Leaf cutter ant foraging*[[1]](#footnote-1)*

****Leaf cutter ants, like the species *Atta columbica* from Central America, live in colonies of several million individuals. These ants farm a particular fungus species in their nests, and feed from the fruiting bodies of the fungus. In order to feed the fungus (which, of course, is itself heterotrophic) they cut pieces of leaves from nearby vegetation and return the pieces to their nest. Like many social insects, they have distinct castes that divide the labor of the collective society: recon ants scout for new leaf sources and establish trails, foragers cut leaves and bring them to the nest, farmers tend the fungus, nursery ants care for larvae, and the queen takes care of the reproduction.

Watch [**this video**](http://www.radford.edu/jmwojdak/images/ant-intro.wmv) (may take a minute or two to download) of ants foraging in a forest in Panama.



Wow! As you can probably now imagine, leaf cutter ants can defoliate a tree in a very short time, and can be significant agricultural pests on tropical farms. Because of their importance to forest ecosystems, to agriculture, and because they are common and conspicuous members of tropical ecosystems, leaf cutter ants have been studied quite a bit.

One of the most amazing things scientists have learned about leaf cutter ants is how they maintain the purity of the fungal colony growing in their nest. The ants have mutualistic bacteria that they tend and cultivate, and the bacteria themselves have anti-microbial properties. Simply put, to keep bad bacteria from taking over their hot, moist, resource-rich nests, the ants tend "good" bacteria.

Despite the long history of research on these ants, there are many questions remaining to be answered. For example, how do the ants choose which species of plant to forage on? You saw in the video that the ants will walk great distances through the forest, passing by thousand of individual plants and dozens and dozens of other species to feed on a particular tree. Then suddenly, after several days of feeding, they will switch to a new species or new individual. We know that the chemical defenses plants incorporate into their leaves deter ants (and many other species) from feeding, and new leaves are often less well defended. But in the green sea of leaves that is a tropical forest, how do ants, with individual brains the size of a grain of sand, survey what foods are available and decide on which to forage on?

**Examine the example picture and video below, while considering what you might be able to learn about leaf cutter ant foraging from a larger collection of similar images/videos**. Videos of the individual ants in the still photos were taken that captured that particular ant running a course of known length. We can measure any 1-D or 2-D dimension of the ants morphology (e.g., body length, leg length, head area) or of the leaves they carry (e.g., leaf area). We could count the number of ants moving in each direction, or we could note how many are carrying something vs. how many are empty-handed. We could measure the ants running speed. What hypotheses can we address with this kind of data?



[**Video example**](http://www.radford.edu/~jmwojdak/imagesets/0010339/0010339.m4v)

# Image Analysis

We are going to use image analysis to help us learn about these amazing leaf cutter ants. **Image analysis**, in the most general sense, just means obtaining information from images. In modern scientific parlance, image analysis refers to using computer software to collect data from images. The images may be obtained from satellites, cameras, imaging technology like CAT or MRI scans, or microscopes. The data may be presence/absence, counts, or measurements of lengths, volumes, or areas. Image analysis is thus a very general tool used by scientists in many disciplines.

Because image analysis is so useful, the National Institutes of Health (NIH) funded the development of free image analysis software, now called [ImageJ](http://imagej.nih.gov/ij/). ImageJ is very powerful, and luckily for us, very simple to use.

The logic of measuring from images is simple - the computer can count pixels in an image, and if we know how many pixels equates to a centimeter, for example, we can measure things in centimeters. This requires us to either include a scale bar in our image, or to know the precise dimensions of the image captured by the camera. Here we will use scale bars:

If the red line in the picture above is 140 pixels long, for example, we would know there are 140 pixels per centimeter, and thus if we were to measure an ant's total length to be 280 pixels, how long would it be in centimeters?

Yep, 2 cm. Nicely done.

# Getting familiar with image analysis in ImageJ

So let's start analyzing some images! As a tutorial for how to use Image J, you can follow the instructions below, and/or watch these videos: [basics](http://www.radford.edu/jmwojdak/ImageJmeasuringlength.mp4), [labeling](http://www.radford.edu/jmwojdak/ImageJlabeling.mp4), [exporting data](http://www.radford.edu/jmwojdak/ImageJExportingdata.mp4).

1. Launch ImageJ. It may be installed in your computer lab. Alternatively, download and install a copy on your local computer via <http://rsb.info.nih.gov/ij/>. {Or, you can run ImageJ in your web browser (without installing on computer lab or personal computers) by using the site <https://qubeshub.org/tools/imagej>. If you aren't a qubeshub member, you will be asked to create an account.}

2. Download [this picture](http://www.radford.edu/jmwojdak/images/_0010376.jpg) to start with. Save it on your local computer. [NOTE: file management will be a little different if you are running ImageJ through QUBESHub.]

3. In ImageJ, go to "*File"*, "*Open"* (menu selections and typed commands will be indicated *with italics* throughout the lab), then select the image you just downloaded.

4. Let's learn to set the scale for an image; we need to tell the computer how big things are in the image. Choose the straight line tool from the menu bar.

a. Use "*Control* *+"* or "*Control* *-"* to zoom in and out.

b. Zoom in on the scale bar, and carefully draw a line that is exactly 2 centimeters. To

draw a line, click the left mouse button, drag the line, then release the mouse button. If you screw up, just try again.

c. Now, select "*Analyze"*, "*Set Scale"*... then type in "2" in the "*known distance"*

box, and "cm" in the "*unit of length"* box. Hit "OK".

d. Select "*Analyze"*, "*Set measurements"*, and unclick all the boxes. Here you

can set what kind of measurements you want the software to calculate for

you, but for now we just want length which it does automatically.

5. Success! Now let's measure! Zoom in on an ant, draw a line along its body, then hit "*m*" for "*Measure"*. A window should pop-up, with a length and the angle of the line (presumably you don't care about the angle, but in some applications you might). Each time you draw a line and hit "*m*" a new data point will be added to this worksheet. It is easy to cut and paste these data into MS Excel or other spreadsheet/statistical software for analysis or graphing.

a. A good way to test that you've successfully set the scale for the image

is to draw another line of known length along the ruler... say 3cm this time, and hit "*m*". It should report the length of 3 (or very close to it).

6. By choosing other tools (ellipse, polygons, etc.) and changing the "*Set Measurements"* setting, you can measure all kinds of features of the images. Try measuring the area of a leaf with the polygon tool, and remember to add *Area* to the "*Set Measurements"* settings.

7. If the image is too bright/dark or otherwise needs adjustments, go to "*Image"*, "*Adjust"*, and "*Brightness/Contrast"* or whichever parameter might help you. These adjustments can be very helpful to make dark ant legs show up against the dark background.

# Hypothesis formation

Now, let's get back to our friends the leaf-cutter ants... What kinds of hypotheses can we test with data we can obtain from these images? There are LOTS of possibilities. To help you out, let's constrain ourselves to data that can be measured on a numerical scale: like # of ants passing by per minute, walking speed, body size, leg length, or leaf area. Avoid categorical variables like worker caste (soldier vs. worker), time of day (morning, noon, evening), or the direction of travel of ants (left/right).

So, pick a numerical variable that you can measure and that you are interested in understanding... then choose another numerical variable (that you could measure) that *you think* may in part determine or influence the first.

Why this constraint? First, we impose this constraint just to help you narrow down what you'd like to study. Second, choosing to look at the relationship between two numerical variables constrains the type of statistical analyses we will need to a single approach, **simple linear regression**. Ordinarily in research, we would follow our biological interests and use whatever statistical tools were needed.

Before you begin collecting data, consider:

* How many individuals will you need to measure?
* Given the structure of the data (i.e. ants photographed during three periods during the day, more ants than you can measure during the lab period), how will you select which ants to measure?
* How will you keep track of the data?
* If you are working in a group, how will you ensure that different people don't measure the same animals?

[**Here**](http://www.radford.edu/~jmwojdak/imagesets/)are all the images and videos, plus some information about how these images were collected ([**metadata**](http://www.radford.edu/jmwojdak/ia8.html)). Start with the metadata so you know what the heck you are looking at.

Depending on your operating system/software, some videos may not open in ImageJ via the "*File*", "*Open*" commands... in that case, try "*Import*", "*Using Quick Time*", then click "*Use Virtual Stack*".

Many instructors may want groups to briefly describe their plan before moving on - pay attention to what your instructor says to do here. Now, go collect that data!

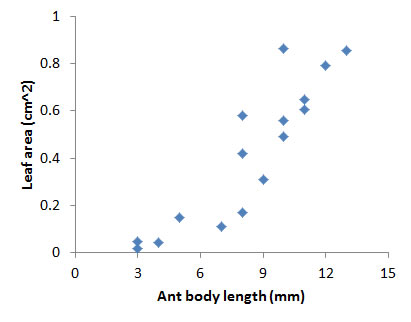
# Introduction to Linear Regression

Excellent. You have *data*! To a scientist, having new data is like having an unwrapped birthday present... so much excitement, so much potential! So, did we get a new toy or a sweater?

