Ch 6: Linear combinations and multiple comparisons of means & Ch7 Simple linear regression: a model for the mean

Class 10: 3/9/09 M

EEOS611

Slide 1 Ch 6: Linear combinations and multiple comparisons of means

& Ch7 Simple linear regression: a model for the mean

NOTES:

HW 8 due Thus 3/12/09 11 am

Submit as Myname-HW8.doc (or *.rtf)

- Read Chapter **7 Comparisons among several samples**
- Comment on Chapter 7 conceptual problems in Blackboard Vista4
- Computation Problem 8
  - Problem 6.22 A biological basis for homosexuality
  - You must use linear contrasts to solve the problem
  - You can assume that the contrasts were specified *a priori*

EEOS611

Slide 2 HW 8 due Thus 3/12/09 11 am

NOTES:

HW 9 due Monday 3/16/09 10 am

Submit as Myname-HW9.doc (or *.rtf)

- Read Chapter **8 A closer look at assumptions for simple linear regression**
- Comment on Chapter 8 conceptual problems in Blackboard Vista4
- Computation Problem 9
  - Problem 7.29 (Sleuth 2nd edition, p. 203) Male displays

EEOS611

Slide 3 HW 9 due Monday 3/16/09 10 am

NOTES:

Student Presentations

Starting at 10:50 (8 minutes each)

- Seth Sheldon for HW 3
  - 2.21 Bumpus's data: weights of Bumpus's birds
- Barry Fradkin for HW 4.
  - **3.28 Pollen removal**

EEOS611

Slide 4 Student Presentations

NOTES:

Ch 6: Linear combinations and multiple comparisons of means

EEOS611

Slide 5 Ch 6: Linear combinations and multiple comparisons of means

NOTES:

Case Study 6.1.1

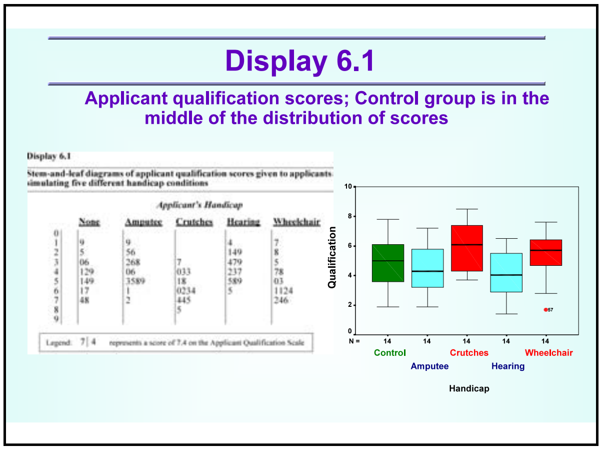
Discrimination against the handicapped

- U.S. Vocational Rehabilitation Act of 1973
- 5 Videotaped job inteviews
  - Applicant appeared with different handicaps
  - Wheelchair
  - Crutches
  - Hearing impaired
  - Amputated
  - No handicap
- 70 undergraduates randomly assigned to view tapes, 14 to each tape.
- Rated on a 1 to 10 applicant qualification scale

EEOS611

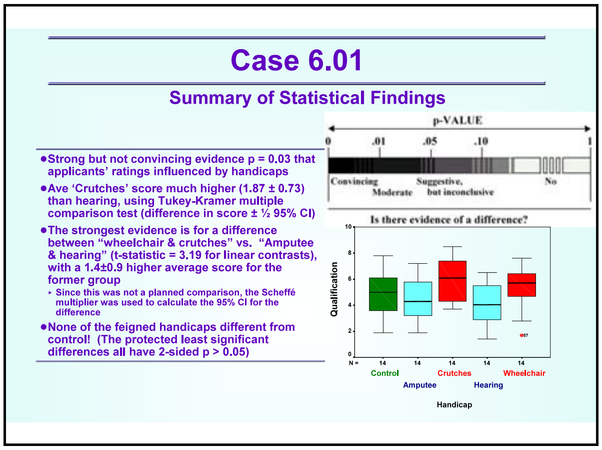
Slide 6 Case Study 6.1.1

NOTES:



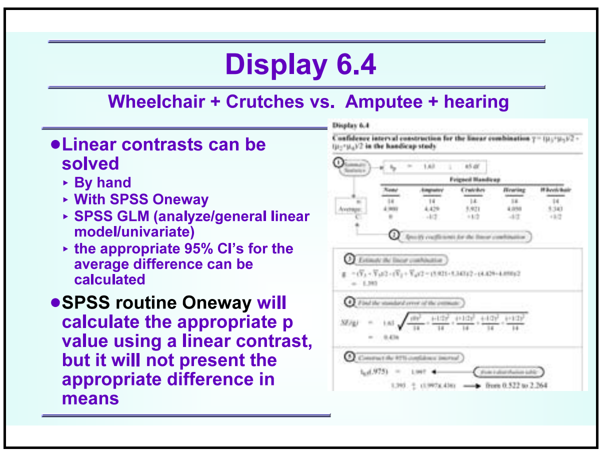
Slide 7 Display 6.1

NOTES:



Slide 8 Case 6.01

NOTES:



Slide 9 Display 6.4

NOTES:

SPSS syntax for Linear contrasts

Display 6.4, p. 149 (1st ed), p. 155 (2nd ed)
If planned, report results as 1.4 ± 0.9
If unplanned report as 1.393 ± 1.382 (Sleuth p. 164)

- * DATA order Control Amputee Crutches Hearing Wheelchair.
- ONEWAY
- score BY code
- /CONTRAST = 0 -1 1 -1 1.
- * This call to GLM does it all.
- UNIANOVA
- score BY code
- /METHOD = SSTYPE(3)
- /INTERCEPT = INCLUDE
- /LMATRIX = "Avg A H vs Avg C W" code 0 -1/2 1/2 -1/2 1/2
- /POSTHOC = code (TUKEY SCHEFFE LSD BONFERRONI)
- /CRITERIA = ALPHA(.05)
- /DESIGN = code .

Contrast Results (K Matrix)[a]

		Dependent Variable Qualification
Contrast Estimate		1.4
Hypothesized Value		0
Difference (Estimate - Hypothesized)		1.4
Std. Error		.44
Sig.		.002
95% Confidence Interval for Difference	Lower Bound	.5
	Upper Bound	2.3

a. Based on the user-specified contrast coefficients (L') matrix: Avg A H vs Avg C W

5 Construct the 95% confidence interval
$t_{65}(.975) = 1.997$ ← from t-distribution table
$1.393 \pm (1.997)(.436)$ → from 0.522 to 2.264

EEOS611

Slide 10 SPSS syntax for Linear contrasts

NOTES:

---

5 Construct the 95% confidence interval
$t_{65}(.975) = 1.997$ ← from t-distribution table
$1.393 \pm (1.997)(.436)$ → from 0.522 to 2.264

Scheffé: 1.393 ± 3.1705 * 0.436 → from 0.011 to 2.775
p=0.0476

ANOVA

Qualification

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	30.5	4	7.6	2.9	.030
Within Groups	173.3	65	2.7		
Total	203.8	69			

Contrast Coefficients

	Handicap				
Contrast	Control	Amputee	Crutches	Hearing	Wheelchair
1	0	-1	1	-1	1

SPSS Oneway doesn't allow fractional contrast coefficients; the estimates are 2x too large, but the p values are ok

Contrast Tests

Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
2.79	.87	3.19	65	.002

Slide 11

NOTES:

Case 6.1.1

Scope of inference, Questions

- Scope of inference
  - Differences exist, but the situation is complicated by having the control having an average in the middle of the group of 5 treatments
  - How should one compare groups?
- Questions:
  - How does one perform linear contrasts in SPSS?
    - Use Oneway with contrasts
    - Use UNIANOVA (GLM) with /Lmatrix
      - The p values and CI's assume planned or *a priori* contrasts
  - What is the Tukey-Kramer procedure?
  - What is "the protected least significant difference"?
  - When should the Bonferonni & Scheffé procedures be used?

Slide 12 Case 6.1.1

NOTES:

Case 6.1.2

Preexisting preferences of fish – a randomized experiment

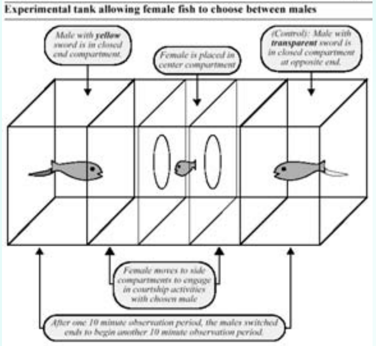
- Sexual selection by females
  - A. L. Basolo
- Southern platyfish: males don't produce the brightly colored sword tail
- Experiment
  - 6 pairs of males surgically given artificial plastic sword tails.
    - 1 individual of each pair received a yellow sword
    - the other a transparent sword.
  - Female fish placed in a compartment
  - Amount of 20 minute periods spent courting with the yellow-sword male recorded.

