The End of the Beginning

First, Centering

Let's say you're going along blisfully with your multiple linear regression...

```
plankton <- read.csv("./data/planktonSummary.csv")
#
library(car)
alm <- lm(DIN ~ SAL*T, data=plankton)
```

First, Centering

Suddenly...

```
vif(alm)

#    SAL      T  SAL:T
#  4.421 27.413 33.299
```

```
cor(plankton$T, plankton$T*plankton$SAL)

# [1] 0.9325
```

Centering Before Nonlinear Transformation Reduces Variance Inflation

Suddenly...

```
cent <- function(x) x-mean(x)

plankton$int <- with(plankton, cent(SAL) * cent(T))

alm2 <- lm(DIN ~ SAL + T + int, data=plankton)
vif(alm2)

#   SAL     T   int
# 1.115 1.033 1.122
```

Interpretation Changes

```
#             Estimate Std. Error t value   Pr(>|t|)
# (Intercept) 299.5726    2.90911  102.98  0.000e+00
# SAL          -7.8375    0.09095  -86.17  0.000e+00
# T            -2.4891    0.09154  -27.19 1.415e-149
# int           0.4166    0.01576   26.43 4.274e-142
```

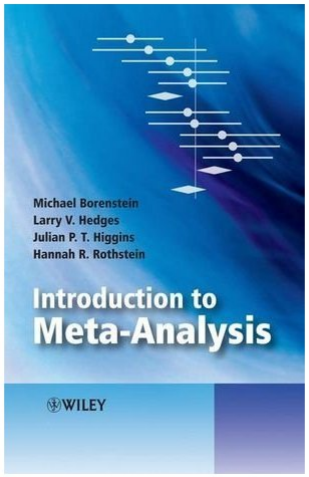
Additive coefficients are evaluated at mean level of each other.

Where to from Here?

Never Stop Reading

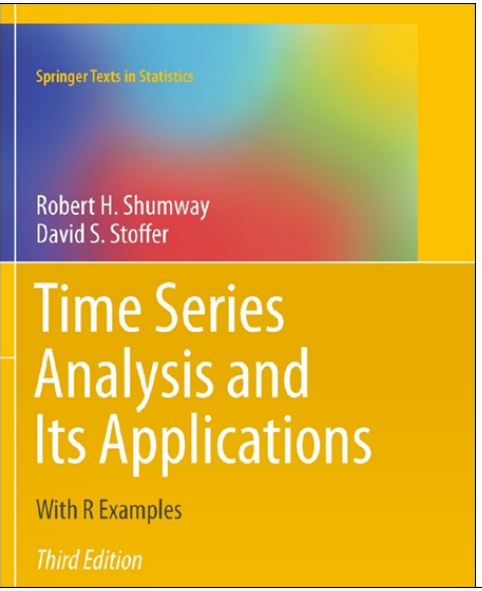
- `http://www.r-bloggers.com/`
- `http://andrewgelman.com/`
- `http://masi.cscs.lsa.umich.edu/~crshalizi/weblog/`
- *Methods in Ecology & Evolution*
- *Significance*

Meta-Analysis



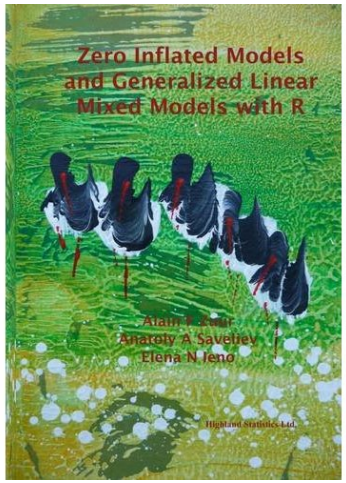
The analysis of results from previous studies - summarizes information to get a grand answer to big questions

Time & Space



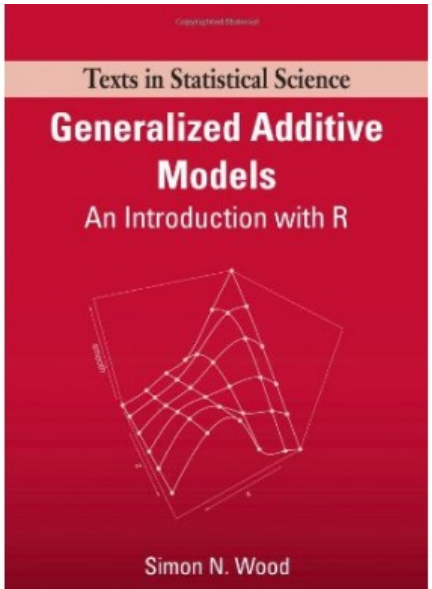
A whole different set of consideration in thinking about correlation structure in complex spatial and temporal landscapes

