Ch 9: Multiple regression
Ch 10 Inferential tools for multiple regression

Class 16: 4/8/09 W

Slide 1 Ch 9: Multiple regression

Ch 10 Inferential tools for multiple regression

NOTES:

HW 10 due Friday 4/10/09

Submit as Myname-HW10.doc (or *.rtf)

- Computation Problem 10, chapter 8
  - 8.16 Meat processing, Must assess lack of fit!
  - Due Friday 4/10 Noon
- Read Chapter 9 on multiple regression
  - Read chapter 9 conceptual problems & solutions
  - Post a question or response about Chapter 9 conceptual problems
- HW 11 9.19: Speed of evolution,
  - Due Monday 4/13/09 10 am
  - This is a TOUGH problem! Weds: ask for help/hints!
- Read Chapter 10
  - 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:

Chapter 9

Multiple regression

EEOS611

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

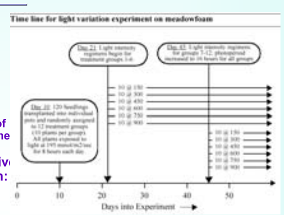
Slide 5 Multivariate vs. Univariate

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**



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

Display 9.2

Numbers of *flowers* per meadowfoam plant, in twelve treatment groups

timing		*intensity* (µmol/m²/sec) 150	300	450	600	750	900
	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

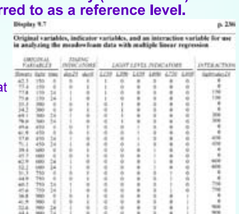
Slide 7 Display 9.2: Two-factor layout

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 conveted 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

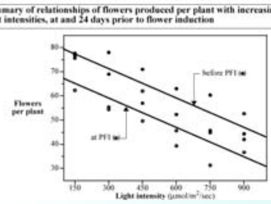
NOTES:

Case 9.1 Statistical summary

- **Increasing light intensity decreased the mean numl of flowers per plant by 4 (: plants per 100 µEm^{-2} s^{-1}.**
- **Beginning light treatment days before PFI increased mean number of flowers by 12.2 (± 5.5) (± ½ 95% CI)**

Display 9.3

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



Slide 9 Case 9.1 Statistical summary

NOTES:

Case 9.2: Brain weight

Covariation of Brain weight and demographic traits

- Brain weight and other length measurements are scaled allometrically
  - $Y = a\ Weight^{B}$
  - Log-log transforms are the rule for allometric data: $\log(Y) = \log(a) + ß \times \log(W)$
- Is there an association between brain weight (response) and litter size, after controlling for the effect of body weight?
- Is there an association between brain weight (response) and gestation length, after controlling for body weight?

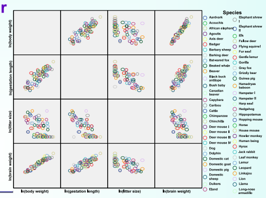
Slide 10 Case 9.2: Brain weight

NOTES:

Case 9.2 Statistical summary

A form of analysis of covariance

- There was convincing evidence that brain weight was associated with either gestation length or litter size after accounting for the effect of body weight (p<0.0001; extra sum of squares F test).
- There was strong evidence that
  - litter size was associated with brain weight after accounting for body weight and gestation (2-sided p value 0.0089) and,
  - Gestation period associated with brain weight after accounting for body weight and litter size (2-sided p value = 0.0038)



EEOS611

Slide 11 Case 9.2 Statistical summary

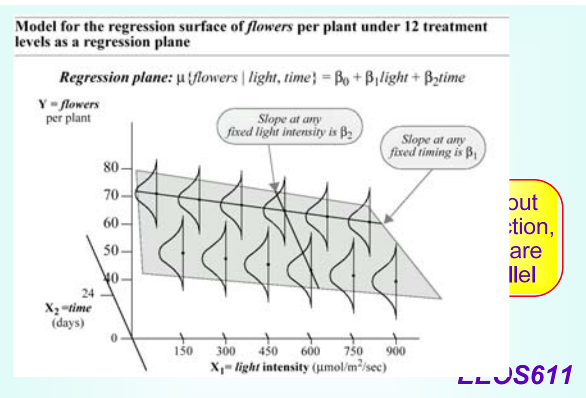
NOTES:

Case 9.1, Theory and calculations

EEOS611

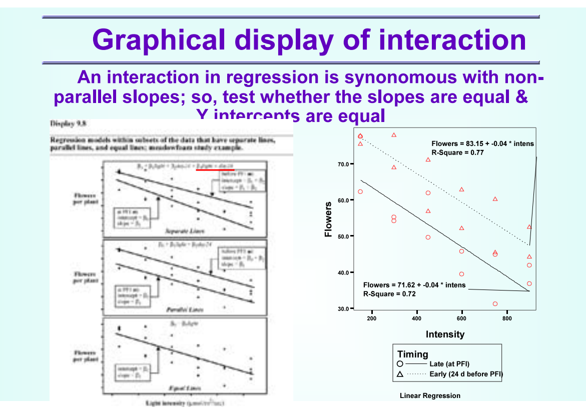
Slide 12 Case 9.1, Theory and calculations

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

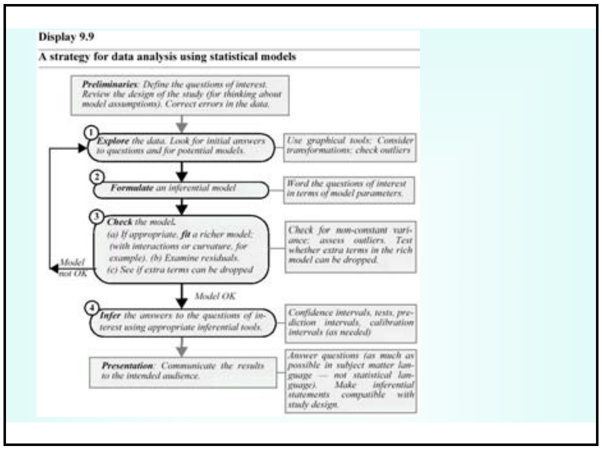
NOTES:

