Lab 10 – self assessment #1

This self assessment explores the different consequences of log transforming X, log transforming Y, and both. Allometry is an sub-area of morphology that examines the relationship between morphological characters. For example, allometry describes why human baby’s faces are so large relative to their body size. The activities in this self assessment are based on tarsus and wing measurements for a species of bee. 300 bees were measured. The data in bees.csv are their tarsus length and their wing length. Allometric relationships are usually described by power-law models, e.g.

(1)

The goal is to estimate *b*. We will (eventually) fit this model, but we will also fit and interpret 3 other models. In linear regression terms, those models are:

(2)

(3)

(4)

(5)

a) Which of these 4 linear regression models provides estimates of *a* and *b* in the power-law model, equation (1)?

b) Plot Y vs X for each of the 4 linear regression models. Based on these plots, which model is the most appropriate one for these data? Briefly explain why you decided on that model.

c) Fit each linear regression, then plot residuals (Y axis) vs predicted values (X axis). Based on these plots, which model is the most appropriate one for these data?

d) For the 3 models that you “reject” as not appropriate, briefly explain what in its residual plot led you to decide that it wasn’t appropriate.

The next 4 questions are exercises in interpreting the slope coefficient in each sort of model. For each question, ignore whether or not that model is appropriate. Answer each question as if that model were appropriate**. For simplicity, you may use causal language even though these data are observational.** For example, answers might be “Doubling tarsus length increases wing length by <number>” or “Increasing tarsus length by 1 multiplies median wing length by <number>”.

Both tarsus and wing lengths are in millimeters (mm). I made up these data so the values may not be biomechanically realistic.

e) Model (2): Report the estimated slope and write a short (1 sentence) conclusion about how tarsus length (X) is related to wing length (Y). Make sure to include units where appropriate.

f) Model (3): Report the estimated slope and write a short (1 sentence) conclusion about how tarsus length (X) is related to wing length (Y). Make sure to include units where appropriate.

g) Model (4): Report the estimated slope and write a short (1 sentence) conclusion about how tarsus length (X) is related to wing length (Y). Make sure to include units where appropriate.

h) Model (5): Report the estimated slope and write a short (1 sentence) conclusion about how tarsus length (X) is related to wing length (Y). Make sure to include units where appropriate.

i) Model (5): Write a more carefully worded conclusion that avoids the causal language used in my example conclusions.

a) Model (5).

Note: Model (5) is obtained by log transforming both sides of model (1).

b) Model (5). It is the closest to a straight line and the vertical spread of Y values looks about the same across the range of X.

Here are my plots:



c) Model (5). No explanation requested (that’s the next question).

Note: My plots are on the next page.

d) Model (2): I see lack of fit and some unequal variance

Model (3): I see lack of fit.

It looks like unequal variance, but that’s a graphical artifact interpreting vertical spread when lines are steep

Model (4): I see lack of fit and unequal variance.

If you didn’t choose model (5) in question c, you probably focus on unequal variance.

That’s overinterpreting the patterns in this plot.

These are simulated data so I know the truth – that is constant error variance for log wing.

Note: So how would I go about choosing a model based on these plots.

What I described in lecture:

use the linear regression version of the power-law model (a biology-based choice)

OR: model (2) shows unequal variance, try model (3), but that’s hard to see that it has equal variances

model (3) has lack of fit so try model (5)

OR: model (2) shows lack of fit, try model (4), shows lack of fit and unequal variance so try model (5)

Here are my residual vs predicted value plots:



e) estimate = 0.129. Increasing tarsus length by 1 mm increases wing length by an average of 0.126 mm.

OR: … increases mean wing length by 0.126 mm.

f) estimate = 0.066. Increasing tarsus length by 1 mm increases median wing length by 6.8% (or multiplies median wing length by 1.068). Note: exp(0.066) = 1.068, no units needed for wing length, because all conclusions are relative.

g) estimate = 0.742. Doubling tarsus length increases wing length by an average of 0.514 mm.

note: 0.742\*log(2) = 0.742\*0.693 = 0.514

h) estimate = 0.402. Doubling tarsus length increases median wing length by 32% (or multiplies median wing length by 1.32). Note: The calculations:

doubling tarsus length increases log wing length by 0.693\*0.402 = 0.279

exp(0.279) = 1.32

i) Here’s one way: When you compare two bees, one with double the tarsus length than the other, the bee with the larger tarsus length has a wing length that is 32% larger than the other.

Note: The non-causal conclusion is a descriptive conclusion. It describes changes between two groups of bees. Its language never implies “changing X results in <this change in Y>”; that implies that all the differences in Y are caused by the differences in X. Appropriate for randomized experiments where everything else is the same except for the experimentally manipulated X. Not appropriate for observational studies. Conclusions from observational studies need to focus on descriptions, i.e. comparisons between two groups of subjects.