SMA Regressions

Background: Productivity (P: g dry mass / m2 / year) and biomass (B: g dry mass / m2) data have been collected and compared among different ecosystems for many years. In principle, P should be predicted from B, and *vice versa* because these are two measures of the same entities (like your height and weight).

So far this semester you have used ordinary least squares (OLS) regressions, *which only predict Y from X*. Today we compare those to standardized major axis (SMA) regressions for P/B data of grasslands (Warton et al. 2012). Because this is allometry, we should also evaluate log-transformations of P and of B data (Kerkhoff & Enquist 2009). Four combinations exist: P/B, P/logB, logP/B, and logP/logB. We evaluate alternative models and:
- test regression assumptions – we should doubt models that do not meet assumptions. Use QQ normality plots and residual plots as before.
- use AICc of OLS and SMA models *for the same response variable* – it is unfair to compare simple P (e.g., 10 g/m^2/y) to logP (e.g., log(10 = 1).

Our Mission:
A) Can a strong relationship be found between P and B? What is the best equation?
B) Does the logP:logB relationship fit the ¾ slope prediction of Metabolic Theory? (for more detail read Brown et al. 2004)
C) And what is the effect, if any, of grazing on this relationship? Grazing may increase slope if plants respond with more growth, but may also reduce slope by removing growth before it is measured.

Preliminary Steps
1. Install and load the `smatr` package in R. Here we use lm to make OLS models and sma in smatr to make SMA.
2. Load the `bbmle` and `ggplot2` packages
3. Import and attach the `PB.txt` data set from the course web page.
4. First let's just look at the data pattern. Plot P as a function of B for each subset:
   ```
   qplot(B,P,colour=Grazed,pch=Grazed)
   ```

   Three questions:
   - Does a strong linear relationship seem apparent?
   - Is there a wedge-shaped scatter that may be fixed with log-transforms?
   - Does it look like grazed and ungrazed P:B relationships differ? [we revisit this later]

5. Make log-10 transforms (log-10 vs. log-e is arbitrary, but log-10 is easier to interpret):
   ```
   logP <- log10(P)
   logB <- log10(B)
   ```
6. Now re-plot using the log-transformed versions and answer the same 3 questions above.
7. Let's compare residuals of the untransformed and transformed versions.
   ```
   untrans <- lm(P~B)
   plot(untrans) # ain't that nice?
   ```

and then repeat for the transformed version:

```
trans <- lm(logP~logB)
plot(trans)
```

- Does log-log transforming make assumptions better for regressions? If so, expect log-log assumption tests and models to work better below.

Models and Model Comparison: We compare models with AICc, _BUT_ _we cannot fairly compare a model of logP to a model of simple P_. So we compare OLS and SMA models in two sets:

- simple P, using [P ~ B] vs. [P ~ logB], each with OLS and SMA models (4 models total)
- logP, using [logP ~ B] vs. [logP ~ logB], each with OLS and SMA models (4 models total)

8. Use this code to compute one OLS and one SMA regression for one model [P ~ B] – comments after # tell you what these do.

```
PBols <- lm(P~B) #run lm on P as function of B
plot(PBols)       #run diagnostic plots

summary(PBols)    #lm output

PBsma <- sma(P~B, method="SMA")  #now run sma model
resPBsma <- residuals(PBsma)     #report resids of sma
model

plot(resPBsma~B)  #this and next line plot resids
abline(a=0,b=0, lty=2)
summary(PBsma)    #sma output
```

9. Use the above as a template for running similar models of the following
   ○ P ~ logB
   ○ logP ~ B
   ○ logP ~ logB

At the end you should have two sets (OLS and SMA), each containing four models (P~B, P~logB, logP~B, and logP~logB). Total = 8 models.

Which model best:
- shows residuals most evenly scattered along the length,
- normal (i.e., balanced and centered around regression line)
- without some other pattern (e.g., an arch)?

10. Finally, compute AICc tables, with weights, using `AICctab` in the `bbmle` package – just like you have used in prior labs. You should compute AICc tables in two sets: one for models of P, and another for models of logP.

- Which model is most plausible in each set?
- Of the two "winners," which model would you use to estimate P, based on assumptions and R^2? Notice that we only now use R^2 !

Grazing Effects: Now use your "best" model to address the effect of grazing: Do grazed and ungrazed P:B relationships differ?
1. repeat your best model (assumed here to be sma of log-log) for just the ungrazed subset by specifying *inside the model parentheses*:

```
subset = PB$Grazed=="n"
```

2. Now enter the slope (e.g., 1) and intercept (e.g., 0.1) values of the ungrazed model in appropriate places in the slope.test and elev.test parts of the sma command. For example:

```
summary(sma(logP~logB, method="SMA", slope.test=1,
elev.test=0.1))
```

- Does grazing change P:B scaling in grasslands?

Metabolic Theory: Now we test to see if the slope of our best log-log model = ¾ as predicted by Metabolic Theory (Brown et al. 2004). Short version: fractal metabolic networks of plants should cause this slope. We specify a slope.test=0.75 in the logP~logB model:

```
summary(sma(logP~logB, method="SMA", slope.test=0.75))
```

- Do the data support the Metabolic Theory of Ecology?

References
Brown, JH, JF Gillooly, AP Allen, VM Savage, GB West. 2004. Toward a metabolic theory of ecology. *Ecology* 85:1771–1789.
Kerkhoff, AJ & BJ Enquist. 2009. Multiplicative by nature: Why logarithmic transformation is necessary in allometry. *J. Theoretical Biology* 257:519–521.
Warton, DI, RA Duursma, DS Falster, S Taskinen. 2012. SMATR 3 – an R package for estimation and inference about allometric lines. *Methods in Ecology and Evolution* 3:257–259.