**Inferential statistics** are used to make inferences - to draw conclusions - from data. Scientists use inferential statistics to make objective decisions about what data tell us, rather than relying solely on their own opinion. That is, well-implemented and interpreted statistics provide evidence for our conclusions that others can evaluate.

As described before, we constrained the type of data you collected to that which could be analyzed with **simple linear regression**, or what is more commonly called simply, **regression**. Regression is a technique that tries to identify a (typically) linear relationship between two numerical variables. That is, as we change the value of one variable, does the value of the other variable change in a predictable way?

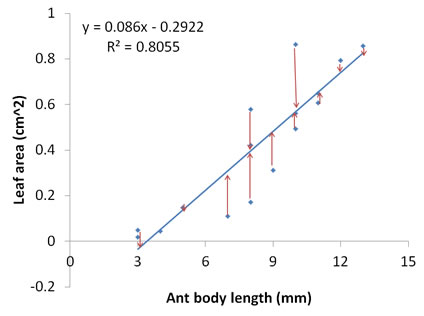
Ordinarily, we have some inkling that one variable responds to the other - the **dependent** variable responds to changes in the **independent** variable. In experiments we might vary the independent variable and look for changes in the dependent variable. Or, in an observational study, we may think, for instance, that ant's body size determines how big a leaf they can carry. By convention we plot the independent variable on the x-axis and the dependent variable on the y-axis.



Look at the data plotted to the left. Here someone measured both the body length of 16 individual ants, and the area of the leaves they were carrying. Do you see a pattern here? How would you describe it in words?

We will leave the details of calculation of linear regression to another day, but it is an easy process to understand visually. The computer will try to fit a straight line to the data that minimizes the (sum of the squared) vertical distances between the points and the line (see graph below).

This is sometimes called the **line of best fit**, or the **regression line**. Imagine, just for comparison, a line on the graph below with a flat (i.e. zero) slope... the vertical distances from each point to the line would be much larger... it wouldn't fit the data particularly well. The line of best fit... well... is the line that fits best. :-)



Regression analysis gives us an equation for the line of best fit, with a slope and a y-intercept. Alternatively (and less precisely) we could estimate that equation visually. Remember ***y = mx + b***, the equation for a straight line? Do you remember estimating the slope (rise over run) and intercept (where the line will cross the y-axis)? That has to bring back memories... your first algebra class? Middle school or high school... hormones... Good times, good times...

The magnitude of the slope is informative - it tells us how much the y-axis variable changes with a unit change in the x-axis variable (again, just rise over run). Moreover, we can use the equation to make predictions: in the graph above, if an ant's body length is 10mm (x=10), then we would expect it to be carrying a leaf with area (y) of 0.086\*(10) - 0.2922, or 0.57 cm2.

Regression analysis also gives us a value called **R2**, **R squared**. This tells us how much of the variation in the y-axis variable's values is accounted for by the variation in the x-axis variable's values.

One way to think about it is this: if you could tell *exactly* what the y-axis value would be if you knew the x-axis value, there is a perfect relationship between the two, and R2 = 1. All the data points would lie exactly on the line of best fit. And you could explain ALL of the variation in the y-axis variable by knowing the value of the x-axis variable.

Alternatively, if you would have no clue what the y-axis value would be given a particular x-axis value, the two variables are unrelated and R2=0. Very roughly speaking, R2 tells us about the strength of the relationship between the two variables.

Well, what about an R2 of 0.3, or 0.8... what do those mean? There is no simple answer. In physics or chemistry, when there is a single very strong cause or mechanism controlling the outcome, we typically expect very high R2 values... like when examining the concentration of a chemical versus the absorbance in a spectrophotometer. More chemical equals more absorbance, and very little else controls or contributes to absorbance. In contrast, in ecology, most of the processes we are interested in are controlled or influenced by many factors; the growth of a tree might respond to average climate, recent weather, local soils, local plant competition, and many other factors. Now, if we were to regress tree growth versus just one of those factors, we wouldn't really expect it to explain all of the variation in growth. Realistically it will only explain a small part. So, there is a bit of an art to interpreting R2 values... but for tree growth explaining 30% of the variation (R2 = 0.30) might be a real achievement!

Lastly, regression analysis gives us a **p-value**. A p-value from a regression analysis, strictly speaking, is the probability that if we repeated the study again and again, we would observe a relationship as strong or stronger than the one we did, assuming the null hypothesis (that there is no relationship between the two variables) were true. A little convoluted, yes, but the logic is this - if the p-value is very low, we are unlikely to have observed data like ours by chance, given two variables that are in fact unrelated. It is still possible, but rather unlikely. Thus, if the null hypothesis seems unlikely to be true, we can be confident that there is actually a relationship between the variables we measured. If the p-value is high, we have little confidence that the variables are actually related. By convention, p-values less than 0.05 (or 5% chance) are taken as evidence for a "significant" relationship - or, in other words, we are sufficiently confident to conclude there is a relationship. P-values larger than 0.05 suggest we should not be confident that a real relationship exists between the variables.

So, if the p-value for a linear regression analysis of the above data is p=0.001, and the R2 is 0.80, what would you conclude about these data?

Lastly, the most often regurgitated axiom of statistics is that "correlation doesn't imply causation" - which means that simply because we see a relationship between two variables doesn't mean we have demonstrated that one factor *causes* the other. This is absolutely true. However, a strong relationship is still *consistent* with a cause-effect relationship.

Need more background? Watch these videos: introduction to regression [1](http://www.radford.edu/jmwojdak/regression.1.mp4) and [2](http://www.radford.edu/jmwojdak/regression.2.mp4).

# Conducting regression analysis

If you have access to the statistical software JMP, follow [these instructions](http://www.radford.edu/jmwojdak/regression-instructions.for.JMP.doc).

If you have access to MS Excel, follow [these instructions](http://www.radford.edu/jmwojdak/regression.in.MSExcel.pptx).

# Interpreting your analysis: Post-lab assignment

Awesome! You started with a hypothesis. You used image analysis to collect data. You analyzed your data with linear regression. The only task left is to draw some inferences and describe what you've learned.

Each person should *independently* complete the [regression practice sheets](http://www.radford.edu/jmwojdak/Linear%20Regression.doc), and turn those in. Then, in small groups (~four people) you should write a short lab report:

1. Describe your hypothesis, and your rationale (why, biologically, did you have that hypothesis?).
2. Describe your results in words, focusing on the biology, using graphs and the regression analysis to provide support for your inferences.

* Ideally your sentences **should be** biology-focused:

"Larger ants, as measured by total length, carried larger leaves than did smaller workers (linear regression, p<0.001, R2=0.56; Figure 1)."

* + Describing the direction or strength of a relationship is better than just saying "there was a relationship".
  + Moreover, referencing a figure showing the data is always more informative than just describing it with words.
* Ideally, again, your sentences **should** **NOT BE** statistics-focused, like this one is:

"There was a significant linear regression between ant size and leaf area (p<0.001, R2=0.56; Figure 1)."

* + Notice this sentence doesn't even tell us if bigger ants carried bigger leaves!

1. Draw a brief conclusion. How does your data compare with what you expected? Delve into your hypothesis and your results, explaining what you think is really happening in nature.

* Super extra bonus points for using information from other published articles (from the bibliography here or from your own literature search) to better understand your results.

# Bibliography

Here are some research articles related to leaf cutter ant foraging, allometry, and movement.

[Moll K, Federle W, Roces F. 2012. The energetics of running stability: costs of transport in grass-cutting ants depend on fragment shape. Journal of Experimental Biology 215:161-168.](http://jeb.biologists.org/content/jexbio/215/1/161.full.pdf)

[Roces F, Hölldobler B. 1994. Leaf density and a trade-off between load-size selection and recruitment behavior in the ant *Atta cephalotes*. Oecologia 97: 1-8.](http://www.jstor.org/stable/4220581)

[Rudolph SG, Loudon C. 1986. Load size selection by foraging leaf-cutter ants (*Atta cephalotes*). Ecological Entomology 11:401-410.](http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2311.1986.tb00319.x/abstract;jsessionid=6EC5988385D4AAC0B6EE43B51D9CC711.f04t02)

[Wirth R, Beyschlag W, Ryel RJ, Holldobler B. 1997. Annual foraging of the leaf-cutting ant *Atta colombica* in a semideciduous rain forest in Panama. Journal of Tropical Ecology 13:741-757.](http://www.jstor.org/stable/2559925?origin=JSTOR-pdf)

1. This lab was created by JM Wojdak at Radford University. [↑](#footnote-ref-1)