Case 9.2 Allometry of brain size

EEOS611

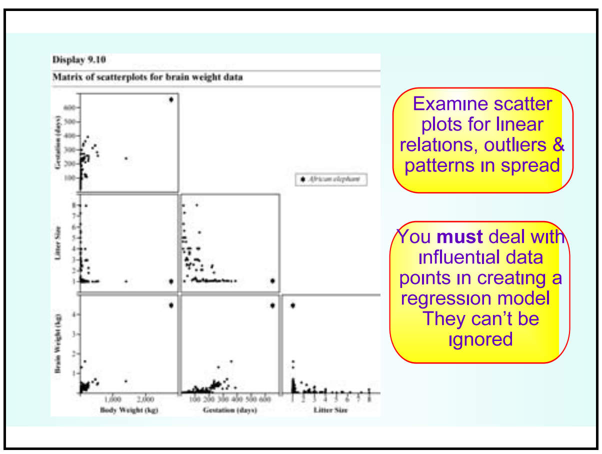
Slide 16 Case 9.2 Allometry of brain size

NOTES:



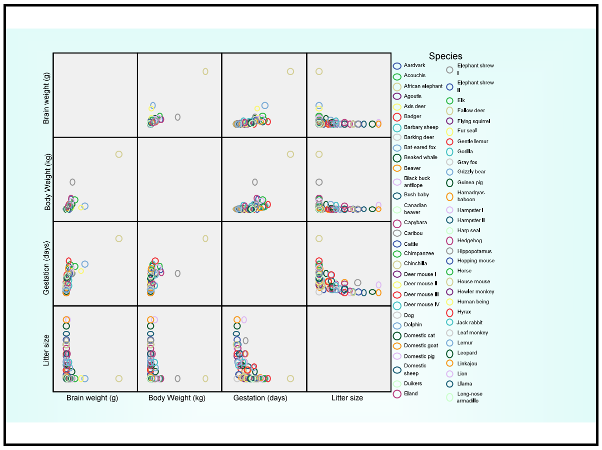
Slide 17 Display 9.9

NOTES:



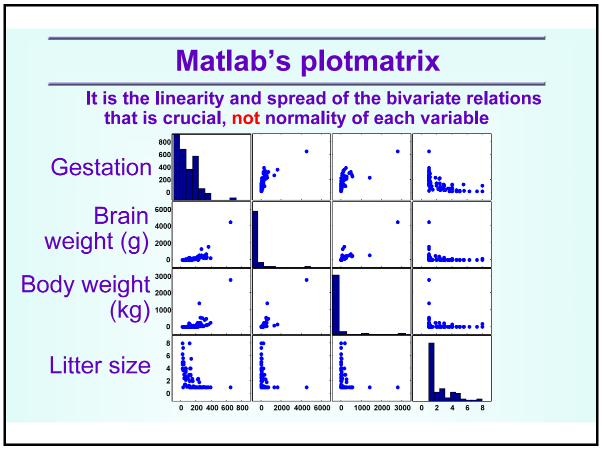
Slide 18 Display 9.10

NOTES:



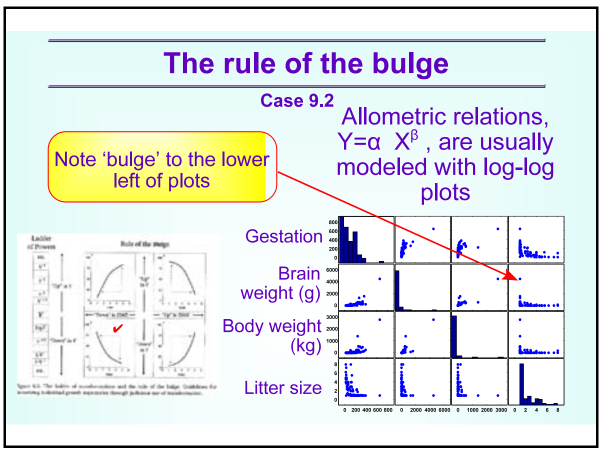
Slide 19

NOTES:



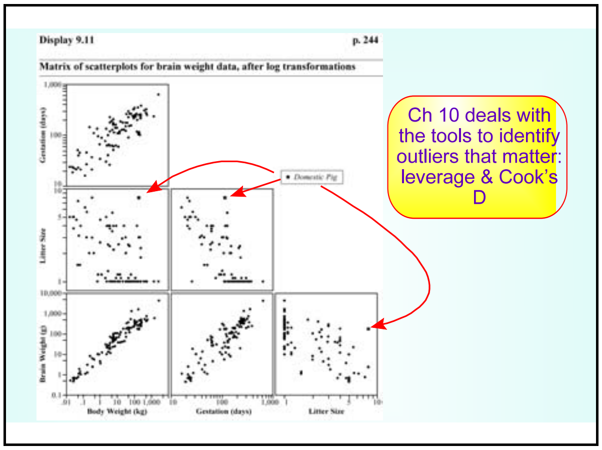
Slide 20 Matlab's plotmatrix

NOTES:



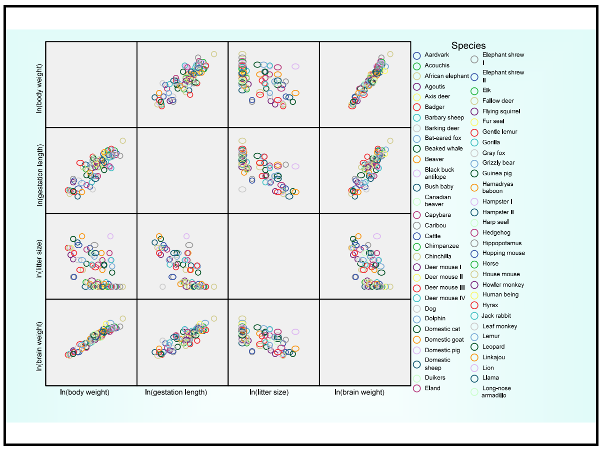
Slide 21 The rule of the bulge

NOTES:



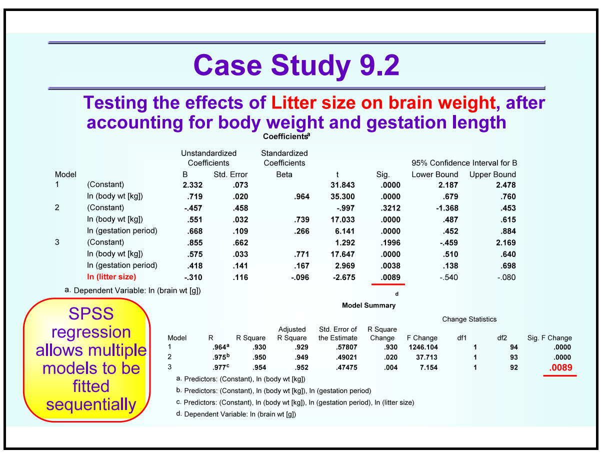
Slide 22 Display 9.11

NOTES:



