

<p style="text-align: center;"><b>Chapter 8: A closer look at assumptions</b>  <b>Chapter 9: Multiple regression</b></p> <hr/> <p style="text-align: center;">Class 15: 4/6/09 M</p>	<p><b>Slide 1 Chapter 8: A closer look at assumptions</b></p> <p>Chapter 9: Multiple regression</p> <p>NOTES:</p>
<p style="text-align: center;"><b>HW 10 due Weds 4/8/09</b></p> <hr/> <p style="text-align: center;">Submit as Myname-HW9.doc (or *.rtf)</p> <ul style="list-style-type: none"> <li>• Computation Problem 10, chapter 8             <ul style="list-style-type: none"> <li>▸ 8.16 Meat processing, Must assess lack of fit!</li> <li>▸ Due Weds 4/8 10 am</li> </ul> </li> <li>• Read Chapter 3 from Draper &amp; Smith on regression, especially designing a standard curve (includes lack of fit from Chapter 8)</li> <li>• Read Chapter 9 on multiple regression             <ul style="list-style-type: none"> <li>▸ Read chapter 9 conceptual problems &amp; solutions</li> <li>▸ Post a question or response about Chapter 9 conceptual problems</li> </ul> </li> <li>• HW 11 9.19: Speed of evolution,             <ul style="list-style-type: none"> <li>▸ Due Monday 4/13/09 10 am</li> <li>▸ This is a TOUGH problem! Weds: ask for help/hints!</li> </ul> </li> </ul>	<p><b>Slide 2 HW 10 due Weds 4/8/09</b></p> <p>NOTES:</p>
<p style="text-align: center;"><b>Homework Solution Presentation</b></p> <hr/> <p style="text-align: center;">Lisa Greber: Homework 7</p> <ul style="list-style-type: none"> <li>• Computation Problem 7             <ul style="list-style-type: none"> <li>▸ Problem 5.25 Duodenal ulcers</li> </ul> </li> </ul> <p style="text-align: right;"><b>EEOS611</b></p>	<p><b>Slide 3 Homework Solution Presentation</b></p> <p>NOTES:</p>

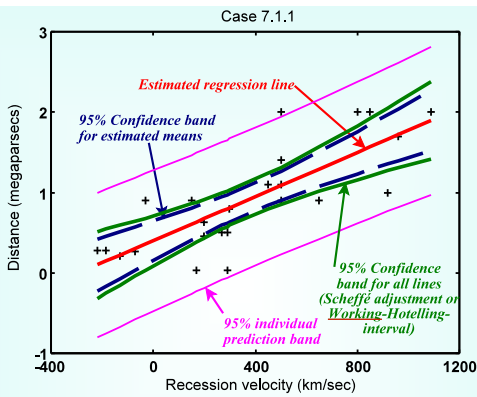
## Chapter 7 (Continued) Simple Linear Regression: a model for the mean

Simple regression = ordinary least squares (OLS)  
regression, Model I regression

**EEOS611**

### Slide 4 Chapter 7 (Continued) Simple Linear Regression: a model for the mean

NOTES:



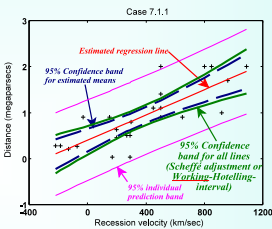
### Slide 5

NOTES:

### Working-Hotelling CI for a regression line

Kendall & Stuart (1979, *Advanced theory of statistics*,  
4th edition, Section 28.24)

"Suppose now that we require a confidence region for an entire regression line, i.e. a region R in the (x,y) plane ...such that there is probability  $1-\alpha$  that the true regression line  $y=x\beta$  is contained in R. ... we are now seeking a confidence region, not an interval, and it covers the whole line, not one point on the line... this problem first solved in the simplest case by Working and Hotelling (1929) in a remarkable paper..." p 388



### Slide 6 Working-Hotelling CI for a regression line

NOTES:

**Display 10.9**

Construction of the 95% confidence band using repeated fits of the multiple regression model with different reference points

Computer Work

Reference Point	Explanatory Variables	Intercept Estimate	Standard Error
Body Mass	Body Mass		
100	$\ln(\text{mass} - \log(100))$	2.2789	0.0604
400	$\ln(\text{mass} - \log(400))$	3.3087	0.0635
non-echo bats	$\ln(\text{mass} - \log(100))$	2.1767	0.1144
400	$\ln(\text{mass} - \log(400))$	3.3064	0.0931
echo bats	$\ln(\text{mass} - \log(100))$	2.2553	0.1277
400	$\ln(\text{mass} - \log(400))$	3.3851	0.1759

Hand Calculations — an Example

Multiplier =  $\sqrt{4 F_{4,16, 0.95}} = 3.468$

Lower limit =  $\exp[2.2789 - (3.468)(0.0604)] = 7.9$

Upper limit =  $\exp[2.2789 + (3.468)(0.0604)] = 12.0$

Note Scheffé multiplier used for Working Hotelling CI's somewhat atypical, but appropriate

**Slide 7 Display 10.9**

---

NOTES:

---



---



---



---



---



---



---

**Display 10.8**

Estimated median energy expenditures for birds, echolocating bats, and non-echolocating bats as functions of body mass: parallel lines model on log scale, with 95% confidence bands

Scheffé multiplier produces broad prediction interval

**Slide 8 Display 10.8**

---

NOTES:

---



---



---



---



---



---



---

1007 [DataView] - SPSS Data Editor

CASE	type	mass	lnmass	energy	lnenergy
1	Bat Non-echo	770	6.66	44	3.79
21	Bat Non-echo	100	4.61		
22	Bats	100	4.61		
23	Bats Echo	100	4.61		
24	Bat Non-echo	400	5.99		
25	Bats	400	5.99		
26	Bats Echo	400	5.99		

This Scheffé prediction interval is based on  $\infty$  samples. There is a further Scheffé adjustment for individual CI's

