Mean Comparisons
Approaches

<u>Qualitative</u> – classification variables
- Multiple Comparison Procedures
- Contrasts

<u>Quantitative</u> – numerical variables
- Orthogonal Polynomial Contrasts
- Curve Fitting

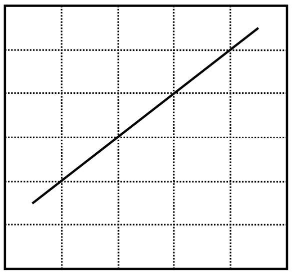
Regression
Polynomial Equations

$$Y = b_0 + b_1x + b_2x^2 + ... + b_nx^n$$

Where:
- b_0 = intercept
- b_n = regression coefficient for x^n

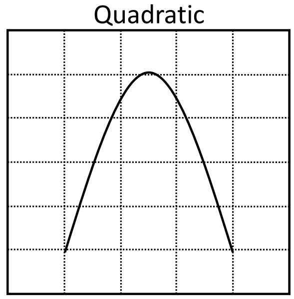
Regression
Polynomial Equations

Linear



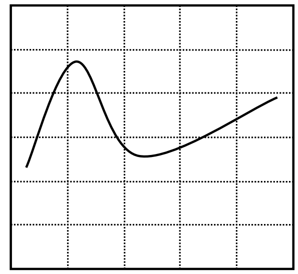
$Y = b_0 + b_1x$

Quadratic



$Y = b_0 + b_1x - b_2x^2$

Cubic



$Y = b_0 + b_1x - b_2x^2 + b_3x^3$

Contrasts
Orthogonal Polynomials

Example: Nitrogen Fertilizer Rate Study

Rates:

0 lbs / acre

50

100

150 Note equal spacing of fertilizer rates

200

Contrasts

Orthogonal Polynomials

Source	df	Coefficients				
Treatment	4	0	50	100	150	200
linear	1	-2	-1	0	1	2
quadratic	1	2	-1	-2	-1	2
cubic	1	-1	2	0	-2	1
quartic	1	1	-4	6	-4	1
Error	15					
Total	19					

* Orthogonal polynomial coefficients are found in Appendix 5 of Lorenzen and Anderson

Orthogonal Polynomial Contrasts

Nitrogen Example

$$SS(Q) = MS(Q) = \frac{r(\sum c_i \overline{Y}_i)^2}{\sum c_i^2}$$

For the linear effect:

$$MS_Q = \frac{4(-2(359.08) - 1(9631.20) + 0(1831.36) + 1(2482.89) + 2(3055.98))^2}{(-2)^2 + (-1)^2 + (0)^2 + (1)^2 + (2)^2}$$

$$MS_Q = \frac{4(6913.495)^2}{10} = 1918565$$

$$F = \frac{19118567}{44831} = 427$$ P > F = < 0.0001

Orthogonal Polynomial Contrasts

Nitrogen Example

ANOVA

Source	DF	Squares	Mean Square	F Value	Pr > F
Rate	4	19209561	4802390.14	107.12	<.0001
linear	1	19118565	19118565.25	426.46	**<.0001**
quad	1	22191.69	22191.69	0.5	0.4925
cubic	1	46916.34	46916.34	1.05	0.3225
quart	1	21887.31	21887.31	0.49	0.4954
Error	15	672467.2	44831.15		
Total	19	19882028			

Polynomial Equations

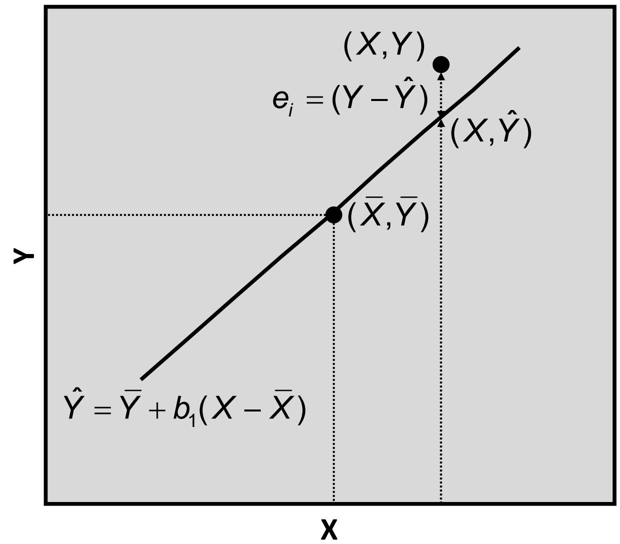
Linear Model

$$Y = b_0 + b_1 X$$

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

$$b_0 = \overline{Y} - b_1 \overline{X}$$

Polynomial Equations
Linear Model



Linear Regression
Nitrogen Example

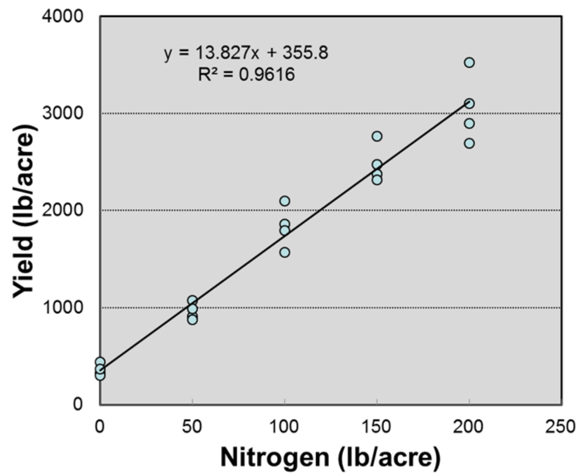
$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{1382699}{100000} = 13.83$$

$$b_0 = \bar{Y} - b_1\bar{X} = 1738.504 - 13.826 * 100 = 355.81$$

$$Y = b_0 + b_1 X = 355.81 + 13.83X$$

Regression

Nitrogen Example



Regression

ANOVA

Source	df	Sum of Squares*	MS	F
Total	n - 1	$\sum(Y_{ij} - \bar{Y})^2$		
Regression	1	$\sum(\hat{Y}_i - \bar{Y})^2$	$\frac{SS_{Reg}}{df_{Reg}}$	$\frac{MS_{Reg}}{MS_{Res}}$
Residual	n – 2	$\sum(Y_{ij} - \hat{Y}_i)^2$	$\frac{SS_{Res}}{df_{Res}}$	
Lack of Fit	t – 2	$\sum(\bar{Y}_{ij} - \hat{Y}_i)^2$	$\frac{SS_{LOF}}{df_{LOF}}$	$\frac{MS_{LOF}}{MS_{PE}}$
Pure Error	n - t	$\sum(Y_{ij} - \bar{Y}_i)^2$	$\frac{SS_{PE}}{df_{PE}}$	

* These formulas are incomplete. They are intended to illustrate which differences are being used to calculate the SS for each source of variation.

Regression

ANOVA

Source	df	Sum of Squares	MS	F	P > F
Total	19	19882028			
Regression	1	19118565	19118565	450.75	<0.0001
Residual	18	763463	42415		
Lack of Fit	3	90995	30332	0.68	0.579
Pure Error	15	672467	44831		

Evaluating Regression Equations

Coefficient of Determination

$$R^2 = \frac{SS_{Reg}}{SS_{Total}} = \frac{19118565}{19882028} = .962$$

$$R^2_{Adj} = \frac{(n-1)R^2 - k}{n-1-k} = \frac{(20-1).962-1}{20-1-1} = .959$$

Where:
n = the number of observations
k = the number of estimated parameters

Evaluating Regression Equations

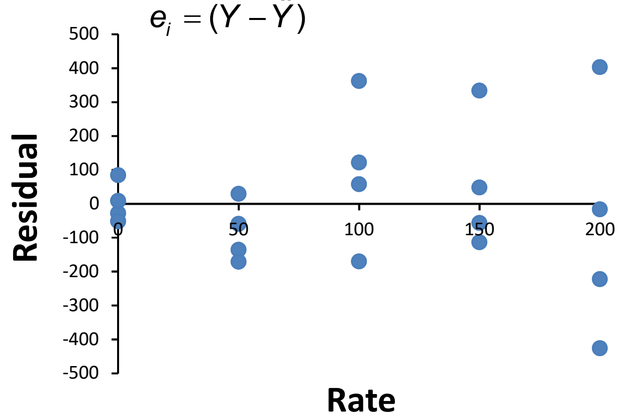
Standard Error of the Estimate

$$S_{y \cdot x} = \sqrt{\frac{SS_{RES}}{n-p}} = \sqrt{\frac{763463}{20-2}} = 205.9$$

Evaluating Regression Equations

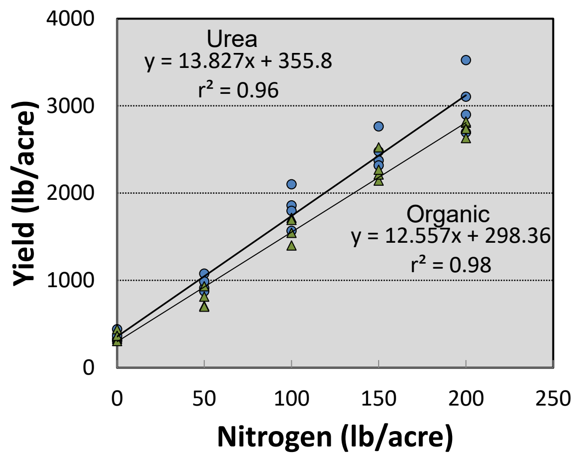
Examining Residuals

$$e_i = (Y - \hat{Y})$$



Comparing Regression Equations

Urea vs Organic N



Comparing Regression Equations

Urea vs Organic Slopes

$$t = \frac{b_1 - b_2}{S_{b1-b2}} = \frac{13.827 - 12.557}{0.794} = 1.599 < t_{0.05,36df} = 2.028$$

$\therefore$ the two slopes are not different

$$S_{b1-b2} = \sqrt{\frac{(S^2_{Y \cdot X})_p}{\sum (X - \bar{X})_1^2} + \frac{(S^2_{Y \cdot X})_p}{\sum (X - \bar{X})_2^2}} = \sqrt{\frac{31543.61}{100000} + \frac{31543.61}{100000}} = 0.794$$

$$(S^2_{Y \cdot X})_p = \frac{(residual\ SS)_1 + (residual\ SS)_2}{(residual\ DF)_1 + (residual\ DF)_2} = \frac{763463 + 372107}{18 + 18} = 31543.61$$

Comparing Regression Equations

Mixed Model Analysis

```
proc mixed;
  class source;
  model yield  = source rate*source / solution noint;
  estimate ‘Organic rate' rate*source 1;
  estimate 'Urea rate' rate*source 0 1;
  estimate ‘Organic v Urea rate' rate*source 1 -1;
run;
```

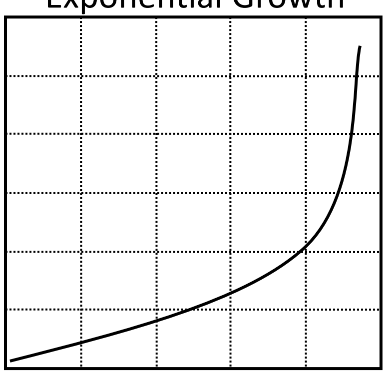
Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
Nitamin rate	12.5571	0.5616	36	22.36	<.0001
Urea rate	13.8270	0.5616	36	24.62	<.0001
Nitamin v Urea rate	-1.2699	0.7943	36	-1.60	0.1186

Nonlinear Equations

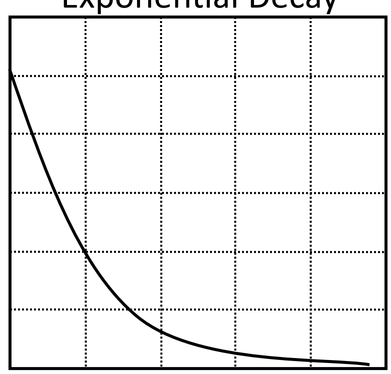
Intrinsically Linear

Exponential Growth



$$Y = ae^{bx}$$
$$\ln Y = \ln a + bx$$

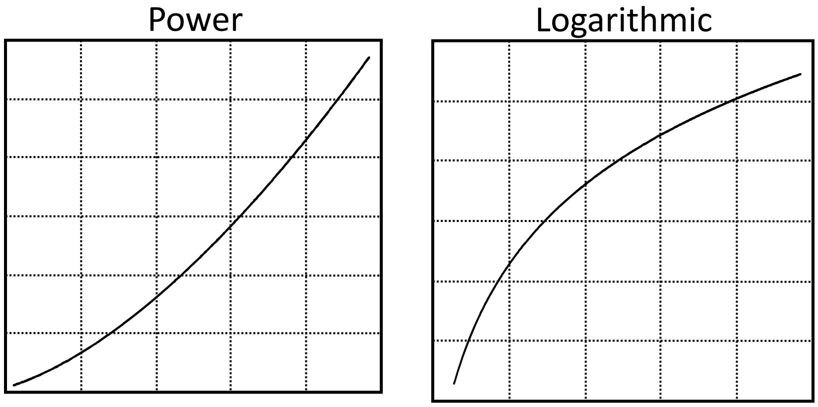
Exponential Decay



$$Y = ae^{-bx}$$
$$\ln Y = \ln a - bx$$

Nonlinear Equations

Intrinsically Linear



$Y = ax^b$

$\ln Y = \ln a + b\ln x$

$Y = a + b\ln x$

Nonlinear Equations

Intrinsically Linear – Smith's Formula

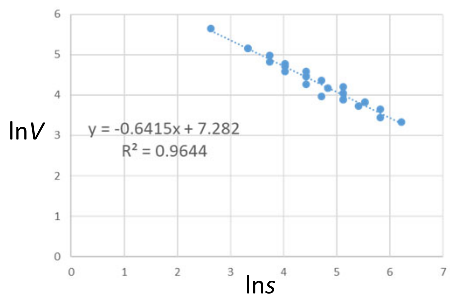
Smith's (Power) Formula:

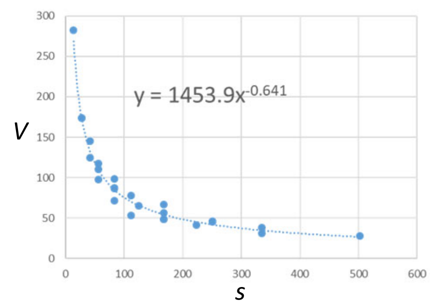
$$V_s = \frac{V}{s^b} \qquad \ln(V_n) = \ln(V_1) - b\ln(n)$$

where:
V_S = variance among plots of size *s*
b = soil heterogeneity index

Intrinsically Linear Equations

Smith's Formula - Calculations





$$V_n = \frac{4339.9}{n^{0.9458}}$$

Intrinsically Linear Equations

Smith's Formula - Calculations

```
proc mixed;
  class site;
  model lnvar = site lnarea*site / solution noint;
  estimate 'isu.ne v isu.se slope' lnarea*site 1 -1;
  estimate 'isu.ne v isu.se intercept' site 1 -1;
run;
```

Details:

- MIXED code for calculating regressions
- Two sites – isu.ne and isu.se listed in CLASS
- *lnvar* = natural log of variance for a plot size *lnarea*
- SOLUTION displays the fixed parameters, intercept and slope
- NOINT prevents calculation of a common intercept
- ESTIMATEs compare the site parameters

Intrinsically Linear Equations

Smith's Formula - Calculations

Solution for Fixed Effects

Effect	site	Estimate	Standard Error	DF	t Value	Pr > |t|
site	isu.ne	8.2250	0.2840	46	28.96	<.0001
site	isu.se	7.2820	0.2840	46	25.64	<.0001
lnarea*site	isu.ne	-0.9257	0.06040	46	-15.33	<.0001
lnarea*site	isu.se	-0.6415	0.06040	46	-10.62	<.0001

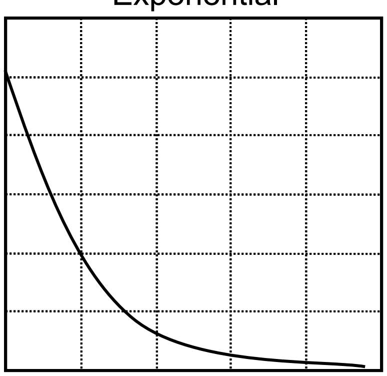
Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
isu.ne v isu.se slope	-0.2842	0.08541	46	-3.33	0.0017
isu.ne v isu.se intercept	0.9430	0.4017	46	2.35	0.0233

Nonlinear Equations

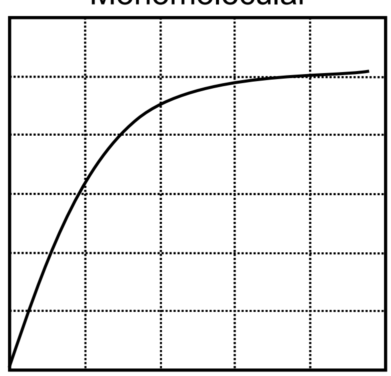
Common Functions

Exponential



$$Y = ae^{-bx}$$

Monomolecular

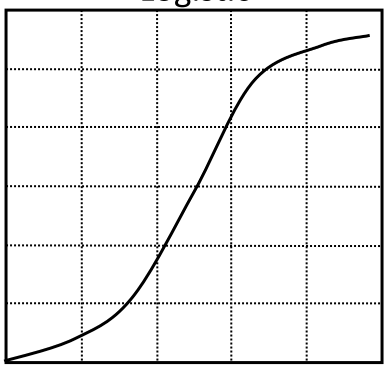


$$Y = a(1-e^{-bx})$$

Nonlinear Equations

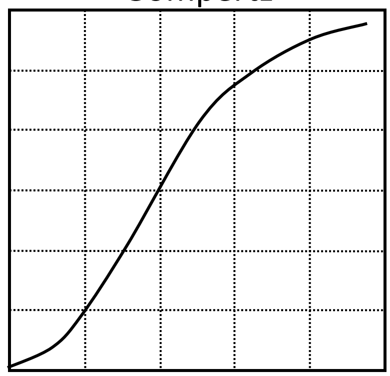
Common Functions

Logistic



$$Y = \frac{a}{1 + be^{-cx}}$$

Gompertz

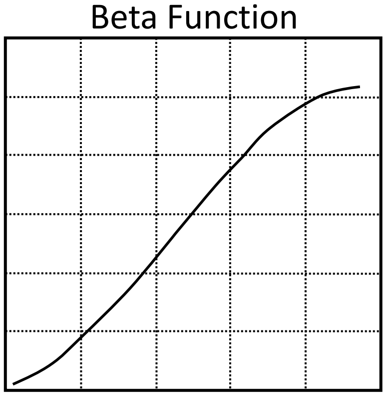


$$Y = ae^{-be^{-cx}}$$

Nonlinear Equations

Common Functions

Beta Function



$$Y = a\left(1 + \frac{b-x}{b-c}\right)\left(\frac{x}{b}\right)^{b/(b-c)}$$

a = Y_{max} = maximum response
b = x to max response
C = x to max rate

Fitting Nonlinear Equations

1. Plot means
2. Evaluate response visually
3. Fit appropriate model(s)
4. Evaluate stats – R^2, Residual SS, SEE
5. Evaluate residuals
6. Choose best model

Nonlinear Equations

Fit Statistics

<u>R^2 Approximation</u>

$$R^2 = 1 - \frac{SS_{Residual}}{SS_{Total}}$$

<u>Standard Error of the Estimate</u>

$$S_{y \cdot x} = \sqrt{\frac{SS_{RES}}{n - p}}$$