

<p style="text-align: center;">Ch 9: Multiple regression Ch 10 Inferential tools for multiple regression</p> <p style="text-align: center;">Class 16: 4/8/09 W</p>	Slide 1 Ch 9: Multiple regression
	Ch 10 Inferential tools for multiple regression
	NOTES:
<p style="text-align: center;">HW 10 due Friday 4/10/09</p> <p style="text-align: center;">Submit as Myname-HW10.doc (or *.rtf)</p> <ul style="list-style-type: none"> • Computation Problem 10, chapter 8 <ul style="list-style-type: none"> ▸ 8.16 Meat processing, Must assess lack of fit! ▸ Due Friday 4/10 Noon • Read Chapter 9 on multiple regression <ul style="list-style-type: none"> ▸ Read chapter 9 conceptual problems & solutions ▸ Post a question or response about Chapter 9 conceptual problems • HW 11 9.19: Speed of evolution, <ul style="list-style-type: none"> ▸ Due Monday 4/13/09 10 am ▸ This is a TOUGH problem! Weds: ask for help/hints! • Read Chapter 10 <ul style="list-style-type: none"> ▸ Read chapter 10 conceptual problems & solutions ▸ Post a question or response about Chapter 9 conceptual problems 	Slide 2 HW 10 due Friday 4/10/09
	NOTES:
<p style="text-align: center;">Chapter 9</p> <p style="text-align: center;">Multiple regression</p> <p style="text-align: right;">EEOS611</p>	Slide 3 Chapter 9
	NOTES:

Major Issues in Chapter 9

- Using multiple explanatory variables
 - Multiple regression is not a multivariate procedure
 - Multiple regression can produce curvilinear plots, but it is still a linear model. It is linear in the parameters. It is a type of general linear model.
- Using indicator variables
 - Also called dummy variables
 - Sometimes called categorical variables
 - Q treatment levels can be coded by Q-1 indicator variables
 - Must pick one level of a treatment to be the reference category (arbitrary choice, but controls are usually set as the reference)
- Student's *t* test for individual regression terms vs. Extra Sum of Squares F test
 - F test can test for the effects of 1 or more terms
 - Student's *t* tests for single terms
- Tests for interactions
 - Interactions are always tests for differences in slope
 - Interactions are created by multiplying main effects variables

Slide 4 Major Issues in Chapter 9

NOTES:

Multivariate vs. Univariate

From Tabachnik & Fidel

- Variables can be separated into two classes: **explanatory** and response (T&F use **independent** and dependent)[**predictor** or **causal** vs. **outcome**, **stimulus-response**, **input-output**]
 - Univariate statistics: a single response
 - Bivariate: neither is a response (Pearson's *r*, tests of independence)
 - Multivariate: simultaneously analyze multiple explanatory and multiple response variables
- Multivariate statistics are the complete or general case, whereas univariate and bivariate statistics are special cases

EEOS611

Slide 5 Multivariate vs. Univariate

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

Time line for light variation experiment on meadowfoam

EEOS611

Slide 6 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

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	55.3	49.6	39.4	31.3	36.8
	24 days before PFI	77.4	54.2	61.9	45.7	44.9	41.9
24 days before PFI	at PFI	77.8	69.1	57.0	62.9	60.3	52.6
	24 days before PFI	75.6	78.0	71.1	52.2	45.6	44.4

ANOVA: Light intensity would be treated as a categorical variable

Slide 7 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).
```

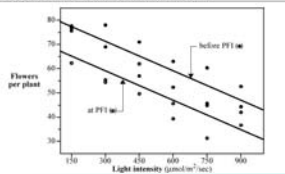
Note that SPSS General linear model (UNIANOVA in syntax) will automatically fit the data, assigning indicator variables for each level of the explanatory variable

Slide 8 SPSS Syntax & dummy variables

NOTES:

Case 9.1 Statistical summary

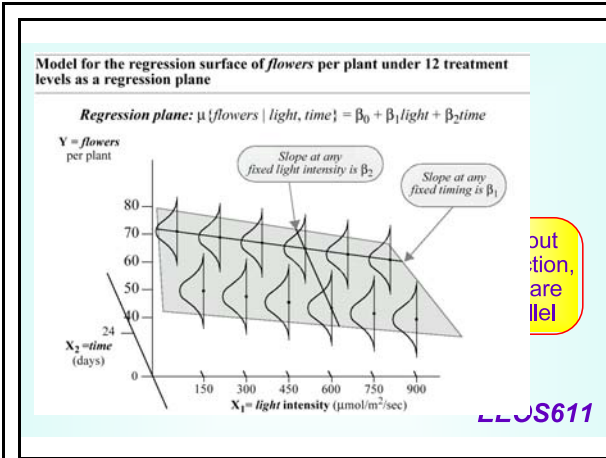
Summary of relationships of flowers produced per plant with increasing light intensities, at and 24 days prior to flower induction



- Increasing light intensity decreased the mean numl of flowers per plant by 4 (plants per 100 $\mu\text{Em}^{-2} \text{s}^{-1}$.
- Beginning light treatment days before PFI increased mean number of flowers by $12.2 (\pm 5.5) (\pm \frac{1}{2} 95\% \text{ CI})$

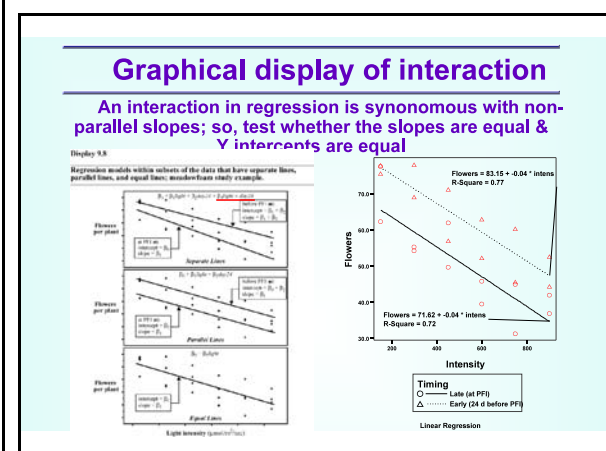
Slide 9 Case 9.1 Statistical summary

NOTES:



Slide 13 Display 9.5

NOTES:



Slide 14 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 15 SPSS: Solving regression problems

NOTES:

Case 9.2 Allometry of brain size

EEOS611

Slide 16 Case 9.2 Allometry of brain size

NOTES:

Display 9.9
A strategy for data analysis using statistical models

Preliminaries: Define the questions of interest. Review the design of the study (or thinking about model assumptions). Correct errors in the data.

1 Explore the data. Look for initial answers to questions and for potential models. *Use graphical tools. Consider transformations; check outliers.*

2 Formulate an inferential model. *Word the questions of interest in terms of model parameters.*

3 Check the model. (a) If appropriate, fit a richer model: (with interactions or curvature, for example); (b) Examine residuals. (c) See if extra terms can be dropped. *Check for non-constant variance; assess outliers. Test whether extra terms in the rich model can be dropped.*

Model OK

4 Infer the answers to the questions of interest using appropriate inferential tools. *Confidence intervals, tests, prediction intervals; calibration intervals (as needed).*

Presentation: Communicate the results to the intended audience. *Answer questions (as much as possible) in subject matter language — not statistical language. Make inferential statements compatible with study design.*

Slide 17 Display 9.9

NOTES:

Display 9.10
Matrix of scatterplots for brain weight data

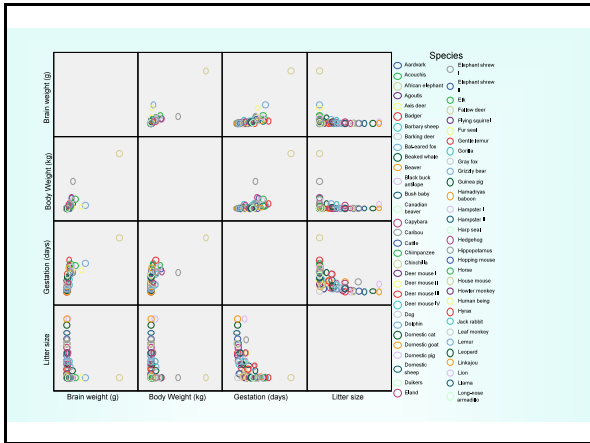
Examine scatter plots for linear relations, outliers & patterns in spread

You must deal with influential data points in creating a regression model. They can't be ignored

Slide 18 Display 9.10

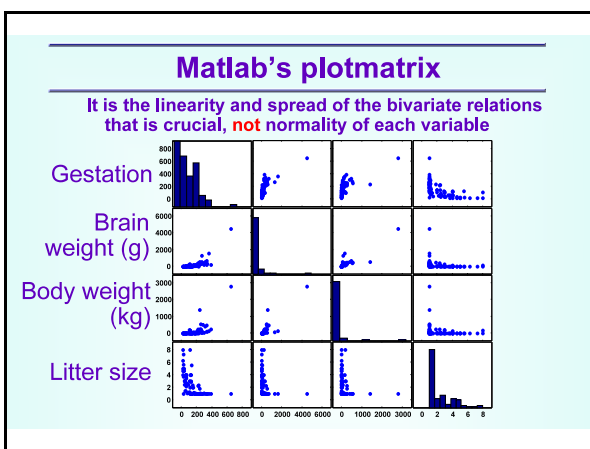
NOTES:

Slide 19



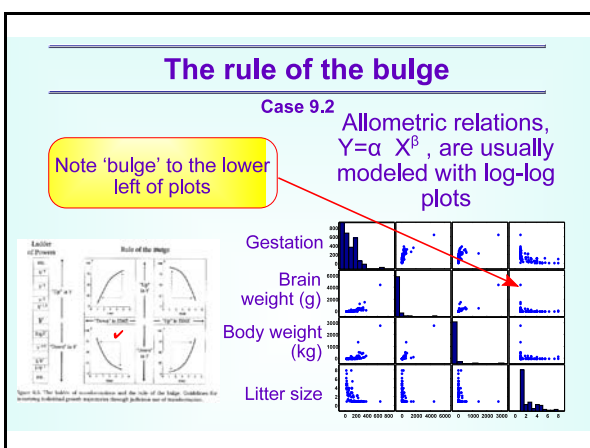
NOTES:

Slide 20 Matlab's plotmatrix

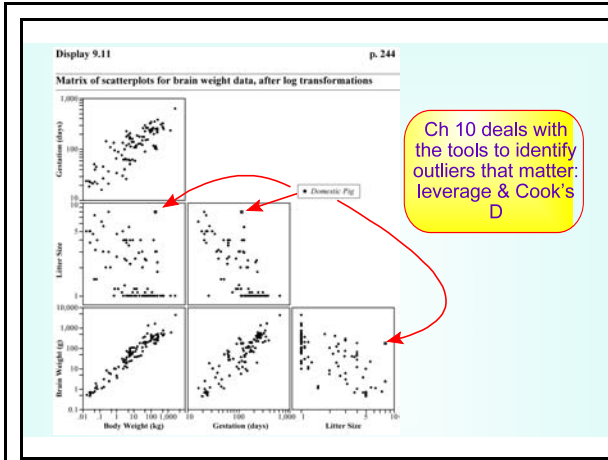


NOTES:

Slide 21 The rule of the bulge

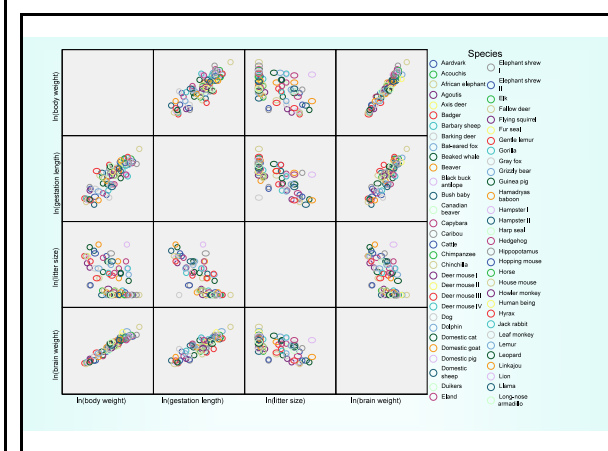


NOTES:



Slide 22 Display 9.11

NOTES:



Slide 23

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	Lower Bound	Upper Bound
1						
	(Constant)	2.332	.073		2.187	2.478
	In (body wt [kg])	.719	.020	.964	35.300	.0000
2						
	(Constant)	-.487	.458		-.997	.3212
	In (body wt [kg])	.591	.022	.739	17.023	.0000
	In (gestation period)	.688	.109	.266	6.141	.0000
3						
	(Constant)	.855	.662		1.292	.1996
	In (body wt [kg])	.575	.033	.771	17.647	.0000
	In (gestation period)	.418	.141	.167	2.969	.0036
	In (litter size)	-.310	.116	-.096	-2.675	.0088

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 ^a	.930	.929	.27607	.330	1246.104	1	84	.0000
2	.975 ^b	.950	.949	.49021	.029	37.713	1	83	.0000
3	.977 ^c	.954	.952	.47475	.004	7.154	1	92	.0089

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

SPSS regression allows multiple models to be fitted sequentially

Slide 24 Case Study 9.2

NOTES:

SPSS syntax Case 9.2

Can be done with GUI's, but syntax quicker

