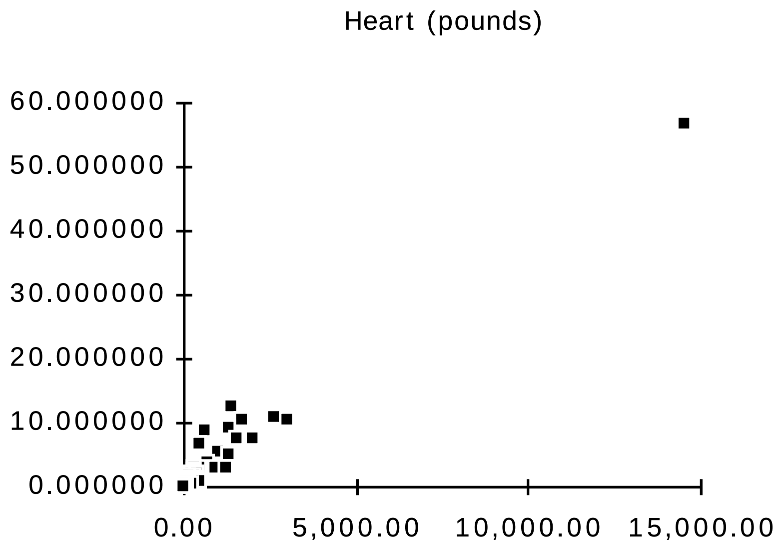
Log Log

“Log log graphs”, as they are called are simple, in principle, but they have a strange reputation — among those people with whom such things have any reputation at all.

I think what happens is that people who use statistical tools, and whose control over the techniques they depend on is tenuous, suddenly realize that things have gotten beyond them with log log analysis. It is too many steps beyond control, and they get scared. The discomfort is expressed as skepticism, but there is actually an event that triggers the expression of: It seems to be a common experience that many kinds of things look linear — when you are looking at them on a log log graph. So people, the same people with whom these things have any reputation at all, say “everything” looks linear on a log log graph, and back off. They dismiss what they see — the idea being that if everything looks linear on a log log graph, then nothing is learned in any particular case.

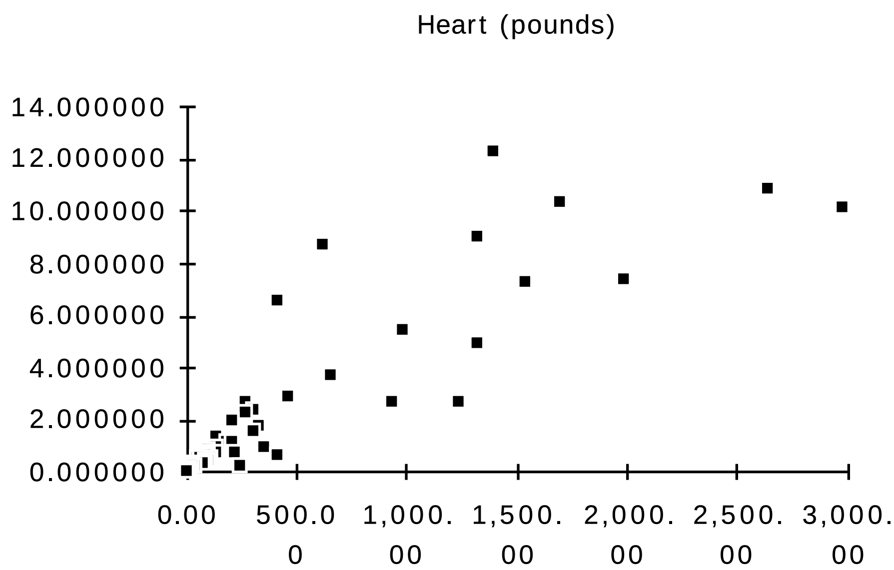
Such stuff somehow passes for sophistication, but really it is a kind of belief in magic. It comes from statistics as magic, that then gets out of control. But if we stay calm and rational, if we use fairly simple math — and believe in it as the tool by which to interpret what you’ve found, there is nothing out of control when logs begin to pop up on both sides of an equation. Would that it were true that “everything” looks linear on a log log graph. That has not been my experience. And if many things do look linear with this analysis then there is something to be learned here about nature.

First let’s take a look: Here, for example is Heart Weight (as “Y” — shown vertically) and Body Weight. Thedata are from ___ describing the average weights of bodies and of the various organs of vertebrates (a very peculiar collection of data). You might thing about lines on such a graph, and then think about slopes — looking for the weight of the heart as a fraction of the weight of the body, the larger the body the larger the heart. That’s what I expect. But all bets are off when you look at the graph of these weights. Looking at the graph, and speaking non-technically, it’s a mess.



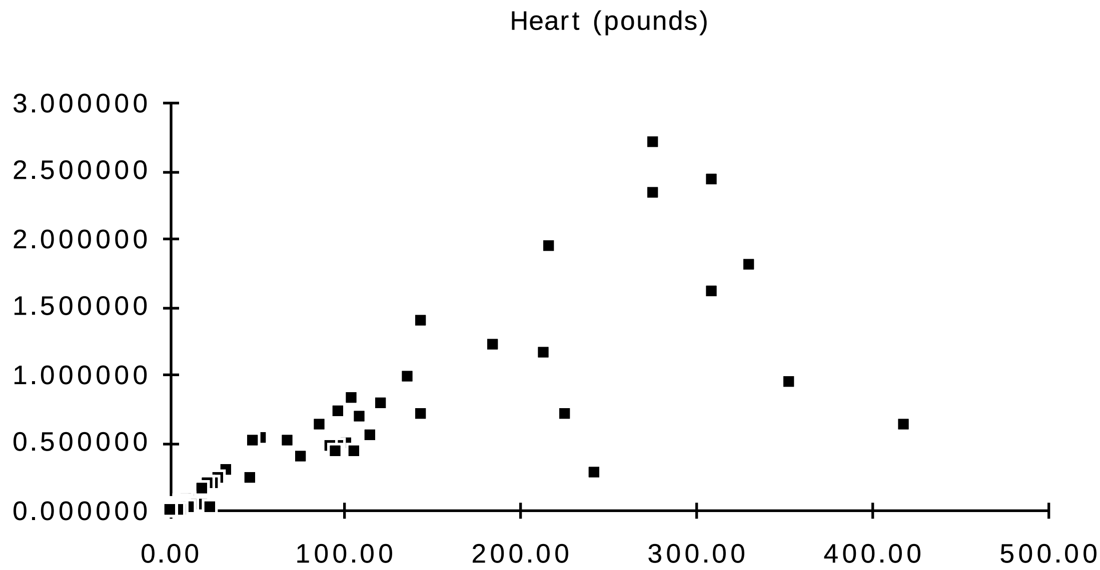
Now, what *you* should do if somebody offers you such a graph, as I have offered it to you is either a), walk away because this person doesn't know how to analyze data, or b) gently walk toward such a person to explain that data analysis begins "at the beginning". And the beginning is one variable analyses — always accompanied by good labels (not a dot at the upper left, but "elephant" (In fact, I would go so far as to say that labels are so important with such things that you will get further faster with these data doing them by hand, given the unlikeliness of getting current software to label properly, by machine.) And there you will discover, or guess, or hypothesize, because you've dealt with such things before, that the intelligent start, for this, would have been logs, not pounds.

But here's what would be done by the pseudo sophisticated. Actually, this person is starting out wrong and then patching and filling — trying to make the patching and filling look like sophistication. Patching and filling, we observe that that point ("Elephant", were it labelled) is an outlier — exclude it from the analysis (or analyze it separately). O.K., excluding the "outlier"



That looks better, until you realize that seeing about 15 points clearly may look better, but that "15" is 15 out of about 150 points, and most of the stuff is still down there at the lower left.

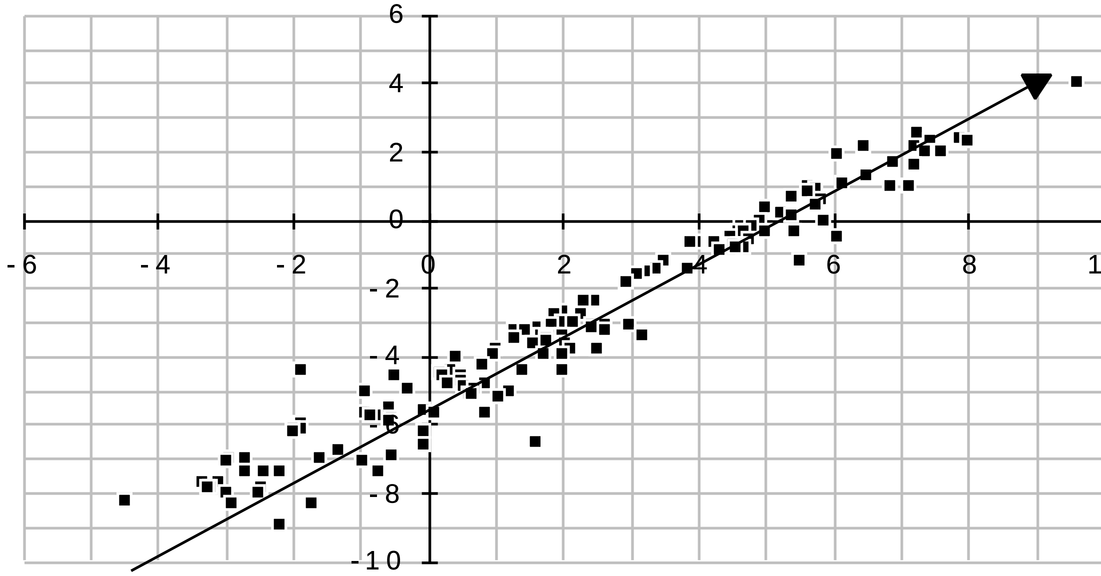
How about excluding just those 15 or so points that *are* the most visible, concentrating on the main body of the data — in effect declaring these 15 as outliers too.



