As for most statistical analyses, it is important to verify that the assumptions underlying the model are fulfilled. Of special importance are the assumptions of proper model specification, homogeneous variance, and lack of outliers. In regression, this can be accomplished by examining the residuals. Additional methods are provided in Chapter 8.

7.10 CHAPTER EXERCISES

Concept Questions

For the following true/false statements regarding concepts and uses of simple linear regression analysis, indicate whether the statement is true or false and specify what will correct a false statement.

1. _____ The need for a nonlinear regression can only be determined by a lack of fit test.
2. _____ The correlation coefficient indicates the change in y associated with a unit change in x.
3. _____ To conduct a valid regression analysis, both x and y must be approximately normally distributed.
4. _____ Rejecting the null hypothesis of no linear regression implies that changes in x cause changes in y.
5. _____ In linear regression we may extrapolate without danger.
6. _____ If x and y are uncorrelated in the population, the expected value of the estimated linear regression coefficient (slope) is zero.
7. _____ If the true regression of y on x is curvilinear, a linear regression still provides a good approximation to that relationship.
8. _____ The x values must be randomly selected in order to use a regression analysis.
9. _____ The error or residual sum of squares is the numerator portion of the formula for the variance of y about the regression line.
10. _____ The term $\hat{\mu}_{y|x}$ serves as the point estimate for estimating both the mean and individual prediction of y for a given x.
11. _____ Useful prediction intervals for y can be obtained from a regression analysis.
12. _____ In a regression analysis, the estimated mean of the distribution of y is the sample mean ($\bar{y}$).
13. _____ All data points will fit the regression line exactly if the sample correlation is either $+1$ or -1.
14. _____ The prediction interval for y is widest when x is at its mean.

15. _____ The standard error of the estimated slope of a regression model becomes larger as the dispersion of x increases.

16. _____ When there is no linear relationship between two variables, a horizontal regression line best describes the relationship.

17. _____ If $r > 0$, then as x increases, y tends to increase.

18. _____ If a regression line is computed for data where x ranges from 0 to 30, you may safely predict y for $x = 40$.

19. _____ The correlation coefficient can be used to detect any relationship between two variables.

20. _____ If r is very close to either $+1$ or -1, then there is a cause and effect relationship between x and y.

Exercises

Note: Exercises 1 through 5 contain very few observations and are suitable for manual computation, which can be checked against computer outputs. The remainder of the problems are best performed by a computer.

1. The data of Table 7.14 represent the thickness of oxidation on a metal alloy for different settings of temperature in a curing oven. The values of temperature have been coded so that zero is the "normal" temperature, which makes manual computation easier.
   (a) Calculate the estimated regression line to predict oxidation based on temperature. Explain the meaning of the coefficients and the variance of residuals.
   (b) Calculate the estimated oxidation thickness for each of the temperatures in the experiment.
   (c) Calculate the residuals and make a residual plot. Discuss the distribution of residuals.
   (d) Test the hypothesis that $\beta_1 = 0$, using both the analysis of variance and t tests.

Data for Hand Calculations Problem →

Table 7.14 Data for Exercise 1

Oxidation	Temperature
4	−2
3	−1
3	0
2	1
2	2

2. The data of Table 7.
   numbers of days afte
   (a) Obtain the estim
   number of days a
   (b) Calculate and pl
   fault in the mode

3. The grades for 15 st
   course are given in T
   (a) Obtain the least-
   ination from the
   regression and in
   (b) It is suggested th
   a final examinati
   intervals.)
   (c) Plot the estimate
   (d) Predict the final s
   Check this calcul
   (e) Compute r and
   squares in part (

4. Given the values in
   variable y:
   (a) Perform the linea
   (b) Note that half o
   Does this sugges
   compare results

5. It is generally believe
   they are better able t
   of a sample of 25 non
   made in a 60-s time p
   (a) Perform a regres
   such a relationsh
   (b) Estimate the nu
   How much confid

Data for Computer Calculations Problem (but note code for entering data is given on course web page) →

Table 7.20 Data for Exercise 7: Heating Costs

Mo	Day	Tavg	Kwh	Mo	Day	Tavg	Kwh
9	19	77.5	45	10	13	68.0	50
9	20	80.0	73	10	14	66.5	37
9	21	78.0	43	10	15	69.0	43
9	22	78.5	61	10	16	70.5	42
9	23	77.5	52	10	17	63.0	25
9	24	83.0	56	10	18	64.0	31
9	25	83.5	70	10	19	64.5	31
9	26	81.5	69	10	20	65.0	32
9	27	75.5	53	10	21	66.5	35
9	28	69.5	51	10	22	67.0	32
9	29	70.0	39	10	23	66.5	34
9	30	73.5	55	10	24	67.5	35
10	1	77.5	55	10	25	75.0	41
10	2	79.0	57	10	26	75.5	51
10	3	80.0	68	10	27	71.5	34
10	4	79.0	73	10	28	63.0	19
10	5	76.0	57	10	29	60.0	19
10	6	76.0	51	10	30	64.0	30
10	7	75.5	55	10	31	62.5	23
10	8	79.5	56	11	1	63.5	35
10	9	78.5	72	11	2	73.5	29
10	10	82.0	73	11	3	68.0	55
10	11	71.5	69	11	4	77.5	56
10	12	70.0	38				

the 48 contiguous states. The data are given in Table 7.21, where Teen denotes the birthrate for teenage mothers and Mort denotes the infant mortality rate.

(a) Perform a regression to estimate Mort using Teen as the independent variable. Do the results confirm the stated hypothesis? Interpret the results.

(b) Construct a residual plot. Comment on the results.

10. In Exercise 13 of Chapter 1, the half-life of aminoglycosides was measured on 43 patients given either Amikacin or Gentamicin. The data are reproduced in different form in Table 7.22.

(a) Perform a regression to estimate Half-Life using DO_MG_KG for each type of drug separately. Do the drugs seem to have parallel regression lines? (A formal test for parallelism is presented in Chapter 11.)

(b) Perform the appropriate inferences on both lines to determine whether the relationship between half-life and dosage is significant. Use $\alpha = 0.05$. Completely explain your results.

Table 7.21 Data

State	Teen	Mo
AL	17.4	13.
AR	19.0	10.
AZ	13.8	9.
CA	10.9	8
CO	10.2	8.
CT	8.8	9.
DE	13.2	11.
FL	13.8	11.
GA	17.0	12.
IA	9.2	8
ID	10.8	11.
IL	12.5	12.
IN	14.0	11.
KS	11.5	8.
KY	17.4	9.
LA	16.8	11.

(c) Draw a scatter diag drug (use A's and G same graph.

11. An experimenter is test known to be accurate) by lb/in.^{2}. The purpose of th of the new gauge. The d

As we saw in Example in determining the effect priate analysis to determ device has a shortcomin analyses to find the shor

12. A research article states vey questionnaire) is re traffic, decayed building

Satisfaction with pc social disorder incre

(a) Calculate the F test the authors justified

(b) Give a 95% confide authors to use the w