Instructor Solutions Manual

The Statistical Sleuth
A Course in Methods of Data Analysis

THIRD EDITION

Fred L. Ramsey
Oregon State University

Daniel W. Schafer
Oregon State University

BROOKS/COLE
CENGAGE Learning

Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

BROOKS/COLE
CENGAGE Learning

For product information and technology assistance, contact us at **Cengage Learning Customer & Sales Support, 1-800-354-9706**

For permission to use material from this text or product, submit all requests online at **www.cengage.com/permissions**
Further permissions questions can be emailed to **permissionrequest@cengage.com**

ISBN-13: 978-1-133-59054-5
ISBN-10: 1-133-59054-3

Brooks/Cole
20 Channel Center Street
Boston, MA 02210
USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at: **www.cengage.com/global**

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

To learn more about Brooks/Cole, visit **www.cengage.com/brookscole**

Purchase any of our products at your local college store or at our preferred online store **www.cengagebrain.com**

NOTE: UNDER NO CIRCUMSTANCES MAY THIS MATERIAL OR ANY PORTION THEREOF BE SOLD, LICENSED, AUCTIONED, OR OTHERWISE REDISTRIBUTED EXCEPT AS MAY BE PERMITTED BY THE LICENSE TERMS HEREIN.

READ IMPORTANT LICENSE INFORMATION

Dear Professor or Other Supplement Recipient:

Cengage Learning has provided you with this product (the “Supplement”) for your review and, to the extent that you adopt the associated textbook for use in connection with your course (the “Course”), you and your students who purchase the textbook may use the Supplement as described below. Cengage Learning has established these use limitations in response to concerns raised by authors, professors, and other users regarding the pedagogical problems stemming from unlimited distribution of Supplements.

Cengage Learning hereby grants you a nontransferable license to use the Supplement in connection with the Course, subject to the following conditions. The Supplement is for your personal, noncommercial use only and may not be reproduced, posted electronically or distributed, except that portions of the Supplement may be provided to your students IN PRINT FORM ONLY in connection with your instruction of the Course, so long as such students are advised that they may not copy or distribute any portion of the Supplement to any third party. You may not sell, license, auction, or otherwise redistribute the Supplement in any form. We ask that you take reasonable steps to protect the Supplement from unauthorized use, reproduction, or distribution. Your use of the Supplement indicates your acceptance of the conditions set forth in this Agreement. If you do not accept these conditions, you must return the Supplement unused within 30 days of receipt.

All rights (including without limitation, copyrights, patents, and trade secrets) in the Supplement are and will remain the sole and exclusive property of Cengage Learning and/or its licensors. The Supplement is furnished by Cengage Learning on an “as is” basis without any warranties, express or implied. This Agreement will be governed by and construed pursuant to the laws of the State of New York, without regard to such State’s conflict of law rules.

Thank you for your assistance in helping to safeguard the integrity of the content contained in this Supplement. We trust you find the Supplement a useful teaching tool.

Printed in the United States of America
1 2 3 4 5 6 7 16 15 14 13 12

Contents

Chapter 1:	Drawing Statistical Conclusions	1
Chapter 2:	Inference Using *t*-Distributions	5
Chapter 3:	A Closer Look at Assumptions	9
Chapter 4:	Alternatives to the *t*-Tools	15
Chapter 5:	Comparisons Among Several Samples	20
Chapter 6:	Linear Combinations and Multiple Comparisons of Means	25
Chapter 7:	Simple Linear Regression: A Model for the Mean	33
Chapter 8:	A Closer Look at Assumptions for Simple Linear Regression	39
Chapter 9:	Multiple Regression	47
Chapter 10:	Inferential Tools for Multiple Regression	55
Chapter 11:	Model Checking and Refinement	65
Chapter 12:	Model Selection with Large Numbers of Explanatory Variables	80
Chapter 13:	The Analysis of Variance for Two-way Classification	85
Chapter 14:	Multifactor Studies	91
Chapter 15:	Adjustment for Serial Correlation	103
Chapter 16:	Repeated Measures	116
Chapter 17:	Exploratory Tools for Summarizing Multivariate Responses	123
Chapter 18:	Comparisons of Proportions or Odds	133
Chapter 19:	More Tools for Tables of Counts	135
Chapter 20:	Logistic Regression for Binary Response Variables	137
Chapter 21:	Logistic Regression for Binomial Counts	143
Chapter 22:	Log-Linear Regression for Poisson Counts	150
Chapter 23:	Elements of Research Design	155
Chapter 24:	Factorial Treatment Arrangements and Blocking Designs	158

Preface: To The Instructor

This solutions manual contains answers and sketches of solutions to all "computational exercises" and "data problems" that appear at the ends of the chapters in *The Statistical Sleuth.* The Student Solutions Manual contains identical information, but for selected "computational exercises" only. About half of the computational exercises answers are provided in the Student Solutions Manual.

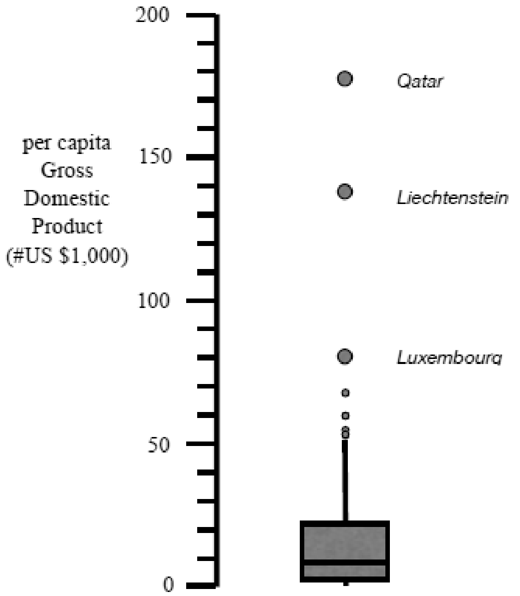
The "data problems" are, of course, important for practical experience at real data analysis and communication of statistical results. Many of these are quite hard—mainly because real data problems can be quite hard. We provide sketches of solutions to the data problems here, but wish to point out that there is often more than one correct approach. We hope that students use the "Statistical Conclusions" sections at the end of each case study in the book as templates for their own wording of results.

We will periodically provide updates and corrections on the web site www.statisticalsleuth.com. There are instructions there for joining our mailing list so that you may receive any updates or news that we believe worthy of broadcasting. You may contact us by e-mail at: ramsey@stat.orst.edu orschafer@stat.orst.edu.

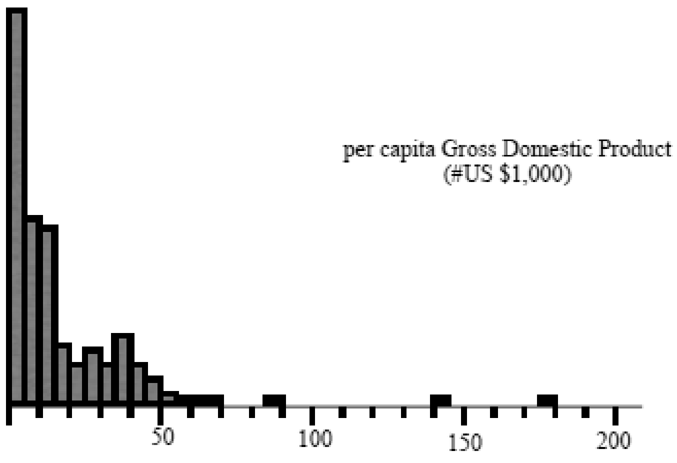
Chapter 1: Drawing Statistical Conclusions

1.16 Gross Deomestic Product (GDP) Per Capita.

a



c



1.17 The difference between averages (A − B) in the observed outcome is 78.00 − 62.67 = +15.33 points. In the list that follows, there are three outcomes (nos. 1, 34, and 35) that have a difference as large or larger in magnitude as the observed difference. The two-sided *p-value* is therefore 3/35 = 0.0857.

Outcome No.	Guide A	A Average	Guide B	B Average	(A − B) Difference
1	53, 64, 68, 71	64.00	77, 82, 85	81.33	−17.33
2	53, 64, 68, 77	65.50	71, 82, 85	79.33	−13.83
3	53, 64, 68, 82	66.75	71, 77, 85	77.67	−10.92
4	53, 64, 68, 85	67.50	71, 77, 82	76.67	−9.17
5	53, 64, 71, 77	66.25	68, 82, 85	78.33	−12.08
6	53, 64, 71, 82	67.50	68, 77, 85	76.67	−9.17
7	53, 64, 71, 85	68.25	68, 77, 82	75.67	−7.42
8	53, 64, 77, 82	69.00	68, 71, 85	74.67	−5.67
9	53, 64, 77, 85	69.75	68, 71, 82	73.67	−3.92
10	53, 64, 82, 85	71.00	68, 71, 77	72.00	−1.00
11	53, 68, 71, 77	67.25	64, 82, 85	77.00	−9.75
12	53, 68, 71, 82	68.50	64, 77, 85	75.33	−6.83
13	53, 68, 71, 85	69.25	64, 77, 82	74.33	−5.08
14	53, 68, 77, 82	70.00	64, 71, 85	73.33	−3.33
15	53, 68, 77, 85	70.75	64, 71, 82	72.33	−1.58
16	53, 68, 82, 85	72.00	64, 71, 77	70.67	+1.33
17	53, 71, 77, 82	70.75	64, 68, 85	72.33	−1.58
18	53, 71, 77, 85	71.50	64, 68, 82	71.33	+0.17
19	53, 71, 82, 85	72.75	64, 68, 77	69.67	+3.08
20	53, 77, 82, 85	74.25	64, 68, 71	67.67	+6.58
21	64, 68, 71, 77	70.00	53, 82, 85	73.33	−3.33
22	64, 68, 71, 82	71.25	53, 77, 85	71.67	−0.42
23	64, 68, 71, 85	72.00	53, 77, 82	70.67	+1.33
24	64, 68, 77, 82	72.75	53, 71, 85	69.67	+3.08
25	64, 68, 77, 85	73.50	53, 71, 82	68.67	+4.83
26	64, 68, 82, 85	74.75	53, 71, 77	67.00	+7.75
27	64, 71, 77, 82	73.50	53, 68, 85	68.67	+4.83
28	64, 71, 77, 85	74.25	53, 68, 82	67.67	+6.58
29	64, 71, 82, 85	75.50	53, 68, 77	66.00	+9.50
30	64, 77, 82, 85	77.00	53, 68, 71	64.00	+13.00
31	68, 71, 77, 82	74.50	53, 64, 85	67.33	+7.17
32	68, 71, 77, 85	75.25	53, 64, 82	66.33	+8.92
33	68, 71, 82, 85	76.50	53, 64, 77	64.67	+11.83
34	68, 77, 82, 85	78.00	53, 64, 71	62.67	+15.33
35	71, 77, 82, 85	78.75	53, 64, 68	61.67	+17.08

1.18 Outcomes will vary with different randomizations. See text Display 1.7

1.19 Coin flips will not divide the subjects in such a way that there is an exact age balance. However, it is impossible to tell prior to the flips which group will have a higher average age.

1.20 The randomization scheme suggested in problem 18 works. So would dealing five red and five black cards after shuffling. Once again it will not guarantee an exact age balance, but the group that gets the higher average is not predictable in advance of the randomization.

1.21 There is no computation involved. This is, however, a sobering exercise.

1.22

Intrinsic ($n = 24$)	Stem	Extrinsic ($n = 23$)
	5	04
	6	1
	7	
	8	
	9	
	10	9
	11	8
900	12	03
6	13	
	14	8
	15	0
6	16	8
52	17	2245
2	18	577
831	19	25
653	20	7
63	21	2
621	22	1
1	23	
30	24	0
	25	
7	26	
	27	
	28	
7	29	
	30	

Legend: | 29 | 7 represents a creativity score of 29.7

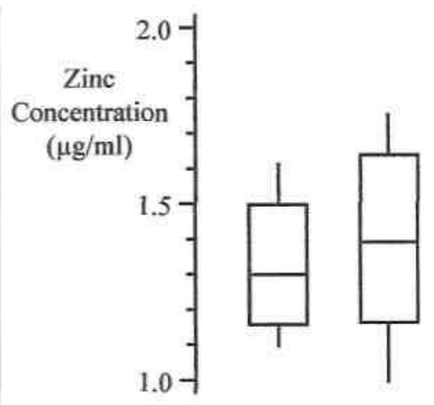
1.23 The box plot should look a bit like the stem and leaf diagram in exercise #22.

1.24 (Int,Ext): Medians are (20.4,17.2); Lower quartiles are (17.35, 12.0); Upper quartiles are (22.4, 19.2); IQRs are (5.05,7.2). There are no extreme points in either group.

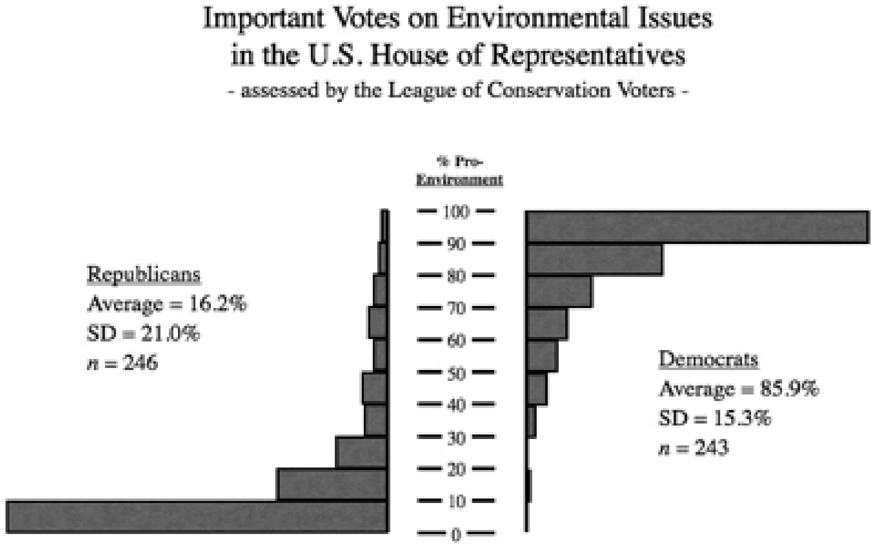
1.25

Group A ($n = 20$)	Stem	Group B ($n = 19$)
	10	03
764210	11	3559
9320	12	0
10	13	349
952	14	77
9320	15	59
2	16	8
	17	1446

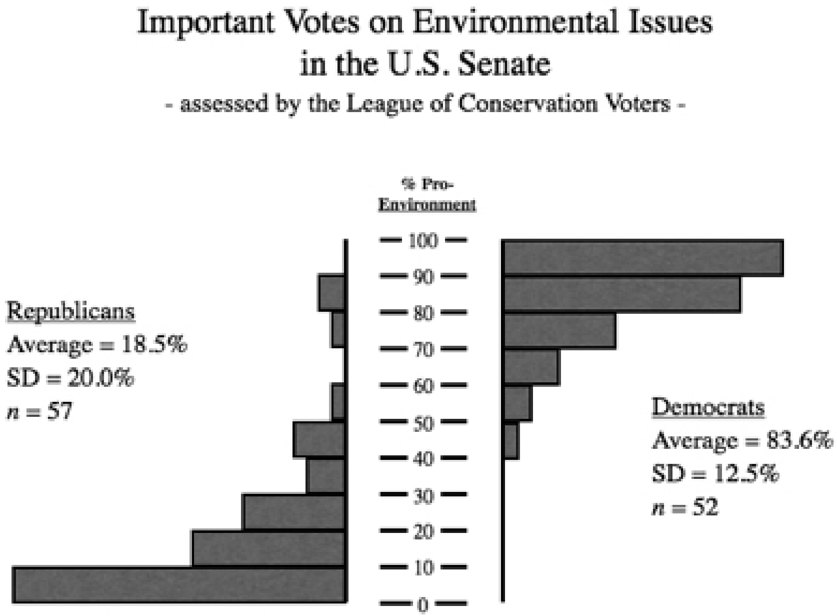
Legend: | 17 | 6 represents 1.76 µg/ml



1.26 Any picture tells the story. There is no need for a statistical test.



1.27 Again, any picture tells the story. There is no need for a statistical test.



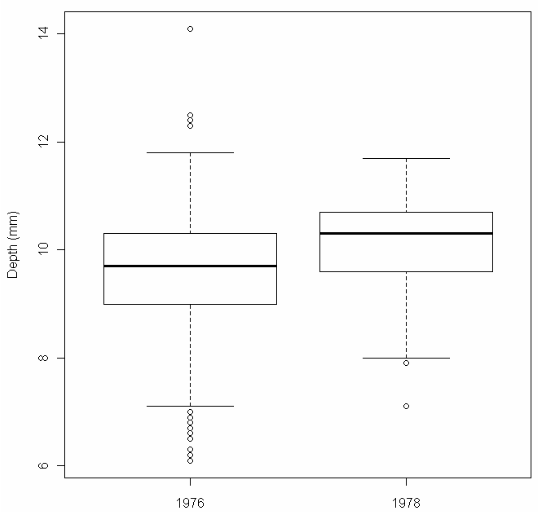
Chapter 2: Inference Using *t*-Distributions

2.12 **a** From 18 8 to 372 grams.
b From 203 to 357 grams. In both these, round *out,* which can be accomplished by rounding the halfwidth up before adding and subtracting.
c *t*-statistic = 6.00. The two-sided *p*-value is <.0001. (It is minuscule.)

2.13 **a** (Fish, Regular): Averages are (6.571, −1.143); SDs are (5.855, 3.185)
b Pooled SD = 4.713
c SE for difference = 2.519
d d.f. = 12; $t_{12}(.975) = 2.179$
e 95% CI from 2.225 to 13.203 mm
f *t*-stat = 3.062
g One-sided *p*-value = .005. Using the table in Appendix 2, locate the d.f. =12 line, and move across the line until the position of the *t*-statistic, 3.062. It is slightly larger than 3.055 so the table tells you that the one-sided *p*-value is slightly smaller than .005.

2.15 *t*-statistic = 9.32, with 174 d.f. Very convincing, indeed.

2.18 **The Grants' Complete Finch Beak Data.**
a



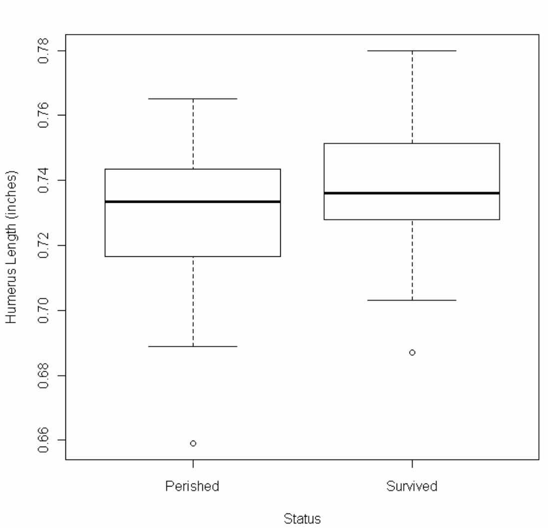
b one-sided *p*-value = 1.617e−06 (equal variance version of two-sample *t*-test)
c two-sided *p*-value = 3.233e−06 (equal variance version of two-sample *t*-test)
d Estimate of 1978 mean minus 1976 mean: 0.54mm; 95% confidence interval: 0.31 to 0.76mm
e Some of the finches may be in both samples or some of the finches in the 1978 sample may be offspring of some in the 1976 sample.

2.19 **a** Average = −1.14; SD = 3.18; d.f. = 6.
b SE = 1.20
c 95% CI: from −4.09 to 1.80
d *t*-statistic = −0.95; two-sided *p*-value = .38. [Using the table: 0.906 < 0.95 < 1.134, so 0.95 is between the 80th and the 85th percentiles. The one-sided *p*-value is therefore between 1−0.80 = 0.20 and 1−0.85 = 0.15, and the two-sided *p*-value is between 0.40 and 0.30 (by doubling).]

2.20 *t*-statistic = 2.97; two-sided *p-value* = 0.025. [Using the table: 2.612 < 2.97 < 3.143, so the one-sided *p*-value is between 1−.99 = .01 and 1−.98 = .02. The two-sided *p*-value is between .02 and .04 (by doubling).] The typical reduction is about 6.6 mm of mercury. The 95% confidence interval is from 1.15 to 11.99 mm.

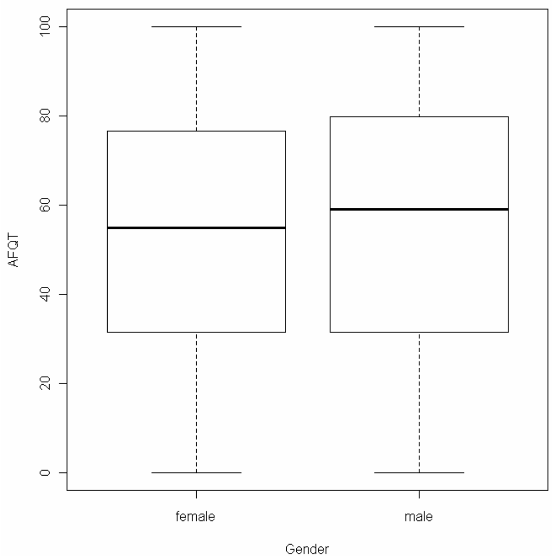
DATA PROBLEMS

2.21 **Bumpus Natural Selection Data**.

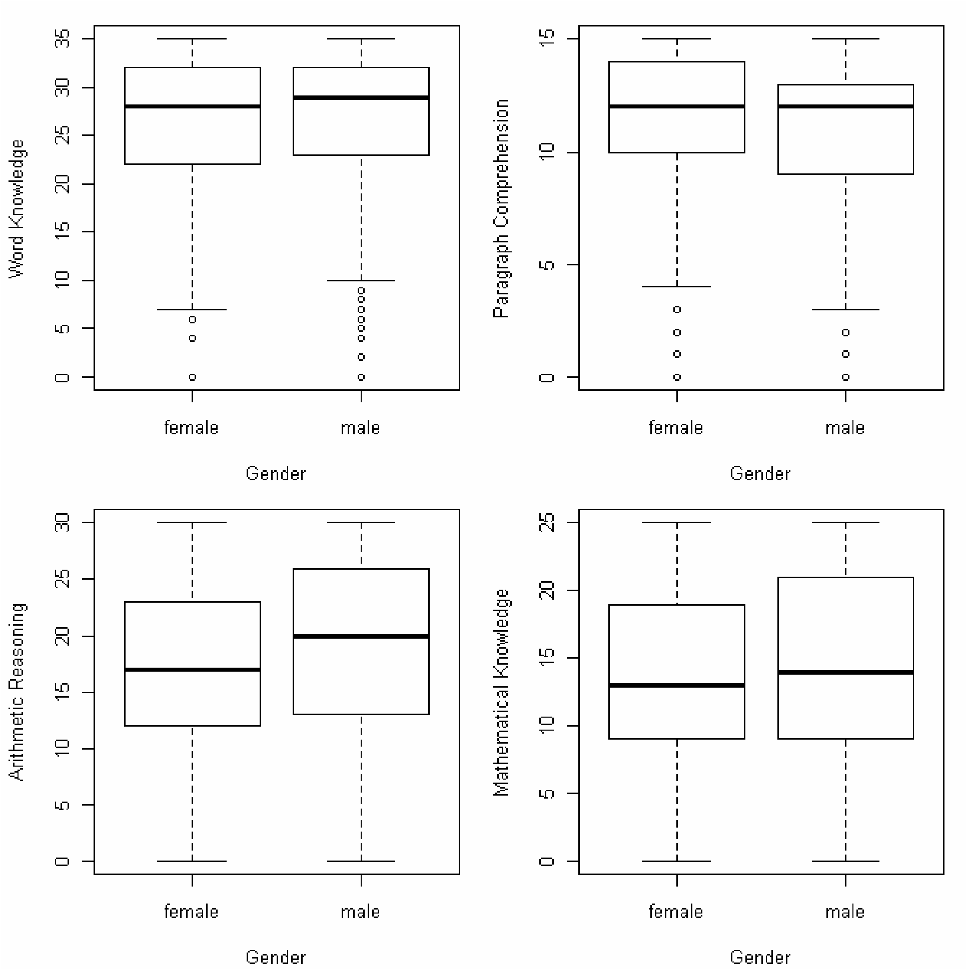


These data provide suggestive but inconclusive evidence that the distribution of humerus lengths differed in the populations of sparrows that perished and survied (2-sided *p*-value = 0.08 from a two-sample *t*-test). The mean for the population that survived is estimated to exceed the mean for the population that perished by 0.0101 inches (95% confidence interval: −0.0214 to 0.0013 inches).

2.22 **Male and Female Intelligence**.



These data provide suggestive but inconclusive evidence that the distribution of AFQT scores for males and females differ (2-sided *p*-value = 0.062). The male mean is estimated to exceed the female mean by 2.04 percentage points (95% confidence interval for male excess: − 0.10 points to 4.18 percentage points).



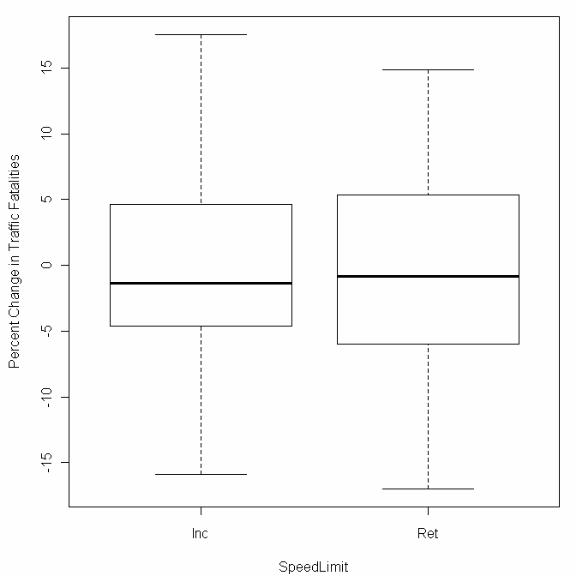
The data provide no evidence of a gender difference on Word Knowledge test scores (2-sided *p*-value = 0.94). The female mean is estimated to exceed the male mean by 0.02 percentage points (95% confidence interval −0.57 to 0.52 percentage points).

The data provide convincing evidence of a gender difference on Paragraph Comprehension test scores (2-sided *p*-value = 0.0000045). The female mean is estimated to exceed the male mean by 0.57 percentage points (95% confidence interval 0.32 to 0.81 percentage points).

The data provide convincing evidence of a gender difference on Arithmetic Reasoning test scores (2-sided *p*-value = 0.0000000000004). The male mean is estimated to exceed the female mean by 2.04 percentage points (95% confidence interval 1.49 to 2.58 percentage points).

The data provide strong evidence of a gender difference on Mathematical Knowledge test scores (2-sided *p*-value = 0.002). The male mean is estimated to exceed the female mean by 0.75 percentage points (95% confidence interval 0.27 to 1.24 percentage points).

2.23 Speed Limits and Traffic Fatalities.



The data provide no evidence that the mean percentage increase was higher in states that increased their speed limit than in states that didn't (1-sided *p*-value = 0.44). The mean percentage increase in traffic fatalities in states that increased their speed limit was estimated to exceed the mean in states that didn't by 0.39 percentage points (95% confidence interval: −4.4 to 5.2 percentage points).

Chapter 3: A Closer Look at Assumptions

3.20 **a** average In-State tuition = $11,600
log(average In-State tuition) = 9.3588
average log(In-State tuition) = 8.6606

b median In-State tuition = $5,000
log(median In-State tuition) = 8.5172
median log(In-State tuition) = 8.5172

c median Out-of-State tuition = $30,000
(median Out)/(median In) = 6.0
median(Out/In) = 3.0
median[log(Out) − log(In)] = 1.0986 = log(3)

3.21 **a** One-sample t-test on differences (observed − expected) for the subset of umpires whose lifetimes were not censored (Censored = 0): t-stat = −0.987, df = 194, p-value = 0.32 (1-sided p-value = .16). A 95 percent confidence interval for mean life length minus expected life length: −1.6 years to 0.54 years.

b This might be a problem if the ones for whom data were unavailable tended to have died young. In any case, the available sample is not a random sample from the population of all umpires.

c This is a considerable problem since with the given sampling routine we are more likely to sample umpires who died young than umpires who died old. For this reason the t-test based on the uncensored lifetimes is not a good idea here. (It is also inappropriate to insert artificial death times for the censored group; more sophisticated techniques of *survival analysis* would be needed.)

3.22 **a**

at 26 kV:	1.756	7.365	7.751		
at 28 kV:	4.231	4.685	4.703	6.055	6.973

b At (26 kV, 28 kV): Averages = (5.624, 5.329), SDs = (3.355, 1.145), *n*'s = (3, 5). Difference in averages = 0.295.

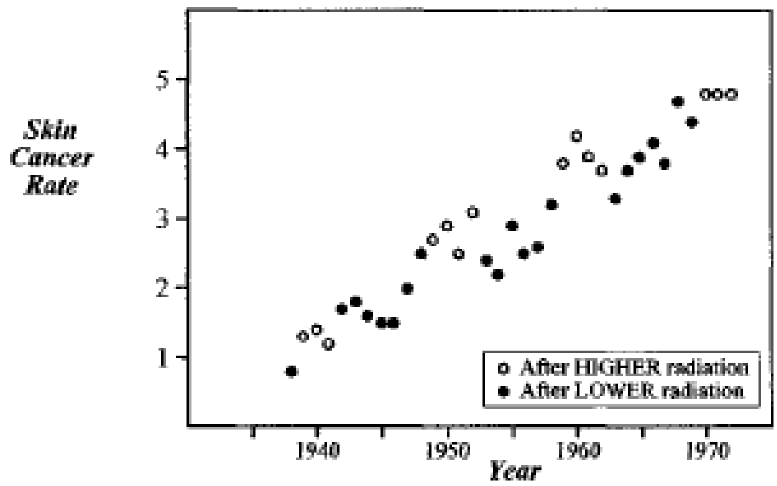
c exp(0.295) = 1.343 estimates the multiplicative effect on time to breakdown of changing voltage level from 28 kV to 26 kV.

d 95% CI goes from −4.138 to 3.549. Antilogs: from 0.016 to 34.8. The multiplicative effect of raising the voltage from 26 kV to 28 kV is estimated to be between 0.016 and 34.8 (95% confidence interval). So, you expected better precision?

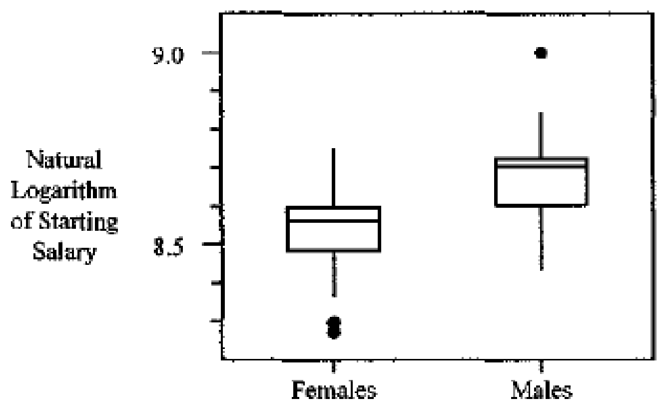
3.23 Refer to Display 3.10.

a Yes. One should expect the rates to follow a time series where serial correlation is present.

b Following is a picture that puts them both together. There is a problem: there is a steady increase, or 'trend', in the series. There is also a somewhat cyclic behavior. The trend and (possibly) the cyclic behavior are most likely unrelated to solar radiation, but they will have a strong influence on the comparison because more of the 'after higher' values fall in the later years.



3.24 **a** The box plots look like this:



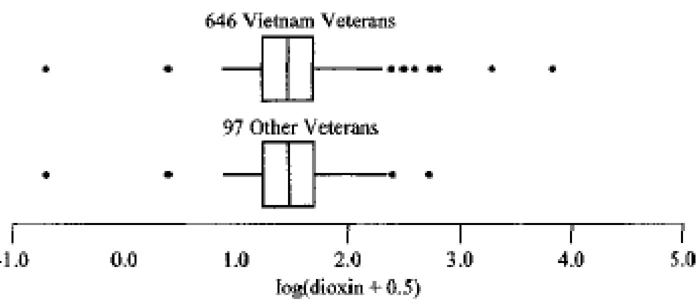
b Based on a *t*-statistic of 6.17 with 91 degrees of freedom, the one-sided *p*-value is <.0001.

c The 95% confidence interval for the ratio of median starting salaries (M/F) is from 1.10 to 1.21. (A 95% confidence interval for the ratio of median starting salaries (F/M) is from 1/1.21 to 1/1.10, or .82 to .91.)

3.25 Use the computer. Refer to Display 3.6.

3.26 **Agent Orange.** Use log(dioxin + 0.5) as the response.

a

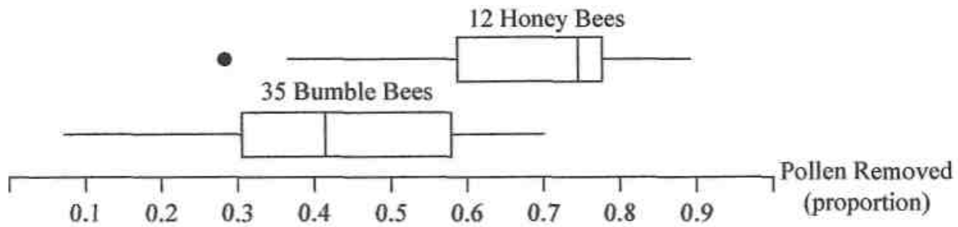


b Two-sided p-value = .3816

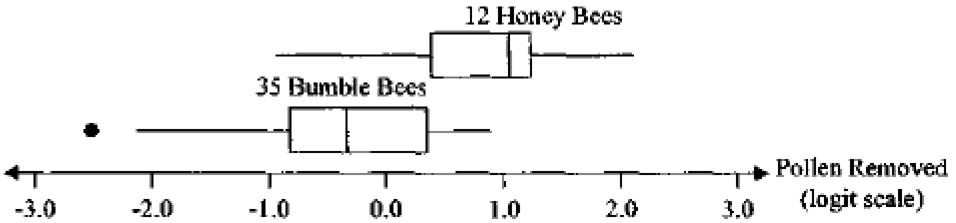
c (−0.0539, +0.1410) on the transformed scale converts to (0.9475, 1.1514) by anti-logs. These are bounds on the ratio of (median + 0.5) values (Vietnam to Other).

3.27

a (i)

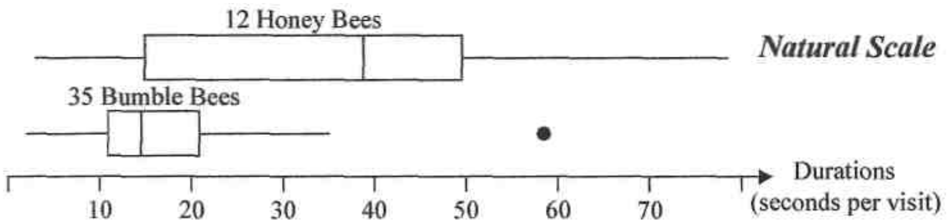


(ii)

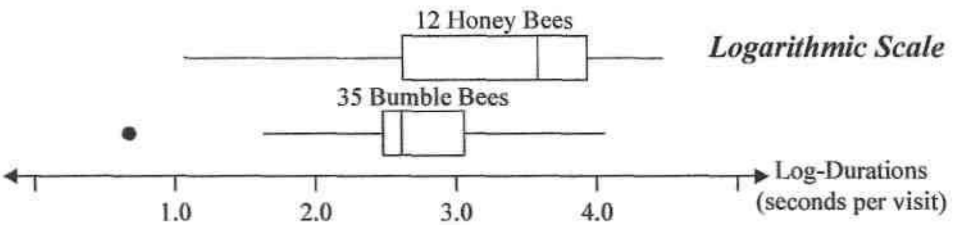


(iii) t = 3.85, two-sided p-value = .0004.

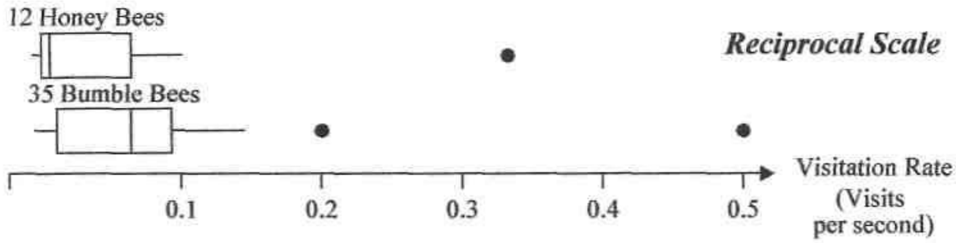
b (i)



(ii)



(iii)

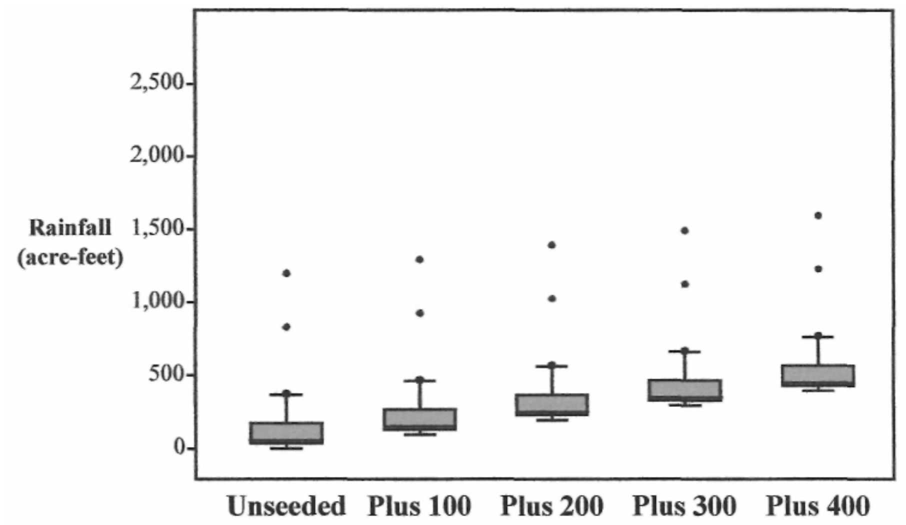


(iv) The log scale looks best (if not ideal). (v) Natural scale: from 10.0 to 29.4. Log scale: from 0.18 to 1.12. Reciprocal scale: from −0.081 to 0.031. (vi) All three are relatively easy to deal with. The reciprocal scale is a natural scale, but converting differences back would not give anything meaningful. (vii) It's very difficult.

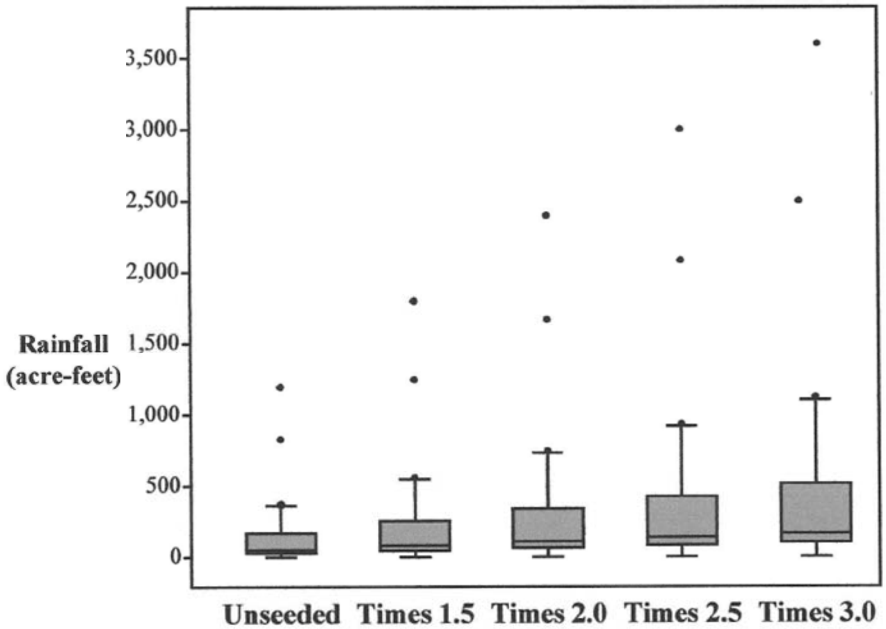
3.28 With all the data, the one-sided *p*-value is 0.0405; without the .659 value, the one-sided *p*-value is .0900. This is a fair swing; the evidence goes from suggestive to none.

3.29 **Cloud Seeding.**

a Additive changes to be unseeded distribution:



b Multiplicative (1.5 − 3.0, not 2 – 5) changes to the unseeded distribution:



c The data (Display 3.2) look more like the multiplicative effect of part **b**.

DATA PROBLEMS

3.30 **Education and Future Income.** The incomes span several orders of magnitude, so they are log-transformed. A test of the difference in average logs results in a two-sided p-value < 0.0001, providing convincing evidence of a real difference (subject to caveats concerning sampling). Upon back-transformation, it is estimated that students completing 16 years of education have 1.77 times (77% higher) the salary of students completing only 12 years. A 95% confidence interval on this multiplicative factor is from 1.60 (60% higher) to 1.96 (96% higher).

3.31 **Education and Future Income II.** Incomes are again log-transformed. A test of the difference in average logs results in a two-sided p-value = 0.165, providing no evidence of a real difference (subject to caveats concerning sampling). Upon back-transformation, it is estimated that students completing >16 years of education have 1.11 times (11% higher) the incomes of students completing only 16 years. A 95% confidence interval on this multiplicative factor is from 0.96 (4% lower) to 1.28 (28% higher).

3.32 **College Tuition.** Box plots or histograms of tuitions suggest log-transformation.

a This is a one-sample problem, with the response of log(Out/In). One-sample t-tools provide convincing evidence (two-sided p-value < 0.0001) that out-of-state tuitions exceed in-state tuitions in public schools. Back-transformation estimates that out-of-state tuition is 2.29 times (129% higher) than in-state tuition. The 95% confidence interval is from 2.01 to 2.61.

b This is a two-sample problem, again with log-transformed tuitions. The two-sided p-value ≤ 0.0001 provides convincing evidence of a difference between in-state tuitions of private and public schools. Back-transformation estimates that in-state tuition at private schools is 3.69 times what it is in public schools (95% confidence interval from 2.90 to 4.69).

c This is also a two-sample problem, and again on the transformed scale. The two-sided p-value = 0.0002 provides convincing evidence of a difference between out-of-state tuitions of private and public schools. Back-transformation estimates that out-of-state tuition at private schools is 1.61 times what it is in public schools (95% confidence interval from 1.27 to 2.05).

3.33 **Brain Size and Litter Size.** Pictures make the log scale an obvious choice. The evidence that brain sizes relative to body sizes are unequal is strong, but not convincing (two-sided *p*-value = .0512, from two-sample t = 1.975 with 94 d.f.). If relative brain weight (RBW) is 1,000 × (brain weight/body weight), it is estimated that the median RBW among species with larger litter sizes (Š2) is 48.7% larger than the median RBW among species with smaller litter sizes (<2). A 95% confidence interval on this factor is (0.2% smaller, 121.7% larger).

Chapter 4: Alternatives to the *t*-Tools

4.14 **O-Ring study**. The one-sided *p-value* from the *t*-test is .0004, compared to .00989 from the permutation test.

4.15 One-sided *p*-value = 2/10 = .20.

4.16

Group 1	Group 2	Aver. Diff.
4,5,6	7,12	−4.50
4,5,7	6,12	−3.67
4,5,12	6,7	+0.50
4,6,7	5,12	−2.83
4,6,12	5,7	+1.33
4,7,12	5,6	+2.17
5,6,7	4,12	−2.00
5,6,12	4,7	+2.17
5,7,12	**4,6**	**+3.00**
6,7,12	4,5	+3.83

One-sided *p*-value = 2/10 = .20 ... again.

4.17 **O-Ring study.** (136 + 170 + 10 + 85 + 10 + 10) / 10,626 = 421/10,626 = .0396.

4.18

Treatment	Control	Rank Sum
1,2,3	4,5,6	6
1,2,4	3,5,6	7
1,2,5	3,4,6	8
1,2,6	3,4,5	9
1,3,4	2,5,6	8
1,3,5	2,4,6	9
1,3,6	2,4,5	10
1,4,5	2,3,6	10
1,4,6	2,3,5	11
1,5,6	2,3,4	12
2,3,4	1,5,6	9
2,3,5	1,4,6	10
2,3,6	1,4,5	11
2,4,5	1,3,6	11
2,4,6	1,3,5	12
2,5,6	1,3,4	13
3,4,5	1,2,6	12
3,4,6	1,2,5	13
3,5,6	**1,2,4**	**14**
4,5,6	1,2,3	15

One-sided *p-value* = 2/20 = 0.10.

4.19 **a** 0.1718
b Normal approximation
c Continuity correction
d *t*-test gives $p = .081$; *t*-test with removal gives $p = .180$; rank sum gave $p = .1718$.
e The rank sum test is valid AND it uses all the data.

4.20 **a** **Trauma and metabolic expenditure.**
b Sum = $T = 82$
c $n_1 = 7$, $n_2 = 8$, $\mu_R = 8.00$, $\sigma_R = 4.4681$, Mean(T) = 56.00, SD(T) = 8.6333, Z = 3.012
d One-sided *p*-value = .0013.

4.21 **Trauma and metabolic expenditure.** $Z = 2.95369$; two-sided *p*-value = .0314

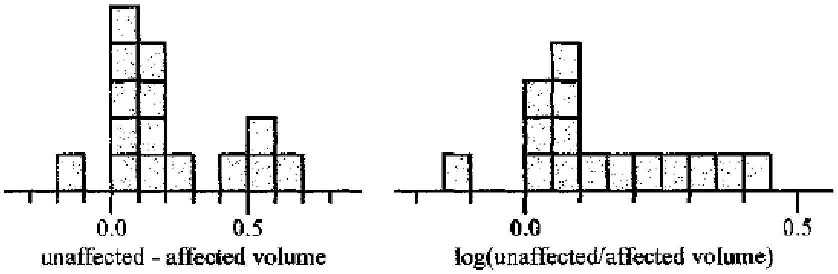
4.22 **Trauma and metabolic expenditure. (1.9, 16.8)**

4.23 **Motivation and creativity.** Two-sided *p*-value = .00643, compared to .00537.

4.24 **Motivation and creativity,** (1.00, 6.60), compared to (1.29, 7.00). (The former is based on the randomization test.)

4.25 **Guinea pig lifetimes.** CI: (39.59, 165.81), based on Welch's *t* with 97 d.f. The halfwidth is 63.11 and the critical *t*-multiplier is 1.9847. $SE_W = 31.80$ makes $t_w = 3.23$, giving a two-sided $p = .0016$. No. It looks like something else is involved.

4.26 **Schizophrenia Study, a** and **b.** There is mild skewness on the natural scale. The log scale is (marginally) better.

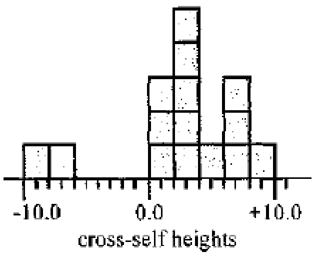


c On the log-scale, $t = 3.20$, for a two-sided $p = .0065$. On the 'natural' scale, the *p*-value is .0061, so there is virtually no difference.

d Estimate = 0.1285; 95% CI:(0.0423, 0.2147), which translates back to the original scale as ... estimate = 1.14; 95% CI: (1.04, 1.24).

4.27 **Schizophrenia Study.** Two-sided *p*-value = .00452, from signed rank test on the log(ratio) values. On the straight difference scale, the signed rank gives .00208 ... close. It is not particularly apparent.

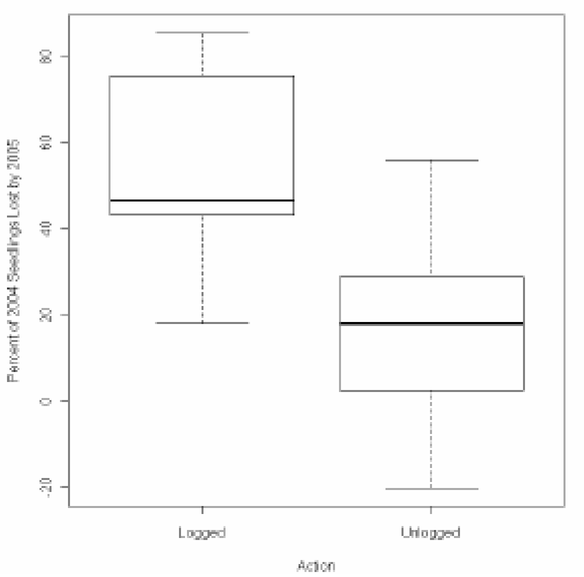
4.28 a Darwin Data.



b .0497
c 95% CI: (0.0033, 5.2301)
d No
e .0438.

DATA PROBLEMS

4.29 Salvage Logging.



The data provide strong evidence that the mean percentage of seedlings lost from 2004 to 2005 differed for the two types of logging (2-sided *p*-value = 0.007 from a two-sample t-test). The mean percentage of seedlings lost in logged plots was estimated to exceed the mean in unlogged plots by 38.1 percentage points (95% confidence interval: 16.7 to 54.8 percentage points).

a 2-sided *p*-value from (exact) Rank Sum test with continuity correction: 0.01154; estimated difference in "location": 33.4 percentage points; 95% confidence interval: 10.8 to 65.1 percentage points.

b 2-sided *p*-value from two-sample t-test: 0.007011; estimated difference in means: 38.1 percentage points; 95% confidence interval: 16.7 to 54.8 percentage points. The t-test gives a smaller *p*-value and narrower confidence interval. (Note: the boxplots do not seem to indicate any problema with using the t-test.)

4.30 **Sunscreen Protection Factor.** This is a one-sample (paired) problem. A starting point may be examining the set of ratios = (During Treatment)/(Pretreatment), which estimate the protection factor of the sunscreen — according to the theory. A stem-and-leaf diagram suggests no particular problem, so one could construct an estimate and confidence interval directly from the ratios. The estimate is 9.22, with a 95% confidence interval of (5.60, 12.85).

Because theory suggests the factor is multiplicative, a better analysis proceeds on the logarithmic scale. With the t-tools and back-transformation, the resulting estimate is 7.88, with a 95% confidence interval from 4.76 to 11.43.

Distributional problems are suggestive enough to warrant a non-parametric check. One can utilize the Wilcoxon signed-rank test to construct a confidence interval. To determine whether ϕ should be in the confidence interval or out, divide each During Treatment value by ϕ. Take the logs of the results and compare them to the logs of the Pretreatment values. If the signed-rank test two-sided *p-value* is greater than or equal to .05, include ϕ. Using trial and error, the resulting interval extends from 4.62 to 12.24.

There are many confounding factors that can enter into this study. It is unfortunate that randomization was not used. Another feature that should be noted in this example is that the 'mean' protection factor affords quite variable protection to different individuals. For this reason, it may be more appropriate to report a 95% *prediction interval* for the protection afforded an individual by the sunscreen. This is not covered in the text. It can be constructed by using the same estimate, but using a standard error of $s\sqrt{(1+1/n)}$. The result: 95% of the protection factors afforded by this liquid to individuals are estimated to be between 1.43 and 37.94.

4.31 **Group Therapy and Breast Cancer.** Looking at the data suggests that the rank-sum test is appropriate. There are a large number of ties, but the test may be performed. The sum of ranks in the control group is $T = 637$. If there is no (additive or multiplicative) treatment effect, the expected sum is 708, and the standard deviation of the sum is 63.25. This gives a continuity-corrected test statistic of $Z = -1.11$ and a two-sided p-value = .2650. It suggests there is no evidence of a treatment effect.

Looking again at the data, one should be a bit perplexed by the result. There certainly are a large number of long survivors in the therapy group and none in the control group. Why does this not show in the test? The stem-and-leaf diagram (below) reveals that there are also a larger proportion of early failures in the therapy group.

Such effects are not uncommon. A beneficial treatment may be harsh in the beginning. Some may respond unfavorably at the outset. However, individuals that do not respond unfavorably may be aided substantially. One may hesitate to parallel group therapy with chemotherapy, but read on.

The sample medians in these data are 17 months in the control group and 21 months in the therapy group. The 4-month difference between medians does not reflect the large difference between the two distributions. But the rank-sum test is looking for a difference in medians, so it is unlikely to pick up the apparent difference.

Control ($n = 24$)		Group Therapy ($n = 34$)
862	0	22444668
888666444220	1	00246668
6220	2	02
864	3	26
80	4	6688
	5	88
	6	6
	7	22
	8	2
	9	
	10	
	11	
	12	2
		2*2*2*

Caution: Data Snooping ahead. Levene's test offers an opportunity to focus on the apparent difference. It runs into problems with the censored information, but an easy way to get around this problem is to use sample medians instead of averages. On a log scale, compute

$$X_{ij} = [\log(Y_{ij} / \text{median}_i)]^2$$

(i referring to group) and perform a two-sample analysis. This does not totally escape the censoring problem, because now the censored observations do not necessarily have the largest responses. Ignoring that fact gives a two-sided *p*-value of .0054. If the actual values of the censored observations were known, the evidence would only be stronger.

Although the *p*-value appears convincing, it cannot be trusted. The data themselves suggested a particular form to the treatment effect — a form that would not have been anticipated prior to seeing the stem-and-leaf diagram. To construct a test to look for a pattern that has already been found is known as *data snooping.* The *p*-value for any such test cannot be relied upon as an accurate measure of uncertainty, because it fails to include the effects of the snooping. More will be said about this in later chapters.

Students should be rewarded if they notice this pattern and suggest this is the reason for lack of a significant result from the rank sum test. They should be advised that reporting the pattern should be phrased descriptively, not inferentially (even though this was a randomized experiment).

Another way to approach the same issue is to use a permutation test based on the ranks of the magnitudes of the differences between survival times and the combined median survival time (18 months). Calculate $|Y_i - \text{Median}(Y_i)|$, rank them all, and perform the same operation as for the rank sum test. Add the ranks (corrected for ties) among the therapy group. The result is *z*-statistic = 3.02, with one-sided *p*-value = .0013. Again, however, this is data snooping, and the .0013 does not incorporate the search for the 'right' test.

4.32 **Therapeutic Marijuana.** For every subject in the study the number of vomiting episodes after the marijuana treatment was less than or equal to the number after the placebo treatment. There is overwhelming evidence of a reduction due to marijuana in these subjects (one-sided p-value = .0005 from a Wilcoxon signed-rank test of the differences). The median decrease in number of episodes is 19. A 95% confidence interval for the marijuana effect is a reduction of 9.2 to 54.5 episodes. (Note 9.2 and 54.5 are the hypothesized values for the marijuana effect that lead to two-sided p-values of nearly .05 with the Wilcoxon signed rank test, found by trial and error).

Chapter 5: Comparisons Among Several Samples

5.13 **Spock Trial.**
- **a** 26.6%.
- **b** All 9.
- **c** 4. (Count the number of negative residuals in the separate means model fit.)

5.14 **Spock Trial.**
- **a** 6.914, with 39 d.f.
- **b** *t*-statistic = 5.056, with 39 d.f., giving one-sided *p*-value < .0001.

5.15 **Spock Trial.**
- **a** 84.256
- **b** 3,791.525
- **c** 47.807 × 39 = 1,864.466
- **d**

Source of Variation	Sum of Squares	d.f.	Mean Square	*F*-Statistic	*p*-value
Between Groups	1,927.059	6	321.177	6.718	.0001
Within Groups	1,864.466	39	47.807		
Total	3,791.525	45			

5.16 Use the computer.

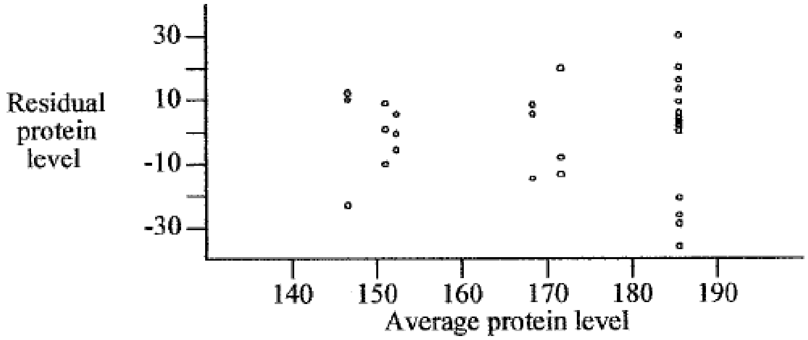
5.17 There are 8 groups. The evidence for different group means is strong

Source of Variation	Sum of Squares	d.f.	Mean Square	*F*-Statistic	*p*-value
Between Groups	35,819	7	5,117	3.50	.0099
Within Groups	35,088	24	1,462		
Total	70,907	31			

5.18 **Fatty Acid.**
- **a**

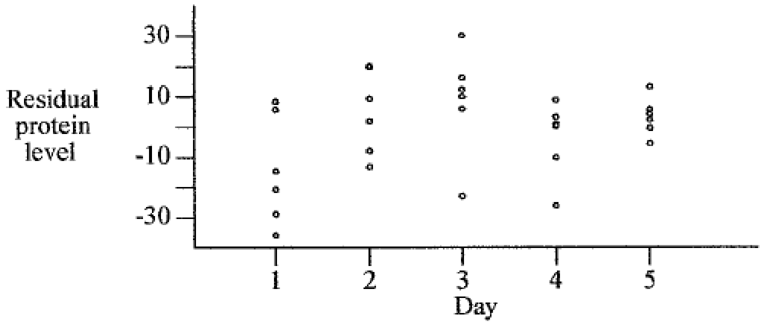
CPFA50	CPFA150	CPFA300	CPFA450	CPFA600	Control
168.3	171.7	146.7	151.0	152.3	185.6

Residuals vs Estimated Means:



There is no suggestion that the size of residuals depends on the average protein level.

Residuals vs. Day:



There is a suggestion that the mean level may change from one day to another.

b

	CPFA50	CPFA150	CPFA300	CPFA450	CPFA600	Control
Day 1	168.3					157.3
Day 2		171.7				195.7
Day 3			146.7			203.3
Day 4				151.0		179.0
Day 5					152.3	192.7

Analysis of Variance Table:

Source of Variation	Sum of Squares	d.f.	Mean Square	*F*-Statistic	*p*-value
Between Groups	11,147.47	9	1,238.607	7.801	.0001
Within Groups	3,175.333	20	158.767		
Total	14,322.800	29			

There is convincing evidence that the means are different

c

Source of Variation	Sum of Squares	d.f.	Mean Square	*F*-Statistic	*p*- value
Between Groups	11,147.470	9	1,238.607	7.801	.0001
Between Treatment	7,222.533	5	1,444.507	9.098	.0001
Between Days in Control	3,924.933	4	981.233	6.180	.0021
Within Groups	3,175.333	20	158.767		
Total	14,322.800	29			

There is ample evidence to suggest that the means under the control treatment are different on different days.

5.19 **Cavity Size and Use.**

a 0.1919

b

Source of Variation	Sum of Squares	d.f.	Mean Square	*F*-Statistic	*p*-value
Between Groups	17.4402	8	2.1800	11.3583	<.0001
Within Groups	54.7007	285	0.1919		
Total	72.1408	293			

c The first term is $127(7.347)^2 + ... + 6(8.297)^2 = 16{,}442.9742$. The grand mean is $Y = (127\times7.347 + ... + 6\times8.297)/294 = 7.4746$, so the second term (the "correction factor") is $294(7.4746)^2 = 16{,}425.5202$. The difference is 17.4540, which is roundoff error away from the ANOVA answer.

d You will need a between group sum of squares for the intermediate model. Following the formula in part c, you first calculate averages in the two sets of species. $= (\overline{Y_1}$ $127\times7.347 + ... + 16\times7.568)/(127 + ... + 16) = 1{,}999.4820/270 = 7.4055$; and $\overline{Y_2} = (11\times8.214 + 7\times8.272 + 6\times8.297)/(11+7+6) = 198.0400/24 = 8.2517$. The between group sum of squares is then $270(7.4055)^2 + 24(8.2517)^2 - 294(7.4746)^2 = 16{,}441.3018 - 16{,}425.5202 = 15.7816$, with 1 d.f. (The correction factor — 16,425.5202 — is the same as in part c.) The extra sum of squares for adding different species means into this intermediate model is $17.4402 - 15.7816 = 1.6586$. Putting these together as in Section 5.3.3 yields the analysis of variance table:

Source of Variation	Sum of Squares	d.f	Mean Square	*F*-Statistic	*p*-value
Between Groups	17.4402	8	2.1800	11.3583	<.0001
Between Sets	15.7816	1	15.7816	82.2248	<.0001
Within Sets	1.6586	7	0.2369	1.2345	.2910
Within Groups	54.7007	285	0.1919		
Total	72.1408	293			

The within-sets *p*-value indicates there is no evidence that the species means are different within the two sets. There are only two different mean values.

5.20 15.24

5.21 Levene's test for the Spock trial data.

Source of Variation	Sum of Squares	d.f.	Mean Square	*F*-Statistic	*p*-value
Between Groups	135.74	6	22.62	1.4459	0.2223
Within Groups	610.18	39	15.65		
Total	745.92	45			

The evidence does not suggest that the spread differs between groups.

5.22. a ANOVA *F*-statistic = 3.3363, with 3 & 76 d.f. => *p*-value ~0.0237

b Conclusion: strong evidence that differences exist.

c It makes inferences model-based; introduces the possibility of biases.

DATA PROBLEMS

5.23 **Tyrannosaurus rex.** A graphical check (below) reveals that the spread is uniform across groups and that there are no suspicious outliers or non-normality features. One can proceed with the standard analysis

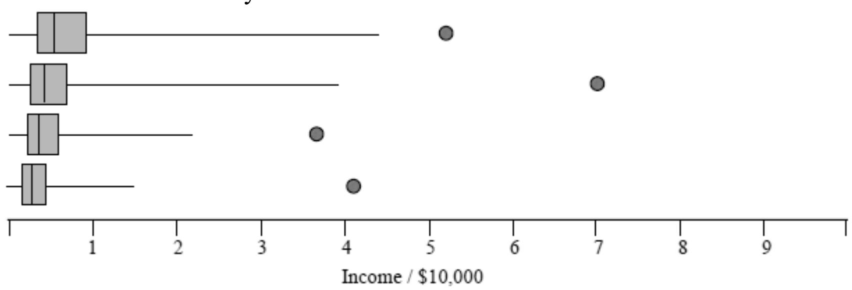
Source of Variation	Sum of Squares	d.f.	Mean Square	*F*-Statistic	*p*- value
Between Groups	6.0675	11	0.5516	7.4268	<.0001
Within Groups	2.9708	40	0.0743		
Total	9.0683	51			

The question of interest is answered affirmatively, with $p < .0001$.

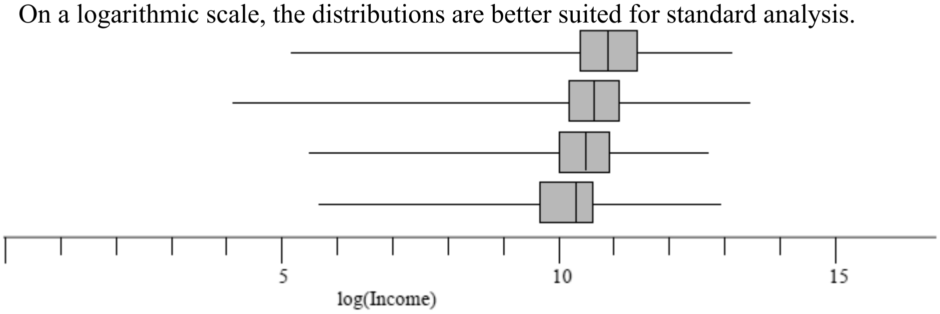
Stem	Dorsal vertebra (1)	Proximal caudal	Rib 16	Dorsal vertebra (2)	Gastralia (1)	Mid-caudal	Tibia	Metatarsal	Gastralia (2)	Phalange	Femur	Distal caudal
5												
6	1											
7												
8	8											
9		3		2								
10		18										
11	2	2	0									
12	4	8	29	0		5						
13		7		0	2	3	3	2				
14	3		9		0	07	1					
15										4		
16				2			2	5	0			
17				0	1				8		09	
18										9		
19						2		6			1	5
20									5	4		1
21							5	5			5	
22												5
23												09
24												
25												

Legend: 23 | 9 represents oxygen isotopic composition = 12.39

5.24 Incomes in each group are positively skewed with larger spread going along with larger center and with at least one very extreme value.



On a logarithmic scale, the distributions are better suited for standard analysis.



Analysis on the log scale provides convincing evidence of group differences.

Source of Variation	Sum of Squares	d.f.	Mean Square	F-Statistic	p-value
Between groups	184.79	3	61.5979	70.1651	<0.0001
Within groups	2,264.98	2,580	0.8779		
Total	2,449.77	2,583			

The difference on the log scale is 0.7347 with a standard error of 0.0521. Back-transformation gives and estimate that median income in the 4th quartile group is 2.08 times (108% greater) than the median income in the 1st quartile group. The 95% confidence interval is from 1.88 (88% greater) to 2.31 (131% greater).

5.25 Again, incomes are transformed to the log scale. Then ANOVA provides convincing evidence of group differences.

Source of Variation	Sum of Squares	d.f.	Mean Square	F-Statistic	p-value
Between groups	217.8538	4	54.4134	62.8695	<0.0001
Within groups	2,232.1204	2,579	0.8655		
Total	2,449.7742	2,583			

Differences are evaluated on the log scale and then back transformed to arrive at the following summary, which provides estimates of the ratio of median incomes in group A to group B.

Group A	Group B	(Median B)/(Median A)	Percentage Greater	95% CI on Percentage
<12	12	1.388	38.8%	17.5% - 64.0%
12	13-15	1.178	17.8%	7.5% - 29.1 %
13-15	16	1.501	50.1%	33.7% - 68.4%
16	>16	1.106	10.6%	-2.9% - 26.1%

Chapter 6: Linear Combinations and Multiple Comparisons of Means

6.12 **Handicap study**. $g = 1.181$; SE(g) = 0.504; t-stat = 2.344, with 65 d.f. For a planned comparison, two-sided p-value = 0.022. For an unplanned comparison, compare to the 5% critical difference of 2.814 (Tukey-Kramer); 2.344 is smaller, so p-value > 0.05.

6.13 **Handicap Study.** For Bonferroni with only 3 groups, $k = 3*2/2 = 3$. So [1–.05/6] = .9917. Then $t_{65}(.9917) = 2.458$, and the SE(diff) = 0.617, giving an interval halfwidth of 1.517. The 95% confidence intervals are: (Amputee – Crutches) between –3.009 and 0.025; (Amputee – Wheelchair) between –2.431 and 0.603; (Crutches – Wheelchair) between –0.939 and 2.095.

6.14 Use a computer.

6.15 **Comparison of Five Teaching Methods.**

a 4.484

b 1/3, –1/2, –1/2, 1/3, 1/3

c g = 3.000, SE(g) = 1.3645; $t_{40}(.975) = 2.021$, HW = 2.758, giving 95% confidence interval: (0.24, 5.76).

6.16

i)	LSD	2.042
ii)	protected LSD	2.042
iii)	Tukey-Kramer	3.041
iv)	Bonferroni	3.030
v)	Sheffe'	3.557

6.17 **Adder Head Size.**

a $s_p = 11.72$, d.f. = 230, $q(.05; 7, 230) = 4.17$, so T-K multiplier is 2.9486. Also, $t = 1.96$. The following table shows the interval halfwidths for differences. The upper right are T-K, and the lower left has the LSD versions.

	Uppsala	In-Fredeln	Inre Hamnskär	Norrpäda	Kärring-boskär	Ängskär	Svenska Högarna
Uppsala	—	9.59	10.80	10.23	15.08	8.45	9.04
In-Fredeln	6.38	—	9.74	9.10	14.34	7.05	7.75
Inre Hamnskär	7.18	6.47	—	10.37	15.18	8.62	9.20
Norrpada	6.80	6.05	6.89	—	14.78	7.90	8.52
Kärringboskär	10.03	9.53	10.09	9.82	—	13.61	13.98
Ängskär	5.62	4.69	5.73	5.25	9.05	—	6.28
Svenska Högarna	6.01	5.15	6.11	5.67	9.29	4.17	—

b Using mean-corrected body sizes (no further multiplier), g = 29.96, SE(g) = 7.27, so t = 4.12 gives a two-sided p-value < .0001; strong evidence of an association.

c g = 207.03, SE(g) = 71.58, t = 2.89, and two-sided p-value = 0.0038, strong evidence.

6.18 **Nest Cavities,** g = 0.803, s_p = 0.4381, d.f. = 285, SE(g) = 0.0984, HW = 0.1929. 95% CI: 0.610 to 0.996. The estimated ratio of medians is 2.23, with 95% CI from 1.84 to 2.71.

6.19 **Diet Restriction.** q(.05; 6, 343) = 4.03, so Multiplier = 2.85. SE(diff) = 1.188, giving HW = 3.387. Estimate = 9.6, so the CI is from 6.213 to 12.987. The LSD multiplier is 1.9669, so the associated interval is narrower. Inappropriate, because this was a (THE) planned comparison.

6.20 **a** g = 0.62; se(g) = 0.2069 => t-statistic = 2.9965; 2-sided p-value ~ 0.005

b

Differences Between Averages

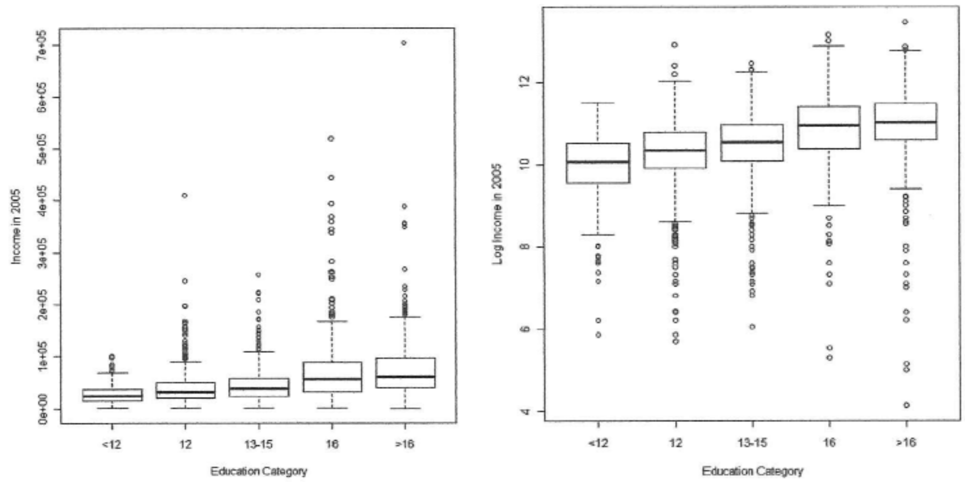
	H	ML	L
ML	0.08		
L	0.10	0.02	
MH	0.13	0.05	0.03

Tukey-Kramer HSDs

	H	ML	L
ML	0.090		
L	0.108	0.112	
MH	0.077	0.083	0.102

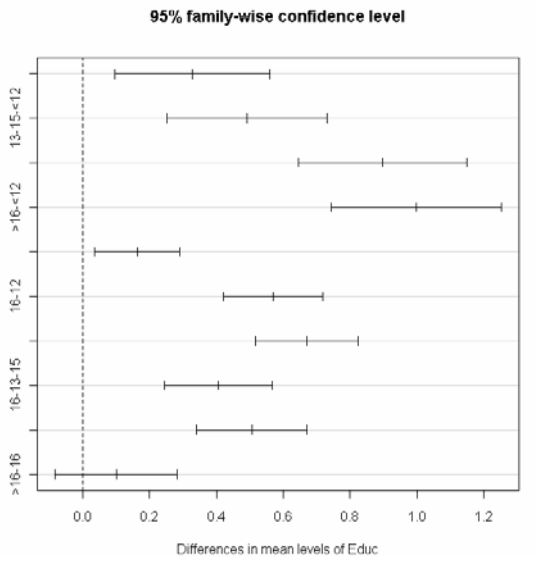
The only difference that appears large enough to be judged real by the HSD is the H vs. MH difference.

6.21 **Education and Future Income.**



The distribution looks more nearly normal with equal spread on the log scale

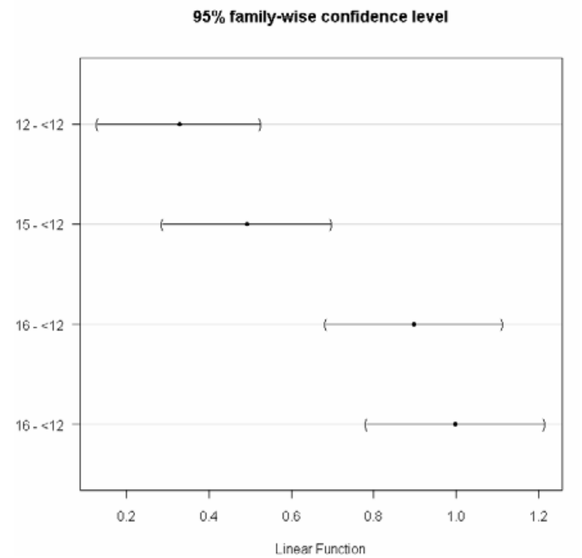
a Tukey Comparisons on logged incomes:



The highlighted rows in the table below show that the only comparison for which there is not some evidence of difference is the >16 years educational category to theto 16 years category.

Comparison	Difference	95% CI: lower	95%CI: upper	Adj. p-value
12−<12	0.3279	0.0961	0.5597	0.0011
13−15−<12	0.4919	0.2523	0.7314	0.0000
16−<12	0.8977	0.6461	1.1493	0.0000
>16−<12	0.9986	0.7443	1.2529	0.0000
13−15−12	0.1640	0.0364	0.2916	0.0042
16−12	0.5699	0.4209	0.7189	0.0000
>16−12	0.6707	0.5172	0.8242	0.0000
16−13−15	0.4059	0.2451	0.5666	0.0000
>16−13−15	0.5067	0.3418	0.6716	0.0000
>16−16	0.1008	−0.0812	0.2828	0.5547

b Dunnett Comparisons to <12 year category, on log scale

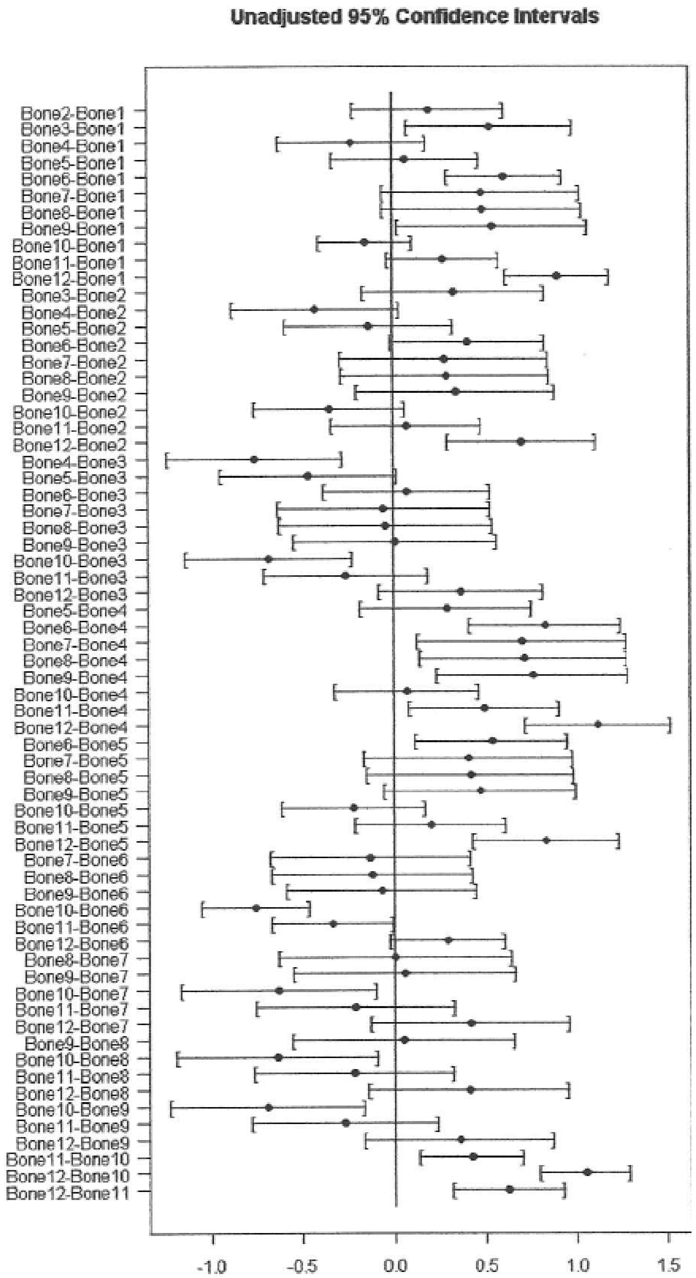


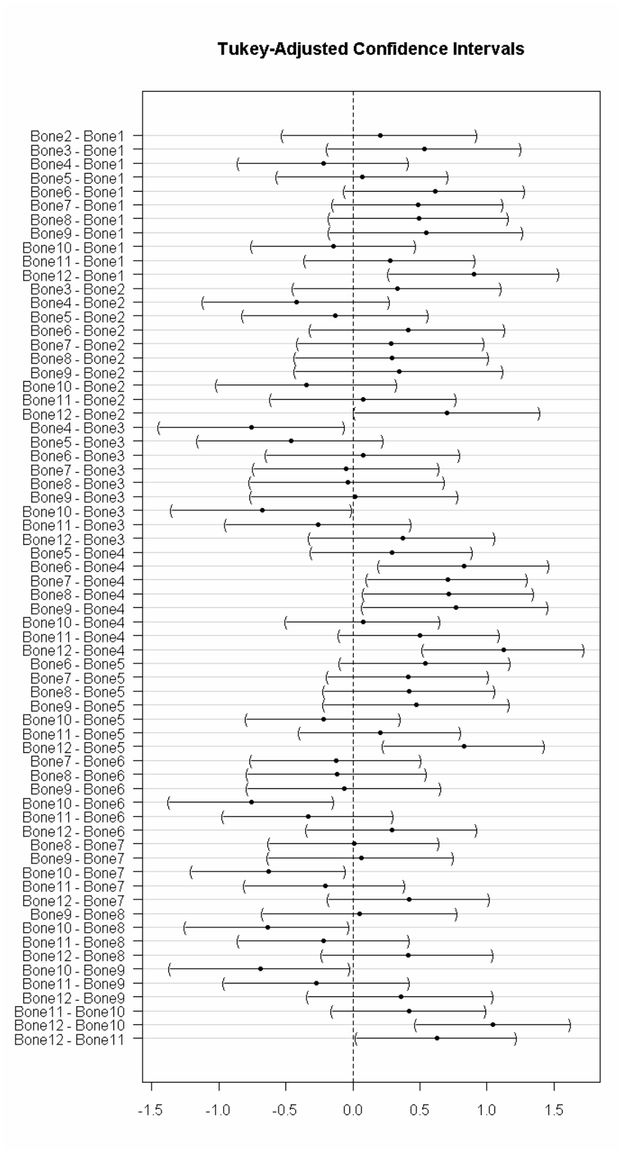
The table below indicates that there is at least some evidence that each of educational groups 12 years, 13–15 years, 16 years, and > 16 years have a mean income greater than that for the <12 year educational group.

Comparison	Estimate	Std. Error	t value	Adj *p*-value
12 − <12 = = 0	0.32787	0.08493	3.861	<0.001
13−15 − <12 = = 0	0.49187	0.08775	5.606	<0.001
16 − <12 = = 0	0.89775	0.09217	9.74	<0.001
>16 − <12 = = 0	0.99856	0.09316	10.719	<0.001

Comparison	Difference	95%CI: lower	95%CI: upper
12 − <12 = = 0	0.3279	0.1307	0.5251
13−15 − <12 = =0	0.4919	0.2881	0.6956
16 − <12 = = 0	0.8977	0.6837	1.1118
>16 − <12 = = 0	0.9986	0.7822	1.2149

6.22 **Dinosaur Bones.** a and b:



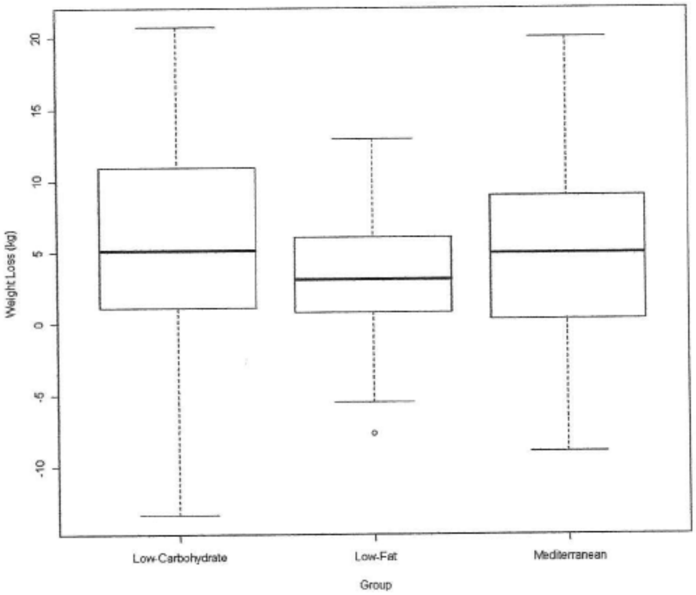


c Number of unadjusted confidence intervals that exclude 0: about 22 or 23

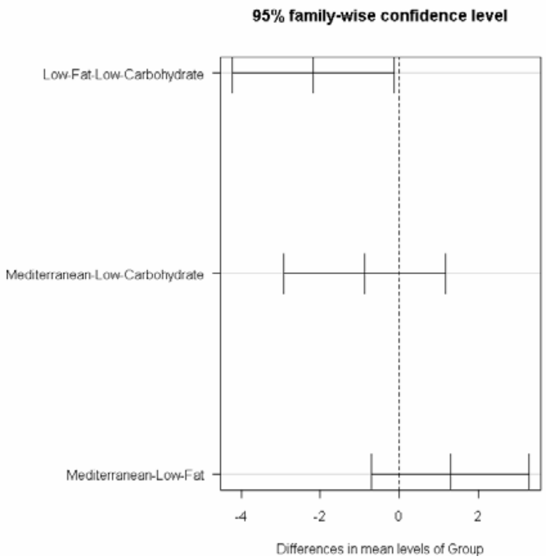
d Number of Tukey-adjusted confidence intervals that exclude 0: about 15 or 14 or 13

e Width of unadjusted 95% CI for comparing bone 2 to bone 1: .8206 (from output not shown). Width of Tukey adjusted 95% CI: 1.44266 (from output not shown). 1.4266/0.8206 = 1.758. The Tukey interval width exceeds the unadjusted inteval width by 76%.

6.23 Diet Wars



As evident in the box plots above, the mean weight loss in the 24 month study ranged from about 3 kg to about 5 kg. The data provide highly suggestive evidence that the mean weight loss is not the same in all three groups (p-value = 0.04 from a one-way ANOVA F-test). In particular, there is moderate evidence that the mean weight loss for the Low-Fat group is different from the mean for the Low-Carb group (2-sided *p*-value = 0.033, from a Tukey-adjusted *t*-test). The mean weight loss for the Low-Carb group is estimated to exceed the mean loss for the Low-Fat group by 2.18 kg (95% confidence interval: 0.14 to 4.22 kg). The graph below shows the Tukey-adjusted 95% confidence intervals for each pairwise comparison.



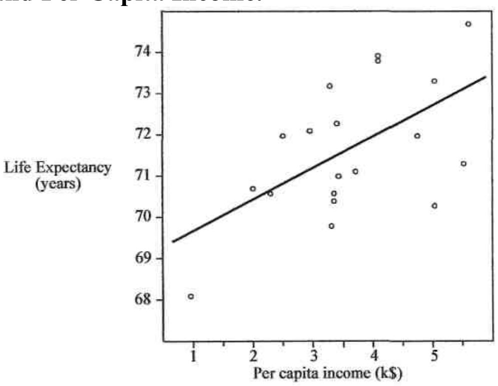
6.24 **Biological Basis for Homosexuality.** There is no evidence that the INAH3 volume is associated with cause of death (2-sided p-value = .11 from at-test for the hypothesis that $(\mu_3 + \mu_3 + \mu_4)/3 - (\mu_2 + \mu_5)/2 = 0$). The cause of death is therefore ignored in the following analysis of the *three* groups: 1: heterosexual males, 2: homosexual males, and 3: heterosexual females. There is strong evidence that homosexual males have a different mean ESJAH3 volume than heterosexual males (2-sided p-value = .0005 from a t-test for the hypothesis that $\mu_{hetero\cdot males} - \mu_{hetero\cdot males} = 0$, based on 55 degrees of freedom.) The mean for heterosexual males is estimated to be 69 mm^3/l000 greater than that for homosexual males (95% confidence interval: 32 to 106). There is also evidence that the mean for heterosexual males is different than that for heterosexual females (2-sided p-value = .02). The mean for heterosexual males is estimated to be greater by 65mm^3/l000 (95% confidence interval: 12 to 117). Finally, there is no evidence of a difference in mean volumes between homosexual males and heterosexual females (2-sided p-value = .86). A 95% confidence interval for the homosexual male mean minus the heterosexual female mean is –56to47mm^3/1000.

Chapter 7: Simple Linear Regression: A Model for the Mean

7.12 **a** 15.53
b 3.94
c 4

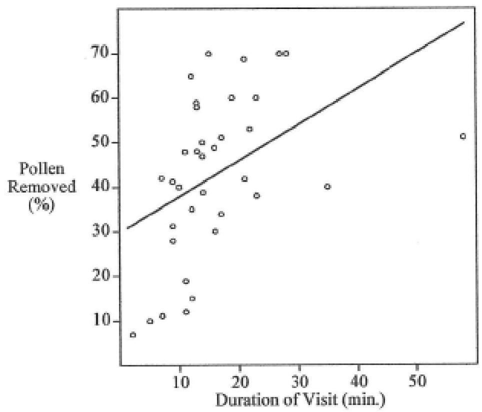
7.13 **Big Bang.** 0.1533 to 0.6449 megaparsecs.

7.14 **a** **Life Expectancy and Per Capita Income.**



b life = 68.87 + 0.00077 income, one-sided p-value = 0.0053
c 0.572.

7.15 **a** **Pollen Removal.** (in percentages, not proportions.)



b Pollen% = 29.52 + 0.8106 Duration. SE(slope) = 0.2831.
The fit looks bad. It underestimates strip means in the middle and overestimates strip means for visit of short duration.

7.16 To calculate the right hand sides of the expressions, refer to this table:

Steer	Time	Y=pH	X= log(Time)	X^2	XY
1	1	7.02	0.0000	0.0000	0.0000
2	1	6.93	0.0000	0.0000	0.0000
3	2	6.42	0.6931	0.4805	4.4500
4	2	6.51	0.6931	0.4805	4.5124
5	4	6.07	1.3863	1.9218	8.4148
6	4	5.99	1.3863	1.9218	8.3039
7	6	5.59	1.7918	3.2104	10.0159
8	6	5.80	1.7918	3.2104	10.3922
9	8	5.51	2.0794	4.3241	11.4577
10	8	5.36	2.0794	4.3241	11.1458
	SUMS:	61.20	11.9013	19.8735	68.6928

According to the formulas, the first expression is = 19.8735 − (1/10)(11.9013)^2 = 5.7094; and the second is = 68.6928 − (1/10)(11.9013)(61.20) = −4.1431. Notice that these calculations can be accomplished using only five storage registers, and only one pass through the data is required. The registers keep running totals of n, Y, X, X^2, and XY. As each steer's data is entered, the totals are updated. [Pocket calculators with statistical functions use a key marked M+ or Σ+ to enter data into the registers. They also have M- or 2- to subtract out mistakes before continuing!] Only the totals enter into the final calculations.

The left hand side expressions require a second pass through the data, because one full pass must be made to get the averages of X and Y. Therefore, either the data must be entered twice or it must be stored the first time through.

7.17 **Meat Processing.**

a Intercept estimate = 6.9836, with SE = 0.0485; slope estimate = −0.7257, with SE = 0.0344; estimate of σ = 0.0823.

b Estimate = 5.8157; SE = 0.0297.

c Same as in part b.

7.18 **Meat Processing.**

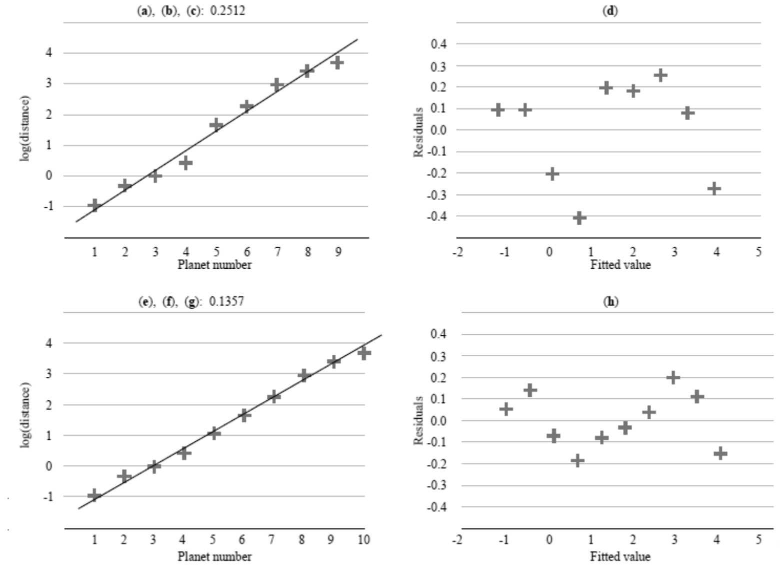
a SE{pred} = 0.0875.

b 5.6139 < mean pH < 6.0175.

7.19 **Meat Processing.** If zero were not a lower limit on time, this would be impossible. However, the predicted time should be between 0 and about 1.3 hours (from Display 7.4).

7.20 **Meat Processing.** About 109.

7.21 Planetary Distances and Order from Sun.



(i) Second set. Residual SD is smaller.

7.22 a Crab Claw Size and Force.

	H. nudus	*L. bellus*	*C. productus*
intercept	0.5191	−3.7800	−1.9673
SE(intercept)	1.1147	1.2842	0.9978
slope	0.4083	2.9737	2.0685
SE(slope)	0.5426	0.6125	0.4275
s_p	0.4825	0.4811	0.2981
d.f.	12	10	10

b *C. productus* vs. *L. bellus: t*-stat = 1.212. two-sided *p*-value = 0.24.
C. productus vs. *H. nudus: t-stdX* = 2.403. two-sided *p*-value = 0.025.

7.23 a This is straightforward. (Hopefully, you realize that differentiation is the correct method for finding the solutions.)

b There are several ways to attack this problem. The most direct way is to solve the two normal equations directly. Solve the first equation for (β_0 in terms of β_1, then substitute the answer into the second equation. Solve the second equation for β_1. A second approach is to substitue the answers into the equations, rearrange terms, and verify that the equations are satisfied.

c Again, there are different methods. One method consists of taking second derivatives and verifying that the matrix of second deriviatives is negative definite (if that is a result known to the student). A more direct method consists of noting that the normal equations have a unique solution. That solution must be a minimum, because the sum of squares is bounded below (by zero) but increases without limit the parameter values increase without limit.

7.24 **a** Confirmed.

b t-statistics: −2.07, −5.71, −4.02, −5.78. The p-values are .04, < .0001, .007, and < .0001. There seems to be real evidence of a decline.

c The standard error of the estimated coefficient is quite small, so the evidence of a non-zero slope is stronger.

d It is apparent that the standard deviation about the regression, σ, is smaller.

e The proportions are based on different numbers of births. Presumably, the proportion for the U.S. is based on a larger sample size, so there is more precision in estimating the mean proportion of male births in any given year.

DATA PROBLEMS

7.25 **Big Bang II.** The fitted equation is:

$$\text{est } \mu\{distance \mid velocity\} = 0.9537 + 0.0016430\ velocity.$$

The standard errors are: 0.5251 for the intercept estimate and 0.0000640 for the slope estimate. The estimate of σ is 1.1042, and R^2 = 98.8%. The evidence against the hypothesis that the intercept is zero comes from the t-statistic =1.82 with 8 d.f., so the two-sided *p*-value is .1069. There is no reason here to reject the simple theory.
Fitting the model with no intercept gives:

$$\text{est } \mu\{distance \mid velocity\} = 0.0017298\ velocity$$

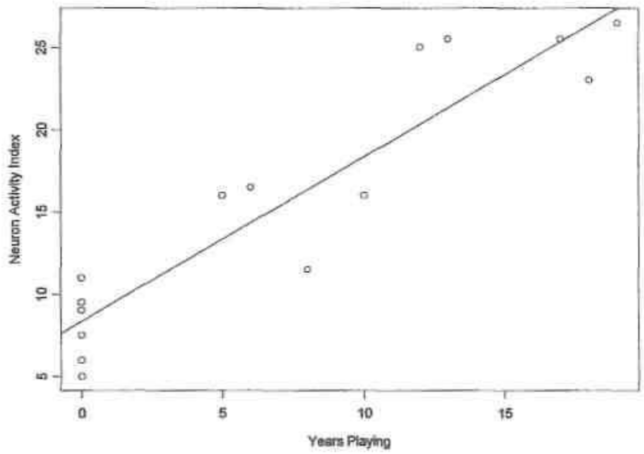
The standard error of the slope estimate is 0.0000477; the estimate of σ is 1.2372; and R^2 = 99.3%. Putting a 95% confidence interval on the slope and converting to years gives an estimate of 1.69 billion years with the interval from 1.59 billion to 1.80 billion years.

7.26 **Origin of the Term "Regression."** b_0 = 19.8261; b_1 = 0.7139, residual SD = 2.2436.
Parent height = 65 in.: Predicted child height = 66.23 in. (95% CI 61.81 to 70.65 in. Parent height = 76 in.: Predicted child height = 74.08 in. (95% CI 69.65 to 78.52 in.)

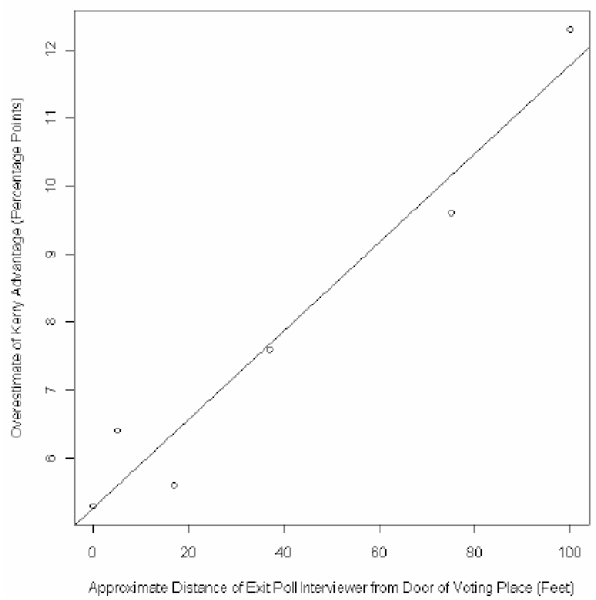
7.27 estimated mean = 0.0875 + 0.0328*MSM
(0.0787) (0.0106)

2-sided *p*-value = .0061 for the reg4ression coefficient.

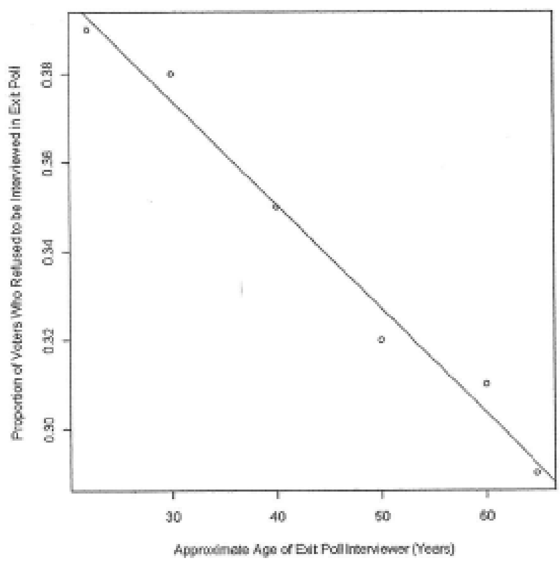
7.28 **Brain activity in violin and string players.** A two-sample t-test shows convincing evidence 7 of a difference in neuron activity index in the controls and musicians (1-sided p-value = .0001, t-test on 13 d.f.). It is estimated that the mean index for stringted musicians is 7.4 to 17.9 points higher than it is for non-musicians (95% confidence interval). As shown in the figure below, there is apparently a linear increase in the neuron activity index with increasing age. It is estimated that the mean neuronal index increases by 1.0 unit (standard error = .11) for each one-year increase in age.



7.29 **Sampling Bias in Exit Polls.** As apparent in the scatter plot below, the data provide strong evidence that the mean Kerry overestimate increases with increasing distance of interviewer from door of voting location (2-sided p-value 0.001). The mean overestimate increases by an estimate of 0.6 percentage points for each additional 10 feet of distance, for distances between 0 and 100 feet (95% confidence interval: 0.44 to 0.85 percentage points per 10 feet of distance).

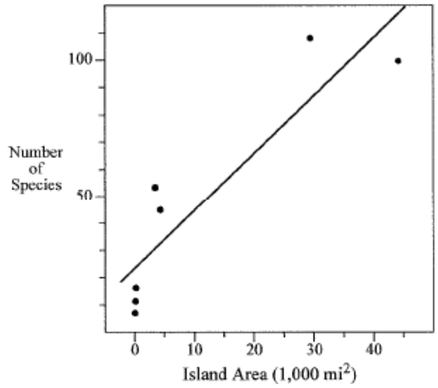


7.30 **Sampling Bias in Exit Polls 2.** The scatterplot below shows the proportion of voters who refuse to be interviewed versus the approximate age of the interviewer. There is convincing evidence that the mean refusal proportion depends on age (2-sided p-value = 0.001). The proportion who refuse decreases by an estimated 2 percentage points for each 10-year increase in age of interviewer, for ages between 20 and 65 (95% confidence interval for decrease: 1.9 to 2.8 percentage points for each 10-year increase in age).

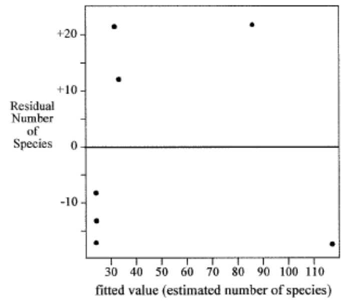


Chapter 8: A Closer Look at Assumptions for Simple Linear Regression

8.15 **a** **Island Size and Species**. Scatter plot:

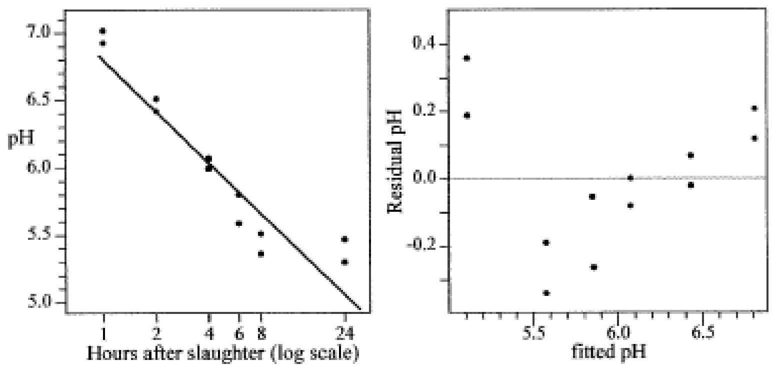


b Estimated mean number of species = 24.04928 + 0.0021 l × Area. The residual plot:



c The regression line does not come near hitting the center of the distribution of species numbers from islands with similar area. There is a pronounced curvature in the residual plot.

8.16 **a** **Meat Processing.** The fitted regression: Estimated mean pH = 6.811 − 0.535 × log(Hour). The straight line fit misses the centers of all pairs in a systematic way (see scatter plot and residual plot, below). The residual plot shows pronounced curvature.

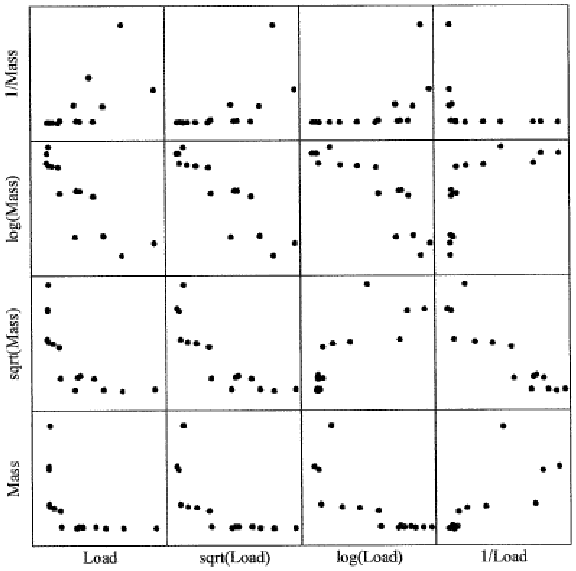


b The analysis of variance table confirms the lack of fit to a straight line.

Source of Variation	Sum of Squares	d.f.	Mean Square	*F*-Statistic	*p*-value
Between Groups	3.91638	5	0.78328	79.59	<.0001
Regression	3.51940	1	3.51940	357.60	<0001
Lack-of-fit	0.39698	4	0.09925	10.08	.0078
Within Groups	0.05905	6	0.00984		
Total	3.97543	11			

c Although the data suggest that a straight line fit is not globally correct, there is no reason to abandon a straight line approximation through the hours 1 through 8 after slaughter. The question of interest concerns when the mean pH first drops to 6.0.

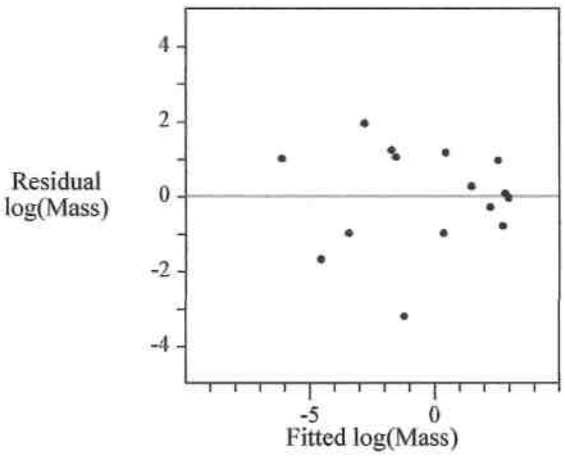
8.17 **a** **Biological Pest Control.** Here are scatter plots, on different scales for both variables:



The panel most easily described by a straight line regression has log(Mass) versus sqrt(Load).

b Estimated mean log(Mass) = 3.797 − 0.262 × sqrt(Load)

c The residual plot looks satisfactory.



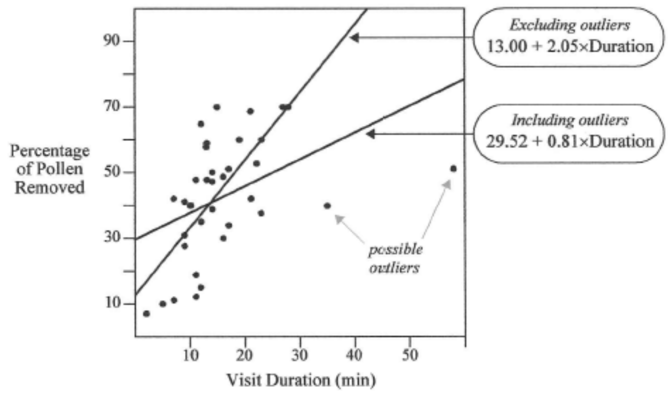
8.18 **Distance and Order from Sun.** Estimated slope in the regression of log(Distance2) on Order2: 0.5369326 (95% CI: 0.5024888 to 0.5713764). Exponeniating: 1.710751 (1.652830 to 1.770703). Subract 1 and multiply by 100 to get answer: 71% (95% CI: 65% to 77%).

8.19 **a** **Pollen Removal.** A residual plot from a fit of all data shows possible curvature and possible outliers.

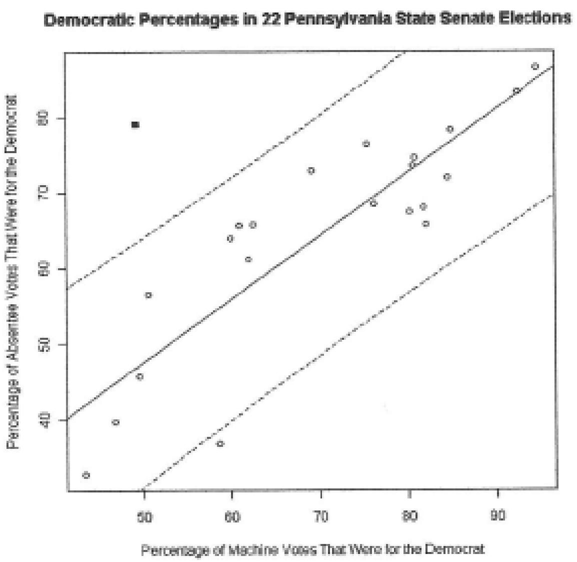
b Transformation of the scales does not accomplish much in clarifying the situation.

c Fits with and without the bees with duration over 30 seconds give quite different results for the regression. Examination of a residual plot from the fit without durations over 30 seconds shows no further problems.

Conclusion: For visits under 30 seconds, a straight line regression appears to give a reasonable summary. That description does not extend to visits over 30 seconds.



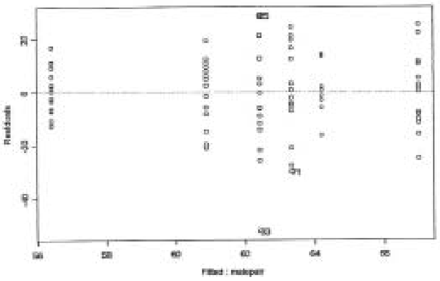
8.20 **Quantifying Evidence for Outlierness. a,b:**

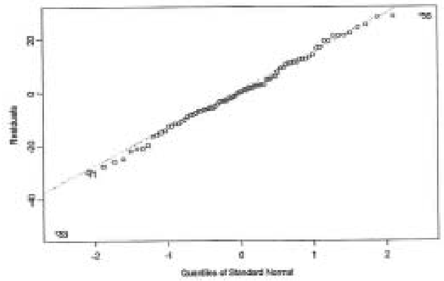


c Prediction at x = 49.3: 46.88664%; SE of prediction at x = 49.3: 7.91503; (79.0 – 46.88664)/7.91503 = 4.05. The observed percentage is 4.05 standard errors of prediction above the predicted value. The proportion of values in a *t*- distribution on 19 degrees of freedom that are as far or farther from 0 than 4.05 is 0.00067.

d Bonferonni-adjusted p-value: 22 × 0.00067 = 0.0148.

8.21 No transformation appears necessary. Case 33 and, possibly, case 35 are potential outliers.





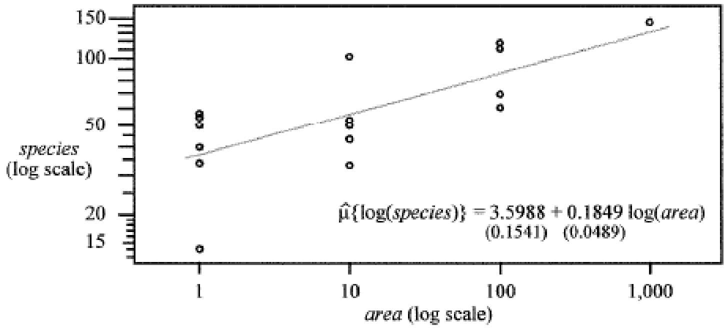
With 33 deleted: $g = -21.81$; se(g) = 50.62; t-statistic = −0.43
ANOVA p-value for group differences = .3744.
The conclusions remain the same with and without case 33.

DATA PROBLEMS

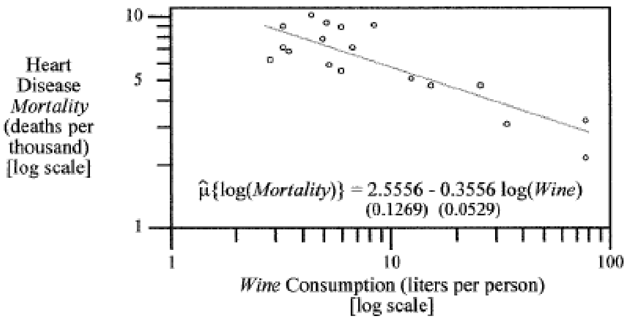
8.22 **Ecosystem Decay.** Various diagrams show that area should be transformed to its logarithm Should species also be log transformed? A scatter plot of species versus log area shows somewhat more equality of variation than does a scatter plot of log species versus log area. The goodness-of-fit of a straight line shows no significant lack of fit for either response. Here is a good place to appeal to theory, where the species-area curve is assumed to be linear on the log-log scale. The analysis of variance table summarizes the fit to 7= *log(species)* using *X*= *log(area)* as the explanatory variable.

	Sum of Squares	d.f.	Mean Square	*F*-statistic	*p-value*
Regression	2.5366	1	2.5366	12.4833	.0041
Lack-of-fit	0.0428	2	0.0214	0.1053	.9009
Within groups	2.4384	12	0.2032		
Total	5.0178	15			

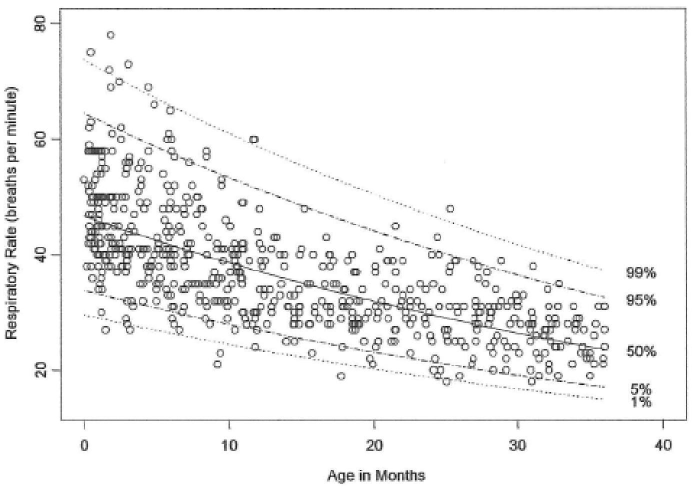
A summary of the fit:



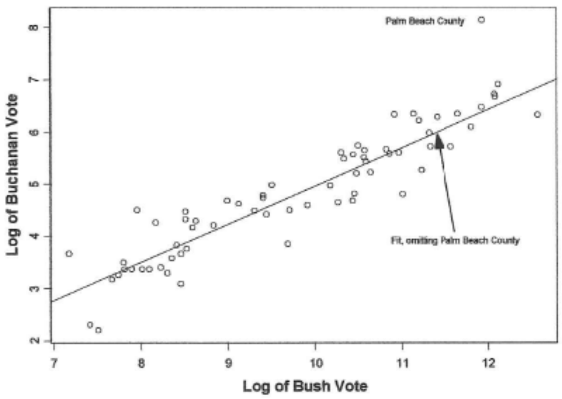
8.23 **Wine Consumption and Heart Disease.** The log-log scale provides the best looking straight line fit. There is strong evidence of an association between wine consumption and mortality — a t-statistic = 6.72 gives a two-sided p-value < .0001. Approximately 74% of the variation in log(*Mortality*) is explained by variations in log(*Wine*). An estimate of the residual standard deviation is 0.2285. This is observational data, so no causation can be inferred. Further, the countries were not randomly selected. The best wording of results would emphasize that the association could not have arisen from a random assignment of mortality numbers to wine consumption values.



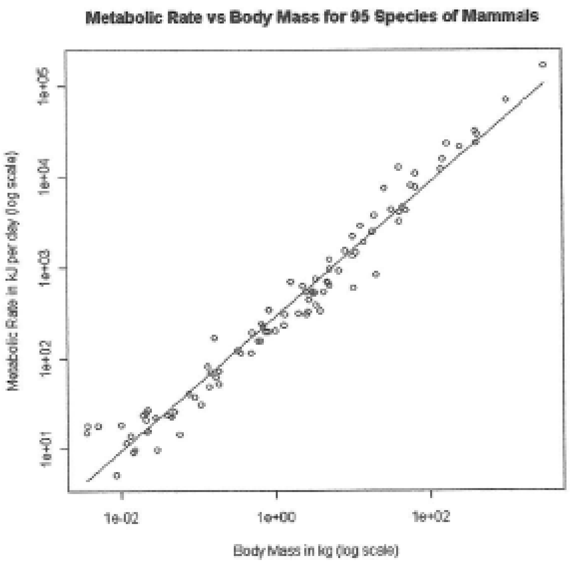
8.24 Shown below is a plot of respiratory rate versus age for the 618 children used in the study. Included on the graph are lines estimating the 1st, 5th, 50th, 95th, and 99th percentiles of respiratory rate as a function of age. These are based on a model in which the log of the respiratory rate is normal with a mean that is a straight line function of age and with a standard deviation that is independent of age. Diagnostic tools indicate this model to be adequate. The use of this plot for evaluating respiratory rate of a child (between 0 and 3 years old) hinges on that child being from the same population of children as the ones used to estimate the model.



8.25 A plot of the Buchanan vote versus the Bush vote, on the log-log scale, is shown below. On the log scale, the predicted number of votes for Buchanan in Palm Beach County, assuming the relationship with the Bush vote is the same as in other counties, is 6.38 and a 95% prediction interval is 5.53 to 7.23. This means the predicted count (not logged) for Buchanan is 592 votes and a 95% prediction interval is 252 to 1390. The actual vote count for Buchanan was 3407. The actual vote count was therefore 2815 votes larger than predicted and 2017 votes larger than the upper limit of a 95% prediction interval.

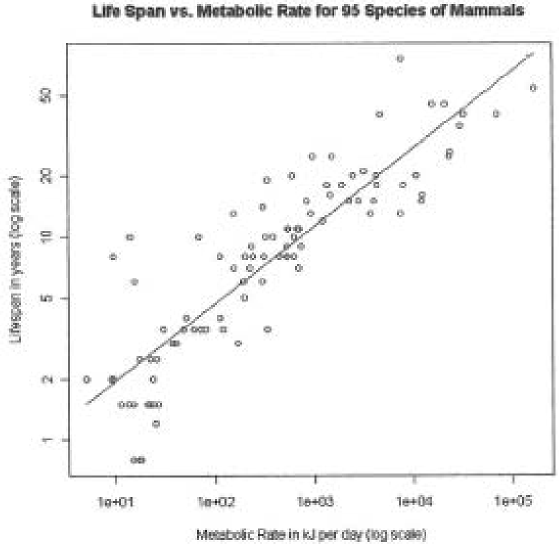


8.26 **Kleiber's Law**. As evident in the log-log plot below, the data are consistent with a model in which the median Metabolic Rate is proportional to Body Mass raised to a power [median Metabolic Rate = exp(β_0)(Body Mass)^β_1]. Furthermore, the data are consistent with the power being 0.75 (2-sided *p*-value = 0.44 for a test that the coefficient of log Body Mass is 0.75 in the linear regression of log Metabolic rate on log Mass). The power is estimated to be 0.739 (95% confidence interval: 0.710 to 0.768).

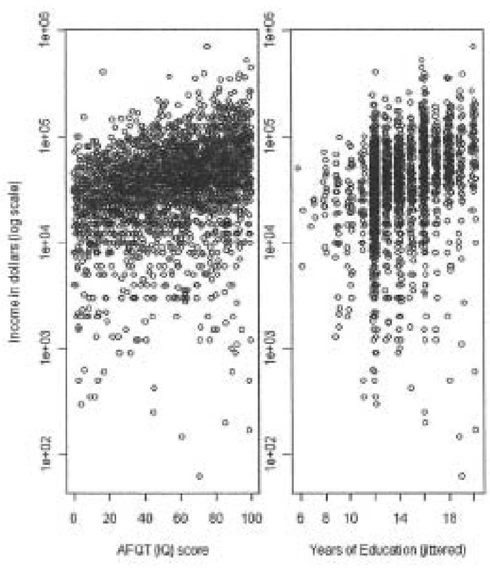


8.27 **Metabolic Rate and Lifespan**. The plot below shows that the mean log life span is linearly increasing function of the log of metabolic rate. The regression on log metabolic rate explains 78% of the variation in log life span. An equation for predicting life span is

$$(0.80782) \times (\textit{Metabolic Rate})^{.38215}$$



8.28 **IQ, Education, and Future Income.** As evident in the graphs below, the distribution of 2005 incomes increases with increasing AFQT score (1-sided p-value < 0.0001 for the coefficient of AFQT in the regression of log 2005 income on AFQT) and with increasing years of education (1-sided p-value < 0.0001 for the coefficient of Educ in the regression of log 2005 income on Educ). Median 2005 income is estimated to increase by 11.1% with each 10 percentage point increase in AFQT score (95% confidence interval: 9.6% to 12.5%) and is estimated to increase by 11.9% with each additional year of education (95% confidence interval: 10.3% to 13.5%). The regression on AFQT explains 8.9% of the variation in log incomes; the regression on education explains 8.3% of the variation.



8.29 **Autism Rates.** As evident in the graph below, the autism rate among 10-year olds increased over the period of 1992 to 2000 (2-sided p-value = 0.00004 for the coefficient of Year in the regression of log Autism Rate on Year). The autism rate was estimated to have increased by 23% per year in this period (95% confidence interval: 22% to 24%).