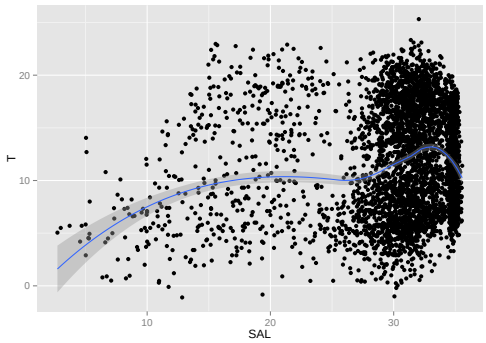
Zero Inflated or Censored Models



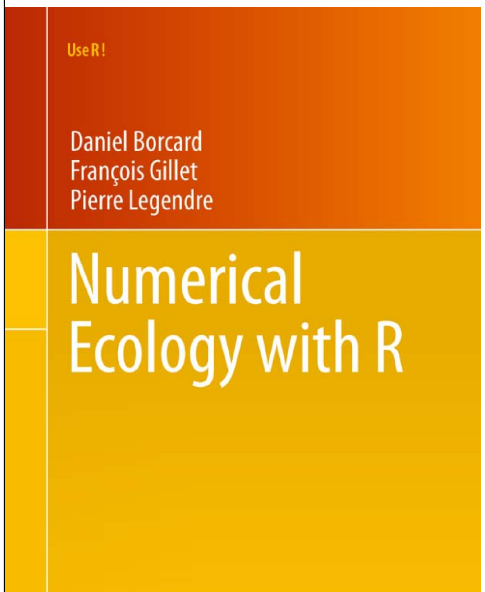
What if you have a LOT of zeroes...but otherwise things look Poisson, Negative Binomial, Normal, or more. What about censored data where an instrument only reads so high or low?

Generalized Additive Models



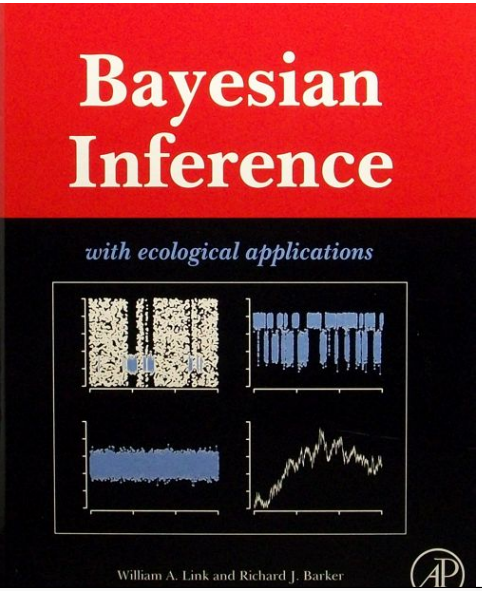


Multivariate Methods



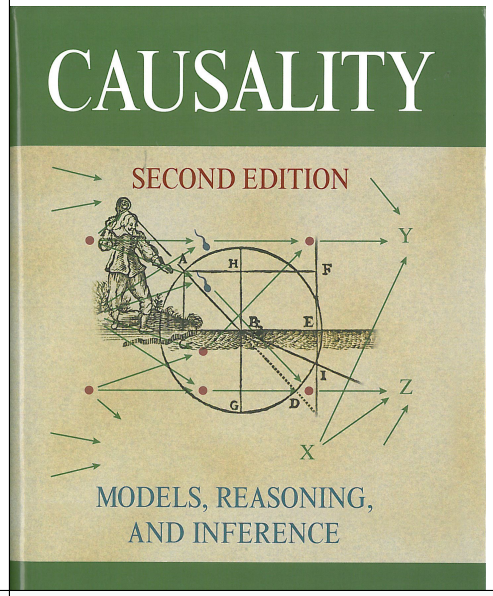
What if you have MANY response variables? Structural Equation Modeling, Clustering, NDMS, and more!

Bayesian Methods



A different kind of inference with a huge amount of flexibility.

Digging into Causality



What is causal inference? When can we draw causal conclusions?

Closing Thoughts

Course Goals

1. Learn how to think about your research in a systematic way to design efficient observational & experimental studies.
2. Understand how to get the most bang for your buck from your data.
3. Make you effective collaborators with statisticians.
4. Make you comfortable enough to learn and grow beyond this class.

We Are Fitting Models

Think Causally - When you Can

Does X Influence Y?
Might X and Y be influences by a common cause?
How can we design a study to cleanly determine the relationship/effect between X and Y?

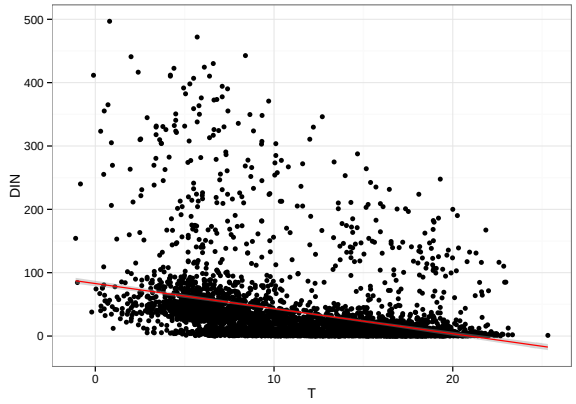
We're Just Fitting a Curve with an Error Distribution

$$Y \sim D(\ F(X)\ )$$

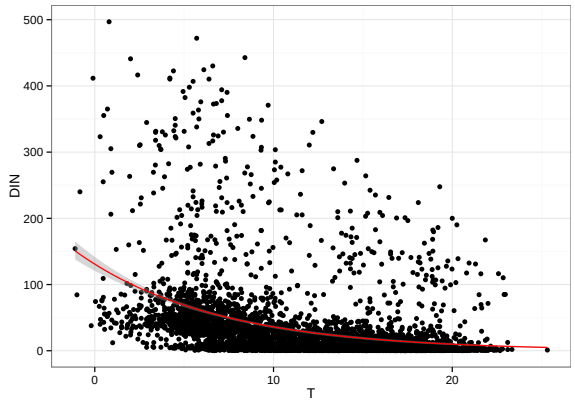
$$Y = \beta X + \epsilon_i$$

- F(X) can take many forms
- D, ϵ_i need not just be normal.

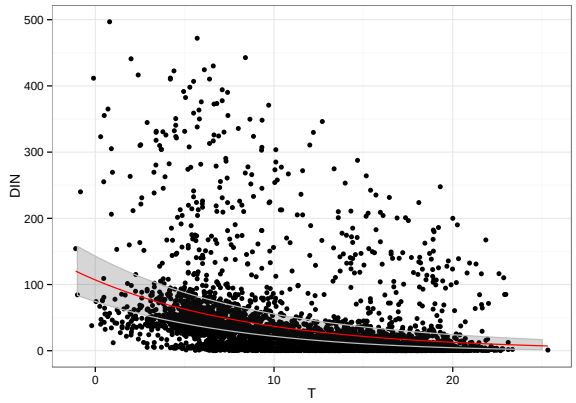
We're Just Fitting a Curve with an Error Distribution



We're Just Fitting a Curve with an Error Distribution



We're Just Fitting a Curve with an Error Distribution



Moving Beyond Simple Error Structures

$$y_i = \alpha_{j[i]} + \beta_{[j]i} X + \epsilon_{ij}$$

$$cor(\epsilon) = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

And this is just a start!

Think About What You Are Doing

Think About What You Are Doing

Think *a priori*

I think that Y is predicted by...

Consider Your Mode of Inference

Am I testing a null hypothesis? Why?

Do I want to evaluate the relative weight of evidence for multiple hypotheses?

Do I have prior information? Do I want to know about my degree of belief?

Can I even make parameteric assumptions about relationships?

Exploration and Verification are Both Valid

Let's go and build a model v. Let's test a single predictive framework

Sample Size

- How many points to fit a probability distribution?
- Ensure that your effect is not a fluke accident
- $\frac{p^{3/2}}{n}$ should approach 0 for Likelihood (Portnoy 1988 Annals of Statistics)
- i.e.,$\sim$10 samples per paramter (1 treatment = 1 parameter, but this is total # of samples)

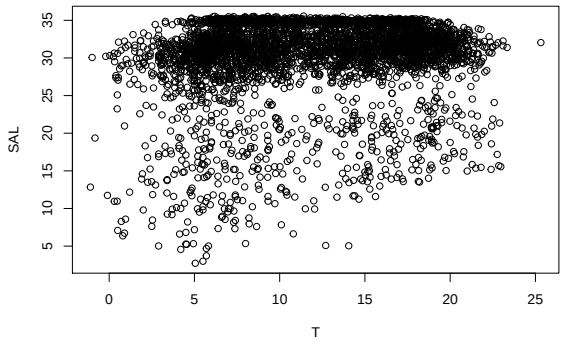
Computational Tools Are Just That - Tools!

Coding Brings You Closer to Your Model

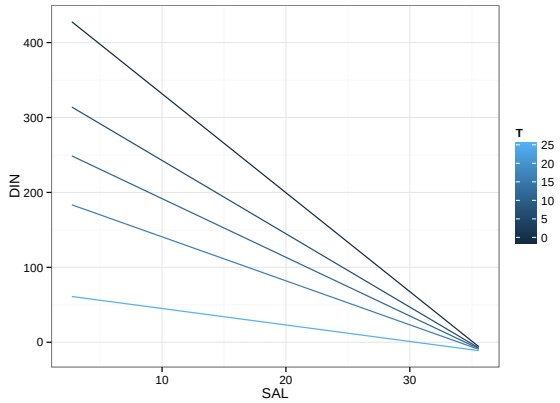
Which helps you understand the model you are fitting better? Select the Model Menu. Select General Linear Model. Click Y as your response. Then click X and your predictor. Click Block. Scroll to a second menu. Select random effects model, and click Block again. Click Run.

```
lme(Y ~ X, random=~ 1|Block, data=mydata)
```

Screen Your Data



Visualize Your Work



It is likely that no one ever masters anything in which he has not known impotence; and if you agree, you will see that this impotence comes not at the beginning of or before the struggle with the subject, but at the heart of it.

- Walter Benjamin