

30 March 2009

UCSB ESM 206B

Stephanie Hampton

National Center for Ecological Analysis & Synthesis

Contact for appointment – [hampton@nceas.ucsb.edu](mailto:hampton@nceas.ucsb.edu)

Lectures & labs

Readings posted to website

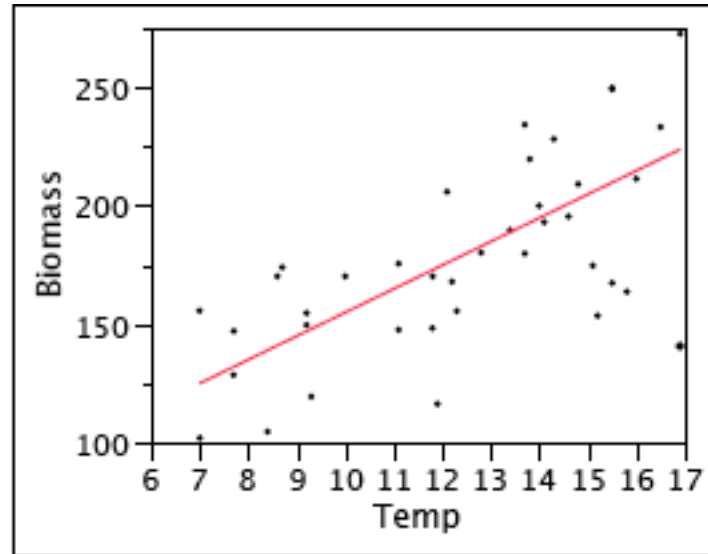
Lab assignments & take-home “micro-exam”

- 30 March Review Multiple Regression
- 1 April Model Selection in Multiple Regression
- 6-8 April Nonlinear Regression
- 13-15 April Logistic Regression
- 20-22 April Bootstrapping & Monte Carlo
- 27-29 April Similarity Metrics for Multivariate Stats

Suggestions?

# 1. Simple regression

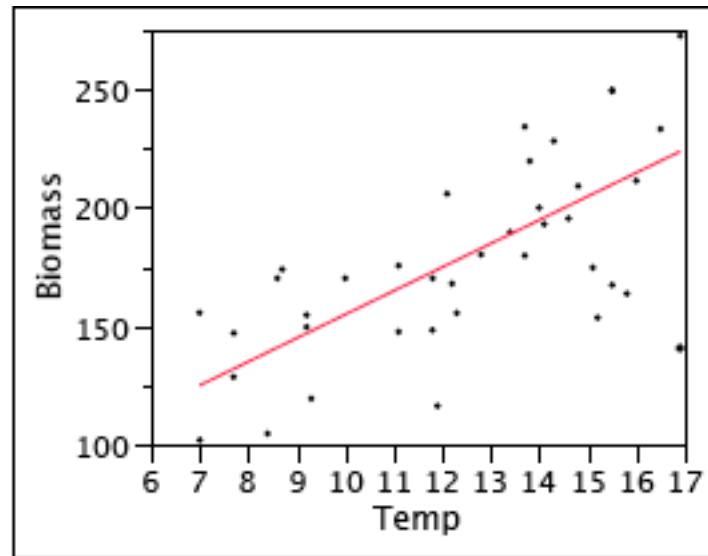
$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$



# 1. Simple regression

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

e.g.



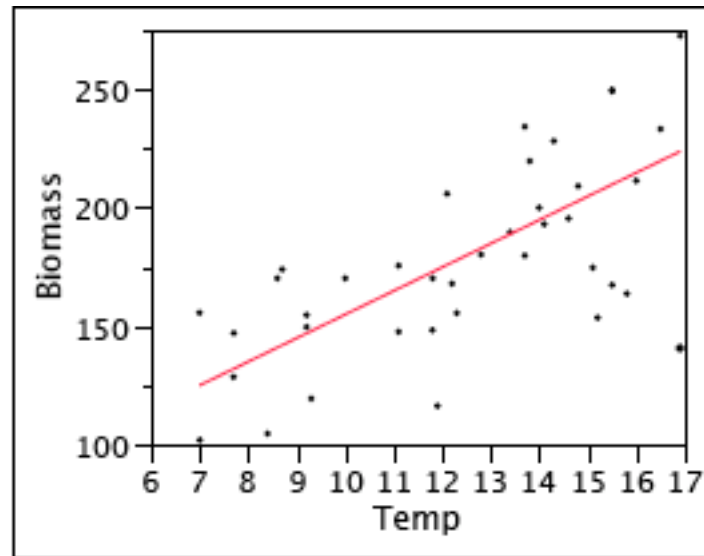
$R^2 = 0.52$   
 $P < 0.001$

$$\text{Log}(\text{Biomass, mg/L}) = -41.49 + 9.02(\text{temperature, C})$$

# 1. Simple regression

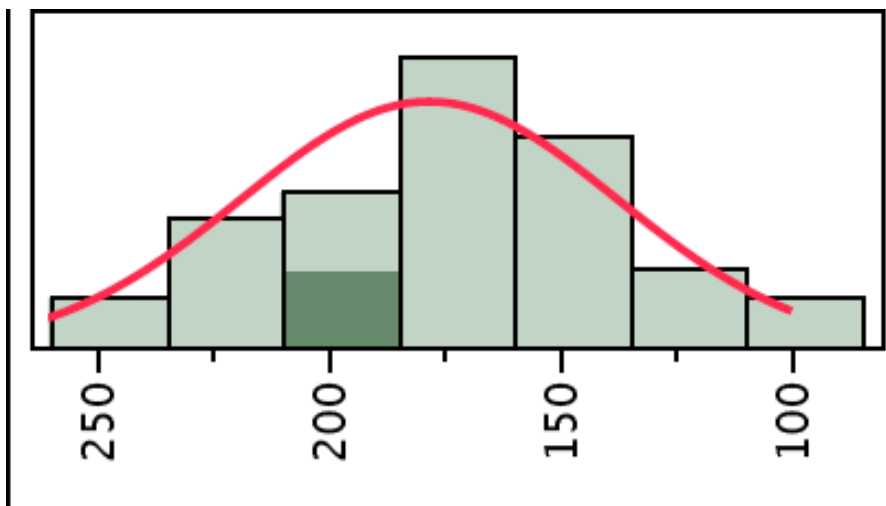
$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

e.g.



