ANCOVAs

ANCOVAs are hybrids of regression and ANOVA; a mix of categorical and continuous predictors for a response variable. ANCOVAs assume the effect of the continuous variable is constant across the categories (i.e., slopes are parallel). You have already analyzed such data, but today we pay particular attention to two important details:

- the order in which you list categorical and continuous predictors matters **for aov (not lm)**.
- ANCOVA assumes the continuous factor does not interact with other treatments (i.e., its slope is the same across other treatments)

The helicopter experiment mixed a categorical treatment (folded / not folded) and continuous treatments (wing length, body width). We also had groups (a block effect) and a covariate (steps). Those are different, and represent a planned, and unplanned structure to the data collection, respectively.

1. Import and attach the ipomopsis data set from the course web site (you used it for homework #7).
   I assume below you name the imported file to be "data" and attach it.

2. Don't sweat assumptions today. Reminder: those are about *residuals*, which we address after you are done with this.

3. Plot fruit weights as a function of root diameter for grazed and ungrazed plants:

```
plot(Root, Fruit, pch=16, col=c("blue","red"))
abline(lm(Fruit[Grazing=="Grazed"] ~ Root[Grazing=='Grazed']), col='blue')
abline(lm(Fruit[Grazing=="Ungrazed"] ~ Root[Grazing=='Ungrazed']), col='red')
```

4. Does this look like it could be two *different* regressions? And are the two regressions about parallel; *a requirement for ANCOVA?*
5. How do you test if slopes are equivalent between Grazed and Ungrazed?
   (a) Option 1. Use dplyr to filter only Grazed data & run a regression. Ditto for Ungrazed. Compare slope coefficients and 1.96 X SE terms (i.e., 95% CIs) – are the slope coefficients similar?
   (b) Option 2. Use dplyr to filter only *Grazed* data & run a regression.
   (c) Then use the **car** package to do this for *Ungrazed*:

```
modelUG <- lm(Fruit ~ Root)
linearHypothesis(model, "x = ___") # where you fill in the slope from the first model into ___
```

Do you get the same answer?

If slopes are parallel, then you have a simple ANCOVA. But if slopes are not parallel, then an interaction is likely. In that case, this is not an ANCOVA but a more complex model.

6. If you are going to include both quantitative and categorical predictors in a model, which gets listed first? Try this:

```
model1 <- lm(Fruit ~ Root + Grazing)
model2 <- lm(Fruit ~ Grazing + Root)
```

Do you get different answers? **The order that variables are listed is important for aov, so list the covariate first, categorical treatment second.** Think of it this way: we level the playing field for a fair comparison among categories by first removing the signal due to the quantitative "covariate."

Does order matter for AICs? Try using the **bbmle** package to compare model1 & model2 – are they different?

Bottom Lines:

A) Analysis of covariance is a specific kind of model, where different groups have parallel effects (thus the grouping effect is one of elevation on the Y axis only). Non-parallel patterns between those groupings = more complex models with interactions.
B) Analysis of variance in general is *partially* useful but limited – it only shows overall effects exist (or not), and it depends on: p values; homogeneity of variance among groups and careful writing of the models.
C) We steer you toward regressions and AIC-based model comparisons the rest of the semester.