Logarithms, Magnitude, and Regression for Biological Scaling: A Computational Science Module

Andrew J. Kerkhoff, Departments of Biology and Mathematics, Kenyon College, Gambier, Ohio, 43022

Math and Science Level: <u>Introductory</u>

Keywords: *Metabolism, Allometry, Caterpillars, Non-majors, Regression, Statistics,*

Module Overview

This computational science module covers the basics of logarithms and exponents (i.e., magnitude scales) and the statistical technique of least-squares regression in the context of biological scaling, or *allometry.* Allometry is the study of how various aspects of life change quantitatively with organism size. The idea of magnitude (and thus logarithms and exponentiation) is essential in this context because organisms span an astounding 21 orders of magnitude in mass. That is, the largest organisms, like blue whales and giant sequoias, are approximately 10,000,000,000,000,000,000,000 (or ten-thousand-million-million-million) times heavier than the smallest bacteria, yet all organisms must abide by the same laws of physics, biochemistry, genetics, physiology, and evolution. Comparing organisms on magnitude or logarithmic scales allows us to better understand how the various aspects of biological form, function, and diversity change with organism size.

The module was originally designed to provide undergraduate non-science majors with the tools necessary for understanding biological scaling and allometric data and models. As such, I assume very little knowledge of biology, statistics, and mathematics. For example, in describing least-squares regression, I do not provide the mathematical details of computing the regression line, as that would entail an understanding of basic statistical concepts of mean, variance, and correlation, as well as the particulars of statistical inference. Instructors could certainly extend this section to include such details, as I do when I include this exercise in my introductory statistics class.

The *Background* part of the module reviews the basic mathematics of logarithms and power functions, with the goal of having students: *i)* understand logarithmic and exponential operations and their algebraic rules, and *ii)* grasp the utility (and meaning) of "magnitude scales." The *Model Details* section provides the basics of graphical relationships between two biological variables using log-log scatterplots and their mathematical description using least-squares regression. In the *Model Assessment* section of the module, the students are then asked to: *i)* construct and qualitatively interpret trends in bivariate scatterplots, *ii*) construct and interpret least-squares regression models, including the regression equation, residual variation and R^2 value, *iii)* apply logarithmic-transformation to data, and *iv)* interpret log-log (power-law or multiplicative) graphs and regressions.

The module uses Microsoft Excel, because it is widely available and generally familiar to students even outside of the sciences. However, it could easily be adapted (and extended) with more powerful statistical software like Minitab, Systat, SAS, SPSS, or the freely available R.

Introduction

Individual organisms span an amazing size range. The ratio of the mass of a redwood or a blue whale to that of a bacterium is approximately 10^{21}; that is, a redwood is ten-thousand-million-million-million times heavier than the smallest bacteria. To put this ratio into perspective, it is similar to the mass ratio of the Moon to a typical human, that of a human to a single molecule of cytochrome-c oxidase (a protein that facilitates cellular respiration), and that of the known universe to our Sun! The scales of ecological interactions between species are yet broader, spanning over 30 *orders of magnitude* (powers of 10) in mass, from the smallest interacting microbes to the entire biosphere (which is estimated to weigh in at $\sim 1.8 \times 10^{19}$ g). Even among the more familiar land mammals, an elephant is almost six orders of magnitude (i.e., 10^6 or one million times) heavier than the smallest mouse. Thus, understanding biodiversity and the ecological complexity of life on Earth is, at least in part, a matter of scale.

An organism's size affects nearly every aspect of how it lives. For example, because they are "small," ants can walk up a vertical surface and survive a fall from any height without injury. The same is clearly not true for humans. A mouse's heart beats 500 times per minute, while an elephant's only beats 28 times, but over their lifetimes, both will experience (on average) approximately the same number of heartbeats – about 1.5 billion. Why 1.5 billion? Understanding how life must change with size is one of the keys to explaining these sorts of mysteries and to developing a more complete understanding of how nature works.

Statement of problem or question

As the ecologist Robert McArthur put it, to do science "is to search for repeated patterns, not simply to accumulate facts." In this case, we are searching for patterns that apply across all creatures, great and small. So the question arises: ***how can we describe the changing characteristics of life over a large range of sizes?***

Interestingly, despite the daunting complexity and variety of biological phenomena, they often exhibit striking mathematical regularities called scaling laws. In the most general sense, scaling laws are simply quantitative descriptions of how some property of a system changes along one of the fundamental dimensions of the system. These dimensions may be physical dimensions like mass, length, or time. Alternatively, the dimension of interest may be more specifically biological, such as the mean population density of a species or the area of an island or habitat patch.