$R^2 = 0.52$   
 $P < 0.001$

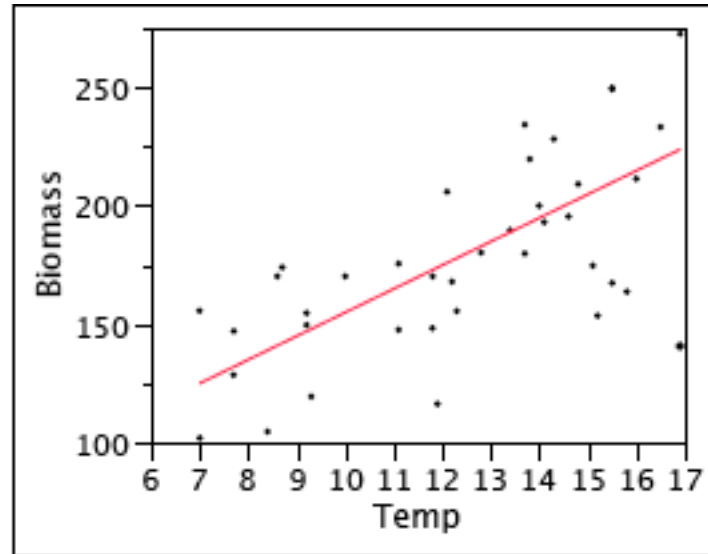
$$\text{Log}(\text{Biomass, mg/L}) = -41.49 + 9.02(\text{temperature, C})$$



# 1. Simple regression

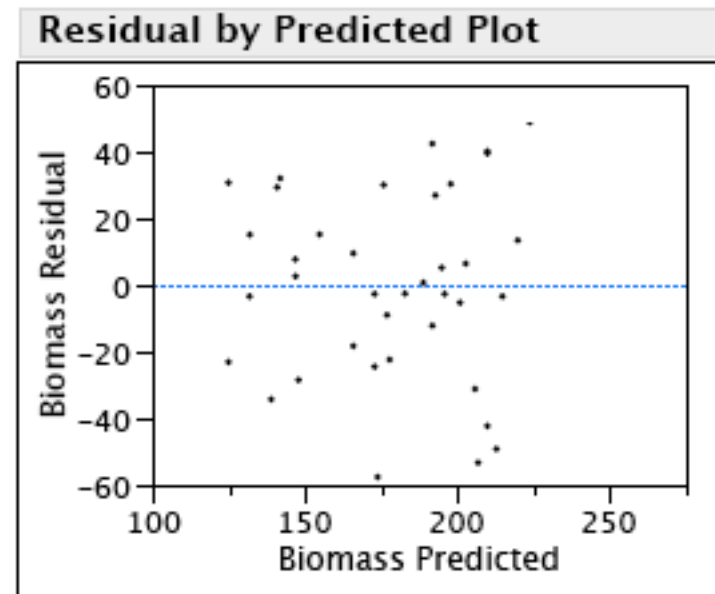
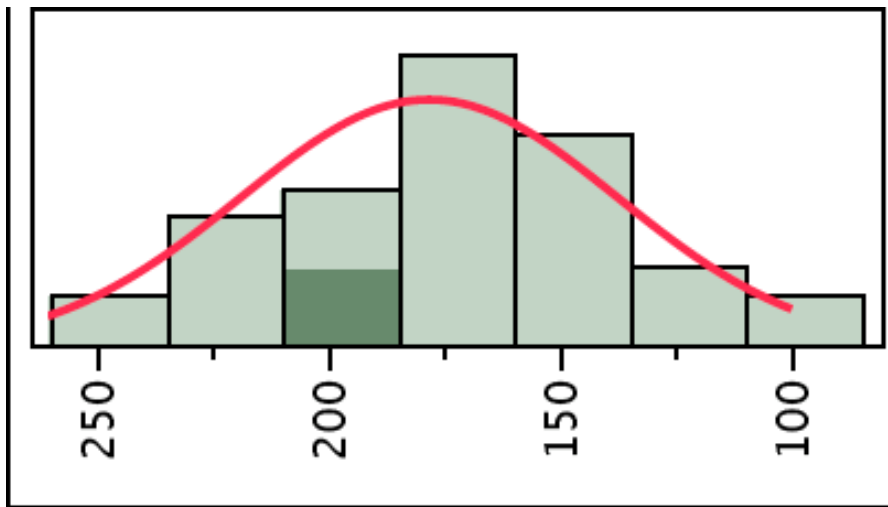
$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

e.g.

