around the same level of 1.0. As reproduction starts rising in mid June as relaxation of social restrictions had an effect, the prediction under estimates infections.

Prediction accuracy

- Does not work well when reproduction is changing



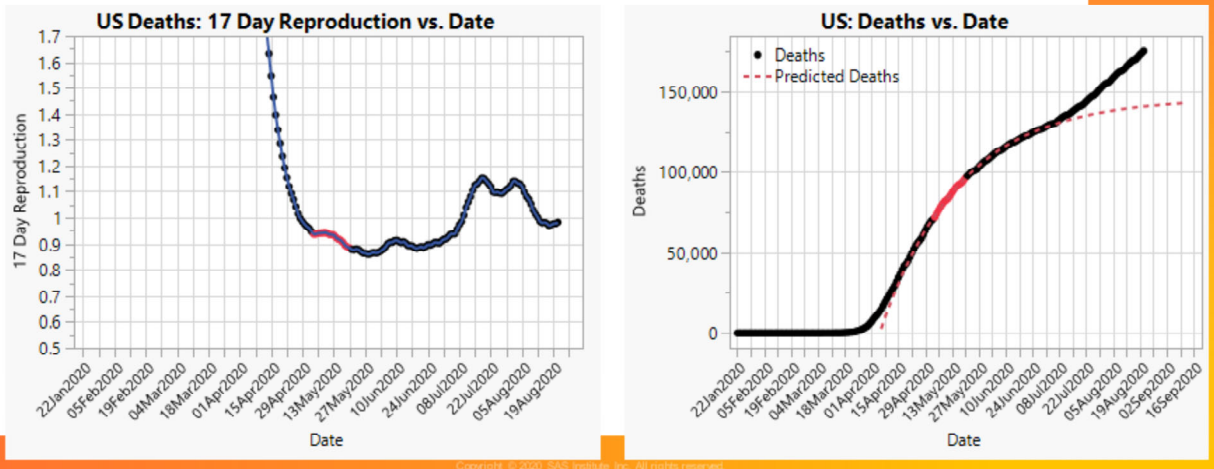


The prediction based on data from late June to mid July when reproduction was at its peak as states were closing down again, that prediction over estimates the infections as reproduction comes down.

This model is good for predicting what will happen if behavior stays the same, but not when behavior is changing.

Deaths Prediction

- The same math applied to deaths also works

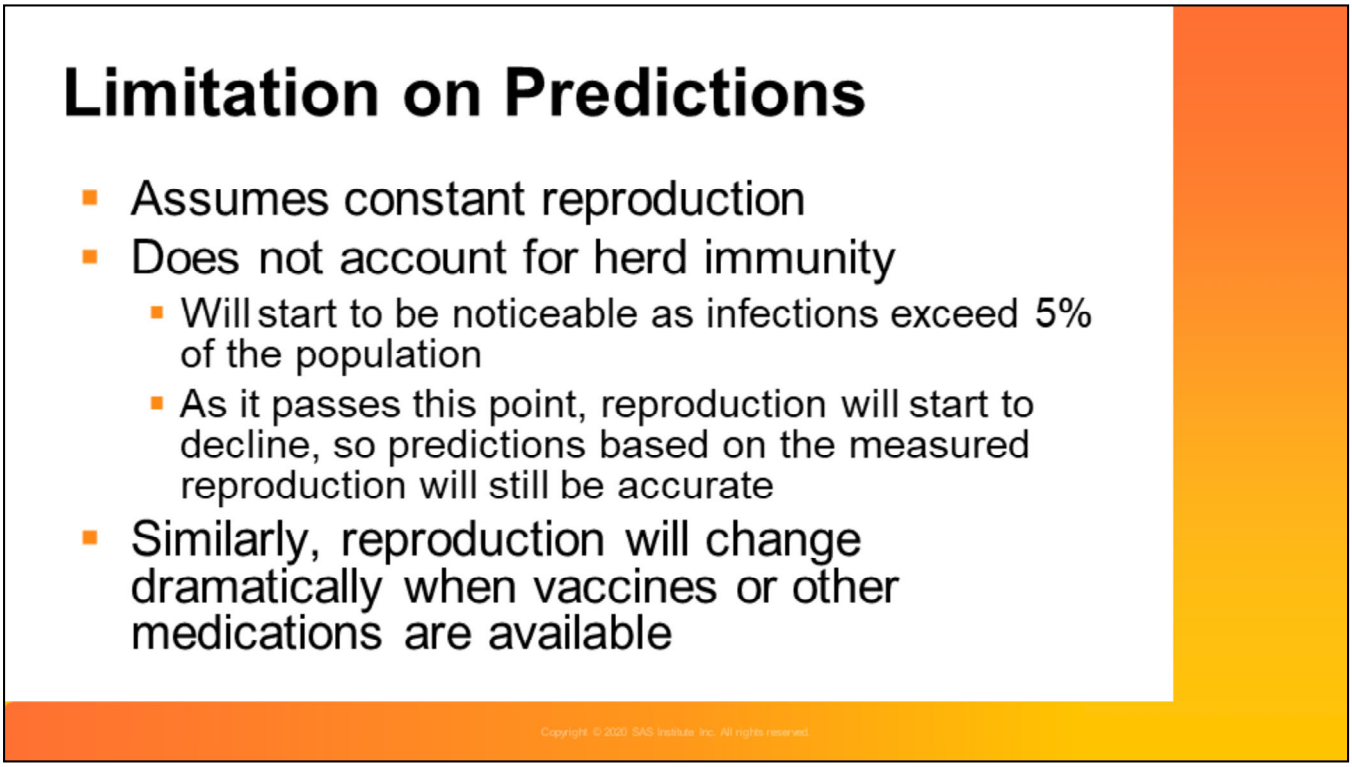


How can we predict deaths?

It should be possible to estimate the delay between infections and deaths, and the proportion of infections that result in deaths, and use this to predict deaths.

