

<p style="text-align: center;">Ch 6: Linear combinations and multiple comparisons of means & Ch7 Simple linear regression: a model for the mean</p> <p style="text-align: center;">Class 10: 3/9/09 M</p> <p style="text-align: right;">EEOS611</p>	<p>Slide 1 Ch 6: Linear combinations and multiple comparisons of means</p> <p>& Ch7 Simple linear regression: a model for the mean</p> <p>NOTES:</p>
<p style="text-align: center;">HW 8 due Thus 3/12/09 11 am</p> <p style="text-align: center;">Submit as Myname-HW8.doc (or *.rtf)</p> <ul style="list-style-type: none"> • Read Chapter 7 Comparisons among several samples • Comment on Chapter 7 conceptual problems in Blackboard Vista4 • Computation Problem 8 <ul style="list-style-type: none"> • 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 <i>a priori</i> <p style="text-align: right;">EEOS611</p>	<p>Slide 2 HW 8 due Thus 3/12/09 11 am</p> <p>NOTES:</p>
<p style="text-align: center;">HW 9 due Monday 3/16/09 10 am</p> <p style="text-align: center;">Submit as Myname-HW9.doc (or *.rtf)</p> <ul style="list-style-type: none"> • Read Chapter 8 A closer look at assumptions for simple linear regression • Comment on Chapter 8 conceptual problems in Blackboard Vista4 • Computation Problem 9 <ul style="list-style-type: none"> • Problem 7.29 (Sleuth 2nd edition, p. 203) Male displays <p style="text-align: right;">EEOS611</p>	<p>Slide 3 HW 9 due Monday 3/16/09 10 am</p> <p>NOTES:</p>

<div data-bbox="349 168 665 199" data-label="Section-Header"> <h3>Student Presentations</h3> </div> <div data-bbox="355 212 652 233" data-label="Text"> <p>Starting at 10:50 (8 minutes each)</p> </div> <div data-bbox="256 245 719 350" data-label="List-Group"> <ul style="list-style-type: none"> • Seth Sheldon for HW 3 <ul style="list-style-type: none"> ▸ 2.21 Bumpus's data: weights of Bumpus's birds • Barry Fradkin for HW 4. <ul style="list-style-type: none"> ▸ 3.28 Pollen removal </div> <div data-bbox="657 506 779 533" data-label="Text"> <p>EEOS611</p> </div>	<div data-bbox="820 132 1232 163" data-label="Section-Header"> <h3>Slide 4 Student Presentations</h3> </div> <div data-bbox="820 254 940 285" data-label="Text"> <p>NOTES:</p> </div>
<div data-bbox="279 688 737 798" data-label="Section-Header"> <h3>Ch 6: Linear combinations and multiple comparisons of means</h3> </div> <div data-bbox="657 995 779 1022" data-label="Text"> <p>EEOS611</p> </div>	<div data-bbox="820 621 1359 693" data-label="Section-Header"> <h3>Slide 5 Ch 6: Linear combinations and multiple comparisons of means</h3> </div> <div data-bbox="820 781 940 812" data-label="Text"> <p>NOTES:</p> </div>
<div data-bbox="381 1184 617 1215" data-label="Section-Header"> <h3>Case Study 6.1.1</h3> </div> <div data-bbox="334 1226 686 1249" data-label="Text"> <p>Discrimination against the handicapped</p> </div> <div data-bbox="256 1260 709 1518" data-label="List-Group"> <ul style="list-style-type: none"> • U.S. Vocational Rehabilitation Act of 1973 • 5 Videotaped job interviews <ul style="list-style-type: none"> ▸ Applicant appeared with different handicaps <ul style="list-style-type: none"> ▸ 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 </div> <div data-bbox="657 1522 779 1549" data-label="Text"> <p>EEOS611</p> </div>	<div data-bbox="820 1148 1159 1180" data-label="Section-Header"> <h3>Slide 6 Case Study 6.1.1</h3> </div> <div data-bbox="820 1270 940 1302" data-label="Text"> <p>NOTES:</p> </div>

Display 6.1

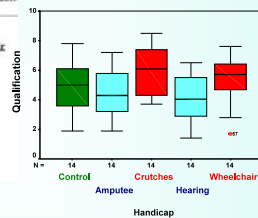
Applicant qualification scores; Control group is in the middle of the distribution of scores

Display 6.1

Stem-and-leaf diagrams of applicant qualification scores given to applicants simulating five different handicap conditions

	None	Amputee	Crutches	Hearing	Wheelchair
0					
1	9	9	4	7	
2	5	56	149	8	
3	06	268	7	479	5
4	129	06	033	237	78
5	149	1589	18	589	03
6	17	1	0234	5	1124
7	88	2	445		246
8			5		
9					

Legend: 7|4 represents a score of 7.4 on the Applicant Qualification Scale



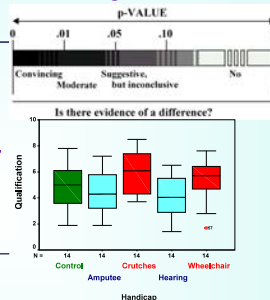
Slide 7 Display 6.1

NOTES:

Case 6.01

Summary of Statistical Findings

- Strong but not convincing evidence $p = 0.03$ that applicants' ratings influenced by handicaps
- Ave 'Crutches' score much higher (1.87 ± 0.73) than hearing, using Tukey-Kramer multiple comparison test (difference in score $\pm \frac{1}{2}$ 95% CI)
- The strongest evidence is for a difference between "wheelchair & crutches" vs. "Amputee & hearing" (t-statistic = 3.19 for linear contrasts), with a 1.4 ± 0.9 higher average score for the former group
 - Since this was not a planned comparison, the Scheffé multiplier was used to calculate the 95% CI for the difference
- None of the feigned handicaps different from control! (The protected least significant differences all have 2-sided $p > 0.05$)



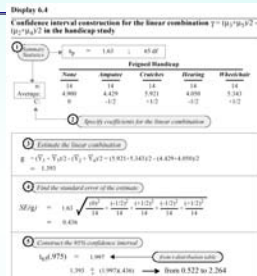
Slide 8 Case 6.01

NOTES:

Display 6.4

Wheelchair + Crutches vs. Amputee + hearing

- Linear contrasts can be solved
 - By hand
 - With SPSS Oneway
 - SPSS GLM (analyze/general linear model/univariate)
 - the appropriate 95% CI's for the average difference can be calculated
- SPSS routine Oneway will calculate the appropriate p value using a linear contrast, but it will not present the appropriate difference in means



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 1 1 -1 1/2 1/2
• /POSTHOC = code (TUKEY SCHEFFE LSD BONFERRONI)
• /CRITERIA = ALPHA(.05)
• /DESIGN = code .

Contrast Results (K Matrix)

Contrast Estimate	Hypothesized Value	Difference (Estimate - Hypothesized)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1.4	0	1.4	.44	.002	.522	2.264

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

Construct the 95% confidence interval
 $t_{(65), .975} = 1.997$ from redistribution table
 $1.393 \pm (1.997)(.436) \rightarrow \text{from } 0.522 \text{ to } 2.264$

EEOS611

Slide 10 SPSS syntax for Linear contrasts

NOTES:

Slide 11

NOTES:

Construct the 95% confidence interval
 $t_{(65), .975} = 1.997$ from redistribution table
ANOVA $1.393 \pm (1.997)(.436) \rightarrow \text{from } 0.522 \text{ to } 2.264$
Scheffé: $1.393 \pm 3.1705 * 0.436 \rightarrow \text{from } 0.011 \text{ to } 2.775$
 $p=0.0476$

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

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

Contrast Tests

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

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

Slide 12 Case 6.1.1

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 CIs assume planned or *a priori* contrasts
 - What is the Tukey-Kramer procedure?
 - What is "the protected least significant difference"?
 - When should the Bonferroni & Scheffé procedures be used?

Case 6.1.2

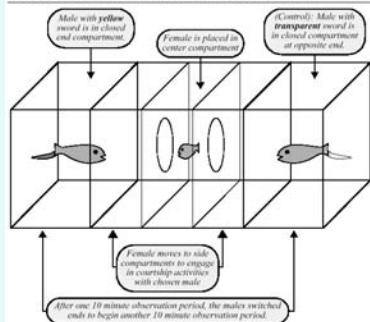
Preexisting preferences of fish – a randomized experiment

- 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

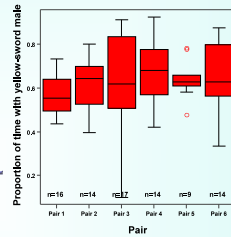
Slide 15 Display 6.3

NOTES:

Sexual preference

Case Study 6.2

- Test for preference for yellow-sword male (expected proportion = $\frac{1}{2}$)
- 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 ($\pm 99.9\%$ 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 ($p\text{-value} = 0.56$, from ANOVA F-statistic).

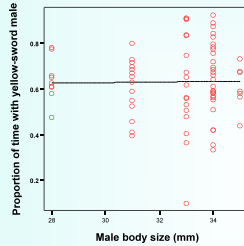
EEOS611

Slide 18 Test for any difference among pairs

NOTES:

Testing a linear contrast

Another way of handling some types of Analysis of Covariance: ANCOVA



Is there a non zero slope between proportion of time spent with yellow sword and male body size?

EEOS611

Slide 19 Testing a linear contrast

NOTES:

Testing for the male fish weight effect

t-test for Linear Effect of Body Size

Group	n	Average (%)	Standard Deviation	Male Body Size (mm)	Coefficient
Pair 1	16	56.41	9.02	35	5
Pair 2	14	60.89	12.48	31	-3
Pair 3	17	62.43	22.29	33	1
Pair 4	14	67.00	14.33	34	3
Pair 5	9	64.21	9.41	28	-9
Pair 6	13	63.34	17.68	34	3
Pooled	84	62.13	15.46	Average = 32.5	

① Calculate the coefficients for the linear combination

$$C_i = 2 * (X_i - 32.5)$$

Gallagher Matlab code:

```
% Case0602 m
BodySize [35 31 33 34 28 34];
mn = mean(BodySize);
% Subtract the mean from each body length
Dev = BodySize - repmat(mn, size(BodySize));
% Solution:
>> Dev =
    2.5000
   -1.5000
    0.5000
    1.5000
   -4.5000
    1.5000
```

Slide 20 Testing for the male fish weight effect

NOTES:

t test for linear contrast

① Calculate the coefficients for the linear combination

$$C_i = 2 * (X_i - 32.5)$$

② Calculate the effect's estimate

$$\hat{\mu} = (5)(56.41) + (-3)(60.89) + (1)(62.43) + (3)(67.00) + (-9)(64.21) + (3)(63.34) = -25.06$$

and its standard error

$$SE(\hat{\mu}) = (15.46) \sqrt{\frac{(5)^2}{16} + \frac{(-3)^2}{14} + \frac{(1)^2}{17} + \frac{(3)^2}{14} + \frac{(-9)^2}{9} + \frac{(3)^2}{14}} = 54.77$$

③ Calculate the t-statistic and determine the p-value

$$t\text{-statistic} = \frac{-25.06}{54.77} = -0.458$$

1-sided p-value = 0.32
(from t-distribution with 78 df)

Conclusion: There is no evidence that the linear association between group means and male body size has a non-zero slope (1-sided p-value = 0.32).

Unless the authors had previous theory, the test should have been performed 2-tailed

Slide 21 t test for linear contrast

NOTES:

Analysis of swordtail linear contrast: ONEWAY or GLM?

Syntax posted on Blackboard/Vista 4

Title 'Case 6.1.2 - Sexual preference in swordtails'.

* This will find the std error but not do the CI

ONEWAY

prop BY code

/CONTRAST = 5 -3 1 3 -9 3.

* This call to GLM does it all.

UNIANOVA

prop BY code

/METHOD = SSTYPE(3)

/INTERCEPT = INCLUDE

/SAVE = PRED RESID

/EMMEANS = TABLES(OVERALL)

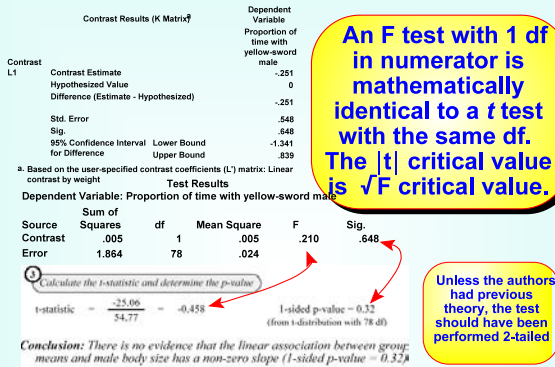
/LMATRIX = "Weight linear contrast" code 5 -3 1 3 -9 3

/CRITERIA = ALPHA(.05)

/DESIGN = code .

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

NOTES:



Slide 23

NOTES:

Matlab, Statbox orthpoly.m

Gordon Smyth's <Free> Statbox 4.2

Also includes all of the major probability distributions and includes a nice routine for Poisson regression.

<http://www.statsci.org/matlab/statbox>.

html>> help orthpoly

ORTHOPOLY ORTHPOLY(X,N) calculates the orthogonal polynomials up to order N corresponding to vector X.
BodySize=[35 31 33 34 28 34];
format rat;orthpoly(BodySize,2)

ans =

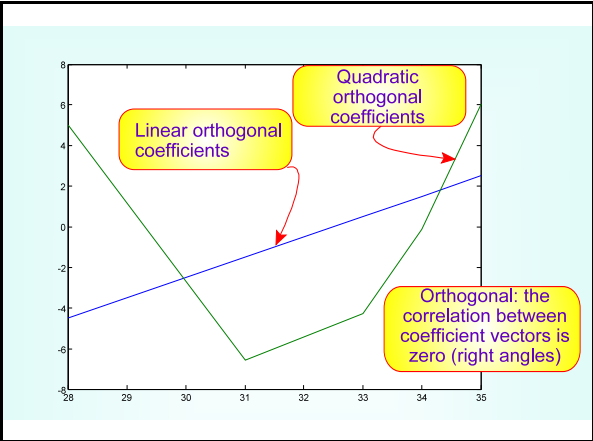
1	5/2	1214/201
1	-3/2	-1318/201
1	1/2	-856/201
1	3/2	-22/201
1	-9/2	1004/201
1	3/2	-22/201

Order-2 polynomial. Is there a quadratic or curved effect?

Slide 24 Matlab, Statbox orthpoly.m

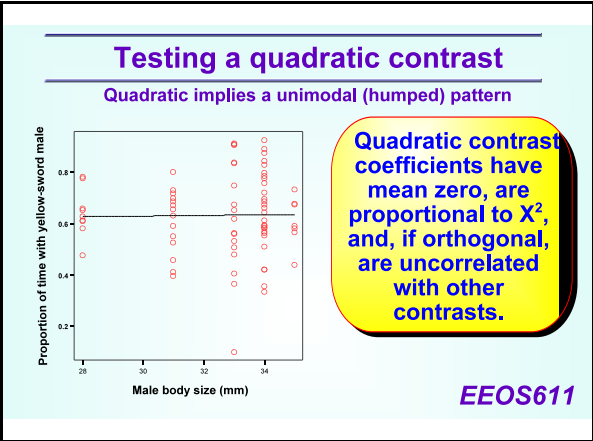
NOTES:

Slide 25



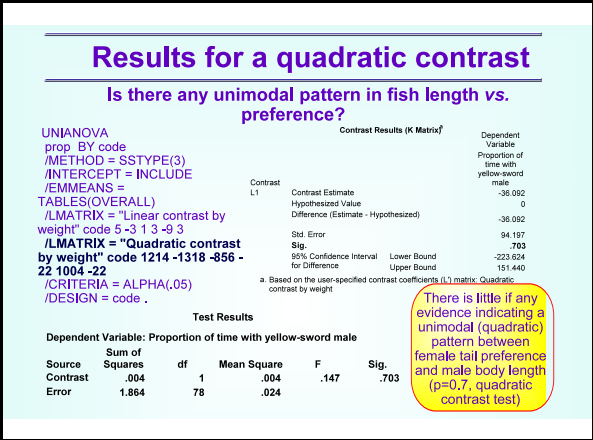
NOTES:

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

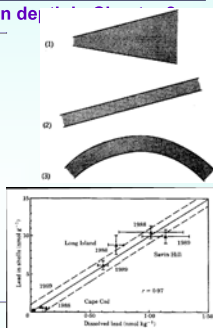
Slide 28 When will environmental scientists need to consider linear contrasts?

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-m² Ted Young modified Van Veen grabs
- Species richness measured with Fisher's α

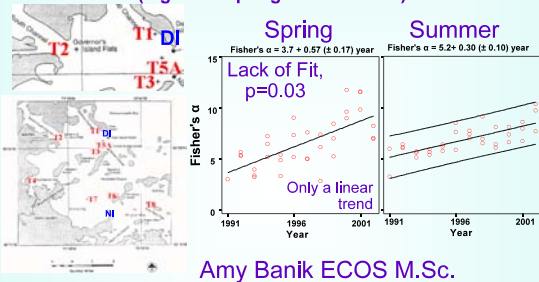


Slide 31 Lack of Fit & Boston Harbor soft-bottom benthic diversity

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-8), also shown in Draper & Smith (1998, p. 467). The test was performed iteratively, with the linear term tested first and the lack of fit resulting after the linear term is removed. The quadratic and cubic terms are tested only if there is no significant lack of fit after the first order polynomial is removed from the model. These orthogonal contrasts were calculated automatically by the SAS procedure generated from NVIS Analysis.

		Year									
		1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Summer data	Linear	-6	-1	-5	-3	-1	1	3	5	7	9
	Quadratic	6	2	-1	-3	-4	-4	-3	-1	2	6
	Cubic	-41	14	15	31	12	-12	-31	-35	-14	41
		Year									
		1992	1993	1994	1995	1996	1997	1998	1999	2000	
Spring data	Linear	-4	-3	-2	-1	0	1	2	3	4	
	Quadratic	25	7	-5	-17	-20	-15	-8	7	25	
	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:

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

a) Fisher's α in Spring at T1					
Source of Variation	SS	df	MS	F	Sig.
Among Years	89.51	8	11.19	4.93 (10%, 8)	0.002
Regression	43.68	1	43.68		
Regression Error	87.21	25	3.49		
Lack of fit	46.43	1	46.43	2.92 (10%, 8)	0.031
Pure Error	40.84	18	2.27		
Linear trend	43.68	1	43.68	15.99	0.0004
Quadratic trend	9.98	1	9.98	4.40	0.0503
Cubic trend	1.21	1	1.21	0.20	0.6563

Slide 34

NOTES:

Conclusion on lack of fit test of Boston Harbor

- There is strong evidence that species richness (as measured by Fisher's α) is increasing in Spring samples [ANOVA linear contrast ($F_{1,18} = 19$, $p < 0.001$)]
- There was significant lack of fit in the OLS regression indicating perhaps non-linear patterns in year-to-year changes in species richness

EEOS611

Slide 35 Conclusion on lack of fit test of Boston Harbor

NOTES:

Estimates of effect size

Sleuth 2e p 152, The mean percentage is 62.4%...

Estimates of effect size, available in GLM/Univariate. Needed to estimate effect across fish pairs in Case 6.2 (62.4 ± 3.4% preferred yellow tails)

2. Grand Mean

Dependent Variable: Proportion of time with yellow-sword male

95% Confidence Interval

Mean	Std. Error	Lower Bound	Upper Bound
.624	.017	.590	.658

Slide 36 Estimates of effect size

NOTES:

<div data-bbox="276 170 747 205" data-label="Section-Header"> <h3>More on Simultaneous Inferences</h3> </div> <div data-bbox="422 214 581 237" data-label="Section-Header"> <h4>Confidence limits</h4> </div> <ul data-bbox="256 247 743 468" style="list-style-type: none"> • 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 <div data-bbox="656 510 779 539" data-label="Text"> <p>EEOS611</p> </div>	<div data-bbox="815 132 1412 168" data-label="Section-Header"> <h3>Slide 37 More on Simultaneous Inferences</h3> </div> <div data-bbox="815 258 938 291" data-label="Text"> <p>NOTES:</p> </div>
<div data-bbox="303 657 714 695" data-label="Section-Header"> <h3>Multiple comparisons (1 of 2)</h3> </div> <div data-bbox="303 699 717 726" data-label="Text"> <p>Interval half width = Multiplier x Standard error</p> </div> <ul data-bbox="256 735 703 984" style="list-style-type: none"> • LSD (Least Significant Difference): Student's t with pooled standard error — no protection against multiple hypothesis testing • F-protected Inference <ul style="list-style-type: none"> ▸ 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 <ul style="list-style-type: none"> ▸ Generalization of Tukey's HSD (Honestly Significant Difference) for unequal sample sizes ▸ Games-Howell more robust to unequal variance <div data-bbox="656 997 779 1024" data-label="Text"> <p>EEOS611</p> </div>	<div data-bbox="815 623 1352 659" data-label="Section-Header"> <h3>Slide 38 Multiple comparisons (1 of 2)</h3> </div> <div data-bbox="815 745 938 779" data-label="Text"> <p>NOTES:</p> </div>
<div data-bbox="303 1148 714 1184" data-label="Section-Header"> <h3>Multiple comparisons (2 of 2)</h3> </div> <div data-bbox="303 1188 717 1215" data-label="Text"> <p>Interval half width = Multiplier x Standard error</p> </div> <ul data-bbox="256 1222 740 1512" style="list-style-type: none"> • Bonferroni, based on the number of comparisons (α/possible tests) <ul style="list-style-type: none"> ▸ A conservative test (most often applied a posteriori test in drug trials for unplanned comparisons) ▸ Test $\alpha = \text{Experiment-wise } \alpha/k$, where k is the number of tests <ul style="list-style-type: none"> ▪ This approximation provides a remarkably accurate estimate of ▪ Experiment-wise alpha: $\alpha_{exp} = 1 - (1 - \alpha_{ind})^k$, where k is the number of tests ▪ For example, 20 groups being tested 2 at a time <ul style="list-style-type: none"> ◦ 20 Choose 2 tests = 190 ◦ Experiment-wise $\alpha = 1 - (1 - 0.05)^{190}$ ◦ Experiment-wise $\alpha = 0.99904$ ◦ But $0.04877683466514 = 1 - (1 - 0.05/190)^{190}$ • 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 <ul style="list-style-type: none"> ▸ Newman-Keuls, SNK, Student-Newman-Keuls; based on studentized range, more powerful (less conservative) than Tukey-Kramer ▸ Duncan's multiple range ▸ Dunnett's, where there is a control group ▸ Dunn's for non-parametric a posteriori contrasts <div data-bbox="656 1484 779 1514" data-label="Text"> <p>EEOS611</p> </div>	<div data-bbox="815 1110 1352 1148" data-label="Section-Header"> <h3>Slide 39 Multiple comparisons (2 of 2)</h3> </div> <div data-bbox="815 1234 938 1266" data-label="Text"> <p>NOTES:</p> </div>

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	1	2	
Tukey HSD ^a				
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	5.9	
Sig.		23	12	
Scheffé ^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	5.9	
Sig.		27		

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. Adjusted R-Square = .99.

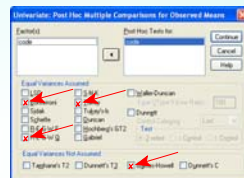
Slide 41 SPSS output from GLM

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: 86) 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:

Slide 43 Quinn & Keough review of multiple comparison tests

- NOTES:

Slide 44 Ryan's test: REGW

NOTES:

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 45 Display 6.7

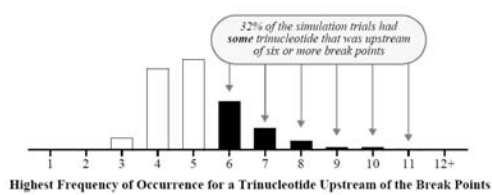
2,436 mononucleotides along a DNA molecule. All 40 occurrences of the trinucleotide TGG appear in bold face. Eleven breaks occurred in the string, at the positions indicated by dashes.

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

NOTES:

Display 6.8

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



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 $\sqrt{\text{Within groups MS}}$ 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:

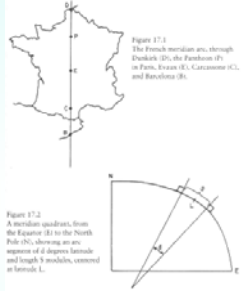
<div data-bbox="328 170 690 207" data-label="Section-Header"> <h3>Conclusions to Chapter 6</h3> </div> <div data-bbox="469 214 526 237" data-label="Text"> <p>3 of 4</p> </div> <div data-bbox="250 228 781 539" data-label="List-Group"> <ul style="list-style-type: none"> • Linear contrasts <ul style="list-style-type: none"> ▸ E.g., Avg (A,B) vs. Avg (C,D,E) ▸ Only Scheffé procedure should be used • Bonferroni <ul style="list-style-type: none"> ▸ 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 <ul style="list-style-type: none"> ■ 20 Choose 2 tests = 190 ■ Experiment-wise $\alpha = 1 - (1 - .05)^{190}$ ■ Experiment-wise $\alpha = 0.99994$ ■ But $0.04877683466514 = 1 - (1 - 0.05/190)^{190}$ <p>EEOS611</p> </div>	<div data-bbox="815 132 1299 172" data-label="Section-Header"> <h3>Slide 49 Conclusions to Chapter 6</h3> </div> <div data-bbox="815 256 941 291" data-label="Text"> <p>NOTES:</p> </div>
<div data-bbox="328 655 690 693" data-label="Section-Header"> <h3>Conclusions to Chapter 6</h3> </div> <div data-bbox="469 701 526 724" data-label="Text"> <p>4 of 4</p> </div> <div data-bbox="250 714 734 1033" data-label="List-Group"> <ul style="list-style-type: none"> • Tukey-Kramer <ul style="list-style-type: none"> ▸ 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 Dunnett's test (only $n-1$ comparisons, not nC_2) • More powerful tests <ul style="list-style-type: none"> ▸ SNK: recommended by Underwood ▸ Ryan's test (REGWF), recommended by Quinn & Keough (in addition to Tukey-Kramer) <ul style="list-style-type: none"> ■ Ryan's test not suitable for unequal group sizes (SPSS) </div>	<div data-bbox="815 621 1299 659" data-label="Section-Header"> <h3>Slide 50 Conclusions to Chapter 6</h3> </div> <div data-bbox="815 743 941 779" data-label="Text"> <p>NOTES:</p> </div>
<div data-bbox="263 1178 734 1289" data-label="Section-Header"> <h2>Chapter 7 Simple Linear Regression: a model for the mean</h2> </div> <div data-bbox="269 1299 758 1352" data-label="Text"> <p>Simple regression = ordinary least squares (OLS) regression, Model I regression</p> </div> <div data-bbox="654 1482 781 1514" data-label="Text"> <p>EEOS611</p> </div>	<div data-bbox="815 1110 1292 1184" data-label="Section-Header"> <h3>Slide 51 Chapter 7 Simple Linear Regression: a model for the mean</h3> </div> <div data-bbox="815 1268 941 1304" data-label="Text"> <p>NOTES:</p> </div>

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:

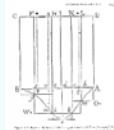
Probability & least squares landmarks

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



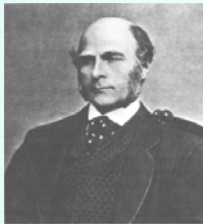
The normal (Gaussian) curve

Slide 53 Probability & least squares landmarks

NOTES:

Sir Francis Galton

Inventor of the Quincunx & 'regression'



Francis Galton (1822-1911)



Figure 5.4. This original quincunx, apparently made for Galton in 1871 by James D. Butler, although it was had on going in the top through which the shot could be poured, the top is now sealed with the shot inside. The glass has become cloudy with lead dust over the years. The caption in Galton's handwriting reads:

Slide 54 Sir Francis Galton

NOTES:

Galton (1885) on filial height

Girls x 1.08

Table 3.1 Galton's 1885 cross-tabulation of 928 adult children born of 205 marriages, by their heights and their midparents' height.

Height of the mid-parent in inches	Height of the child																Total no. of children	Total no. of mid-parents	Medians
	61.7	62.2	63.2	64.2	65.2	66.2	67.2	68.2	70.2	71.2	72.2	73.2	74.2	75.2	76.2	77.2			
> 72.0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
72.5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
71.5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
70.5	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
69.5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
68.5	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
67.5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
66.5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
65.5	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
64.5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
63.5	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
62.5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
61.5	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
60.5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Total	5	7	32	59	48	117	189	129	107	99	81	41	17	11	928	205	—	—	—
Medians	—	—	66.5	67.8	67.9	67.7	67.9	68.5	68.6	69.0	69.0	70.0	—	—	—	—	—	—	—

Source: Galton (1885).

Note: All female heights were multiplied by 1.08 before tabulation. Galton added an explanatory footnote to the table: "In calculating the Medians, the extreme have been taken as corresponding to the middle of the squares in which they stand. The squares with the headings 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, 74.2, 75.2, 76.2, 77.2, are, as far as the observations are unequally distributed between 62 and 65, 65 and 68, 68 and 71, 71 and 74, 74 and 77, there being a strong bias in favor of central squares. After careful consideration, I concluded that the headings, as printed, best satisfied the conditions. This irregularity was not apparent in the case of the Male Parents. Galton equalized these data in 1885, when they were referred to the B.P.P. (see General Family Recollections for the account that the first row must be normal adult children, except here for sons of parents, but he claimed that 'the bottom line, which looks suspicious, is correct' (p. 298).

Slide 55 Galton (1885) on filial height

NOTES:

Galton's regression to mediocrity

Campbell & Kenny, 1999 p. 2

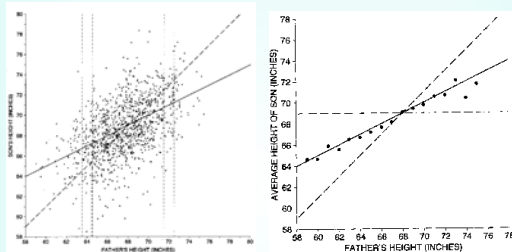
- Galton (1822-1911) measured the heights of 928 parents and children
- Multiplied female heights by a scaling factor
- Galton (1886)
 - Tall parents tended to have tall children, but the children of the tallest parents were, on average, not quite as tall as the parents. Nor, were the children of the shortest parents as short as their parents
 - "tall 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)

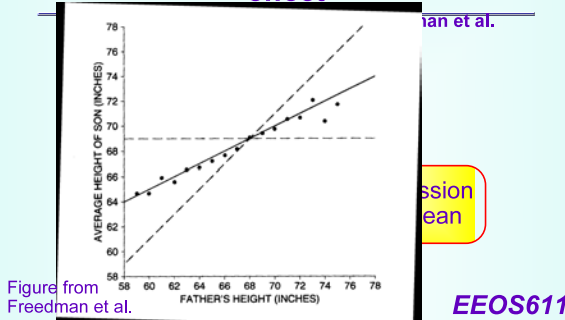


Pearson's data on 1078 fathers & sons at maturity

Slide 57 Galton's regression to mediocrity

NOTES:

RTM (Regression to the mean) effect



Slide 58 RTM (Regression to the mean) effect

NOTES:

Correlation & regression history

Galton (1886) introduced the correlation coefficient

- The method of least squares had been described completely by Legendre (1805) and Gauss in 1810; Gauss introduced the normal equations to solve least squares problems in 1822

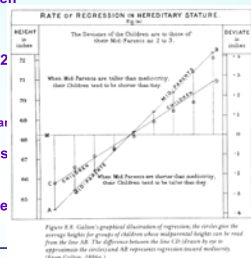
• Galton's regression

- did not formally use the method of least squares!
- described the relation between correlation and 'regression to the mean' in 1888

- Method of least squares and Gauss's normal equations first used for 'regression' by Yule (1897)

- The term regression was later applied to any fit of continuous variables by the method of least squares

My history from Bell (1937), Stigler & Campbell & Kenny (1999)



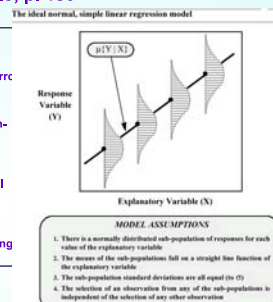
Slide 59 Correlation & regression history

NOTES:

OLS regression invented by Yule

Display 7.5, p. 180

- Linearity
- Constant variance
 - Estimators still unbiased, but p values in error
- Independence of errors
 - Cluster effects,
 - serial correlation: 1 type tested with Durbin-Watson tests
- Normality of errors
 - (not of explanatory variables)
 - Estimators (e.g., of slope & Y intercept) still unbiased if normality assumption violated
 - P values robust to violations of normality
- [X variable measured without error]
 - This is an assumption involved in minimizing residuals



Slide 60 OLS regression invented by Yule

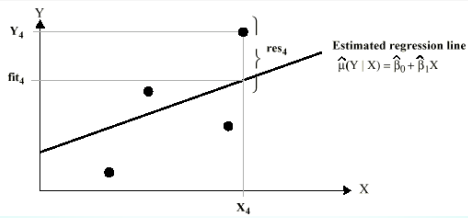
NOTES:

OLS, find parameters to minimize sum of squared residuals

Display 7.6 n. 181

Display 7.6

Illustration of the residual and fitted value for observation (X_4, Y_4) in a hypothetical data set of size 4

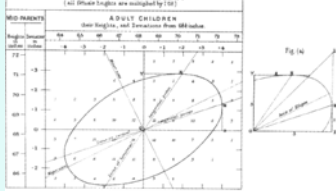


Slide 61 OLS, find parameters to minimize sum of squared residuals

NOTES:

Plot of interpolated estimates

Galton did not use least squares to fit filial height line



Model II regress on See Legendre & Legendre (1998), Quinn & Keough (2002), Not really available in SPSS

Method now called Model II regress on and is appropriate when both X & Y measured with error

Slide 62 Plot of interpolated estimates

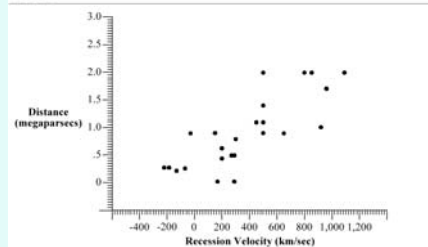
NOTES:

Display 7.1

Hubble (1929)

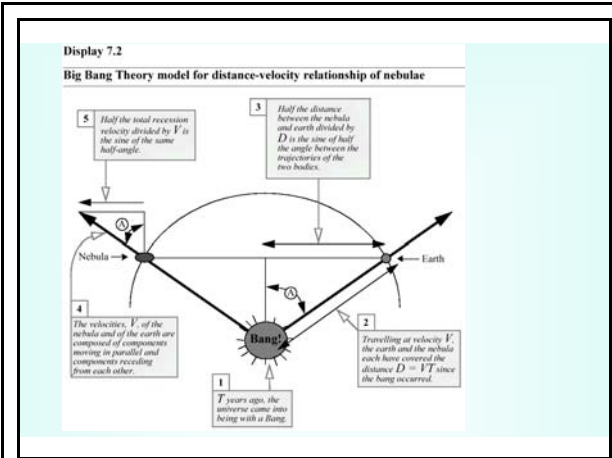
Display 7.1

Scatterplot of measured distance versus velocity for 24 extra-galactic nebulae



Slide 63 Display 7.1

NOTES:



Slide 64 Display 7.2

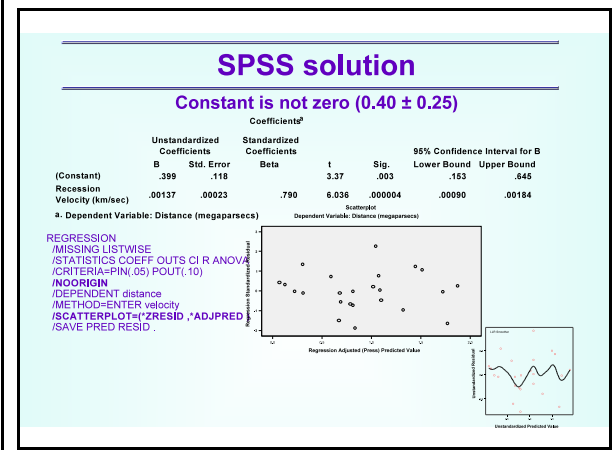
NOTES:

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:

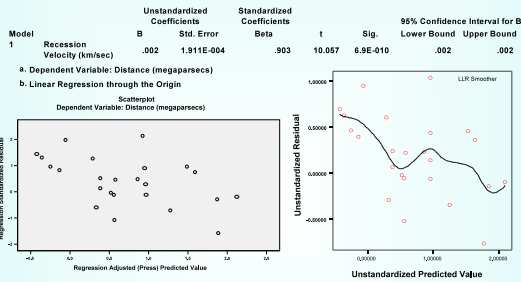


Slide 66 SPSS solution

NOTES:

Forcing regression through origin

Generally, not advisable. Poor fit to slope.



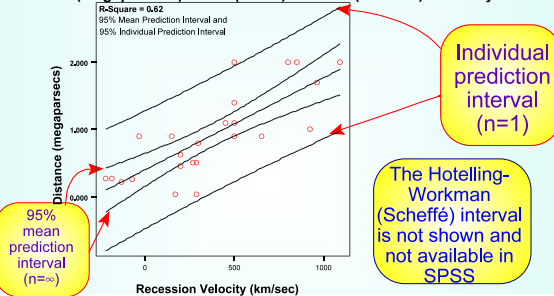
Slide 67 Forcing regression through origin

NOTES:

Case 7.1: Scatterplot

See Sleuth Display 7.11, p 189

Distance (megaparsecs) = $0.40 (\pm 0.25) + 0.0014 (\pm 0.0005) * \text{velocity}$



Slide 68 Case 7.1: Scatterplot

NOTES:

Case 7.2 Meat processing & pH

Postmortem muscle pH and time since slaughter
How many hours (with CI's to obtain a pH of 6)?

Display 7.3

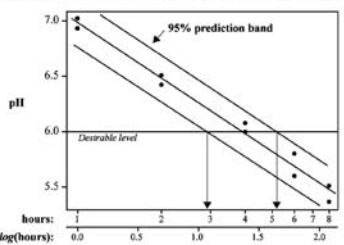
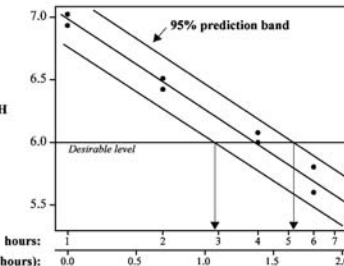
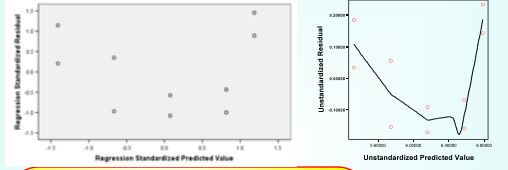
pH of 10 steer carcasses measured at 5 different times after slaughter

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

EEOS611

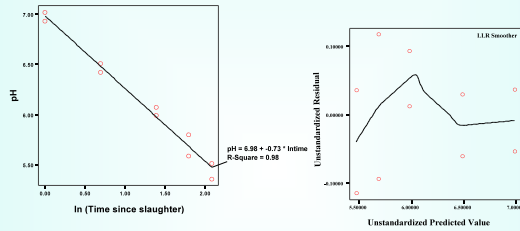
Slide 69 Case 7.2 Meat processing & pH

NOTES:

<div data-bbox="251 168 747 210"><h3>Conclusion from Case 7.2</h3></div> <div data-bbox="235 214 776 535"><p>pH of 6.0 at 3.9 h. 95% calibration interval: 2.94 & 5.1 h</p><p>Display 7.4</p><p>Meat processing data with estimated regression line (from the simple linear regression of pH on log time after slaughter) and a 95% prediction band</p><p>The appropriate 95% CI must be used; for individual measurements, the individual prediction band is appropriate</p><p>Fiducial Limits</p></div>	<div data-bbox="828 136 1412 178"><h3>Slide 70 Conclusion from Case 7.2</h3></div> <div data-bbox="828 252 1412 569"><p>NOTES:</p></div>
<div data-bbox="251 630 747 672"><h3>How to design a standard curve</h3></div> <div data-bbox="235 676 776 1050"><p>From Draper & Smith (1981) Chapter 3: assigned for</p><p>Display 7.4</p><p>Meat processing data with estimated regression line (from the simple linear regression of pH on log time after slaughter) and a 95% prediction band</p></div>	<div data-bbox="828 621 1412 663"><h3>Slide 71 How to design a standard curve</h3></div> <div data-bbox="828 735 1412 1054"><p>NOTES:</p></div>
<div data-bbox="251 1144 747 1186"><h3>Residual plot with smoother</h3></div> <div data-bbox="235 1190 776 1522"><p>Save predicted and residual values: untransformed</p><p>Plot the Residuals (standardized or unstandardized) vs Predicted values and examine the plot for patterns. Save residuals and use a smoother (Lowess fit for patterns, but LOWESS not available in SPSS)</p></div>	<div data-bbox="828 1106 1412 1148"><h3>Slide 72 Residual plot with smoother</h3></div> <div data-bbox="828 1218 1412 1539"><p>NOTES:</p></div>

Log transform 'time since slaughter'

SPSS LLR smoother plot available in scatterplot

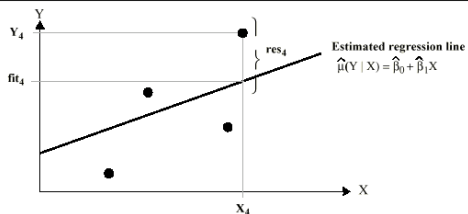


Slide 73 Log transform 'time since slaughter'

NOTES:

Display 7.6

Illustration of the residual and fitted value for observation (X_4, Y_4) in a hypothetical data set of size 4



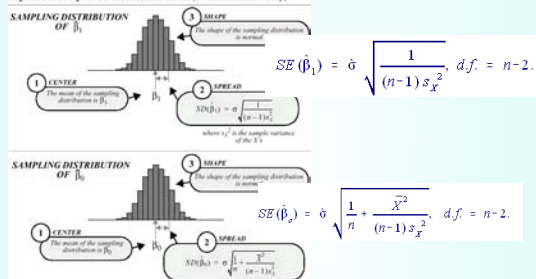
Slide 74 Display 7.6

NOTES:

Standard error of b_0 & b_1

Display 7.7

Facts about the sampling distributions of the least squares estimates of slope and intercept in the ideal normal model (from statistical theory)



Slide 75 Standard error of b_0 & b_1

NOTES:

Display 7.9

Regression parameter estimates for the Big Bang study

Variable	Coefficient	Standard Error	t-Statistic	p-Value
Constant	0.3991	.1185	3.369	.0028
Velocity	.001373	.000227	6.036	.0000045

Estimate of $\sigma^2 = 0.4050$ (22 d.f.)

Ratio of coefficient estimates to their standard errors

For hypotheses that the coefficients = 0

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	5.975	1	5.975	36.4	4.5e-6
Residual	3.608	22	.164		
Total	9.583	23			

$\sigma^2 = \sqrt{\text{Res M}}$
 $= \sqrt{0.164}$
 $= 0.4050$

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients		t	Sig.	95% Confidence Interval for B		
	B	Std. Error	Beta				Lower Bound	Upper Bound	
1	(Constant)	.399	.118		3.369	.003	.163	.645	
	Velocity (km/sec)	.0014	.0002		.790	6.036	<1e-8	.0009	.0018

a. Dependent Variable: Distance (megaparsecs)

$\sigma^2 = 0.4050$

$$\sigma^2 = \sqrt{\text{Res MS}} \\ = \sqrt{0.164} \\ = 0.4050$$

$$\sqrt{36.4} = 6.036$$

Slide 76 Display 7.9

NOTES:

Display 7.10

95% Confidence interval for the estimated mean pH of steers 4 hours after slaughter (from the estimated regression of pH on $\log(\text{time})$ after slaughter for the meat processing data)

$$\hat{\mu}\{Y | 1.386\} = 6.9836 - 0.7257 \times 1.386 = 5.98$$

$$SE[\hat{\mu}\{Y | 1.386\}] = 0.08226 \sqrt{\frac{1}{10} + \frac{(1.386 - 1.190)^2}{9(0.6244)}}$$

$$= 0.0269$$

Upper limit: $5.98 + 2.306 \times 0.0269 = 6.04$
 Lower limit: $5.98 - 2.306 \times 0.0269 = 5.92$

Sleuth computer trick subtract 4 from each X value, look at Y intercept and CI for Y intercept, but there is a better way in SPSS

Slide 77 Display 7.10

NOTES:

Predicting Y, given X in SPSS

Enter X as an additional case, 5.5 hours and 4 hours

Case#002.sav - SPSS Data Editor

	time	pH	Intercept	PRED_1	BES_1	LMCI_1	UMCI_1	LCI_1	UCI_1
1	1.0	7.02	.00	7.0	.04	6.87	7.10	6.76	7.20
2	1.0	6.93	.00	7.0	-.06	6.87	7.10	6.76	7.20
3	2.0	6.42	.89	6.5	-.06	6.41	6.65	6.20	6.68
4	2.0	6.51	.89	6.5	.03	6.41	6.65	6.20	6.68
5	4.0	6.07	1.99	6.0	.09	5.92	6.04	5.78	6.18
6	4.0	5.99	1.99	6.0	.01	5.92	6.04	5.78	6.18
7	6.0	5.59	1.79	5.7	-.09	5.61	5.76	5.48	5.88
8	6.0	5.80	1.79	5.7	.12	5.61	5.76	5.48	5.89
9	8.0	5.51	2.06	5.5	.04	5.30	5.57	5.26	5.69
10	8.0	5.36	2.06	5.5	-.11	5.30	5.57	5.26	5.69
11	5.5		1.70	5.7		5.67	5.82	5.54	5.95
12	4.0		1.39	6.0		5.92	6.04	5.78	6.18

Leave Response variable blank

Sleuth trick subtract 5.5 or 4 from each value of X and use the upper & lower 95% CI for the Y intercept

Slide 78 Predicting Y, given X in SPSS

NOTES:

≥3 confidence intervals

Individual ($n=1$), mean ($n=\infty$), for the line (Scheffé, $n=\infty$)

Display 7.11

The 95% confidence band on the population regression line, the 95% confidence interval band for single mean estimates, and a 95% prediction interval band for the Big Bang example

Distance (megaparsecs)

Recession Velocity (km/sec)

Estimated Regression Line

95% confidence band for estimated means

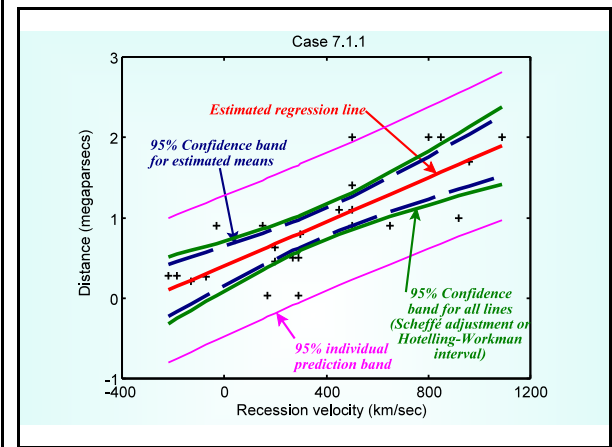
95% confidence band for estimated means, with Scheffé adjustment

95% Prediction Interval for Unknown Distance

CI for single observations
CI for mean & line (Scheffé adjustment)
assumes very large (∞) sample size

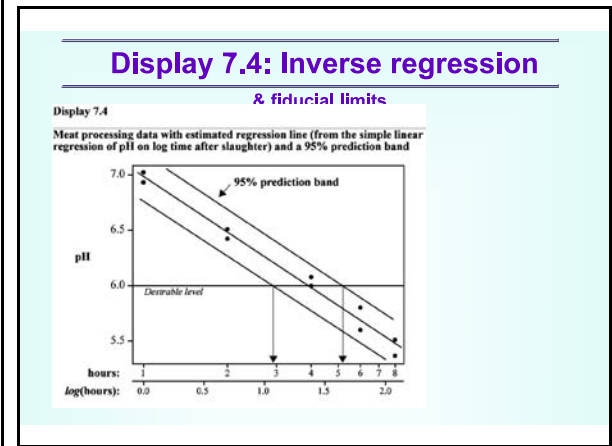
Slide 79 ≥3 confidence intervals

NOTES:



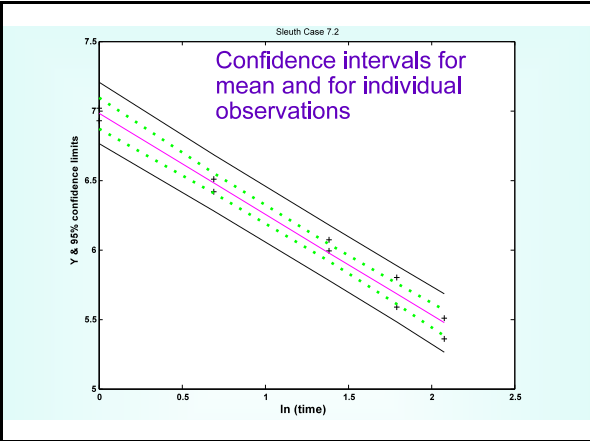
Slide 80

NOTES:



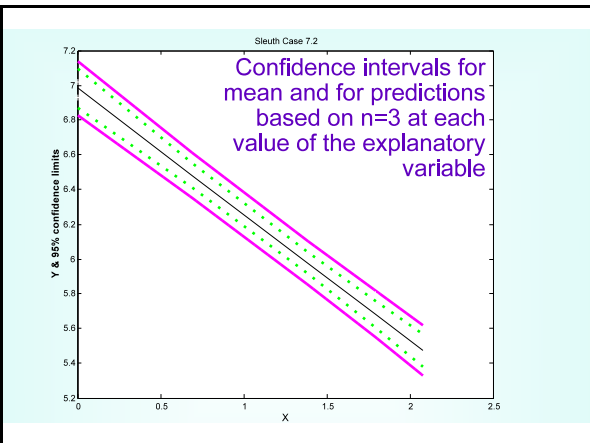
Slide 81 Display 7.4: Inverse regression

NOTES:



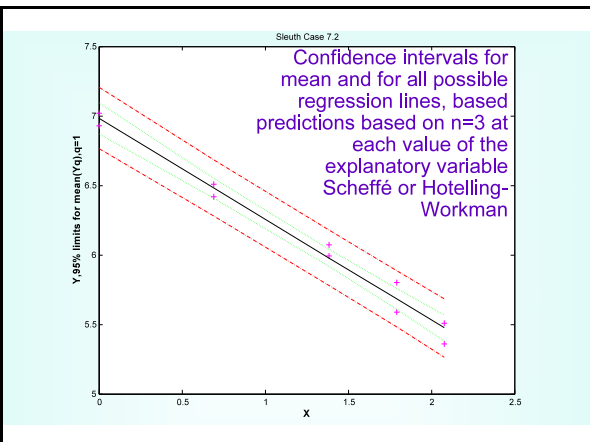
Slide 82

NOTES:



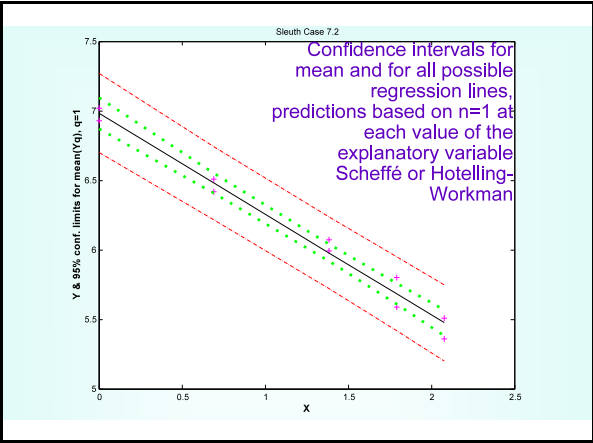
Slide 83

NOTES:



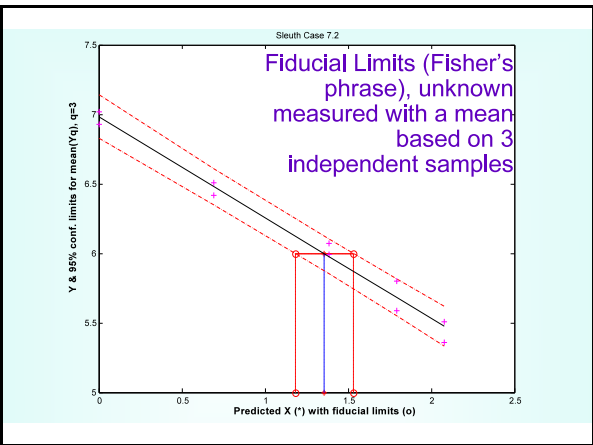
Slide 84

NOTES:



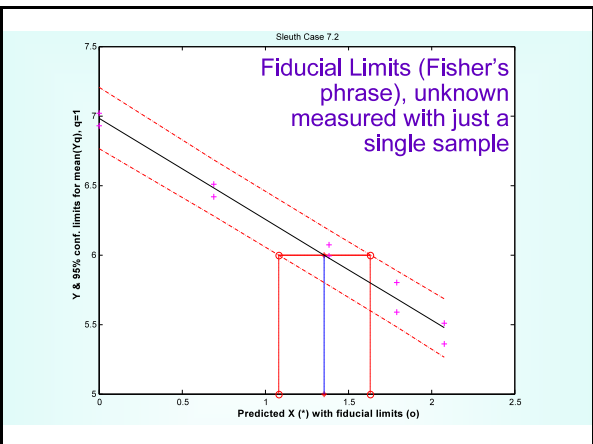
Slide 85

NOTES:



Slide 86

NOTES:

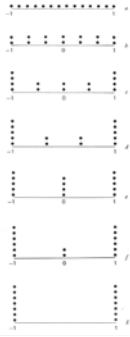


Slide 87

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) $Sd.(b_1)/s_r$:	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.

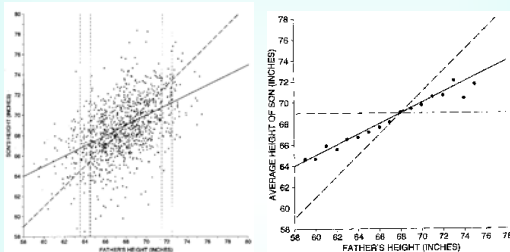
EEOS611

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)



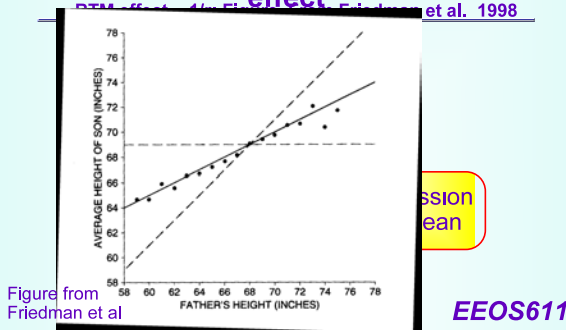
Pearson's data on 1078 fathers & sons at maturity

Slide 90 Galton's regression to mediocrity

NOTES:

RTM (Regression to the mean) effect

RTM Effect 4th Edition Friedman et al. 1998

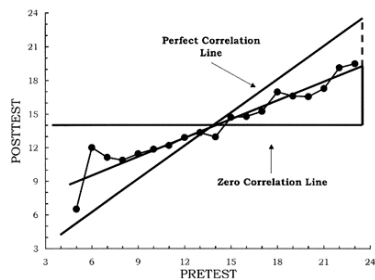


Slide 91 RTM (Regression to the mean) effect

NOTES:

The regression effect or regression artifact

Campbell & Mann 1990

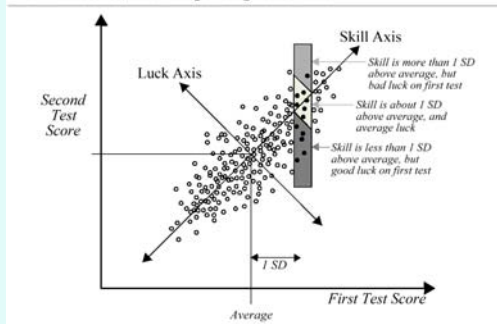


Slide 92 The regression effect or regression artifact

NOTES:

Display 7.13

Test-retest scores, illustrating the regression effect



Slide 93 Display 7.13

NOTES:

The Galton Squeeze diagram

Don't use improvement to fire employees or award \$

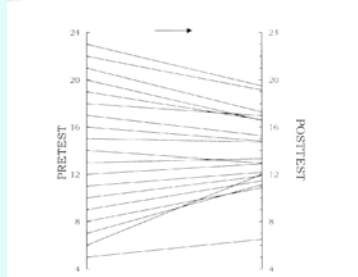


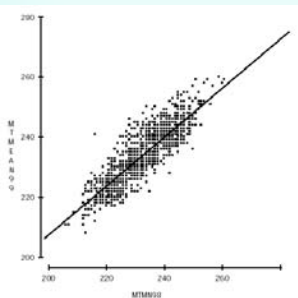
FIGURE 1.8. Galton squeeze diagram for the data set with 500 cases using pretest to predict posttest.

Slide 94 The Galton Squeeze diagram

NOTES:

4th Grade MCAS scores

1998 vs. 1999



Lake Woebeguaranteed:
Misuse of Test Scores in
Massachusetts, Part I

Walt Haney¹
Boston College

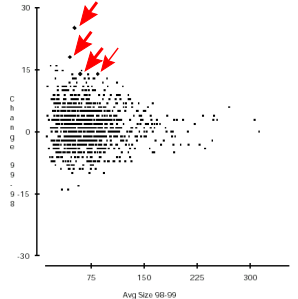
Citation: Haney, W. (2002, May 6). Lake
Woebeguaranteed: Misuse of test scores in
Massachusetts, Part I. *Education Policy Analysis
Archives*, 10(24). Retrieved [date] from
<http://epaa.asu.edu/epaa/v10n24/>

Slide 95 4th Grade MCAS scores

NOTES:

4th Grade Change in MCAS scores

1998 vs. 1999

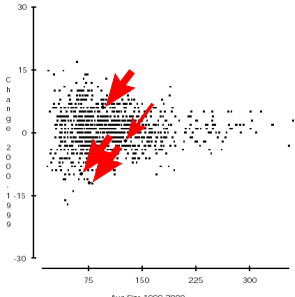


The mean **change** in
math scores from 1998
to 1999 for all schools
with 4th grade classes
are plotted vs. average
class size. Four
principals heading the
schools with the greatest
improvement (↗) were
given \$10,000 cash
awards from a private
foundation grant.

EEOS611

Slide 96 4th Grade Change in MCAS scores

NOTES:

<div data-bbox="207 134 792 562"> <h3>4th Grade Change in MCAS scores</h3> <p>1998 vs. 1999, from Walt Haney BC</p>  <p>3 of the 4 award-winning schools declined in performance in their post-award year</p> <p>Why?</p> <p>EEOS611</p> </div>	<div data-bbox="824 134 1396 210"> <h3>Slide 97 4th Grade Change in MCAS scores</h3> </div> <div data-bbox="824 294 1396 336"> <p>NOTES:</p> </div>
<div data-bbox="207 663 792 1092"> <h3>The regression artifact</h3> <ul style="list-style-type: none"> What major statistical principal must be considered when analyzing test and retest data of this sort? [A 2003 midterm question] Two related statistical problems <ul style="list-style-type: none"> 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 </div>	<div data-bbox="824 663 1396 705"> <h3>Slide 98 The regression artifact</h3> </div> <div data-bbox="824 789 1396 831"> <p>NOTES:</p> </div>
<div data-bbox="207 1146 792 1575"> <h3>Luck or skill in awarding MCAS winners (1 of 3)</h3> <ul style="list-style-type: none"> 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. <ul style="list-style-type: none"> 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. </div>	<div data-bbox="824 1146 1396 1222"> <h3>Slide 99 Luck or skill in awarding MCAS winners (1 of 3)</h3> </div> <div data-bbox="824 1306 1396 1348"> <p>NOTES:</p> </div>

<div data-bbox="292 163 734 231"> Luck or skill in awarding MCAS winners (2 of 3) </div> <div data-bbox="254 247 742 499"> <ul style="list-style-type: none"> • 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. <ul style="list-style-type: none"> ▪ 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. </div> <div data-bbox="657 510 779 539"> EEOS611 </div>	<div data-bbox="816 134 1422 210"> Slide 100 Luck or skill in awarding MCAS winners (2 of 3) </div> <div data-bbox="816 296 940 327"> NOTES: </div>
<div data-bbox="292 688 734 756"> Luck or skill in awarding MCAS winners (3 of 3) </div> <div data-bbox="254 772 716 1008"> <ul style="list-style-type: none"> • 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 </div> <div data-bbox="657 1035 779 1064"> EEOS611 </div>	<div data-bbox="816 663 1422 739"> Slide 101 Luck or skill in awarding MCAS winners (3 of 3) </div> <div data-bbox="816 823 940 854"> NOTES: </div>