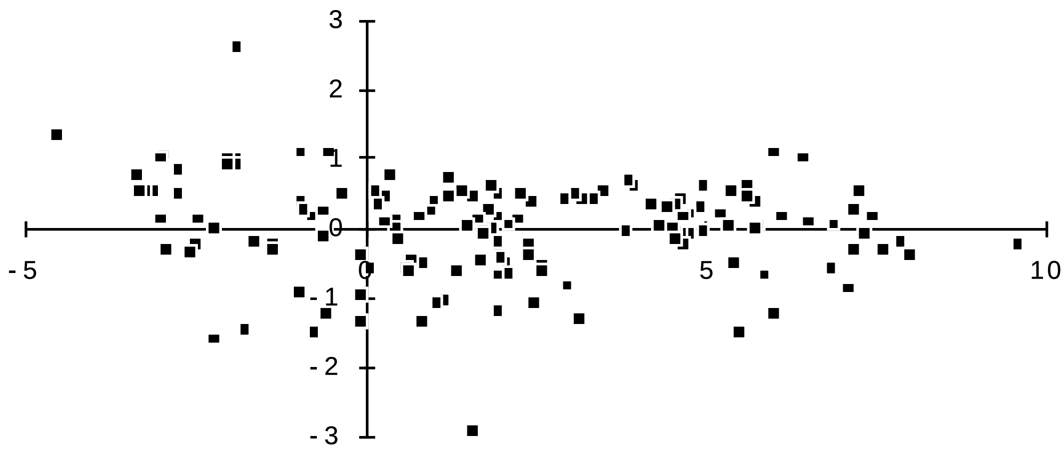
Well — better. But I'd venture to say that if this were not already the third graph in the sequence, if we were starting fresh with it, then we would observe that the 20 or so points that are visible at the upper right seem to be on a different scale from the bulk of the data, about 120 point, at the lower left. And we would proceed to remove them from the data, just as we have already removed 16 points inorder to get at the heart of the data.

When you find yourself getting into this kind of dissatisfying loop, cutting out data, cutting out more data, and still not really solving the problem, then it is time to drop back and think. And thinking, or guessing, or starting with the logs in the first place, gets this (using logs base e).

Ln Heart Wt



observed - expected



There you see a line, or at least a sort-of-linear cloud. And, you can understand why someone starting with the first figure (which they should not have) would be in awe of this log log picture derived from the same data: Junk has become orderly and, if you have let the mathematics get the better of you, then the conversion of the chaos in the first figure into the order of the second figure, might seem like magic. Magic had nothing to do with it.

Below it, observing the line and then looking at residuals, I've computed residuals using the numbers

$$\ln(Heart) = .9845 \ \ln(Body) - 5.15925$$

So, if we're so smart, what does this log log equation mean? I find out by trusting the math and using it to decode the meaning. First exponentiate

$$e^{\ln(Heart)} = e^{.9845 \ \ln(Body) - 5.15925}$$

$$e^{\ln(Heart)} = e^{(.9845 \ \ln(Body) - 5.15925)}$$

and then simplify

$$Heart = e^{-5.15925} \ Body^{.9845}$$

$$Heart = e^{(-5.15925)} e^{(.9845 \ \ln(Body))}$$

and

$$Heart = .005746 \bullet Weight^{.9845}$$

$$Heart = .005746 \ Weight^{.9845}$$

It says: The weight of the heart, on the left, is proportional to the .98th power of the weight of the organism. Observing that .9845 is close to one, the equation says that the weight of the heart is proportional to the weight of the body.

$$Heart \approx .005746 \ Body$$

So the equation implies that the weight of the heart is to the weight of the body as .005746.

$$\frac{Heart}{Body} \approx .005746$$

That is, weight of the heart is approximately one half of one percent of the weight of the body (0.57 %). That is the decoding of the slope and the intercept of this log log linear relation.

Thinking about this relation, I find the relation surprising: It says that the weight of the heart is *directly proportional* to the weight of the body. I'm not sure exactly what I expected (so much for falsifiable hypotheses), but I was thinking that big animals tend to be warm blooded, that should make a difference in demands on heart. Or, there should be some efficiencies of size when a heart muscle has to push blood around. Too bad: what differences there are in the residuals, they are not a function of total weight.

How big are the residuals?

You can read that right off of the graph of the residuals: The observed values are within plus or minus one of the predicted values. (Imagine trying to estimate the size of the residuals using the first graph, without logs). And errors of plus or minus one in logs, base e, correspond to factors of 2.7 above or below the predicted weight of the heart. So

The weight of the heart is proportional the weight of the body, averaging about one half of one percent of body weight. However, there is a large variation around that average, amounting to a factor of nearly three in either direction, meaning that heart weight typically valls within a range of 0.2 percent to 1.6 percent of body weight (multiplying and dividing .005746 by 2.7).