However, changes in behavior, such as increasing testing and tracing, skews the number of infections detected. To avoid this skew also feeding into the prediction for deaths, we can use the exact same mathematics on deaths that we used on infections. As with infections, the deaths graphs show accurate predictions when deaths "reproduction" is stable. Note that contagious and reproduction numbers for deaths do not represent anything real. This method works because deaths follow infections and so follows the same trends and the same mathematics.

Again, keep it simple



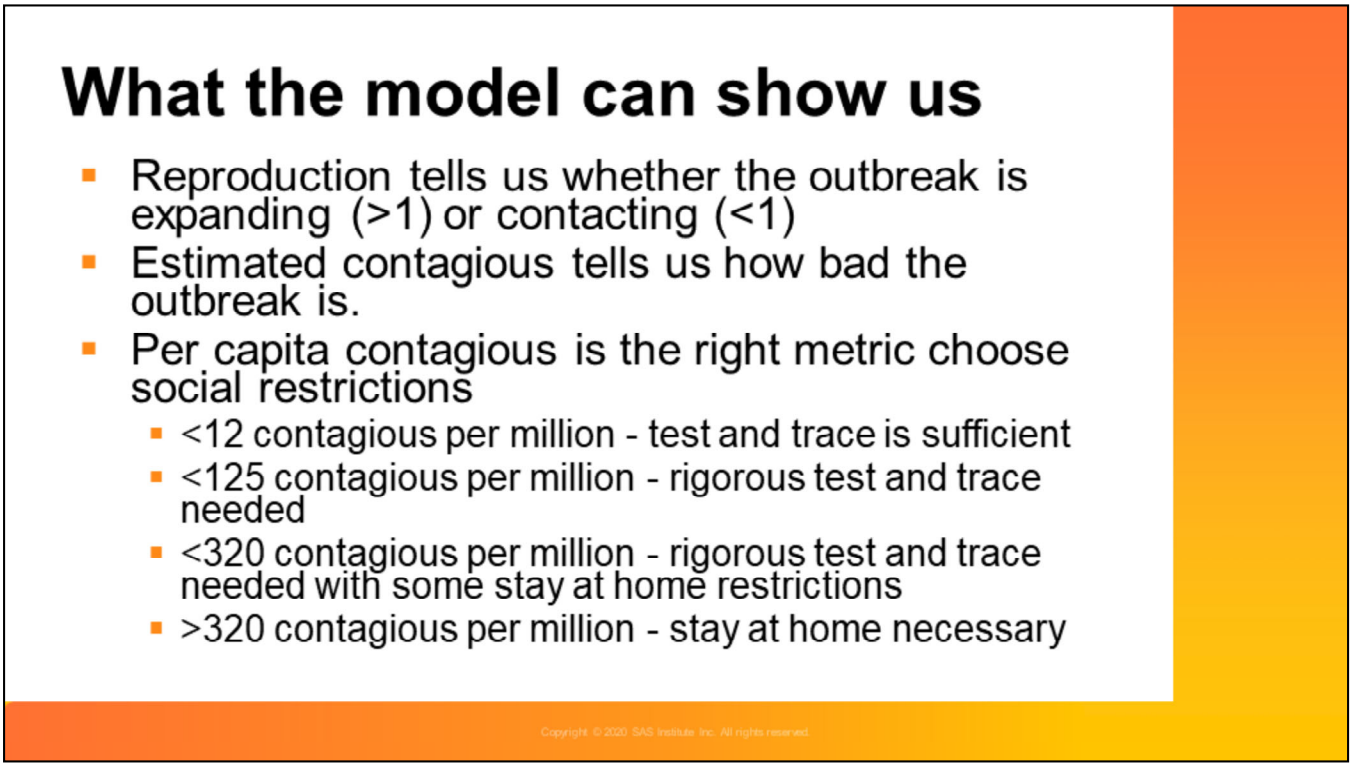
We have already seen that the model assumes constant reproduction.

It also does not take into account herd immunity. We are fitting an exponential, but the outbreak really follows a binomial distribution. Binomial and a fitted exponential differ by less than 2% with up to 5% of the population infected. Graphs demonstrating this are in the appendix.

When more than 5% of the population is no longer susceptible (due to previous exposure or vaccination), transmission and reproduction will naturally decline, so predictions based on recent reproduction numbers will still be accurate. However, long term predictions based on an old reproduction number with significantly less herd immunity will over estimate the number of infections.

[Changes due to new medications will similarly change reproduction, but again will be accurate when the new reproduction number is used.]

On 21st of Aug, the US had per capita infections of 1.7%. If only 34% of infected people have been diagnosed as infected (there is data that indicates this is likely) we are at the 5% level where herd immunity begins to have a measurable effect (reduces reproduction by ~2%)



What the model can show us.

Reproduction tells us whether the outbreak is expanding (>1, accelerating) or contacting (<1, decelerating). Estimated number of people contagious tells us how bad the outbreak is (how fast we are traveling). Per capita contagious is the right metric to choose appropriate social restictions

Recommendation for social restrictions are adapted from those published by the Harvard Global Health Institute (see appendix for details)

At the time of writing, the US had 1,290 contagious per million, down from a peak of 1,860 in late July

Reopening Comparison

Country/State	Date Reopened	Per Capita Contagious at Time of Reopening
China	21st of March	0.03 contagious per million
South Korea	28th of March	2 contagious per million
Germany	15th of May	85 contagious per million
Italy	18th of May	126 contagious per million
Spain	7th of June	63 contagious per million
France	11th of June	63 contagious per million
UK	4th of July	59 contagious per million
Arizona	16th of May	511 contagious per million
Texas	22nd of May	363 contagious per million
Florida	5th of June	412 contagious per million
California	12th of June	691 contagious per million

It is instructional to look at the per capita contagious in various countries and states when they decided to re-open.

China and South Korea had only a hand full of people contagious

Europe was in the tens of people, except for Italy

The US was in the hundreds. It should not have re-opened in May, it was an emotional decision, not a data driven decision.

