I. Reflection on lab 5 and the week prior
   1. Approaching the end of Part II.
   2. Expect errors. Think through them and work to correct the errors. In lab, be sure to ask the demonstrators and me for help – but also observe how we correct the problem (usually!).
   3. Midterm project directions now posted (due April 9).
   4. There’s an interesting discussion about open data going on in PLOS right now (this is one of the world’s largest scientific publishers).

II. Linear Models
   1. What do we do if we have a categorical variable and a continuous variable? Both of which we think affects the response? Consider limpets.
      a. H0 = Eggs are NOT affected by density or season
      b. H1 = Eggs vary with density but not season
      c. H1 = Eggs vary with density and season additively
      d. H1 = Eggs vary with density and season multiplicatively
   2. Review of ANOVA (seasons and limpets)
      a. Examine the model sum of squares (between groups, i.e. the differences between the group means and the grand mean).
      b. Residual sum of squares are the within group differences (i.e. differences between the points and the group mean).
      c. The mean square error is the SS divided by DF (i.e. variance).
      d. The F ratio is MS of the effect over SS of the residuals/ error.
      e. Note also that you can run a regression using categorical variables (e.g. season) – this is the same as ANOVA!
   3. Quick review of simple linear regression (density and limpets)
      a. As we learned last week, part of the linear regression analysis is an ANOVA which determines significance.
   4. Linear models: ANOVA $\leftrightarrow$ Regression $\leftrightarrow$ Linear Models
      a. lm(EGGS ~ DENSITY, data=Limpets)
      b. lm(EGGS ~ SEASON, data=Limpets)

III. ANCOVA (another linear model example)
   1. We can examine the interaction of two variables (season and density)
      a. lm(EGGS ~ DENSITY*SEASON, data=Limpets)
         i. Lines “cross”, i.e. there is an interaction
         ii. If no interaction term, remove the multiplicative term
      b. lm(EGGS ~ DENSITY+SEASON, data=Limpets)
      c. Evaluate the null and alternative hypotheses as appropriate
      d. Construct linear regression equations as needed.
         i. For an ANCOVA, the R output provides the slope and intercept for one group (ordered alphanumerically).
         ii. It then provides the difference between both groups such that each line can be constructed.
      e. Check homoscedasticity and normality assumptions with residual vs. fitted and QQ plots.
IV. Multiple linear regression
1. Now that we’ve seen the common threads between ANOVA and linear regressions we can scale up to multiple regression (in fact ANCOVA is just a specialized multiple regression).
2. Logic is the same – just more variables!
3. Instead of fitting lines now we’re fitting planes.
4. Do three variables (e.g. PREC, TEMVAR and PET) have an effect on a response variable (e.g. woody plant species richness)?
   a. The regression will tell you a coefficient of determination and which variables are significant.
   b. $R^2$ vs. Adjusted $R^2$: The former is the $R^2$ is traditionally defined whereas adjusted $R^2$ is adjusted for sample size.
5. Multiple regression can also include polynomial terms.
   a. Quadratic (and higher, e.g. cubic) terms introduce curvature into the lines essentially.
   b. Note that “I” in R is just an indicator to R to evaluate the terms within the parenthesis (e.g. I(PET^2)).

V. What to do with outliers
1. Transform the data
2. Use non-parametric methods
3. Remove outliers

VI. Steps in Environmental Data Analysis
1.) Import the data
2.) Plot the data (often with a histogram)
3.) Descriptive statistics (mean, variance)
4.) Test the statistical significance of any differences between the means or variances of populations.
5.) Check validity of the analysis (e.g. residuals)
EMBRACE ERROR