Homework assignment 4: regression

[brouwern@gmail.com](mailto:brouwern@gmail.com)

February 13, 2017

# Homework assignment 4: regression

The goal of this assignment is to independently go through the steps of building an analyzing regression models for the milk data.

**Workin in Pairs, do the following:**

## 1) Pick a response variable (y)

* fat
* protein.num
* sugar.num
* energy.num

## 2) Pick a numeric predictor (x)

* mass.female
* mass.litter
* repro.output
* relative gestation length

## 3) Pick a categorical (factor) predictor

* diet (ominviores & herbivores pooled)
* arid
* biome

If you use fat then use one of the predictors we haven't used yet (mass.litter, repro.output)

## 4) Build 5 models as in the "interactions" exercise

* null (1 intercept; flat line)
* categorical predictor only (2 intercepts; 2 flat lines)
* numeric predictor
* numeric + categorical predictor (additive model)
* numeric\*categorical (interaction model)

## 5) Compare your 5 models w/AICc

**Questions:**

* Which of your models model is the "best" according to typical AICc criteria?
  + ignore the concept of "model averaging"
* Which model is next best?

## 6) Compare models w/ANOVA

**Questions:**

* Compare the "interaction" model to the "additive model" using anova(); is the interaction term "significant"? at the convetional p < 0.05 level?
* Write the mathematical equation for the interaction model
  + eg, response = intercept + slope\*x + ...

## 7) Calculate R2 and F by hand

* Consult the "field guide" to anova() output
* Check against the anova() output

**QUESTIONS**

* For the "best" model according to AIC, Write out the calculations needed to calcualte R2 and F using your results (ie, obtain the numbers and give the final result)

## 8) Compare the best fitting model via AIC to what anova() tells you

**QUESTION**

* Does the ANOVA result match the AIC result?

## 9) Assess the residuals

* Make a 2 x 2 panel of the residuals

**QUESTION**

* Do any of the residual plots indicate a problem?
* What assumptions of linear regression might be violated?

## 10) Plot 4 of the models in a single panel

* Make a 4 x 4 plot of the models (except the null)
* Do the models seems to fit the data well?
* Are violations of the assumptions apparent in these plots?

# Write up

Produce a short write up in RMarkdown with the following elements

## 1) Written portion of write up:

Write answers to the above questions in paragraph form, kind of like in a methods section. Such as "we assess the residuals of our best model, which contained an interaction between ... and found evidence of non-costant variance in the residual vs. fitted plot. The R2 value for this best model was calcualted as xxx/yyy = zzzz and the F statistic as qqq/ttt = aaa."

## 2) Statistical ouput

* The summary() output
* The ICtab() output

This output does NOT need to be formatted.

## 3) Figures

* A 4-panel plot of the models (except the null)
* A 4-panel plot with the residuals

There does not need to be any special formatting.

## 4) Code

For this assignment you don't have to submit code, just the output. Using Rmarkdown this can be accomplished by

* Easiest: Rendering all you work to a Word file then deleting by hand anything not needed
* Almost as easy: you can include at the start of your chunk " {r, echo = F} " instead of just "{r} " and RMarkdown will not print the output. You can use this to keep the ouput hidden for the writeup. You will probably still have to delete things by hand.
* Harder, and not necessary: after you've done your work, create a new file that contains only the necessary code to load the data and plotting functions, create the graphs and produce the R output. Add appropriate labels. This isn't hard, just would be an extra step. I am increasingly using this approach when I finalize an analysis - creating a new, clean file that contains only the essentials and removes either completely or via echo = F anything not needed to get the point of the analysis.

# NOTES

## Transformations

**NOTE**

* My plotting function plot.ANCOVA() (from the file "fnxn\_plot\_ANCOVA.R") needs to have the transformations hardcoded!
* Be sure to do this!
* eg, "daty)"
* if you do this, it will cause problems: lm(log(y) ~ log(x))
* If you use plot.ANCOVA() and have a problem email me - I have not extensively tested this function

# Loading my plotting function plot.ANCOVA()

* Make sure the file is saved to you working directory
* use source() to load it

# Formatting

There does not need to be any special formatting of the statistical output or the graphs.