of units:

i:	1	2	3	4	5	6	7	8	9	10
X_i:	0	1	2	3	4	5	6	7	8	9
Y_i:	98	135	162	178	221	232	283	300	374	395

a. Prepare a scatter plot of the data. Does a linear relation appear adequate here?

b. Use the Box-Cox procedure and standardization (3.36) to find an appropriate power transformation of Y. Evaluate SSE for $\lambda = .3, .4, .5, .6, .7$. What transformation of Y is suggested?

c. Use the transformation $Y' = \sqrt{Y}$ and obtain the estimated linear regression function for the transformed data.

d. Plot the estimated regression line and the transformed data. Does the regression line appear to be a good fit to the transformed data?

e. Obtain the residuals and plot them against the fitted values. Also prepare a normal probability plot. What do your plots show?

f. Express the estimated regression function in the original units.

3.18. **Production time.** In a manufacturing study, the production times for 111 recent production runs were obtained. The table below lists for each run the production time in hours (Y) and the production lot size (X).

i:	1	2	3	...	109	110	111
X_i:	15	9	7	...	12	9	15
Y_i:	14.28	8.80	12.49	...	16.37	11.45	15.78

a. Prepare a scatter plot of the data. Does a linear relation appear adequate here? Would a transformation on X or Y be more appropriate here? Why?

b. Use the transformation $X' = \sqrt{X}$ and obtain the estimated linear regression function for the transformed data.

c. Plot the estimated regression line and the transformed data. Does the regression line appear to be a good fit to the transformed data?

d. Obtain the residuals and plot them against the fitted values. Also prepare a normal probability plot. What do your plots show?

e. Express the estimated regression function in the original units.

Exercises

3.19. A student fitted a linear regression function for a class assignment. The student plotted the residuals e_i against Y_i and found a positive relation. When the residuals were plotted against the fitted values $\hat{Y}_i$, the student found no relation. How could this difference arise? Which is the more meaningful plot?

3.20. If the error terms in a regression model are independent $N(0, \sigma^2)$, what can be said about the error terms after transformation $X' = 1/X$ is used? Is the situation the same after transformation $Y' = 1/Y$ is used?

3.21. Derive the result in (3.29).

3.22. Using (A.70), (A.41), and (A.42), show that $E\{MSPE\} = \sigma^2$ for normal error regression model (2.1).

3.23. A linear regression model with intercept $\beta_0 = 0$ is under consideration. Data have been obtained that contain replications. State the full and reduced models for testing the appropriateness of the regression function under consideration. What are the degrees of freedom associated with the full and reduced models if $n = 20$ and $c = 10$?

Projects

3.24. **Blood pressure.** The following data were obtained in a study of the relation between diastolic blood pressure (Y) and age (X) for boys 5 to 13 years old.

i:	1	2	3	4	5	6	7	8
X_i:	5	8	11	7	13	12	12	6
Y_i:	63	67	74	64	75	69	90	60

a. Assuming normal error regression model (2.1) is appropriate, obtain the estimated regression function and plot the residuals e_i against X_i. What does your residual plot show?

b. Omit case 7 from the data and obtain the estimated regression function based on the remaining seven cases. Compare this estimated regression function to that obtained in part (a). What can you conclude about the effect of case 7?

c. Using your fitted regression function in part (b), obtain a 99 percent prediction interval for a new Y observation at $X = 12$. Does observation Y_7 fall outside this prediction interval? What is the significance of this?

3.25. Refer to the **CDI** data set in Appendix C.2 and Project 1.43. For each of the three fitted regression models, obtain the residuals and prepare a residual plot against X and a normal probability plot. Summarize your conclusions. Is linear regression model (2.1) more appropriate in one case than in the others?

3.26. Refer to the **CDI** data set in Appendix C.2 and Project 1.44. For each geographic region, obtain the residuals and prepare a residual plot against X and a normal probability plot. Do the four regions appear to have similar error variances? What other conclusions do you draw from your plots?

3.27. Refer to the **SENIC** data set in Appendix C.1 and Project 1.45.

a. For each of the three fitted regression models, obtain the residuals and prepare a residual plot against X and a normal probability plot. Summarize your conclusions. Is linear regression model (2.1) more apt in one case than in the others?

b. Obtain the fitted regression function for the relation between length of stay and infection risk after deleting cases 47 ($X_{47} = 6.5$, $Y_{47} = 19.56$) and 112 ($X_{112} = 5.9$, $Y_{112} = 17.94$). From this fitted regression function obtain separate 95 percent prediction intervals for new Y observations at $X = 6.5$ and $X = 5.9$, respectively. Do observations Y_{47} and Y_{112} fall outside these prediction intervals? Discuss the significance of this.

3.28. Refer to the **SENIC** data set in Appendix C.1 and Project 1.46. For each geographic region, obtain the residuals and prepare a residual plot against X and a normal probability plot. Do the four regions appear to have similar error variances? What other conclusions do you draw from your plots?

3.29. Refer to **Copier maintenance** Problem 1.20.

a. Divide the data into four bands according to the number of copiers serviced (X). Band 1 ranges from $X = .5$ to $X = 2.5$; band 2 ranges from $X = 2.5$ to $X = 4.5$; and so forth. Determine the median value of X and the median value of Y in each of the bands and develop

Cited References

4.1. Miller, R. G., Jr. *Simultaneous Statistical Inference.* 2nd ed. New York: Springer-Verlag, 1991.
4.2. Fuller, W. A. *Measurement Error Models.* New York: John Wiley & Sons, 1987.
4.3. Berkson, J. "Are There Two Regressions?" *Journal of the American Statistical Association* 45 (1950), pp. 164–80.
4.4. Cox, D. R. *Planning of Experiments.* New York: John Wiley & Sons, 1958, pp. 141–42.

Problems

4.1. When joint confidence intervals for β_0 and β_1 are developed by the Bonferroni method with a family confidence coefficient of 90 percent, does this imply that 10 percent of the time the confidence interval for β_0 will be incorrect? That 5 percent of the time the confidence interval for β_0 will be incorrect and 5 percent of the time that for β_1 will be incorrect? Discuss.

4.2. Refer to Problem 2.1. Suppose the student combines the two confidence intervals into a confidence set. What can you say about the family confidence coefficient for this set?

*4.3. Refer to **Copier maintenance** Problem 1.20.

a. Will b_0 and b_1 tend to err in the same direction or in opposite directions here? Explain.

b. Obtain Bonferroni joint confidence intervals for β_0 and β_1, using a 95 percent family confidence coefficient.

c. A consultant has suggested that β_0 should be 0 and β_1 should equal 14.0. Do your joint confidence intervals in part (b) support this view?

*4.4. Refer to **Airfreight breakage** Problem 1.21.

a. Will b_0 and b_1 tend to err in the same direction or in opposite directions here? Explain.

b. Obtain Bonferroni joint confidence intervals for β_0 and β_1, using a 99 percent family confidence coefficient. Interpret your confidence intervals.

4.5. Refer to **Plastic hardness** Problem 1.22.

a. Obtain Bonferroni joint confidence intervals for β_0 and β_1, using a 90 percent family confidence coefficient. Interpret your confidence intervals.

b. Are b_0 and b_1 positively or negatively correlated here? Is this reflected in your joint confidence intervals in part (a)?

c. What is the meaning of the family confidence coefficient in part (a)?

*4.6. Refer to **Muscle mass** Problem 1.27.

a. Obtain Bonferroni joint confidence intervals for β_0 and β_1, using a 99 percent family confidence coefficient. Interpret your confidence intervals.

b. Will b_0 and b_1 tend to err in the same direction or in opposite directions here? Explain.

c. A researcher has suggested that β_0 should equal approximately 160 and that β_1 should be between -1.9 and -1.5. Do the joint confidence intervals in part (a) support this expectation?

*4.7. Refer to **Copier maintenance** Problem 1.20.

a. Estimate the expected number of minutes spent when there are 3, 5, and 7 copiers to be serviced, respectively. Use interval estimates with a 90 percent family confidence coefficient based on the Working-Hotelling procedure.

b. Two service calls for preventive maintenance are scheduled in which the numbers of copiers to be serviced are 4 and 7, respectively. A family of prediction intervals for the times to be spent on these calls is desired with a 90 percent family confidence coefficient. Which procedure, Scheffé or Bonferroni, will provide tighter prediction limits here?

c. Obtain the family of prediction intervals required in part (b), using the more efficient procedure.

a. Obtain a 90 percent confidence interval for the student's ACT test score. Interpret your confidence interval.

b. Is criterion (4.33) as to the appropriateness of the approximate confidence interval met here?

4.20. Refer to **Plastic hardness** Problem 1.22. The measurement of a new test item showed 238 Brinell units of hardness.

a. Obtain a 99 percent confidence interval for the elapsed time before the hardness was measured. Interpret your confidence interval.

b. Is criterion (4.33) as to the appropriateness of the approximate confidence interval met here?

Exercises

4.21. When the predictor variable is so coded that $\bar{X} = 0$ and the normal error regression model (2.1) applies, are b_0 and b_1 independent? Are the joint confidence intervals for β_0 and β_1 then independent?

4.22. Derive an extension of the Bonferroni inequality (4.2a) for the case of three statements, each with statement confidence coefficient $1 - \alpha$.

4.23. Show that for the fitted least squares regression line through the origin (4.15), $\sum X_i e_i = 0$.

4.24. Show that $\hat{Y}$ as defined in (4.15) for linear regression through the origin is an unbiased estimator of $E\{Y\}$.

4.25. Derive the formula for $s^2\{\hat{Y}_h\}$ given in Table 4.1 for linear regression through the origin.

Projects

4.26. Refer to the **CDI** data set in Appendix C.2 and Project 1.43. Consider the regression relation of number of active physicians to total population.

a. Obtain Bonferroni joint confidence intervals for β_0 and β_1, using a 95 percent family confidence coefficient.

b. An investigator has suggested that β_0 should be -100 and β_1 should be .0028. Do the joint confidence intervals in part (a) support this view? Discuss.

c. It is desired to estimate the expected number of active physicians for counties with total population of $X = 500$, 1,000, 5,000 thousands with family confidence coefficient .90. Which procedure, the Working-Hotelling or the Bonferroni, is more efficient here?

d. Obtain the family of interval estimates required in part (c), using the more efficient procedure. Interpret your confidence intervals.

4.27. Refer to the **SENIC** data set in Appendix C.1 and Project 1.45. Consider the regression relation of average length of stay to infection risk.

a. Obtain Bonferroni joint confidence intervals for β_0 and β_1, using a 90 percent family confidence coefficient.

b. A researcher suggested that β_0 should be approximately 7 and β_1 should be approximately 1. Do the joint intervals in part (a) support this expectation? Discuss.

c. It is desired to estimate the expected hospital stay for persons with infection risks $X = 2, 3, 4, 5$ with family confidence coefficient .95. Which procedure, the Working-Hotelling or the Bonferroni, is more efficient here?

d. Obtain the family of interval estimates required in part (c), using the more efficient procedure. Interpret your confidence intervals.