Among the most well known scaling laws in biology are so-called *allometric* relationships between organism size and various aspects of form, physiology, and ecology. The term *allometry*, which translates from Greek as, "other measure," was coined by Julian Huxley and George Tessier in 1936. While the term was originally applied to changes in the morphological proportions of organisms as they grew, it now is used to describe the study of size-related changes in living things more generally. Here we will model developmental changes in the metabolism of the tobacco hawkmoth caterpillar, *Manduca sexta*, but similar allometric analyses can be applied to a wide variety of biological and ecological phenomena. Metabolism is the sum of biochemical processes carried out in the cells of living things by which energy is provided and materials assimilated for all vital processes and activities. The fundamental importance of metabolism to all living things has made it the focus of many

allometric studies. Furthermore, the statistical method we will be using, called *least-squares regression*, is used in a wide variety of other situations.

Background information

Allometric scaling laws typically take the form of a power function:

$$Y = aX^b$$

where Y is the property of interest (e.g., metabolic rate), X is the size of the observed entity along the dimension of interest (e.g., organism body mass), and a and b are called the scaling coefficient and exponent, respectively. The equation provides a *mathematical model* of the relationship between two measurable, biological variables. This exercise is meant to help you understand how these mathematical models are constructed and assessed, based on actual data. Since we are going to be working with power functions, it is useful to review some of the basic mathematics of powers.

First, remember that 10^b simply means "multiply *10* times itself b times," just as $b \cdot 10$ means "add *10* to itself b times." In this way, powers are to multiplying as multiplying is to adding. Just like with multiplication, the same operation applies even when the exponent b is negative, or a fraction. Thus, even though it is hard to imagine multiplying *10* times itself -1/4 times, it can be done.

Second, recall that just as addition has subtraction and multiplication has division, powers also have an inverse function: logarithms. Logarithms are the inverse of powers in the sense that $Y = 10^Z$ means the same thing as $Z = \log_{10}Y$. Note that the base-10 logarithm of Y ($\log_{10}Y$) is simply the number of times the base (10) is multiplied by itself to yield Y. That is, *taking the logarithm of a number yields an exponent*. In this exercise, we will always work with base-10 logartithms. Base-10 logarithms are useful because they describe the *order of magnitude* of whatever process or phenomena is under consideration.

Finally, let's review a few of the algebraic "rules" associated with powers and logarithms that will come in handy. For powers, recall the following:

1. A product raised to a power equals the product of each term raised to that power: $(a \cdot b)^x = a^x \cdot b^x$
2. The product of a common base raised to two different powers equals that base raised to the sum of those powers: $a^x \cdot a^w = a^{(x+w)}$
3. The quotient (or ratio) of a common base raised to two different powers equals that base raised to the difference of those powers: $\dfrac{a^x}{a^w} = a^{(x-w)}$
4. Negative exponents are reciprocals: $a^{-x} = \dfrac{1}{a^x}$

Likewise, for logarithms:

1. The logarithm of a product equals the sum of the logarithms: $\log(a \cdot b) = \log a + \log b$
2. The logarithm of a quotient equals the difference of the logarithms: $\log\left(\dfrac{a}{b}\right) = \log a - \log b$
3. The logarithm of a power function equals the exponent times the logarithm of the base: $\log(a^x) = x \cdot \log a$

4. The logarithm of the reciprocal of a number is the negative of the logarithm of that number: $\log\left(\dfrac{1}{a}\right) = -\log a$

Note that addition of logarithms is the same as multiplication of the original numbers. This is why logarithms were originally invented – to simplify complex multiplicative calculations. It also means that logarithms change in multiplicative, rather than additive increments, with each integer change, say from 3 to 4, corresponding to another power of the logarithmic base, in this case from 1,000 to 10,000. Furthermore, on this *logarithmic scale* (also called a *magnitude scale*), the step from 1,000 to 10,000 is the same as the step from 0.001 to 0.01 or the step from 10^{50} to 10^{51}. Thus, while ordinary additive increments tell us "how much more," logarithmic increments tell us "how many times more."

Returning to our allometric equation from above $Y = aX^b$, we can show that taking the logarithm of both sides produces the equation for a straight line $\log Y = \log a + b \cdot \log X$, where a is the intercept (i.e., the value of $\log X$ where $\log Y = 0$) and b is the slope. For the last two steps in the derivation, identify the relevant algebraic rule from above.
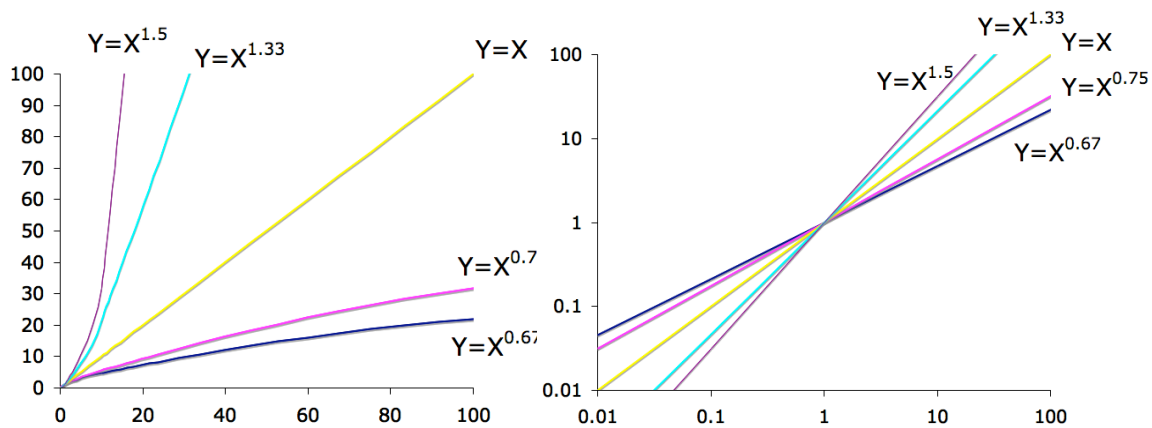
$$Y = a \cdot X^b$$

$$\log Y = \log(a \cdot X^b)$$

$$\log Y = \log a + \log X^b$$

$$\log Y = \log a + b \cdot \log X$$

Thus, a power function relationship between two variables X and Y means that their logarithms are linearly related, with a slope equal to the exponent of the power function. Recall that a slope is defined as the "rise over run" or, in this case (since we are working with base-10 logarithms), the change in the *magnitude* of Y for every 10-fold change in X. Looking at some graphs of power functions on both standard *arithmetic* and *logarithmic scales* can help to understand them better.

The left hand graph shows power functions with five different exponents. In all cases, the coefficient (a in $Y=aX^b$) is set to one, for simplicity. Note that when the exponent is less

than one, the line climbs at an ever-decreasing rate, while when the exponent is greater than one, it climbs at an ever-increasing rate. In both cases, the curves are not straight lines, that is, they are *non-linear*. Of course, when the exponent is equal to one, we do have a straight line, in this case $Y=X$. Finally, note that when viewed on an arithmetic scale, the five lines all appear to converge towards zero.

Now examine the same five power functions on logarithmic scales (right hand graph), on which each equal increment is a power of ten. As we would predict from the derivation above, each function is linear, and since the coefficient *a* is always one, the lines only differ in their slopes, and they all cross at the point [1, 1]. The constant slopes indicate that for every 10-fold increment in *X*, *Y* changes 10^b-fold. Viewed on logarithmic scales, we can also see that the curves diverge on the small end as well as on the large end. These differences are not apparent on the left hand graph because they are compressed down into the space between zero and one.

In a way, these two graphs represent different ways of looking at the world, and some of the differences are subtle. For example, the way in which the left-hand graph "minimizes" the differences between the power functions near the origin shows that arithmetic scales are sensitive to the units of measure: 1 mm is very different from 1 km, but the space between zero and one is always the same on the graph. Logarithmic scales, on the other hand, present proportional changes that are insensitive to the units of measure, because ten times longer is ten times longer, whether in mm or km. This difference does not make one quantitative view superior to another, but it does mean that one or the other may be more appropriate, depending on the situation. In our case, logarithmic scales are the appropriate choice.
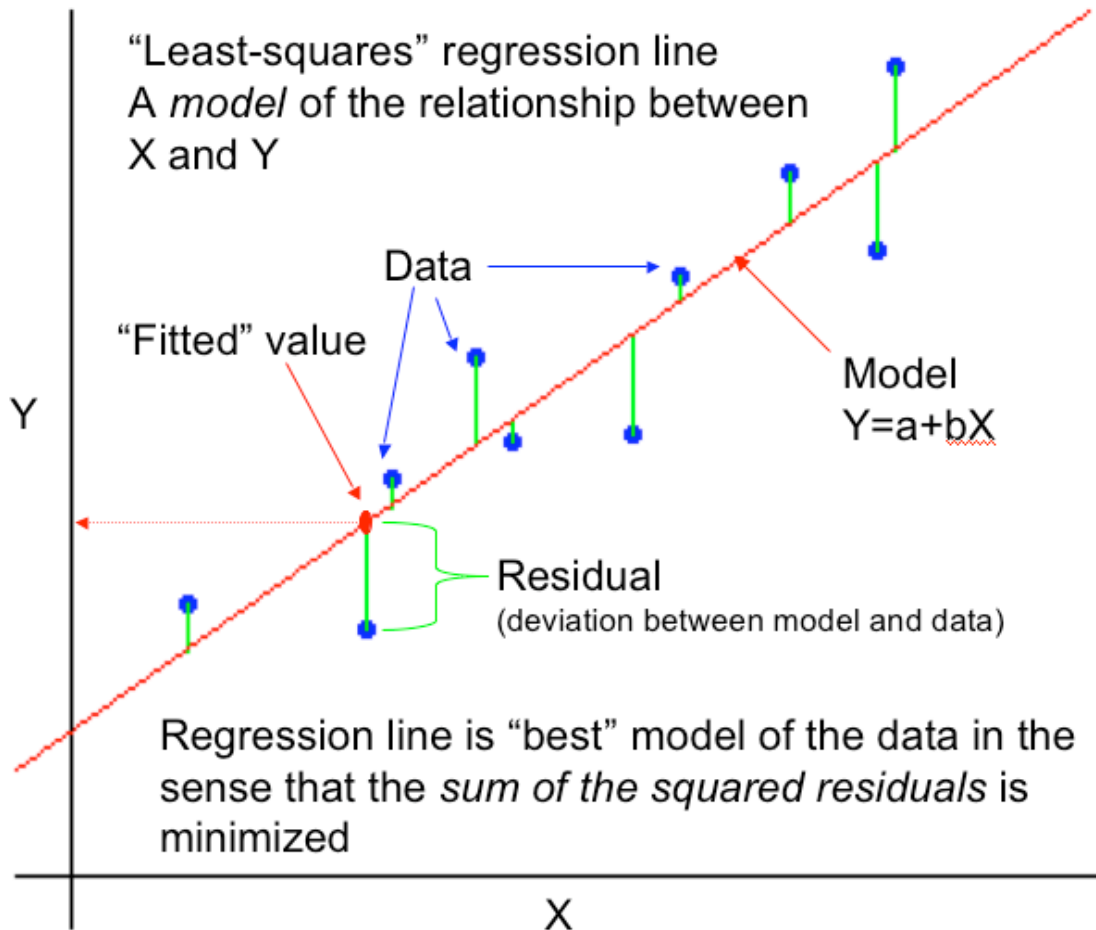
Model Details
Allometric scaling laws are not (generally) derived from basic theoretical considerations, like the laws of physics or population genetics, nor are they just arbitrary mathematical equations. Instead they are statistical descriptions of empirical data; useful simplifications that help us to understand and make predictions about biological phenomena.

For example, in the 1930s, physiologists Francis Benedict, Max Kleiber and others made careful measurements of metabolic rate for mammals ranging from a common house mouse to an African elephant. They then plotted the logarithm of the metabolic rate as a function of the logarithm of the body weight of the animal, and found a very regular relationship between the two.
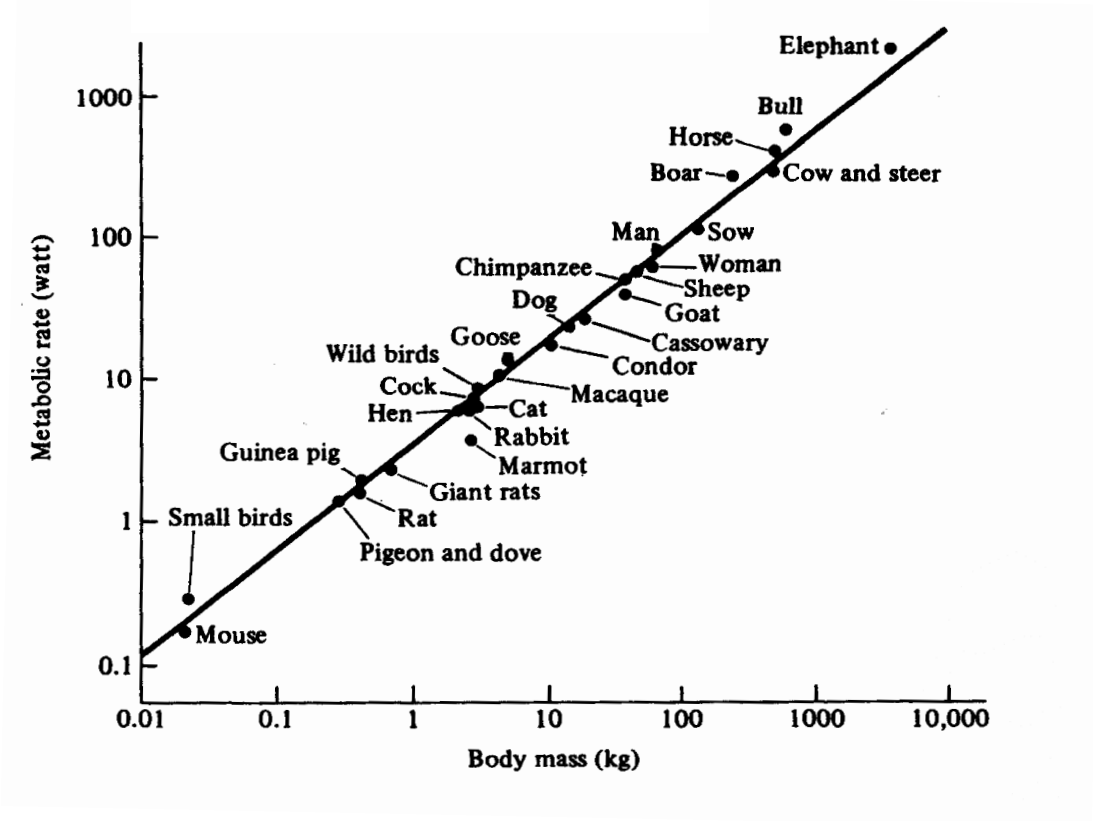
It is clear from the graph that a line, that is a power function, can be used as a *mathematical model* of the data, a compact representation of the relationship between the measured variables, metabolic rate and body mass. There are many possible models (just think of all the possible lines you could draw through the data!), so how do we choose the "best" one. Of course, "best" depends on what you want to use the model for, but the most common method for describing the relationship between two variables Y and X is called *least-squares regression*.

You can think of regression as using X to ***predict*** Y, and the least-squares method is just a way of minimizing the likely error of the prediction. From each X value, you can use the equation for a line to compute a predicted Y value. Then you can compare the *predicted* Y value to the *actual* Y value. The difference between the two (in units of the Y variable), is called the *residual*. The least-squares regression line is the line that *minimizes the sum of the squared residuals*. Thus, the least-squares regression line, minimizes the error you can expect to make in predicting Y, based on knowledge of X. The features of the least-squares regression line are summarized below.

"Least-squares" regression line
A *model* of the relationship between X and Y

Data

"Fitted" value

Y

Model
Y=a+bX

Residual
(deviation between model and data)

Regression line is "best" model of the data in the sense that the *sum of the squared residuals* is minimized

X

It is easy to imagine a case where even the "best" line is not a very good description of the relationship between the variables, simply because the data themselves do not really fall along a tidy, straight line. We can use a value called the *coefficient of determination* (symbolized R^2 and said "R squared") to tell us how well the model "fits" the data. The value of R^2 varies from zero to one, where zero means that *X* tells you nothing about *Y*, and one means that all of the *Y* values fall *exactly* on the regression line. Technically speaking, R^2 is *the fraction of variation in Y that is explained by variation in X*.
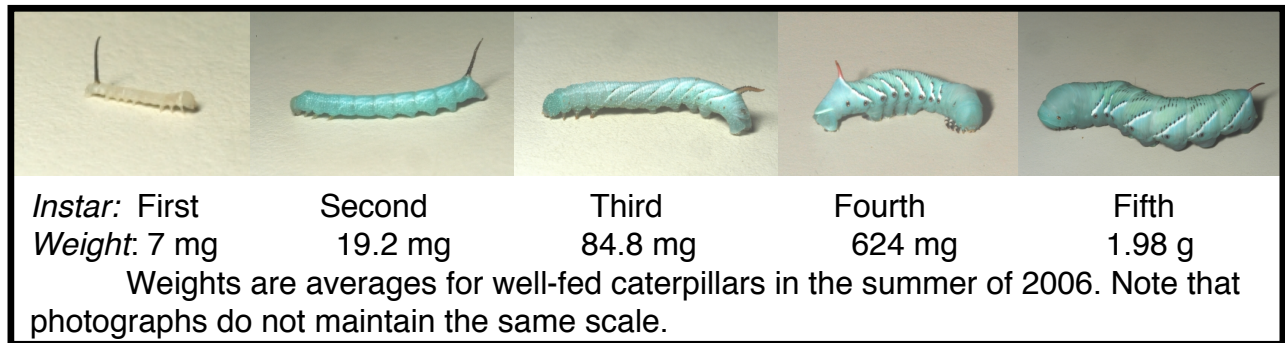
The least-squares regression line is relatively straightforward, but tedious, to calculate, so we will let the computer do the work for us. Specifically, we will use the

spreadsheet program Microsoft Excel to compute the least-squares regression line relating metabolic rate to body weight in larvae of the tobacco hawkmoth *Manduca sexta.*

Methodology

Natural History: Tobacco hornworms (*Manduca sexta*) are the caterpillar larvae of the Tobacco hawkmoth. They are a common agricultural pest of tobacco fields (hence the name) and tomato gardens across North America, but in their moth stage, they are also an important pollinator of many night-blooming plants. *Manduca* is a common laboratory organism for the study of the nervous and endocrine systems and how they change over development. They are easily reared in the laboratory and their large size (for an insect larva) makes them relatively easy to work with.

Here, we are going to use *Manduca* to study the scaling of metabolism, similar to the studies of Benedict and Kleiber described above. The caterpillars are useful in this case because they grow from a 1 mg egg to a 10+ g caterpillar (i.e., about four orders of magnitude in weight) in about 18 days! *Manduca* caterpillars develop through five *instars* or stages, between which they *molt* or shed their outer skin and other tissues. Examine the pictures below:

Instar: First   Second   Third   Fourth   Fifth
Weight: 7 mg   19.2 mg   84.8 mg   624 mg   1.98 g
     Weights are averages for well-fed caterpillars in the summer of 2006. Note that photographs do not maintain the same scale.

Measuring Metabolic Rate: We measured metabolic rate by placing individual larvae in a chamber and measuring the rate at which they release CO_2 into the air (see below).
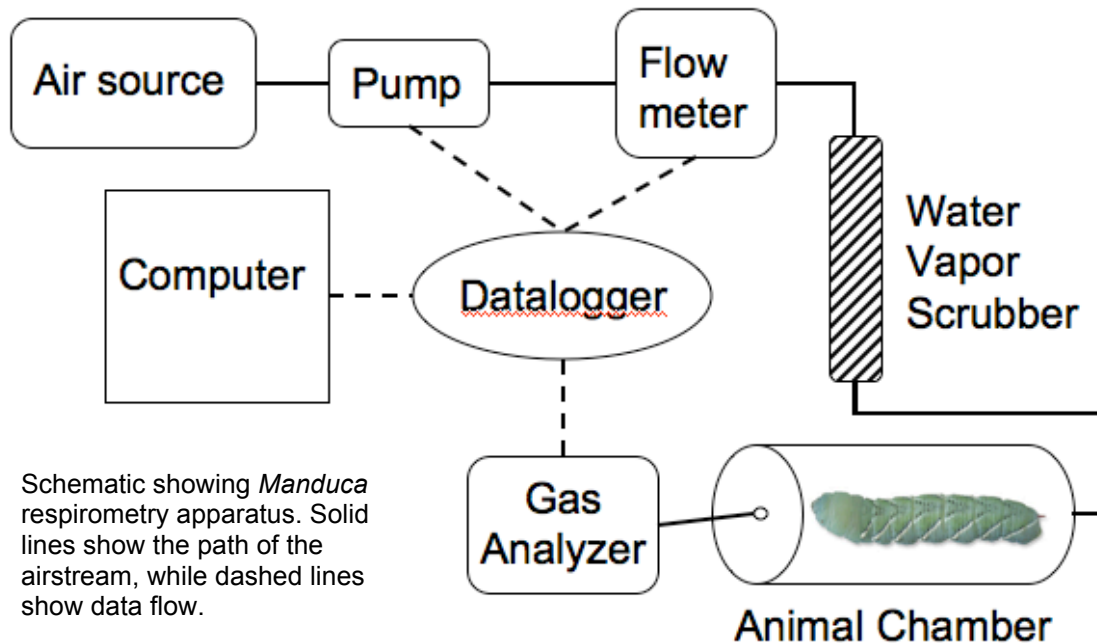
Schematic showing *Manduca* respirometry apparatus. Solid lines show the path of the airstream, while dashed lines show data flow.

All we do is take a baseline reading with the chamber empty, then introduce the animal into the chamber. Over time, as the animal respires the CO_2 concentration rises. Since we know the rate at which air is flowing through the chamber, we can use the difference between the baseline CO_2 reading and that with the animal in the chamber to calculate the metabolic rate of the whole animal.

Physiologists distinguish between whole-organism metabolic rate and so-called "specific metabolic rate," which is the metabolic rate per unit of mass. Thus, metabolic rate and specific metabolic rate have different dimensions. For example, the respirometry system described above provides measurements of metabolic rate in units of micro-mols of CO_2 per minute (μmol CO_2 min^{-1}). Mols (or moles) are just a molecule-counting unit (1 mol = 6.02×10^{23} molecules), so metabolic rate here is just the rate of carbon dioxide release. The specific metabolic rate can then be found by dividing an animal's metabolic rate by its mass, giving units, in the case of our *Manduca,* of micro-mols of CO_2 per minute per gram (μmol CO_2 min^{-1} g^{-1}). Specific metabolic rate is useful because, by "standardizing" the measurement on a gram-for-gram basis, it tells you how fast an organism's cells are actually working. Small organisms live at a faster pace than large organisms in the sense that their *specific metabolic rate*, and thus the rate at which each of their cells work, is higher. At the whole-organism level, even though their cells are working more slowly, larger organisms tend to have a higher metabolic rate simply because they are larger, and thus have many more cells.

By measuring multiple larvae every day as they grow and develop, we can collect a lot of data in a short amount of time. Our job here is to analyze data from the summer of 2006, and draw some inferences about metabolic scaling in *Manduca,* and maybe even more generally. As described above, you will use log-log scatterplots and least-squares regression to estimate the parameters (i.e., the coefficient and scaling exponent) of the

power function. You will use least-squares regression to estimate the scaling relationships from log-log scatterplots of the data.

Model Assessment
Data in the Excel file are in three columns. *Instar* is the developmental stage of the caterpillar, *Body Size* is the weight of the animal (in grams) at the time of measurement, and *Metabolic Rate* is the measured in micromols of CO_s per minute.

The first step is to use the spreadsheet to calculate the logarithm of each body size and metabolic rate data, as demonstrated in class. Make a plot of the log-transformed whole-organism metabolic rate data, and a separate plot of the untransformed data. *Compare the two plots. Describe how the two graphical representations provide different information about how metabolic rate varies with body weight.* Now, use the "Trend Line" utility (right click on a data point for the menu) to fit the least squares regression line to the log-transformed data. Make sure to select the options for displaying the equation and the R^2 value on the graph. Use the graph to answer the following questions.

What is the least-squares regression equation?

Using the rules for logarithms and powers described above rewrite the regression equation as a power law. (Remember that $Y = aX^b$ is the same as $\log Y = \log a + b \log X$ and $10^{\log a} = a$.)

What does this equation tell you about how metabolic rate changes as Manduca *larvae get larger?*

How much of the variation in metabolic rate is explained by body size?

Next, analyze the metabolic rate data for each instar separately in order to see whether metabolic scaling changes over the life of the larvae. Plot the data for each instar as a separate series on a single scatterplot, then perform the least-squares regression analysis on each instar. For each instar, record the least-squares regression equation and R^2 value.

Describe how the scaling relationships change as the larvae develop from first through fifth instars.

Write a few sentences summarizing the patterns of metabolic scaling in Manduca sexta *in relation to the patterns we have observed for other organisms.*

Finally, use the spreadsheet to calculate specific metabolic rate, then calculate its logarithm. Make a scatterplot of the log(specific metabolic rate) as a function of log(body weight) (no need to use separate instars here) and perform another least-squares regression analysis.

State and interpret (as in 1, 3, and 4 above) the regression equation and R^2 value.

Conceptual Questions

- *Compare the regression models for whole-organism and specific metabolic rate. In developing* Manduca sexta *caterpillars, does body size tell you more about one or the other? How do you know?*

- *On your original graph of metabolic rate as a function of body size, identify the two data points with the largest positive and negative residuals. Like all instruments, the respirometry equipment used to make the measurements of metabolic rate has limited accuracy, especially when the changes in CO_2 concentration are relatively small. How might this source of measurement error be reflected in the observed pattern of residual variation?*

- *Based on the mammal data shown in the figure above, Francis Benedict (who collected most of the data) remarked "It is obvious that this apparent straight-line relationship is of no physiological significance… It seems illogical to make use of complicated mathematics in the attempt to unravel the end results of the pooled activities of millions of cells." Max Kleiber, the other physiologist most closely associated with this pattern, disagreed strongly with Benedict's interpretation. Respond to Benedict's statement, taking Kleiber's point of view, and recalling McArthur's assertion that to do science, "is to search for repeated patterns, not simply to accumulate facts."*

Problems and Projects

Max Rubner, Max Kleiber, and **Manduca***:* By the late 19th century, it was clear to physiologists that whole-organism metabolic rate changed as a power function of body mass with an exponent different from one; that is, the relationship was non-linear. One of the foremost physiologists of the time, Max Rubner, proposed that the exponent should be around 2/3, based on the theoretical geometric relationship between the mass of an animal and its ratio of surface area to volume. Careful work with dogs of various sizes seemed to confirm his result.

The careful analyses by Max Kleiber of the data collected by Benedict indicated that across all mammals, the exponent was substantially higher than 2/3, somewhere very close to 3/4. While there was no theoretical basis for the difference at the time, a number of theories have been put forward since then, and this area of biological science remains an exciting field full of open questions.

The goal of this project is to compare the observed exponents to Rubner's "Surface Law" (where the exponent is 2/3) and to "Kleiber's Rule" (where the exponent is 3/4).

What scaling relationships are expected between body weight (W, the X variable) and **whole-organism** *metabolic rate (B, the Y variable) based on the Surface Law and Kleiber's Rule?*

    *Surface Law:*_____   *Kleiber's Rule:*_____

*What corresponding relationships are expected between body weight (W, the X variable) and **mass-specific** metabolic rate (B/W, the Y variable) based on the Surface Law and Kleiber's Rule?*

*Surface Law:*_____ *Kleiber's Rule:*_____

Based on your results, is either Rubner's Surface Law or Kleiber's Rule supported in the metabolic scaling of Manduca sexta*? Explain.*

Empirical Data and Solutions

Available from the author: kerkhoffa@kenyon.edu

Suggestions to Instructors

Statistical Extensions: The elementary level at which least-squares regression is handled in this module leaves ample room for more technically advanced applications. To list only a few:
1. F-statistics and inference techniques could be used to assess the significance of the regression model parameters.
2. 95% confidence intervals on the regression slope could be used to compare empirical scaling exponents to expectations, e.g., 2/3, 3/4.
3. Analysis of covariance (ANCOVA) could be used to assess differences in slope and intercept between instars.
4. Alternative regression models (e.g., reduced or standardized major axis regression) could be used to estimate slope values.

Biological Extensions: The same analytical techniques can be applied to a variety of problems in biological scaling, and both the primary literature and the internet provide a wealth of available data and problems for student analysis.

For example, students could compare the *Manduca* data analyzed here directly to the available data for mammals of all sizes compiled from the literature. In this case, the metabolic rate data will need to be transformed to watts to have a common basis for measurement. Since watts are J/s, we first have to divide by 60 to get micromoles of CO_s per second. We can then use a standard "respiratory quotient" of 0.82 mol CO_2 per mol O_2 and the corresponding energetic yield of respiration (20.2 J per mol O_2) to find *Manduca* metabolic rate in watts. These data can then be combined with the available mammal data to compare metabolic scaling in a single species of caterpillar to that of the entire class mammalia. The comparison can either be made qualitatively, as in the **Projects and Problems** section above, or more objectively using more advanced statistical methods like ANCOVA.

Please contact the module author for data and other suggestions.

Acknowledgements

Glossary of Terms

- *Allometry:* The study of size-related changes in form and function in living things, focused on the development and explanation for quantitative descriptions of patterns across a wide array of sizes.
- *Manduca sexta:* Tobacco hornworms/hawkmoths.
- *Metabolism:* The sum of biochemical processes carried out in the cells of living things by which energy is provided and materials assimilated for all vital processes and activities. May be accounted for on a whole-organism or mass-specific basis.
- *Order of magnitude:* Power of ten.
- *Respirometry:* The measurement of respiration, and thus metabolism, via the measurement of oxygen or carbon dioxide exchange between organism and environment.

References

Bonner, J.T. 2006. Why Size Matters. Princeton University Press. 161 p.

Calder, W.A. 1984. Size, Function, and Life History. Harvard University Press. 431 p.

Kleiber, M. 1961. The Fire of Life. Wiley. 454 p.

McMahon, T.A. and J.T. Bonner. 1983. On Size and Life. Scientific American Press, Freeman, Inc.. 255 p.

Peters, R.H. 1986. The Ecological Implications of Body Size. Cambridge University Press. 329 p.

Schmidt-Nielsen, K. 1984. Scaling: Why Animal Size is So Important. Cambridge University Press. 241 p.

Whitfield, J. 2006. In the Beat of a Heart: Life, Energy, and the Unity of Nature. Joseph Henry Press. 270 p.