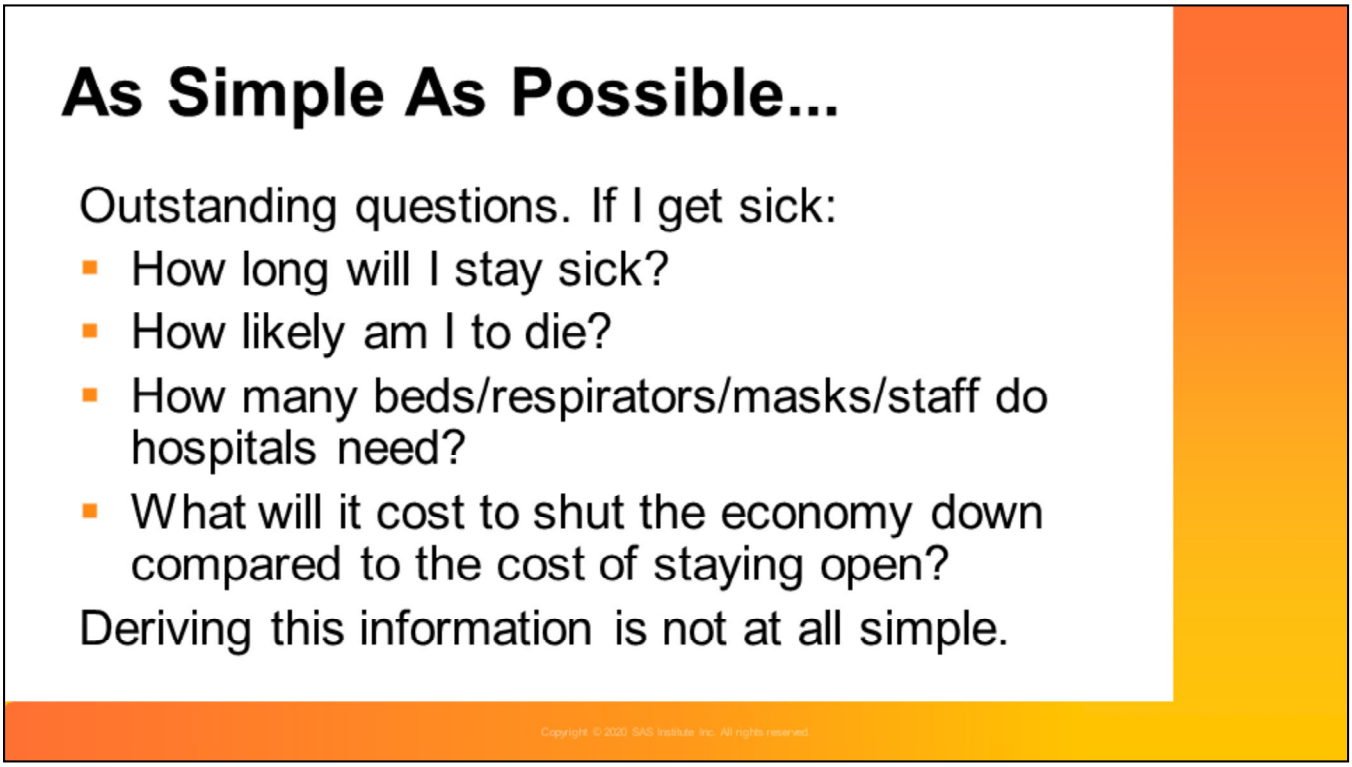
US Specific Reopening

- Per capita contagious for US as of the 21st of Aug was 1,290 per million with a reproduction of 0.94
- With this per capita contagions and reproduction it will take until 7/Dec to get below 320 contagious per million
- Lowest reproduction during the April lock down was 0.86.
- Even with a reproduction of 0.86 it will take until 8/Oct to get below 320 contagious per million
- If the USA had not reopened in May, we would have dropped below 320 contagious per million by the 20/Jun, and would have plateaued out at 135K deaths
- Currently deaths are projected to plateau out at around 270K

Prediction

Country	Infections		Deaths	
	21/Aug	13/Oct	21/Aug	13/Oct
Worldwide	23M	36M	800K	1.18M
USA	5.6M	7.3M	175K	220K
Brazil	3.5M	5.4M	113K	160K
India	3.0M	7.7M	56K	131K
Russia	940K	1.16M	16.1K	20K
South Africa	600K	640K	12.8K	18.7K

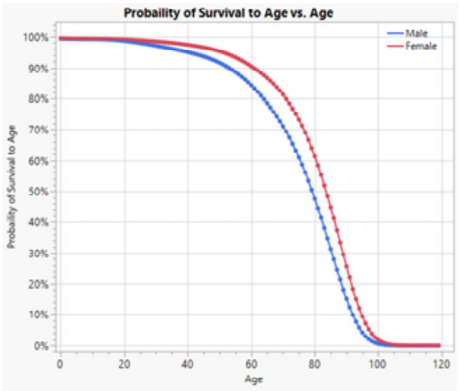
Based on our model, as of the time of writing, here is the model prediction on where the infections and deaths will be in the 5 most infected countries as the JMP Discover Summit opens in October.



We have a reasonably accurate way to predict the course of the outbreak, assuming that population behavior does not change significantly.

The model just gives us projected infections and projected deaths. This is good information, but not enough to plan the response to the outbreak. To do this we need to answer these "how" questions.

Actuarial and Survival Functions



- Similar graphs for Covid-19 give what we need
- Requires tracking individual patients, or does it...

Actuarial tables have been around since the Roman Empire (first documented use by Ulpain who was either calculating pension costs of retiring legionaries, or the value of slaves, it is not clear which). They tell us the likelihood of a person of a given age surviving the next year.

Simple math can convert these actuarial numbers to a survival function, the likelihood of a new born surviving to any given age.

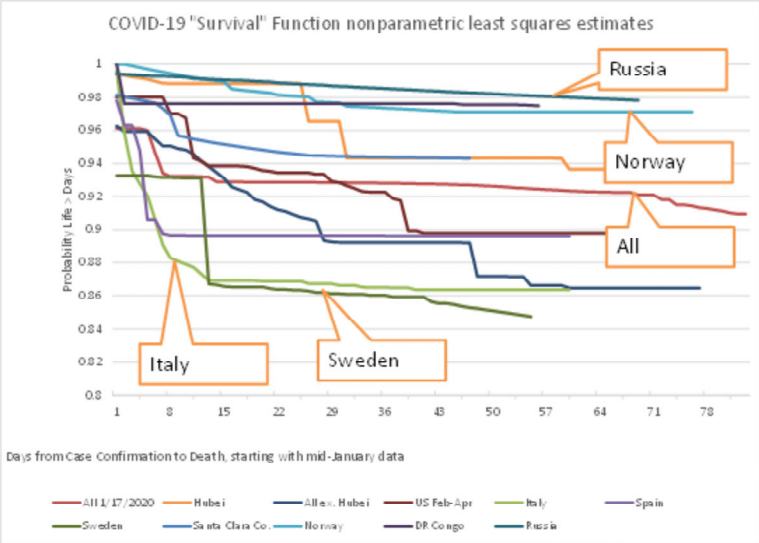
These graphs are the actuarial and survival functions for the USA as of 2017.

Similar survival graphs of Covid-19 patients would give us the logistical data that we are looking for, but generating them requires tracing individual patients though their disease to death or recovery.

However, the survival function can be estimated from infection and death/recovery data using non-parametric maximum likelihood or least squares optimization.

Covid-19 Survival Functions

- Can be estimated from infection, death and recovery numbers. Life data not needed.
- JMP does not have the functionality to do this (was pitched to SAS in 2006 by Larry George and Mark Felthauser)
- Excel solver used for this chart

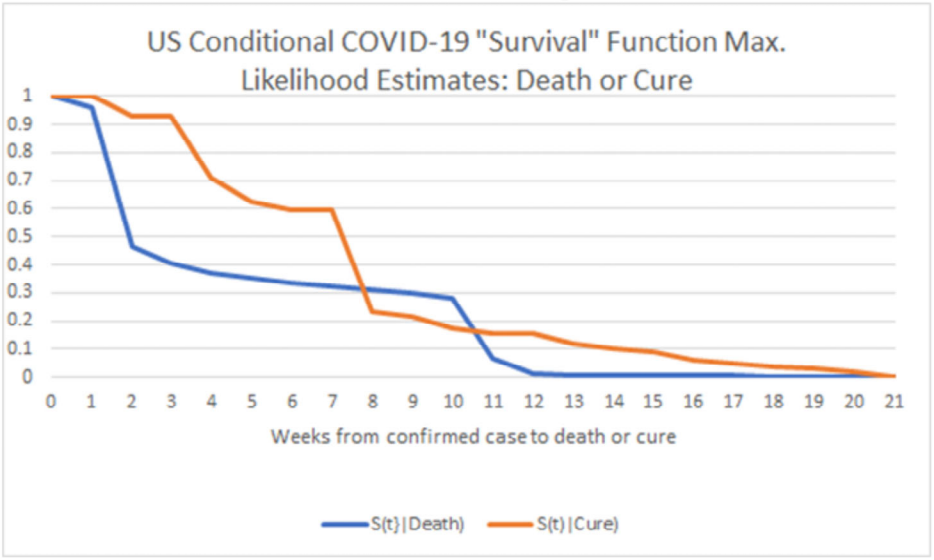


Generating this date is not at all simple. See references in the appendix for how to do this [Oscarsson and Hallberg (Ericsson) and Harris and Rattner (Virginia AIDS cases) used least squares to estimate survival functions from case and death counts.) See Larry George's web site for max. likelihood article (1973). Ren and Schuhegger simulated survival times and plugged them into Kaplan-Meier survival function estimator until they minimized SSE (sum of squared errors)]

JMP does not have the functionality to generate this data, it was generated using Excel solver. It was pitched to SAS in 2006 by Larry George and Mark Felthauser. If anyone watching wants to revisit adding it to JMP, please get in touch with Larry.

It was Larry who introduce me to this technique, and who kindly allowed me to present his work. He originally used it to estimate component reliability when components are not individually tracked. It has turned out to be highly applicable to disease characterization, modeling and planning.

Death and Recovery



Interestingly, death and recovery data can be plotted from the same data set. [case, death, and recovery counts, by max. likelihood, assuming M(t)/G/infinity (nonstationary Poisson case counts) for both].

Going back to its original use on component reliability, a similar method can be used to estimate survival function by failure mode.

Looking at the graphs, if you are going to die, it is most likely that you will die during week 2 after diagnosis, with week 11 being the next most likely.

If you are going to recover, it is most likely that you will recover during week 8, though week 4 is the next most likely.

Generating survival function from infections, deaths and recovery data is possible, but is very computationally intensive, and susceptible to defects in the data. More accurate data can be derived by tracking individual patients, but this costs money. Is the extra accuracy worth the expense?

[Zhou Wang got data from more than 1000 hospital cases to estimate survival functions by sex and age, from life data. Shigui Ruan's Lancet article reported

corona virus baseball statistics from China life data to The WHO.

Slides don’t mention transient Markov SEIR approximation that uses actuarial death and recovery rates as transition probabilities. It’s for forecasting and exploring effects of control measures. Reference by Yaesoubi and Cohen and my PhD thesis associate differential equations and Markov approximations for queuing systems analyses. Oli, Brown, and Venkataraman did too.]

Larry has used this technique working with the government of the Democratic Republic of the Congo during the Ebola outbreak, and also during the SARS and MERS outbreaks.

Summary

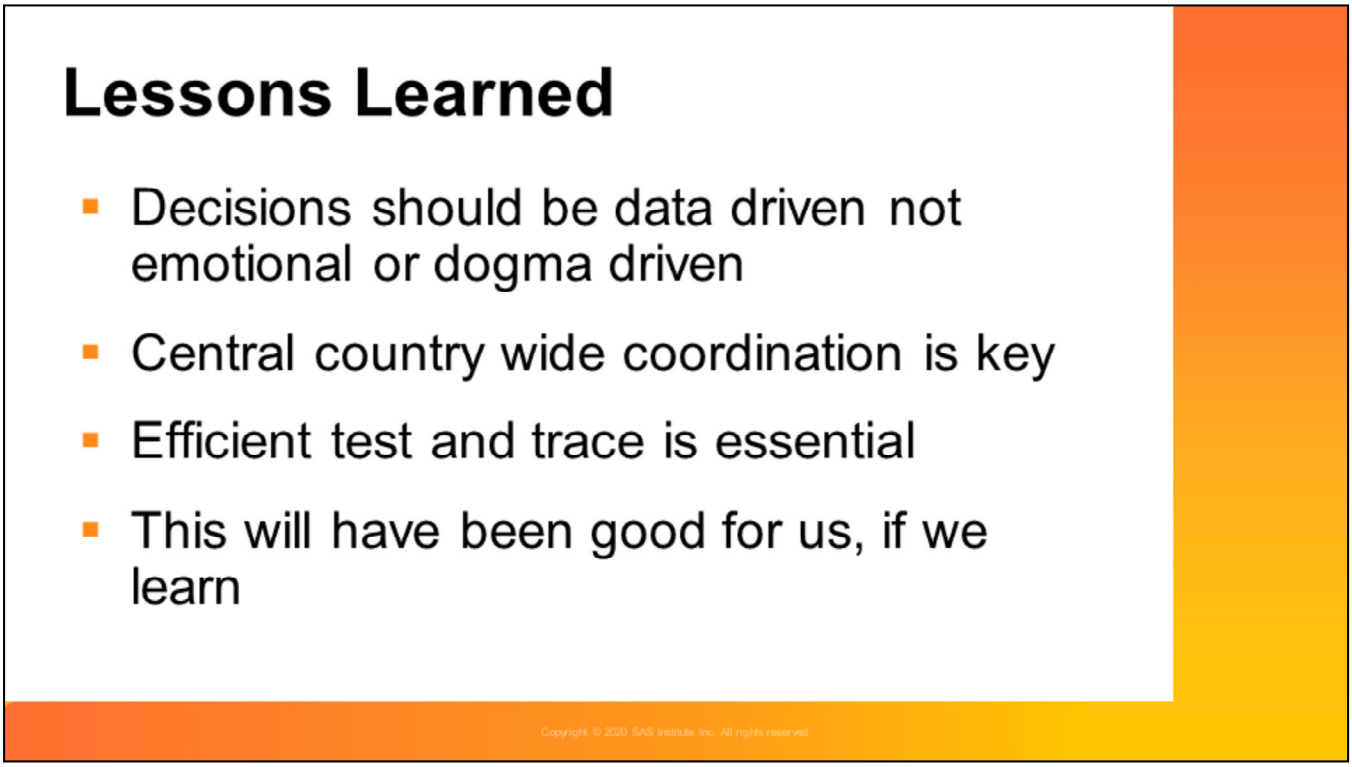
- Simpler is often better
- Understanding the system mechanics is key to allowing appropriate simplicity
- Understanding model limitations is key to seeing when predictions will be inaccurate
- Some areas are not simple, but should still be "as simple as possible"

Using the tools within JMP without considering the mechanics of the system can lead to much more complex and less satisfactory models

Understanding the mechanics allows simpler modeling, avoiding trying to model imperfections in the data.

Understanding the limitations of the simple model is key to seeing when it will give inaccurate results, and being able to adjust for those inaccuracies

Sometimes "as simple as possible" is far from being simple, but should still be "as simple as possible"



I hope that what we are experiencing with the pandemic is providing some learnings for us as a country.

- Experts should not be mistrusted because they are experts
- Personal freedoms are well worth fighting for, but should be tempered by their potential negative effect on society
    - The societal benefits of mask wearing far outweigh the loss of personal freedom
- Emotional decisions often lead to bad outcomes. Emotions should be listened to, but should be checked against real world objective data
    - Emotional decisions on reopening have significantly increased deaths, and the length of time that our economy is compromised

I believe that in the long run this situation will be good for us. It has shown the significant decision making and leadership weaknesses of the current US administration. Looking at the raw data, the outbreak has been handled extremely badly in the US.

Another good thing coming from the outbreak. There are much worse things coming down the pipe in the next 20 years.

- Significant food insecurity

- Mass migration
- Likely world conflict
- Significant social upheaval

This has been a good dry run for us. We must make sure that we learn from our mistakes to be able to better handle what is to come.

Appendix: Derivation of Transmission

C_n = number contagious on day n

PGc_n = Percentage Growth in Contagious on day n

$$C_n = C_{n-1} \times (PGc_n + 1)$$

NGi_n = Numeric Growth in Infections on day n

I_n = Infections on day n

DC = Days Contagious

$$C_n = \sum_{i=n-DC+1}^{n} NGi_i$$

$$NGi_n = I_n - I_{n-1}$$

$$C_n = I_n - I_{n-DC}$$

T_n = Transmission on day n

$$T_n = \frac{NGi_n}{C_{n-1}}$$

Make transmission (T) and Percentage growth in contagious (PGc) constants (so no suffix)

$$Newly\ Contagious = NGi_n = C_{n-1} \times T$$

$$Expiring\ Contagious = NGi_{n-DC} = C_{n-DC-1} \times T$$

$$C_n = C_{n-1} + C_{n-1} \times T - C_{n-DC-1} \times T$$

$$C_{n-1} \times (PGc + 1) = C_{n-1} + C_{n-1} \times T - \frac{C_{n-1}}{(PGc+1)^{DC}} \times T$$

$$(PGc + 1) = 1 + T - \frac{1}{(PGc+1)^{DC}} \times T$$

$$PGc = T \times \left(1 - \frac{1}{(PGc+1)^{DC}}\right)$$

$$PGc = T \times \left(\frac{(PGc+1)^{DC}}{(PGc+1)^{DC}} - \frac{1}{(PGc+1)^{DC}}\right)$$

$$PGc = T \times \left(\frac{(PGc+1)^{DC} - 1}{(PGc+1)^{DC}}\right)$$

$$T = PGc \times \left(\frac{(PGc+1)^{DC}}{(PGc+1)^{DC} - 1}\right)$$

Appendix: Converting from Reproduction to Contagious % Growth (1)

