Module 4F
Supervised Machine Learning

Different flavors of REGRESSION and General Linear Models

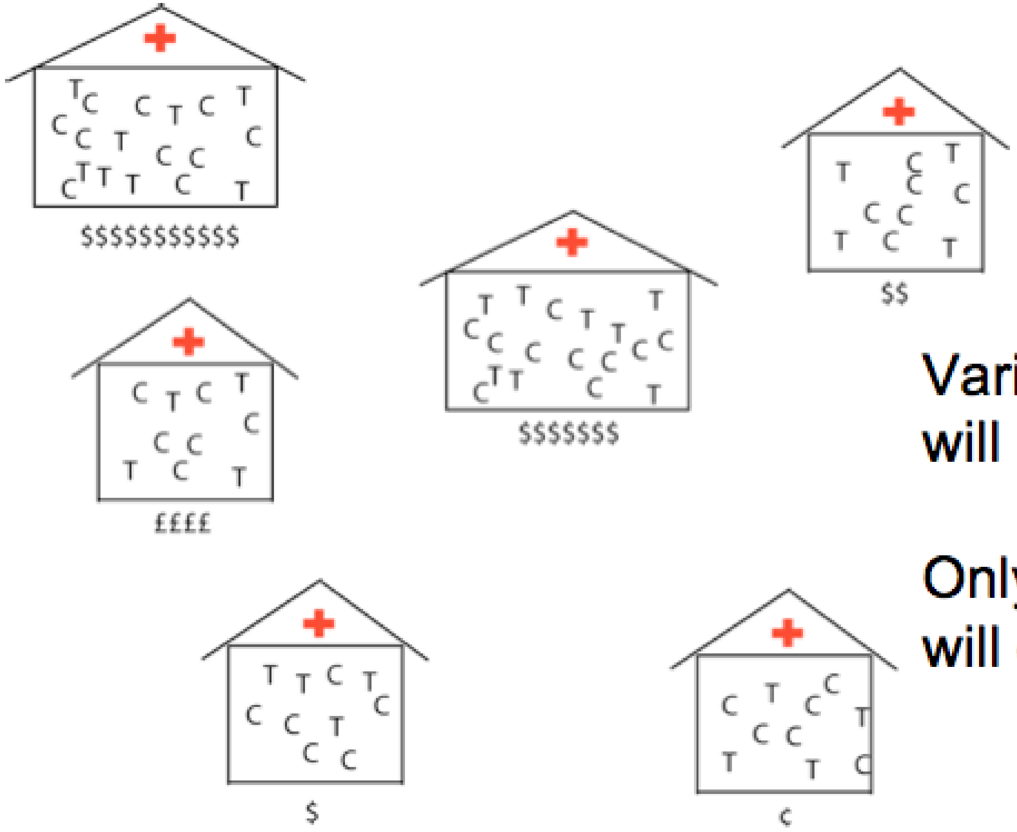
Blocking

- Results in an additional variable, a block, that must be included in analysis
  - Can no longer use simple one-factor ANOVA
- Randomized block design
  - **Paired design** for > 2 treatments
  - Example:
    - Every treatment is replicated **once** within each block
    - Minimize "noise"
- Accounting for any variation caused by blocking **can improve our treatment effect detection** (i.e., increase the power of our test)
- Treatment effects are assessed by different treatments **within** each block so there is no interaction term

Goals of experiments:

determine how explanatory variable (treatment) affects response variable

- Eliminate Bias
- Reduce Sampling Error
  - Blocking:



C = Control
T = Treated

Variance among hospitals will not contribute to SE.

Only variance within hospitals will contribute to "noise"

Main Principle of Blocking

Response = Constant + Treatment + Block

H_0: Response = Constant + Block
H_A: Response = Constant + Block + Treatment

- Determine significance via ANOVA table which includes a row for the **block**
- Calculates a F value for block - examines how much better fit is with the block versus without the block

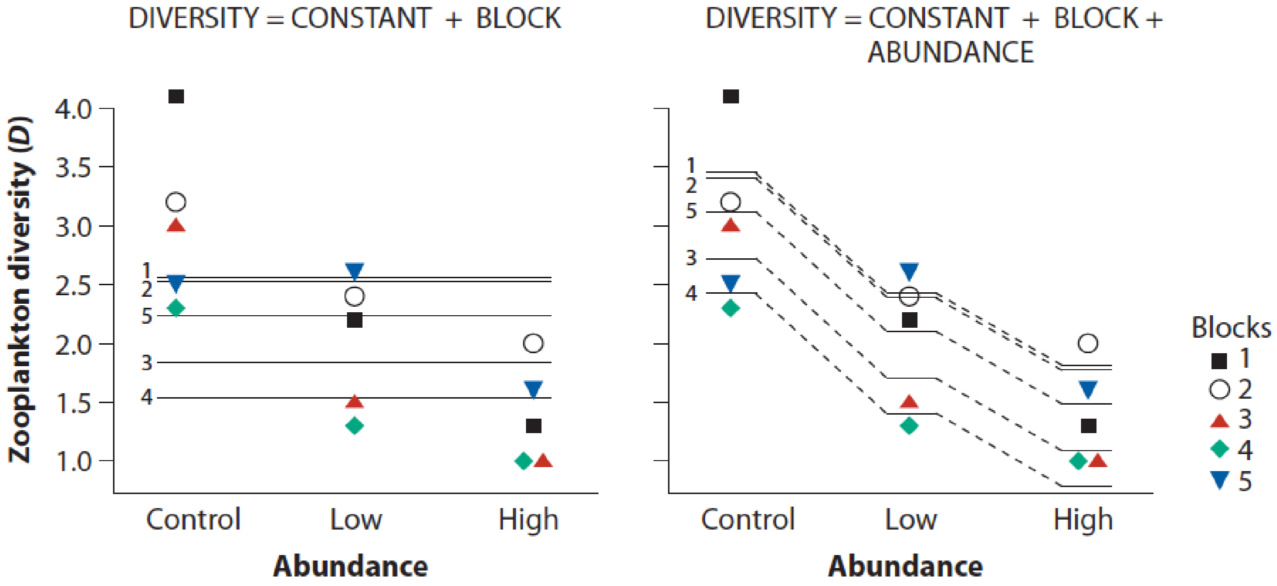
Example of Blocking:

Response = Constant + Treatment + Block

H_0: Response = Constant + Block

H_A: Response = Constant + Block + Treatment

$$F = \frac{H_A}{H_0} = \frac{\text{Constant+Block+Treatment}}{\text{Constant + Block}} = \frac{\text{residual+location+fish Abundance}}{\text{residual+location}}$$



Source of variation	Sum of Squares	df	Mean Square	F	P
BLOCK	2.340	4	0.5850		
Treatment	6.8573	2	3.4287	16.37	0.001
Residual	1.6760	8	0.2095		
Total	10.8733	14			

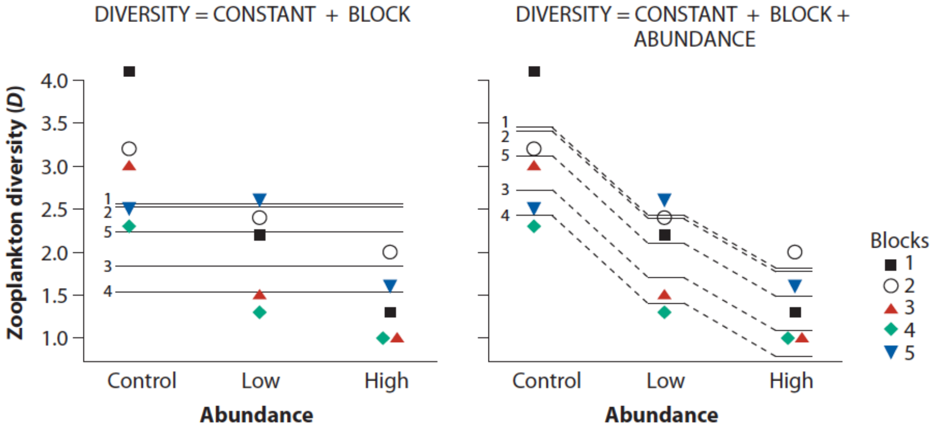
Example of Blocking:

Response Full = Constant + Treatment + Block

Treatment:

H_0: Response = Constant + Block

H_A: Response = Constant + Block + Treatment



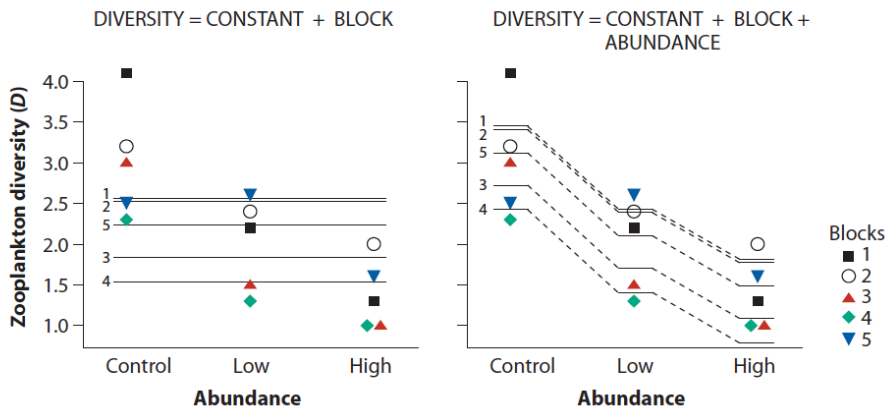
$$F = \frac{H_A}{H_0} = \frac{\text{Constant+Block+Treatment}}{\text{Constant + Block}}$$

$$= \frac{\text{residual+location+fish Abundance}}{\text{residual+location}} = \frac{MS_{\text{treatment}}}{MS_{\text{block}}} = \frac{3.43}{0.59} = \mathbf{16.37}$$

$F_{0.05(1),2,14}$ **= 3.74** so we reject the H_O

Source of variation	Sum of Squares	df	Mean Square	F	P
BLOCK	2.340	4	0.5850		
Treatment	6.8573	2	3.4287	16.37	0.001
Residual	1.6760	8	0.2095		
Total	10.8733	14			

Example of Blocking:



Treatment:

H_0: Response = Constant + Block

H_A: Response = Constant + Block + Treatment

$$F = \frac{H_A}{H_0} = \frac{\text{Constant+Block+Treatment}}{\text{Constant + Block}}$$

$$= \frac{\text{residual+location+fish Abundance}}{\text{residual+location}} = \frac{MS_{treatment}}{MS_{block}} = \frac{3.43}{0.59} = \mathbf{16.37}$$

$\mathbf{F_{0.05(1),2,14} = 3.74}$ so we reject the H_O

Block:

$$F_{Block} = \frac{H_A}{H_0} = \frac{\text{Residual+treatment+Block}}{\text{Residual +treatment}} = \frac{MS_{block}}{MS_{residual}} = \frac{0.5850}{0.2095} = 2.79$$

$\mathbf{F_{0.05(1),4,8} = 3.84}$ so we fail to reject the H_O

Source of variation	Sum of Squares	df	Mean Square	F	P
BLOCK	2.340	4	0.5850		
Treatment	6.8573	2	3.4287	16.37	0.001
Residual	1.6760	8	0.2095		
Total	10.8733	14			