Slide 13 Case 6.1.2

NOTES:

Display 6.2 p. 145

Experimental tank allowing female fish to choose between males



EEOS611

Slide 14 Display 6.2

NOTES:

Display 6.3

Percent of courtship time spent by 84 females with the yellow-sword male; body sizes of the males are shown in parentheses

	Pair 1 (35 mm)	Pair 2 (31 mm)	Pair 3 (33 mm)	Pair 4 (34 mm)	Pair 5 (28 mm)	Pair 6 (34 mm)
	43.7	52.5	91.0	72.2	78.3	33.4
	54.0	65.6	62.0	58.5	66.0	42.2
	49.8	68.5	10.0	51.0	47.7	35.6
	65.5	45.9	83.8	56.8	77.5	79.9
	53.1	80.2	91.3	92.4	58.3	59.0
	53.0	67.0	56.3	55.3	61.1	58.1
	62.3	73.0	83.6	59.3	65.1	64.2
	49.4	71.7	53.3	42.0	62.9	82.8
	45.7	55.0	36.5	68.5	61.0	75.7
	56.6	70.0	65.4	78.4		66.3
	59.0	63.2	48.1	69.6		56.3
	67.8	39.6	50.6	89.2		84.5
	73.3	41.0	40.4	67.3		61.1
	43.8	59.2	90.6	77.5		87.6
	67.4		74.9			
	58.1		56.0			
			67.5			
Average:	56.41	60.89	62.43	67.00	64.21	63.34
SD:	9.02	12.48	22.29	14.33	9.41	17.68
n:	16	14	17	14	9	14

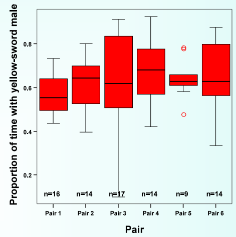
Slide 15 Display 6.3

NOTES:

Sexual preference

Case Study 6.2

- Test for preference for yellow-sword male (expected proportion = ½)
- Test for differences among pairs
- Test for the covariate of male fish weight using a linear contrast



Slide 16 Sexual preference

NOTES:

6.1.2

SUMMARY OF FINDINGS

- No evidence that the mean percentage of time with the yellow-sword male differed from one male pair to another [$P(F_{5,78} \geq 0.79)) = 0.56$]
- No evidence for linear relationship with male body size, from a linear contrast
  - Contrast available with one-way or general linear model
- Mean proportion (± 99.9% CI) with yellow sword is 62.4 (± 5.9) %
- This study provide convincing evidence that the mean percentage of time with the yellow tail exceeds the lack of preference value (50%)

EEOS611

Slide 17 6.1.2

NOTES:

Test for any difference among pairs

Page 158, Sleuth 2nd edition

Display 6.5

Analysis of the pre-existing preference example: F-test for differences in mean percent of time with yellow-tailed male and t-test for linear effect of male body size

ANOVA F-test

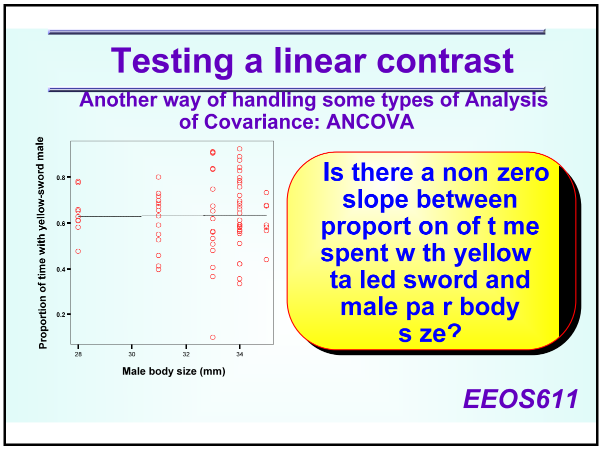
Source of Variation	Sum of Squares	df	Mean Square	F-Statistic	p-value
Between Male Groups	938.75	5	187.75	0.786	0.56
Within Groups	18,636.68	78	238.93		
Total	19,575.43	83			

Conclusion: *There is no evidence that the group means are different for different pairs of males (p-value = 0.56, from ANOVA F-statistic).*

EEOS611

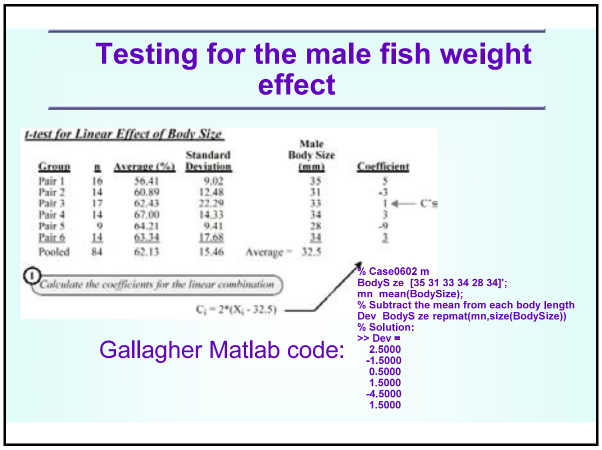
Slide 18 Test for any difference among pairs

NOTES:



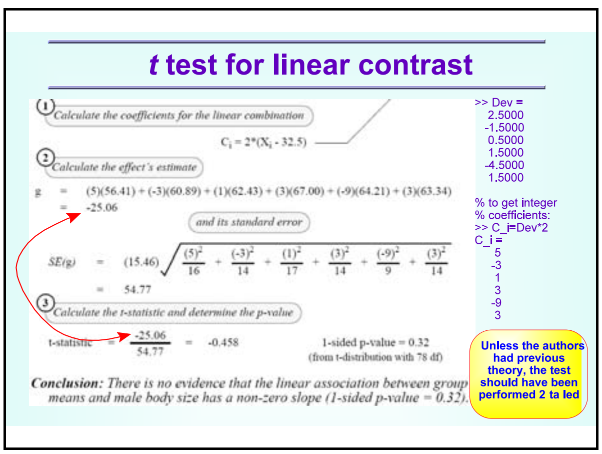
Slide 19 Testing a linear contrast

NOTES:



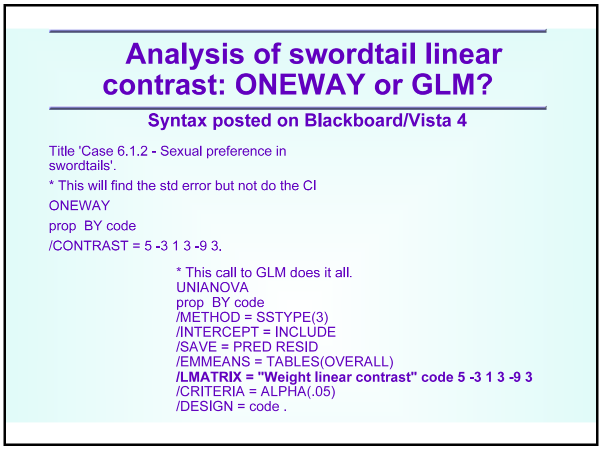
Slide 20 Testing for the male fish weight effect

NOTES:



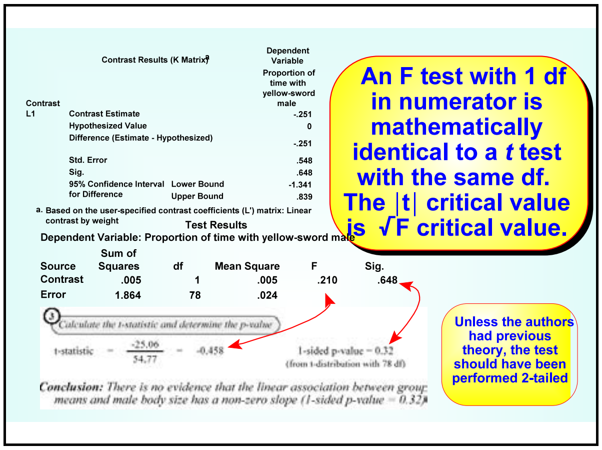
Slide 21 t test for linear contrast

NOTES:



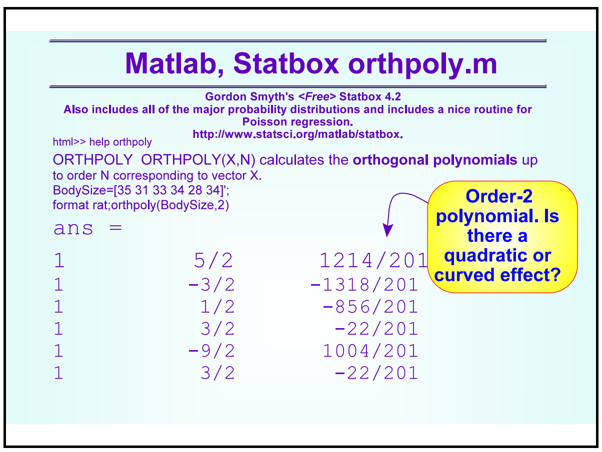
Slide 22 Analysis of swordtail linear contrast: ONEWAY or GLM?

NOTES:



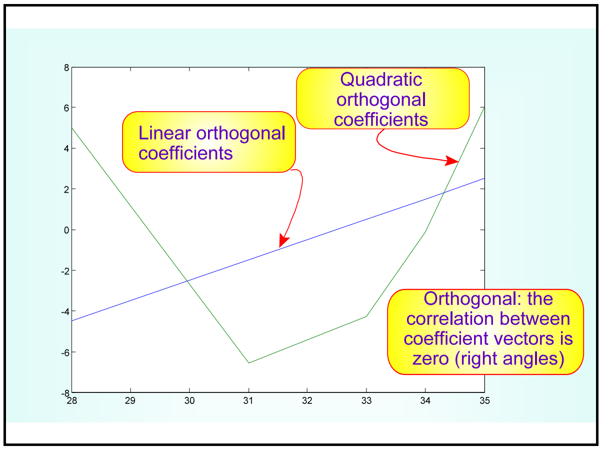
Slide 23

NOTES:



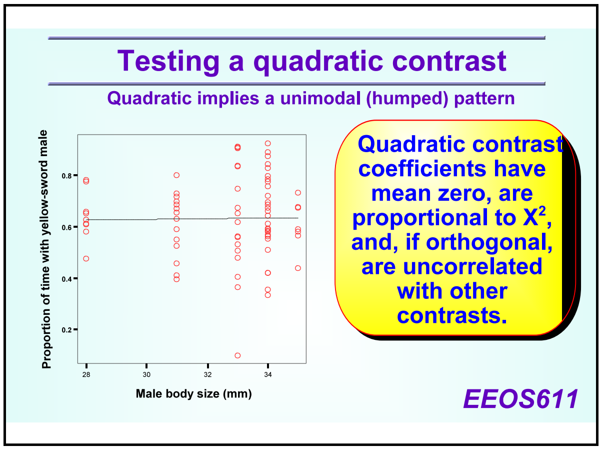
Slide 24 Matlab, Statbox orthpoly.m

NOTES:



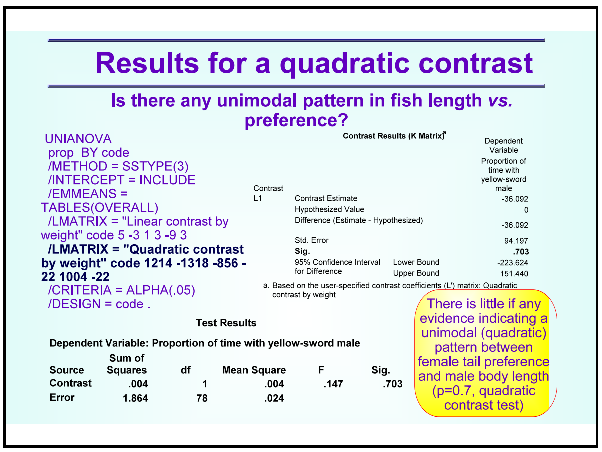
Slide 25

NOTES:



Slide 26 Testing a quadratic contrast

NOTES:



Slide 27 Results for a quadratic contrast

NOTES:

When will environmental scientists need to consider linear contrasts?