```

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 25 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	Standardized Coefficients	t	Sig.	95% Confidence Interval for B			
					B	Std. Error	Lower Bound	Upper Bound
1	(Constant)	2.332	.073	31.843	3.8E-052	2.187	2.478	
	In(body weight)	.719	.020	35.300	4.9E-056	.679	.760	
2	(Constant)	2.798	.100	27.945	5.0E-047	2.599	2.997	
	In(body weight)	.652	.021	31.339	3.2E-051	.610	.693	
	In(litter size)	-.538	.090	-.166	-.5.953	4.4E-008	-.718	-.359
3	(Constant)	.855	.682	1.292	.200	-.459	2.169	
	In(body weight)	.575	.033	.771	17.847	2.8E-031	.510	.640
	In(litter size)	-.310	.116	-.096	-2.675	.009	-.540	-.080
	In(gestation length)	.418	.141	.167	2.969	.0038	.138	.698

a. Dependent Variable: In(brain weight)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	F Change	df1	df2	Sig.	Change Statistics	
									Change in R Square	Sig. F Change
1	.964 ^a	.930	.929	.57807	930	1246	94	.000		
2	.974 ^b	.949	.948	.49430	910	39	92	.000		
3	.977 ^c	.954	.952	.47475	.004	8.813	1	.92	.0038	

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

Slide 26 Case Study 9.2

NOTES:

Extra Sum of Squares F test

= t test for 1 term, but F test can be used to test for the importance of 1 or several terms

10.3.2 F-Test for the Joint Significance of Several Terms

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

Model	Sum of Squares	df	Mean Square	F	Sig.	
1	Regression	416.400	1	416.400	1246.104	.000 ^a
	Residual	31.411	94	.334		
	Total	447.812	95			
2	Regression	426.089	2	212.545	869.815	.000 ^b
	Residual	20.723	93	.224		
	Total	447.812	95			
3	Regression	427.076	3	142.359	631.604	.000 ^c
	Residual	20.736	92	.225		
	Total	447.812	95			

a. Predictors: (Constant), Inbody weight
 b. Predictors: (Constant), Inbody weight, Inlitter size
 c. Predictors: (Constant), Inbody weight, Inlitter size, Ingestation length
 d. Dependent Variable: Inbrain weight

EEOS611

$$[(22.723 - 20.736) / 1] / .225$$

$$\text{ans} = 8.8311$$

$$2.97 = \text{sqrt}(8.8311)$$

Slide 27 Extra Sum of Squares F test

NOTES:

Comparing Full & Reduced models

Equivalence of t tests and F tests for 1 parameter

ANOVA ^a					
Model	Sum of Squares	df	Mean Square	F	Sig.
1	Regression 31411 Total 447.812	1 95	416.400 .346	1246.104	.000 ^b
2	Regression 148.881 Residual 22.723 Total 447.812	2 93	212.888 .244	689.915	.000 ^b
3	Regression 427.276 Residual 20.726 Total 447.812	3 92	142.359 .225	631.604	.000 ^b

$$[(22.723 - 20.726) / 1] / .225$$

$$\text{ans} = 8.8311$$

$$2.97 = \sqrt{8.8311}$$

a. Predictors: (Constant), ln(body weight), ln(litter size)

b. Predictors: (Constant), ln(body weight), ln(litter size), ln(gestation length)

c. Dependent Variable: ln(brain weight)

Model	B	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
	ln(body weight)	.72	.02	36.3	.000	.68	.76
	ln(litter size)	.65	.02	31.3	.000	.61	.69
2	(Constant)	2.80	.10	27.9	.000	2.60	3.00
	ln(body weight)	.65	.02	31.3	.000	.61	.69
	ln(litter size)	-.54	.09	-6.0	.000	-.72	-.36
3	(Constant)	.85	.06	1.3	.200	-.46	2.17
	ln(body weight)	.68	.03	17.6	.000	.61	.74
	ln(litter size)	-.31	.12	-2.7	.009	-.54	-.08
	ln(gestation length)	.42	.14	3.07	.004	.14	.70

a. Dependent Variable: ln(brain weight)

Slide 28 Comparing Full & Reduced models

NOTES:

Case Study 9.2

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

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error				Lower Bound	Upper Bound
1	(Constant)	2.332	.073		31.843	3.9E-052	2.187	2.478
	ln(body weight)	.719	.020	.964	35.300	4.9E-056	.679	.760
	(Constant)	2.798	.100		27.945	5.0E-047	2.599	2.997
2	ln(body weight)	.652	.021	.874	31.339	3.2E-051	.610	.695
	ln(litter size)	-.538	.090	-.166	-5.963	4.4E-008	-.716	-.359
	(Constant)	2.855	.062		1.282	.200	-.459	2.169
3	ln(body weight)	.575	.033	.771	17.647	2.8E-031	.510	.640
	ln(litter size)	-.310	.116	-.096	-2.675	.009	-.540	-.080
	ln(gestation length)	.418	.141	.167	2.969	.0038	.136	.696
	(Constant)	2.855	.062		1.282	.200	-.459	2.169

a. Dependent Variable: ln(brain weight)

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	F Change	df1	df2	Sig. F Change	
1	.964 ^a	.930	.929	.87807	830	1248.104	1	94	.0000
2	.974 ^a	.948	.948	.84930	819	33.382	1	93	.0000
3	.977 ^a	.954	.952	.84775	804	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])

Change Statistics: t-test

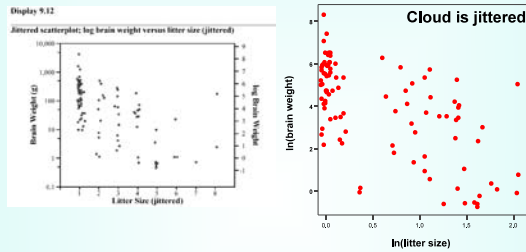
Extra Sum of Squares F test

Slide 29 Case Study 9.2

NOTES:

Display 9.12: Jittered plots

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



EEOS611

Slide 30 Display 9.12: Jittered plots

NOTES:

Inferential Tools for Multiple Regression

Sleuth Chapter 10

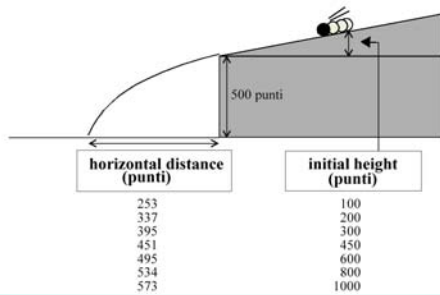
EEOS611

Slide 31 Inferential Tools for Multiple Regression

NOTES:

Display 10.1

Galileo's experimental results Published 1609



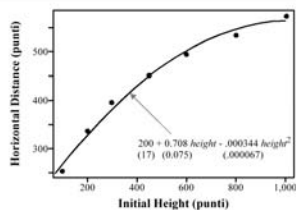
Slide 32 Display 10.1

NOTES:

Statistical summary: while there is evidence for a cubic term in the regression model, the quadratic model explains 99.03% of the variation in horizontal distance, and the cubic term explains only an additional 0.91% of the variation in distance

Display 10.2

Scatterplot of Galileo's horizontal distances versus initial heights, with estimated quadratic regression model (with standard errors in parentheses)



Slide 33 Display 10.2

NOTES:

Display 10.11

Analysis of variance table for Galileo's data: fit of the data to the quadratic model: $\mu(\text{distance} | \text{height}) = \beta_0 + \beta_1 \text{height} + \beta_2 \text{height}^2$; based on 7 observations

① Residual sum of squares from reduced model: $\text{Mean}(\text{distance}) = \beta_0$

② Residual sum of squares from full model: $\text{Mean}(\text{distance}) = \beta_0 + \beta_1 \text{height} + \beta_2 \text{height}^2$

③ The extra sum of squares is the total sum of squares minus the residual sum of squares (the amount of variation explained by the 2 explanatory variables).

④ Mean Squares are always sums of squares divided by their degrees of freedom. The Residual Mean Square is the estimate of σ^2 .

⑤ The F-statistic (for overall significance of regression) is the Regression Mean Square divided by the Residual Mean Square.

⑥ $p\text{-value} = \text{Pr}(F_{2,4} > 205.03)$. The small p-value here indicates overwhelming evidence that the coefficient of at least one of the explanatory variables height and height² is different from zero.

Source of Variation	Sum of Squares	df	Mean Square	F-Statistic	p-value
Regression	76277.92	2	38138.96	205.03	<.0001
Residual	744.08	4	186.02		
Total	77022.00	6			

Slide 34 Display 10.11

NOTES:

SPSS: Cubic & Quadratic fits

Strong evidence that the coefficient for the cubic term is not zero ($p=0.007$)

Unstandardized Predicted Value

Initial height (punti)

Quadratic fit

Cubic fit

Cubic term, X^3 , models an S-shaped pattern; Quadratic term, X^2 , models a unimodal (hump-shaped) pattern

Slide 35 SPSS: Cubic & Quadratic fits

NOTES:

Display 10.13

Partial output from the regression of distance on height, height-squared, and height-cubed, for Galileo's Data

Variable	Coefficient	Standard Error	t-statistic	p-value
Constant	155.78	8.33	18.71	0.0003
height	1.1153	0.0657	16.98	0.0004
height-squared	-0.001245	0.000138	-8.99	0.0029
height-cubed	5.477x10 ⁻⁷	0.838x10 ⁻⁷	6.58	0.0072

Estimate of standard deviation about the regression: 4.011 on 3 degrees of freedom

$R^2 = 99.94\%$

Model	Unstandardized Coefficients	Standardized Coefficients	t	Sig.	95% Confidence Interval for B		
					Beta	Lower Bound	Upper Bound
1	(Constant)		11.09	.000	207.215	332.209	
	Initial Height (punti)	.333	.042	7.93	.001	.225	.441
2	(Constant)		11.93	.000	153.381	246.444	
	Initial Height (punti)	.703	.075	9.47	.001	.501	.916
	heightsq	.000	.000	-1.112	.007	-.001	.000
3	(Constant)		18.71	.000	129.279	182.272	
	Initial Height (punti)	1.115	.066	3.220	0.003	.966	1.324
	heightsq	-.001	.0001	-4.028	.003	-.002	-.001
	heightcb	5.48E-007	8.3E-008	6.58	.007	2.8E-007	8.1E-007

a. Dependent Variable: Horizontal Distance (punti)

Slide 36 Display 10.13

NOTES:

Scatterplots indicate problems

Quadratic fit, left, Cubic fit right
Cook's D, (Ch 11: D>1 cause for concern, Sleuth p 320)

Both the quadratic and cubic models are strongly influenced by one datum: Case 7, with a high Cook's D (identifies an influential datum).

Slide 37 Scatterplots indicate problems

NOTES:

Extra sum of squares F test

R² change

Model	Sum of Squares	df	Mean Square	F	Sig.
1	Regression 71350.794	1	71350.794	62.906	.001 ^a
	Residual 5671.206	5	1134.241		
	Total 77022.000	6			
2	Regression 75277.922	2	38138.961	205.027	9.33E-009 ^b
	Residual 744.078	4	186.020		
	Total 77022.000	6			
3	Regression 78973.740	3	25957.915	1595.189	2.69E-005 ^c
	Residual 48.254	3	16.085		
	Total 77022.000	6			

$(744.078 - 48.254) / 1 = 43.26$
 Test with $F_{(1,3 \text{ df})}$
 $p = 0.0072$
 Equivalent to t test
 $6.577 = \sqrt{43.260}$

EEOS611

Slide 38 Extra sum of squares F test

NOTES:

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.962 ^a	.9264	.9116	33.678	.5264	62.906	1	5	.001
2	.995 ^b	.9903	.9855	13.639	.0640	26.487	1	4	.007
3	1.000 ^c	.9994	.9987	4.011	.0090	43.260	1	3	.007

a. Predictors: (Constant), Initial Height (punti)
 b. Predictors: (Constant), Initial Height (punti), heightsq
 c. Predictors: (Constant), Initial Height (punti), heightsq, heightscb
 d. Dependent Variable: Horizontal Distance (punti)

Linear Regression: Statistics

Regression Coefficients: Model fit

Estimates R-squared change

Confidence intervals Descriptives

Collinearity statistics Fit and partial correlations Collinearity diagnostics

Buttons: Continue, Cancel, Help

$R^2 = 100 \text{ (Total Sum of Squares) - (Residual Sum of Squares) \%}$
 $\text{Total Sum of Squares}$
 = **Coefficient of Determination**
 Percentage of total response variation explained by the regression

$\text{Adjusted } R^2 = 100 \text{ (Total Mean Square) - (Residual Mean Square) \%}$
 Total Mean Square

Adjusted R² is affected by the number of parameters

Slide 39

NOTES:

Slide 40

$$R^2 = 100 \frac{\text{Total Sum of Squares} - (\text{Residual Sum of Squares})}{\text{Total Sum of Squares}} \%$$

$$\text{Adjusted } R^2 = 100 \frac{\text{Total Mean Square} - (\text{Residual Mean Square})}{\text{Total Mean Square}} \%$$

ANOVA ^d						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	71350.8	1	71351	63	.001 ^a
	Residual	5671.2	5	1134		
	Total	77022.0	6			
2	Regression	75277.9	2	38139	205	9.33E-005 ^b
	Residual	744.1	4	186		
	Total	77022.0	6			
3	Regression	76973.7	3	25658	1595	2.66E-005 ^c
	Residual	48.3	3	16		
	Total	77022.0	6			

$$\text{Adjusted } R^2 = 100 * \frac{(77022/6) - 16}{(77022/6)} = 99.875\%$$

- a. Predictors: (Constant), Initial Height (punti)
- b. Predictors: (Constant), Initial Height (punti), heightsq
- c. Predictors: (Constant), Initial Height (punti), heightsq, heightcub
- d. Dependent Variable: Horizontal Distance (punti)

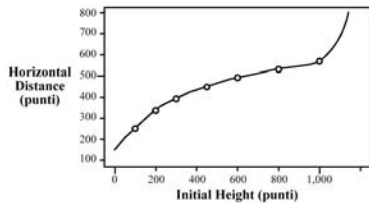
Adjusted R² introduces a penalty for number of parameters

NOTES:

Slide 41 Display 10.14

Display 10.14

Scatterplot of Galileo's horizontal distances versus initial heights, with estimated sixth-order polynomial regression curve (R² = 100%)



7 data can always be fit perfectly with a 6th-order polynomial (Y=Constant+B*X+...+X⁶). A high order polynomial model, while offering a better fit, often is a poor predictive model

NOTES:

Slide 42 Estimating the predicted value and standard error for 250 punti

Estimating the predicted value and standard error for 250 punti

Display 10.7

Estimates of polynomial coefficients with two different references levels of height, in Galileo's study

(Reference height = 0)				
Variable	Coefficient	Standard Error	t-Statistic	2-Sided p-Value
CONSTANT	199.91	16.76	11.93	.0003
height	0.7083	0.0748	9.47	.0007
height ²	-0.0003437	0.0000668	5.15	.0068
R-squared = 99.0%		adj. R-squared = 98.6%	Estimated SD = 13.6	
$\hat{\mu}\{\text{distance} \mid \text{height} = 250\}$ $SE\{\hat{\mu}\{\text{distance} \mid \text{height} = 250\}\}$				
(Reference height = 250)				
Variable	Coefficient	Standard Error	t-Statistic	2-Sided p-Value
CONSTANT	355.51	4.62	53.66	<.0001
height - 250	0.5365	0.0430	12.48	.0002
(height - 250) ²	-0.0003437	0.0000668	5.15	.0068
R-squared = 99.0%		adj. R-squared = 98.6%	Estimated SD = 13.6	

NOTES:

Plotting a supplemental case

1 of 3

SPSS readily allows the analysis of supplemental cases, cases with new values for the explanatory variables. The estimates & standard errors are calculated

	distance	height	heightsq	heighttcb
1	253	100	10000.00	1000000
2	337	200	40000.00	8000000
3	395	300	90000.00	27000000
4	451	450	202500.0	91125000
5	495	600	360000.0	2.16E+08
6	534	800	640000.0	5.12E+08
7	573	1000	1000000	1.00E+09
8	.	250	62500.00	15625000

Slide 43 Plotting a supplemental case

NOTES:

Case 10.1: Plotting predicted values

2 of 3

Slide 44 Case 10.1: Plotting predicted values

NOTES:

SPSS will solve for the predicted value, standard error, & the CI's for the mean and for individual data

3 of 3

	distance	height	heightsq	heighttcb	pre_1	std. error	lower	upper
1	253	100	10000.00	1000000	307.30813	10.55175	218.77907	315.83720
2	337	200	40000.00	8000000	327.92928	7.36396	284.78926	370.96930
3	395	300	90000.00	27000000	351.47715	6.52482	339.49926	423.45503
4	451	450	202500.0	91125000	443.92989	7.72136	436.54612	492.57486
5	495	600	360000.0	2.16E+08	521.17661	8.20107	498.99922	548.36400
6	534	800	640000.0	5.12E+08	548.93989	7.69997	523.24999	598.96469
7	573	1000	1000000	1.00E+09	564.54168	12.63250	513.11486	615.96936
8	.	250	62500.00	15625000	365.51280	6.62490	313.41407	397.61113

Standard error for prediction Lower & upper individual prediction limits

Slide 45 SPSS will solve for the predicted value, standard error, & the CI's for the mean and for individual data

NOTES:

Slide 46 Case 10.2

Case 10.2

Energy of echolocating bats:
Do they require more energy than non-echolocating bats or birds, after accounting for the effects of body mass on energy consumption?

EEOS611

NOTES:

Slide 47 Display 10.3

Display 10.3

Mass and in-flight energy expenditure for 4 non-echolocating bats (Type = 1), 12 non-echolocating birds (Type = 2), and 4 echolocating bats (Type = 3)

Species	Mass (g)	Type	Flight Energy Expenditure (W)
<i>Pteropus gouldii</i>	779	1	21.7
<i>Pteropus poliocephalus</i>	628	1	34.8
<i>Myiagnotus monstrosus</i>	258	1	23.3
<i>Lidolon helvum</i>	315	1	22.4
<i>Meliphaga virescens</i>	24.3	2	2.46
<i>Melipittacus undulatus</i>	35	2	3.93
<i>Sturnus vulgaris</i>	72.8	2	9.15
<i>Falco sparverius</i>	120	2	13.8
<i>Falco tinnunculus</i>	213	2	14.6
<i>Corvus ossifragus</i>	275	2	22.8
<i>Larus atricilla</i>	370	2	26.2
<i>Columba livia</i>	384	2	25.9
<i>Columba livia</i>	442	2	29.5
<i>Columba livia</i>	412	2	43.7
<i>Columba livia</i>	330	2	34.0
<i>Corvus cryptoleucus</i>	480	2	27.8
<i>Phyllostomus hastatus</i>	93	3	8.83
<i>Plecotus auritus</i>	8	3	1.35
<i>Pipistrellus pipistrellus</i>	6.7	3	1.12
<i>Plecotus auritus</i>	7.7	3	1.02

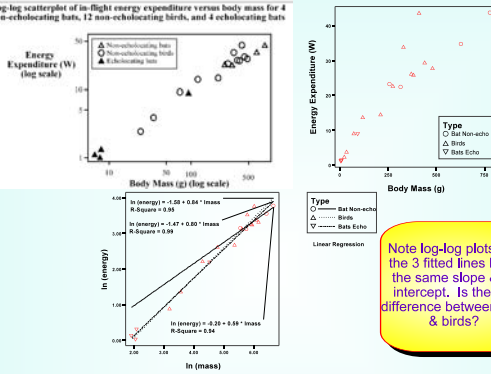
Note that all type is categorical variable with 3 categories. It need 2 indicator variables to code.

NOTES:

Slide 48 Display 10.4

Display 10.4

Log-log scatterplot of in-flight energy expenditure versus body mass for 4 non-echolocating bats, 12 non-echolocating birds, and 4 echolocating bats



Note log-log plots: Do the 3 fitted lines have the same slope & Y-intercept. Is there a difference between bats & birds?

NOTES:

Display 10.5

The parallel regression lines model for the bat echolocation data

1) Test whether the 3 slopes differ (2 interaction terms)
2) Test whether 3 Y-intercepts differ

Slide 49 Display 10.5

NOTES:

Display 10.12

The extra sum of squares F-test comparing the separate regression lines model to the parallel regression lines model; bat echolocation data

The models differ in 2 interaction terms

1) FIT FULL MODEL: $\mu(\text{energy} | \text{mass}, \text{TYPE}) = \beta_0 + \beta_1 \text{mass} + \beta_2 \text{bird} + \beta_3 \text{bat} + \beta_4 \text{mass} \times \text{bird} + \beta_5 \text{mass} \times \text{bats}$

Source of Variation	Sum of Squares	df	Mean Square	F-Statistic	p-value
Regression	29.46993	5	5.89399	163.4	<.0001
Residual	.50487	14	.03606		
Total	29.97480	19			

2) FIT REDUCED MODEL: $\mu(\text{energy} | \text{mass}, \text{TYPE}) = \beta_0 + \beta_1 \text{mass} + \beta_2 \text{bird} + \beta_3 \text{bat}$

Source of Variation	Sum of Squares	df	Mean Square	F-Statistic	p-value
Regression	29.42148	3	9.80716	283.6	<.0001
Residual	.55332	16	.03458		
Total	29.97480	19			

Slide 50 Display 10.12

NOTES:

Display 10.12

The extra sum of squares F-test comparing the separate regression lines model to the parallel regression lines model; bat echolocation data

3) The extra sum of squares is the difference between residual sums of squares

Extra SS = .55332 - .50487 = .04845

5) Calculate the F-Statistic

$$F\text{-Statistic} = \frac{.04845}{2} \div \frac{.03606}{16} = .672$$

6) Look up $\text{Pr}(F_{2,14} > 0.672)$ → p-value = 0.53

Conclusion: There is no evidence that the association between energy expenditure and body size is different for the three types of flying vertebrates (p-value = 0.53).

Slide 51

NOTES:

t tests: Are interaction terms zero?

Parallel slopes model, Can't rely on the t statistic alone to judge whether both interaction terms can be dropped

Model		Unstandardized Coefficients		Standardized Coefficients		95% Confidence Interval for B		
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	-1.468	.137		-10.705	3.1E-009	-1.76	-1.18
	In (mass)	.809	.027	.990	30.127	7.4E-017	.75	.86
	Birds	-.152	.114	.041	.998	.364	-.14	.34
2	(Constant)	-1.576	.287		-5.488	5.0E-005	-2.19	-.97
	In (mass)	.815	.045	.998	18.297	3.8E-012	.72	.91
	Birds	-.102	.114	.041	.998	.364	-.14	.34
3	(Constant)	-2.02	1.261		-.161	.875	-2.91	2.50
	In (mass)	.590	.206	.722	2.861	.013	.15	1.03
	Birds	-1.378	1.295	-.552	-1.064	.305	-4.16	1.40
	Echolocating bats	-1.268	1.285	-.414	-.987	.341	-4.03	1.49
	In(bn (Mass, Birds v. Bats)	.246	.213	.536	1.151	.269	-.21	.70
	In(bn (Mass, Echolocating bats)	.215	.224	.204	.961	.353	-.28	.69

Two 1-coefficient-at-a-time t tests can have p values > 0.05, but both together can have p<0.05. Need extra Sum of Squares F test with combined df in numerator.

Slide 52 t tests: Are interaction terms zero?

NOTES:

Extra sum of squares F test (=Partial F test) for both interaction terms

Model	R	R Square	Adjusted R Square		Change Statistics		F Change	df1	df2	Sig. F Change
			the Estimate	R Square Change	F Change	df1				
1	.990 ^a	.981	.979	.17955	.981	907.638	1	18	.000	
2	.991 ^b	.982	.978	.18596	.001	.428	2	16	.659	
3	.992 ^c	.983	.977	.18990	.002	.672	2	14	.527	

- a. Predictors: (Constant), In (mass)
- b. Predictors: (Constant), In (mass), Birds not Bats, Bats (echo) v. Bats (non-echo)
- c. Predictors: (Constant), In (mass), Birds not Bats, Bats (echo) v. Bats (non-echo), In(bn (Mass, Twobats), In(bn (Mass, Birds v. Bats)
- d. Dependent Variable: In (Energy)

Slide 53 Extra sum of squares F test (=Partial F test) for both interaction terms

NOTES:

Do echolocating bats differ from non-echolocating bats in energy expenditure?

Non-echolocating bats are the reference category: Little evidence (t test, p=0.7) that the difference in Y intercepts ≠ 0, so conclude that there is little evidence that echolocating and non-echolocating bats differ in energy expenditure.

Model		Unstandardized Coefficients		Standardized Coefficients		95% Confidence Interval for B		
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	-1.468	.137		-10.705	.000	-1.766	-1.180
	In (mass)	.809	.027	.990	30.127	.000	.752	.865
	Birds	-.152	.114	.041	.998	.364	-.140	.344
2	(Constant)	-1.576	.287		-5.488	.000	-2.185	-.967
	In (mass)	.815	.045	.998	18.297	.000	.721	.909
	Birds	-.102	.114	.041	.998	.364	-.140	.344
3	(Constant)	-2.02	1.261		-.161	.875	-2.908	2.503
	In (mass)	.590	.206	.722	2.861	.013	.148	1.032
	Birds	-1.378	1.295	-.552	-1.064	.305	-4.156	1.400
	Echolocating bats	-1.268	1.285	-.414	-.987	.341	-4.029	1.469
	Bats x mass	.246	.213	.536	1.151	.269	-.212	.703
	Ebat x mass	.215	.224	.204	.961	.353	-.285	.694

a. Dependent Variable: In (Energy)

Slide 54 Do echolocating bats differ from non-echolocating bats in energy expenditure?

NOTES:

Display 10.10

The extra-sum-of-squares F-test for testing equality of intercepts in the parallel regression lines model; bat echolocation data

1. Fit the FULL model: $\mu(\text{energy} | \text{Inmass, TYPE}) = \beta_0 + \beta_1 \text{Inmass} + \beta_2 \text{bird} + \beta_3 \text{bat}$

Sum of squared residuals = .55332 d.f. = 16 $\hat{\sigma}^2 = .03458$

2. Fit the REDUCED model: $\mu(\text{energy} | \text{Inmass, TYPE}) = \beta_0 + \beta_1 \text{Inmass}$

Sum of squared residuals = .58289 d.f. = 18

3. The extra sum of squares is the difference between the two residual sums of squares

→ Extra SS = .58289 - .55332 = .02957

4. Numerator degrees of freedom are the number of β 's in the full model minus the number of β 's in the reduced model.

→ Numerator d.f. = 4 - 2 = 2

5. Calculate the F-Statistic

→ F-Statistic = $\frac{.02957}{2} \div \frac{.03458}{.03458} = \frac{.014785}{.03458} = .428$

6. Find $P(F_2, 16) > .428$ from table, computer, or calculator

→ p-Value = .66

Conclusion: There is no evidence that mean log energy differs for birds, echolocating bats, and non-echolocating bats, after accounting for body mass.

Extra sum of squares F test for different Y intercepts

Model	Sum of Squares	df	Mean Square
1	29.392	1	29.392
Residual	.553	16	.032
Total	29.975	19	
2	29.421	3	9.807
Residual	.553	16	.035
Total	29.975	19	

Slide 55 Display 10.10

NOTES:

Display 10.6

Partial summary of the least squares fit to the regression of log energy expenditure on log body mass, an indicator variable for bird, and an indicator variable for echolocating bat

Variable	Coefficient	Standard Error	t-Statistic	2-Sided p-Value
CONSTANT	-1.5764	0.2872	5.4880	<0.0001
Inmass	0.8150	0.0445	18.2966	<0.0001
bird	0.1023	0.1142	0.8956	0.3837
ebat	0.0787	0.2027	0.3881	0.7030

Estimate of $\sigma = 0.1860$, df = 16

Coefficients^a

Unstandardized Coefficients	B	Std. Error	t	Sig.	95% Confidence Interval for B	
					Lower Bound	Upper Bound
(Constant)	-1.576	.287	-5.488	4.96E-005	-2.19	-.97
Echolocating bats	.079	.203	.388	.703	-.35	.51
Birds	.102	.114	.896	.384	-.14	.34
In (mass)	.815	.045	18.297	3.76E-012	.72	.91

a. Dependent Variable: ln (energy)

Slide 56 Display 10.6

NOTES:

Regression σ estimated from $\sqrt{\text{Residual mean square}}$

Model Summary^a

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics	F	Sig.	F Change
1	.990 ^c	.981	.979	.17995	1	907.638	.000	
2	.991 ^b	.982	.978	.18078	2	16	.659	
3	.992 ^c	.983	.977	.18092	2	14	.527	

a. Predictors: (Constant), ln (mass)

b. Predictors: (Constant), ln (mass), Birds not Bats, Bats (echo) v. Bats (non-echo)

c. Predictors: (Constant), ln (mass), Birds not Bats, Bats (echo) v. Bats (non-echo), ln (Mass, Twobats), ln (Mass, Birds v. Bats)

d. Dependent Variable: ln (Energy)

Note that mean squares are variances, the square root of the residual mean square provides the standard error for the regression.

Model	Sum of Squares	df	Mean Square	F	Sig.
1	29.392	1	29.392	907.638	7.44E-011 ^c
Residual	.553	16	.032		
Total	29.975	19			
2	29.421	3	9.807	283.589	4.46E-014 ^b
Residual	.553	16	.035		
Total	29.975	19			
3	29.470	5	5.894	183.440	6.70E-012 ^c
Residual	.553	14	.036		
Total	29.975	19			

a. Predictors: (Constant), ln (mass)

b. Predictors: (Constant), ln (mass), Birds not Bats, Bats (echo) v. Bats (non-echo)

c. Predictors: (Constant), ln (mass), Birds not Bats, Bats (echo) v. Bats (non-echo), ln (Mass, Twobats), ln (Mass, Birds v. Bats)

d. Dependent Variable: ln (Energy)

Slide 57 Regression σ estimated from $\sqrt{\text{Residual mean square}}$

NOTES:

Display 10.6

Partial summary of the least squares fit to the regression of log energy expenditure on log body mass, an indicator variable for bird, and an indicator variable for echolocating bat

Variable	Coefficient	Standard Error	t-Statistic	2-Sided p-Value
CONSTANT	-1.5764	0.2872	5.4880	<0.0001
lmass	0.8150	0.0445	18.2966	<0.0001
bird	0.1023	0.1142	0.8956	0.3837
ebat	0.0787	0.2027	0.3881	0.7030

Estimate of $\sigma = 0.1860$, $df = 16$

How do you estimate sigma, σ , standard error of the estimate, root mean square error for the regression, standard error for the regression?

Slide 58

NOTES:

Model Summary^a Change Statistics

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.990 ^a	.981	.979	.17995	.981	907.638	1	18	.000
2	.991 ^b	.982	.978	.18596	.001	.428	2	16	.659
3	.992 ^c	.983	.977	.18990	.002	.672	2	14	.527

a. Predictors: (Constant), ln (mass)
 b. Predictors: (Constant), ln (mass), Birds not Bats, Bats (echo) v. Bats (non-echo)
 c. Predictors: (Constant), ln (mass), Birds not Bats, Bats (echo) v. Bats (non-echo), ln(xn (Mass, Twobats), ln(xn (Mass, Birds v. Bats))

ANOVA^a

Model	Sum of Squares	df	Mean Square	F	Sig.
1	Regression 29.392	1	29.392	907.638	7.44E-017*
	Residual .583	18	.032		
	Total 29.975	19			
2	Regression 29.421	3	9.807	283.589	4.46E-014*
	Residual .553	16	.035		
	Total 29.975	19			
3	Regression 29.470	5	5.894	163.440	6.70E-012*
	Residual .505	14	.036		
	Total 29.975	19			

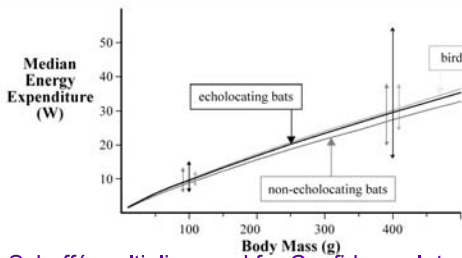
a. Predictors: (Constant), ln (mass)
 b. Predictors: (Constant), ln (mass), Birds not Bats, Bats (echo) v. Bats (non-echo)
 c. Predictors: (Constant), ln (mass), Birds not Bats, Bats (echo) v. Bats (non-echo), ln(xn (Mass, Twobats), ln(xn (Mass, Birds v. Bats))
 d. Dependent Variable: ln (Energy)

Slide 59

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-log scale, with 95% confidence bands



Scheffé multiplier used for Confidence Intervals

Slide 60 Display 10.8

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 CI's: somewhat atypical, but appropriate

Slide 61 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 62 Display 10.8

NOTES:

1007 [DataView1] - 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 ∞ sample size. 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 63

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.

Slide 64 Syntax: Scheffé multiplier

NOTES:

Variance formulae for linear contrasts

COMPUTE Schlnt = SEP_1 * Scheffemultiplier

Display 18.15

Inference about $\beta_2 - \beta_3$, the coefficient of the indicator variable for birds minus the coefficient of the indicator variable for echolocating bats

Estimate of $\beta_2 - \beta_3$ from: .01304 - .04108 = -0.02804

Estimated variance-covariance matrix (from regression):

	(Constant)	In (mass)	In (energy)	Birds	Echolocating bats
(Constant)	.08320	-.01211	-.05923	.00687	-.04108
In (mass)	-.01211	.00198	.00173	-.01864	-.05056
In (energy)	-.05923	.00173	.01304	-.01864	-.05056
Birds	.00687	-.01864	-.01864	.01304	-.04108
Echolocating bats	-.04108	-.05056	-.05056	-.04108	.01304

Estimated variance of $\beta_2 - \beta_3$: $.01304 + .04108 - 2*0.0444 = .02484$
 $SE(\beta_2 - \beta_3) = (.02484)^{.5} = .1576$ (df = 16)

Slide 65 Variance formulae for linear contrasts

NOTES:

Echolocating Bats as Reference

Model	Unstandardized Coefficients		Standardized Coefficients		t	Sig.	95% Confidence Interval for B	
	B	Std. Error	Beta				Lower Bound	Upper Bound
1								
	(Constant)	-1.498	.1499		-9.993	.000	-1.815	-1.180
	In (mass)	.815	.0445	.998	18.297	.000	.721	.909
	Non-echolocating bats	-.0787	.0227	-.026	-.388	.703	-.508	-.351
	Birds	.0236	.1576	.009	.150	.882	-.310	.358

a. Dependent Variable: In (energy)

Display 18.15

Inference about $\beta_2 - \beta_3$, the coefficient of the indicator variable for birds minus the coefficient of the indicator variable for echolocating bats

Estimate of $\beta_2 - \beta_3$ from: .01304 - .04108 = -0.02804

Estimated variance-covariance matrix (from regression):

	(Constant)	In (mass)	In (energy)	Birds	Echolocating bats
(Constant)	.08320	-.01211	-.05923	.00687	-.04108
In (mass)	-.01211	.00198	.00173	-.01864	-.05056
In (energy)	-.05923	.00173	.01304	-.01864	-.05056
Birds	.00687	-.01864	-.01864	.01304	-.04108
Echolocating bats	-.04108	-.05056	-.05056	-.04108	.01304

Estimated variance of $\beta_2 - \beta_3$: $.01304 + .04108 - 2*0.0444 = .02484$
 $SE(\beta_2 - \beta_3) = (.02484)^{.5} = .1576$ (df = 16)

Slide 66 Echolocating Bats as Reference

NOTES:

Birds vs. Echolocating Bats

Display 10.15

Inference about $\beta_2 - \beta_3$, the coefficient of the indicator variable for birds minus the coefficient of the indicator variable for echolocating bats

D Estimate the linear combination of coefficients on the same linear combination of estimated coefficients

Estimate of $\beta_2 - \beta_3$, from 1.1023 + 0.7977 = 1.9000

Model		Unstandardized Coefficients		Standardized Coefficients		t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta				Lower Bound	Upper Bound
1	(Constant)	-1.576	.287			-5.488	.000	-2.185	-.967
	In (mass)	.815	.045	.998		18.297	.000	.721	.909
	Birds	.102	.114	.041		.896	.384	-.140	.344
	Echolocating bats	.079	.203	.026		.388	.703	-.351	.508

a. Dependent Variable: ln (energy)

Slide 67 Birds vs. Echolocating Bats

NOTES:

Using indicator variables judiciously

- By changing the reference category to echolocating bats, instead of non-echolocating bats, the bird coefficient will test the hypothesis that the bird Y-intercept differs from echolocating bat Y-intercept
 - If non-echolocating bats were the reference, the standard error for the difference in Y intercepts could be calculated using the 'propagation of error variance' formula:
 - $Var(a - b) = variance(a) + variance(b) - 2 cov(a, b)$
- By coding the animal type using non-integer coding, ebats as $\frac{1}{2}$ and nebats as $-\frac{1}{2}$, "Birds vs. Both types of Bats" can be tested. Draper & Smith cover non-integer coding schemes.

Slide 68 Using indicator variables judiciously

NOTES:

Birds vs. Both types of bats

Weighted average of the 2 bat types using $-\frac{1}{2}$ and $\frac{1}{2}$ for the indicator variables for ebats and bats

Model		Unstandardized Coefficients		Standardized Coefficients		t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta				Lower Bound	Upper Bound
1	(Constant)	-1.537	.205			-7.481	.000	-1.973	-1.101
	In (mass)	.815	.045	.998		18.297	.000	.721	.909
	Birds	.063	.093	.025		.676	.509	-.134	.260
	Bats (echo) v.								
	Bats (non-echo)	.079	.203	.020		.388	.703	-.351	.508

a. Dependent Variable: ln (energy)

Statistical summary: After accounting for body mass effects, little evidence that the energy consumption by echolocating bats differs from non-echolocating bats (t-test, $p=0.7$), nor is there much evidence that birds differ from bats in general in energy expenditure (t test, $p=0.51$)

Slide 69 Birds vs. Both types of bats

NOTES: