regeneration ratio is used as a response, the variance of the residuals is roughly constant and there is no evidence of an interaction between blocks and treatments. The interaction is investigated via the extra sum of squares F-test comparing the nonadditive to the additive model. This is most conveniently obtained with an analysis of variance procedure in a statistical computer program. Analysis of variance is then used on a fit to the additive model, to show that the treatment effects are significant. Although the subsequent analysis can be based on tests and confidence intervals for appropriate linear combinations of treatment means (Section 13.3.4), it is more easily couched in terms of a regression model with indicator variables (Section 13.3.5). In particular, indicator variables for the presence of large fish, small fish, and limpets permit the analyst to obtain confidence intervals and tests that can be directly reported in the summary of statistical findings.

Pygmalion Effect

Since the F-test for interaction between companies (blocks) and treatment is insignificant, the additive model is used for inferences. A single term describes the additional average score of a pygmalion-treated platoon over a control platoon. Section 13.4.4 demonstrates, for this example, how a p-value is tied directly to the chance involved in the randomization process. Although this is a very good approach, it is typically easier and quite adequate to use the usual regression approach as an approximation.

13.7 EXERCISES

Conceptual Exercises

1. Seaweed. If, instead of the five treatment variables in the regression formulation of Section 13.3.5, the control treatment is used as a reference and five indicators for the other five treatments (grazer combinations) are defined, how would the results be different?

2. Seaweed. Display 13.11 shows that R^2 from the fit to the nonadditive model is 92.84%. R^2 from the additive model is 85.34%. Does this indicate that the nonadditive model is better?

3. Pygmalion. Why are the average scores of the platoon used as the response variable, rather than the scores of the individual soldiers?

4. Why is there so little interest in how the mean response is associated with blocks in a randomized block experiment?

5. Why is balance important?

6. What does it mean when there are significant interactions but no significant main effects?

7. If the F-test for treatments is not significant but the t-test for one of the contrasts is significant, is it proper to report the contrast?

8. Is it possible to examine whether an additive model is appropriate in cases where there is only one observation in each cell?

9. Seaweed. Why is the residual SD from the additive model a better choice than the residual SD from the full interactive model for calculating standard errors for the seaweed study linear combinations (Section 13.3.4)?

10. What is the difference between a randomized block design and a completely randomized design with factorial treatment structure? How might the analyses differ?

Computational Exercises

11. Pygmalion. Analyze the Pygmalion data using the regression approach described in Section 13.4. (a) Obtain the p-value for the test that interactions are zero. (b) Obtain a p-value, from the additive model, for testing the hypothesis that the treatment effect is zero.

12. Seaweed. (a) Calculate the sample variance of the six treatment averages in Display 13.13. Multiply this by 16, and verify that the result equals the Treatments Mean Square in Display 13.11. (b) Verify that the Blocks Mean Square in Display 13.11 equals 12 times the sample variance of the block averages from Display 13.12. (c) Verify that the Between Groups Mean Square (i.e., the Model Mean Square) in Display 13.10 is 2 times the sample variance of the 48 cell averages from Display 13.12. (*Note*: *The operative rule here is that the sample variance of a set of averages gets multiplied by the number of observations going into each average.*) (d) You can easily work backward from the mean squares to the sums of squares. Then verify that the Interactions Sum of Squares is the difference between the Between Groups Sum of Squares and the sum of the Blocks and Treatments Sum of Squares.

13. Seaweed. Carry out the analysis of the seaweed data with multiple linear regression, as outlined in Section 13.3.5. Using the regression fit, obtain a set of fitted values and residuals. Construct a residual plot. Take the antilogarithms of the fitted values to get estimated medians of percent regeneration scale and construct a plot of these versus block number, ordered as in Display 13.7.

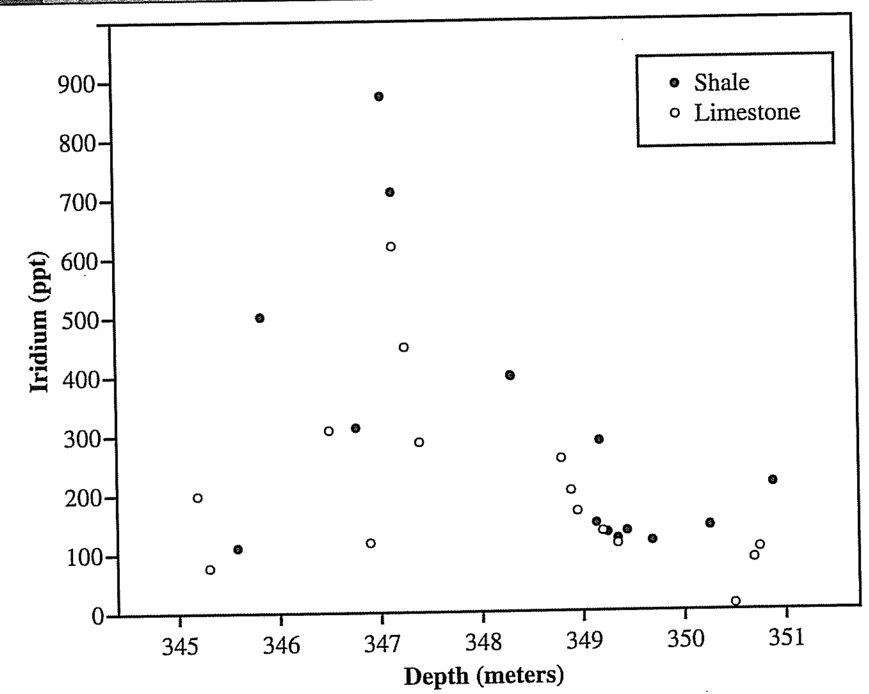
14. Seaweed. Using the fit to the regression model described in Section 13.3.5, estimate the mean log regeneration ratio on plots in the reference block with both small and large fish allowed but limpets excluded. Back-transform the estimate and word a statement about the median regeneration ratio.

15. Blood–Brain Barrier. Analyze the effect of the design, variables—sacrifice time and treatment—on the log of the ratio of brain count to liver count in the data set described in Section 11.1.2 (file: case1102). (a) Ignore the covariates and use an analysis of variance procedure to fit the data. Fit a model that includes interaction terms; plot the residuals versus the fitted values. (b) Test whether there is an interactive effect of treatment and sacrifice time. What are the F-statistic, the degrees of freedom, and the p-value? (c) If there are no interactive effects, test whether there are main effects of treatment and sacrifice time. (d) Complete the analysis by describing the effects of treatment and sacrifice time, either by estimating the appropriate contrasts or by using a regression procedure with indicator variables to model treatment (one indicator) and sacrifice time (three indicators).

16. Toxic Effects of Copper and Zinc. Reconsider the data in Display 10.20 (file: ex1014). (a) Is the experiment a randomized block design or a completely randomized design with factorial treatment structure? (b) Analyze the data using two-way analysis of variance. (c) How does the analysis of variance compare to the regression analysis (i.e., what are the issues in deciding which analysis is more appropriate)?

17. Dinosaur Extinctions—An Observational Study. About 65 million years ago, the dinosaurs suffered a mass extinction virtually overnight (in geologic time). What happened in this period, the Cretaceous–Tertiary (KT) boundary, that could have produced such calamity? Among many clues, one that all scientists regard as crucial is a layer of iridium-rich dust that was deposited over much of the earth at that time. Data from one of the first discoveries of this layer are shown in Display 13.22. The diagram traces iridium concentrations in soil samples taken from extensive shale and limestone

DISPLAY 13.22 Iridium trace at Gubbio, Italy



deposits at Gubbio, Italy. Iridium (parts per trillium) is graphed against the depth at which samples were taken, with depth giving a crude measure of historic time.

Iridium is a trace element present in most soils. But concentrations as high as those at the peak near 347 meters, at the KT boundary, can only occur in association with highly unusual events, such as volcanic eruptions or meteor impacts. So the theory is that such an event caused a massive dust cloud that blanketed the earth for years, killing off animals and their food sources. But was the cause a meteor (coming down on the Yucatan peninsula in central America) or volcanic eruptions (centered in southern China)? Two articles debating the issue appeared in the October 1990 issue of *Scientific American*—W. Alvarez and F. Asaro, "What Caused the Mass Extinction? An Extraterrestrial Impact," *Scientific American* 263(4): 76–84, and E. Courtillot, "What Caused the Mass Extinction? A Volcanic Eruption," *Scientific American* 263(4): 85–93. A crucial issue in the debate is the shape of the iridium trace because the timing and extent of the source give clues to its nature.

The raw data are provided in Display 13.23. (a) Fit the two-way, saturated model to the untransformed data and draw a plot of the residuals versus estimated means to see if a transformation is suggested. (b) Fit the two-way model (after transformation if appropriate) and test for interaction, using a multiple regression routine. (c) If appropriate with your statistical software, repeat part (b) using an analysis of variance routine.

18. El Niño and Hurricanes. Reconsider the El Niño and Hurricane data set from Exercise 10.28. (a) Regress the log of the storm index on West African wetness (treated as a categorical factor with 2 levels) and El Niño temperature (treated as a categorical factor with 3 levels); retain the sum of squared residuals and the residual degrees of freedom. (b) Regress the log of the storm index on West African wetness (treated as categorical with 2 levels), El Niño *temperature* (treated as numerical),

DISPLAY 13.23 Iridium (ppt) in samples from two strata at Gubbio, Italy, by depth category

	Depth category (meters)					
	345–346	346–347	347–348	348–349	349–350	350–351
Limestone	75 200	120 310	290 450 620	170 205 260	120 135	5 90 105
Shale	110 501	315	710 875	400	120 130 135 150 290	145 215

and the square of El Niño *temperature*. Retain the sum of squared residuals and the residual degrees of freedom. (c) Explain why the answers to (a) and (b) are the same. (d) Explain why a test that the coefficient of the temperature-squared term is zero can be used to help decide whether to treat temperature as numerical or categorical.

Data Problems

19. Nature–Nurture. A 1989 study investigated the effect of heredity and environment on intelligence. From adoption registers in France, researchers selected samples of adopted children whose biological parents and adoptive parents came from either the very highest or the very lowest socioeconomic status (SES) categories (based on years of education and occupation). They attempted to obtain samples of size 10 from each combination: (1) high adoptive SES and high biological SES, (2) high adoptive SES and low biological SES, (3) low adoptive SES and high biological SES, and (4) low SES for both parents. It turned out, however, only eight children belonged to combination three. The 38 selected children were given intelligence quotient (IQ) tests. The scores are reported in Display 13.24. (Data from C. Capron and M. Duyme, "Children's IQs and SES of Biological and Adoptive Parents in a Balanced Cross-fostering Study," *European Bulletin of Cognitive Psychology* 11(3) (1991): 323–48.) Does the difference in mean scores for those with high and low SES biological parents depend on whether the adoptive parents were high or low SES? If not, how much is the mean IQ score affected by the SES of adoptive parents, and how much is it affected by the SES of the biological parents? Is one of these effects larger than the other? Analyze the data and write a report of the findings.

20. Gender Differences in Performance on Mathematics Achievement Tests. Display 13.25 shows a sample of five rows from a data set on 861 ACT Assessment Mathematics Usage Test scores from 1987. The test was given to a sample of high school seniors who met one of three profiles of high school mathematics course work: (a) Algebra I only; (b) two Algebra courses and Geometry; and (c) two Algebra courses, Geometry, Trigonometry, Advanced Mathematics, and Beginning Calculus. Analyze the data to determine whether male scores are distributed differently than female scores, after accounting for coursework profile, and whether the difference is the same for all profiles. Write a brief summary of statistical findings and include a graphical display. (These data were generated from summary statistics for one particular form of the test, as reported in A. E. Doolittle, "Gender Differences in Performance on Mathematics Achievement Items," *Applied Measurement in Education*, 2 (2) 1989: 161–77.)

DISPLAY 13.24 IQ scores for adopted children whose biological and adoptive parents were categorized either in the highest or the lowest socioeconomic status category

SES of adoptive parents	SES of biological parents	IQ scores of adopted children									
High	High	136	99	121	133	125	131	103	115	116	117
High	Low	94	103	99	125	111	93	101	94	125	91
Low	High	98	99	91	124	100	116	113	119		
Low	Low	92	91	98	83	99	68	76	115	86	116

DISPLAY 13.25 ACT mathematics test scores for a set of high school seniors having one of three mathematics course profiles; 5 sample rows from 861

Gender	Profile	Score
male	a	5
male	b	13
female	a	18
female	c	28
male	c	32

21. Pygmalion. The term *Pygmalion effect* (used in the Section 13.1.2 data problem) originated with the 1960s' work of psychologist Robert Rosenthal and elementary school principal Lenore Jacobson, who conducted a randomized block experiment on the children in Jacobson's elementary school. In each classroom (i.e., for each block of students), they assigned roughly 20% of the students to a "likely to excel" treatment group and the remainder to a control group. After all students took a standardized test at the begining of the school year, Rosenthal and Jacobson identified to teachers those students that had been indicated by the intelligence test as very likely to excel in the coming year. This was false information, though. The "likely to excel" students were simply those who had been assigned by a chance mechanism to the "likely to excel" treatment group. The researchers deceived the teachers with this false information to create artificial expectations and to explore the effect of those expectations on student test score gains. Display 13.26 is a partial listing of a data set simulated to match the summary statistics and conclusions from Table A-7 of Rosenthal and Jacobson's report (R. Rosenthal, and L. Jacobson, 1968, *Pygmalion in the Classroom: Teacher Expectation and Pupil's Intellectual Development*, Holt, Rinehart, and Winston, Inc.). Analyze the data to summarize the evidence that teachers' expectations affect student test score gains. Write a statistical report of your conclusions.

22. Pygmalion; Mixed Effects Model. The blocks in the Pygmalion study of Excercise 21 are the different classrooms in the school. To model possible differences among the 17 classrooms (such as the effects of the different classroom teachers) as a categorical factor, 16 parameters are needed. Some researchers would prefer to think of the classrroom effects as random, which would mean acting as if the 17 teachers were a random sample from some larger population of teachers. The treatment effect would be handled in the usual way and would be referred to as a "fixed effect." A model that includes both fixed and random effects is called a *mixed effects model*. Analyze the data of Exercise 21 but treat the block effects as random by using a mixed linear model routine in a statistical computer program.

DISPLAY 13.26 Results of a randomized block experiment to see whether teacher's expectations affect student's intelligence test score gains, with Grade (1 through 6), Class (to indicate distinct classes at each grade), Treatment (either control or Pygmalion), and Gain (standardized test score at end of year minus score on test at beginning of year); first 5 of 320 rows

Student	Grade	Class	Treatment	Gain
1	1	1a	control	−4
2	1	1a	control	0
3	1	1a	control	−19
4	1	1a	control	24
5	1	1a	control	19

Answers to Conceptual Exercises

1. This would be almost exactly like the analysis of Section 13.3.5. Instead of measuring interesting contrasts, the coefficients would measure effects of applying treatment combinations. To tease out the separate effects of each grazer, it would be necessary to consider linear combinations of the regression coefficients. (Notice how the judicious choice of indicator variables in Section 13.3.5 avoided the messy contrast approach.)

2. No. As usual, R^2 is always larger for a model with more terms, whether or not those terms are significantly different from zero.

3. The experimental unit—that which receives the treatment—is the platoon, not the individual soldier.

4. The association is not an issue. Blocks are formed to provide similar units on which to compare treatments. Also, the randomization is not conducive to drawing inferences about block differences.

5. The simple row and column means in a balanced table produce simple estimates of model parameters. To see that the treatment effects are not influenced by block differences, try adding a certain fixed amount to all responses in one block. It will not change any of the treatment differences in a balanced table. But it will in an unbalanced table.

6. It simply means that there are significant interactions. When there is nonadditivity, the best strategy is to avoid talking about main effects.

7. If the contrast was planned in advance, report it. If it is a comparison that the data suggested, you should account for the data snooping when attaching any significance to it.

8. The analysis of variance table will have no within-group entry if there is only one observation in each cell. So there will be no formal analysis of variance test for nonadditivity. But the answer to the question is yes, and the next chapter will discuss how.

9. If there are interactions, there is no sense in estimating the linear combinations.

10. In the randomized block design the random allocation of treatment levels to experimental units is conducted separately for each block. In a completely randomized design the treatment levels are randomly allocated to all experimental units at once, but the levels correspond to all possible combinations of the levels of two distinct treatments. In the randomized block data there is generally no interaction expected and there is little interest in the block effects. In the design with factorial treatment arrangement there may be interest in an interactive effect and there is typically interest in the effects of both treatments.