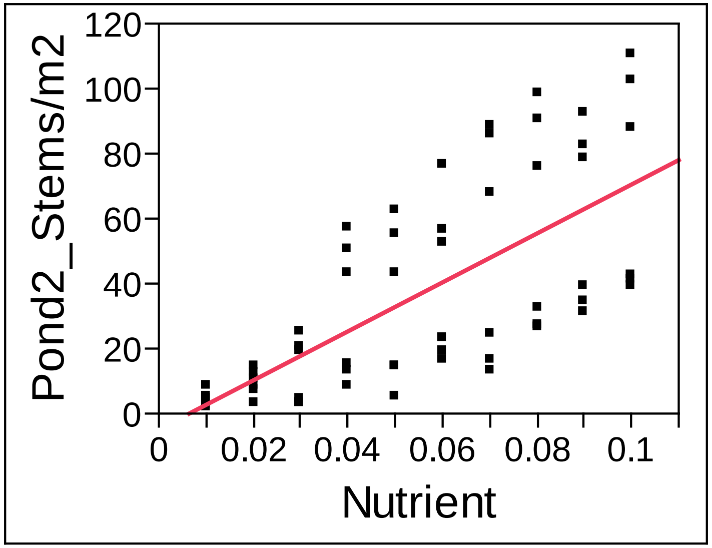
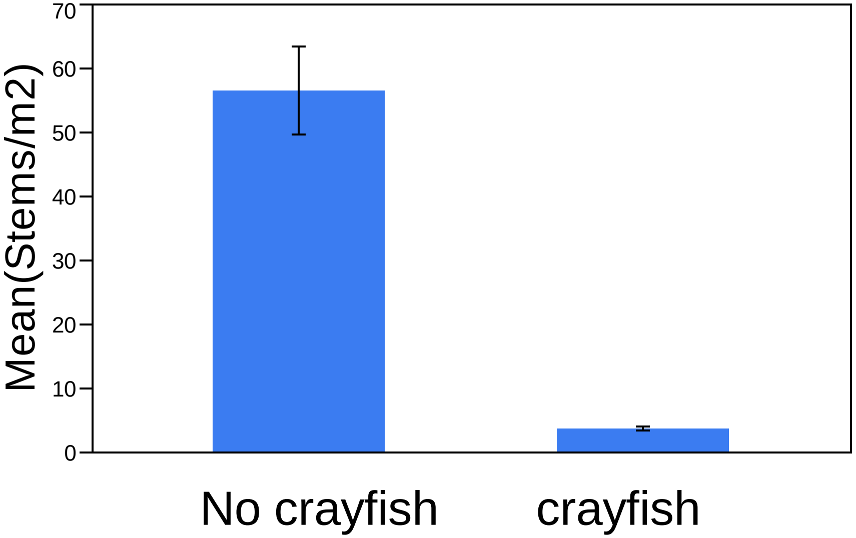
Analysis of Covariance (ANCOVA)

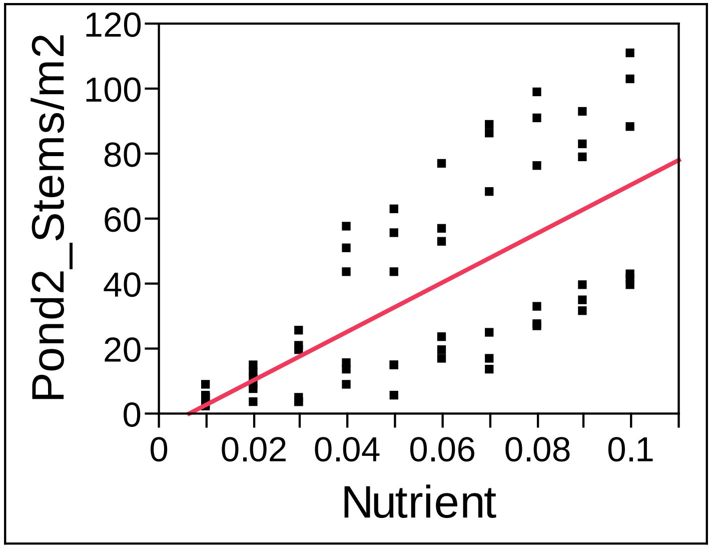
Hybrid of ANOVA and regression



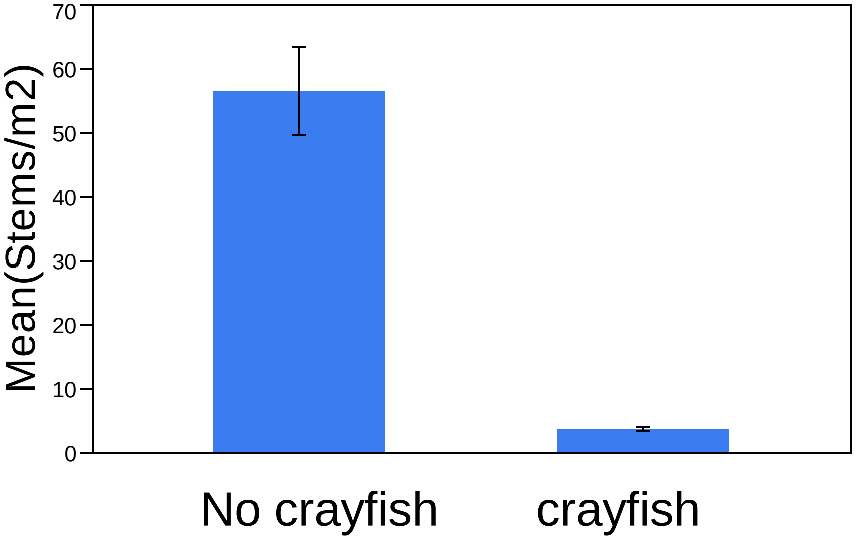


Analysis of Covariance (ANCOVA)

Hybrid of ANOVA and regression

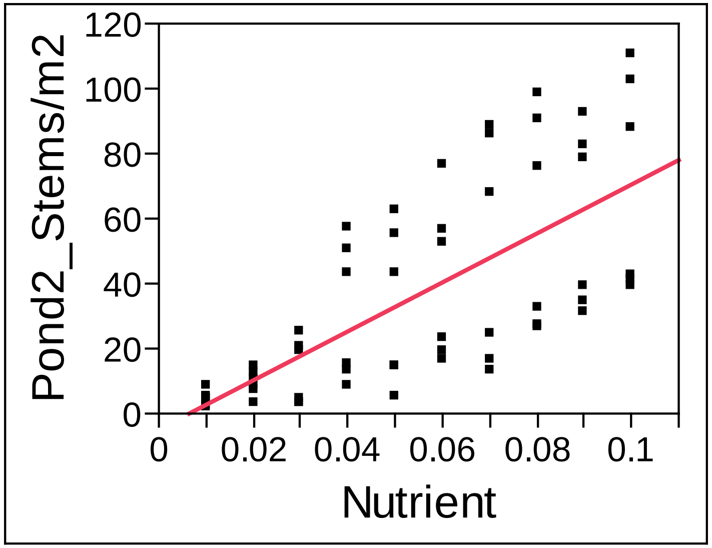


You have both categorical and continuous X variables

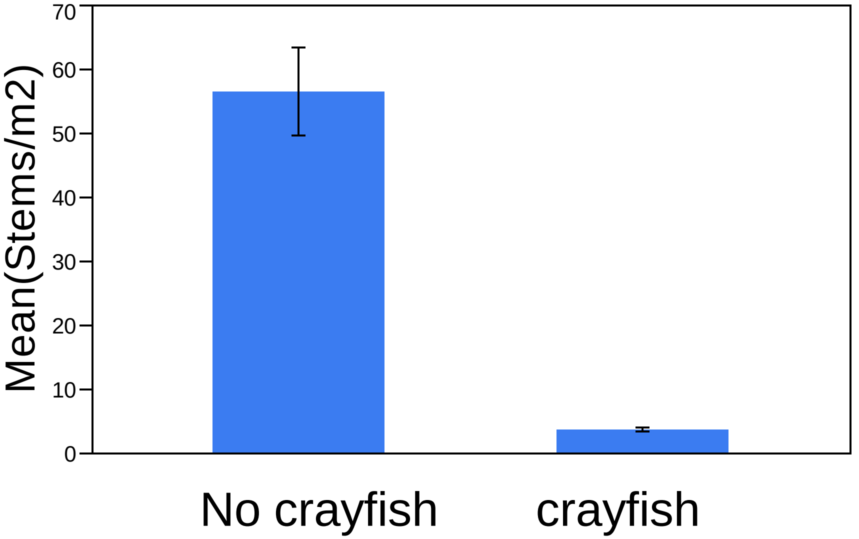


Analysis of Covariance (ANCOVA)

Hybrid of ANOVA and regression

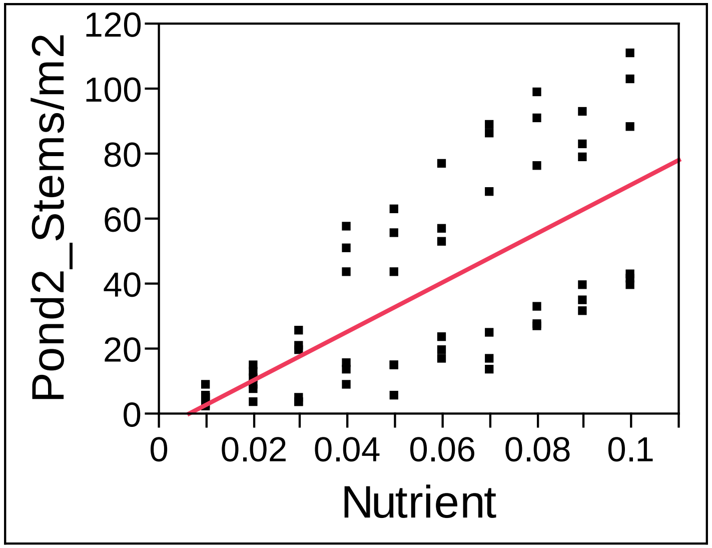


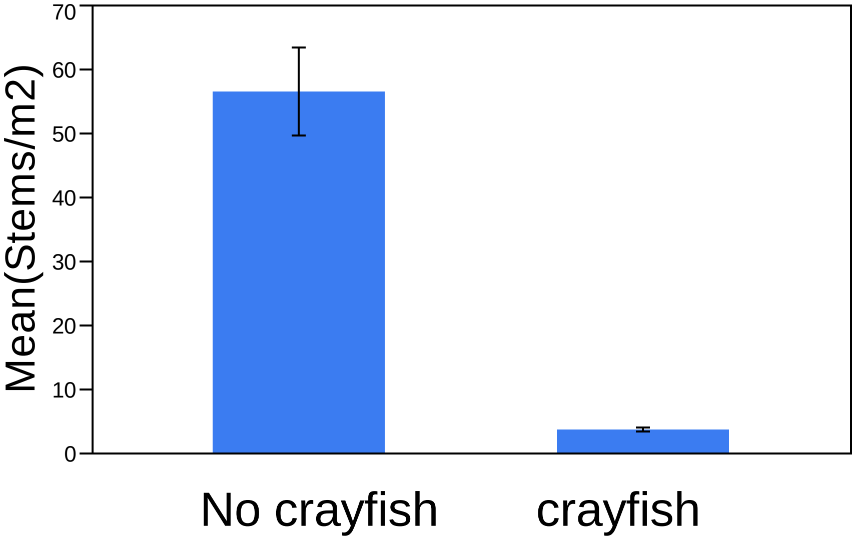
You have both categorical and continuous X variables



ANOVA design, with an additional continuous variable (covariate)

Analysis of Covariance (ANCOVA)

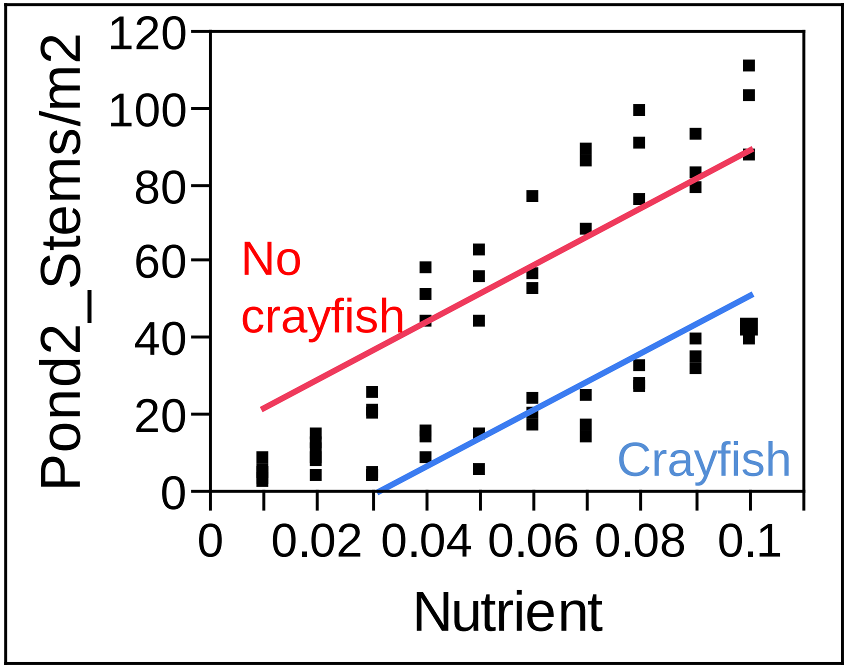


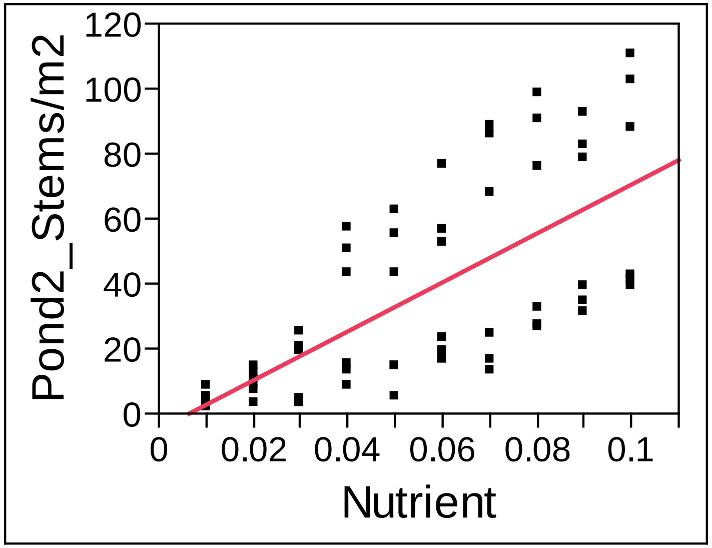


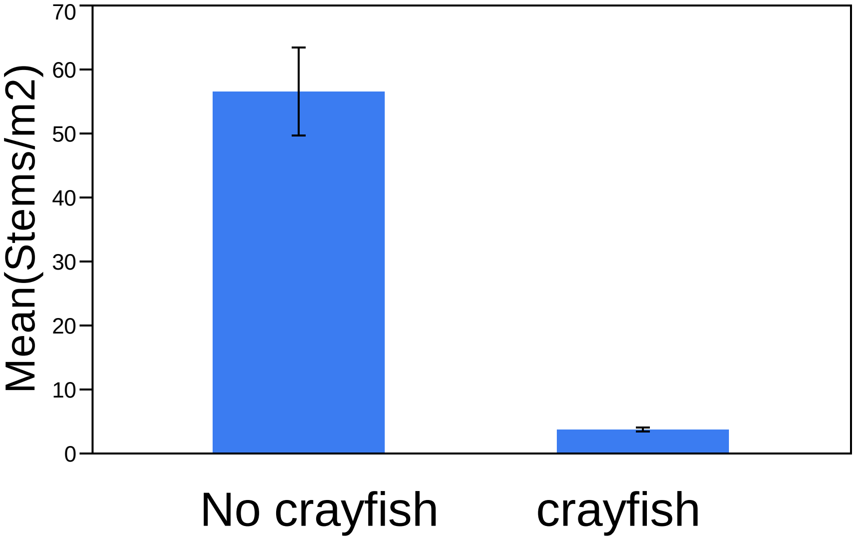
ANOVA design, with an additional continuous variable (covariate)

Photo: E. Engbretson, U.S. Fish and Wildlife Service

Analysis of Covariance (ANCOVA)







ANOVA design, with an additional continuous variable (covariate)

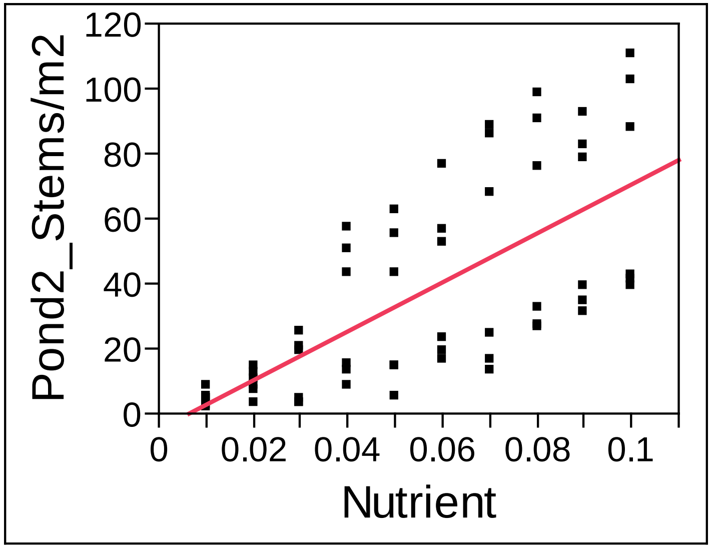
fake_crayfish.jmp

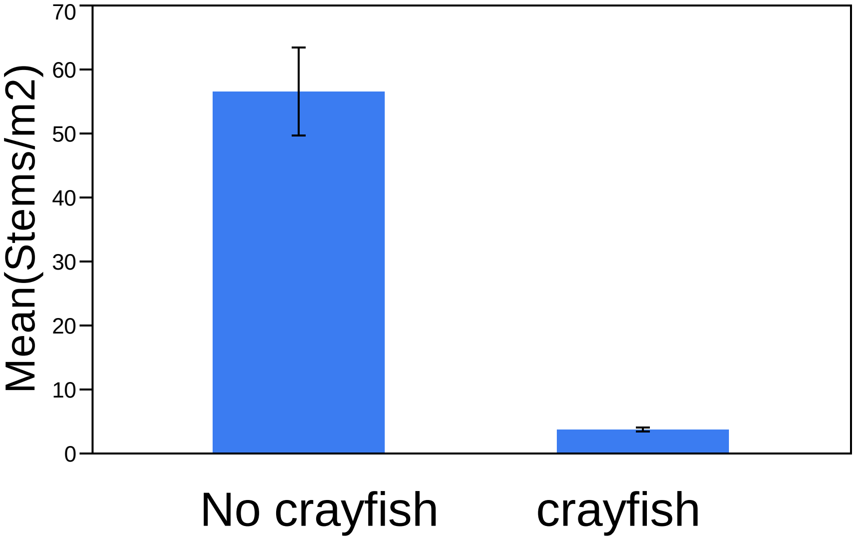
The modification of ANOVA that allows ANCOVA

ANOVA

$Y_{ij} = \mu + A_{ij} + \varepsilon_{ij}$

Observations Y are a combo of some mean level μ + the positive or negative average effect A of each level i plus error ε





fake_crayfish.jmp

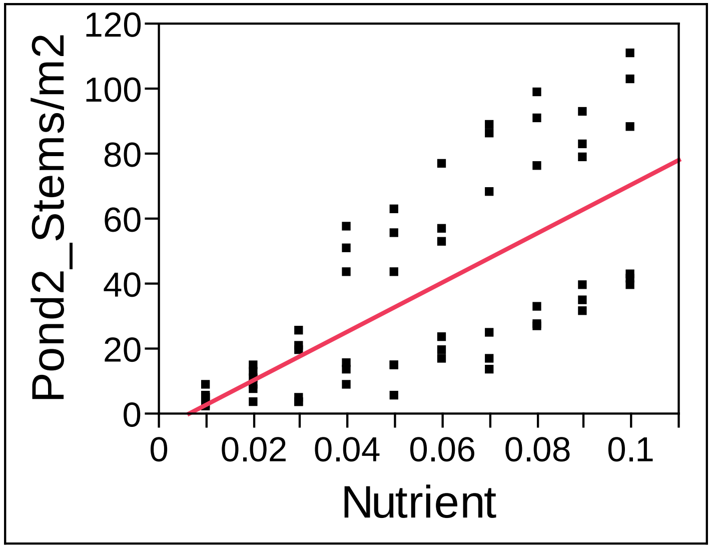
The modification of ANOVA that allows ANCOVA

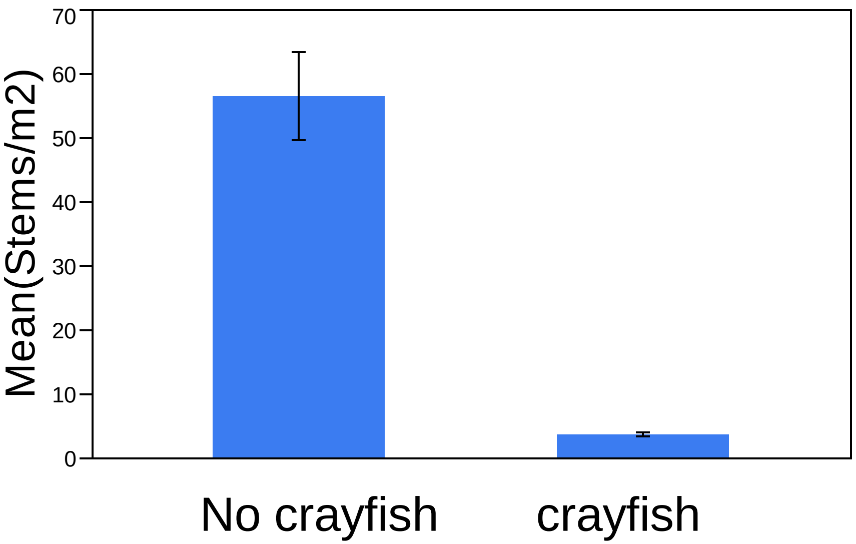
ANOVA

$Y_{ij} = \mu + A_{ij} + \varepsilon_{ij}$

Observations Y are a combo of some mean level μ + the positive or negative average effect A of each level i plus error ε

