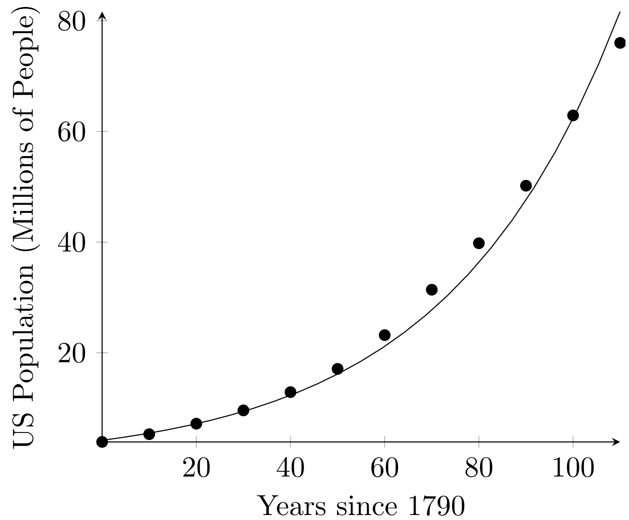
Logarithm Plots

Purpose In this lab you will learn how to determine if data points can be modeled by either a power function $y = ct^k$ or by an exponential function $y = ce^{kt}$ and you will learn how to choose the constants c and k.

Preview The eighteenth and nineteenth century United States census data in the table below can be well-represented by the function $P(t) = 4.19e^{0.027t}$, where t is the number of years since 1790. The plot shows the function $P(t)$ superimposed on a plot of the data from the table.

Year	Population (millions)
1790	3.9
1800	5.3
1810	7.2
1820	9.6
1830	12.9
1840	17.1
1850	23.2
1860	31.4
1870	39.8
1880	50.2
1890	62.9
1900	76.0



You will learn how to test such data to determine if an exponential or power function fits the data well and you will see how to determine the constants in the function. The function is called a 'model' for the data, and the constants are called the model's 'parameters'.

Overview We begin by creating some "data" points using an exponential function, but we will pretend our data were gathered from an experiment. This will allow us to discover the original functional relationship between the variables. Once we have the technique and concept in hand, we can apply this knowledge to real data.

Part I: Semilog Plots—Beginning of the Experiment

Open the spreadsheet for this lab, make a copy, and rename it with your group names.

1. We will use the function $y(t) = 3 \cdot 2^t$ to create some data points. In Column B of the first tab of the spreadsheet for this lab, compute values for this function at the values of t given in Column A.

2. Insert a chart for y vs. t into your spreadsheet.
3. In Column C, insert values for $\ln(y)$.
4. Plot $\ln(y)$ vs. t[1]. What do you observe about this new graph?

An Explanation

First, a reminder of properties of logs:

Properties of Logs

$$\ln(AB) = \ln A + \ln B \quad \ln\left(\frac{A}{B}\right) = \ln A - \ln B \quad \ln(A^p) = p\ln(A)$$

The type of graph that you just made is called a "semilog plot" because you used the logarithms of the second coordinates in your plot rather than the original y values. (The prefix "semi" refers to the fact that we did not take the logarithms of the first coordinate.)

In this case we created our own "data" from a known function so we could use that function to explain what happened in the second graph.

The equation that expresses the relationship between y and t is

$$y = 3 \cdot 2^t.$$

Take the (natural) logarithm of both sides of the equation to get the new equation

$$\ln y = \ln(3 \cdot 2^t).$$

Applying some properties of logarithms, we see that

$$\ln y = \ln 3 + t \cdot \ln 2.$$

The last equation tells us that $\ln y$ is a linear function of t; i.e., a plot of $(t, \ln y)$ must be linear!

Looking for Exponential Models

The explanation above gives us the information we need to test data points for an exponential fit. Suppose we have some data points (t, y) and suppose also that the semilog plot of these data points is linear. This linearity of the semilog plot implies that

$$\ln y = mt + b \text{ for some constants } m \text{ and } b.$$

Making both sides of the equation an exponent of e, we get the equation

$$e^{\ln y} = e^{mt+b}.$$

[1] To select non-contiguous columns, hold down the ctrl key (or the command key on a Mac).

Simplifying this equation gives us

$$y = e^b e^{mt} = ce^{mt}$$

where the two-symbol constant e^b is replaced by the single symbol constant c.

We have shown that if the semilog plot is linear, then $y = ce^{mt}$ for some constants c and m!

Part II: Semilog Plots—Completion of the Experiment

We pretend here that we do not know the precise relationship between t and y for the "data" points in Part I. We have only the numbers in your spreadsheet. We deduce from the semilog plot that a function of the form $y = Ce^{mt}$, for some constants m and C, should fit the data. We must now find the constants m and C.

5. Using the semilog plot, fill in the following blank. You can use a linear trendline added to your graph:

$$\ln(y) = \underline{\hspace{3cm}}.$$

6. Solve your equation for y and simplify it to to find an equation of the form $y = Ce^{mt}$.

7. Fill in the blanks. If we can fit a linear function to a semilog plot, $\ln(y) = mt + b$, then we can recover an exponential, $y = Ce^{mt}$, where:

   - m is the _____________ of the semilog plot;
   - If b is the _____________ of the semilog plot, then $C =$ _____________.

8. We have outlined the procedure for testing data to determine if it can be fitted with an exponential function and we have seen how to find such a function. If you've followed along with the computations, then you probably found a function similar to

$$y = 3e^{0.69t} = 3(1.9937)^t.$$

   But you know in this case that we used the function $y = 3 \cdot 2^t$ to create the data which we then used to "rediscover" the function. Explain why the function we fitted to our data is slightly different from the original one.

Part III: US Population Data

9. Refer to the US population data shown in the preview. It is reproduced in the second tab of your spreadsheet. Check to see that the semilog plot is approximately linear and construct your own exponential function to fit the data.

Part IV: Log-Log Plots—Power Function Fits

Another type of function that we often use to model data is the family of power functions $y = ct^k$. It is as easy to test for this type of fit as it was for the exponential fit. Indeed, we use the same method to discover what kind of plot to use: take logarithms of the equation for a power function, and use log properties to simplify the expression.

Assume that $y = ct^k$. By taking logarithms of both sides, it follows that

$$\begin{aligned} \ln y &= \ln(ct^k) \\ &= \ln(c) + \ln(t^k) \\ &= \ln(c) + k\ln(t). \end{aligned}$$

Because c and k are constants, this equation implies $\ln(y)$ is a linear function of $\ln(t)$. The implication here is that if $y = ct^k$, then a plot of the points $(\ln(t), \ln(y))$ will be a line and the slope of this line is the degree of the power function. Plots of the points $(\ln(t), \ln(y))$ are called "log-log plots."

We can summarize these results as follows (fill in the blank):

- If the log-log plot is linear, then $y = ct^k$ for some constants c and k.
- The ______________ of the log-log plot is the exponent k.

10. The following data is found in the third tab of your spreadsheet. By adding data for the natural log of columns A and B in columns C and D and plotting, show that the log-log plot of the data is linear. Find a function that fits the data well and superimpose the function over a scatter plot of the data. Make a semilog plot of the same data and compare that plot to the log-log plot. Explain why the semilog plot has the shape that it does.

t	.09	.24	.39	.54	.69	.84	.99
y	1.64	2.09	2.37	2.57	2.73	2.87	2.99

Part V: Concluding Cautions and Comments

In an earlier lab you learned how to find a linear function to fit data which appears to be approximately linear. In this lab you learned how to fit exponential and power functions to data. But you should note that the only type of function we can identify simply by looking at its graph is a linear function. Indeed, we found the exponential and power functions by recognizing a line on a semilog or log-log plot.

11. (a) Test the data in the following tables (found in the fourth tab of your spreadsheet) to determine if an exponential function could provide a good fit.

    (b) If an exponential function is appropriate, then find it and make a graph of the exponential function superimposed on a scatter plot of the data.

(c) If using an exponential function is not appropriate, then explain carefully what conclusions you can draw from the semilog plot. For example, if an exponential is not appropropirate, comment on whether the function increases (or decreases) faster (or slower) than an exponential.

(d) If using an exponential function is not appropriate, repeat the three questions above using a log-log plot to analyze whether a power functions fits.

t	1	2	3	4	5	6	7
y	3.65	3.18	2.77	2.40	2.09	1.82	1.58

t	1	2	3	4	5	6	7	8	9
y	.111	.125	.143	.167	.200	.250	.333	.500	1.00

In general, many functions are not linear, exponential, or power. The world of functions covers a far more general set of relationships than these three. Nonetheless, these relationships (especially linear and exponential ones) arise so often in real world data that it is useful to understand them specifically.

Project: NC COVID Hospitalization Data - A Complex Log Model

The Data

This project will deal with COVID-19 hospitalization data over a number of weeks in 2020. You will be given a seven day running average of the number of people hospitalized with COVID-19 in North Carolina. That is, each data point will be the average number of people hospitalized with COVID-19 over the previous seven days. You will create various plots of the data, and use ideas from the lab to gain insight into the growth of COVID-19 in NC. You will then research factors that may explain certain trends in the data, and try to gain insight into the decisions policymakers made as the disease progressed. The data can be found in the spreadsheet associated to this lab, in the NC COVID-19 Data tab.

Mathematical Preliminaries: Piecewise linear functions

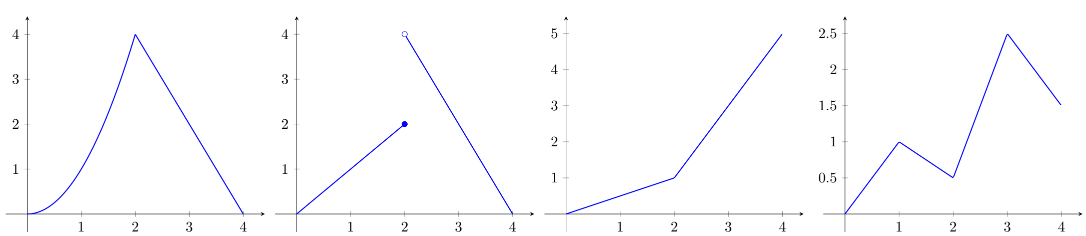
Prior to beginning work with the data, we will need to review *piecewise linear functions.* The answers to this section are not part of the report. However, you should complete this section prior to beginning analysis of the data.

A function is *piecewise linear* if it can be written as piecewise function, where each piece is linear. For example:

$$f(x) = \begin{cases} x+4 & \text{when } 0 < x < 4 \\ 2x+1 & \text{when } 4 \le x < 8 \end{cases} \qquad g(x) = \begin{cases} x+4 & \text{when } 0 < x < 4 \\ x^2 & \text{when } 4 \le x < 8 \end{cases}$$

Piecewise Linear — Not piecewise linear

1. Which of the following graphs represent piecewise linear functions on the domain $0 \le x \le 4$? For each function that is piecewise linear, compute a formula for it.



2. Any linear function can be considered piecewise linear (with only one piece). However, the reverse statement is not true: piecewise linear functions are not necessarily linear. Briefly explain why this is the case, drawing one or more examples to illustrate your argument. Your explanation should use the word ‘slope’.

Looking at the Data

In the spreadsheet provided, you will see a list of dates starting April 14 2020, and ending June 25 2020. The second column is the seven day average number of people hospitalized in North Carolina with COVID-19 [2]. As you may know, the early spread of a disease is often exponential.

3. If the spread of the disease over the given time period is indeed exponential, which type of plot do you expect to be linear?

4. Create the plot from your previous answer for the given data (be sure to title your plot and label the axes). Make the minimum of the y-axis 5.5. Was the growth in COVID-19 hospitalizations over this period exponential? Explain your answer.

Breaking Down the Data

5. Looking at your plot again, is it reasonable to say that it is is *piecewise* linear? Explain your answer. At a glance, how many pieces would you divide the plot into?

6. We will focus on the three periods of growth of hospitalization. Find the dates on which each of them starts and ends. You should decide what constitutues a beginning and end of a growth period. Be sure to write down your criteria as well as your results.

7. Create three new plots, one for each of the three growth periods of the disease. Fit each with a linear trendline and consider its equation to three significant figures (if you do this right, that turns out to be four decimal places).

8. A linear log plot corresponds an equation $\ln y = mx + c$. Noting that in an exponential function $y = Ce^{kx}$, the constant k is the *per capita growth rate* of the function (for example, $y = e^{0.02x}$ has a per capita growth rate of 0.02 per unit of time, or 2%), find the per capita growth rates of hospitalizations during each of the three periods of growth. Express your answers in percentages.

9. To get a handle on what the per capita growth rates mean, compute the *doubling time* (in days) of hospitalizations during each of the three growth periods.

Lagging Indicators

It is often said that hospitalizations are a *lagging indicator* of disease infection. Since most people do not exhibit symptoms of COVID-19 for a number of days after infection, and do not exhibit severe symptoms a few days subsequent to initial ones, they are highly unlikely to be hospitalized soon after infection. Therefore, the number of hospitalizations on a given day (or an average of them), does not indicate currently occuring infections on that day. The notion of lagging indicator originates from economics, but is very useful in epidemiology as well.

10. Do some research on the concept of lagging indicators (and their companion, *leading* indicators), both in economic and epidemidological terms, then answer the following questions:

    (a) Why might a lagging indicator not be useful in predicting current or future trends?

    (b) What might be a leading indicator of COVID-19 infection?

[2] Source: https://covid19.ncdhhs.gov/dashboard/about-data

(c) Explain why lagging indicators are relatively easy to measure, but best at evaluating past policy decisions, rather than directly indicating at future ones. Likewise, explain why leading indicators are difficult to measure, but useful in making policy decisions about the near future.

11. In the case of COVID-19, hospitalizations lag infections by around 10 days. Looking back at your division into periods of growth of hospitalizations in NC, find when they indicate periods of growth of *infection* began and ended.

Your Report

You will use the data analysis above and research into NC COVID-19 policy (as well as other factors affecting COVID-19 growth) to attempt to explain the periods of growth you found. You should look into the details of NC's shutdown policy, and evaluate the decisions made. For example:

- Does the data indicate that the shutdown was effective was in slowing infection?
- Does the data indicate that the end of the strictest phase of shutdown led to increased infection?
- What event(s) might have led to the beginning of each of the growth periods? (Note: the first growth period is simply the early infections in NC.)

You should keep in mind the concept of a lagging indicator, and that hospitalizations lag infection by around ten days.

Your report should present your data analysis regarding the three periods of growth, and an evaluation of policy choices made in the light of it. Be sure to explain why your answers are best seen in the light of evaluating past policy and events, and explain what measurements might be more useful in setting future policy.

Your report will be assessed on the following criteria:

- **Format:** adheres to 'Guidelines for Technical Writing in Math', including legibility and structure, figure labels and titles, citations, etc.
- **Clarity:** adheres to 'Guidelines for Technical Writing in Math', including grammar, appropriate attention to detail, awareness of audience, etc.
- **Data Analysis:** includes complete and correct mathematics, clear description of construction of the model and criteria applied.
- **Policy Evaluation:** clear and well-supported discussion of the details of NC's shutdown policy and recommendations for the future based on your data analysis, including answering all questions posed in Report instructions and demonstrating understanding of your model's parameters, limitations, and predictive power.