- Create a data table that derives transmission and reproduction from a given percentage growth in contagious

	PGc	Transmission	Reporduction
6	-5.00%	0.0852262885	0.7670365967
7	-4.00%	0.0900956836	0.8108611525
8	-3.00%	0.095120459	0.8560841312
9	-2.00%	0.1002991325	0.9026921929
10	-1.00%	0.1056299922	0.9506699299
11	0.00%	0.1111111111	1
12	1.00%	0.1167403628	1.0506632656
13	2.00%	0.1225154374	1.1026389363
14	3.00%	0.128433857	1.1559047132
15	4.00%	0.1344929927	1.2104369343
16	5.00%	0.14069008	1.2662107198

Transmission Formula

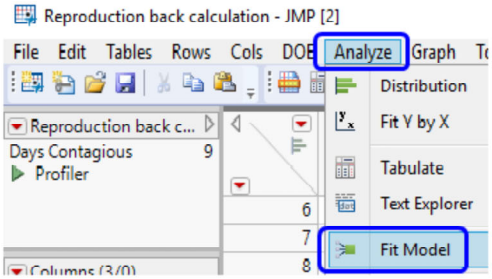
$$\frac{\left(PGc \cdot \left(PGc + 1\right)^{Days\ Contagious}\right)}{\left(\left(PGc + 1\right)^{Days\ Contagious} - 1\right)}$$

Reproduction Formula

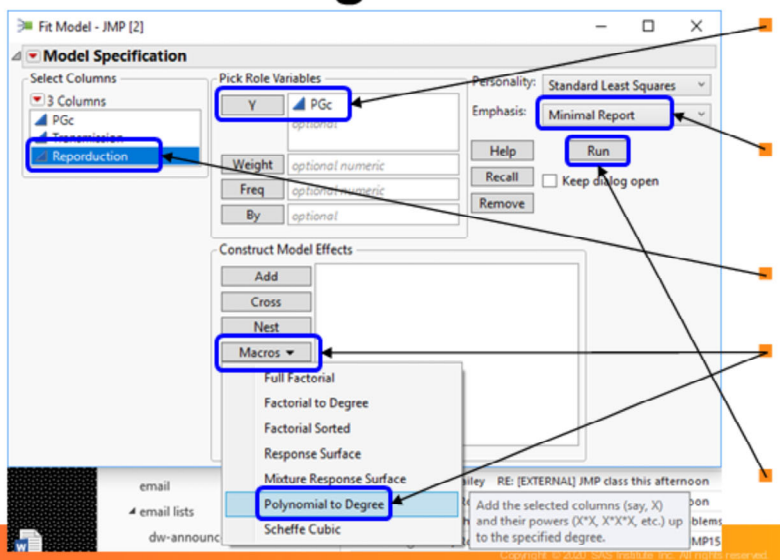
$$Transmission \cdot Days\ Contagious$$

Appendix: Converting from Reproduction to Contagious % Growth (2)

- Bring up the Fit Model tool



Appendix: Converting from Reproduction to Contagious % Growth (3)



Set Y to Percentage Growth in Contagious

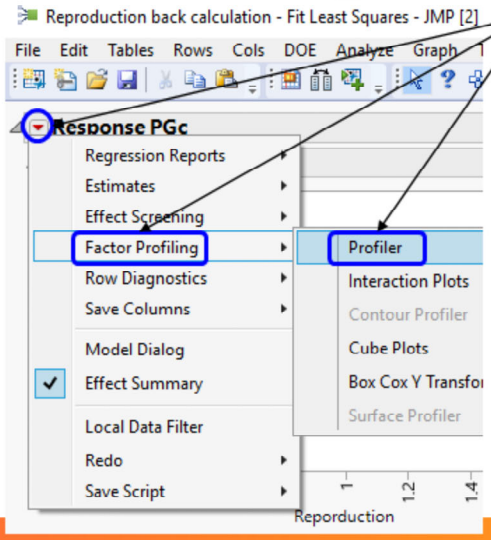
Set Emphasis to "Minimal Report"

Select "Reproduction"

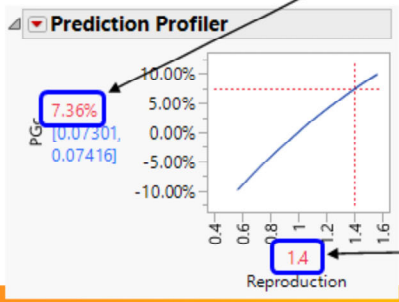
Set Macros to "Polynomial to Degree"

Hit Run

Appendix: Converting from Reproduction to Contagious % Growth (4)

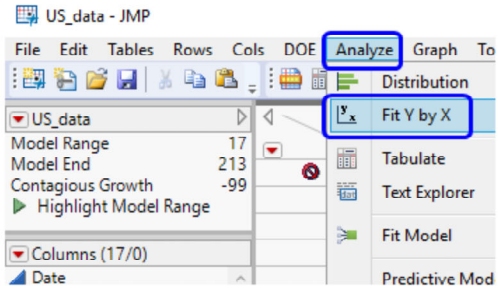


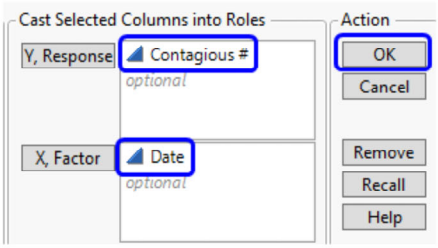
- Turn on the profiler
- Enter the Reproduction number, the profiler will return the percentage growth in contagious



Appendix: Fitting an Exponential using Fit Y by X (1)

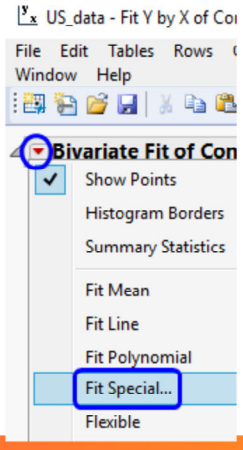
- Bring up the Fit Y by X tool
- Enter date as X and chosen Y parameter and click OK



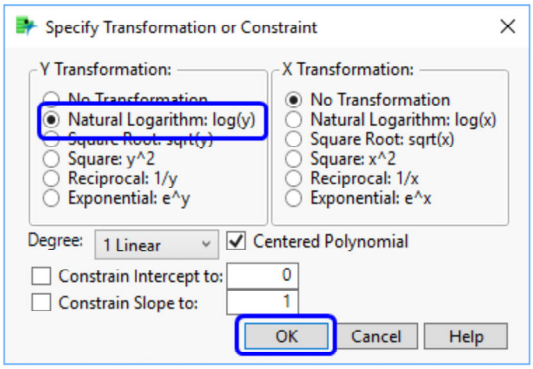


Appendix: Fitting an Exponential using Fit Y by X (2)

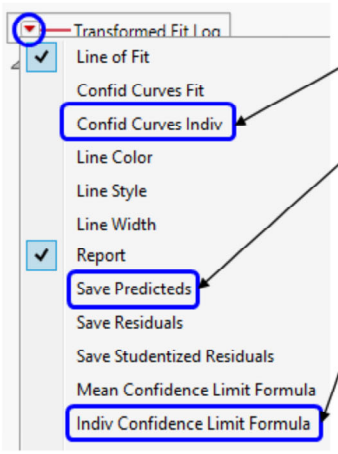
- Select Fit Special



- Select Y Transformation "Natural Logarithm: log(y)



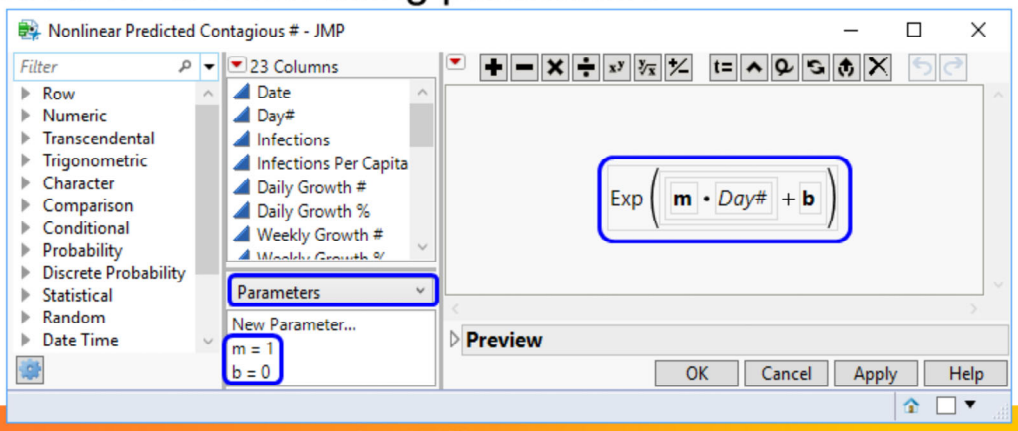
Appendix: Fitting an Exponential using Fit Y by X (3)



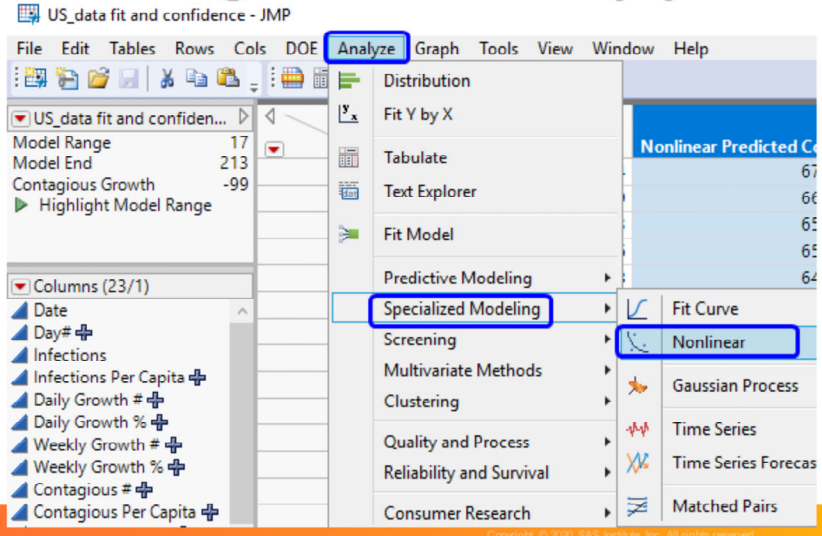
- Add confidence curves
- Save prediction line to data table
- Save confidence lines to data table if desired

Appendix: Fitting an Exponential using Nonlinear (1)

- Create a new column with a formula consisting of the desired fit function using parameters

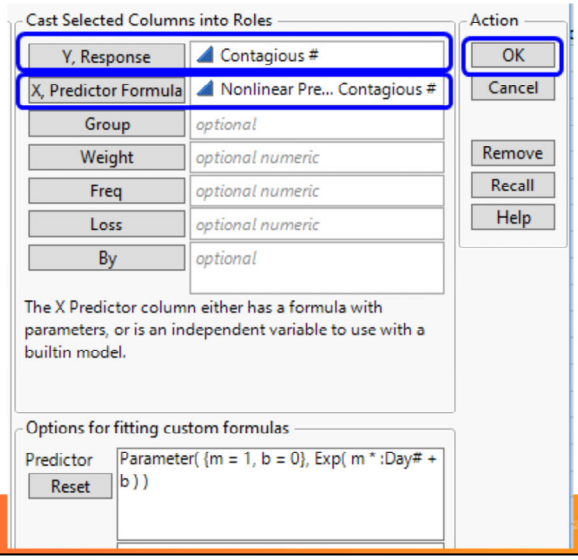


Appendix: Fitting an Exponential using Nonlinear (2)



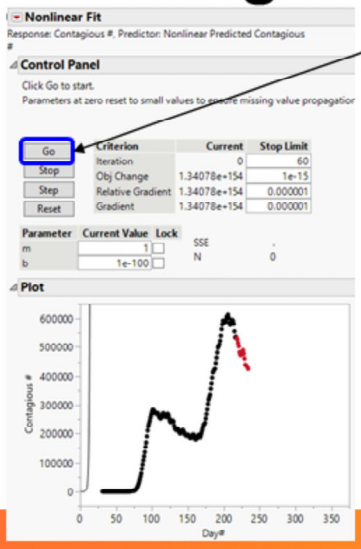
- Bring up the Nonlinear tool

Appendix: Fitting an Exponential using Nonlinear (3)

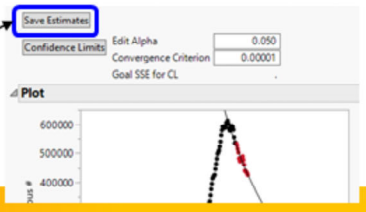


- Set the X, Predictor Formula to the new column you just created
- Set the Y, Response to the data that you want to fit
- Click OK

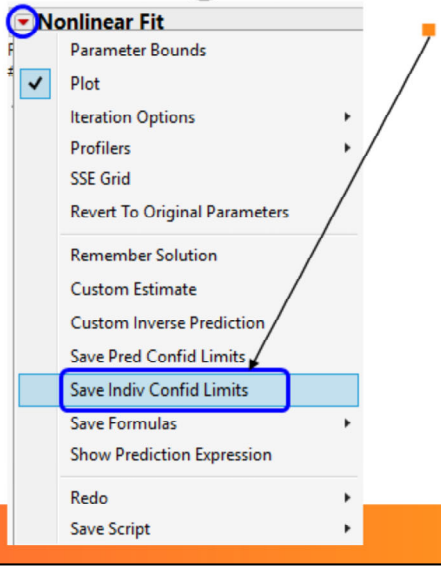
Appendix: Fitting an Exponential using Nonlinear (4)



- Click on "Go" to have tool adjust the parameters to match the data
- If the initial estimates were poor, you may have to click several times.
- Keep clicking until the "Save Estimates" button shows, which you should click to save the parameter values to the data table

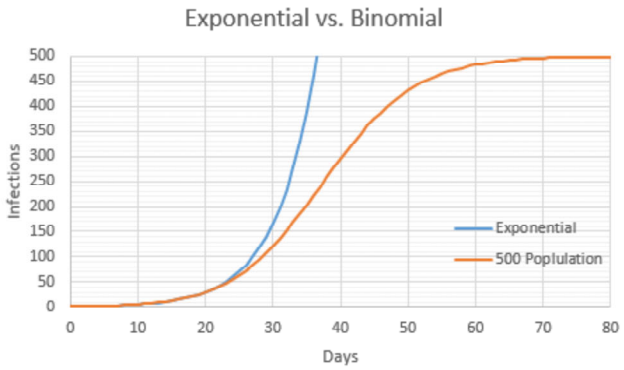


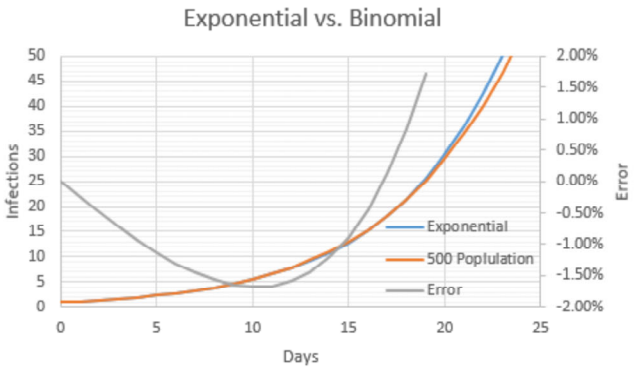
Appendix: Fitting an Exponential using Nonlinear (5)



Click on Save Indiv Confid Limits to save the confidence lines to the data table if desired

Appendix: Exponential Approximation of Binomial





Appendix: References

- Johns Hopkins University COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE): https://coronavirus.jhu.edu/map.html
- Johns Hopkins University COVID-19 Data Repository: https://github.com/CSSEGISandData/COVID-19
- Harvard Global Health Institute: Pandemics Explained: https://globalepidemics.org/key-metrics-for-covid-suppression/
- George, L. L., please see https://sites.google.com/site/fieldreliability/corona-virus-survival-analysis/
- Han, Henry, (2020) "Estimate the incubation period of coronavirus 2019 (COVID-19)," 10.1101/2020.02.24.20027474
- Neumann, Michael H., "Absolute regularity and ergodicity of Poisson count processes," *Bernoulli* 17(4), 2011, 1268-1284, DOI: 10.3150/10-BEJ313
- Oli, Madan, Mary B. Brown, and Meenakshi Venkataraman, "Population Dynamics of Infectious Diseases: a Discrete Time Model," Ecological Modeling, Sept. 2006, DOI: 10.1016/j.ecolmodel.2006.04.007

Appendix: References (cont)

- Oscarsson, Patric and Örjan Hallberg, "ERIVIEW 2000 – A Tool for the Analysis of Field Statistics," Ericsson Telecom AB, (2000)
- Ren, Zhimei, https://zren.shinyapps.io/covid19/, March 15, 2020
- Ruan, Shigui, "Likelihood of Survival of coronavirus disease 2019," https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30257-7/fulltext, March 30, 2020, DOI:https://doi.org/10.1016/S1473-3099(20)30257-7
- Schuhegger, Christian , https://weisser-zwerg.dev/posts/covid-19-data-analysis/, March 11, 2020
- Wang, Zhou et al., "Survival analysis of hospital length of stay of novel coronavirus (COVID-19) pneumonia patients in Sichuan, China," https://doi.org/10.1101/2020.04.07.20057299
- Reza Yaesoubi and Ted Cohen, "Generalized Markov Models of Infectious Disease Spread: A Novel Framework for Developing Dynamic Health Policies," Eur J Oper Res. Dec. 2011; 215(3): 679–687, doi:10.1016/j.ejor.2011.07.016