Fitting Power Laws and Exponential Relationships using Linear Regression

Nicholas Kortessis

Introduction

Many relationships in biology cannot be fit with straight lines. Lines that are not straight are known as **nonlinear**. Simple linear regression is not well suited for these kinds of problems. However, the tools of linear regression can be very helpful for fitting some classes of nonlinear curves that are commonly found in biology.

We will talk about two such curves: **power laws and exponential relationships**

Power laws have the following mathematical form

$$Y = aX^b.$$

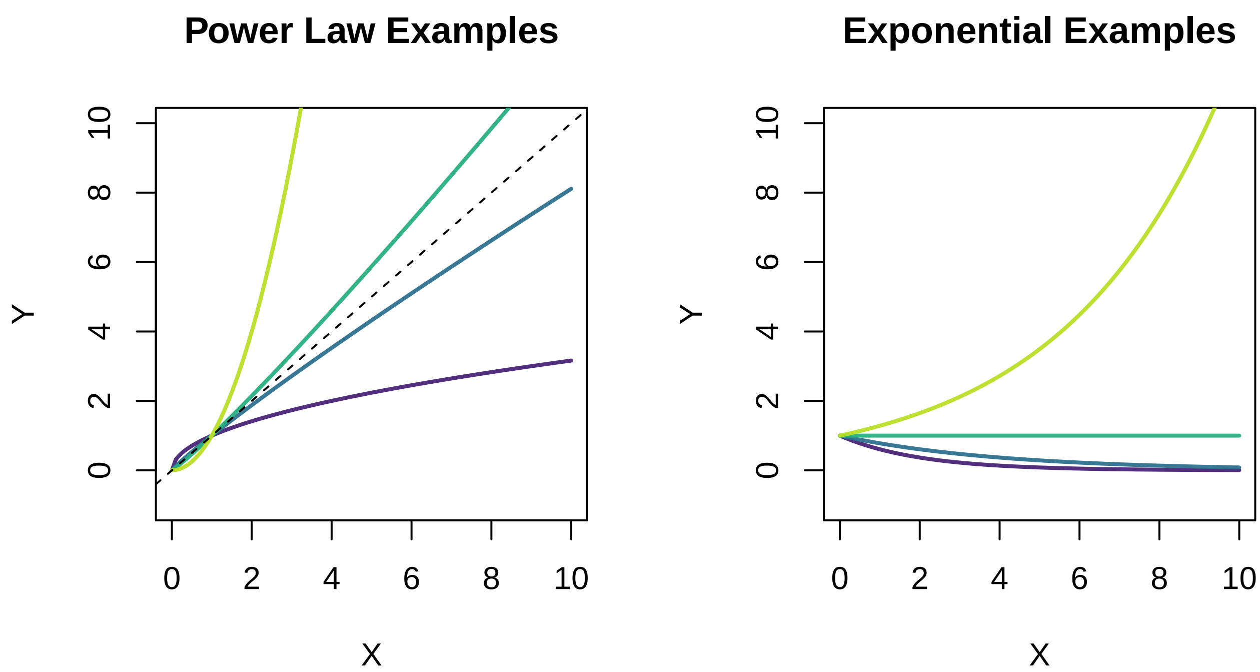
The parameters here are a and b. The parameter a is the value of Y when $X = 1$, and the parameter b is a parameter that controls the shape of the relationship between X and Y. Power laws are pretty flexible and include many common curves you may be familiar with. For example, when $b = 2$, the relationship between X and Y is quadratic. When $b = 1/2$, the relationship is that of a square root function.

Exponential relationships have the following mathematical form

$$Y = ae^{bX},$$

where a is a parameter giving the value of Y when $X = 0$ and the parameter b is a coefficient illustrating how much Y changes with X. Larger values of b indicate faster exponential relationships. Positive values of b indicate positive relationships and negative values of b indicate negative relationships. The special case of $b = 0$ indicates no relationship between Y and X because $Y = ae^{0X} = ae^0 = a$.

Below are some examples of power laws and exponential relationships.



These can come in many different shapes and most clearly are not straight lines. How are we supposed to fit such shapes?

Transforming Data To Linear Form

A key feature of these two relationships is that they are linear on a different scale. Let's see how this works.

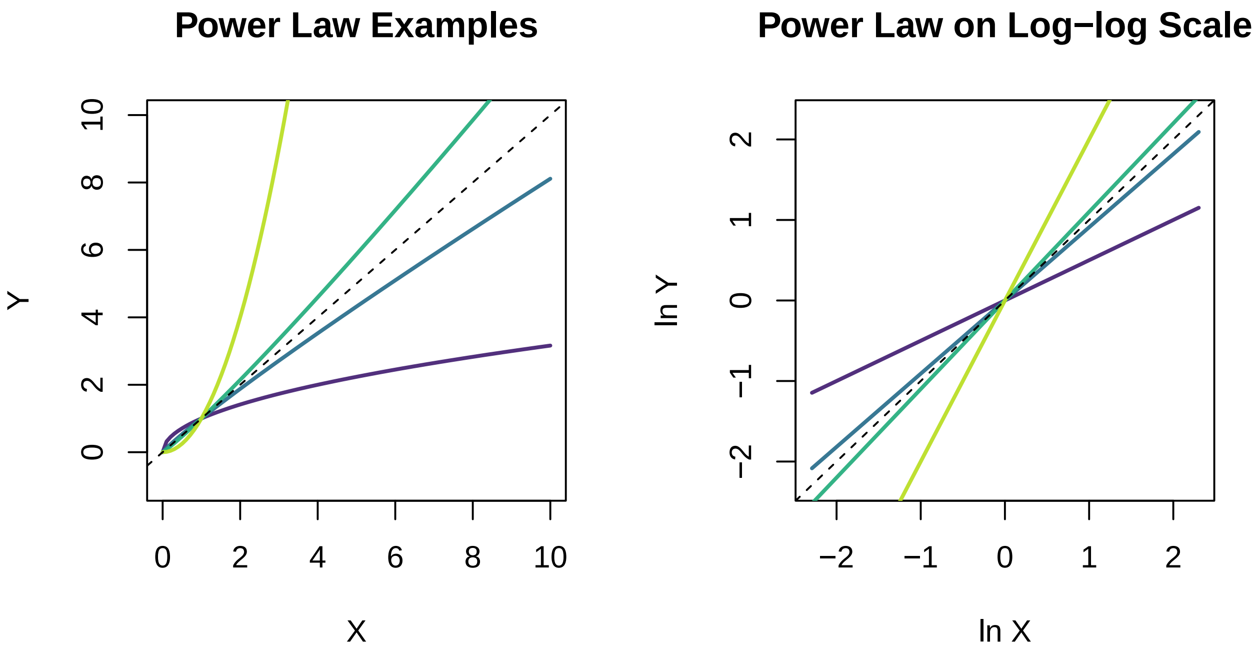
Power Laws on the Log-Log scale

For power laws, let's log-transform the Y values. (Any logarithm will do. I will use the natural log, ln, which is most common in science. Log base 10 and log base 2 are other options that you can see in certain subfields. For example, genomics used log base 2 to represent 'fold change'. All work exactly the same as the math here.)

Taking the log of the power analysis gives

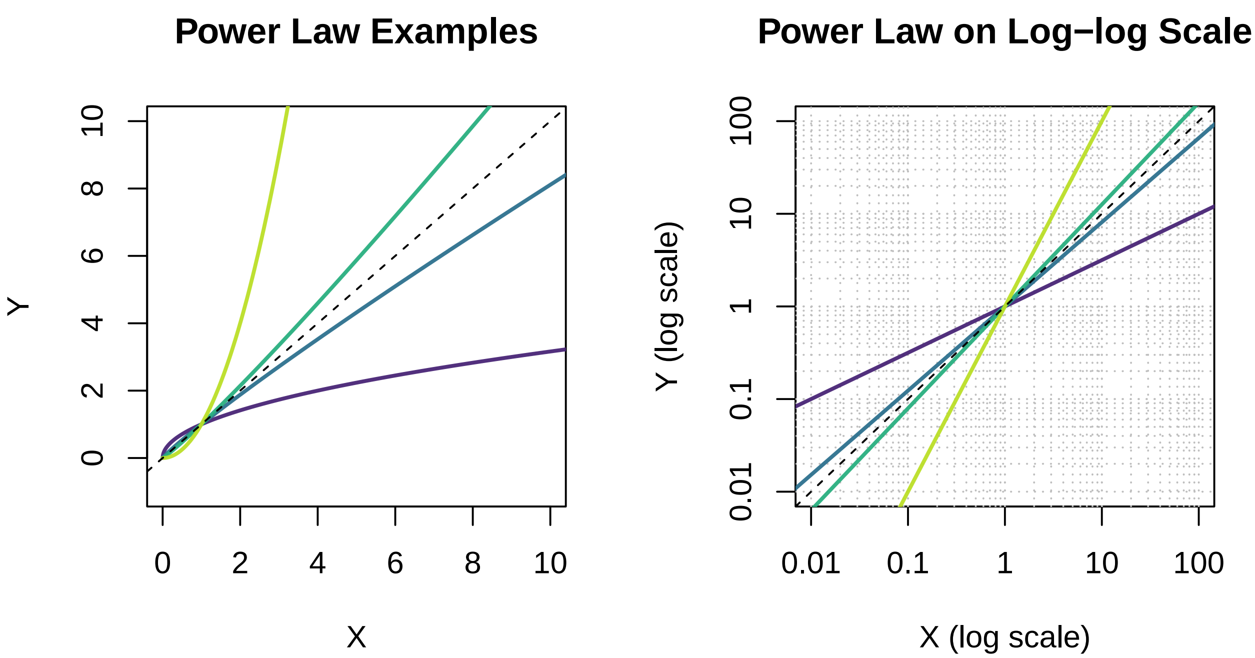
$$\underbrace{\ln Y}_{\text{new Y}} = \ln\{aX^b\} = \underbrace{\ln a}_{\text{intercept}} + \underbrace{b}_{\text{slope}} \times \underbrace{\ln X}_{\text{new X}}$$

What this says is that if we plot ln Y against ln X, power laws look like a straight line! The intercept of the straight line is equal to ln a and the slope of this line is equal to b. Log transforming X and Y on the examples above gives



You can see that the intercept for all cases is 0 and from the equation above, we know that the intercept is ln a, meaning that $\ln a = 0 \rightarrow a = 1$. The slopes of these lines also give the value of b in the power law relationship.

One thing that might help is putting the actual values of X and Y on the figure, like this.



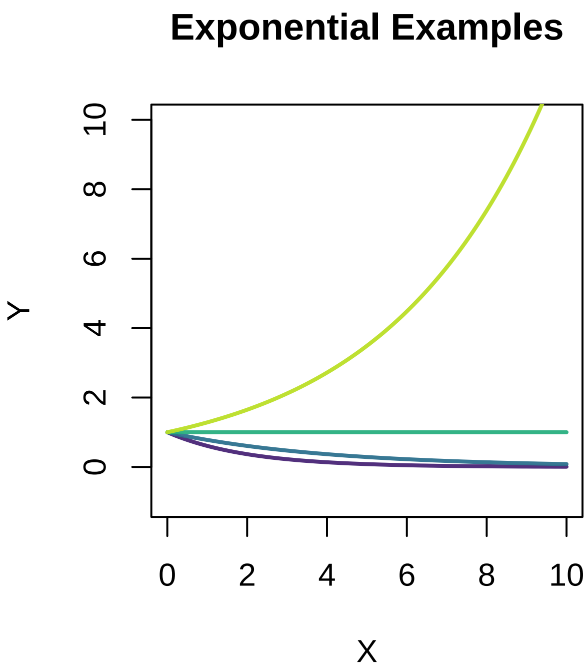
Exponential Relationships

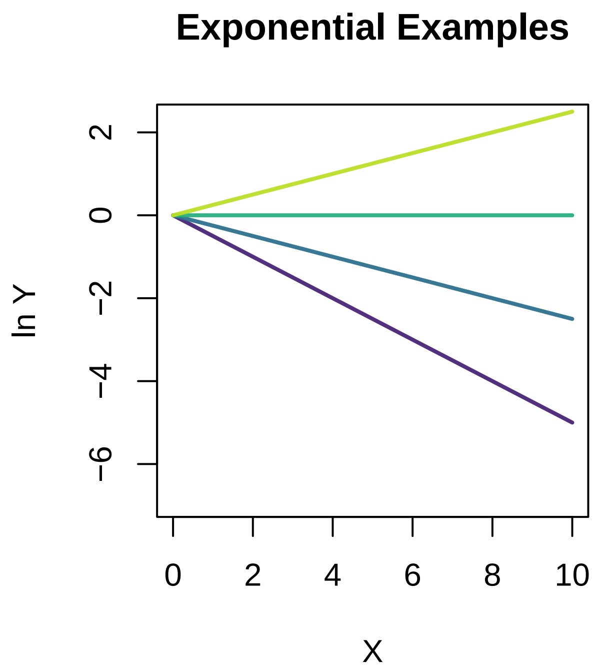
We can also find a scale at which exponential relationships look linear. Again, we take the natural log of the exponential relationship to find

$$\underbrace{\ln Y}_{\text{new Y}} = \ln\{ae^{bX}\} = \underbrace{\ln a}_{\text{intercept}} + \underbrace{b}_{\text{slope}} X.$$

This transformation shows that a plot of $\ln Y$ against X is a straight line for exponential relationships. The slope of a straight line relationship gives the parameter b and the intercept is the natural log of the parameter a.

Here is the original figure of the exponential relationship but where the y-axis is plotted on the log scale.





You can see that that all the curves have an intercept at 0 and since the intercept is $\ln a$, then $\ln a = 0 \rightarrow a = 1$.

Identifying Power Laws and Exponential Relationships In Data

If you want to know whether data fit a power law or an exponential relationship, you have an easy test.

1. Log transform the y values. If you get a straight line relationship between ln y and x, then you have a decent fit for an exponential relationship.
2. Log transform the y and x values. If you get a straight line relationship between ln x and ln y, then you have a decent fit for a power law relationship.

Power Law Example: Metabolic Scaling Analysis

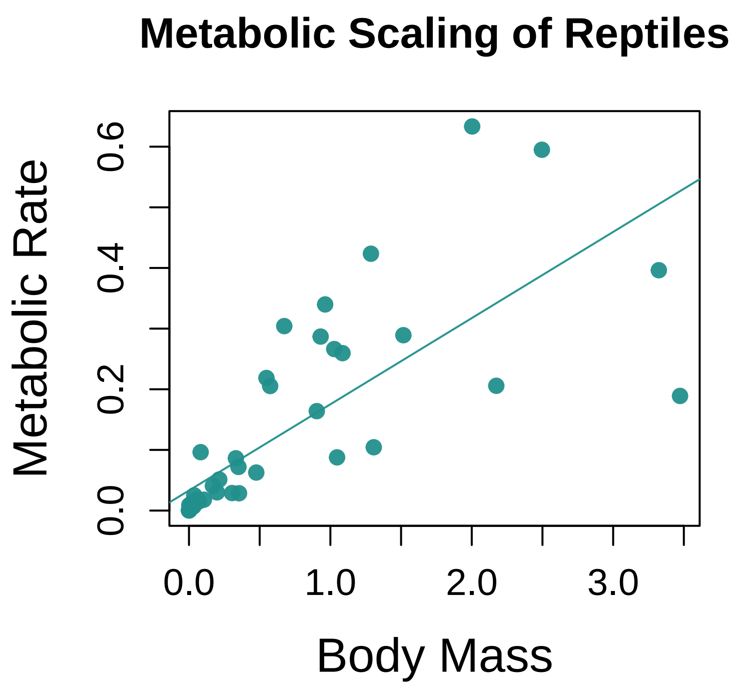
Load and Prepare the Data

We start by analyzing the metabolic scaling of reptiles. Metabolic scaling relationships often follow power laws, with good biological theory behind it. In fact, most body size relationships follow power laws, and the study of body scaling relationships is called allometry.

For the study of allometry, we often want to know whether one component of the body increases greater than proportional to body size or less than proportional to body size. For greater than proportional scaling, $b > 1$ and the body size component makes up a larger proportion of body size for bigger individuals. Fiddler crab claws are larger on larger fiddler crabs, but they also make up a larger proportion of the body size of crabs.

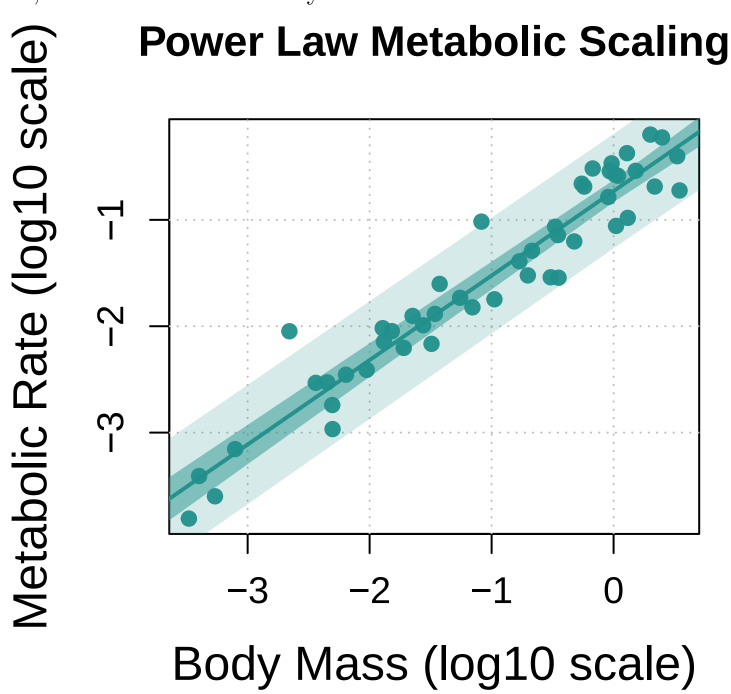
Human heads are also bigger for larger humans, but our heads make up a smaller proportion of our body size as we grow. This is an example where the scaling coefficient $b < 1$.

Metabolic rate works the same way and there is decent theory to suggest many organisms follow a 3/4 exponent scaling law between body size and basal metabolic rate. Here is some data for many different reptile species, shown on the normal scale.



This looks like it might fit a linear regression but it has got the problem that the residuals are not constant for the line. Based on allometric arguments, let's look at the log-log scale.

Log-Log Plot for Power Law Fitting A log-log transformation allows us to evaluate whether a power-law relationship might fit the data well. The best fit as well as the confidence interval and prediction interval are given in the figure below, as well as the summary of the model fit.




```
##
## Call:
## lm(formula = log10(metabolic.rate) ~ log10(body.mass), data = metab.subset.df)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.45609 -0.21705  0.00919  0.14590  0.79324
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.72986    0.05133  -14.22   <2e-16 ***
## log10(body.mass)  0.79412    0.03340   23.78   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2623 on 47 degrees of freedom
## Multiple R-squared:  0.9232, Adjusted R-squared:  0.9216
## F-statistic: 565.3 on 1 and 47 DF,  p-value: < 2.2e-16
```

It seems to work pretty well. The residuals are much more evenly distributed around the line and seem to have equal variance regardless of the value of the predictors. The estimated slope here is 0.79 with a standard error of 0.033 meaning that the theoretical expectation of 3/4 slope is within the 95% confidence interval (shown below). Moreover, the intercept is estimated at -0.73, meaning that the basic metabolic rate with a body size of 1 $e^{-0.73} = 0.482$ and the confidence interval for the basal metabolic rate is between 0.435 and 0.534.

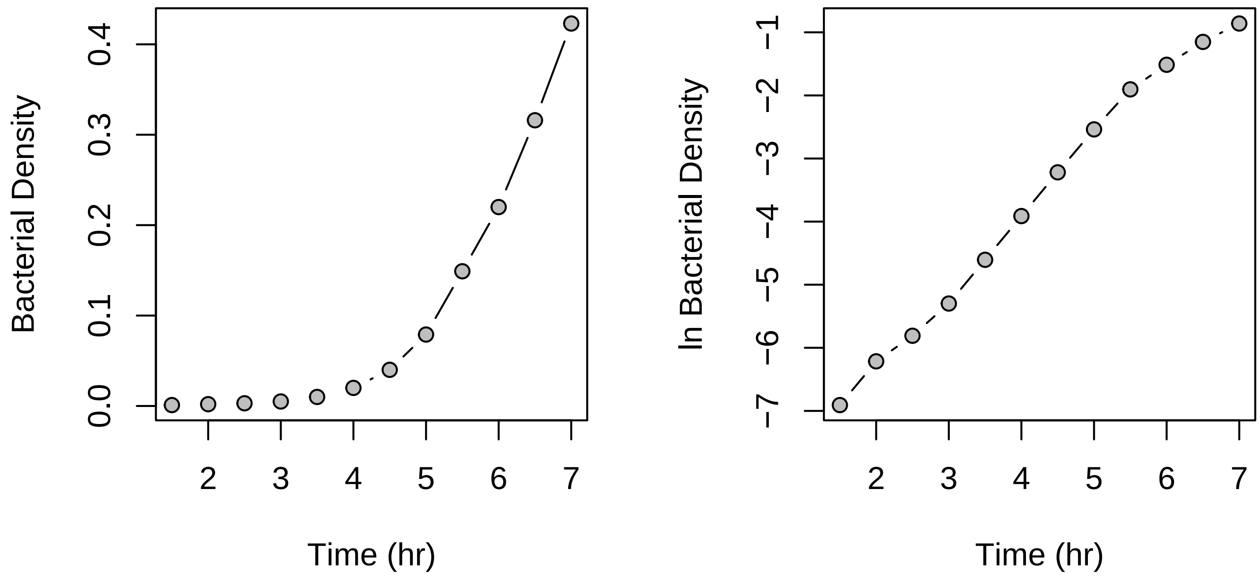
```
## [1] "b CI =  0.726928406800624" "b CI =  0.861314660175588"
```

```
## [1] "a CI =  0.434690828020528" "a CI =  0.534409508950099"
```

Exponential Growth Analysis

Exponential relationships are very common in population growth data because populations reproduce multiplicatively. If each individual in a population has two offspring over some time period, the population doubles each time period. That is multiplicative growth as opposed to additive growth. Populations that grow additively add the same number of individuals over time. Other examples of exponential growth include the spread of disease, the spread of cancer, and the growth of investments.

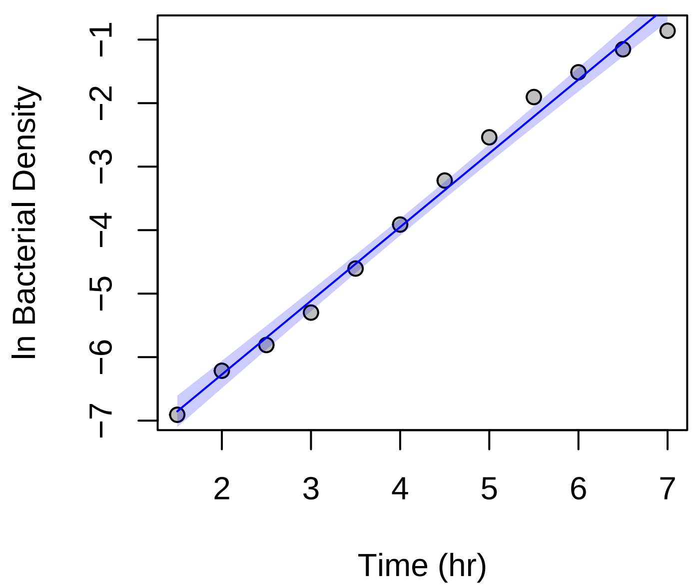
Here, we will use a dataset of bacterial growth over time. The data are from a study of the growth of a bacterium called *Klebsiella pneumoniae* in a laboratory setting. Let's take a look on the normal and log scales.



The line on the left panel surely looks like exponential growth and the panel on the right more or less confirms that it is because we get a mostly straight line between ln bacterial density and time. This is a plot of log transformed y against x, and a straight line on that scale indicates exponential growth.

Let's see how fast the population grows by fitting a linear model.

Fit and Visualize the Exponential Model Using linear regression, we fit an exponential growth model and add confidence intervals to the plot.




```
##
## Call:
## lm(formula = log(No.36.1) ~ time..h., data = growth.df, subset = (time..h. <=
##     7 & time..h. >= 1.5))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.39178 -0.10627 -0.00725  0.12459  0.30621
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.59525    0.15709  -54.72 1.01e-13 ***
## time..h.     1.16095    0.03425   33.90 1.18e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2048 on 10 degrees of freedom
## Multiple R-squared:  0.9914, Adjusted R-squared:  0.9905
## F-statistic:  1149 on 1 and 10 DF,  p-value: 1.18e-11
```

The exponential growth rate estimated by this model is given by the slope, which is estimated to be 1.16 per hour, meaning the population doubles every 0.59 hours (or about 36 minutes). In evolutionary biology, this is called *Malthusian fitness*. The 95% confidence interval for the growth rate is between 1.08 and 1.24 per hour (33.6 - 38.3 minutes).

Conclusion

This document demonstrated how to fit power laws and exponential growth models using linear regression. These techniques are powerful tools for analyzing relationships in biological and ecological data.