Regression lack of fit: a huge, but largely unrecognized, problem with environmental regression analyses

EEOS611

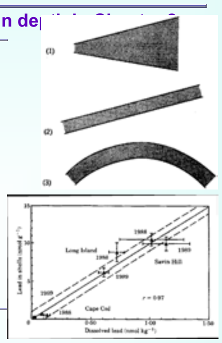
Slide 28 When will environmental scientists need to consider linear contrasts?

NOTES:

Testing for Lack of Fit

This topic will be covered in de

- 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
    - Return to regression 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 29 Testing for Lack of Fit

NOTES:

What to do if there is lack of fit!

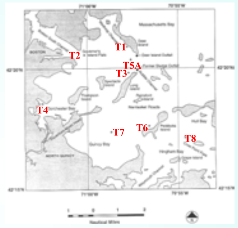
- You may still estimate the slope & Y intercept using regression: OLS regression still provides unbiased estimators
- You can **NOT** use the variance estimates and *p* values based on the error mean square from the OLS linear regression
- Fit a richer or different model
  - Consider testing higher order interaction terms: quadratic & cubic, if warranted
  - Add other explanatory variables
- You **may** analyze the data as an ANOVA model with linear contrasts.
  - Linear contrasts allows tests for linear trend, quadratic trend (hump shaped), cubic trends (S-shaped) and higher order polynomials
  - The variance estimate doesn't assume equal spreads around the regression line, just equal spreads around means

Slide 30 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^2 Ted Young modified Van Veen graps
- Species richness measured with Fisher's α

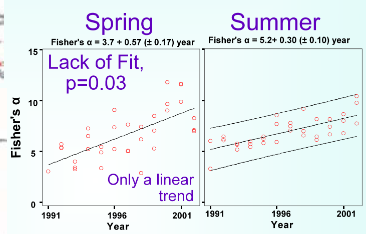


Slide 31 Lack of Fit & Boston Harbor soft-bottom benthic diversity

NOTES:

T1: Deer Island Flats

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



Amy Banik ECOS M.Sc.

Slide 32 T1: Deer Island Flats

NOTES:

Coefficients from Winer et al. (1991, Table D.10), also shown in Draper & Smith (1998, p. 467). The tests are performed hierarchically, with the linear term tested first and the lack-of-fit remaining after the linear term is removed. The quadratic and cubic terms are tested only if there is a significant lack of fit after the inclusion of the lower order polynomial contrast in the model. These orthogonal contrasts are calculated automatically by the SPSS procedure general linear ANOVA model.

a) Summer data

	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	-42	14	35	31	12	-12	-31	-35	-14	42

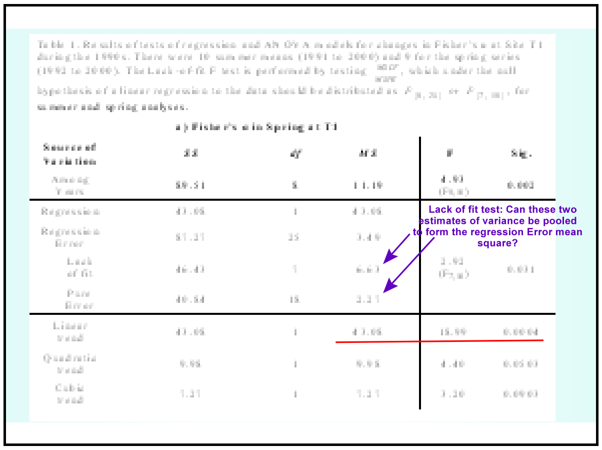
a) Spring data

	1992	1993	1994	1995	1996	1997	1998	1999	2000
Linear	-4	-3	-2	-1	0	1	2	3	4
Quadratic	28	7	-8	-17	-20	-17	-8	7	28
Cubic	-14	7	13	9	0	-9	-13	-7	14

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

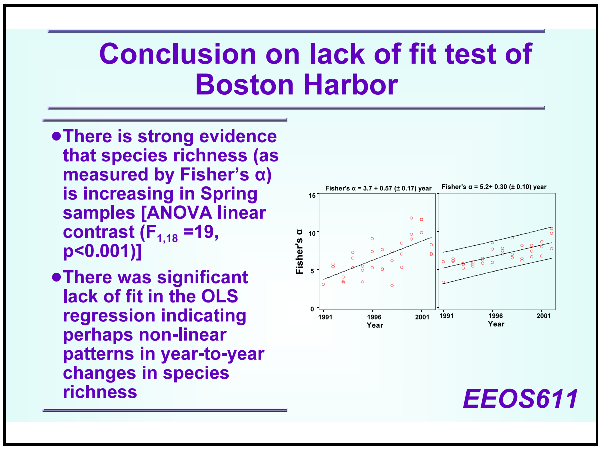
Slide 33

NOTES:



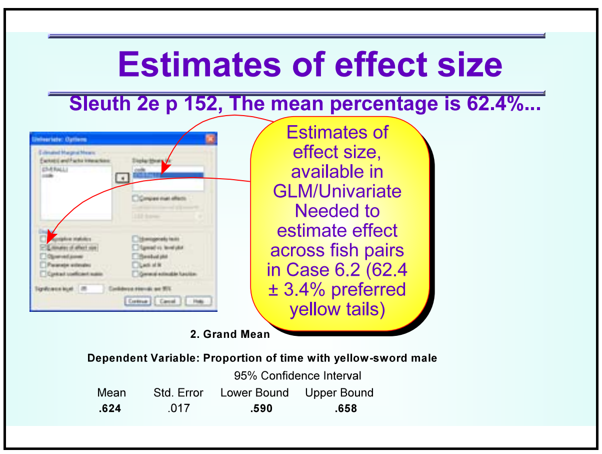
Slide 34

NOTES:



Slide 35 Conclusion on lack of fit test of Boston Harbor

NOTES:



Slide 36 Estimates of effect size

NOTES:

More on Simultaneous Inferences

Confidence limits

- **Individual** (pairwise) confidence level is the frequency with which a single interval captures its parameter.
- **Overall** (familywise or experiment-wise) confidence level is the frequency with which all intervals simultaneously capture their parameters.
- Planned *vs.* Unplanned comparisons

EEOS611

Slide 37 More on Simultaneous Inferences

NOTES:

Multiple comparisons (1 of 2)

Interval half width = Multiplier x Standard error

- LSD (Least Significant Difference): Student's *t* with pooled standard error — **no protection against multiple hypothesis testing**
- F-protected Inference
  - Fisher's protected Least Significant Difference
  - Don't claim a difference if the overall F statistic is not significant
- **Tukey-Kramer,** Studentized range Table A.5
  - Generalization of Tukey's HSD (Honestly Significant Difference) for unequal sample sizes)
  - Games-Howell more robust to unequal variance

EEOS611

Slide 38 Multiple comparisons (1 of 2)

NOTES:

Multiple comparisons (2 of 2)

Interval half width = Multiplier x Standard error

- **Bonferroni**, based on the number of comparisons (α/possible tests)
  - A conservative test (most often applied a posteriori test in drug trials for unplanned comparisons)
  - Test α = Experiment-wise α/k, where k is the number of tests
    - This approximation provides a remarkably accurate estimate of
    - Experiment-wise alpha: $\alpha_{exp} \approx 1-(1 - \alpha_{test})^k$, whre k is the number of tests
    - For example, 20 groups being tested 2 at a time
      - 20 Choose 2 tests = 190
      - Experiment-wise $\alpha = 1-(1- \mathbf{0.05})^{190}$
      - Experiment-wise α = 0.99994
      - But $0.04877683466514 = 1-(1- \mathbf{0.05/190})^{190}$
- **Scheffé**, based on the number of linear contrasts: most conservative of the widely used multiple comparison tests
- **Others** Sokal & Rohlf's Biometry, Quinn & Keough and Toothacker provide comprehensive listing
  - Newman-Keuls, SNK, *Student-Newman-Keuls*; based on studentized range, more powerful (less conservative) than Tukey-Kramer
  - Duncan's multiple range
  - **Dunnet's, where there is a control group**
  - Dunn's for non-parametric *a posteriori* contrasts

EEOS611

Slide 39 Multiple comparisons (2 of 2)

NOTES:

Display 6.6

Summary of 95% confidence interval procedures for differences between treatment means in the handicap study

Group	Average	Difference with ... *hearing*	*amputee*	*control*	*wheelchair*
crutches	5.921	1.871	1.492	1.021	0.578
wheelchair	5.343	1.293	0.914	0.443	
control	4.900	0.850	0.471		
amputee	4.429	0.379			
hearing	4.050				

Procedure	95% Interval Halfwidth
LSD	1.233
Tukey-Kramer	1.735
Bonferroni	1.794
Scheffé	1.957

A confidence interval is centered at a difference with halfwidth given by one of the procedures

Slide 40 Display 6.6

NOTES:

SPSS output from GLM

```
UNIANOVA
score  BY code
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/LMATRIX = "Avg A H vs Avg C W" code 0 -1/2 1/2 -1/2 1/2
/POSTHOC = code ( TUKEY SCHEFFE LSD BONFERRONI )
/CRITERIA = ALPHA(.05)
/DESIGN = code .
```

Qualification

	Handicap	N	Subset 1	Subset 2
Tukey HSD[a,b]	Hearing	14	4.1	
	Amputee	14	4.4	4.4
	Control	14	4.9	4.9
	Wheelchair	14	5.3	5.3
	Crutches	14		5.9
	Sig.		.23	.12
Scheffe[a,b]	Hearing	14	4.1	
	Amputee	14	4.4	
	Control	14	4.9	
	Wheelchair	14	5.3	
	Crutches	14	5.9	
	Sig.		.07	

Means for groups in homogeneous subsets are displayed.
Based on Type III Sum of Squares
The error term is Mean Square(Error) = 2.666.
a. Uses Harmonic Mean Sample Size = 14.000.
b. Alpha = .05.