Slide 23

NOTES:



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 lnbrain
/METHOD=ENTER lnbody
/METHOD=ENTER lngest lnbody
/METHOD=ENTER lngest lnbody lnlitt
/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

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	95% Confidence Interval for B Upper Bound
1	(Constant)	2.332	.073		31.843	3.8E-052	2.187	2.478
	ln(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
	ln(body weight)	.652	.021	.874	31.339	3.2E-051	.610	.693
	ln(litter size)	-.538	.090	-.166	-5.963	4.4E-008	-.718	-.359
3	(Constant)	.855	.662		1.292	.200	-.459	2.169
	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	.138	.698

a. Dependent Variable: ln(brain weight)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics R Square Change	F Change	df1	df2	Sig. F Change
1	.964[a]	.930	.929	.57807	.930	1246.104	1	94	.0000
2	.974[b]	.949	.948	.49430	.019	35.562	1	93	.0000
3	.977[c]	.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])

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

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

ANOVA[d]

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	425.089	2	212.545	869.915	.000[b]
	Residual	22.723	93	.244		
	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), 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)

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

EEOS611

Slide 27 Extra Sum of Squares F test

NOTES:

Comparing Full & Reduced models

Equivalence of *t* tests and F tests for 1 parameter

ANOVA[d]

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	425.089	2	212.545	869.915	.000[b]
	Residual	22.723	93	.244		
	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), 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)

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

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients 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	.96	35.3	.000	.68	.76
2	(Constant)	2.80	.10		27.9	.000	2.60	3.00
	ln(body weight)	.65	.02	.87	31.3	.000	.61	.69
	ln(litter size)	-.54	.09	-.17	-6.0	.000	-.72	-.36
3	(Constant)	.85	.66		1.3	.200	-.46	2.17
	ln(body weight)	.58	.03	.77	17.6	.000	.51	.64
	ln(litter size)	-.31	.12	-.10	-2.7	.009	-.54	-.08
	ln(gestation length)	.42	.14	.17	2.97	.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 B	Std. Error	Standardized Coefficients 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
	ln(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
	ln(body weight)	.652	.021	.874	31.339	3.2E-051	.610	.693
	ln(litter size)	-.538	.090	-.166	-5.963	4.4E-008	-.718	-.359
3	(Constant)	.855	.662		1.292	.200	-.459	2.169
	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	.138	.698

a. Dependent Variable: ln(brain weight)

t-test

Model Summary[d]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change
1	.964[a]	.930	.929	.57807	.930	1246.104	1	94	.0000
2	.974[b]	.949	.948	.49430	.019	35.562	1	93	.0000
3	.977[c]	.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])

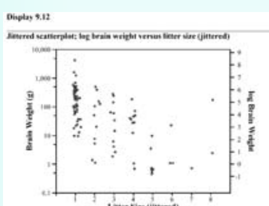
Extra Sum of Squares F test

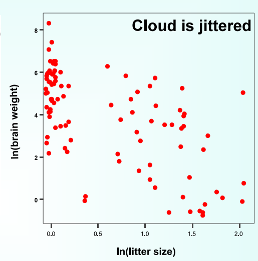
Slide 29 Case Study 9.2

NOTES:

Display 9.12: Jittered plots

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

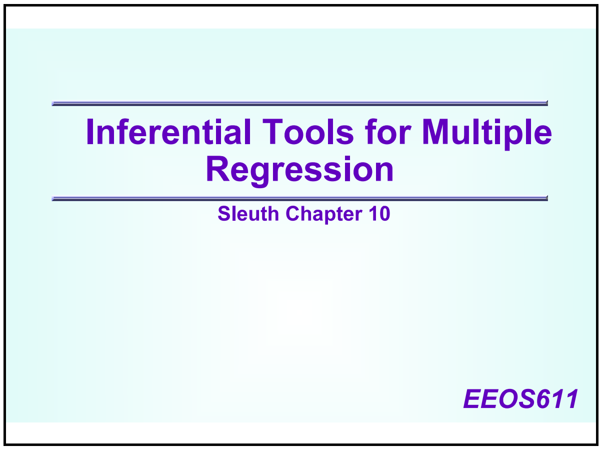




EEOS611

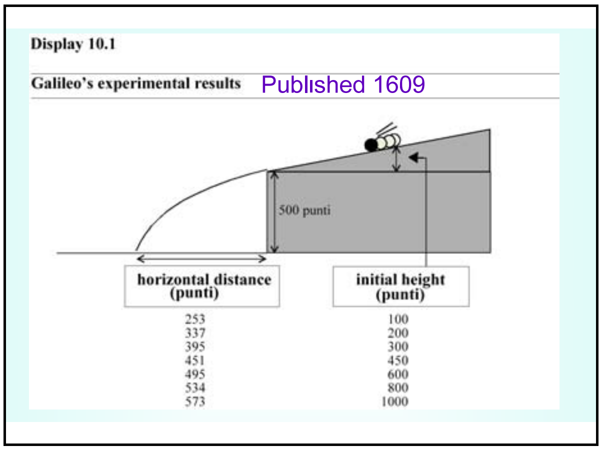
Slide 30 Display 9.12: Jittered plots

NOTES:



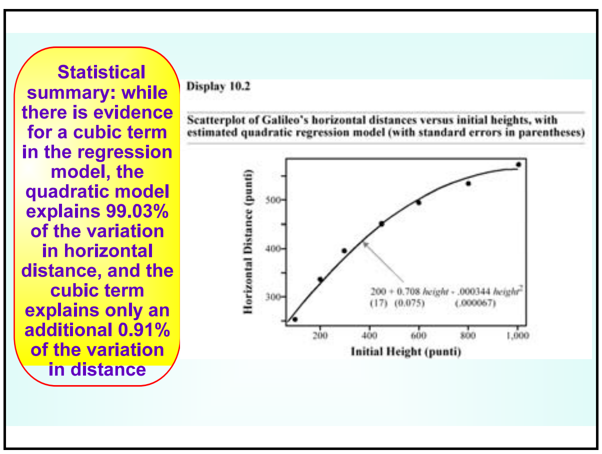
Slide 31 Inferential Tools for Multiple Regression

NOTES:



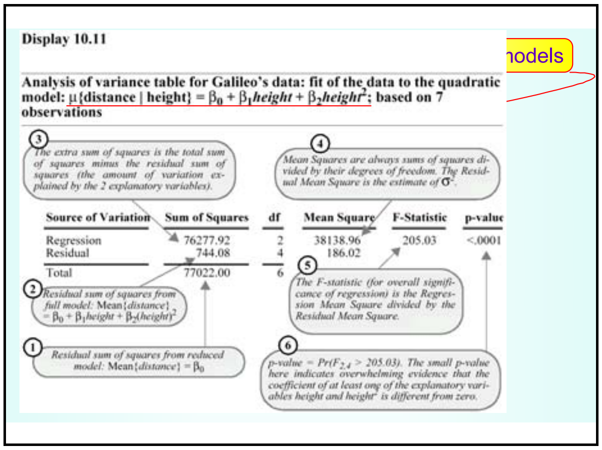
Slide 32 Display 10.1

NOTES:



Slide 33 Display 10.2

NOTES:



Display 10.11

odels

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

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			

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

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

(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).

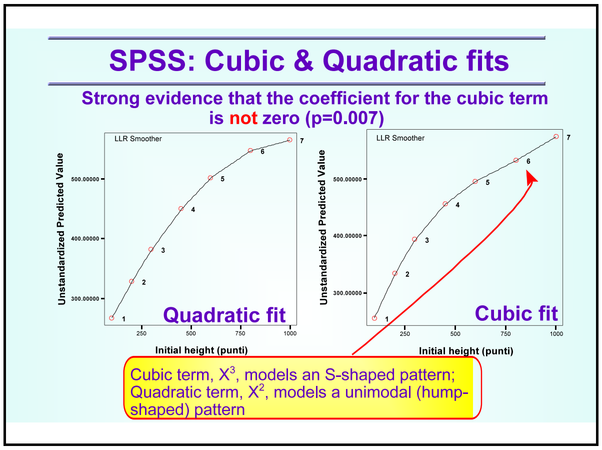
(4) Mean Squares are always sums of squares divided by their degrees of freedom. The Residual Mean Square is the estimate of σ^2.

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

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

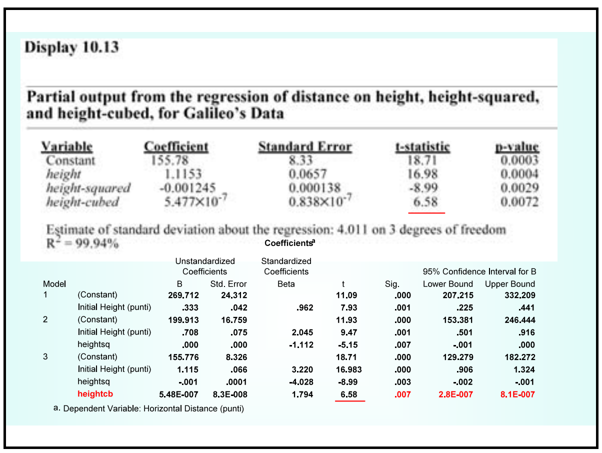
Slide 34 Display 10.11

NOTES:



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	5.477×10^{-7}	0.838×10^{-7}	6.58	0.0072

Estimate of standard deviation about the regression: 4.011 on 3 degrees of freedom
$R^2 = 99.94\%$

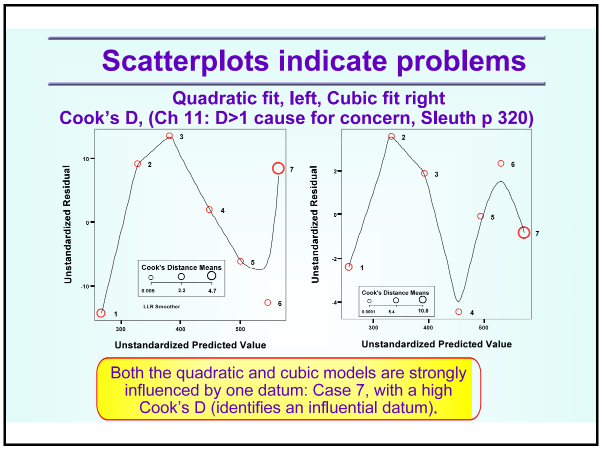
Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	95% Confidence Interval for B Upper Bound
1	(Constant)	269.712	24.312		11.09	.000	207.215	332.209
	Initial Height (punti)	.333	.042	.962	7.93	.001	.225	.441
2	(Constant)	199.913	16.759		11.93	.000	153.381	246.444
	Initial Height (punti)	.708	.075	2.045	9.47	.001	.501	.916
	heightsq	.000	.000	-1.112	-5.15	.007	-.001	.000
3	(Constant)	155.776	8.326		18.71	.000	129.279	182.272
	Initial Height (punti)	1.115	.066	3.220	16.983	.000	.906	1.324
	heightsq	-.001	.0001	-4.028	-8.99	.003	-.002	-.001
	heightcb	5.48E-007	8.3E-008	1.794	6.58	.007	2.8E-007	8.1E-007

a. Dependent Variable: Horizontal Distance (punti)

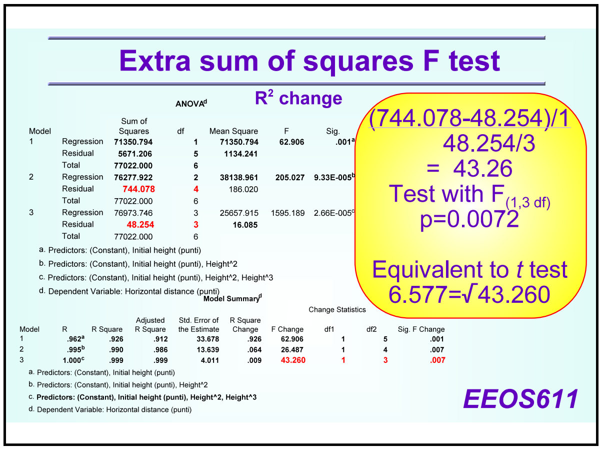
Slide 36 Display 10.13

NOTES:



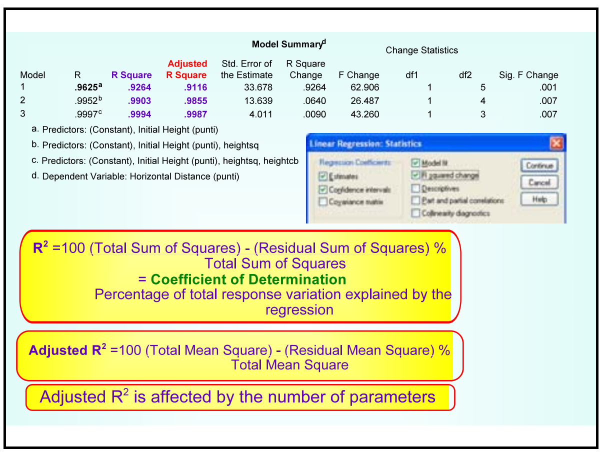
Slide 37 Scatterplots indicate problems

NOTES:



Slide 38 Extra sum of squares F test

NOTES:



Slide 39

NOTES:

R^2 =100 (Total Sum of Squares) - (Residual Sum of Squares) % / Total Sum of Squares

Adjusted R^2 =100 (Total Mean Square) - (Residual Mean Square) % / 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	76277.9	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			

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

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

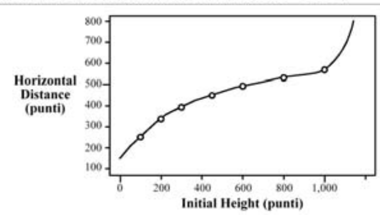
Adjusted R^2 introduces a penalty for number of parameters

Slide 40

NOTES:

Display 10.14

Scatterplot of Galileo's horizontal distances versus initial heights, with estimated sixth-order polynomial regression curve (R^2 = 100%)



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

Slide 41 Display 10.14

NOTES:

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

$\hat{\mu}\{distance \mid height = 250\}$ $SE(\hat{\mu}\{distance \mid height = 250\})$

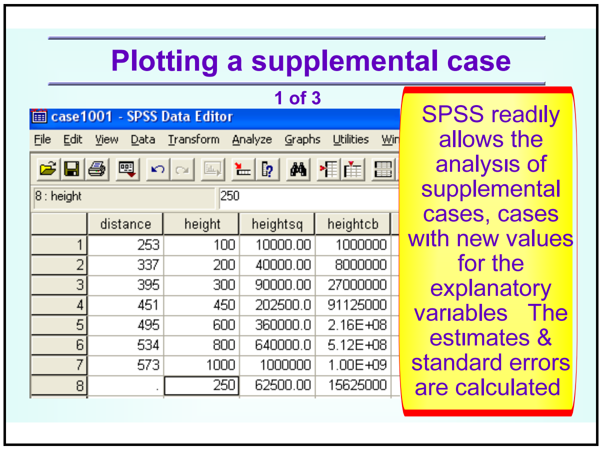
Reference height = 250

Variable	Coefficient	Standard Error	t-Statistic	2-Sided p-Value
CONSTANT	355.51	6.62	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

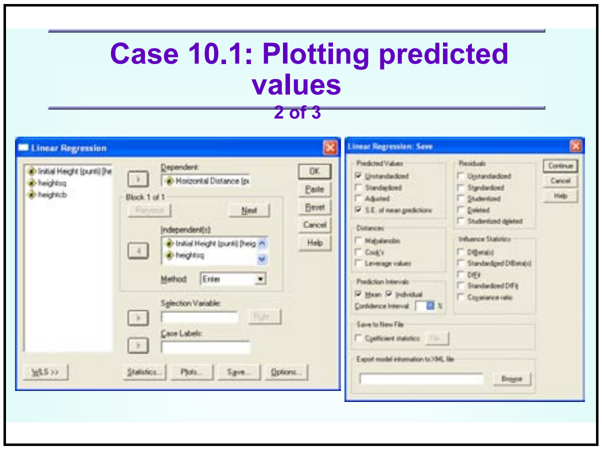
Slide 42 Estimating the predicted value and standard error for 250 punti

NOTES:



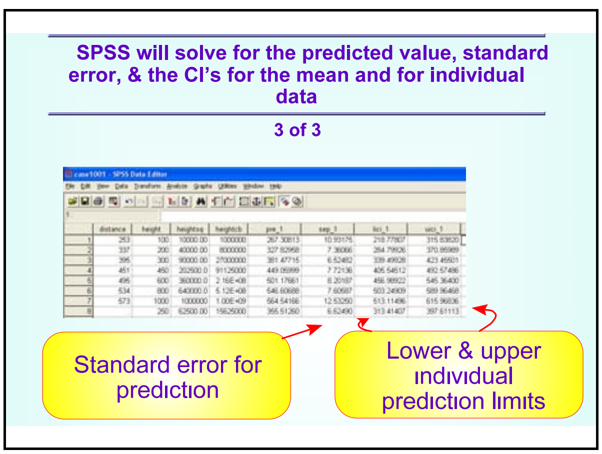
Slide 43 Plotting a supplemental case

NOTES:



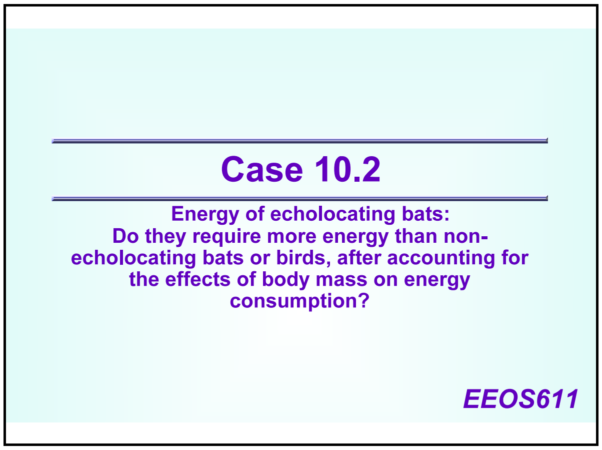
Slide 44 Case 10.1: Plotting predicted values

NOTES:



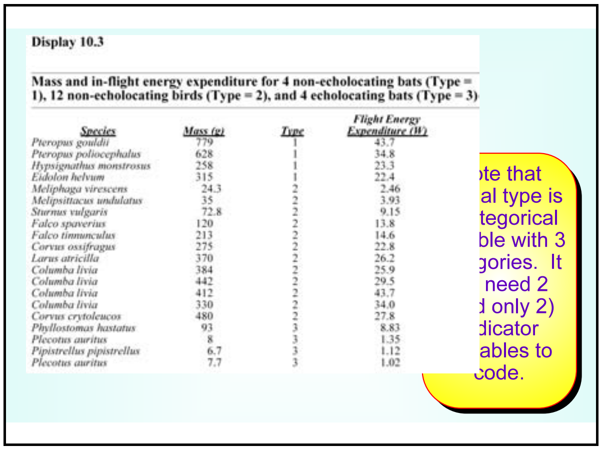
Slide 45 SPSS will solve for the predicted value, standard error, & the CI's for the mean and for individual data

NOTES:



Slide 46 Case 10.2

NOTES:



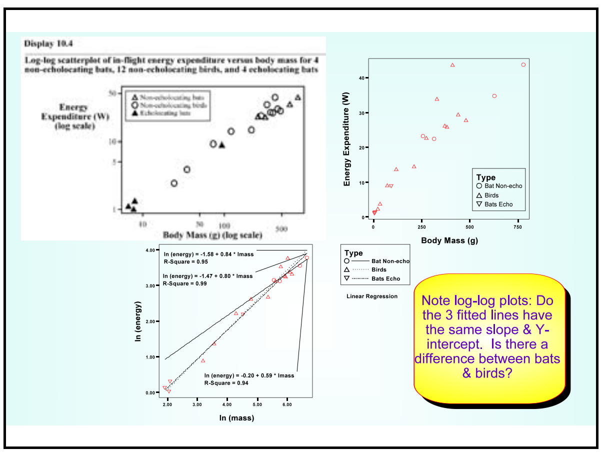
Display 10.3

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)*
Pteropus gouldii	779	1	43.7
Pteropus poliocephalus	628	1	34.8
Hypsignathus monstrosus	258	1	23.3
Eidolon helvum	315	1	22.4
Meliphaga virescens	24.3	2	2.46
Melipsittacus undulatus	35	2	3.93
Sturnus vulgaris	72.8	2	9.15
Falco spaverius	120	2	13.8
Falco tinnunculus	213	2	14.6
Corvus ossifragus	275	2	22.8
Larus atricilla	370	2	26.2
Columba livia	384	2	25.9
Columba livia	442	2	29.5
Columba livia	412	2	43.7
Columba livia	330	2	34.0
Corvus crytoleucos	480	2	27.8
Phyllostomas hastatus	93	3	8.83
Plecotus auritus	8	3	1.35
Pipistrellus pipistrellus	6.7	3	1.12
Plecotus auritus	7.7	3	1.02

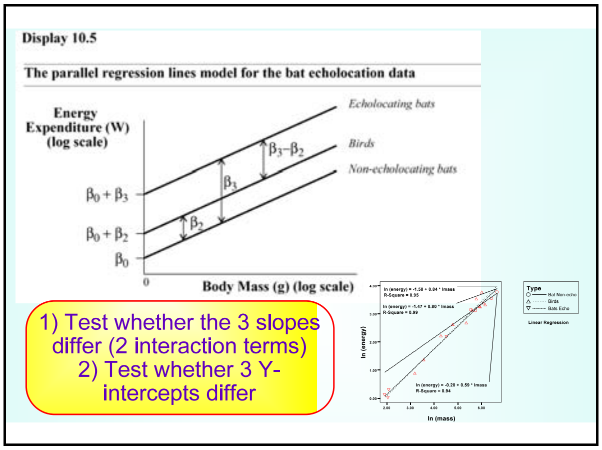
Slide 47 Display 10.3

NOTES:



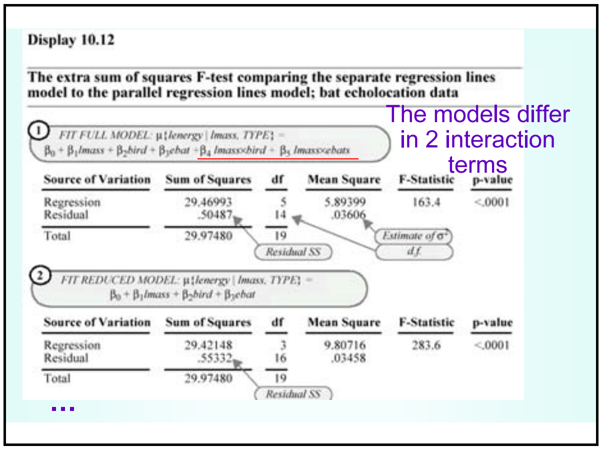
Slide 48 Display 10.4

NOTES:



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: μ{lenergy | lmass, TYPE} = $\beta_0 + \beta_1 lmass + \beta_2 bird + \beta_3 ebat + \beta_4 lmass{\times}bird + \beta_5 lmass{\times}ebat$

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			

Residual SS; Estimate of σ^2; d.f.

② FIT REDUCED MODEL: μ{lenergy | lmass, TYPE} = $\beta_0 + \beta_1 lmass + \beta_2 bird + \beta_3 ebat$

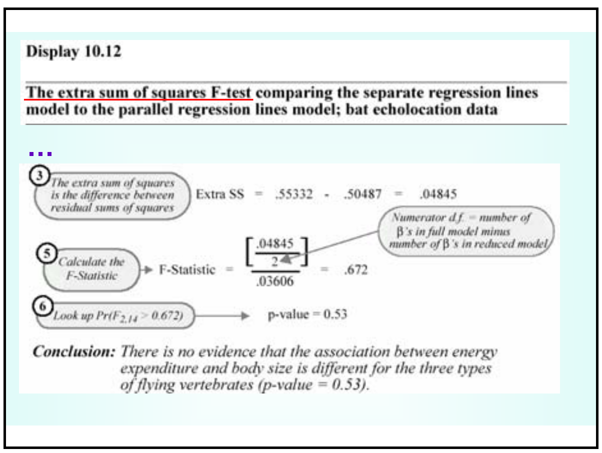
Source of Variation	Sum of Squares	df	Mean Square	F-Statistic	p-value
Regression	29.42148	3	9.80716	283.6	<.0001
Residual	.55332	16	.03458		
Total	29.97480	19			

Residual SS

…

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

Numerator d.f. = number of β's in full model minus number of β's in reduced model

⑤ Calculate the F-Statistic → F-Statistic = $\frac{[.04845 / 2]}{.03606}$ = .672

⑥ Look up $Pr(F_{2,14} > 0.672)$ → 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

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	Upper Bound
1	(Constant)	-1.468	.137		-10.705	3.1E-009	-1.76	-1.18
	ln (mass)	.809	.027	.990	30.127	7.4E-017	.75	.86
2	(Constant)	-1.576	.287		-5.488	5.0E-005	-2.19	-.97
	ln (mass)	.815	.045	.998	18.297	3.8E-012	.72	.91
	Birds	.102	.114	.041	.896	.384	-.14	.34
	Echolocating bats	.079	.203	.026	.388	.703	-.35	.51
3	(Constant)	-.202	1.261		-.161	.875	-2.91	2.50
	ln (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
	Intxn (Mass, Birds v. Bats)	.246	.213	.536	1.151	.269	-.21	.70
	Intxn (Mass,Echolocating bats)	.215	.224	.204	.961	.353	-.26	.69

a. Dependent Variable: ln (energy)

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 Summary[d]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: 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), Intxn (Mass,Twobats), Intxn (Mass, Birds v. Bats)
d. Dependent Variable: ln (Energy)

ANOVA[d]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	29.392	1	29.392	907.638	7.44E-017[a]
	Residual	.583	18	.032		
	Total	29.975	19			
2	Regression	29.421	3	9.807	283.589	4.46E-014[b]
	Residual	.553	16	.035		
	Total	29.975	19			
3	Regression	29.470	5	5.894	163.440	6.70E-012[c]
	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), Intxn (Mass,Twobats), Intxn (Mass, Birds v. Bats)
d. Dependent Variable: ln (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.

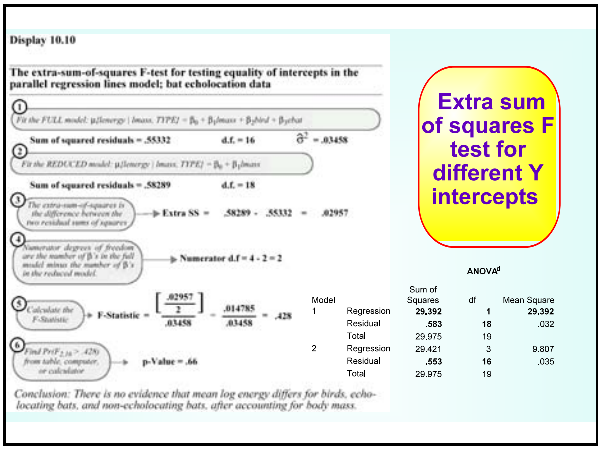
Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	Upper Bound
1	(Constant)	-1.468	.137		-10.705	.000	-1.756	-1.180
	ln (mass)	.809	.027	.990	30.127	.000	.752	.865
2	(Constant)	-1.576	.287		-5.488	.000	-2.185	-.967
	ln (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
3	(Constant)	-.202	1.261		-.161	.875	-2.908	2.503
	ln (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	-4.025	1.489
	Bird x mass	.246	.213	.536	1.151	.269	-.212	.703
	Ebat x mass	.215	.224	.204	.961	.353	-.265	.694

a. Dependent Variable: ln (Energy)

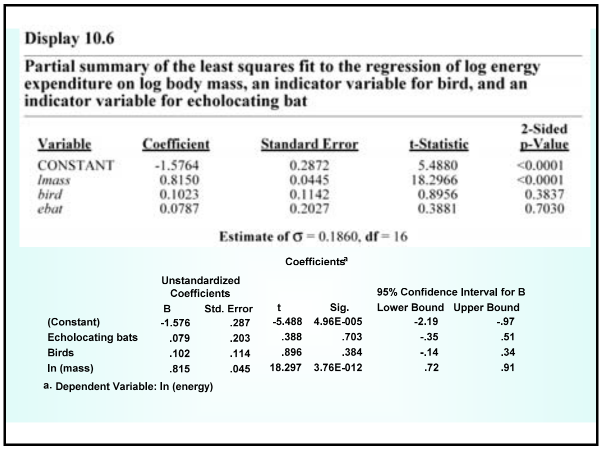
Slide 54 Do echolocating bats differ from non-echolocating bats in energy expenditure?

NOTES:



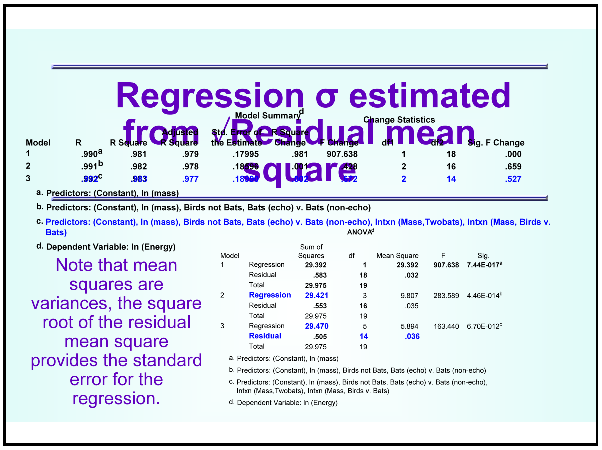
Slide 55 Display 10.10

NOTES:



Slide 56 Display 10.6

NOTES:



Slide 57 Regression σ estimated from √Residual mean square

NOTES:

Slide 58

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 σ = 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?

NOTES:

Slide 59

Model Summary[d]

					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), Intxn (Mass,Twobats), Intxn (Mass, Birds v. Bats)

d. Dependent Variable: ln (Energy)

ANOVA[d]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	29.392	1	29.392	907.638	7.44E-017[a]
	Residual	.583	18	.032		
	Total	29.975	19			
2	Regression	29.421	3	9.807	283.589	4.46E-014[b]
	Residual	.553	16	.035		
	Total	29.975	19			
3	Regression	29.470	5	5.894	163.440	6.70E-012[c]
	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), Intxn (Mass,Twobats), Intxn (Mass, Birds v. Bats)

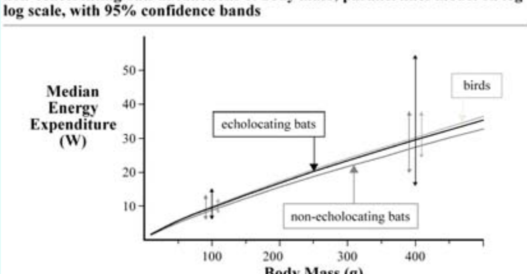
d. Dependent Variable: ln (Energy)

NOTES:

Slide 60 Display 10.8

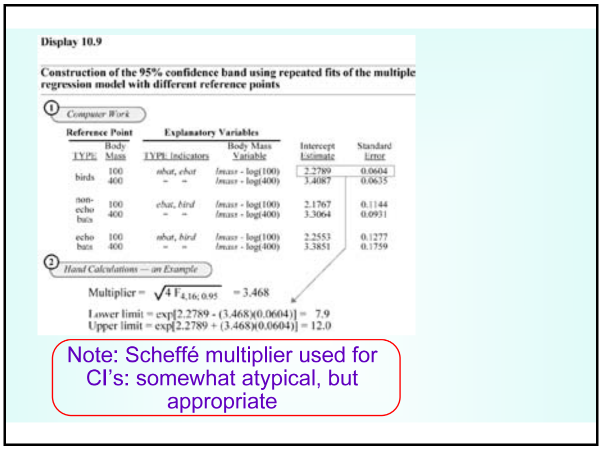
Display 10.8

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



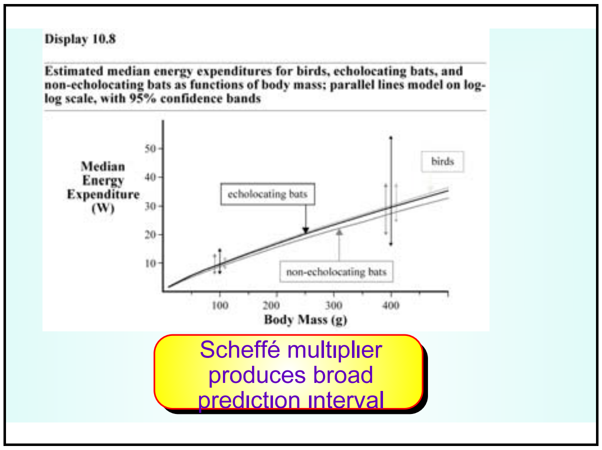
Scheffé multiplier used for Confidence Intervals

NOTES:



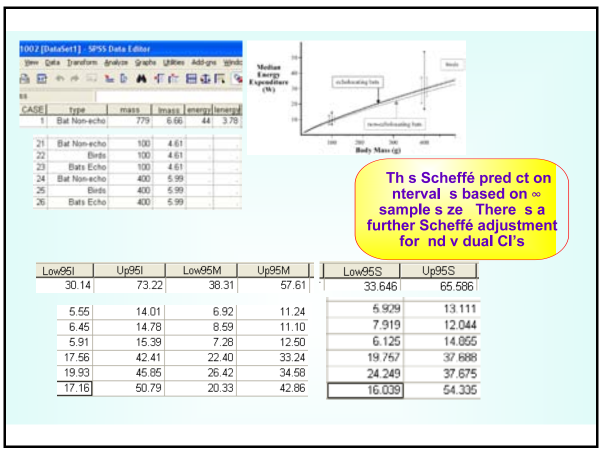
Slide 61 Display 10.9

NOTES:



Slide 62 Display 10.8

NOTES:



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 producded by regression.

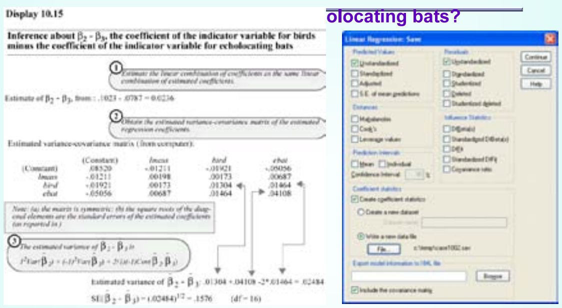
```
COMPUTE SchInt = SEP_1 * Scheffemultiplier
```

Slide 64 Syntax: Scheffé multiplier

NOTES:

Variance formulae for linear contrasts

olocating bats?



Slide 65 Variance formulae for linear contrasts

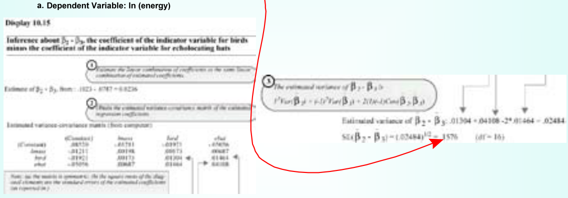
NOTES:

Echolocating Bats as Reference

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	95% Confidence Interval for B Upper Bound
1	(Constant)	-1.498	.1499		-9.993	.000	-1.815	-1.180
	ln (mass)	.815	.0445	.998	18.297	.000	.721	.909
	Non-echolocating bats	-.0787	.2027	-.026	-.388	.703	-.508	.351
	Birds	.0236	.1576	.009	.150	.883	-.310	.358

a. Dependent Variable: ln (energy)



Slide 66 Echolocating Bats as Reference

NOTES:

Birds *vs.* Echolocating Bats

Display 10.15

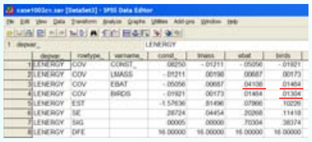
Inference about $\beta_2 - \beta_3$, the coefficient of the indicator variable for birds minus the coefficient of the indicator variable for echolocating bats

(1) *Estimate the linear combination of coefficients as the same linear combination of estimated coefficients.*

Estimate of $\beta_2 - \beta_3$, from : .1023 - .0787 = 0.0236

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	Upper Bound
1	(Constant)	**-1.576**	**.287**		**-5.488**	**.000**	**-2.185**	**-.967**
	ln (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)



	depvar_	rowtype_	varname_	const_	lmass	ebat	birds
1	LENERGY	COV	CONST_	.08250	-.01211	-.05056	-.01921
2	LENERGY	COV	LMASS	-.01211	.00198	.00687	.00173
3	LENERGY	COV	EBAT	-.05056	.00687	.04136	.01464
4	LENERGY	COV	BIRDS	-.01921	.00173	.01464	.01304
5	LENERGY	EST		-1.57636	.81496	.07866	.10226
6	LENERGY	SE		.28724	.04454	.20268	.11418
7	LENERGY	SIG		.00005	.00000	.70304	.38074
8	LENERGY	DFE		16.00000	16.00000	16.00000	16.00000

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 'propogation of error variance' formula:
  - Var(a- b)=variance(a)+variance(b)-2 cov(a,b)
- By coding the animal type using non-intiger coding, ebats as ½ and nebats as -½, "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 -½ and ½ for the indicator variables for ebats and bats

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	Upper Bound
1	(Constant)	-1.537	.205		-7.481	.000	-1.973	-1.101
	ln (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: