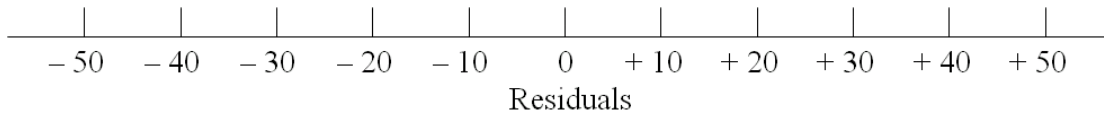


Text Exercise Set 12

NAME:

12-1 The Temperature and Bacteria Data, displayed as Table 12-4, is used in a study of the prediction of the number of thousands of bacteria that will develop in a particular environment from temperature.

- (a) Identify the response variable Y and the explanatory variable X .
- (b) Verify that the least squares line to predict number of bacteria (bct) from temperature (tmp) is $bct = -139.7321 + 4.8821(tmp)$.
(A programmable calculator or appropriate statistical software will be helpful.)
- (c) Find the five-number summary and interquartile range for the residuals.



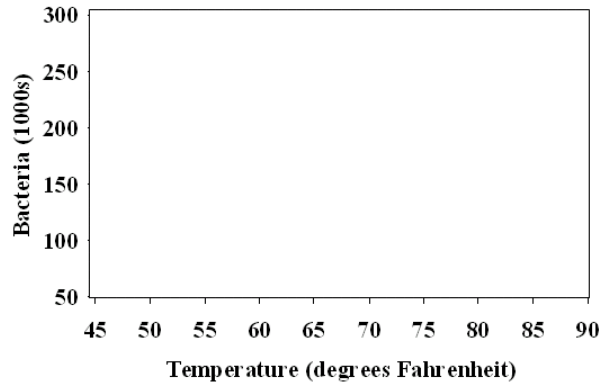
- (d) Decide whether there appear to be any candidates for outliers, and complete the construction of the modified box plot of the residuals.

- (e) Complete the construction of the scatter plot of the data, and graph the least squares line on the scatter plot.
- (f) From the scatter plot constructed in part (e), decide whether or not you think the residual plot will look random.

Table 12-4	
Temperature and Bacteria Data	
Similar colonies of bacteria are placed into environments which are identical except for temperature. The temperature and the number of new bacteria (thousands) at the end of a 24-hour period are recorded for each colony.	
Temperature(°F)	Bacteria(1000s)
50	71
50	93
55	127
55	149
60	180
60	196
65	197
65	175
70	176
70	199
75	208
75	182
80	243
80	259
85	298
85	284

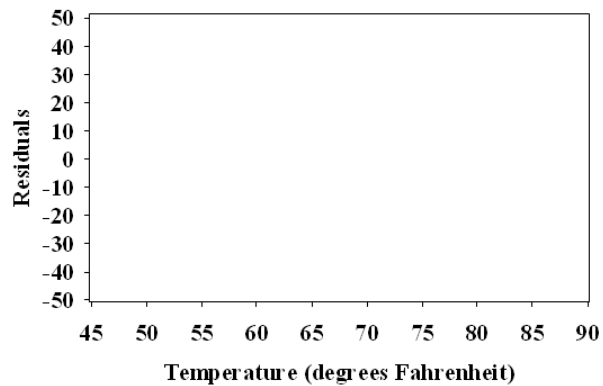
12-1 - continued

- (g) Complete the construction of the residual plot.
- (h) Decide whether or not the linearity assumption appears to be reasonable, and state why or why not.



12-2 The Study Time and Exam Score Data, displayed as Table 12-5, is used in a study of the prediction of exam score from study time.

- (a) Identify the response variable Y and the explanatory variable X .



- (b) Verify that the least squares line to predict exam score (exm) from study time (std) is

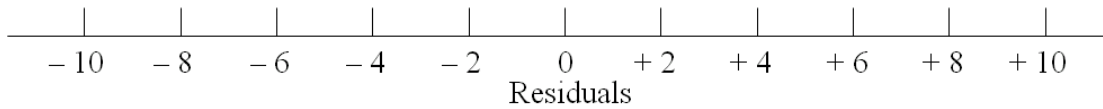
$$exm = 21.02273 + 2.4318(std)$$

(A programmable calculator or appropriate statistical software will be helpful.)

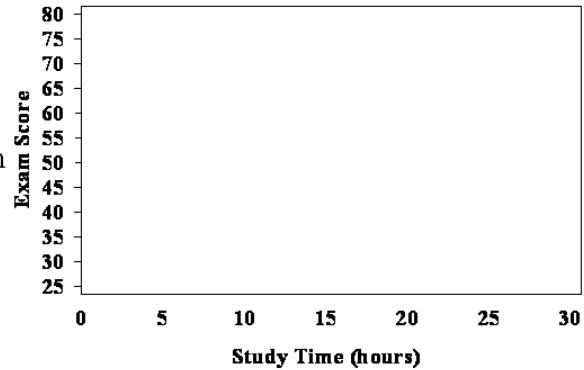
- (c) Find the five-number summary and interquartile range for the residuals.
- (d) Decide whether there appear to be any candidates for outliers, and complete the construction of the modified box plot of the residuals.

Study Time (hours)	Exam Score (%)
5	26.0
5	30.0
10	46.0
10	41.5
10	49.0
15	64.5
15	67.0
15	60.5
20	66.0
20	73.0
20	71.0
25	73.5
25	79.5

12-2 - continued



- (e) Complete the construction of the scatter plot of the data, and graph the least squares line on the scatter plot.
- (f) From the scatter plot constructed in part (e), decide whether or not you think the residual plot will look random.



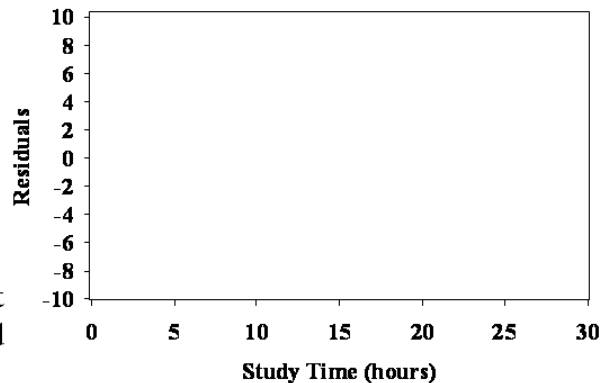
- (g) Complete the construction of the residual plot.
- (h) Decide whether or not the linearity assumption appears to be reasonable, and state why or why not.

12-3 In Exercises 10-5 and 11-1, the Biotin and Weight Data, displayed in Table 10-5, is used in a study of the relationship between the amount of biotin added per kilogram of feed and the weight in grams of newly hatched turkeys after three weeks, with interest in the prediction of weight from amount of biotin. The least squares line was found to be

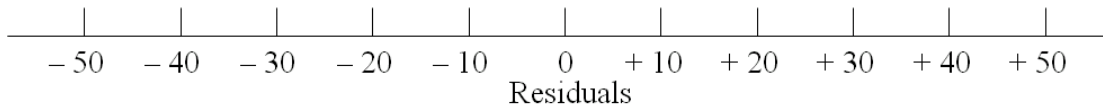
$$wt = 191 + 1.2(btn) ,$$

where wt represents weight in grams and btn represents amount of biotin per kilogram of feed

- (a) Find the residuals.
- (b) Find the five-number summary and interquartile range for the residuals.

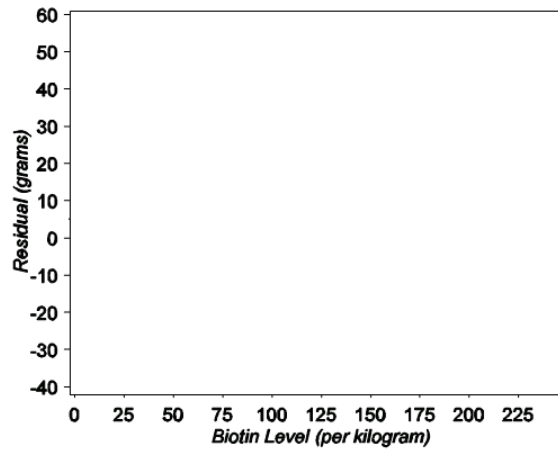


12-3 - continued



- (c) Decide whether there appear to be any candidates for outliers, and complete the construction of the modified box plot of the residuals.

- (d) From the scatter plot constructed in Exercise 11-1(h), decide whether or not you think the residual plot will look random; then, complete construction of the residual plot.



- (e) Decide whether or not the linearity assumption appears to be reasonable, and state why or why not.

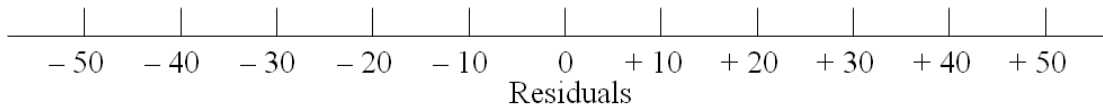
12-4 In Exercise 11-4, the Temperature and Strength Data, displayed in Table 11-2, is used to obtain the least squares line for the prediction of the breaking strength of an alloy from temperature. The least squares line was found to be

$$brk = 416.5 - 1.5(tmp) ,$$

where *brk* represents breaking strength in pounds of force and *tmp* represents temperature in degrees Fahrenheit.

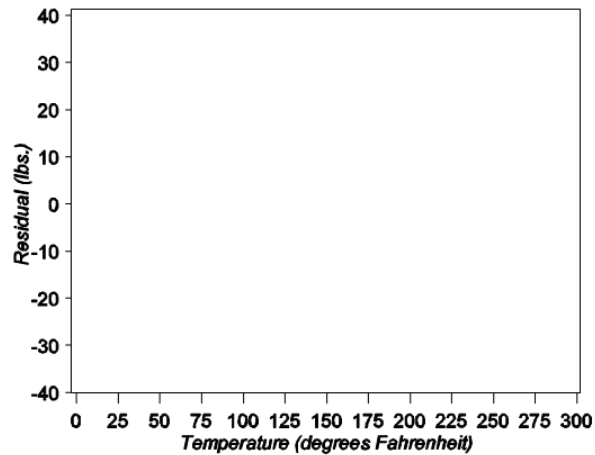
- (a) Find the residuals.
- (b) Find the five-number summary and interquartile range for the residuals.

12-4 - continued



- (c) Decide whether there appear to be any candidates for outliers, and complete the construction of the modified box plot of the residuals.

- (d) From the scatter plot constructed in Exercise 11-4(f), decide whether or not you think the residual plot will look random; then, complete construction of the residual plot.



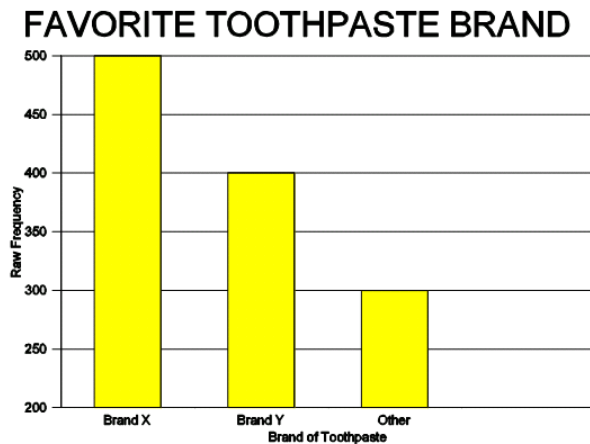
- (e) Decide whether or not the linearity assumption appears to be reasonable, and state why or why not.

12-5 In Exercise 11-3, the Task Training Data, displayed in Table 11-1, is used to obtain the least squares line for the prediction of the number of minutes necessary to perform a job from the number of hours of training to perform the job.

- (a) Why would it be difficult to make a decision about whether the linearity assumption is reasonable from a residual plot of this data?
- (b) The number of hours of training to perform the job in the data ranged from 15 to 45 hours. Do you think the linearity assumption would still be reasonable if the training hours in the data had ranged from 5 to 100 hours? Why or why not?

12-6 In Exercises 10-6 and 11-2, the Stride and Height Data, displayed in Table 10-6, is used in a study of the relationship between a person's height and a person's length of stride, with interest in the prediction of height from length of stride. Why would it be difficult to make a decision about whether the linearity assumption is reasonable from a residual plot of this data?

12-7 Each of 1200 people in a survey is asked to name which brand of toothpaste he or she uses most often. The results are displayed in the bar chart on the right. Explain why this bar chart is a misleading display of the results of the survey.

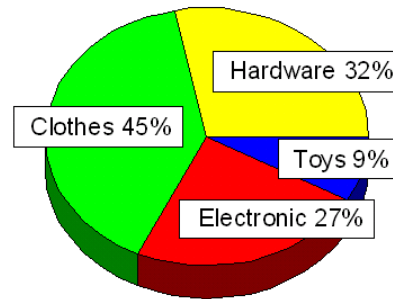


12-8 In Self-Test Problem 12-1, the Age and Grip Strength Data, displayed in Table 10-4, is used in a study concerning the prediction of grip strength from age. The residual plot displayed as Figure 12-6 does not appear to show any identifiable nonrandom pattern among the residuals, and consequently, we have no reason to doubt the linearity assumption. The ages in the data ranged from 11 to 25 years old. Do you think the linearity assumption would still be reasonable if the ages in the data had ranged from 5 to 55 years? Why or why not?

12-9 A line graph is constructed with five different companies represented on the horizontal axis and number of employees represented on the vertical axis. Explain why this graph is an inappropriate and misleading representation of the data.

12-10 The pie chart displayed as Figure 12-10 represents the distribution of sales at the JunkMart department store. What is wrong with the display of the data in this graph?

Figure 12-10
Sales Distribution at JunkMart



[This page is intentionally left blank]