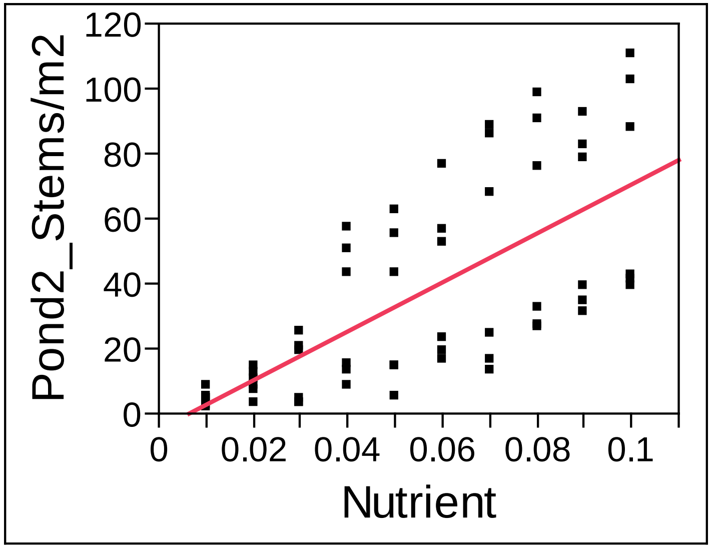
If a continuous variable is affecting Y, its effects end up in the error term ε





fake_crayfish.jmp

The modification of ANOVA that allows ANCOVA



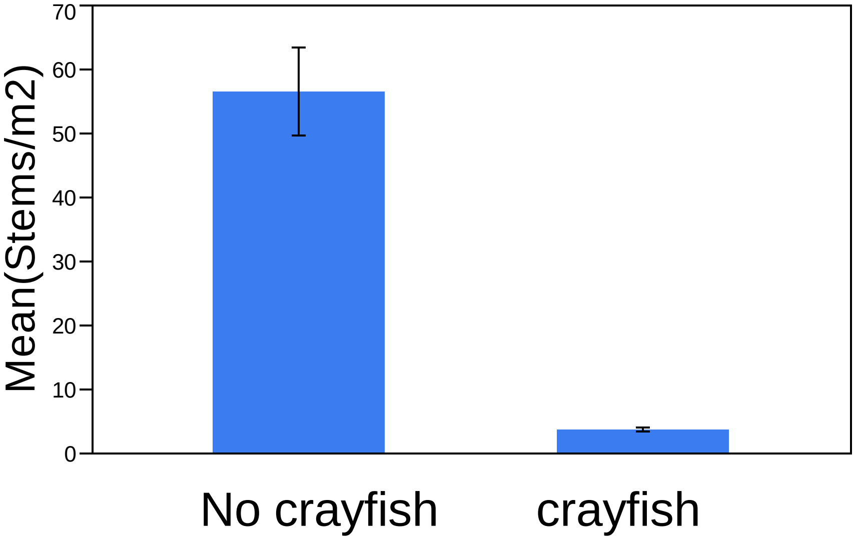
ANOVA

$Y_{ij} = \mu + A_{ij} + \varepsilon_{ij}$

Observations Y are a combo of some mean level μ + the positive or negative average effect A of each level i plus error ε

If a continuous variable is affecting Y, its effects end up in the error term ε

So… correlate Y with the continuous variable first, and then use ANOVA to analyze residuals



fake_crayfish.jmp

The modification of ANOVA that allows ANCOVA

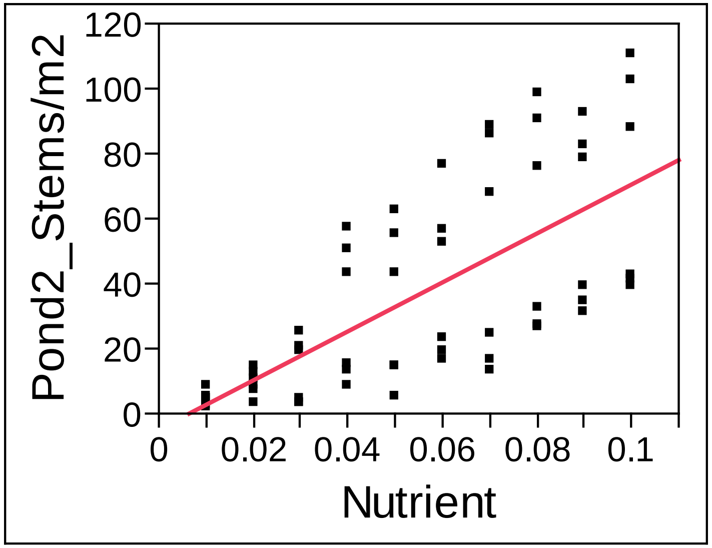
ANOVA

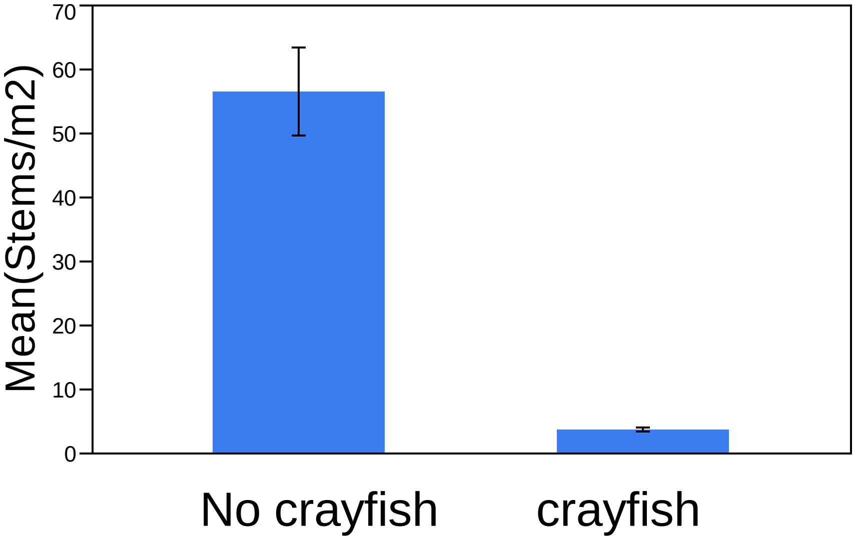
$Y_{ij} = \mu + A_{ij} + \varepsilon_{ij}$

So… correlate Y with the continuous variable first, and then use ANOVA to analyze residuals

ANCOVA

$Y_{ij} = \mu + A_i + \beta_i(X_{ij} - \bar{X}_i) + \varepsilon_{ij}$





fake_crayfish.jmp

The modification of ANOVA that allows ANCOVA

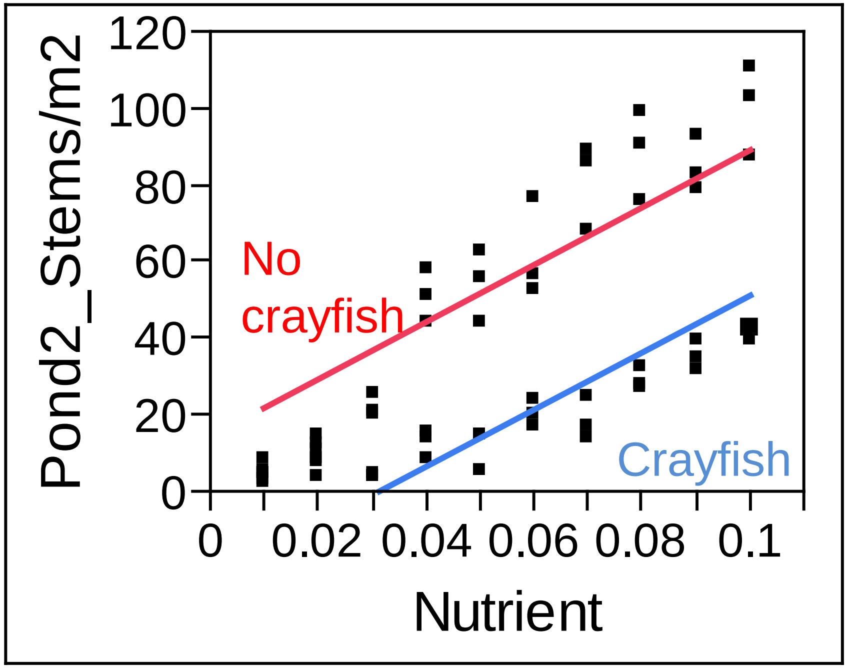
ANOVA

$Y_{ij} = \mu + A_{ij} + \varepsilon_{ij}$

So… correlate Y with the continuous variable first, and then use ANOVA to analyze residuals

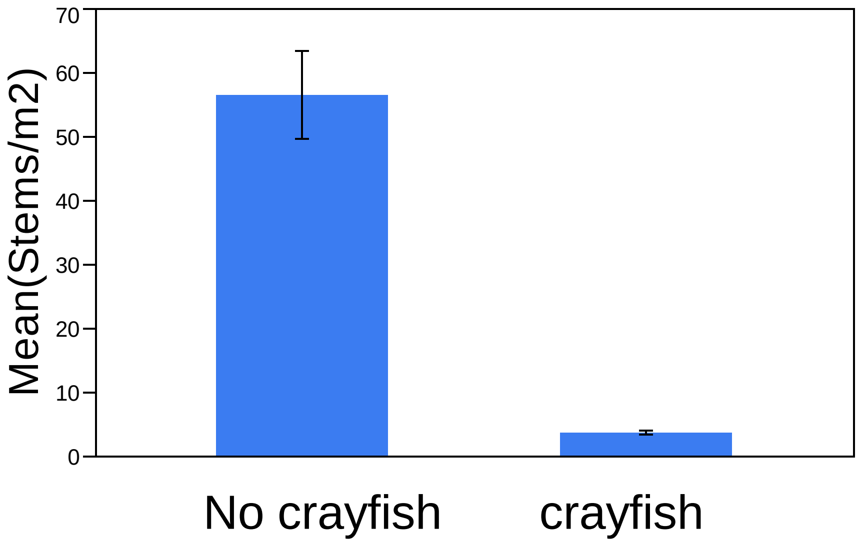
ANCOVA

$Y_{ij} = \mu + A_i + \beta_i(X_{ij} - \bar{X}_i) + \varepsilon_{ij}$



fake_crayfish.jmp

The modification of ANOVA that allows ANCOVA



ANOVA

$Y_{ij} = \mu + A_{ij} + \varepsilon_{ij}$

So… correlate Y with the continuous variable first, and then use ANOVA to analyze residuals

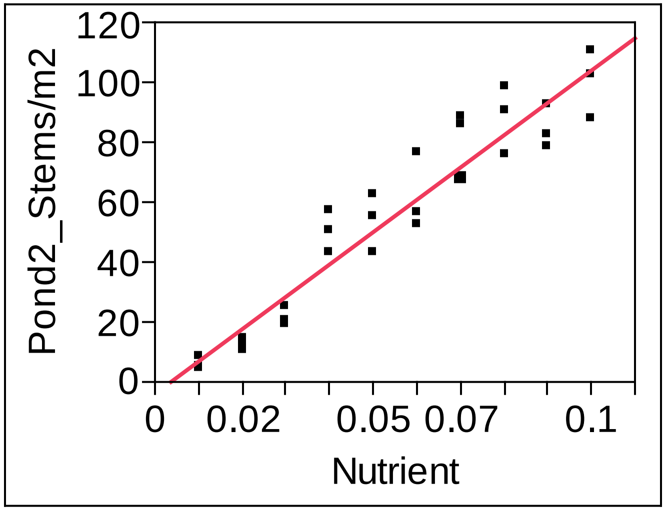
ANCOVA

$Y_{ij} = \mu + A_i + ß_i(X_{ij} - \overline{X}_i) + \varepsilon_{ij}$

If ß is non-significant, it is zero, so the equation reduces to a 1-way ANOVA

fake_crayfish.jmp

The modification of ANOVA that allows ANCOVA



ANOVA

$Y_{ij} = \mu + A_{ij} + \varepsilon_{ij}$

So… correlate Y with the continuous variable first, and then use ANOVA to analyze residuals

ANCOVA

$Y_{ij} = \mu + A_i + \beta_i(X_{ij} - \bar{X}_i) + \varepsilon_{ij}$

If ß is non-significant, it is zero, so the equation reduces to a 1-way ANOVA

If the treatment effect (A) is non-significant, the equation may reduce to a linear regression

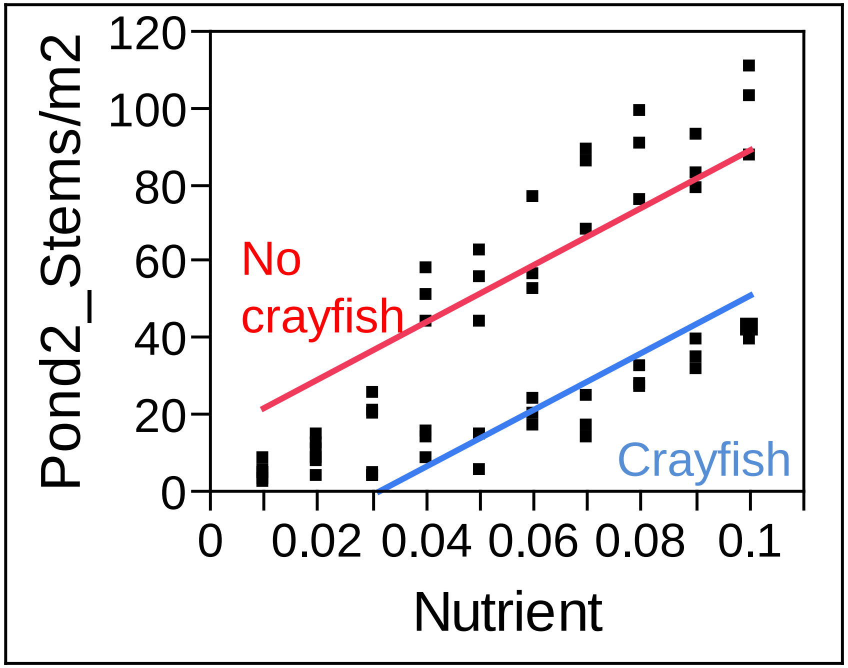
fake_crayfish.jmp

ANCOVA

$Y_{ij} = \mu + A_i + \beta_i(X_{ij} - \bar{X}_i) + \varepsilon_{ij}$

Generally, the algorithm first tests to see whether slopes differ between/among treatments

If not, a common slope is used – the simplest case



fake_crayfish.jmp

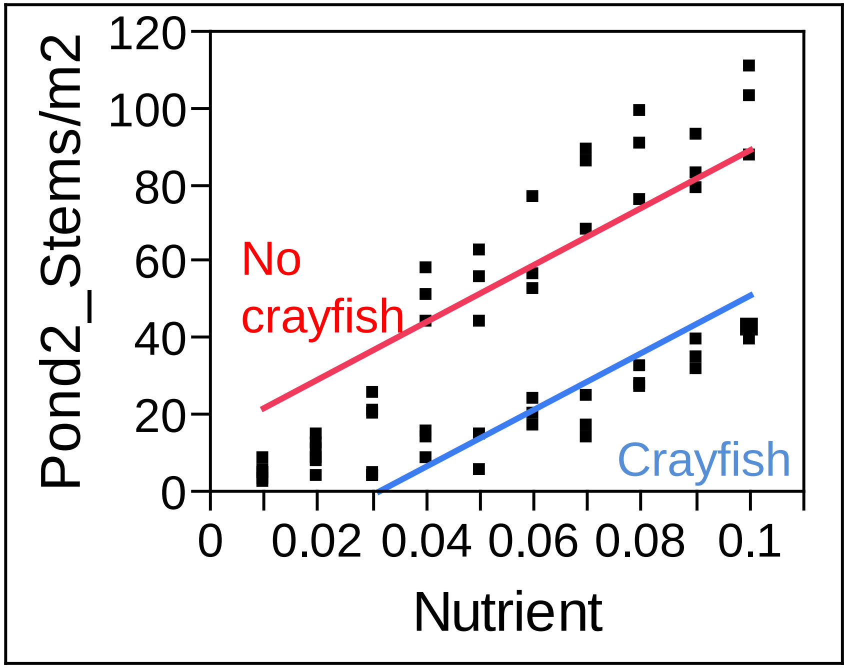
ANCOVA

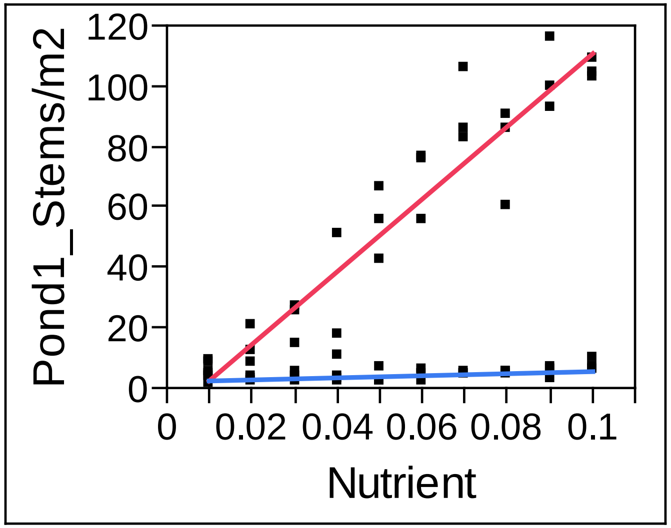
$Y_{ij} = \mu + A_i + \beta_i(X_{ij} - \bar{X}_i) + \varepsilon_{ij}$

Generally, the algorithm first tests to see whether slopes differ between/among treatments

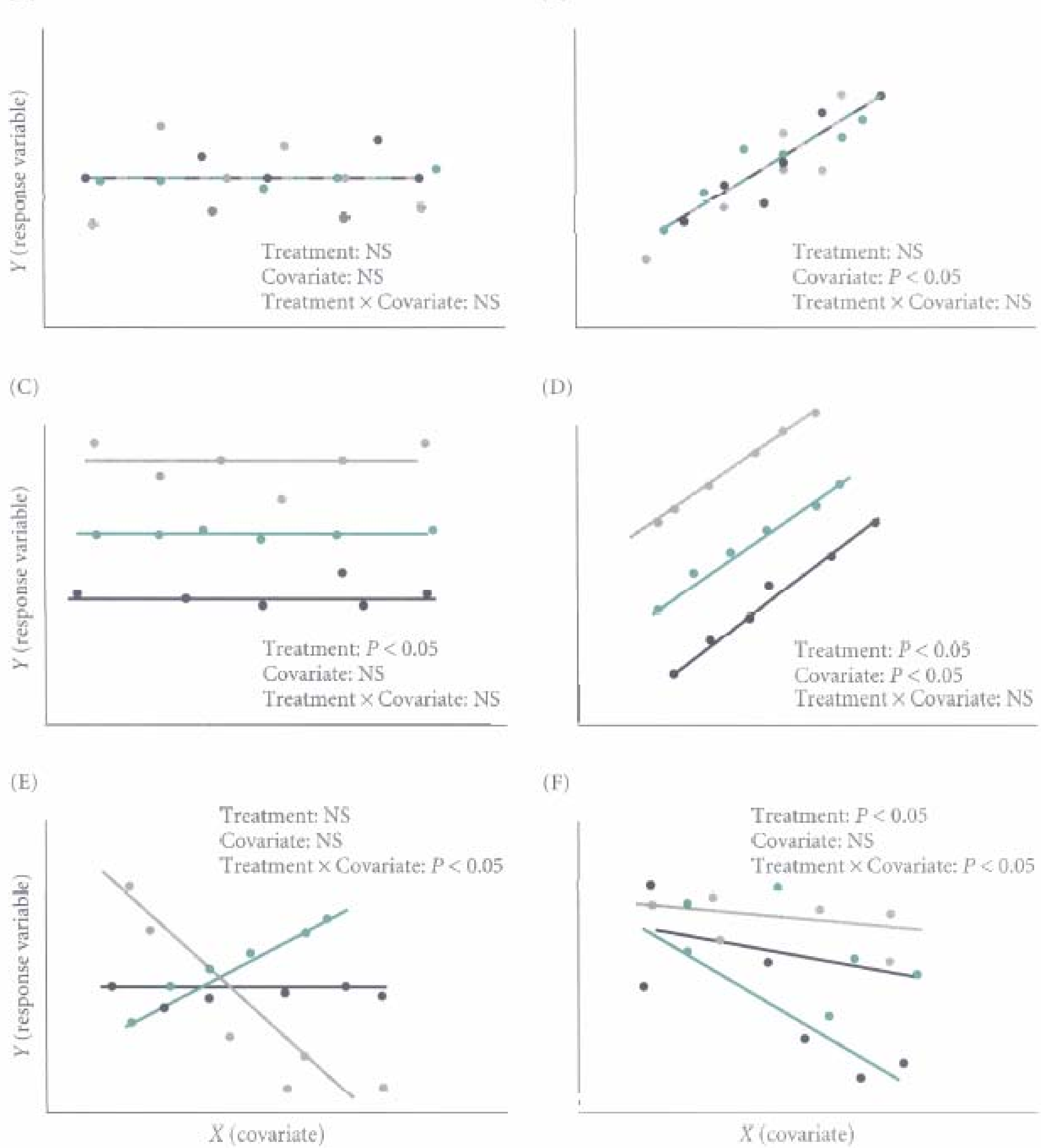
If not, a common slope is used – the simplest case

Interaction term is significant when slopes differ significantly





fake_crayfish.jmp

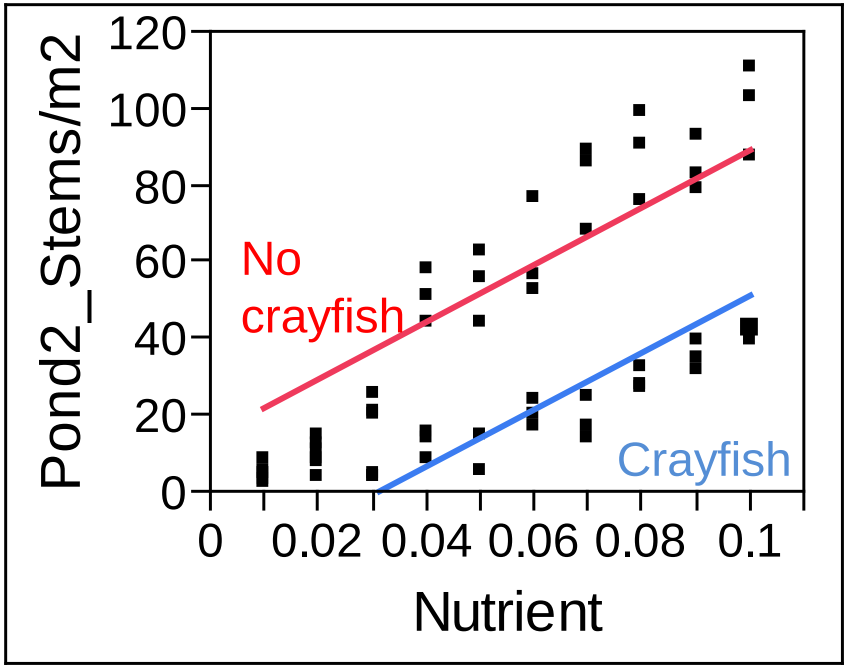


Gotelli & Ellison Fig. 10.4

What are important assumptions in ANCOVA?

Normality/homoscedasticity

Independent samples

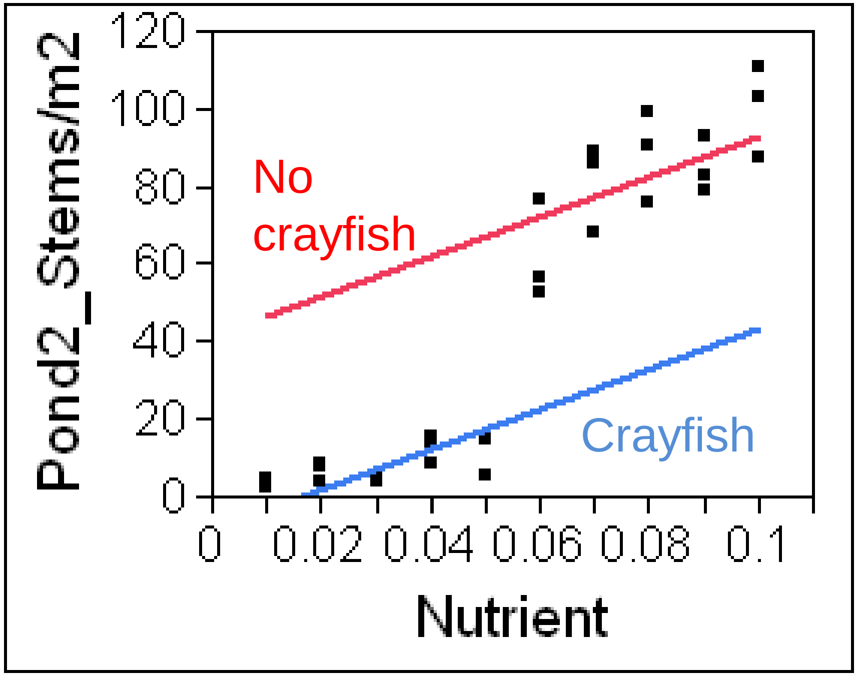


fake_crayfish.jmp

What are important assumptions in ANCOVA?

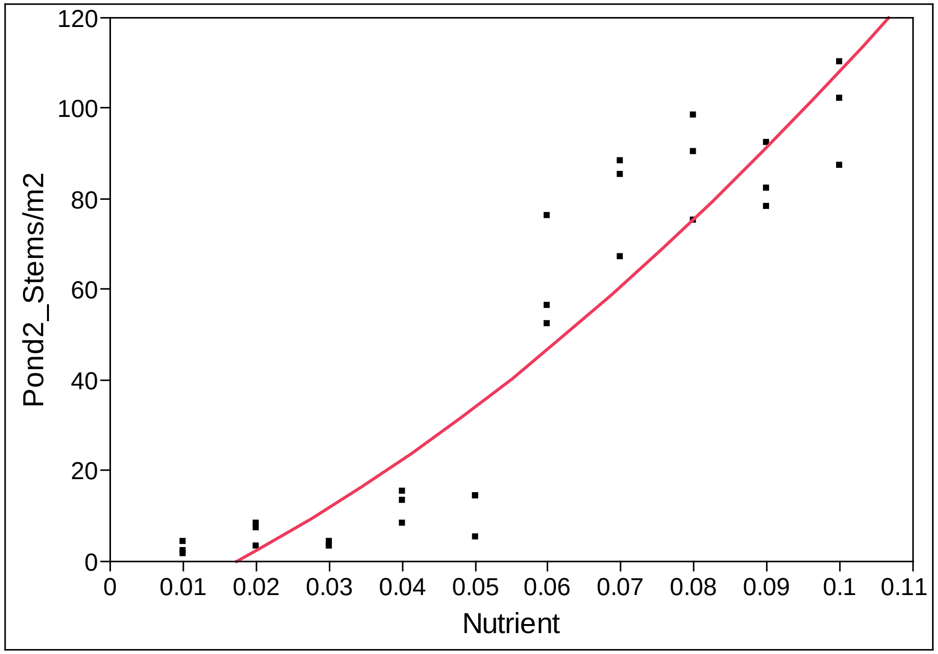
Normality/homoscedasticity

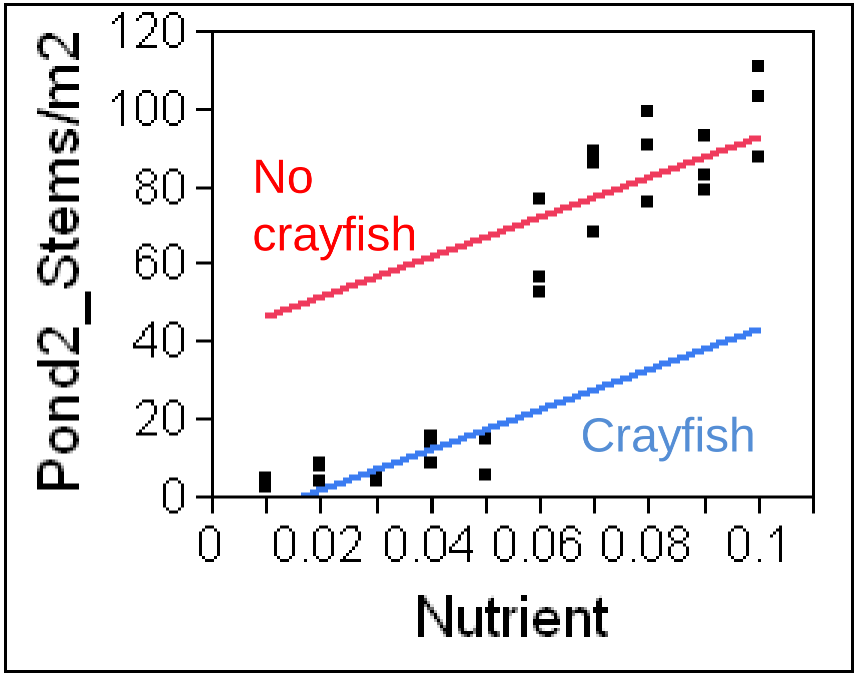
Independent samples



Careful that effects are not confounded!

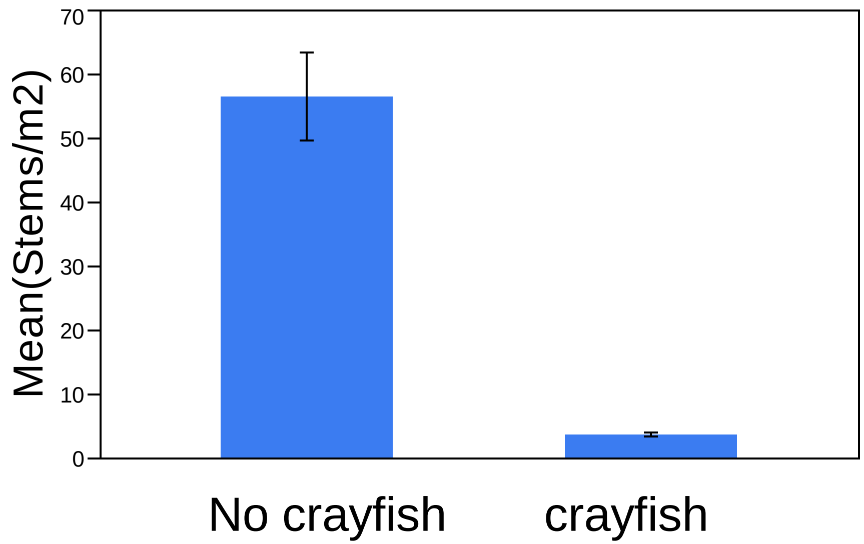
fake_crayfish.jmp

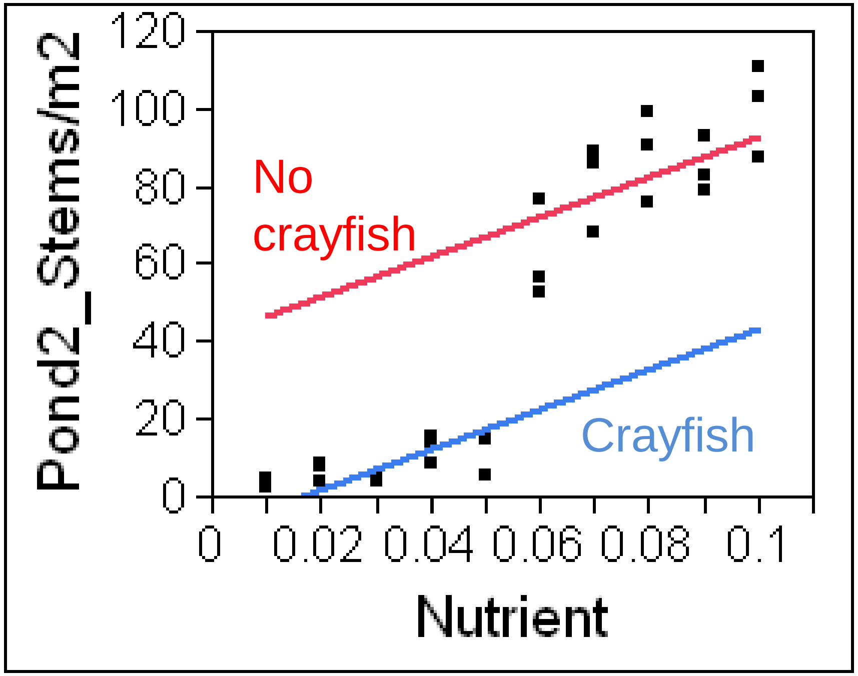


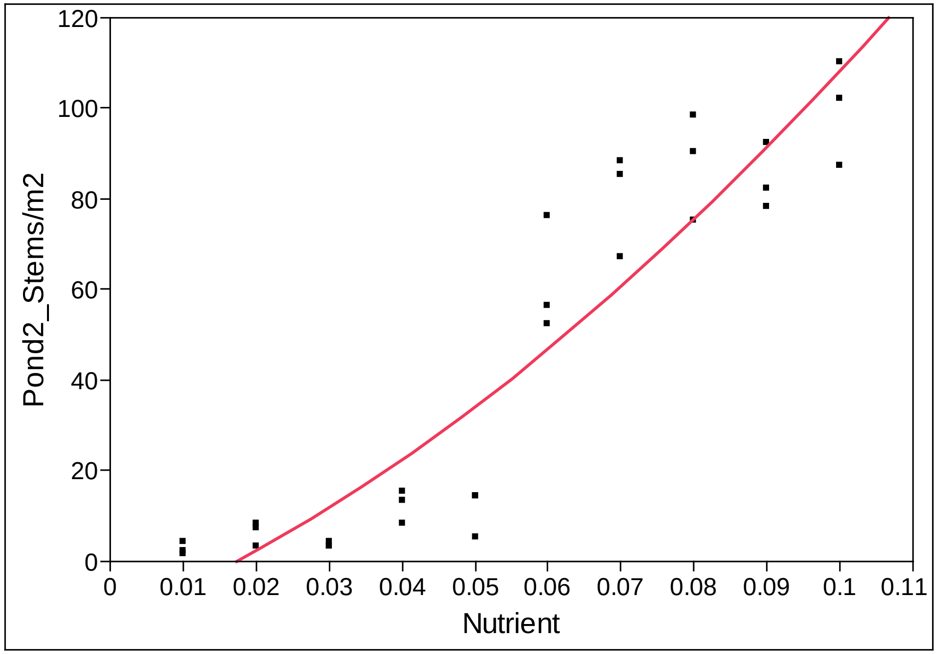


Careful that effects are not confounded!

fake_crayfish.jmp







Careful that effects are not confounded!

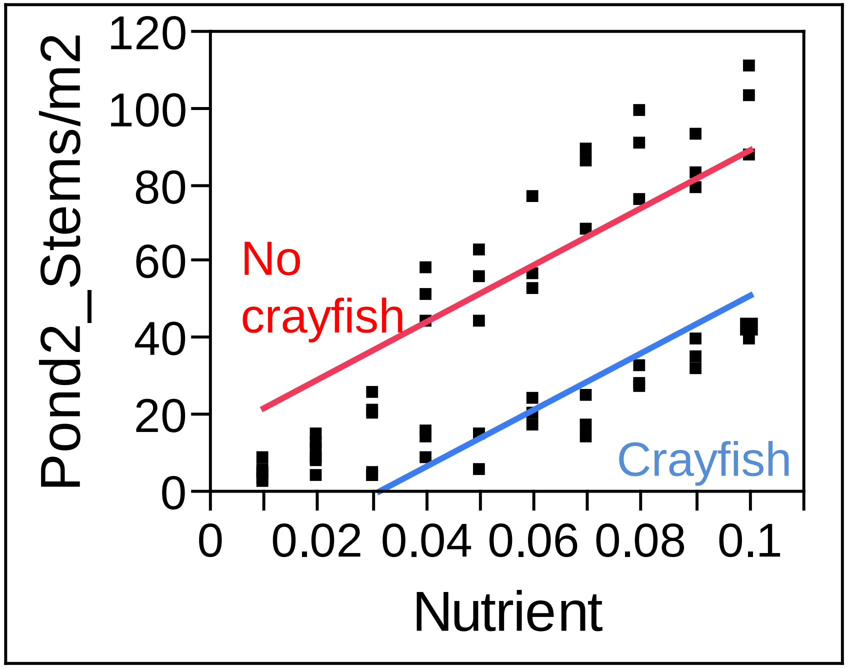
fake_crayfish.jmp

ANCOVA

$Y_{ij} = \mu + A_i + ß_i(X_{ij} - X_i) + \varepsilon_{ij}$

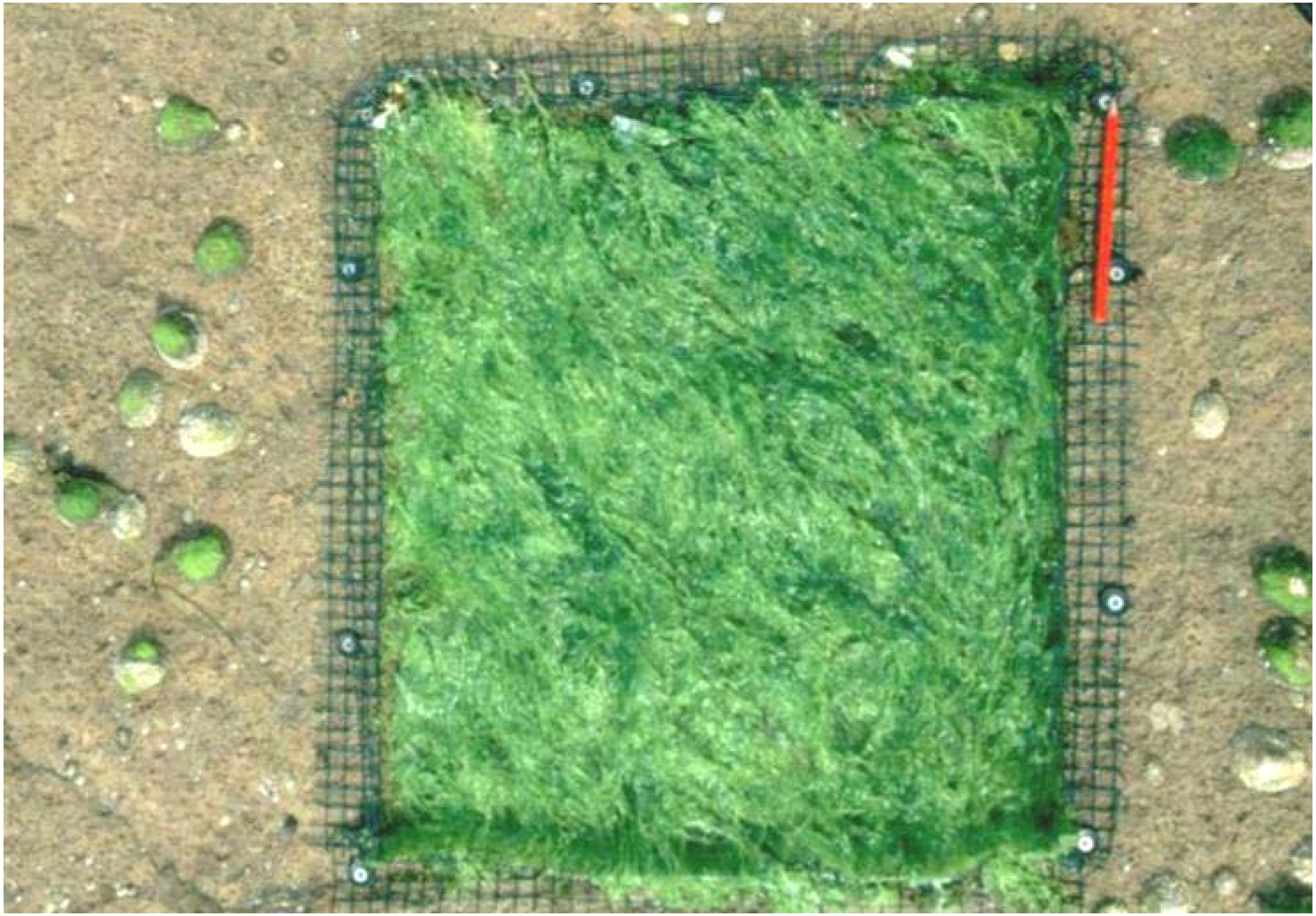
May be used in a range of situations, e.g.,

-to explain "extra" variation in experiment where the ANOVA design is of greatest interest



fake_crayfish.jmp

e.g.,
Presence/Absence of grazer in field exclosures
+ naturally varying temperature, slope, or nutrients….



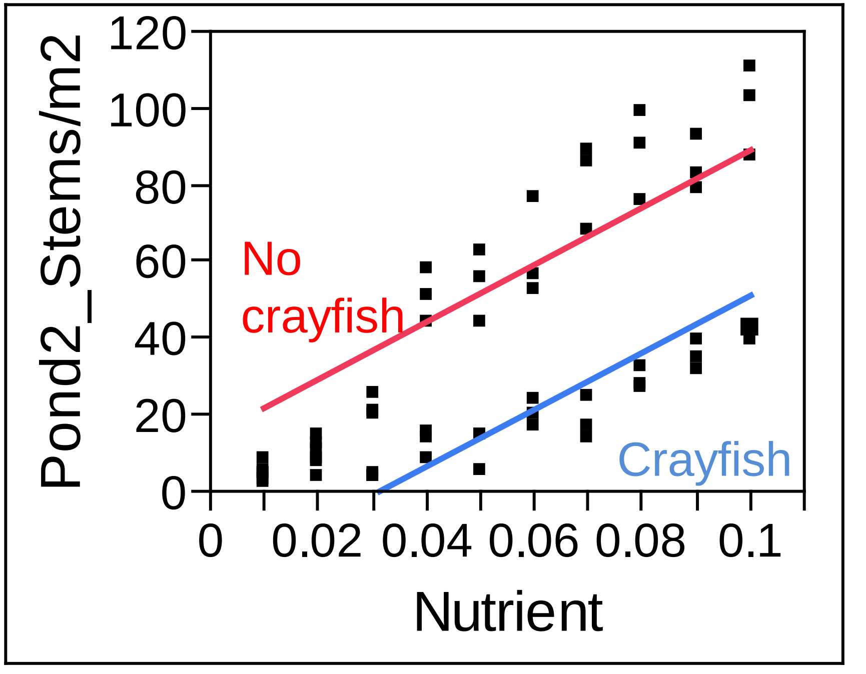
http://www.field-studies.org.uk/limpet_grazing.htm

ANCOVA

$Y_{ij} = \mu + A_i + \beta_i(X_{ij} - X_i) + \varepsilon_{ij}$

May be used in a range of situations, e.g.,

-to explain "extra" variation in experiment where the ANOVA design is of greatest interest

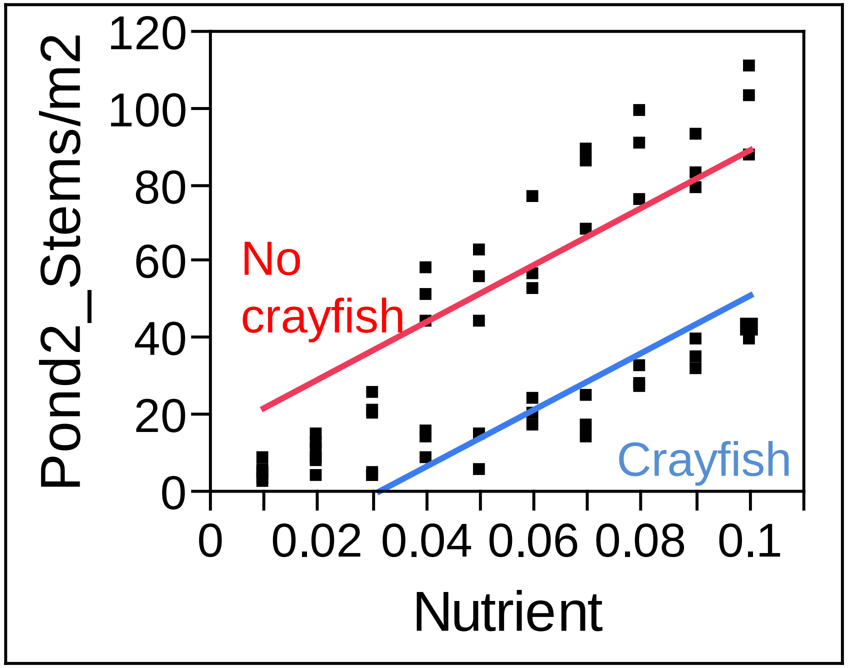


-experimentally, where continuous and categorical variables are both manipulated

See: Gotelli & Ellison P. 197

ANCOVA

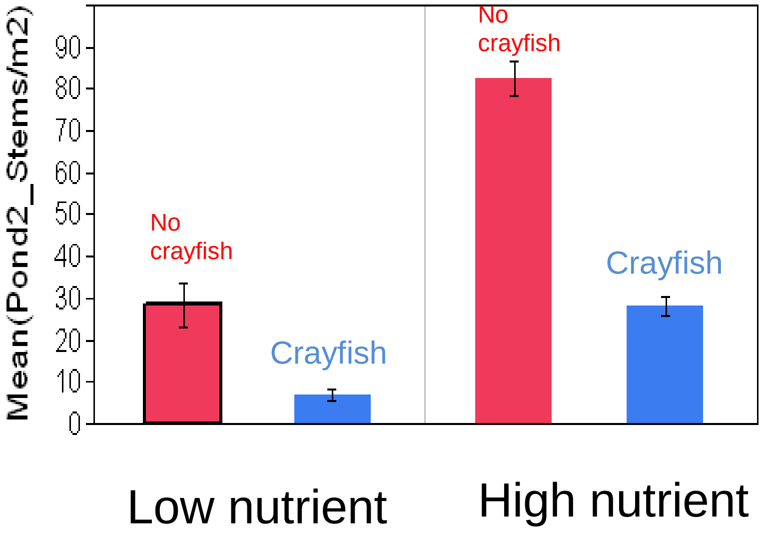
$Y_{ij} = \mu + A_i + \beta_i(X_{ij} - \bar{X}_i) + \varepsilon_{ij}$



May be used in a range of situations, e.g.,

-to explain "extra" variation in experiment where the ANOVA design is of greatest interest

-experimentally, where continuous and categorical variables are both manipulated



If you try to "shoehorn" a continuous variable into discrete categories for a 2-way ANOVA, you will need to treat that factor as **random** rather than **fixed**

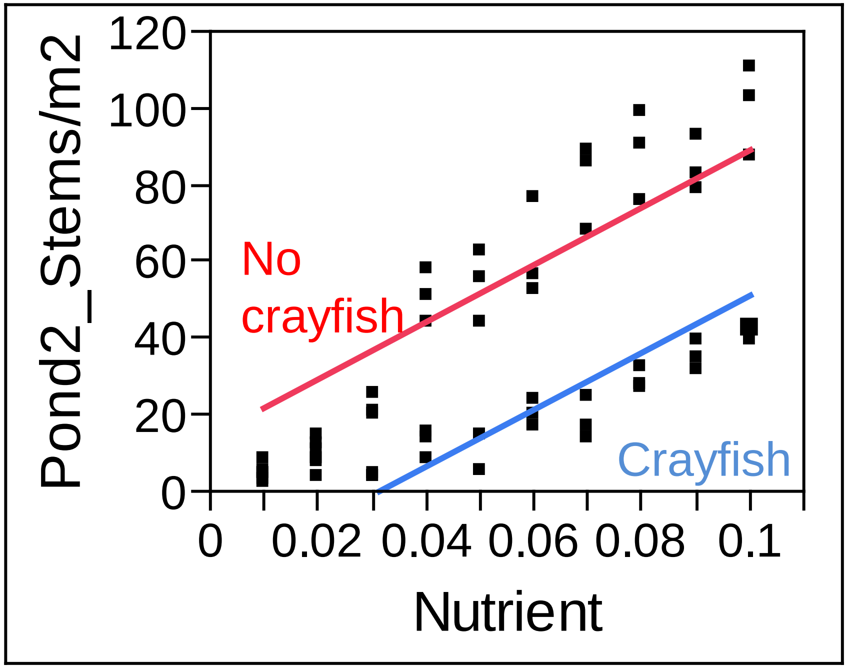
See: Gotelli & Ellison P. 199, Footnote

ANCOVA

$Y_{ij} = \mu + A_i + \beta_i(X_{ij} - X_i) + \varepsilon_{ij}$

May be used in a range of situations, e.g.,

-to explain "extra" variation in experiment where the ANOVA design is of greatest interest



-experimentally, where continuous and categorical variables are both manipulated

Random – the treatment levels represent a random sample of all possible levels – inferences valid across other levels

If you try to "shoehorn" a continuous variable into discrete categories for a 2-way ANOVA, you will need to treat that factor as **random** rather than **fixed**

Fixed – the levels included are the ONLY levels of interest

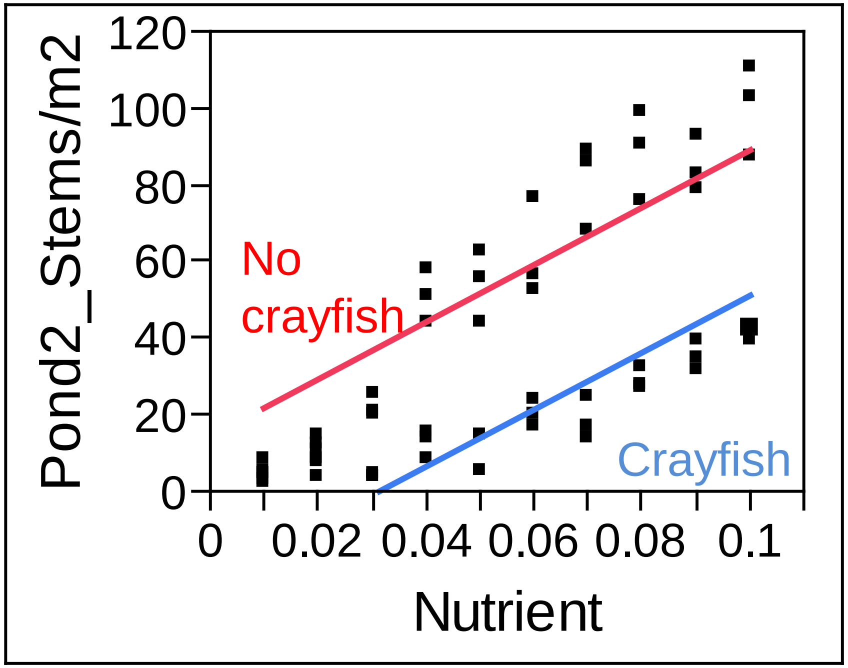
See: Gotelli & Ellison P. 199, Footnote 10

ANCOVA

$Y_{ij} = \mu + A_i + \beta_i(X_{ij} - X_i) + \varepsilon_{ij}$

May be used in a range of situations, e.g.,

-to explain "extra" variation in experiment where the ANOVA design is of greatest interest



-experimentally, where continuous and categorical variables are both manipulated

If you try to "shoehorn" a continuous variable into discrete categories for a 2-way ANOVA, you will need to treat that factor as **random** rather than **fixed**

Random – the treatment levels represent a random sample of all possible levels – inferences valid across other levels

Fixed – the levels included are the ONLY levels of interest

Who cares?!

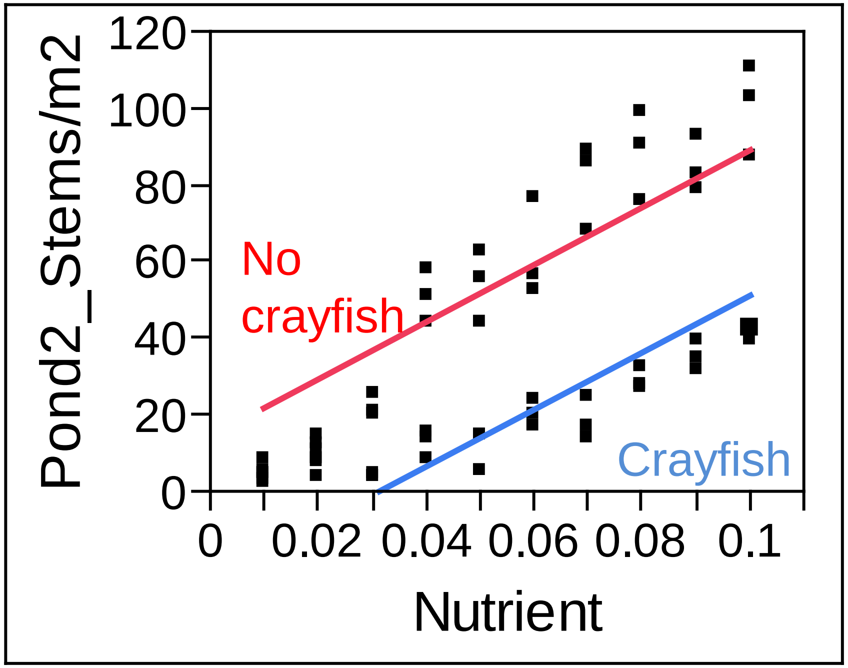
See: Gotelli & Ellison P. 199, Footnote 10

ANCOVA

$Y_{ij} = \mu + A_i + \beta_i(X_{ij} - X_i) + \varepsilon_{ij}$

May be used in a range of situations, e.g.,

-to explain "extra" variation in experiment where the ANOVA design is of greatest interest



-experimentally, where continuous and categorical variables are both manipulated

If you try to "shoehorn" a continuous variable into discrete categories for a 2-way ANOVA, you will need to treat that factor as **random** rather than **fixed**

Random – the treatment levels represent a random sample of all possible levels – inferences valid across other levels

Fixed – the levels included are the ONLY levels of interest

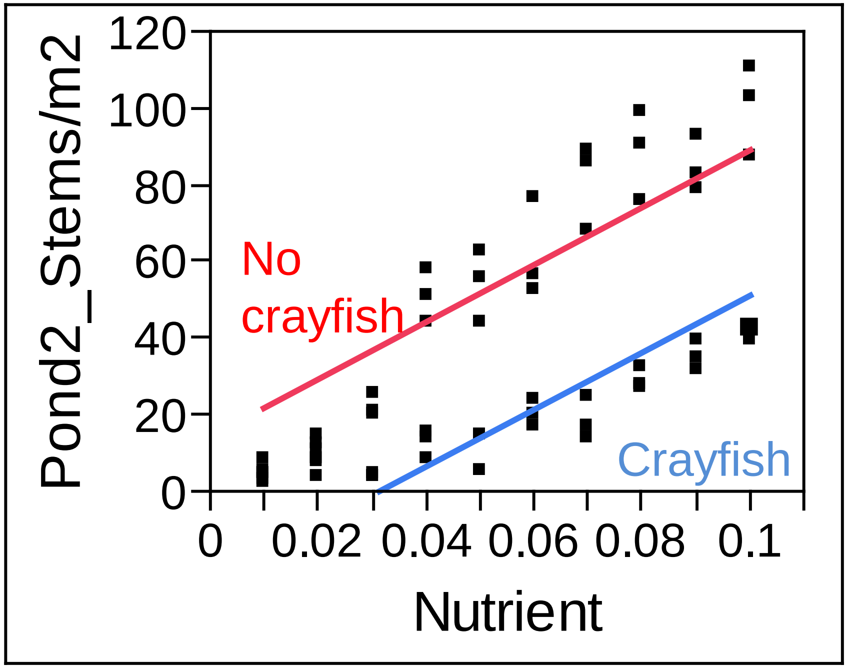
Who cares?! Calculations differ, significance values change dramatically

Why is it important to identify fixed vs. random factors?

Constructing F-values:

Fixed effect:

$$\frac{\text{Mean square}_{factor}}{\text{Mean square error}}$$



Random – the treatment levels represent a random sample of all possible levels – inferences valid across other levels

Fixed – the levels included are the ONLY levels of interest

Who cares?! Calculations differ, significance values change dramatically

Why is it important to identify fixed vs. random factors?

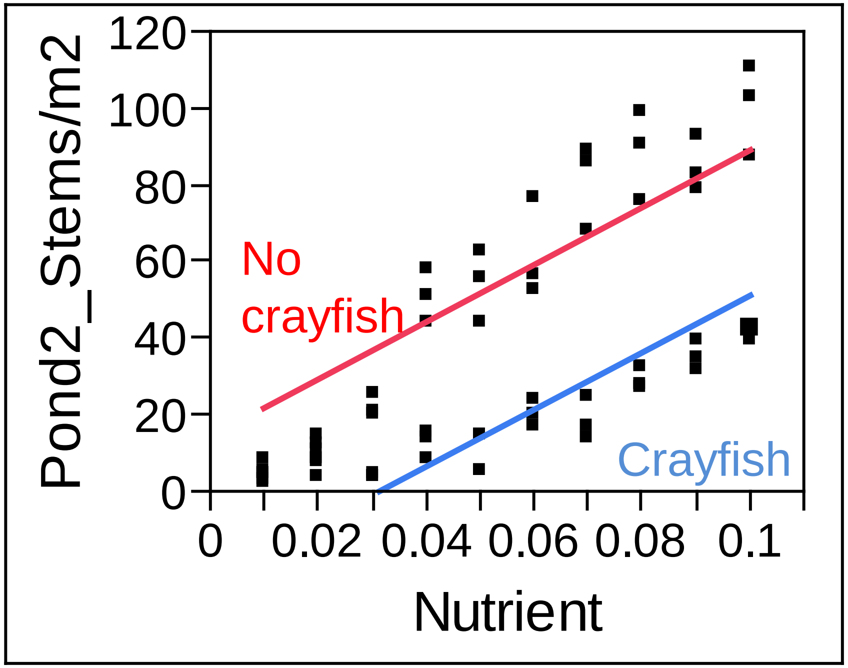
Constructing F-values:

Fixed effect:

$$\frac{\text{Mean square}_{factor}}{\text{Mean square error}}$$

Random effect:

$$\frac{\text{Mean square}_{factor}}{\text{Mean square}_{interaction\ term}}$$



Random – the treatment levels represent a random sample of all possible levels – inferences valid across other levels

Fixed – the levels included are the ONLY levels of interest

Who cares?! Calculations differ, significance values change dramatically

Neter et al. p. 982
Gotelli & Ellison p. 319

Why is it important to identify fixed vs. random factors?

Constructing F-values:

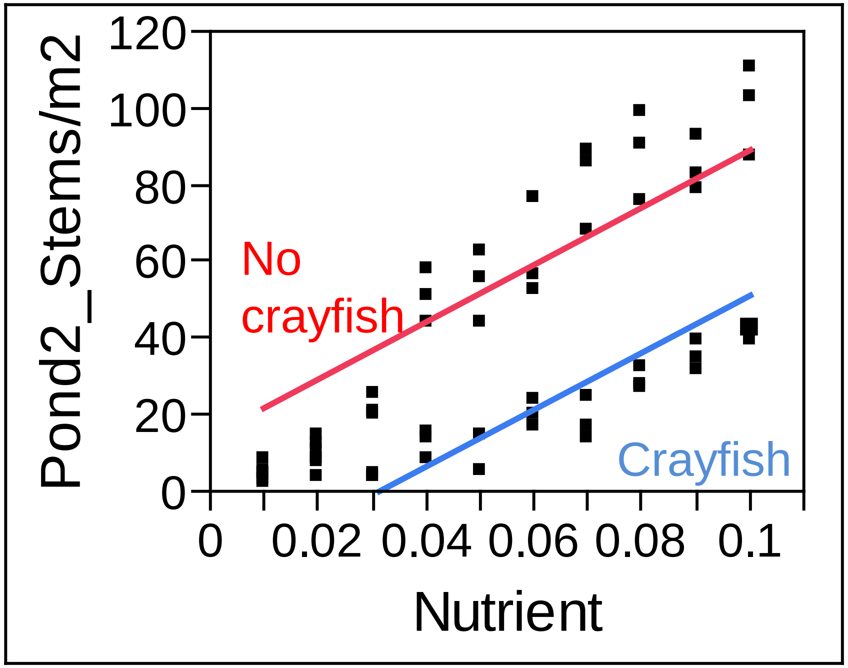
Fixed effect:

$$\frac{\text{Mean square}_{\text{factor}}}{\text{Mean square error}}$$

Random effect:

$$\frac{\text{Mean square}_{\text{factor}}}{\text{Mean square}_{\text{interaction term}}}$$

Mean $\text{square}_{\text{interaction term}}$ is strongly affected by # levels and not by *n* – usually a bigger number, so F-value smaller (less likely to be significant)



Random – the treatment levels represent a random sample of all possible levels – inferences valid across other levels

Fixed – the levels included are the ONLY levels of interest

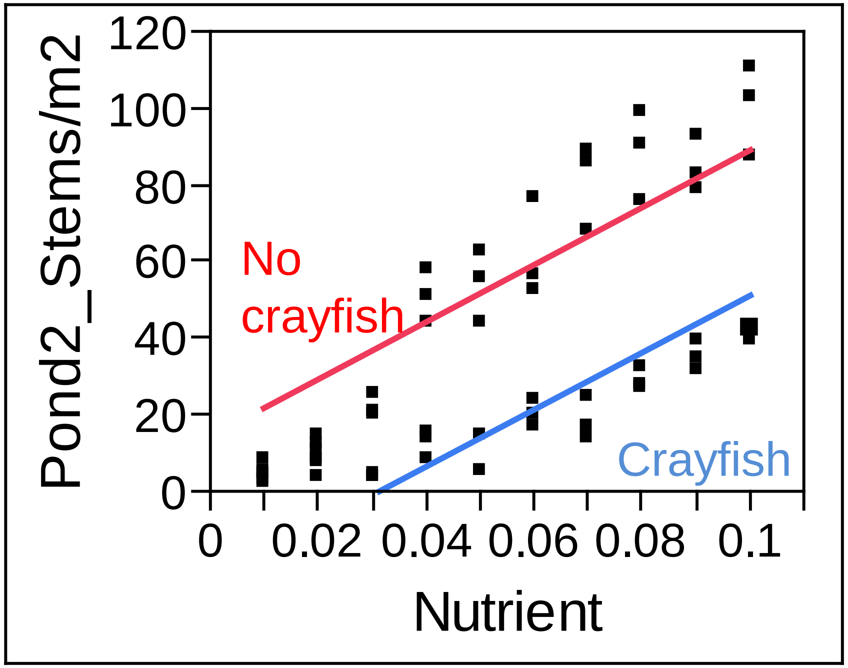
Who cares?! Calculations differ, significance values change dramatically

Moral:

Do regression when regression makes sense

Do ANOVA when ANOVA makes sense

Do ANCOVA when ANCOVA makes sense



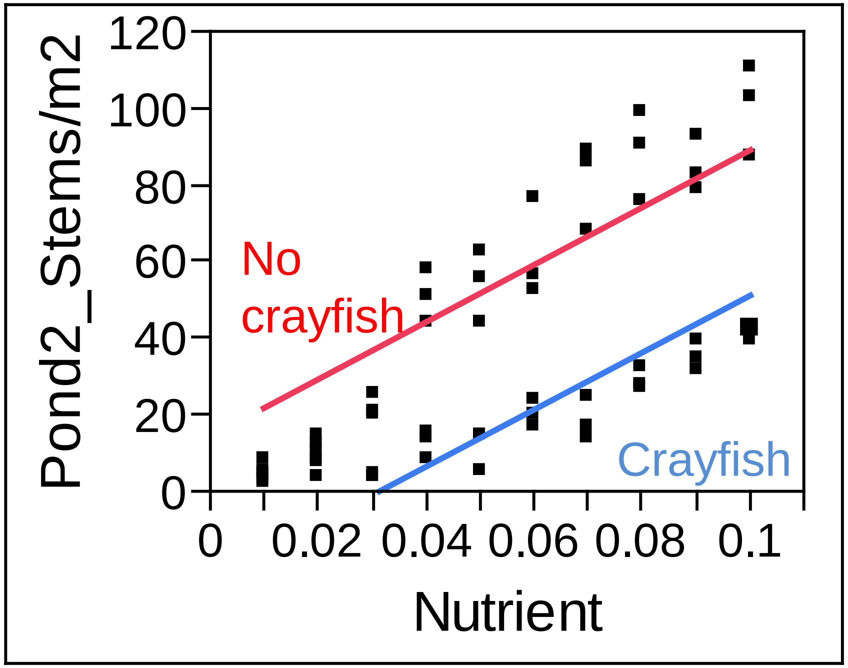
Moral:

Do regression when regression makes sense

Do ANOVA when ANOVA makes sense

Do ANCOVA when ANCOVA makes sense

Think about whether effects are random or fixed BEFORE you do your study – it may change your sampling approach



Random – **the treatment levels represent a random sample of all possible levels – inferences valid across other levels**

Fixed – **the levels included are the ONLY levels of interest**

Who cares?! **Calculations differ, significance values change dramatically**

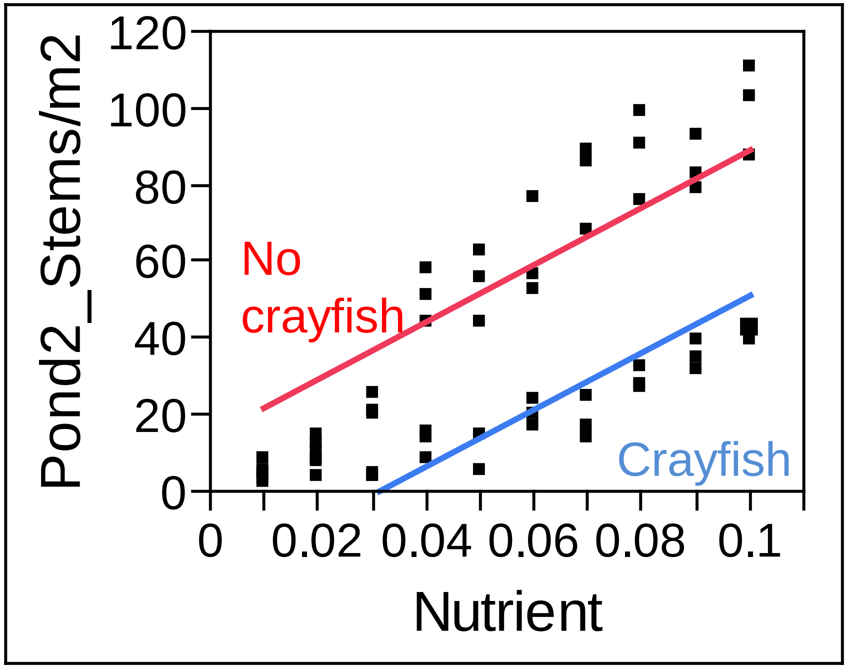
See: Gotelli & Ellison P. 319

More resources

ANCOVA, fixed/random effects:

Gotelli & Ellison. A Primer of Ecological Statistics. Sinauer.

Neter et al. Applied Linear Statistical Models. McGraw Hill.



Fixed/random effects in designing experiments with regression:
Cottingham et al (2005) Knowing when to draw the line: designing more informative ecological experiments. Frontiers in Ecology and the Environment. 3(3): 145–152
http://microbes.kbs.msu.edu/Publications/Cottingham%20et%20al.%202005.pdf

More about what fixed and random effects are, especially for ecologists:
Bennington & Thayne. 1994. Use and Misuse of Mixed Model Analysis of Variance in Ecological Studies. Ecology 75(3): 717-722.
http://www.jstor.org/stable/1941729