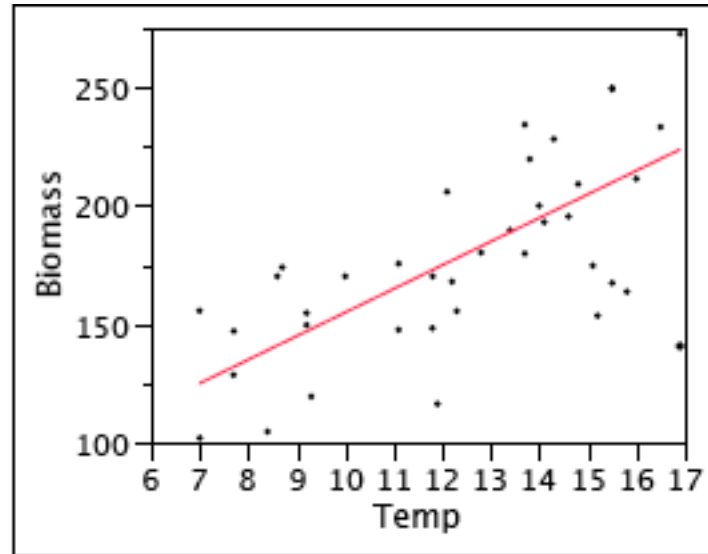


$$R^2 = 0.52$$
$$P < 0.001$$

$$\text{Log}(\text{Biomass, mg/L}) = -41.49 + 9.02(\text{temperature, C})$$

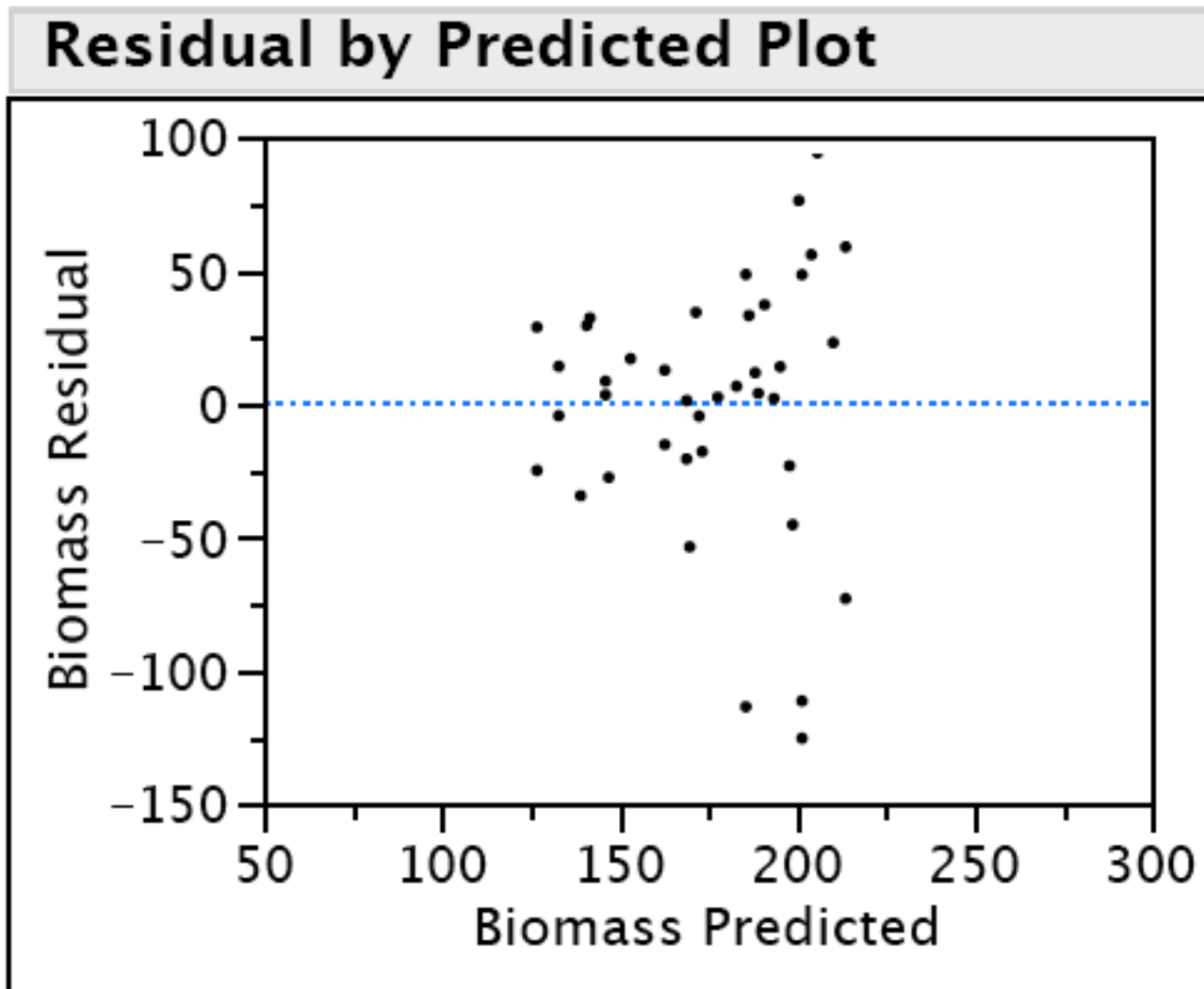


# 1a. Bad residuals?



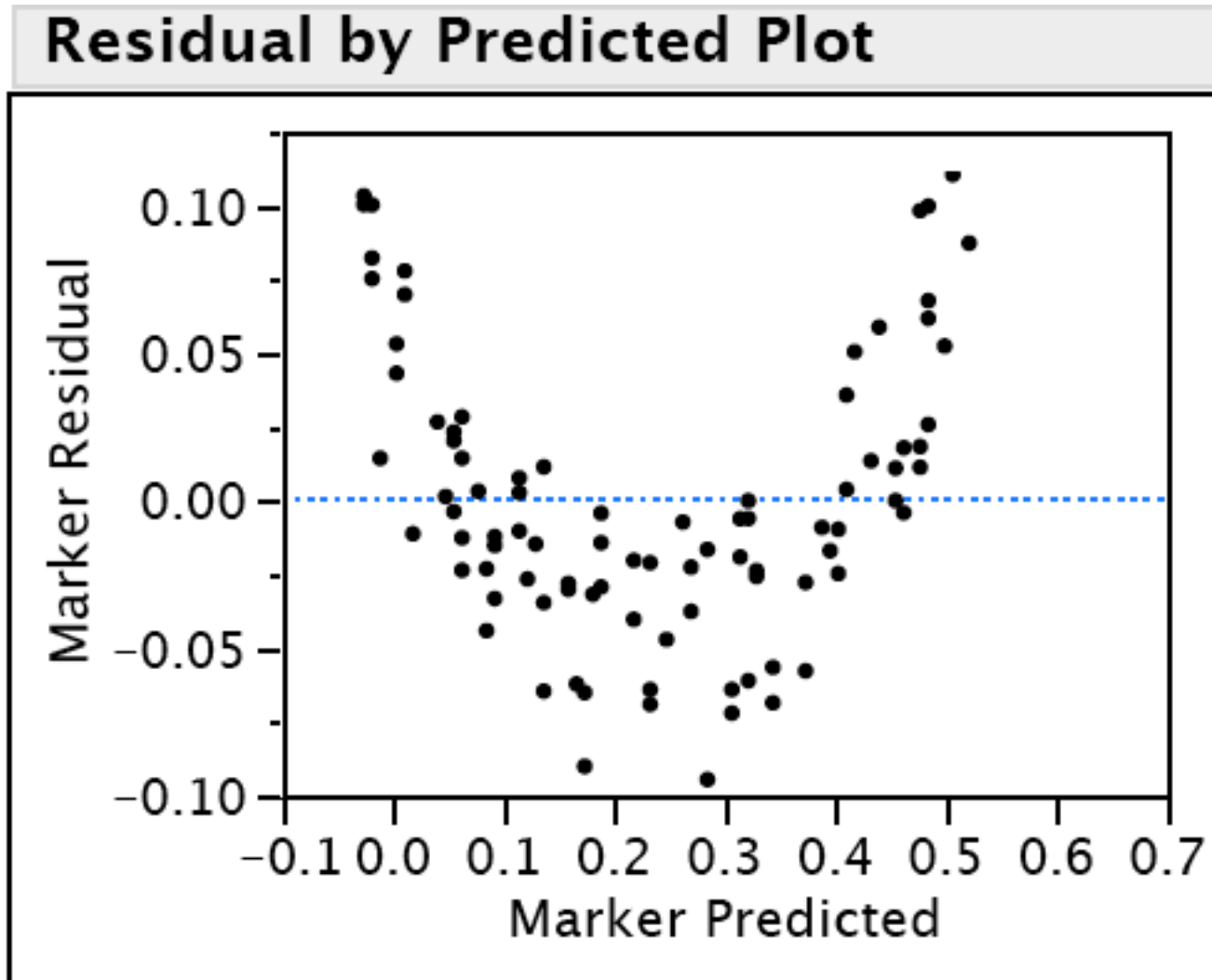
## 1a. “Bad” residuals?

Pattern indicates non-independence or changing variation across gradient



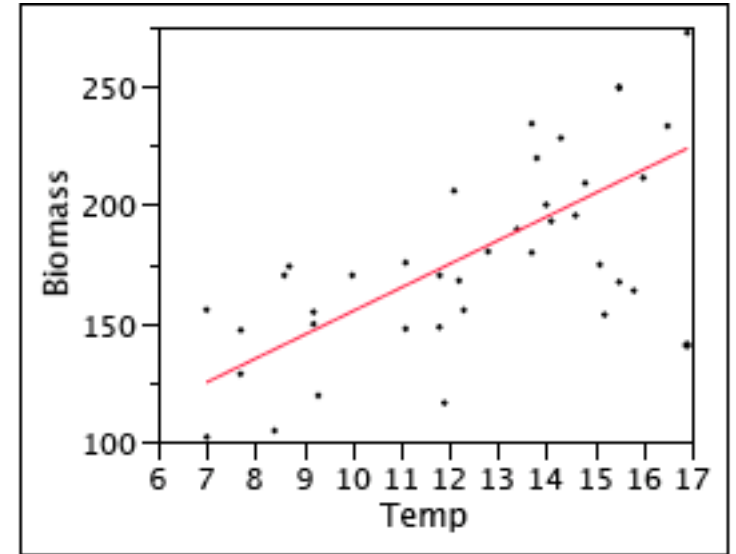
# 1a. "Bad" residuals?

Pattern indicates non-independence or changing variation across gradient



1b. Where do parameters come from?

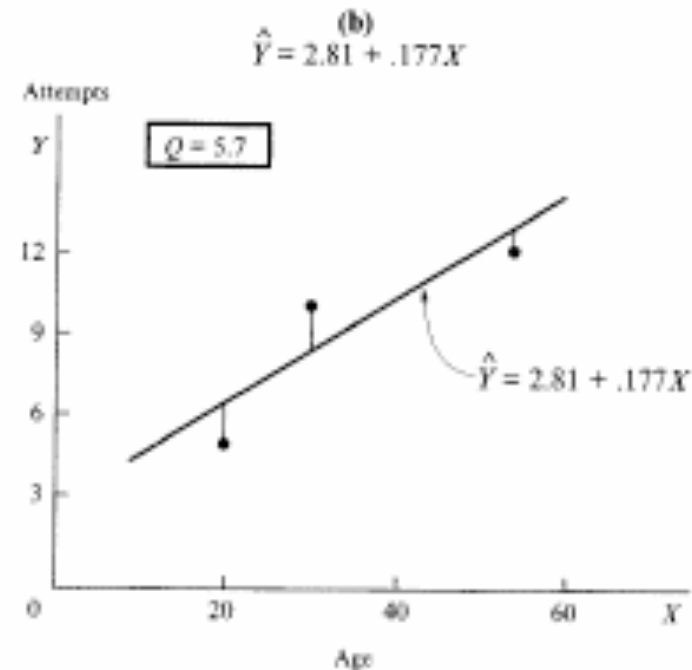
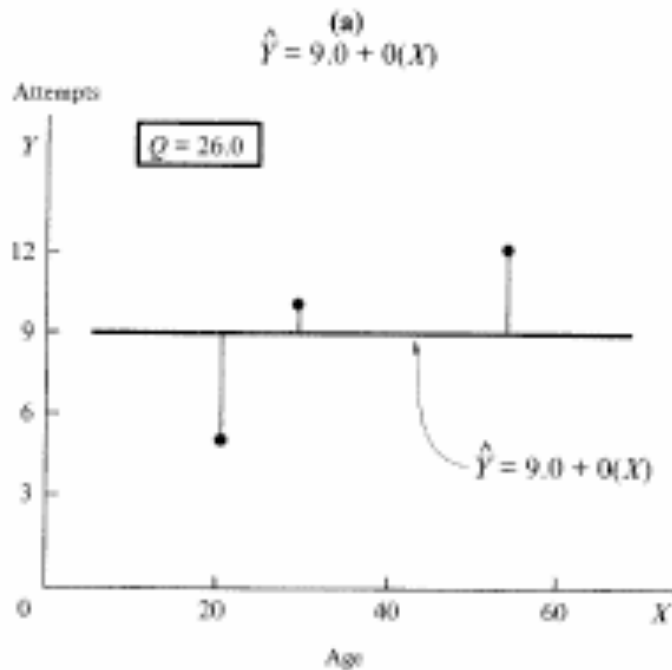
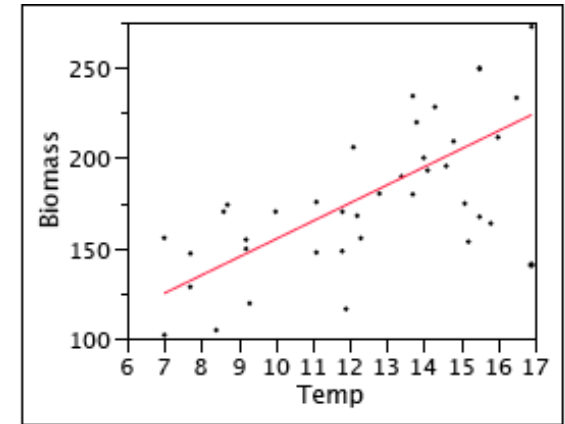
$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$



# 1b. Where do parameters come from?

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

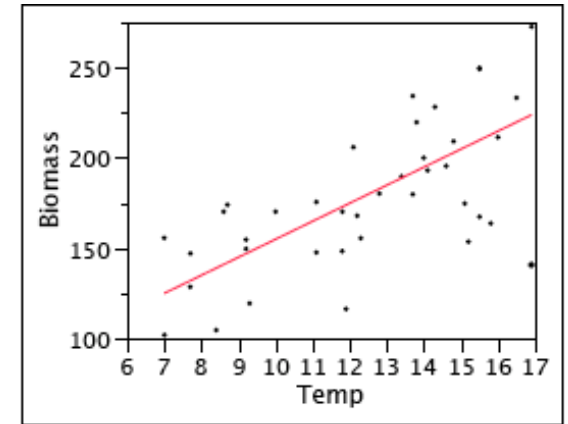
$$SS = \sum(Y_i - \beta_0 - \beta_1 X_i)$$



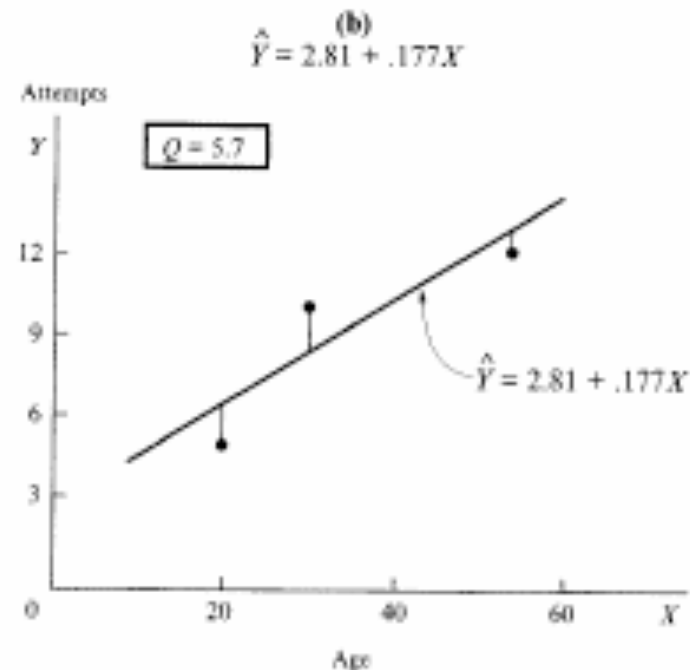
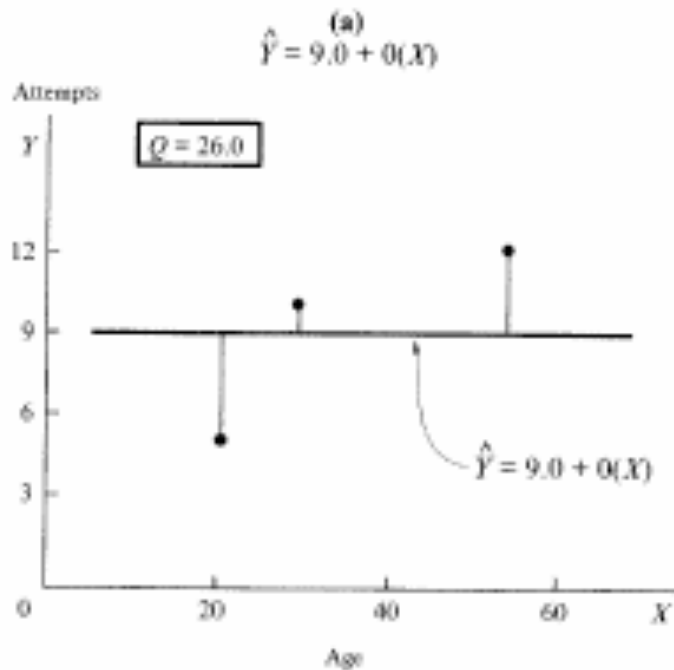
1b. Where do parameters come from?

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$SS = \sum (Y_i - \beta_0 - \beta_1 X_i)^2$$



Minimize this number = “least squares” method

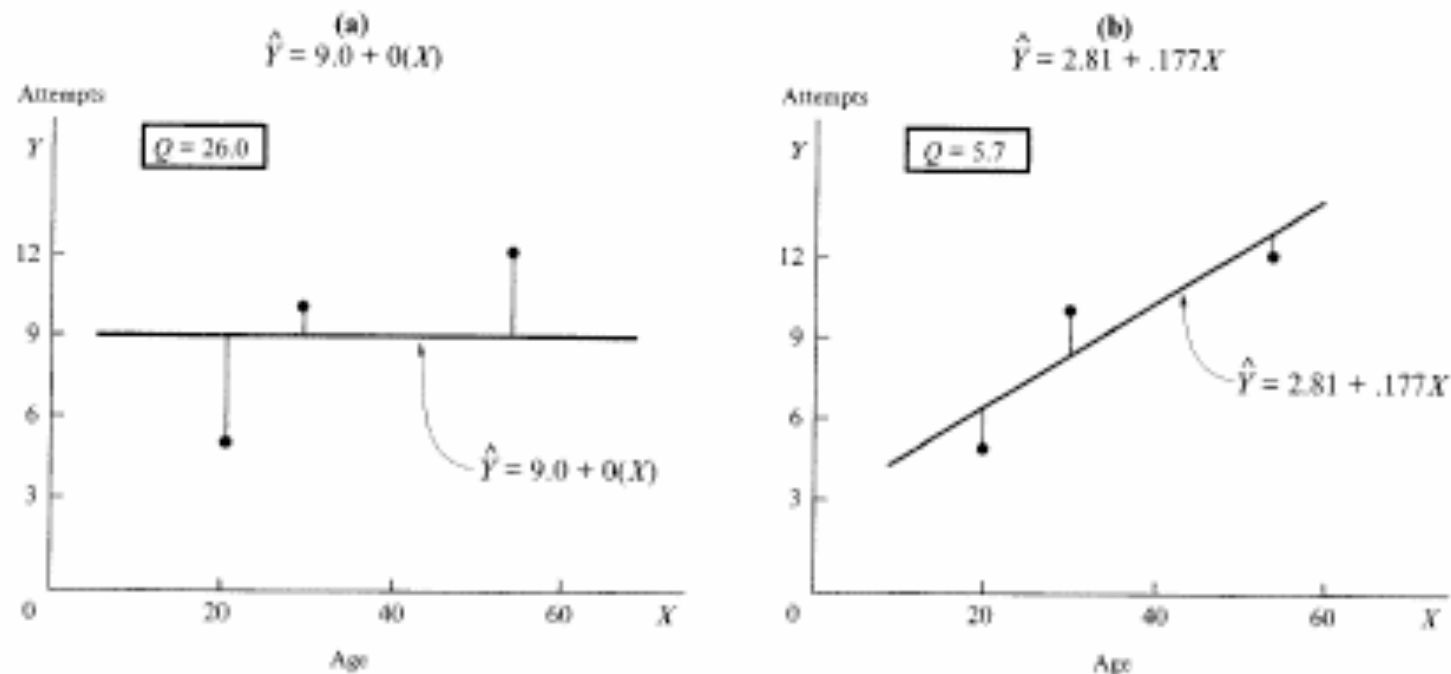


$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$SS = \sum(Y_i - \beta_0 - \beta_1 X_i)^2$$

Minimize this number = “least squares” method

- 1) Computational – iteration
- 2) Analytical for simple models

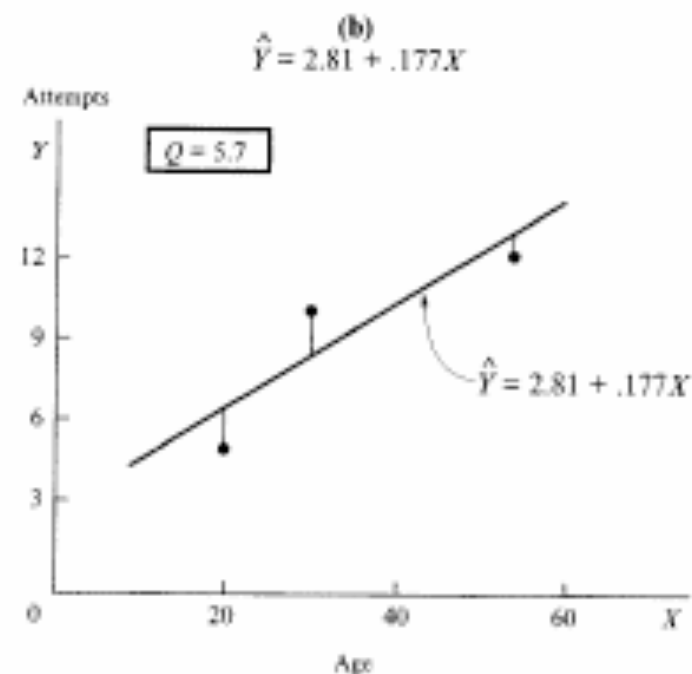
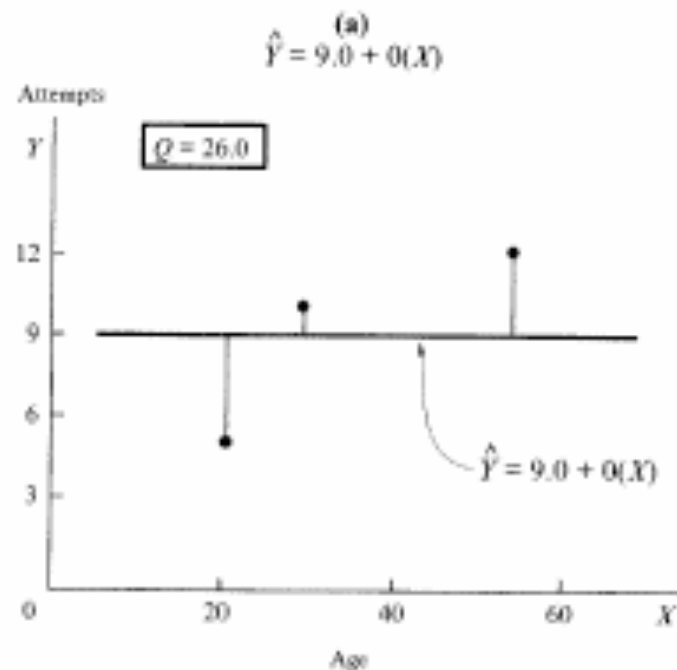


$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Analytical solution for least squares estimation of  $\beta_0$  and  $\beta_1$

$$b_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

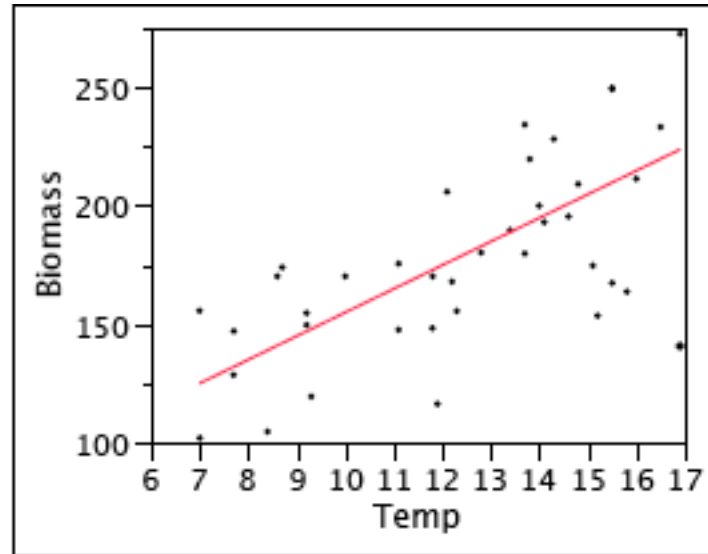
$$b_0 = \frac{1}{n} \left( \sum Y_i - b_1 \sum X_i \right) = \bar{Y} - b_1 \bar{X}$$



1b. Least squares

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

e.g.



$$\text{Log}(\text{Biomass, mg/L}) = -41.49 + 9.02(\text{temperature, C})$$

2. Multiple regression (1<sup>st</sup> order = no interactions)

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \varepsilon_i$$

## 2. Multiple regression (1<sup>st</sup> order = no interactions)

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \varepsilon_i$$

e.g.

Log(Biomass, mg/L) =

$$-41.49 + 9.02(\text{temperature, C}) + 7.10(\text{Phosphorus, ug/L})$$

## 2. Multiple regression (1<sup>st</sup> order = no interactions)

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \varepsilon_i$$

e.g.

Log(Biomass, mg/L) =

$$-41.49 + 9.02(\text{temperature, C}) + 7.10(\text{Phosphorus, ug/L})$$

Partial Regression Parameters



2. Multiple regression (1<sup>st</sup> order = no interactions)

$$Y_i = \beta_0 + \beta_{1,i} X_1 + \beta_{2,i} X_2 + \varepsilon_i$$

Log(Biomass, mg/L) =

$$-41.49 + 9.02(\text{temperature, C}) + 7.10(\text{Phosphorus, ug/L})$$

Partial Regression Parameters

Parameters differ from simple regression

2. Multiple regression (1<sup>st</sup> order = no interactions)

$$Y_i = \beta_0 + \beta_{1,i} X_1 + \beta_{2,i} X_2 + \varepsilon_i$$

Log(Biomass, mg/L) =

$$-41.49 + 9.02(\text{temperature, C}) + 7.10(\text{Phosphorus, ug/L})$$

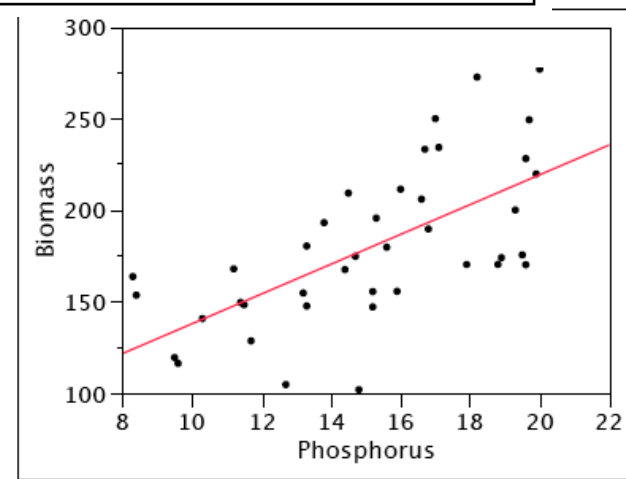
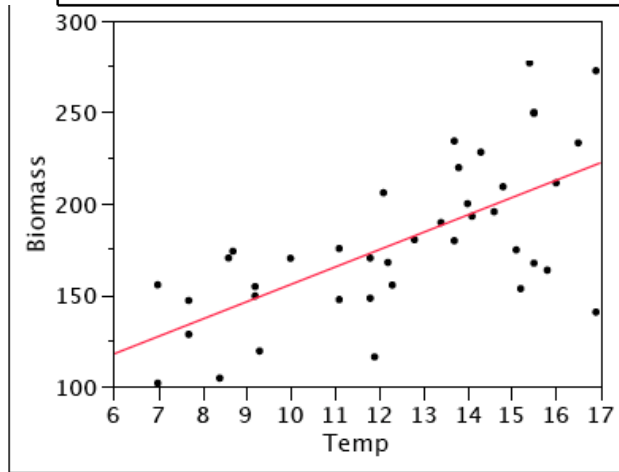
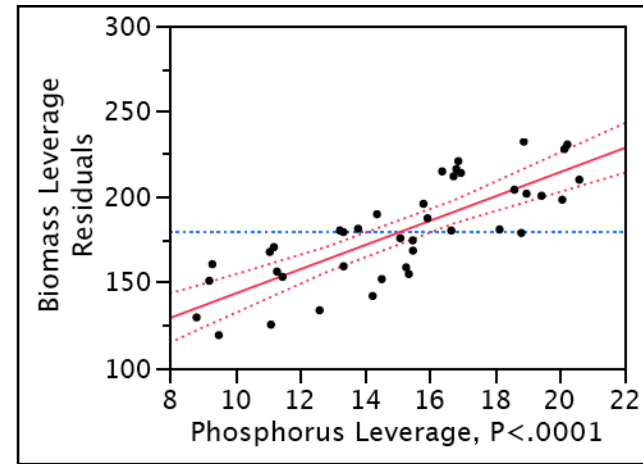
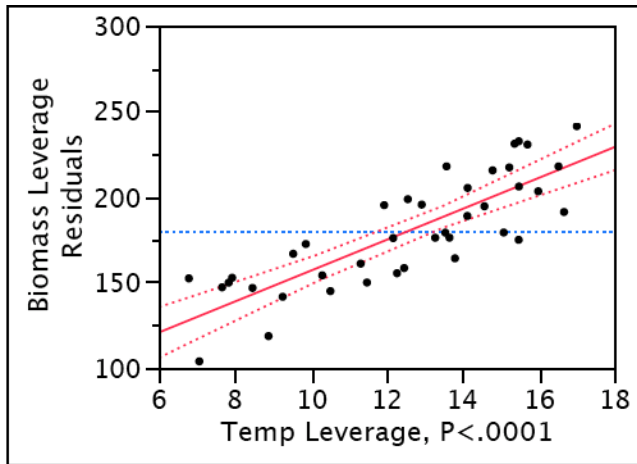
Partial Regression Parameters

Parameters differ from simple regression

$$\text{Log(Biomass, mg/L)} = 60.72 + 9.49(\text{temperature, C})$$

$$\text{Log(Biomass, mg/L)} = 56.44 + 8.14(\text{Phosphorus, ug/L})$$

$$-41.49 + 9.02(\text{temperature, C}) + 7.10(\text{Phosphorus, ug/L})$$



$$\text{Log}(\text{Biomass, mg/L}) = 60.72 + 9.49(\text{temperature, C})$$

$$\text{Log}(\text{Biomass, mg/L}) = 56.44 + 8.14(\text{Phosphorus, ug/L})$$

### 3. Multicollinearity – correlated $X_n$ variables

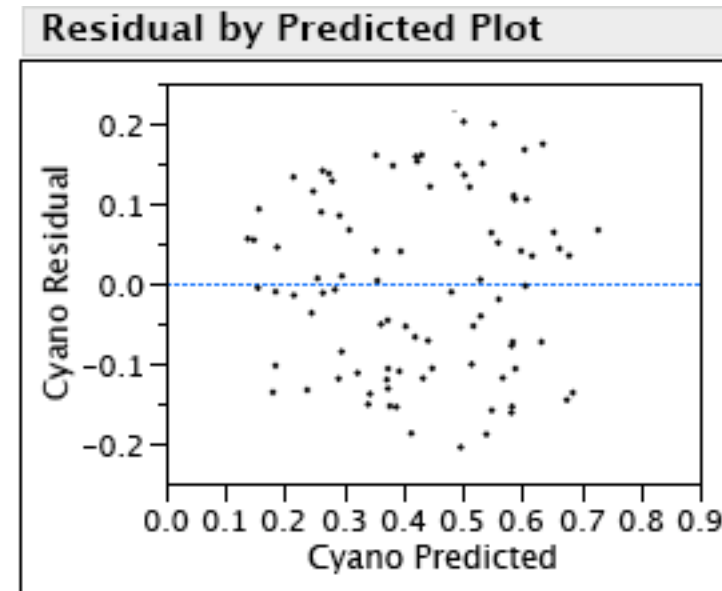
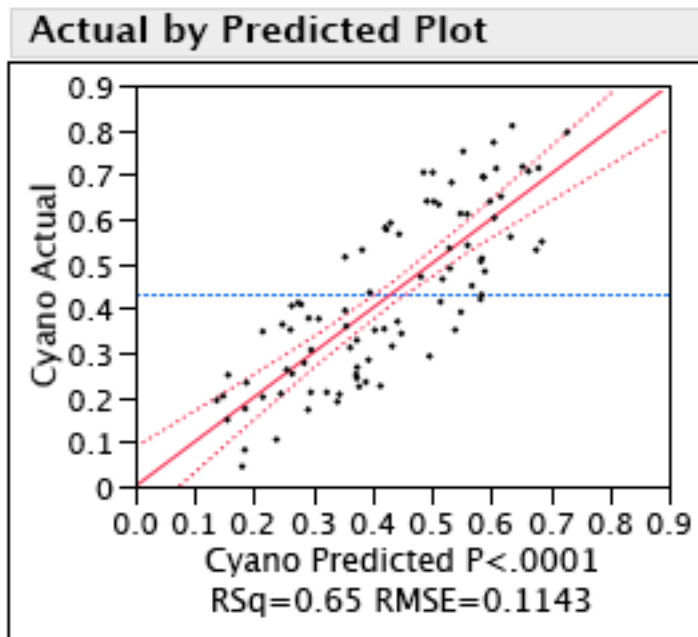
$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \varepsilon_i$$

3. Multicollinearity – correlated  $X_n$  variables

$$\text{Cyanobacteria} = \beta_1 \text{Phosphorus} + \beta_2 \text{Caffeine}$$

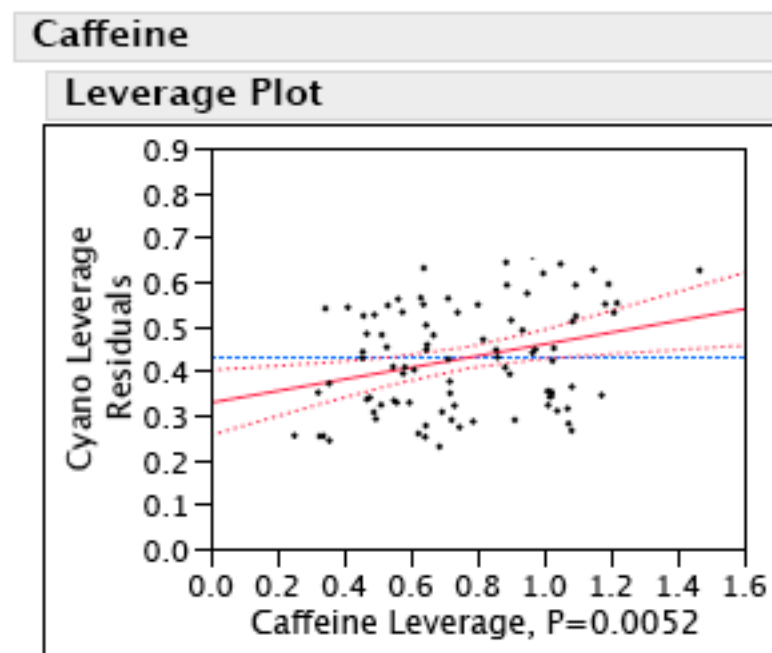
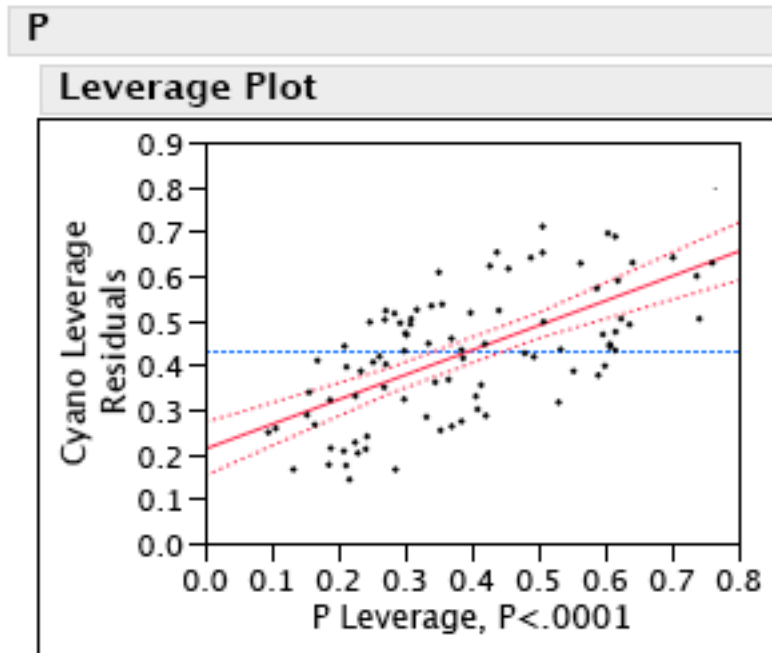
### 3. Multicollinearity – correlated $X_n$ variables

$$\text{Cyanobacteria} = \beta_1 \text{Phosphorus} + \beta_2 \text{Caffeine}$$



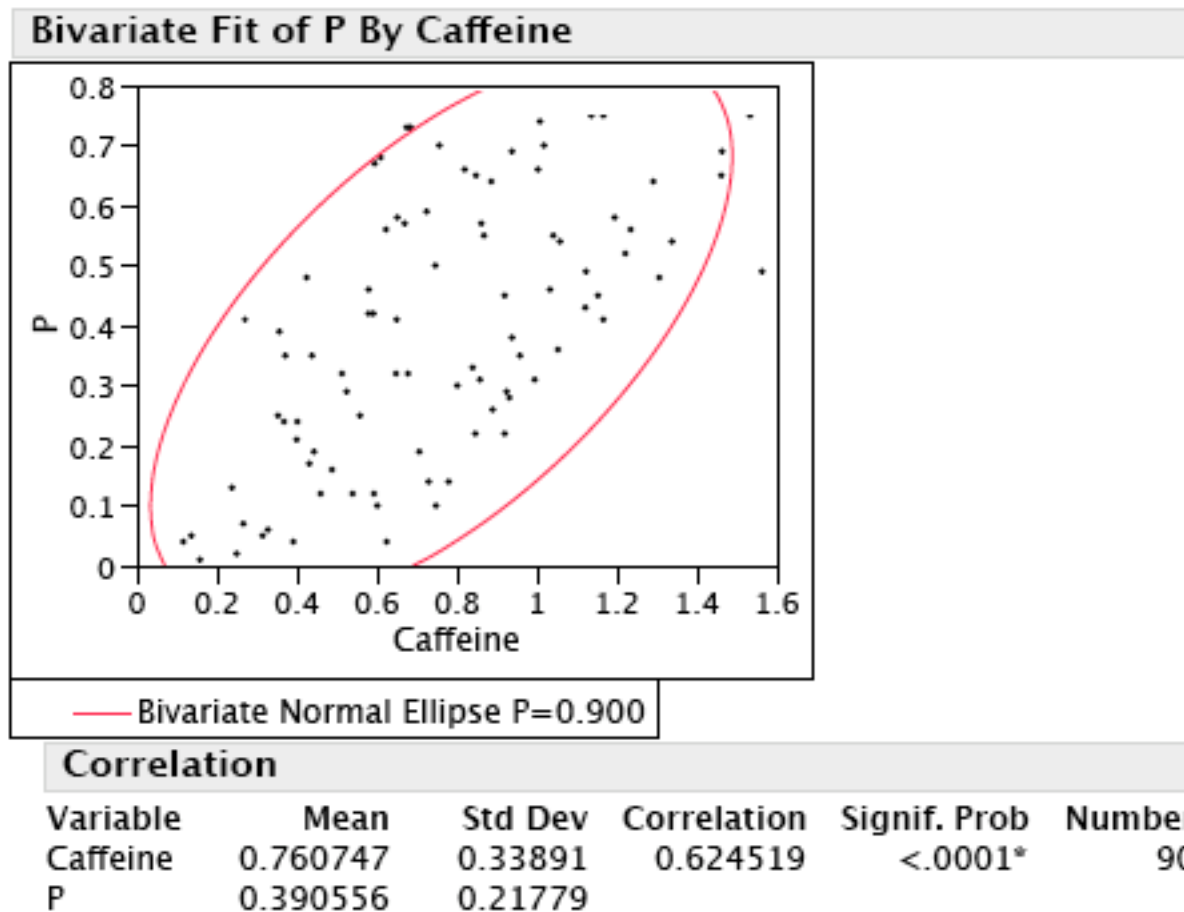
### 3. Multicollinearity – correlated $X_n$ variables

$$\text{Cyanobacteria} = \beta_1 \text{Phosphorus} + \beta_2 \text{Caffeine}$$



### 3. Multicollinearity – correlated $X_n$ variables

$$\text{Cyanobacteria} = \beta_1 \text{Phosphorus} + \beta_2 \text{Caffeine}$$



### 3. Multicollinearity – correlated $X_n$ variables

$$\text{Cyanobacteria} = \beta_1 \text{Phosphorus} + \beta_2 \text{Caffeine}$$

Generally increases parameter estimates

Confounds interpretation