Introduction

P and N functions. The cumulative total number of Covid-19 cases confirmed up to and including day t is denoted by $P(t)$. We allow $t = 0$ to correspond to any given date.

The number of new cases confirmed on N on day t is given by

$$N(t) = P(t) - P(t-1).$$

We sometimes also call $N(t)$ the...

- daily increase in cases on day t.
- daily cases on day t.
- new cases.

Note that with these definitions

$$\begin{aligned} P(t) &= P(t-1) + N(t) \\ &= (\text{Total cases confirmed up to and including day } t) + (\text{New cases confirmed on day } t) \end{aligned}$$

Data sources. All data is sourced from the *COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University* accessible at the following link

- https://github.com/CSSEGISandData/COVID-19.

We reference this source as JHU in short.

Getting the data for your own explorations, or for students to explore. Raw data for countries and US states, together with their 7-day smoothing (see below) can be downloaded using a couple of websites that Isaac Flath built for the Consortium.

- https://isaac-flath.shinyapps.io/coronavirus2/
- https://isaac-flath.shinyapps.io/Covid-19/

These two sites also generate plots for any given time range for all the quantities we discuss in the problems we have written (P, N, relative change of P, log plots of P, etc).

Data smoothing. Covid-19 data (specially N-data) is very noisy, so in all the problems we have written we use the 7-day moving average of the data which we compute as follows ($\widetilde{P}(t)$ is the raw time series obtained JHU):

$$\begin{aligned} P(t) &= (\widetilde{P}(t) + \widetilde{P}(t-1) + \ldots + \widetilde{P}(t-6))/7 \\ N(t) &= P(t+1) - P(t) \end{aligned}$$

so, the new cases time series $N(t)$ gets smoothed through the smoothing of $P(t)$ and does not get smoothed separately. This makes the relation $N(t) = P(t+1) - P(t)$ continue to hold after the smoothing. We also make sure that we pick up JHU data well before $t = -7$, so the data is smoothed using the above formula even for $t = 0$ using the previous 7 data points in the JHU time series.

Figures 1 and 2 show the effect of this smoothing for the cumulative cases P and new cases N for Romania, with t in days since March 1, 2020 (thick blue is the smoothed data, thin black is the raw data). A side product of this smoothing is that data gets shifted forward (to the left) by 3.5 days.

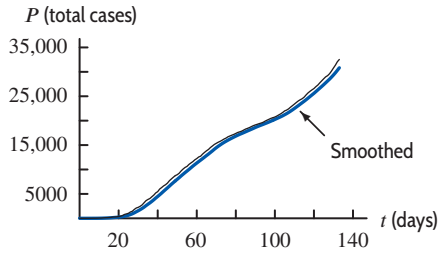

Figure 1

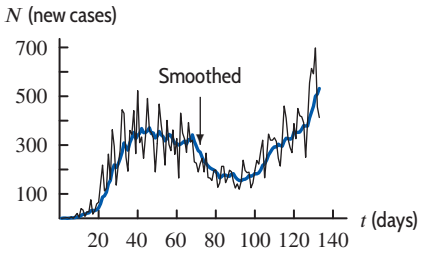

Figure 2

We will not use raw data $\widetilde{P}(t)$ (or $\widetilde{N}(t) = \widetilde{P}(t+1) - \widetilde{P}(t)$) in what follows unless explicitly stated otherwise.

Expected shape of P and N. The Covid-19 outbreaks have occurred in waves in many countries. Each wave has an N-plot which is bell-shaped and that will correspond to a sigmoid shaped P-curve.

Subsequent waves stack on top of one another and result in N-plots that look like a sum of bell-shaped curves (which may or may not drop all the way down to zero between the bells), and this results in a P-plot that will look like a sequence of stacked sigmoid s-curves. Figures 3 and 4 are the P and N plots for Australia between March 1 and August 16, 2020.

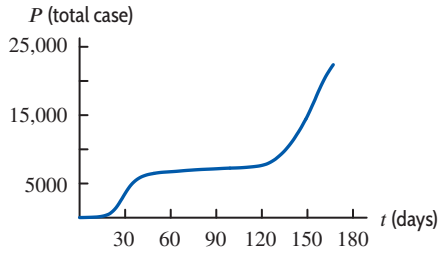

Figure 3

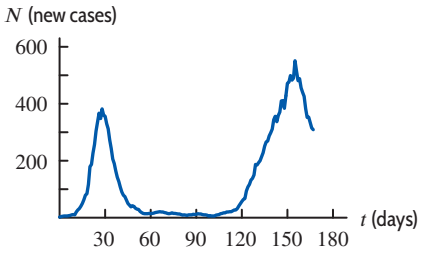

Figure 4

Of course, the outbreaks in many countries do not look so distinctly as a sum of waves. Besides possible data collecting issues or changes in data collecting strategies, drastic measures implemented to control the outbreak have noticeable effects on the way the outbreak developed. Figures 1 and 2 are good examples of expected shapes P and N plots when the waves are not separate.

What students need to have seen to do these problems.

- Definitions of P and N

Many more problems and problem ideas in the Instructor notes package. Visit `https://mcwg.github.io/covid` for many more problems and problem ideas.

Sample problems - Quantitative Reasoning - see the Instructor notes package for more

1. Figure 5 and 6 show data from the Covid-19 outbreak in Switzerland[1] during the Spring of 2020. Which graph is the graph of the function $P(t)$, total number of cases confirmed up to and including day t, and which is the graph of the function $N(t)$, number of new cases confirmed on day t?

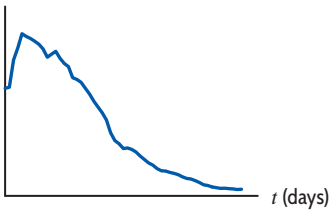

Figure 5

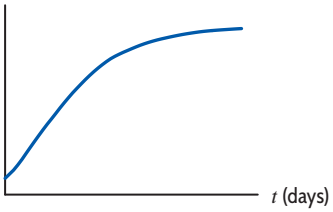

Figure 6

2. Let $P(t)$ be the total number of Covid-19 cases in Minnesota confirmed up to and including day t, where $t = 0$ is March 5, 2020.[2] Let $N(t)$ be the number of new cases confirmed on day t. Explain the meaning of the statements in the context of the outbreak.

(a) $P(37) - P(36) = 80$
(b) $(P(37) - P(36))/P(36) = 0.074 = 7.4\%$
(c) $(P(37) - P(36))/P(37) = 0.034 = 3.4\%$
(d) $P(35) + N(36) + N(37) = 1164$

3. A database shows that 21,417 people had been confirmed by a test to have Covid-19 on or before April 3, 2020. Suggest reasons that the true number of people who have caught the disease by April 3 is probably not exactly 21,417. For each of your reasons, say whether it implies that the true number of cases is higher or lower than 21,417.

4. When comparing between regions the severity of Covid-19 outbreaks, it does not make sense to simply compare the number of confirmed cases: a region with a very large population, like the US, will automatically have more cases than a region with a very small population, like San Marino. It makes more sense to compare the ratio of confirmed cases to total population. In this problem, we explore the various ways to express this ratio and its meaning in some examples.

(a) By September 15, 2020, 2.083% of the population of Bangladesh had been confirmed to have acquired the Covid-19 virus. The percentage 2.083% means there were 2.083 confirmed cases for every 100 people, or 2.083 confirmed cases per 100 people.

(i) How many cases were there per one thousand people? per 10,000? per 100,000? per million?

(ii) Why might giving the cases as a number per 100,000 people instead of a percentage be useful?

(iii) In general, how would you compute the number of cases per million from the total number of confirmed cases? How about cases per 1000? How are they related to the percentage?

(b) By September 15, 2020, Oregon had 8 cases per thousand people. Does this mean that in every group of 1000 people there are 8 people who may have, or have had, the disease?

(c) Table 1 contains the number of confirmed cases[3] by September 15, 2020 for several countries and US states, together with their populations[4].

(i) Compute the percentage of the population that has been confirmed to have acquired the disease by this date, the number of cases per thousand people, and the number of cases per million people. Round your percentages to three decimal places, and your other calculations to the nearest integer.

(ii) According to your calculations, in which region has the outbreak been most severe when considering cases as a percentage? per thousand people? per million people?

(iii) Your computed numbers of cases per thousand in Wyoming and Belgium are the same. Is the severity of the outbreak the same when

[1] Data from Johns Hopkins University, downloaded from https://github.com/CSSEGISandData/COVID-19 accessed on May 23, 2020.

[2] Data from Johns Hopkins University, downloaded from https://github.com/CSSEGISandData/COVID-19 accessed on January 17, 2021.

[3] Data from Johns Hopkins University, downloaded from https://github.com/CSSEGISandData/COVID-19 accessed on September 15, 2020.

[4] Data from https://data.worldbank.org/indicator/SP.POP.TOTL through https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_Unite accessed on June 15, 2020, and https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html accessed on September 22, 2020.

you look at cases per million? Why, or why not?

(iv) Panama and Belgium had roughly the same number of confirmed cases. What can you tell from the cases per thousand people?

(v) Ukraine and Ohio also had roughly the same number of confirmed cases. What can you tell in this case?

(vi) The scale of the outbreak in the US and Panama when comparing total number of cases is very different. Is this the case when you look at cases per million people?

(vii) California has about the same population as one of the countries in the table. How does the total number of cases compare between California and this country? How is this reflected in the cases per million?

Table 1

Country or US state	Cumulative cases by 15 September, 2020	Population estimate
Panama	100,987	4,176,873
US	6,478,876	326,687,501
Belgium	92,404	11,433,256
Ukraine	155,152	44,622,516
Florida	660,787	21,477,737
California	758,165	39,512,223
Wyoming	4299	563,626
Ohio	136,331	11,689,100

5. By October 9, 2020, there had been approximately 8 confirmed Covid-19 cases per thousand people in Portugal. The population of Portugal in 2020 was approximately 10 million.

(a) Approximately how many confirmed Covid-19 cases were there in Portugal on October 9, 2020?

(b) Was it possible to find a group of 1000 people in Portugal on this date all of whom had, or had had Covid-19? Was it possible to find a group of 1000 people none of whom had had Covid-19? What is the meaning of "8 cases per thousand" for actual groups of 1000 people?

6. Figures 7 and 8 show the reported daily number of newly confirmed Covid-19 cases in South Africa and Qatar.[5]

(a) Which of the two figures shows a taller bell? In which country was the maximum number of daily new cases higher?

(b) Both plots show a clear "wave" of daily cases going up and then down during a time period of about 150 days. Based on the plots, estimate approximately how many times more cases there were during these waves in one country than in the other.

(c) To make a better comparison of the severity of the outbreak during those waves we should take into account the population of the two countries and think about cases per capita. The population of South Africa in 2020 was about 20 times larger than the population of Qatar. What does this tell you about which country had a more severe outbreak during the wave?

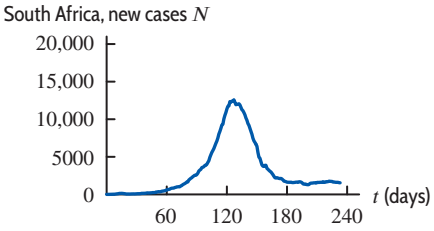

Figure 7

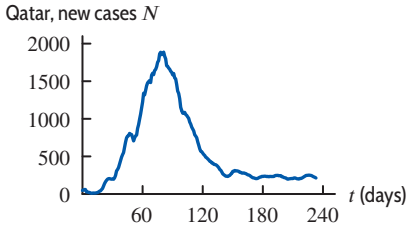

Figure 8

[5]Data from Johns Hopkins University, downloaded from https://github.com/CSSEGISandData/COVID-19 accessed on October 22, 2020. Data smoothed using a 7-day moving average.