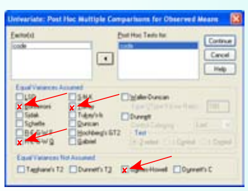
Slide 41 SPSS output from GLM

NOTES:

Multiple Comparisons available in SPSS

See Garson web site

- Bonferroni: a conservative test (beware Type II error)
- Tukey-Kramer
  - In SPSS, if you ask for the Tukey test and sample sizes are unequal, you will get the Tukey-Kramer test, using the harmonic mean.
- Games-Howell, a modified Tukey-Kramer appropriate when the homogeneity of variances assumption is violated, controls for unequal sample sizes
- Ryan test (REGWQ): modified Newman-Keuls test
  - Toothaker (1993: 56) calls Ryan the "best choice" among tests supported by major statistical packages because maintains good alpha control (ex., better than Newman-Keuls) while having at least 75% of the power of the most powerful tests (ex., better than Tukey HSD).



Slide 42 Multiple Comparisons available in SPSS

NOTES:

Quinn & Keough review of multiple comparison tests

- Use planned (*a priori*) contrasts whenever possible for testing specific differences among groups
- "If unplanned comparisons must be used, Ryan's REGW or Tukey's tests are recommended, the latter if simultaneous confidence intervals are required." (P. 207)
  - REGW: Ryan, Einot, Gabriel & Welch procedure.

EEOS611

Slide 43 Quinn & Keough review of multiple comparison tests

NOTES:

Ryan's test: REGW

From SPSS algorithms

Ryan, Einot, Gabriel, and Welsch (R-E-G-W) developed two multiple step-down range tests. Multiple step-down procedures first test whether all means are equal. If all means are not equal, subsets of means are tested for equality. R-E-G-W F is based on an F test and R-E-G-W Q is based on the Studentized range. These tests are more powerful than Duncan's multiple range test and **Student-Newman-Keuls** (which are also multiple step-down procedures), **but they are not recommended for unequal cell sizes.** *<emphasis added by Gallagher>*

Slide 44 Ryan's test: REGW

NOTES:

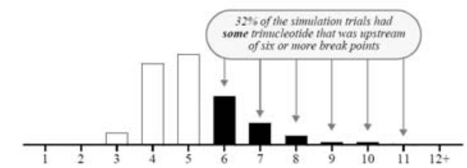
TGG before break in line 7; 6 of the 11 breaks occured 'downstream' of TGG, p= 0.000243

Slide 45 Display 6.7

NOTES:

Display 6.8

Simulated estimate of the distribution of the highest frequency of occurrence of a trinucleotide upstream of eleven randomly-selected breaks



Highest Frequency of Occurrence for a Trinucleotide Upstream of the Break Points

P=0.32 from a Monte Carlo simulation, but p=0.000243 from a test that didn't take into account the number of possible tests

Slide 46 Display 6.8

NOTES:

Conclusions to Chapter 6

1 of 4

- ANOVA is a subset of regression and both are subsets of general linear models
  - SPSS UNIANOVA is the standard GLM package in SPSS
  - GLM/UNIANOVA & regression have the greatest flexibility
- Linear contrasts: can be called through GLM or ANOVA
  - SPSS's Oneway only allows integer contrasts
    - With integer contrasts, p values are identical for any contrast vector multiplied by a scalar (effect sizes and standard errors increase proportionately)
  - With fractional contrasts in GLM/univariate, effect sizes & standard errors don't need to be rescaled
  - Matlab's orthpoly.m (statbox toolbox) solves orthogonal contrasts for any vector of explanatory variables
  - Can be used as an accepted alternative to regression when there is 'lack of fit' due to cluster effects
    - E.g., Boston Harbor regression of biodiversity vs. Year

EEOS611

Slide 47 Conclusions to Chapter 6

NOTES:

Conclusions to Chapter 6

2 of 4

- Planned and unplanned comparisons
  - Always try to specify hypotheses *a priori*, and use the LSD test (or equivalent linear contrast test) at a predetermined experiment-wise α level (usually 0.05)
  - Use the √(Within groups MS) as the pooled estimate of population *s* for these tests
- Unplanned comparisons
  - Also called: *ad hoc*, *a posteriori*, multiple comparison tests
  - Experiment-wise (family-wise) error levels usually set at 0.05

EEOS611

Slide 48 Conclusions to Chapter 6

NOTES:

Conclusions to Chapter 6

3 of 4

- Linear contrasts
  - E.g., Avg (A,B) vs. Avg (C,D,E)
  - Only Scheffé procedure should be used
- Bonferroni
  - A conservative test
  - Test α = Experiment-wise α/ k, where k is the number of tests
  - Experiment-wise $\alpha \approx 1-(1-\text{Test } \alpha)^k$
  - For example, 20 groups being tested 2 at a time
    - 20 Choose 2 tests = 190
    - Experiment-wise $\alpha = 1-(1-.05)^{190}$
    - Experiment-wise α = 0.99994
    - But $0.04877683466514 = 1-(1- 0.05/190)^{190}$

EEOS611

Slide 49 Conclusions to Chapter 6

NOTES:

Conclusions to Chapter 6

4 of 4

- Tukey-Kramer
  - Tukey's HSD ('Honestly significant difference') with adjustments for unequal sample sizes
  - Assumes equal variance (Games-Howell protects for unequal variance)
- Treatment vs. Control: use Dunnet's test (only n-1 comparisons, not ${}_nC_2$)
- More powerful tests
  - SNK: recommended by Underwood
  - Ryan's test (REGWF), recommended by Quinn & Keough (in addition to Tukey-Kramer)
    - **Ryan's test not suitable for unequal group sizes (SPSS)**

Slide 50 Conclusions to Chapter 6

NOTES:

Chapter 7 Simple Linear Regression: a model for the mean

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

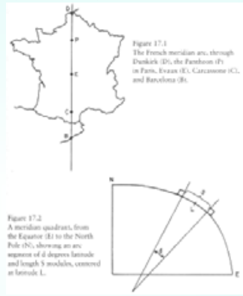
EEOS611

Slide 51 Chapter 7 Simple Linear Regression: a model for the mean

NOTES:

Legendre (1805) & least squares

Used to define the meter, 1/10,000,000 meridional arc



See: Stigler (1986) on statistical history and Alder's 'The Measure of all Things' (2002) on why the meridional circumference is now 40,007.849 km 0.02% deviation from true value

Slide 52 Legendre (1805) & least squares

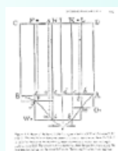
NOTES:

Probability & least squares landmarks

Legendre (1805) gets priority for the method of Least Squares

- **Jacob Bernoulli (1654-1705)**
  - His work led to the binomial distribution
- **De Moivre (1667-1754)**
  - 1733 described what would later be called the normal curve
- **Bayes (1764): Bayes theorem**
- **Legendre (1805) least squares**
- **Gauss (1809)**
  - Reported using least squares since 1795
- **Laplace**
  - 1810: central limit theorem
  - 1827: least squares theory

Gauss



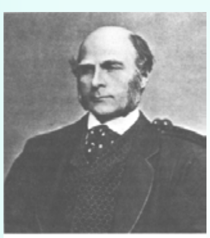
The normal (Gaussian) curve

Slide 53 Probability & least squares landmarks

NOTES:

Sir Francis Galton

Inventor of the Qunicunx & 'regression'



Francis Galton (1822–1911)

Slide 54 Sir Francis Galton

NOTES:

Galton (1885) on filial height

Girls x 1.08

Table 8.1 Galton's 1885 cross-tabulation of 928 adult children born of 205 midparents, by their height and their midparent's height.

Height of the mid-parent in inches	Height of the adult child <61.7	62.2	63.2	64.2	65.2	66.2	67.2	68.2	69.2	70.2	71.2	72.2	73.2	>73.7	Total no. of adult children	Total no. of mid-parents	Median
>73.0	—	—	—	—	—	—	—	—	—	—	—	1	3	—	4	5	—
72.5	—	—	—	—	—	—	—	1	2	1	2	7	2	4	19	6	72.2
71.5	—	—	—	—	1	3	4	3	5	10	4	9	2	2	43	11	69.9
70.5	1	—	1	—	1	1	3	12	18	14	7	4	3	3	68	22	69.5
69.5	—	—	1	16	4	17	27	20	33	25	20	11	4	5	183	41	68.9
68.5	1	—	7	11	16	25	31	34	48	21	18	4	3	—	219	49	68.2
67.5	—	3	5	14	15	36	38	28	38	19	11	4	—	—	211	33	67.6
66.5	—	3	3	5	2	17	17	14	13	4	—	—	—	—	78	20	67.2
65.5	1	—	9	5	7	11	11	7	7	5	2	1	—	—	66	12	66.7
64.5	1	1	4	4	1	5	5	—	2	—	—	—	—	—	23	5	65.8
<64.0	1	—	2	4	1	2	2	1	1	—	—	—	—	—	14	1	—
Totals	5	7	32	59	48	117	138	120	167	99	64	41	17	14	928	205	—
Medians	—	—	66.3	67.8	67.9	67.7	67.9	68.3	68.5	69.0	69.0	70.0	—	—	—	—	—

Source: Galton (1886a).

Note: All female heights were multiplied by 1.08 before tabulation. Galton added an explanatory footnote to the table: "In calculating the Medians, the entries have been taken as referring to the middle of the squares in which they stand. The reason why the headings run 62.2, 63.2, &c., instead of 62.5, 63.5, &c., is that the observations are unequally distributed between 62 and 63, 63 and 64, &c., there being a strong bias in favour of integral inches. After careful consideration, I concluded that the headings, as adopted, best satisfied the conditions. This inequality was not apparent in the case of the Mid-parents." Galton republished these data in 1889, where they are referred to as the R.F.F. Data (Record of Family Faculties); he then noted that the first row must be in error (four children cannot have five sets of parents), but he claimed that "the bottom line, which looks suspicious, is correct" (p. 208).

Slide 55 Galton (1885) on filial height

NOTES:

Galton's regression to mediocrity

Campbell & Kenny, 1999 p. 2

- Galton (1822-1911) measured the heights of 928 parents and children
  - Multiplied female heights by a scaling factor
  - Galton (1886)
    - Tall parents tended to have tall children, but the children of the tallest parents were, on average, not quite as tall as the parents. Nor, were the children of the shortest parents as short as their parents
    - "filial regression toward mediocrity" (Galton 1886, p. 246)
    - Galton (1879, 1886) reasoned that there must be a biological force that made people move toward the mean, and he called that force **regression**.
    - Galton himself soon realized that the cause was statistical, not biological!
    - Stigler: the most remarkable discovery in all of statistics

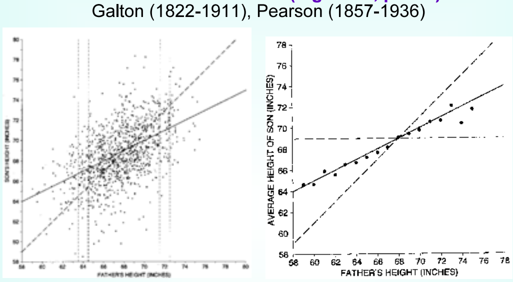
Slide 56 Galton's regression to mediocrity

NOTES:

Galton's regression to mediocrity

Friedman et al. 1998 (Fig. 10.5, p. 171)

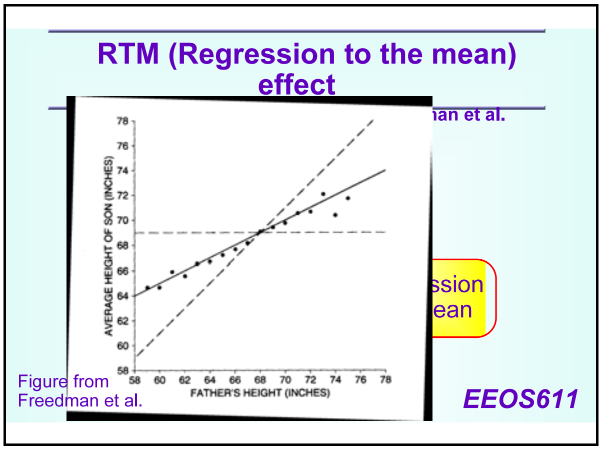
Galton (1822-1911), Pearson (1857-1936)



Pearson's data on 1078 fathers & sons at maturity

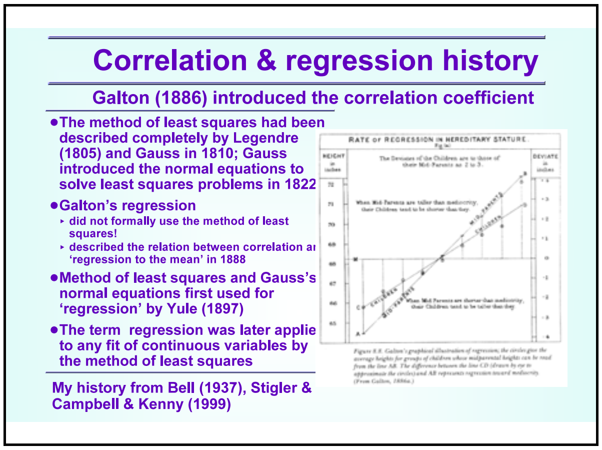
Slide 57 Galton's regression to mediocrity

NOTES:



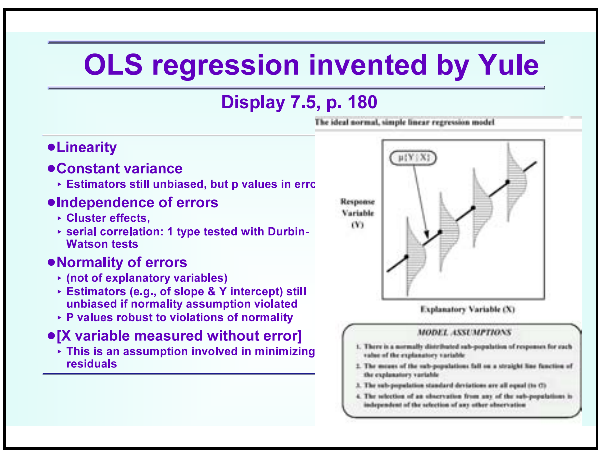
Slide 58 RTM (Regression to the mean) effect

NOTES:



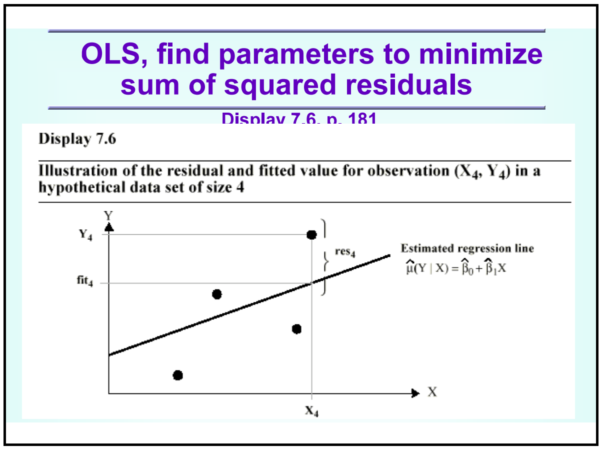
Slide 59 Correlation & regression history

NOTES:



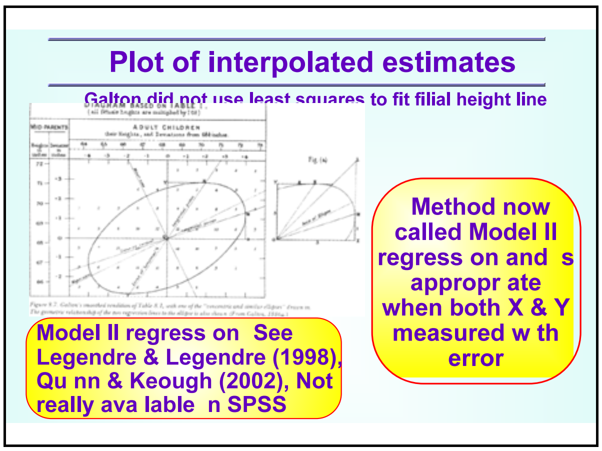
Slide 60 OLS regression invented by Yule

NOTES:



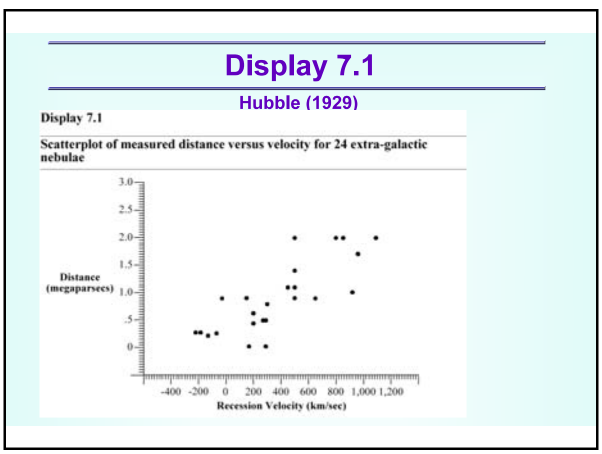
Slide 61 OLS, find parameters to minimize sum of squared residuals

NOTES:



Slide 62 Plot of interpolated estimates

NOTES:

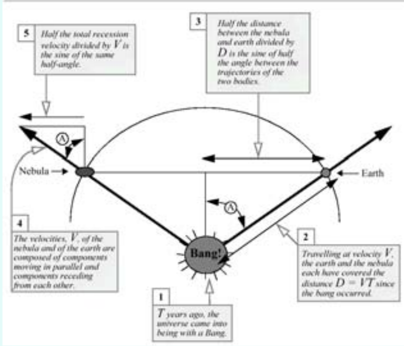


Slide 63 Display 7.1

NOTES:

Display 7.2

Big Bang Theory model for distance-velocity relationship of nebulae



Slide 64 Display 7.2

NOTES:

Questions from Hubble's data

- Is the relationship between distance & velocity a straight line?
- Is the y-intercept zero, as the Big Bang theory predicts?
- How old is the universe?
- Findings
  - Age of the universe: 1.88 billion years (with 95% CI of 1.5 to 2.27 billion years)
    - Current estimates of the Universe's age: 10 to 15 billion years
  - Probability that the Y intercept is zero is **0.0028**
  - Scope: Not a random sample of stars and errors in measuring velocities not included in p values

Slide 65 Questions from Hubble's data

NOTES:

SPSS solution

Constant is not zero (0.40 ± 0.25)

Coefficients[a]

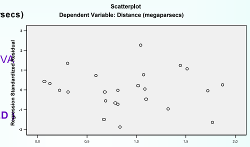
	Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
	B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
(Constant)	.399	.118		3.37	.003	.153	.645
Recession Velocity (km/sec)	.00137	.00023	.790	6.036	.000004	.00090	.00184

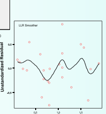
a. Dependent Variable: Distance (megaparsecs)

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT distance
  /METHOD=ENTER velocity
  /SCATTERPLOT=(*ZRESID ,*ADJPRED )
  /SAVE PRED RESID .
```

Scatterplot

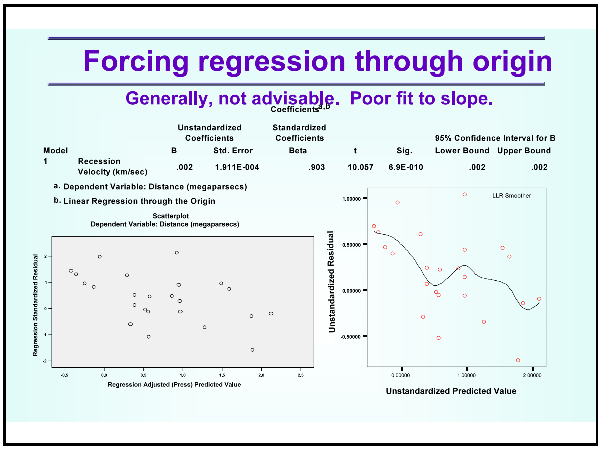
Dependent Variable: Distance (megaparsecs)





Slide 66 SPSS solution

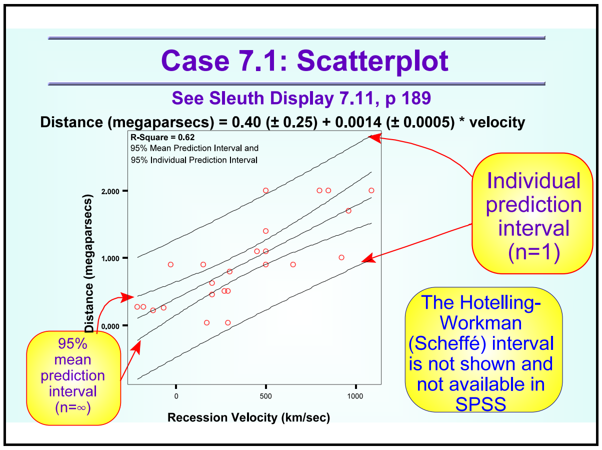
NOTES:



Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	Upper Bound
1	Recession Velocity (km/sec)	.002	1.911E-004	.903	10.057	6.9E-010	.002	.002

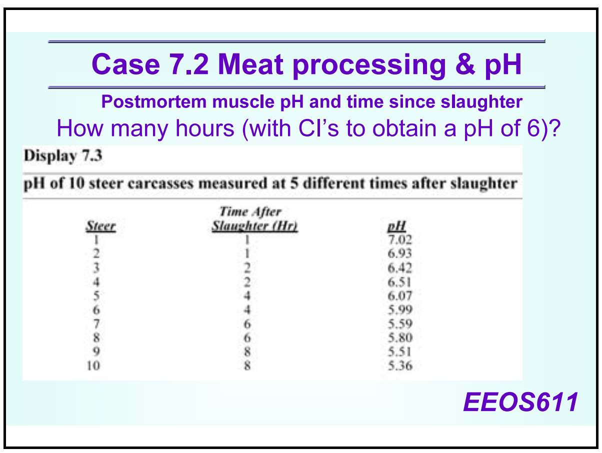
Slide 67 Forcing regression through origin

NOTES:



Slide 68 Case 7.1: Scatterplot

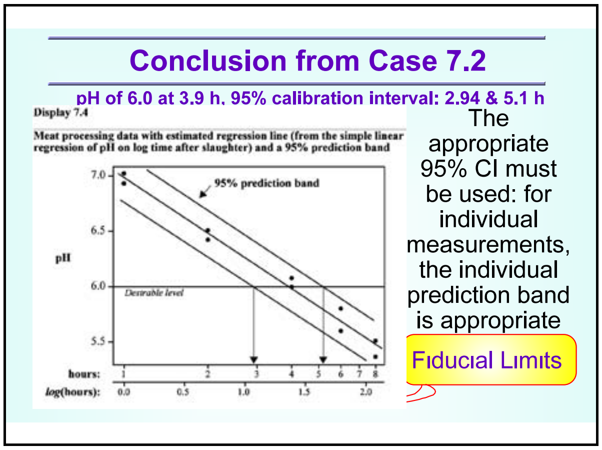
NOTES:



Steer	Time After Slaughter (Hr)	pH
1	1	7.02
2	1	6.93
3	2	6.42
4	2	6.51
5	4	6.07
6	4	5.99
7	6	5.59
8	6	5.80
9	8	5.51
10	8	5.36

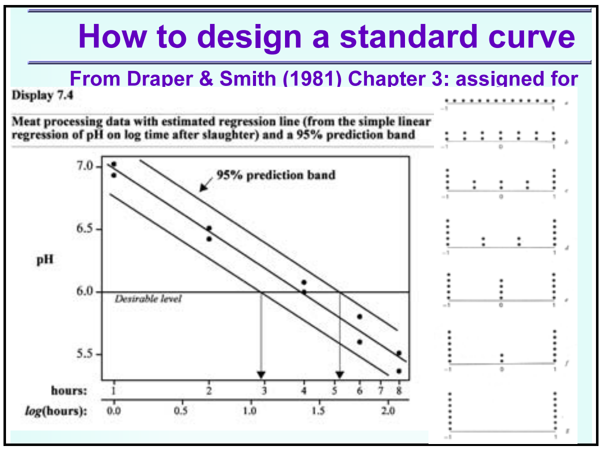
Slide 69 Case 7.2 Meat processing & pH

NOTES:



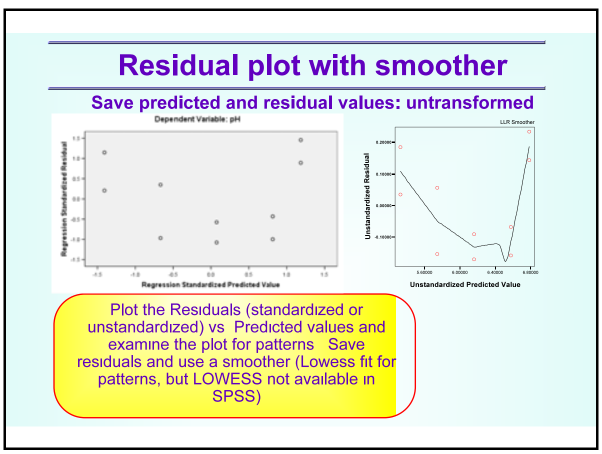
Slide 70 Conclusion from Case 7.2

NOTES:



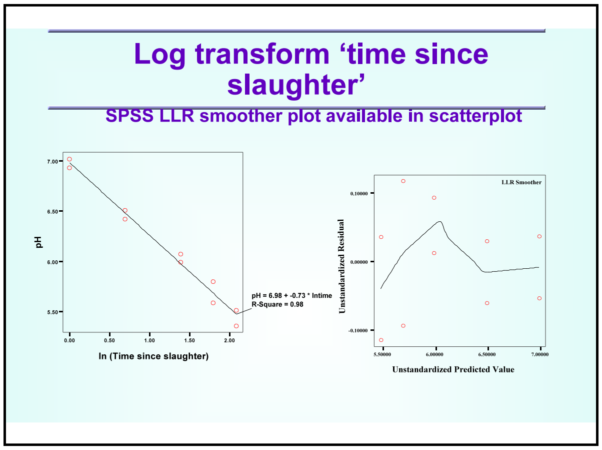
Slide 71 How to design a standard curve

NOTES:



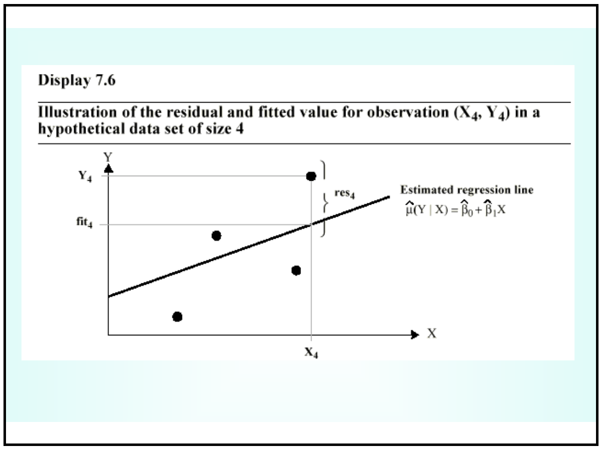
Slide 72 Residual plot with smoother

NOTES:



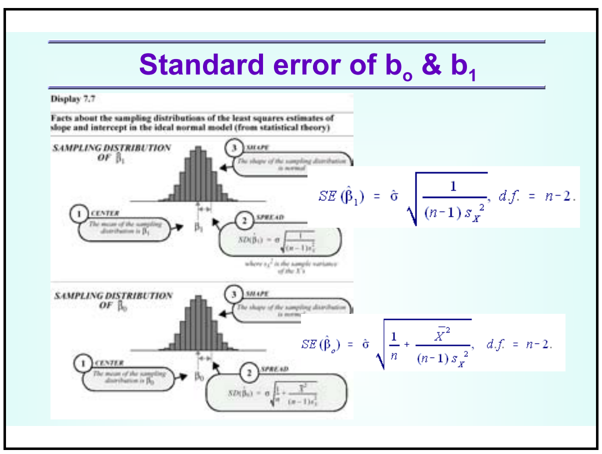
Slide 73 Log transform 'time since slaughter'

NOTES:



Slide 74 Display 7.6

NOTES:

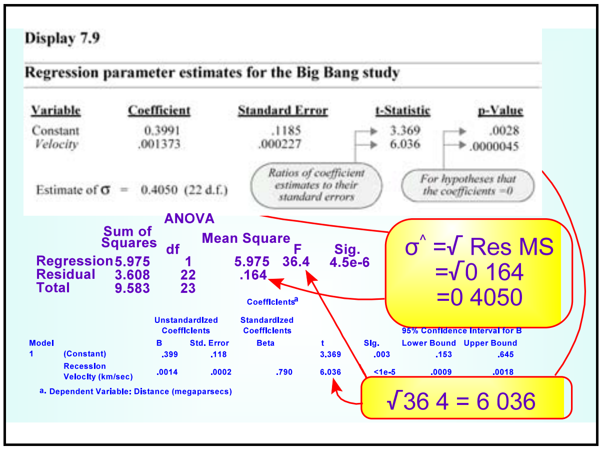


$$SE(\hat{\beta}_1) = \hat{\sigma}\sqrt{\frac{1}{(n-1)s_X^2}},\quad d.f. = n-2.$$

$$SE(\hat{\beta}_o) = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{(n-1)s_X^2}},\quad d.f. = n-2.$$

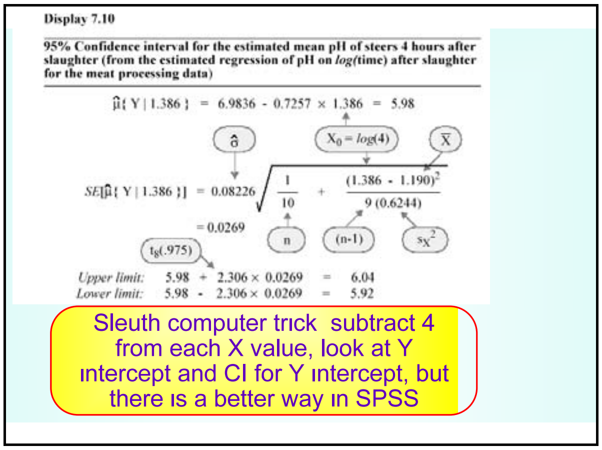
Slide 75 Standard error of bo & b1

NOTES:



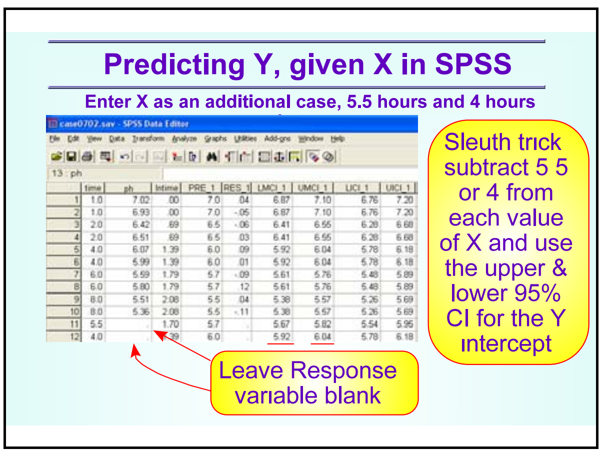
Slide 76 Display 7.9

NOTES:



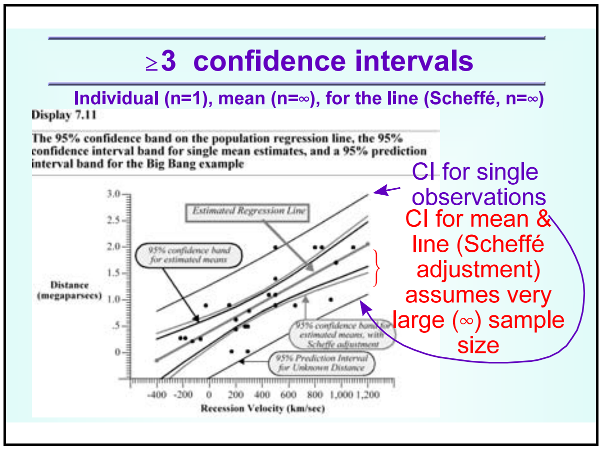
Slide 77 Display 7.10

NOTES:



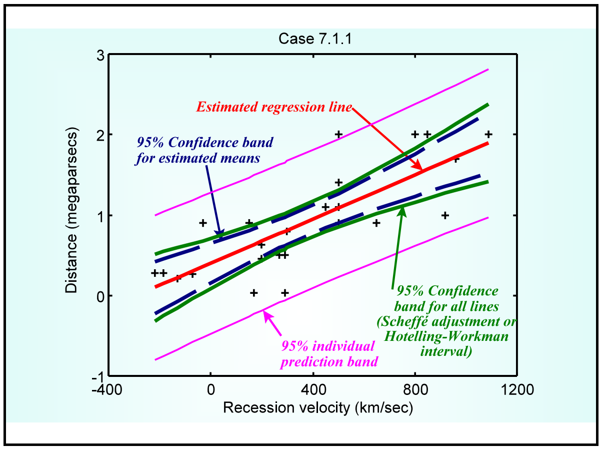
Slide 78 Predicting Y, given X in SPSS

NOTES:



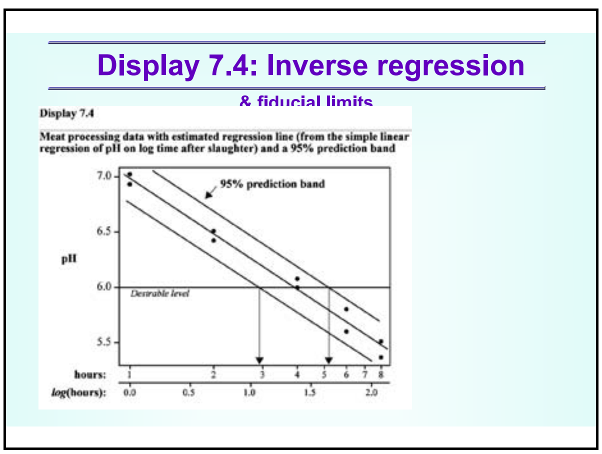
Slide 79 ≥3 confidence intervals

NOTES:



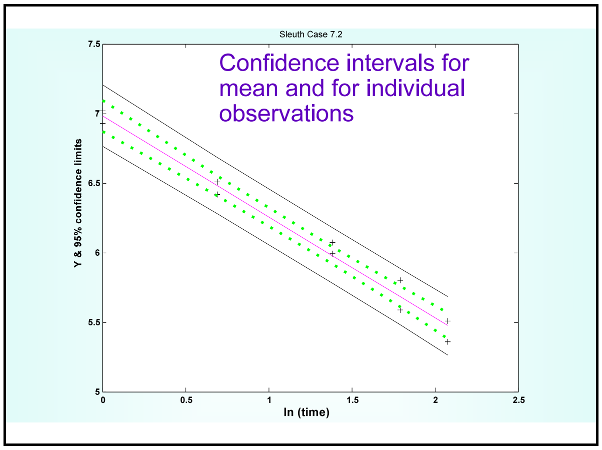
Slide 80

NOTES:



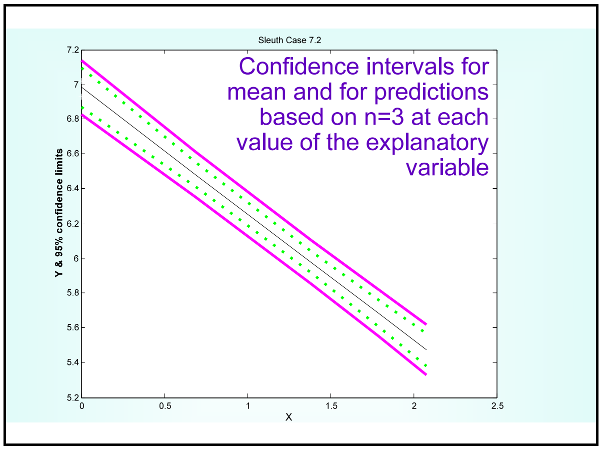
Slide 81 Display 7.4: Inverse regression

NOTES:



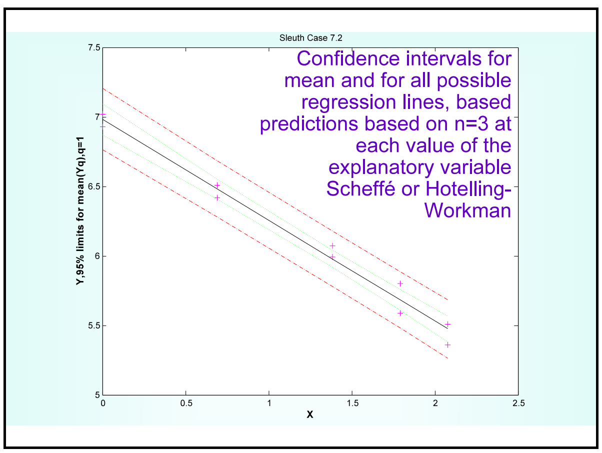
Slide 82

NOTES:



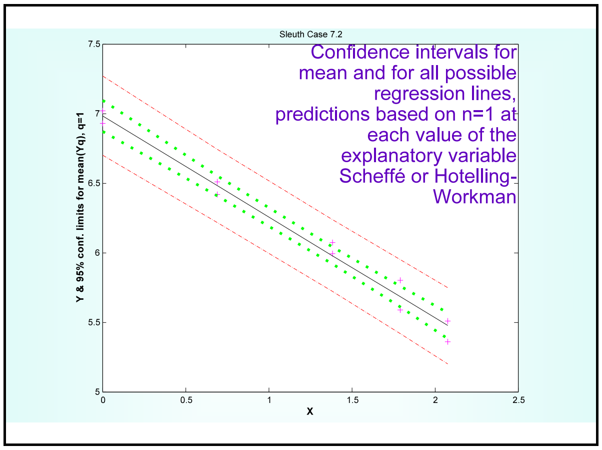
Slide 83

NOTES:



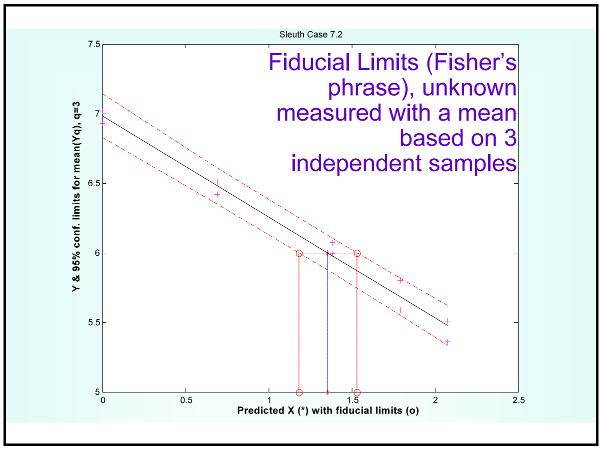
Slide 84

NOTES:



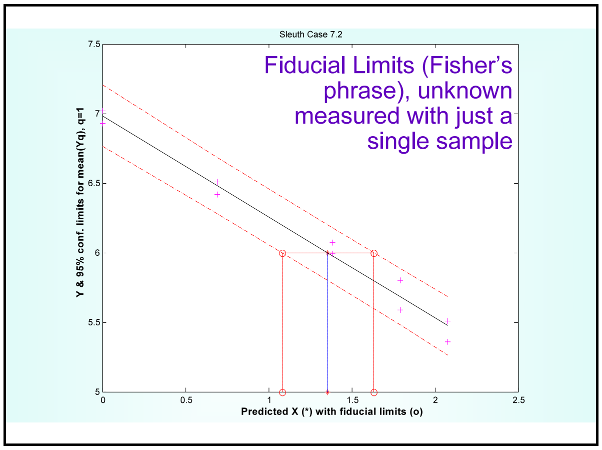
Slide 85

NOTES:



Slide 86

NOTES:

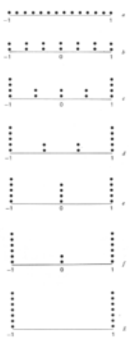


Slide 87

NOTES:

How to design a standard curve

From Draper & Smith (1981)



54 FITTING A STRAIGHT LINE BY LEAST SQUARES

Table 1.9 Characteristics of Various Strategies Depicted in Figure 1.13

	(a)	(b)	(c)	(d)	(e)	(f)	(g)
(1) Lack of fit df:	12	5	3	2	1	1	0
(2) Pure error df:	0	7	9	10	11	11	12
(3) S.d.$(b_1)/\sigma$:	0.43	0.40	0.33	0.31	0.32	0.29	0.27
(4) p sites:	14	7	5	4	3	3	2

f & *d* are the best choices, with f being preferred if a quadratic alternative is all that is being considered.

EEOS611

Slide 88 How to design a standard curve

NOTES:

More regression to the mean

Important note: regression to the mean is a group phenomenon. Groups of individuals will show a consistent regression to the mean, but individuals may not. There are correction factors that take into account the RTM phenomenon. 'Empirical Bayes Estimators' are now the accepted correction procedures to correct the accuracy of predictions for the RTM phenomenon.

EEOS611

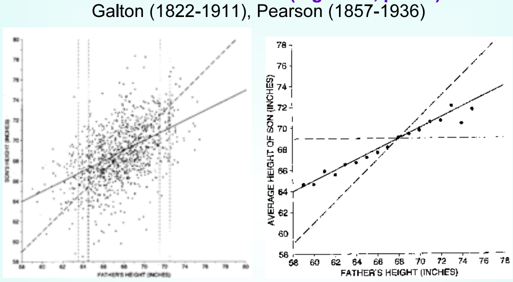
Slide 89 More regression to the mean

NOTES:

Galton's regression to mediocrity

Friedman et al. 1998 (Fig. 10.5, p. 171)

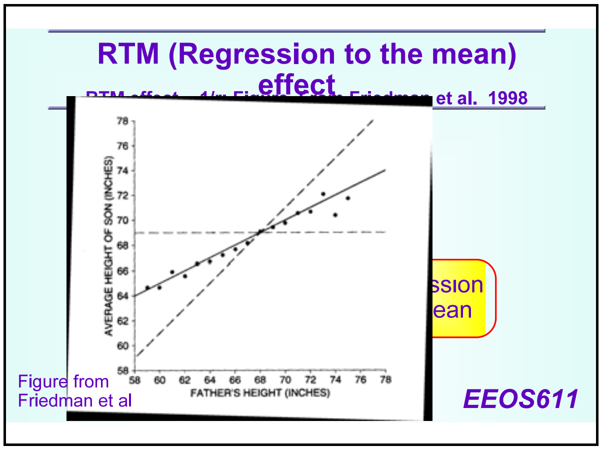
Galton (1822-1911), Pearson (1857-1936)



Pearson's data on 1078 fathers & sons at maturity

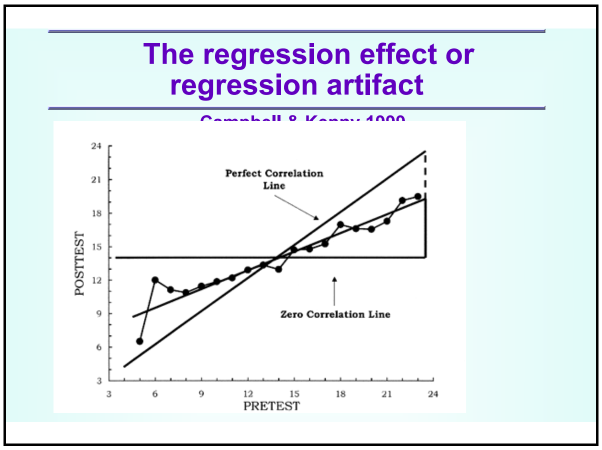
Slide 90 Galton's regression to mediocrity

NOTES:



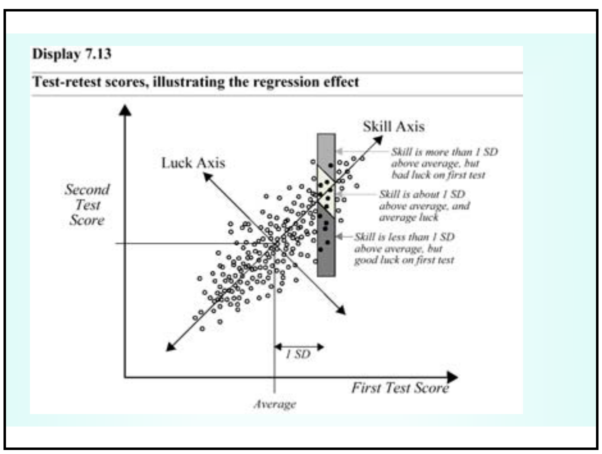
Slide 91 RTM (Regression to the mean) effect

NOTES:



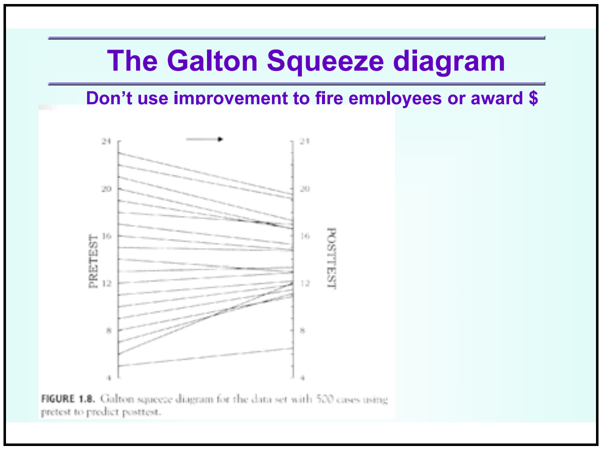
Slide 92 The regression effect or regression artifact

NOTES:



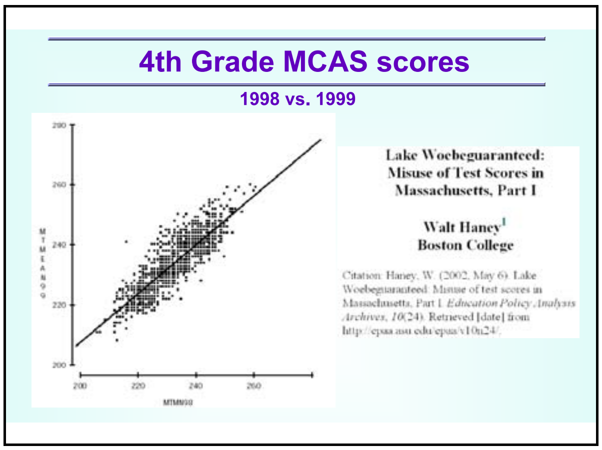
Slide 93 Display 7.13

NOTES:



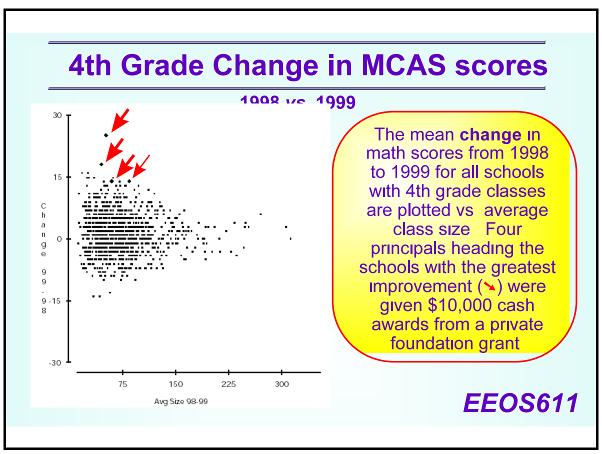
Slide 94 The Galton Squeeze diagram

NOTES:



Slide 95 4th Grade MCAS scores

NOTES:

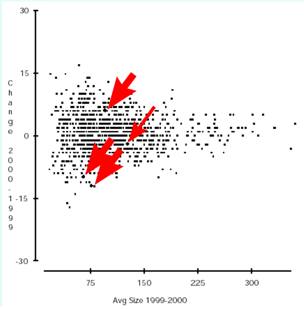


Slide 96 4th Grade Change in MCAS scores

NOTES:

4th Grade Change in MCAS scores

1998 *vs.* 1999, from Walt Haney BC



3 of the 4 award-winning schools declined in performance in their post-award year

Why?

EEOS611

Slide 97 4th Grade Change in MCAS scores

NOTES:

The regression artifact

- What major statistical principal must be considered when analyzing test and retest data of this sort? [A 2003 midterm question]
  - Two related statistical problems
    - Regression to the mean, which is a strong function of the correlation between tests. The weaker the correlation between tests, the more the regression to the mean phenomenon
    - The effects of sample size on the difference in averages.
  - Note that RTM is a group phenomenon, *"You cannot tell which way an individual's score will move based on the regression to the mean phenomenon. Even though the group's average will move toward the population's average, some individuals in the group are likely to move in the other direction."* Quote from Trochim's RTM web site

Slide 98 The regression artifact

NOTES:

Luck or skill in awarding MCAS winners (1 of 3)

- As discussed on page 192 in Sleuth, in a test-retest situation (and many other situations of repeated measures on subjects) the change scores are composed of a true "skill" effect, the improvement in student performance and error.
- The error is reflected in the lack of perfect correlation between the 1st and 2nd tests.
  - In this case, the correlation is 0.86 between 1998 and 1999. The lack of perfect correlation could be due to differences in the teaching quality between schools, but some is due to just test-to-test variability.

Slide 99 Luck or skill in awarding MCAS winners (1 of 3)

NOTES:

Luck or skill in awarding MCAS winners (2 of 3)

- The RTM effect is directly proportional to (1-r), with r being the test-to-test correlation.
  - With perfect correlation, there is no RTM effect.
  - The Dept of Education identified schools based on their change scores on the 1998 to 1999 exams, and most of these schools had small class sizes.
  - Smaller class sizes will be associated with sample averages that deviate from the true mean to a far greater extent than large schools.
    - The extent of this deviation is assessed with the standard error of the difference in averages, with standard errors proportional to $(1/n_1 + 1/n_2)$, where n_1 and n_2 are the class sizes for the two exams.

EEOS611

Slide 100 Luck or skill in awarding MCAS winners (2 of 3)

NOTES:

Luck or skill in awarding MCAS winners (3 of 3)

- Take into account the standard error of the difference, use p values based on change/(standard error of change) instead of absolute differences
- Use Empirical Bayes estimators (James-Stein estimators) to adjust for the chance element in assessing change (used for batting averages & hospital mortality by Effron & Morris)
- Use hierarchical longitudinal models, assessing change in individual student performance

EEOS611

Slide 101 Luck or skill in awarding MCAS winners (3 of 3)

NOTES: