4.2 Building Linear Models from Data

Learning Objectives

1. Draw and Interpret Scatter Diagrams
2. Distinguish between Linear and Nonlinear Relations
3. Use a Graphing Utility to Find the Line of Best Fit

Example 1 (1 of 3)

Drawing and Interpreting a Scatter Diagram

In baseball, the on-base percentage for a team represents the percentage of time that the players safely reach base. The data given in Table 6 on next slide, represent the number of runs scored y and the on-base percentage x for teams in the National League during the 2013 baseball season.

(a) Draw a scatter diagram of the data, treating on-base percentage as the independent variable.

(b) Use a graphing utility to draw a scatter diagram.

(c) Describe what happens to runs scored as the on-base percentage increases.

Example 1 (2 of 3)

Table 6

Team	On-Base Percentages, *x*	Runs Scored, y	(*x*, y)
Arizona	32.3	685	(32.3, 685)
Atlanta	32.1	688	(32.1, 688)
Chicago Cubs	30.0	602	(30.0, 602)
Cincinnati	32.7	698	(32.7, 698)
Colorado	32.3	706	(32.3, 706)
LA Dodgers	32.6	649	(32.6, 649)
Miami	29.3	513	(29.3, 513)

Example 1 (3 of 3)

Table 6 [continued]

Team	On-Base Percentages, *x*	Runs Scored, y	(*x*, y)
Milwaukee	31.1	640	(31.1, 640)
NY Mets	30.6	619	(30.6, 619)
Philadelphia	30.6	610	(30.6, 610)
Pittsburgh	31.3	634	(31.3, 634)
San Diego	30.8	618	(30.8, 618)
San Francisco	32.0	629	(32.0, 629)
St. Louis	33.2	783	(33.2, 783)
Washington	31.3	656	(31.3, 656)

Source: espn.go.com

Solution 1

data is approximately linear...
build the model manually using first and last points

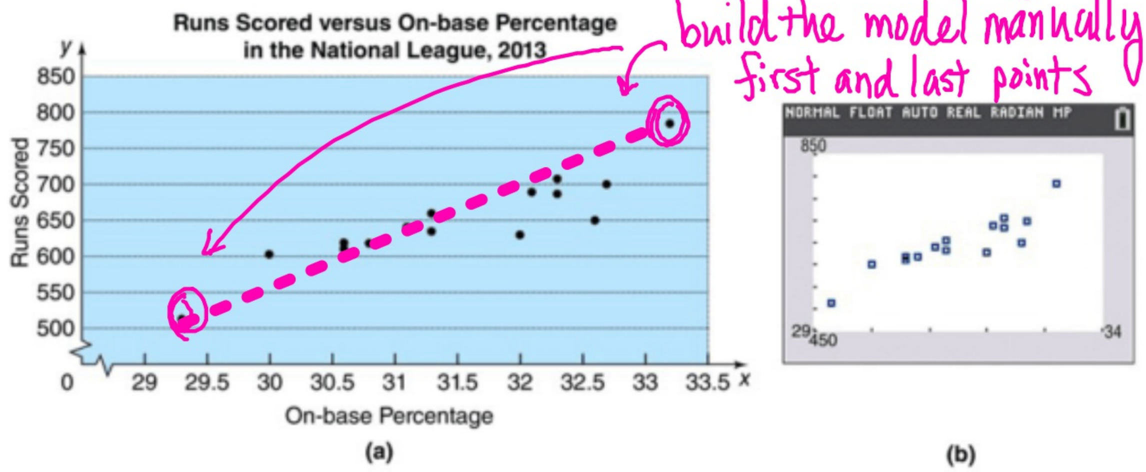


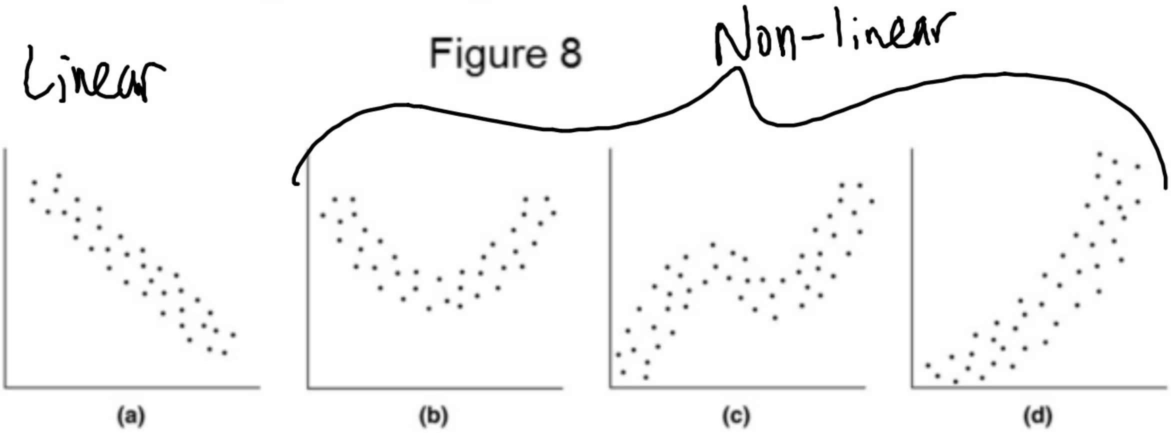
Figure 6

Example 2

Distinguishing between Linear and Nonlinear Relations

Determine whether the relation between the two variables in each scatter diagram in Figure 8 is linear or nonlinear.

Figure 8



Example 3

Finding a Model Manually for Linearly Related Data

Use the data in Table 6 from Example 1.

(a) Select two points and find an equation of the line containing the points.

(b) Graph the line on the scatter diagram obtained in Example 1(a).

Linear Model (Manually)

1. Select two points: $(29.3, 513)$ and $(33.2, 783)$

2. $$m = \frac{783 - 513}{33.2 - 29.3} = \frac{270}{3.9} \approx 69.23$$

3. $$y - y_1 = m(x - x_1)$$

$$y - 513 = 69.23(x - 29.3)$$

$$y - 513 = 69.23x - 2028.44$$

$$\boxed{y = 69.23x - 1515.44}$$

Example 4

Finding a Model using Technology for Linearly Related Data

Use the data in Table 6 (see slides: 6 and 7) from Example 1.

(a) Use a graphing utility to find the line of best fit that models the relation between on-base percentage and runs scored.

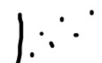
(b) Graph the line of best fit on the scatter diagram obtained in Example 1(b).

(c) Interpret the slope.

(d) Use the line of best fit to predict the number of runs a team will score if their on-base percentage is 33.1.

Linear Model (technology - TI-84/89)

1. Enter Data in calculator: (STAT), Edit..., enter X-values in L_1, y-values in L_2

2. Scatter Plot: (2nd) (y=) STAT PLOT, ON, 
   (change window settings to fit the data)

3. (STAT) (→) CALC option 4: LinReg(ax+b)

$$y = 49.4x - 906.3$$

model from calculator is "line of best fit"

Correlation Coefficient *r*

- The correlation coefficient *r* is used in statistics to determine the relative strength of the linear relationship in the data. Its value ranges from -1 to 1, with -1 being perfect negative correlation and 1 being perfect positive correlation. The closer the value of *r* gets to 0 the weaker the linear relationship is in the data.
- You can have your calculator determine the value of *r*. (turn "DiagnosticOn" before doing "LinReg")