Low95I	Up95I	Low95M	Up95M	Low95S	Up95S
30.14	73.22	38.31	57.61	33.646	65.586
5.55	14.01	6.92	11.24	5.929	13.111
6.45	14.78	8.59	11.10	7.919	12.044
5.91	15.39	7.28	12.50	6.125	14.055
17.56	42.41	22.40	33.24	19.767	37.688
19.93	45.85	26.42	34.58	24.249	37.675
17.16	50.79	20.33	42.86	16.039	54.335

**Slide 9**

---

NOTES:

---



---



---



---



---



---



---

### Syntax: Scheffé multiplier

\* regpars is the number of parameters in the final model, with 16 df in the residual.  
 Compute regpars=4.  
 Compute residdf=16.  
 exe.

```

COMPUTE FScheffe = IDF.F(0,95,regpars,residdf) .
EXECUTE .
COMPUTE Scheffemultiplier = sqrt(regpars*FScheffe) .
EXECUTE .
    
```

\* Scheffe interval is Scheffe multiplier times the standard error for each predicted value, SEP\_1 was produced by regression.  
 COMPUTE Schint = SEP\_1 \* Scheffemultiplier .  
 EXECUTE .  
 COMPUTE L95S = PRE\_1 - Schint.  
 COMPUTE U95S = PRE\_1 + Schint.  
 EXE.  
 COMPUTE PredE = Exp(PRE\_1).  
 COMPUTE Low95S = Exp(L95S) .  
 COMPUTE Up95S = Exp(U95S) .  
 EXECUTE .

### Slide 10 Syntax: Scheffé multiplier

NOTES:

### Lack of fit tests (8.5.3) using the Regression ANOVA model

(using Case 8.2 as an example)

**Chem sts & phys cal oceanographers don't use ANOVA, t s s a d They use regress on. But, they should analyze the regress on ANOVA table and test for lack of f t**

EEOS611

### Slide 11 Lack of fit tests (8.5.3) using the Regression ANOVA model

NOTES:

### How to recognize lack of fit

And what to do about it

**No obvious transformation**

Pitts,  
EEOS  
Ph.D.  
EEOS611

### Slide 12 How to recognize lack of fit

NOTES:

### Case 8.2 Testing for lack of fit

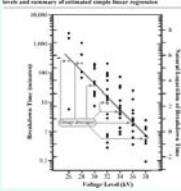
Display 8.4 (p. 210) Semilog plot:  $\ln(Y)$  vs. Linear  $X$

*Estimated Regression Line*

$$\hat{\mu}\{\log(\text{BDT})\} = 18.96 - 0.507 \text{ Voltage}$$

$$\hat{\sigma}D \text{ of } \log(\text{BDT}) = 1.56$$

Combine ANOVA tables  
from regression &  
separate means models



Coefficients<sup>a</sup>

	Unstandardized Coefficients		Standardized Coefficients		95% Confidence Interval for B	
	B	Std. Error	Beta	t	Sig.	Lower Bound Upper Bound
(Constant)	18.965	1.910		9.9	3.1E-015	15.1 22.8
Voltage (KV)	-.507	.057	-.717	-8.8	3.3E-013	-.6 -.4

a. Dependent Variable: Ln(Time)

### Slide 13 Case 8.2 Testing for lack of fit

NOTES:

### Display 8.8

p. 210

Analysis of variances tables for the insulating fluid data from a simple linear regression analysis and from a separate-means (one-way ANOVA) analysis

**(A): ANALYSIS OF VARIANCE TABLE FROM A SIMPLE LINEAR REGRESSION ANALYSIS**

Source	Sum of Squares	df	Mean Square	F-Statistic	p-value
Regression	190.1514	1	190.1514	78.14	<.0001
Residual	180.0745	74	2.4334		
Total	370.2258	75			

*Annotations for (A):*  
 - Residual sum of squares, regression model: 180.0745  
 -  $\sigma^2$  in regression model: 2.4334  
 - compares regression and equal-means models: p-value <.0001

**(B): ANALYSIS OF VARIANCE TABLE FROM A ONE-WAY ANALYSIS OF VARIANCE**

Source	Sum of Squares	df	Mean Square	F-Statistic	p-value
Between Groups	196.4774	6	32.7462	13.00	<.0001
Within Groups	173.7484	69	2.5181		
Total	370.2258	75			

*Annotations for (B):*  
 - Residual sum of squares, separate-means model: 173.7484  
 -  $\sigma^2$  in separate-means model: 2.5181  
 - compares separate-means and equal-means models: p-value <.0001

### Slide 14

NOTES:

### Display 8.9

Reductions in sums of squared residuals in hierarchical models for mean responses in the insulating fluid study

**CONSTANT MEAN**  
(1 parameter)

Adding 1 parameter lowers residual SS by 190.1514

→

**SIMPLE LINEAR REGRESSION**  
(2 parameters)

Adding 5 more parameter lowers residual SS by 6.3260

→

**SEPARATE MEANS**  
(7 parameters)

Adding 6 parameter lowers residual SS by 196.4774

→

**EEOS611**

### Slide 15 Display 8.9

NOTES:

1.5. LACK OF FIT AND PURE ERROR 37

Draper & Smith (Chapter 1)

Figure 1.9 Breakup of residual sum of squares into lack of fit and pure error sum of squares.

**Slide 16**

NOTES:

Display 8.10

Composite analysis of variance table with F-test for lack-of-fit

Source of Variation	Sum of Squares	df	Mean Square	F-Statistic	p-value
Between Groups	196.4774	6	32.7462	13.00	<.0001
Regression	190.1514	1	190.1514	75.51	<.0001
<b>Lack of Fit</b>	<b>6.3260</b>	<b>5</b>	<b>1.2652</b>	<b>0.50</b>	<b>.78</b>
Within Groups	173.7484	69	2.5181		
Total	370.2258	75			

(by subtraction)

LEGEND  
 Normal type items come from Regression Analysis (A)  
 Italicized items come from separate-means Analysis (B)  
 Bold face items are new and calculated here

$$s_r = \hat{\beta}_0 + \hat{\beta}_1 X_j = [E + \hat{\beta}_1] + [\hat{\beta}_0 + \hat{\beta}_1 X_j]$$

Residual = Pure error + Lack of fit

$$F \text{ stat} = \frac{SS_{LR} - SS_{SM}}{SS_{SM} / (a - 1)} \cdot \frac{1}{a - 1} \cdot \frac{1}{SS_{SM}}$$

**Slide 17**

NOTES:

**How to test lack of fit using SPSS**

1 of 2

- You must have true replicates at 1 or more values of the explanatory variables
  - [With no replicates, you can still evaluate lack of fit visually with residual plots. Curvature in residuals is a common cause of lack of fit.]
  - 1) Do ANOVAs to test for differences among means of replicated groups (Use SPSS's oneway or general linear model (UNIANOVA))
  - 2) Do a regression analysis
  - 3) Combine Regression and 'separate means' ANOVA tables
    - A) Subtract regression sum of squares (1df) from the ANOVA among group sum of squares (#replicated groups -1) to obtain 'lack of fit' sum of squares & df (# replicated groups -2)
    - B) Divide sums of squares by appropriate df and evaluate 'lack of fit' F statistic

**Slide 18 How to test lack of fit using SPSS**

NOTES:

### How to test lack of fit using SPSS

2 of 2

- OR, use SPSS General linear model/univariate/ options/lack of fit, enter explanatory variables as covariate
- OR, for equally spaced explanatory variables, use linear contrasts in ONEWAY or enter contrasts in oneway or general linear model
  - You can find the linear contrast values in Winer et al. (1991) or Draper & Smith or use the Matlab's orthpoly.m (in Smyth's Statbox from Matlab users' files)
  - Quadratic, cubic and quartic polynomials also provided (each set of polynomials is uncorrelated with those of lower order)

### Slide 19 How to test lack of fit using SPSS

NOTES:

### Strategy for dealing with lack of fit

- You must have true replicates
- Examine scatterplots
  - Are transformations or quadratic explanatory variables needed?
- Fit linear regression model
  - Examine residuals
  - Transform data, add quadratic or cubic explanatory terms if needed
  - Add other explanatory terms (Ch 9...)
- Perform lack of fit test
  - If LOF significant with linear model, consider tests of higher order (quadratic & cubic) trends in ANOVA model
    - Add quadratic or cubic terms to regression model if quadratic or cubic trend found
    - LOF could be due to cluster & serial effects
- Report effect size with regression or ANOVA
  - Regression slope is still an unbiased estimator of true slope
  - Use linear contrast in ANOVA to determine effect size (GLM Unianova)

### Slide 20 Strategy for dealing with lack of fit

NOTES:

### What to do if there is lack of fit!

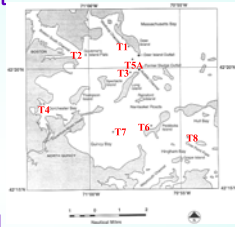
- You may still estimate the slope & Y intercept using regression: OLS regression still provides unbiased estimators
  - ANOVA linear contrast will also provide estimators for linear slope & confidence intervals
- You can **NOT** use the variance estimates and *p* values based on the error mean square from the OLS linear regression
- Fit a richer or different model
  - Consider testing higher order interaction terms: quadratic & cubic, if warranted
  - Add other explanatory variables
- You **may** analyze the data as an ANOVA separate means model with a linear contrast (see sleuth yellow-tail fish weights as an example)

### Slide 21 What to do if there is lack of fit!

NOTES:

### Lack of Fit & Boston Harbor soft-bottom benthic diversity

- Eight sampling stations: not chosen randomly!
  - Historically important sites
  - Severely limits the statistical inference possible
- Stations sampled in May & Aug each year, starting in Aug 1991
- 3 replicate 0.043-m<sup>2</sup> Ted Young modified Van Veen grabs
- Species richness measured with Fisher's  $\alpha$

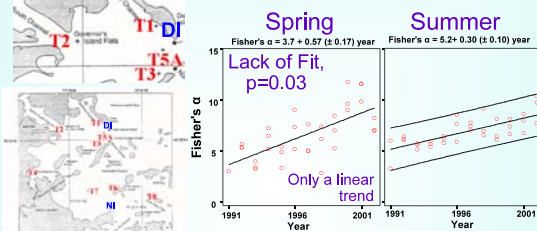


### Slide 22 Lack of Fit & Boston Harbor soft-bottom benthic diversity

NOTES:

### T1: Deer Island Flats

Very high rates of increase in richness (higher in spring than summer)



Amy Banik ECOS M.Sc.

### Slide 23 T1: Deer Island Flats

NOTES:

Coefficients from Winer et al. (1991, Table D-8), also shown in Draper & Smith (1998, p. 467). The tests are performed sequentially, with the linear term tested first, and the lack of fit resulting after the linear term is removed. The quadratic and cubic terms are tested only if there is a significant lack of fit after the full suite of the lower order polynomials is included in the model. These orthogonal contrasts are also obtained automatically by the SAS procedure general linear model.

	Year									
	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Linear	-9	-7	-5	-3	-1	1	3	5	7	9
Quadratic	6	2	-1	-3	-4	-4	-3	-1	2	6
Cubic	-41	14	35	31	12	-12	-31	-35	-14	41

	Year									
	1992	1993	1994	1995	1996	1997	1998	1999	2000	
Linear	-4	-3	-2	-1	0	1	2	3	4	
Quadratic	25	7	-5	-17	-20	-17	-5	7	25	
Cubic	-14	7	13	0	0	-9	-13	7	-14	

Orthogonal polynomials in Winer et al. (See syllabus) or in Matlab's orthpoly m (from Smyth's statbox)

### Slide 24

NOTES:



Table 1. Results of tests of regression and ANOVA models for changes in Fisher's  $\alpha$  at Site T1 through the 1990s. There were 10 summer means (1991 to 2000) and 9 for the spring series (1992 to 2000). The lack-of-fit F test is performed by testing  $H_0: \sigma^2_{\text{residual}} = \sigma^2_{\text{error}}$  which under the null hypothesis of a linear regression to the data should be distributed as  $F_{(n-2), (n-2)}$  for summer and spring analyses.

Source of Variation	SS	df	MS	F	Sig.
Among Years	59.31	8	11.19	4.93 (19.8)	0.002
Regression	43.08	1	43.08		
Regression Error	93.21	25	3.49		
Lack of fit	46.43	1	46.43	2.92 (19.8)	0.031
Pure Error	40.54	18	2.25		
Linear trend	43.08	1	43.08	15.99	0.0004
Quadratic trend	0.98	1	0.98	4.40	0.0503
Cubic trend	1.21	1	1.21	3.20	0.0803

Lack of fit test: Can these two estimates of variance be pooled to form the regression Error mean square?

**Slide 25**

---

NOTES:

---



---



---



---



---



---



---

**Conclusion on lack of fit test of Boston Harbor**

- Very strong evidence that species richness (as measured by Fisher's  $\alpha$ ) is increasing in spring T1 samples [ANOVA linear contrast ( $F_{1,10} = 19, p < 0.001$ )]
- There was significant lack of fit in the OLS regression due to non-linear patterns in year-to-year increases in species richness or cluster effects

Amy Banik ECOS M.Sc. **EEOS611**

**Slide 26 Conclusion on lack of fit test of Boston Harbor**

---

NOTES:

---



---



---



---



---



---



---

**T2: Governor's Island Flats**

Similar rapid rates of increase in spring & summer

Lack of fit

Amy Banik M.Sc. **EEOS611**

**Slide 27 T2: Governor's Island Flats**

---

NOTES:

---



---



---



---



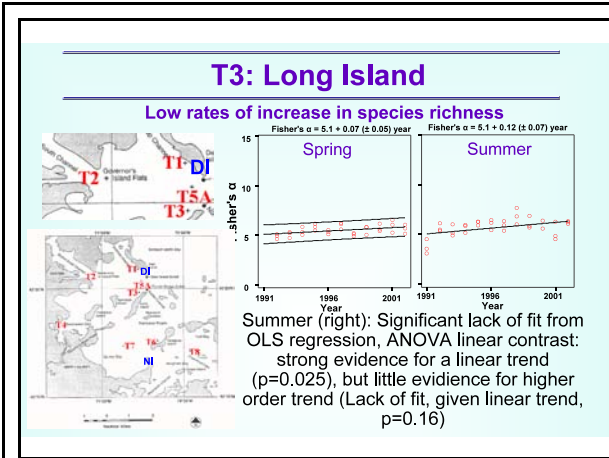
---



---



---



**Slide 28 T3: Long Island**

---



---

NOTES:

---



---



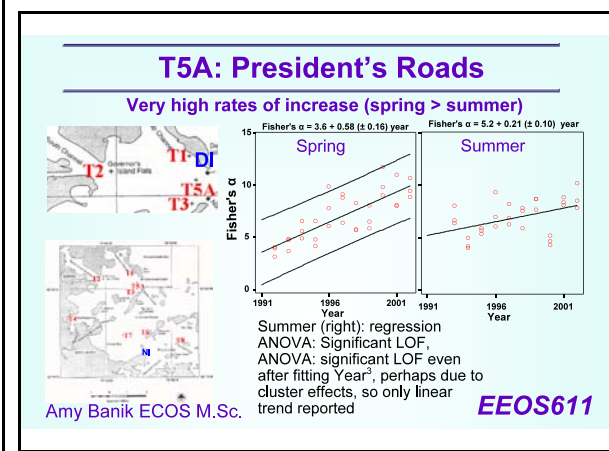
---



---



---



**Slide 29 T5A: President's Roads**

---



---

NOTES:

---



---



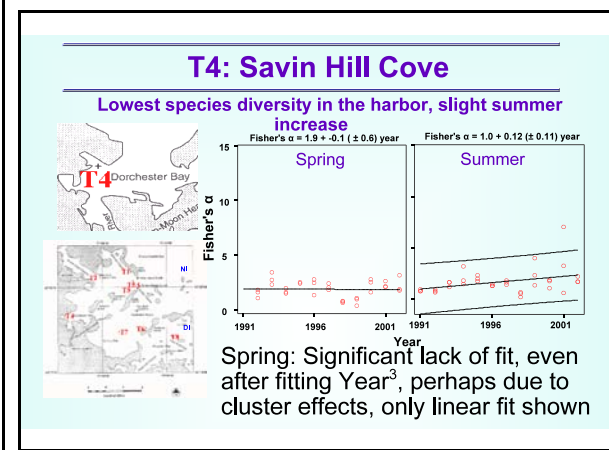
---



---



---



**Slide 30 T4: Savin Hill Cove**

---



---

NOTES:

---



---



---



---



---

**Site T6: Peddock's Island**

A dense amphipod assemblage, a slight increase in species richness only in summer

Amy Banik M.Sc. EOS611

**Slide 31 Site T6: Peddock's Island**

---



---

NOTES:

---



---



---



---



---

**Matrix approach to linear regression & the normal equations**

Available in Matlab

*EOS611*

**Slide 32 Matrix approach to linear regression & the normal equations**

---

NOTES:

---



---



---



---



---

**Linear algebra of regression**

Equations from Draper & Smith (1981), Kendall & Stuart

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_I \end{bmatrix} \quad \begin{bmatrix} X_{11} & \dots & X_{1J} \\ X_{21} & \dots & X_{2J} \\ \vdots & \ddots & \vdots \\ X_{I1} & \dots & X_{IJ} \end{bmatrix}$$

*y* variable *j*.

*J* - Number of explanatory variables.

*I* - Number of cases.

**Slide 33 Linear algebra of regression**

---

NOTES:

---



---



---



---



---

<p style="text-align: center;"><b>The Normal Equations &amp; Matlab</b></p> $Y = X\theta + \epsilon,$ $E(\epsilon) = 0,$ $V(\epsilon) = E(\epsilon\epsilon') = \sigma^2 I.$ <p><i>S</i> = scalar sum of squares.  <i>LS</i> method requires minimization of scalar sum of squares, <i>S</i>.</p> $S = (y - X\theta)'(y - X\theta)$ <p>To minimize <i>S</i>, set <math>\frac{\partial S}{\partial \theta} = 0</math>.</p> <p>Differentiating, <math>2X'(y - X\theta) = 0</math>.</p> $\hat{\theta} = (X'X)^{-1}X'y.$ <p><b><math>\hat{B} = (X'X)^{-1}X'Y</math> or <math>\hat{B} = X \backslash Y</math>.</b> <span style="float: right;"><b>EEOS611</b></span></p>	<p style="text-align: center;"><b>Slide 34 The Normal Equations &amp; Matlab</b></p> <hr/> <p>NOTES:</p> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>
<pre> b=X\Y; Yest=X*b; resid=Y-Yest; n=length(Y); SSb1b0=b'*X'*Y-Y'*ones(n,1)*ones(n,1)'*Y/n; % Here is the full ANOVA table from Draper &amp; Smith (p. 85) with d.f. RegSS=b'*X'*Y; ResSS=Y'*Y-b'*X'*Y; TotSS=Y'*Y; Meanb0=n*(mean(Y)^2); TotcSS=TotSS-Meanb0; RF=SSb1b0/(ResSS/(n-length(b))); % The F statistic for the regression: PRF=fprob(1,n-length(b),RF); SS=[SSb1b0 1 Ssb1b0 ResSS n-length(b) ResSS/(n-length(b)) TotcSS n-1 TotcSS/(n-1)]; F=[SS(1,3) SS(2,3) SS(1,2) SS(2,2) RF PRF]; s2=SS(2,3); R2=(b'*X'*Y-Meanb0)/(Y'*Y-Meanb0); % This is R^2 Vb=inv(X*X)*s2; % Equation 2.3.2 in Draper &amp; Smith % calculation of the variance of Yest (Draper &amp; Smith, p. 28 &amp; 83) VYest=diag(Xo*inv(X*X)*Xo*s2);         </pre>	<p style="text-align: center;"><b>Slide 35 Matlab least squares regression</b></p> <hr/> <p>NOTES:</p> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>
<p style="text-align: center;"><b>Excursis: Rule of the Bulge vs. Non-linear modeling; from Singer &amp; Willett (2003) Applied longitudinal data analysis.</b></p> <hr/> <p style="text-align: center;">Non-linear &amp; Weighted regression (will be covered in more detail at the end of Sleuth Chapter 11)</p> <p style="text-align: right;"><b>EEOS611</b></p>	<p style="text-align: center;"><b>Slide 36 Excursis: Rule of the Bulge vs. Non-linear modeling; from Singer &amp; Willett (2003) Applied longitudinal data analysis.</b></p> <hr/> <p>NOTES:</p> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>

### The rule of the bulge, Case 8.2

Case 8.1 (Scatterplot with smoother on left)

Note 'bulge' to the lower left of the plot

**Slide 37 The rule of the bulge, Case 8.2**

NOTES:

### Rule of the Bulge & Ladder of Powers

Singer

Case 8.2 bulge to the lower left

There is often more than 1 way to transform X's and Y

Figure 6.5. The ladder of transformations and the rule of the bulge. Guidelines for increasing individual growth trajectories through judicious use of transformation.

**Slide 38 Rule of the Bulge & Ladder of Powers**

NOTES:

Display 8.4

Scatterplot of breakdown times (natural logarithm scale) versus voltage levels and summary of estimated simple linear regression

Horn-shaped residual plot still present after square root transform, so try log transform

Display 8.7

Scatterplot of the square root of breakdown time versus voltage and a residual plot based on the simple linear regression fit

**Slide 39**

NOTES:

### The rule of the bulge

Case 8.1 (Scatterplot with smoother [4.0 scaling on X,Y] on left)

Note 'bulge' to the upper left of the plot

Note the spread around the lines very important  
Transformations of the response affect the spread around the line

### Slide 40 The rule of the bulge

NOTES:

### The rule of the bulge

For Case 8.1, 1st step indicates species should be squared but subsequent graphs indicate log on X

Case 8.1 bulge to the upper left, ln(area)

Iterative fits to the rule of the bulge residuals & theory indicates that Y should be log transformed

Figure 6.5. The ladder of transformations and the rule of the bulge. Guidelines for linearizing individual growth trajectories through judicious use of transformation.

### Slide 41 The rule of the bulge

NOTES:

### Recall Display 8.6

Examine the spread of the residuals

Don't use 'The Rule of the Bulge' naively need to also consider the spread

Spread ok, transform X

Spread increasing with predicted value, ok to transform Y

Weighted Regression, p 328

### Slide 42 Recall Display 8.6

NOTES:

### The Michaelis-Menten Equation

**1905: Used to model the rate of enzyme reactions**  
 Many of the linear transforms are not appropriate since variance is unequal after the linear transform

$$v = \frac{V_{max} S}{K_M + S}$$

where,  $K_M$  = Half-saturation constant.  
 =  $S$  at which  $v = \frac{1}{2} V_{max}$ .  
 $S$  = Substrate concentration.  
 $v$  = Reaction velocity.  
 =  $\frac{-dS}{dt} = \frac{d\text{Product}}{dt}$   
 $V_{max}$  = Maximum  $v$ .

**S611**

### Slide 43 The Michaelis-Menten Equation

NOTES:

### Monod (1948): $\mu=f(S_{\text{external}})$

Adapted the Michaelis-Menten equation

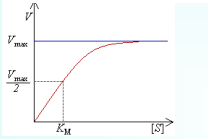
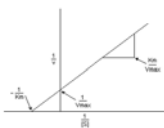
**EEOS611**

### Slide 44 Monod (1948): $\mu=f(S_{\text{external}})$

NOTES:

### Michaelis-Menten, Lineweaver-Burke plots & Eadie-Hofstee plots

$$\frac{d[P]}{dt} = k_2[E_0] \frac{[S]}{K_m + [S]} = V_{max} \frac{[S]}{K_m + [S]}$$

Better approaches nonlinear fits or Weighted regression (Ch 11, Sleuth)

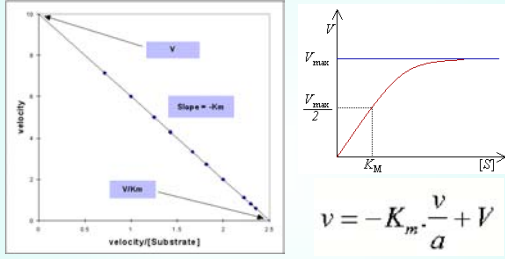
Don't use Lineweaver Burke errors for 1/S and 1/V inflated! Use weighted (Sleuth 11.6.1) or nonlinear regression

### Slide 45 Michaelis-Menten, Lineweaver-Burke plots & Eadie-Hofstee plots

NOTES:

**Michaelis Menten & Eadie-Hofstee**

Don't use Eadie-Hofstee: Error variance affected. The ratio of 2 normals isn't normal!. Use nonlinear or weighted regression (Sleuth 11.6.1)

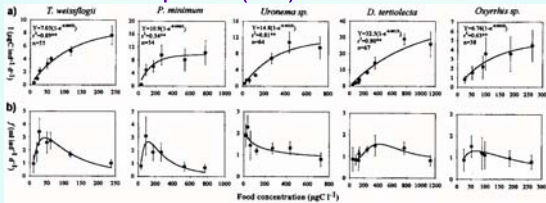


**Slide 46 Michaelis Menten & Eadie-Hofstee**

NOTES:

**Acartia feeding responses**

Besiketep & Dam (2002) MEPS 229: 151-164



Diatom

Ciliate

lvlev feed ng models usually f t w th non l near regress on (Sleuth Chapter 11)

**Slide 47 Acartia feeding responses**

NOTES:

**Chapter 9**

Multiple regression

EEOS611

**Slide 48 Chapter 9**

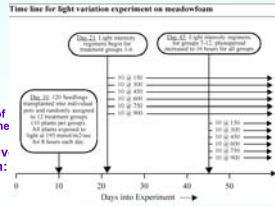
NOTES:



## Multiple Regression Case 9.1

### Timing and light intensity

- **Response variable:** Average number of flowers per meadowfoam plant
- **Explanatory variables**
  - Light intensity: 6 levels
    - Treated as a continuous variable
    - Could have been treated as 6 categories of light level (SPSS GLM: Unianova creates the indicator variables automatically)
  - Timing of light intensity change relative to PFI (Photoperiodic Floral Induction: increase of light from 8 to 16 hours)
- Tests whether the slopes are parallel (is there an interaction?)
- Estimate effect sizes



EEOS611

## Slide 49 Multiple Regression Case 9.1

NOTES:

## Display 9.2: Two-factor layout

Can be tested as either a linear regression or ANOVA, but regression is the more general & powerful method

Display 9.2

Numbers of flowers per meadowfoam plant, in twelve treatment groups

		intensity ( $\mu\text{mol}/\text{m}^2/\text{sec}$ )					
		150	300	450	600	750	900
timing	at PFI	62.3 77.4	55.3 54.2	49.6 61.9	39.4 45.7	31.3 44.9	36.8 41.9
	24 days before PFI	77.8 75.6	69.1 78.0	57.0 71.1	62.9 52.2	60.3 45.6	52.6 44.4

ANOVA: Light intensity would be treated as a categorical variable

## Slide 50 Display 9.2: Two-factor layout

NOTES:

## SPSS Syntax & dummy variables

Reference level: all but one level of the explanatory variables is included in the regression model as a dummy (or indicator) variable, the one left out is referred to as a reference level.

```
* Time (1,2) must be converted to a (0,1) dummy.
COMPUTE Timing = Time-1 .
EXECUTE .
* Create the interaction variable and format it.
COMPUTE Intxn = intens*timing .
EXECUTE .
format Intxn (f1.0).
compute L150 = (intens=150).
compute L300 = (intens=300).
compute L450 = (intens=450).
compute L600 = (intens=600).
compute L750 = (intens=750).
compute L900 = (intens=900).
exe.
formats L150 to L900 (f1.0).
```

Note that SPSS General linear model (UNIANOVA in syntax) will automatically fit the data, assigning indicator variables for each level of the explanatory variable

## Slide 51 SPSS Syntax & dummy variables

NOTES:



**Case 9.1, Theory and calculations**

*EEOS611*

**Slide 55 Case 9.1, Theory and calculations**

NOTES:

Model for the regression surface of *flowers per plant* under 12 treatment levels as a regression plane

**Regression plane:**  $\mu(\text{flowers} | \text{light}, \text{time}) = \beta_0 + \beta_1 \text{light} + \beta_2 \text{time}$

$Y = \text{flowers per plant}$

$X_2 = \text{time (days)}$

$X_1 = \text{light intensity } (\mu\text{mol/m}^2/\text{sec})$

*EEOS611*

**Slide 56 Display 9.5**

NOTES:

**Graphical display of interaction**

An interaction in regression is synonymous with non-parallel slopes: so test whether the slopes are equal & slopes are equal

Display 9.8

Regression models within subsets of the data that have separate lines, parallel lines, and equal lines: meadowlark study example.

Flowers per plant

Intensity

Timing: Late (at PFI), Early (24 d before PFI)

Linear Regression

Flowers =  $83.15 + 0.04 \cdot \text{Intensity}$   
R-Square = 0.77

Flowers =  $71.62 + 0.04 \cdot \text{Intensity}$   
R-Square = 0.72

**Slide 57 Graphical display of interaction**

NOTES:

**SPSS: Solving regression problems**

- /analyze/regression
  - Solves all regression problems
  - Multiple models can be set up in advance
  - Indicator (dummy variables) must be computed manually (in syntax or use copy & paste)
- /analyze/general linear model
  - Continuous variables must be entered as covariates; interactions will be calculated for main effects.
  - More tests available, including scatterplots, power analysis, and lack of fit tests
  - With fixed effects, will do analysis as if each level of an explanatory variable was coded as a dummy (=indicator) variable.

**Slide 58 SPSS: Solving regression problems**

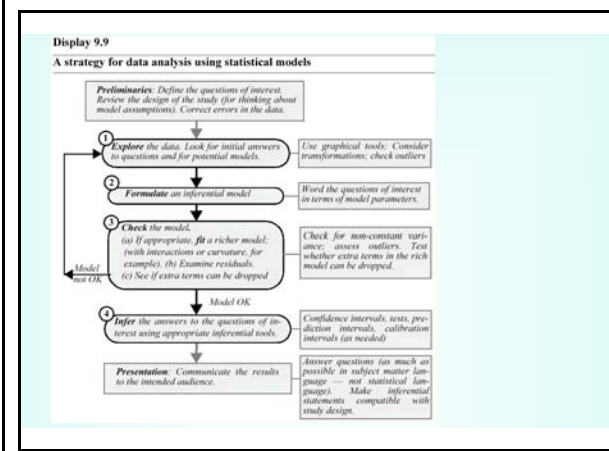
NOTES:

**Case 9.2 Allometry of brain size**

*EEOS611*

**Slide 59 Case 9.2 Allometry of brain size**

NOTES:



**Slide 60 Display 9.9**

NOTES:



### The rule of the bulge

Case 9.2

Note 'bulge' to the lower left of plots

Litter of 20 mice

Rule of the bulge

Note 9.1: The bulge of outlying points on the left of the plots. Question 9.1

### Slide 64 The rule of the bulge

---

NOTES:

---



---



---



---



---



---

Display 9.11 p. 244

Matrix of scatterplots for brain weight data, after log transformations

Ch 10 deals with the tools to identify outliers that matter: leverage & Cook's D

### Slide 65 Display 9.11

---

NOTES:

---



---



---



---



---



---

Species

- Amur leopard
- Asian elephant
- Asian elephant II
- Asian elephant III
- Asian elephant IV
- Asian elephant V
- Asian elephant VI
- Asian elephant VII
- Asian elephant VIII
- Asian elephant IX
- Asian elephant X
- Asian elephant XI
- Asian elephant XII
- Asian elephant XIII
- Asian elephant XIV
- Asian elephant XV
- Asian elephant XVI
- Asian elephant XVII
- Asian elephant XVIII
- Asian elephant XIX
- Asian elephant XX
- Asian elephant XXI
- Asian elephant XXII
- Asian elephant XXIII
- Asian elephant XXIV
- Asian elephant XXV
- Asian elephant XXVI
- Asian elephant XXVII
- Asian elephant XXVIII
- Asian elephant XXIX
- Asian elephant XXX
- Asian elephant XXXI
- Asian elephant XXXII
- Asian elephant XXXIII
- Asian elephant XXXIV
- Asian elephant XXXV
- Asian elephant XXXVI
- Asian elephant XXXVII
- Asian elephant XXXVIII
- Asian elephant XXXIX
- Asian elephant XL

### Slide 66

---

NOTES:

---



---



---



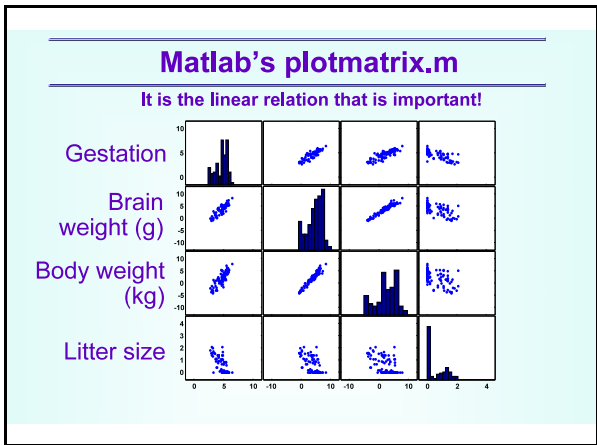
---



---



---



### Slide 67 Matlab's plotmatrix.m

NOTES:

### Case Study 9.2

Testing the effects of **Litter size on brain weight**, after accounting for **body weight and gestation length**

Model	Unstandardized Coefficients		Standardized Coefficients		95% Confidence Interval for B		
	B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	2.332	.073		31.843	.000	2.187	2.478
	(Constant)						
2	.719	.020	.964	35.300	.000	.679	.760
	(Constant)						
	-.487	.458		-.997	.3212	-1.368	.453
	In (body wt (kg))						
	.551	.032	.739	17.033	.000	.487	.615
	In (gestation period)						
	.688	.109	.265	6.141	.000	.452	.924
	(Constant)						
	.855	.662		1.282	.1996	-.459	2.169
	In (body wt (kg))						
	.575	.033	.771	17.647	.000	.510	.640
	In (gestation period)						
	.418	.141	.167	2.989	.0036	.138	.698
	In (litter size)						
	-.310	.116	-.266	-2.675	.0088	-.540	-.080
	In (litter size)						

a. Dependent Variable: In (brain wt (g))

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					F Change	df1	df2		
1	.964 <sup>a</sup>	.930	.929	.07807	.930	1546.104	1	94	.0000
2	.975 <sup>b</sup>	.959	.948	.06021	.030	27.713	1	93	.0000
3	.977 <sup>c</sup>	.964	.952	.04745	.004	7.154	1	92	.0089

a. Predictors: (Constant), In (body wt (kg))  
 b. Predictors: (Constant), In (body wt (kg)), In (gestation period)  
 c. Predictors: (Constant), In (body wt (kg)), In (gestation period), In (litter size)  
 d. Dependent Variable: In (brain wt (g))

SPSS regression allows multiple models to be fitted sequentially

### Slide 68 Case Study 9.2

NOTES:

### SPSS syntax Case 9.2

```

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF CI R ANOVA CHANGE
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT Inbrain
/METHOD=ENTER Inbody
/METHOD=ENTER Ingest Inbody
/METHOD=ENTER Ingest Inbody Inlitt
/SCATTERPLOT=(*ZRESID,*ZPRED)
/SAVE PRED COOK RESID .
    
```

### Slide 69 SPSS syntax Case 9.2

NOTES:

### Case Study 9.2

**Testing the effects of gestation length on brain weight, after accounting for body weight and litter size**

Model	Unstandardized Coefficients	Std. Error	Beta	t	Sig.	95% Confidence Interval for B		
						Lower Bound	Upper Bound	
1 (Constant)	2.332	.073		31.843	3.8E-052	2.187	2.478	
2 In(body weight)	-.719	.020	.964	35.300	4.9E-056	.679	.760	
2 (Constant)	2.798	.100		27.945	5.0E-047	2.599	2.997	
3 In(body weight)	.652	.021		31.339	3.2E-051	.610	.693	
3 In(litter size)	-.538	.090		-1.66	5.363	4.4E-008	-.718	-.359
3 (Constant)	3.555	.082		4.292	.000	3.459	3.651	
3 In(body weight)	.575	.033		17.647	2.8E-031	.510	.640	
3 In(litter size)	-.310	.116		-.996	2.675	.009	-.540	-.080
3 In(gestation length)	.418	.141		1.67	2.969	.0038	.138	.698

a. Dependent Variable: ln(brain weight)

### Slide 70 Case Study 9.2

NOTES:

### Extra Sum of Squares F test

= t test for 1 term, but F test can be used for 1 or several terms

**10.3.2 F-Test for the Joint Significance of Several Terms**

$$F\text{-statistic} = \frac{\text{Extra sum of squares}}{\text{Number of betas being tested}} \div \text{Estimate of } \sigma^2 \text{ from full model}$$

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	416.600	1	416.600	1246.104	.000 <sup>a</sup>
1 Residual	31.411	94	.334		
1 Total	447.812	95			
2 Regression	438.689	2	219.344	669.915	.000 <sup>b</sup>
2 Residual	22.723	93	.244		
2 Total	447.812	95			
3 Regression	427.076	3	142.359	631.604	.000 <sup>c</sup>
3 Residual	20.736	92	.225		
3 Total	447.812	95			

$[(22.723 - 20.736) / 1] / .225$   
 ans = 8.8311  
 $2.97 = \sqrt{8.8311}$

a. Predictors: (Constant), ln(body weight)  
 b. Predictors: (Constant), ln(body weight), ln(litter size)  
 c. Predictors: (Constant), ln(body weight), ln(litter size), ln(gestation length)  
 d. Dependent Variable: ln(brain weight)

**EEOS611**

### Slide 71 Extra Sum of Squares F test

NOTES:

### Comparing Full & Reduced models

Equivalence of t tests and F tests for 1 parameter

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	416.600	1	416.600	1246.104	.000 <sup>a</sup>
1 Residual	31.411	94	.334		
1 Total	447.812	95			
2 Regression	438.689	2	219.344	669.915	.000 <sup>b</sup>
2 Residual	22.723	93	.244		
2 Total	447.812	95			
3 Regression	427.076	3	142.359	631.604	.000 <sup>c</sup>
3 Residual	20.736	92	.225		
3 Total	447.812	95			

$[(22.723 - 20.736) / 1] / .225$   
 ans = 8.8311  
 $2.97 = \sqrt{8.8311}$

Model	Unstandardized Coefficients	Std. Error	Beta	t	Sig.	95% Confidence Interval for B	
						Lower Bound	Upper Bound
1 (Constant)	2.33	.07		31.8	.000	2.19	2.48
1 ln(body weight)	-.72	.02		-.96	.000	-.68	-.76
2 (Constant)	2.80	.10		27.9	.000	2.60	3.00
2 ln(body weight)	-.65	.02		-.81	.000	-.61	-.69
2 ln(litter size)	-.54	.09		-.57	.000	-.72	-.35
3 (Constant)	.85	.06		1.3	.000	.68	1.07
3 ln(body weight)	-.68	.03		-.77	.000	-.61	-.64
3 ln(litter size)	-.31	.12		-.27	.009	-.54	.08
3 ln(gestation length)	.42	.14		2.97	.004	.14	.70

a. Dependent Variable: ln(brain weight)

### Slide 72 Comparing Full & Reduced models

NOTES:



Slide 73 Case Study 9.2

Case Study 9.2

Testing the effects of **gestation length on brain weight**, after accounting for **body weight and litter size**

Model	Unstandardized Coefficients		Standardized Coefficients		95% Confidence Interval for B		
	B	Std. Error	Beta	1	Sig.	Lower Bound	Upper Bound
1	2.332	.073		31.843	3.8E-052	2.187	2.478
	.719	.020	.964	35.200	4.9E-056	.679	.760
2	2.798	.100		27.945	5.0E-047	2.599	2.997
	.852	.021	.874	31.539	3.2E-051	.810	.893
	-.533	.090	-.166	5.363	4.4E-008	-.715	-.359
3	.855	.062		1.292	.200	-.459	2.169
	.575	.033	.771	17.647	2.8E-031	.510	.640
	-.310	.116	-.096	2.675	.009	-.540	-.080
	.418	.141	.167	2.969	.0038	-.138	.696

a. Dependent Variable: ln(brain weight)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.964 <sup>a</sup>	.930	.929	37607	.930	1248.104	1	94	.0000
2	.974 <sup>b</sup>	.948	.948	49430	.018	35.663	1	93	.0000
3	.977 <sup>c</sup>	.954	.952	47475	.004	8.813	1	92	.0038

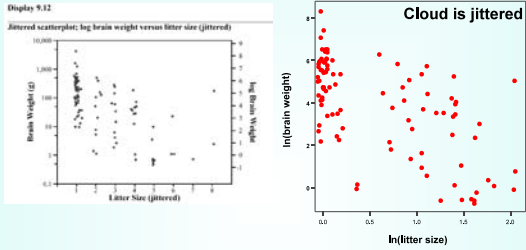
a. Predictors: (Constant), ln (body wt [kg])  
 b. Predictors: (Constant), ln (body wt [kg]), ln (litter size)  
 c. Predictors: (Constant), ln (body wt [kg]), ln (litter size), ln (gestation period)  
 d. Dependent Variable: ln (brain wt [g])

NOTES:

Slide 74 Display 9.12: Jittered plots

Display 9.12: Jittered plots

\Coincident=jitter (5), available in the output editor too



EEOS611

NOTES: