SO5032 Lecture 7

Brendan Halpin

April 1, 2025

Outline

SO5032 Lecture 7

SO5032 Lecture 7

Logarithms

Logarithms

Logarithms allow us to move between multiplicative equations and additive ones.

Logs are defined relative to a base number. If we take 10 as the base then $y = log_{10}(x)$ means $10^x = y$.

It's easy to calculate the log of powers of 10:

log(10) = 1	$10^1 = 10$
log(100) = 2	$10^2 = 100$
log(1000) = 3	$10^3 = 1000$
log(1000000) = 6	$10^6 = 1000000$

10^0 is defined as 1, so the log of 1 is zero.

From 0 to 1

For numbers between 1 and 0, logs are negative

$$\frac{1}{10} = 10^{-1} \qquad \log(0.1) = -1$$
$$\frac{1}{100} = 10^{-2} \qquad \log(0.01) = -2$$
$$\frac{1}{1000} = 10^{-3} \qquad \log(0.001) = -3$$

The $\log_{10}$ of powers of 10 are integers, but we can raise 10 to non-integer powers too, to get the log of any number greater than zero. For instance, $10^{2.09}$ is 123, so the log of 123 is 2.09.

Multiply by adding

We can see with round powers of 10 than using logs we can move between multiplication and addition:

$100 \times 1000 = 100000$

$10^2 \times 10^3 = 10^5 = 10^{2+3}$

Calculate A × B

Thus do calculate A × B we do as follows:

- Calculate log(A)
- Calclate log(B)
- Calculate log(C) = log(A) + log(B)
- Take the anti-log of log(C), i.e., $10^{\log(C)} = C$

sociology UNIVERSITY OF LIMERICK

Example

Multiply 12345 by 67890

$\log(12345) = 9.421$

$\log(67890) = 11.126$

$9.421 + 11.126 = 20.547$

$10^{20.547} = 838102050$

sociology
UNIVERSITY OF LIMERICK

An application

If you have a certain quantity (e.g., money in a bank account), whose value increases by a constant proportion every year, its value in any year depends on a multiplicative relationship.

Let's say the increases is α (i.e., a 10% increase means $\alpha = 1.1$)

Compound interest

Year 0	100
Year 1	$100 \times \alpha$
Year 2	$100 \times \alpha \times \alpha$
Year 3	$100 \times \alpha \times \alpha \times \alpha$
Year 4	$100 \times \alpha \times \alpha \times \alpha \times \alpha$
Year 5	$100 \times \alpha \times \alpha \times \alpha \times \alpha \times \alpha$

In short, the value in year t is $100 \times \alpha^{t}$

$$y_t = 100 \times \alpha^t$$

sociology
UNIVERSITY OF LIMERICK

Constant proportional increase

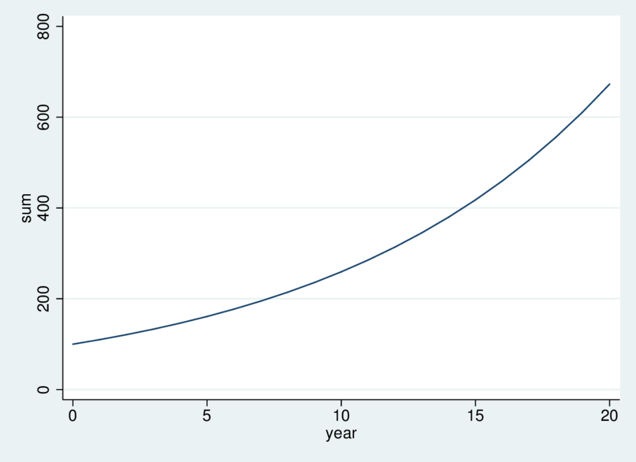


Figure 1: A constant proportional increase

Convert to logs

But if we convert to logs we can calculate it as follows

$$log(y_t) = log(100) + t \times log(\alpha)$$

In other words, rather than multiplying by α every year, we add log(α).

sociology
UNIVERSITY OF LIMERICK

Plot

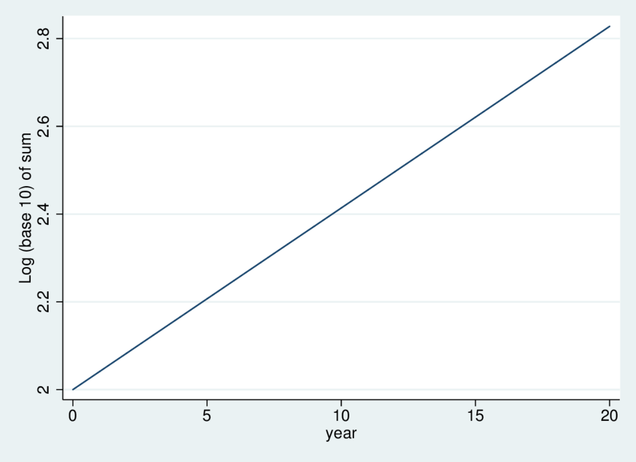


Figure 2: Taking the base-10 log of the sum: a straight line

Straight line

This gives a straight line relationship (see Fig 2).

Thus we can use logs to move between multiplicative and additive (straight-line) relationships.

Other bases

Logs to the base 10 are easy to understand, but the base number need not be 10.
A log to the base n is defined thus:

$$y = log_n(x) \Leftrightarrow n^y = x$$

Natural logs

Computer scientists often use $\log_2$, but the most common log base is the special number $e \approx 2.7183$. This has some special mathematical properties that make certain calculations easier.

Logs to base e are called natural logs, often written ln(x) etc:

$$y = ln(x) \Leftrightarrow e^y = x$$

See Fig 3, which shows that the natural log also gives a straight line.

Natural log straight line

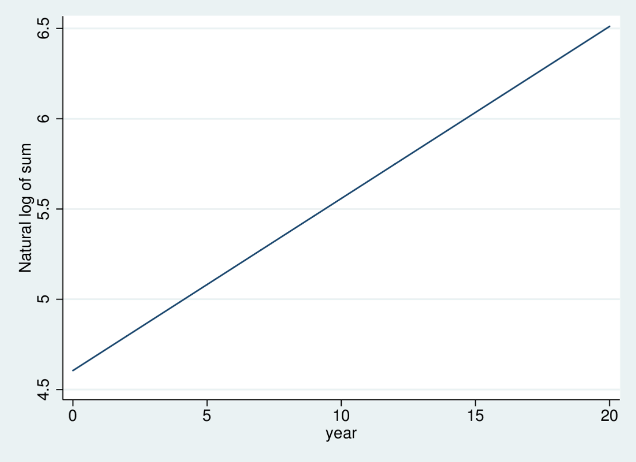


Figure 3: Taking the natural log of the sum: also a straight line

Natural log

- Fig 4 shows the natural log of X from 0.1 (-2.303) to 100 (4.605).
- For X = 1, the log is 0.
- As X approaches 0, the log falls faster and faster.
- As X rises above 1, the log rises, but more slowly as it goes.
- Note that the log rises from X = 5 to 10 as much as it does from X = 40 to 80.

X vs ln(X)

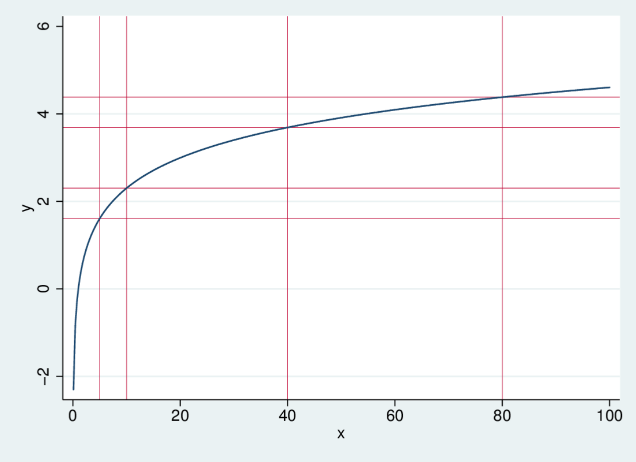


Figure 4: The natural log of X for X from 0.1 to 100

SO5032 Lecture 7

Early pandemic: exponential curves

Logs and COVID-19

- In the early stage of an epidemic, infections tend to increase at a steady rate
- On average each infected person infects others at a given rate, e.g., one person every four days
- So numbers of cases tend to rise at a steady percentage
    - New infections are proportional to existing infections
    - 100 today means 125 tomorrow, 156 the next day, etc.

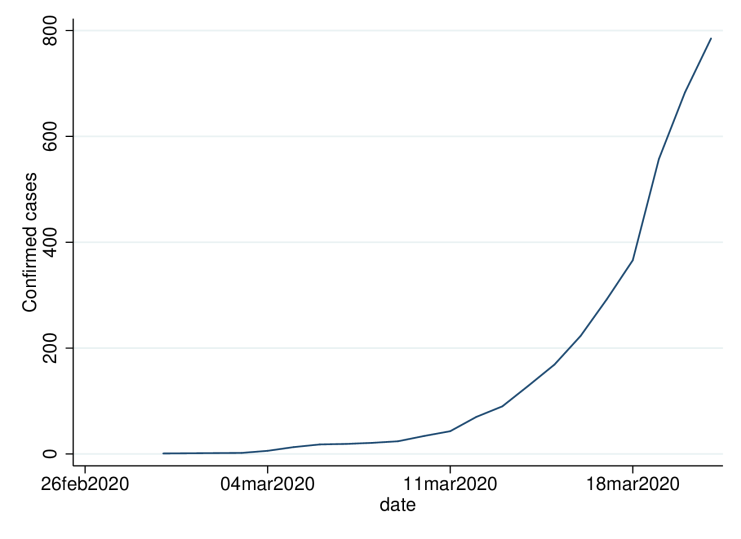
Confirmed cases in Ireland

If we look at the raw number of cases in Ireland:

- it starts off very low
- stays there for a while
- but then starts rising
- and rising faster and faster

```
line cases date
```

Confirmed cases in Ireland



sociology
UNIVERSITY OF LIMERICK

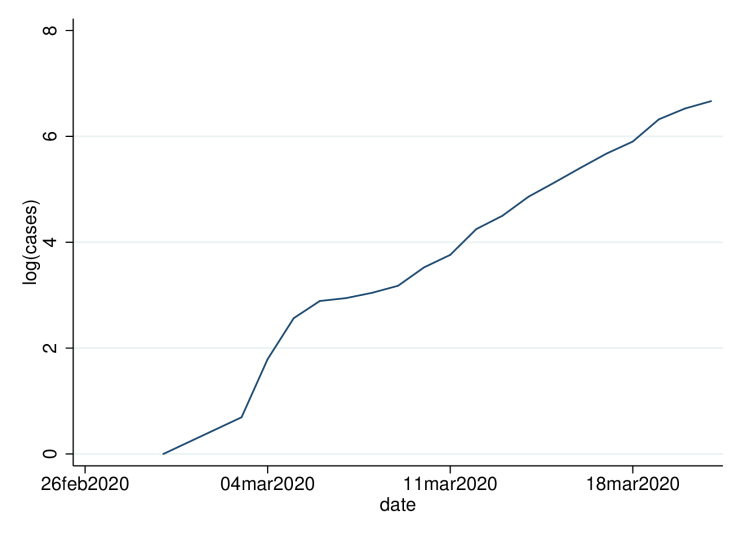
Log cases

If we plot the log of the cases we see a different picture

- wobbly to begin with
- then approximating a straight line

```
gen lcases = log(cases)
line lcases date
```

Log cases



log(cases)
0
2
4
6
8
26feb2020
04mar2020
11mar2020
18mar2020
date

Log cases: straight => exponential

A straight line in logs means `log(ncases)` increases by more or less a set amount very day

That means `ncases` rises by a set proportion every day: exponential rise

Exponential: even if it starts small, if given long enough, will get very very big!

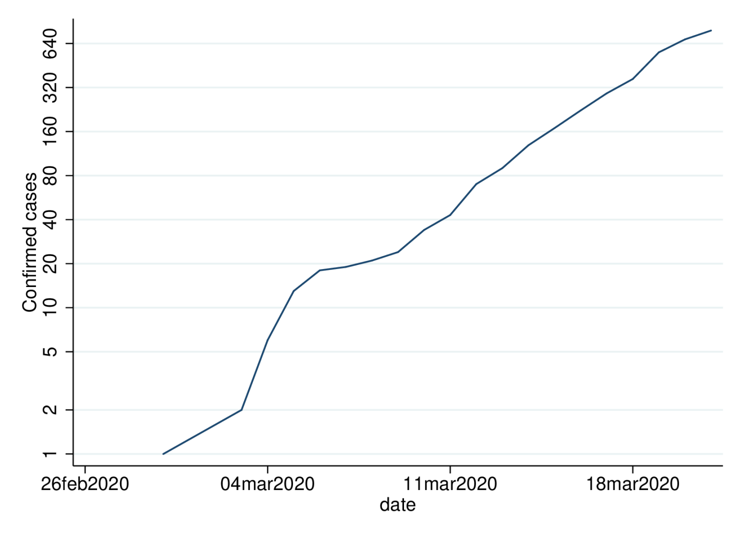
Log scale, real cases

We can graph `log(cases)` but we can also graph `cases` with a Y log-scale

```
line cases date, yscale(log) ylabel(1 2 5 10 20 40 80 160 320 640)
```

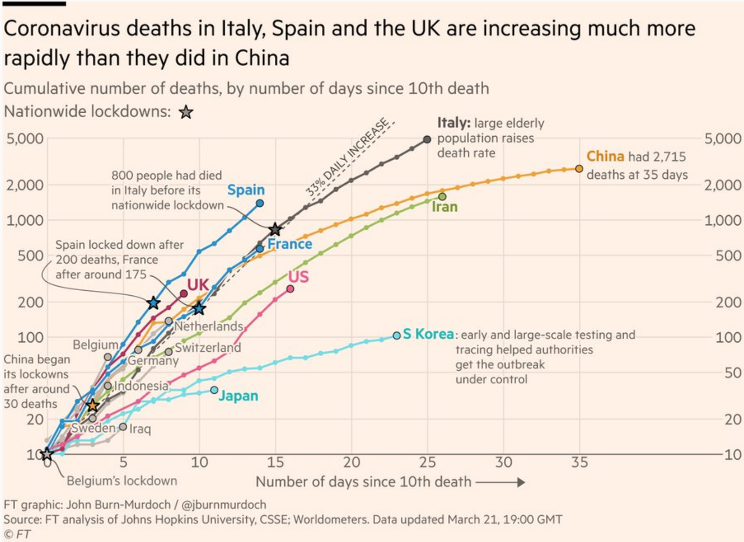
This gives the advantages of the logging while retaining the real numbers on the axis

Log scale, real cases



sociology
UNIVERSITY OF LIMERICK

Log-scale graphic in the wild



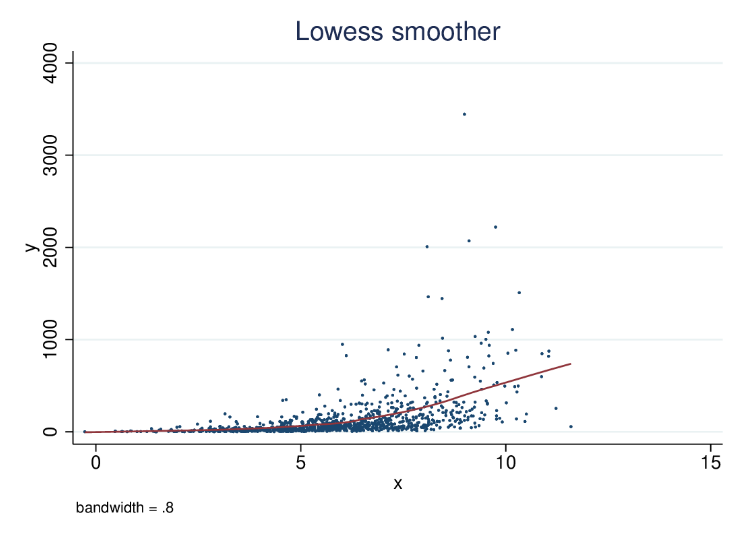
SO5032 Lecture 7

Log regression

Multiplicative relationship

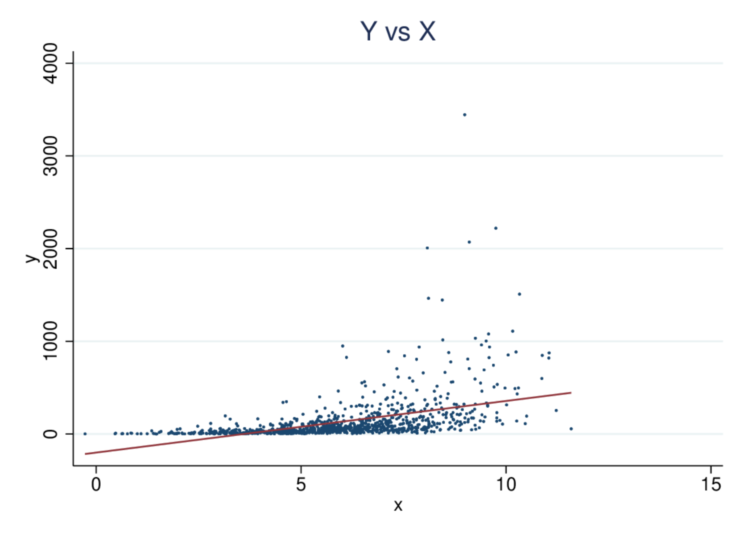
- Where the underlying relationship is multiplicative, linear regression doesn't work well
- Implies an additive increase where a multiplicative one is better
- If we take the log of the dependent variable:
  - better estimates
  - often cures heteroscedasticity

Simulation: Y increases 65% for X +1



Linear regression

Predictions



Log(Y)

sociology
UNIVERSITY OF LIMERICK

Interpretation

- For a 1 unit change in X, $log(\hat{Y})$ rises by 0.4933914
- Thus for a 1 unit change in X, Y rises by $e^{0.4933914} = 1.638$
- $e^{0.4933914}$ is the antilog of 0.4933914

Predictions

Predicted values

- Where the dependent variable is logged the prediction of the Y value is not simply the anti-log of the predicted log(Y)
- When we take the anti-log we must take account of the fact that residuals above the line expand by more than residuals below the line
- Thus a small correction

$$\hat{log(Y)} = a + bX$$

$$\hat{Y} = e^{\hat{log(Y)}} * e^{\mathrm{RMSE}^2/2}$$

- where RMSE is the standard deviation of the regression

sociology
UNIVERSITY OF LIMERICK

Calculations

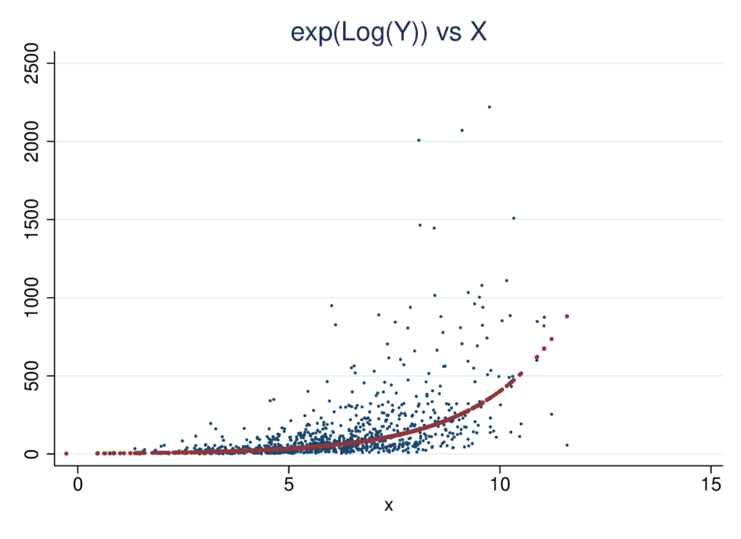
```
gen ly = log(y)
reg ly x

predict lyhat
gen elyh = exp(lyhat)
gen elyh2 = elyh * exp(rmse^2/2)
```

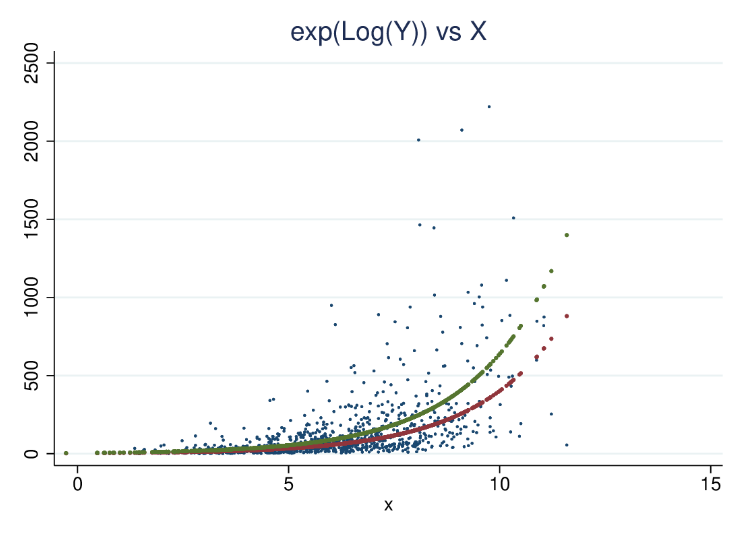
Predictions: predict log(Y) on log scale

sociology
UNIVERSITY OF LIMERICK

Predictions: only $e^{\hat{log(Y)}}$



Predictions: with correction



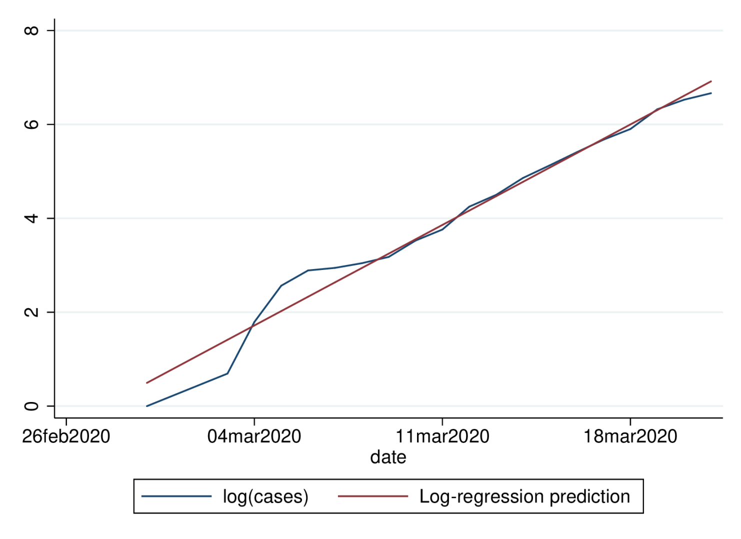
Predicting COVID-19

- We can apply log regression to the COVID-19 data
- A straight line on a log scale means a constant proportional increase.
- We can estimate this increase, regressing log(cases) on date.
- The slope, b, is the amount by which $\log \widehat{\text{cases}}$ rises per day
- e^b is then the multiplier by which `cases` rises per day

```
reg lcases date
```

Stata output

Logs with log regression



sociology
UNIVERSITY OF LIMERICK

Steady increase

The log of `cases` rises by 0.3058 per day

This means `cases` rises by a factor of $e^{0.3058} = 1.358$

The increase is 1.358 - 1 = 0.358, or almost 36% per day

Implies a doubling about every 2.6 days

But exponential increase is temporary

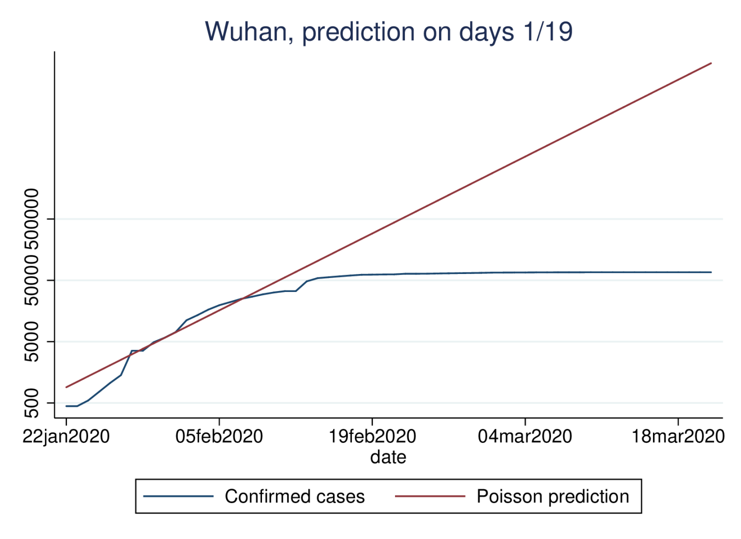
Exponential increase cannot go on indefinitely

Even if nothing is done, the rate of increase will decline as fewer people are left unexposed

And interventions (isolation, tracing) will reduce the rate

See China, for example

Wuhan, with prediction based on 1st 19 days



Summary

If there is a constant rate of increase, logs give us straight lines

Graph the log, or use a log scale on the Y-axis

Log regression allows us to estimate the rate

Exponential increase isn't forever, but modelling the exponential helps us see where the rate starts to drop

Code available here: http://teaching.sociology.ul.ie/so5032/irecovid.do