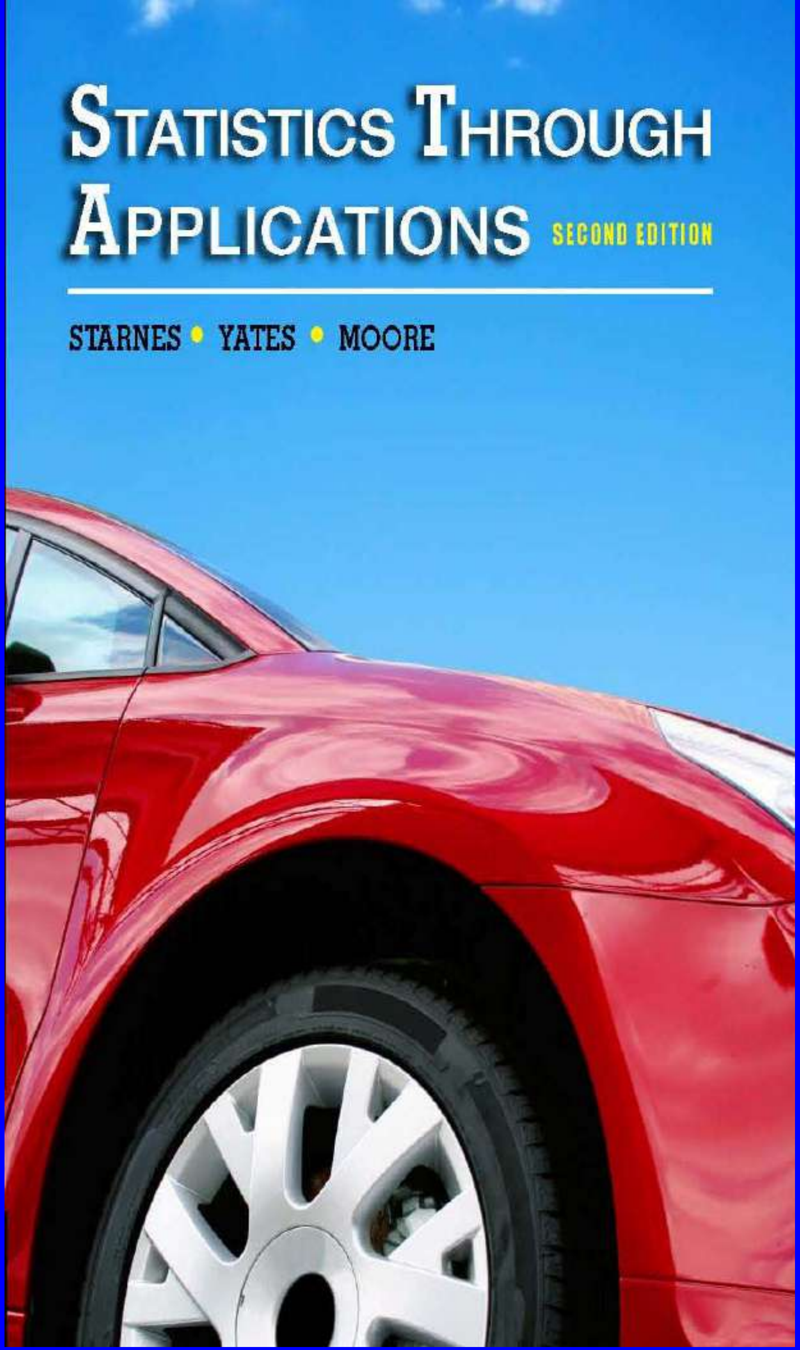



# Chapter 4

# Describing Relationships

# Section 4.2
# Regression and Prediction

## Regression line

a straight line that describes how a response variable y changes as an explanatory variable x changes.

We often use a regression line to predict the value of y for a given value of x.

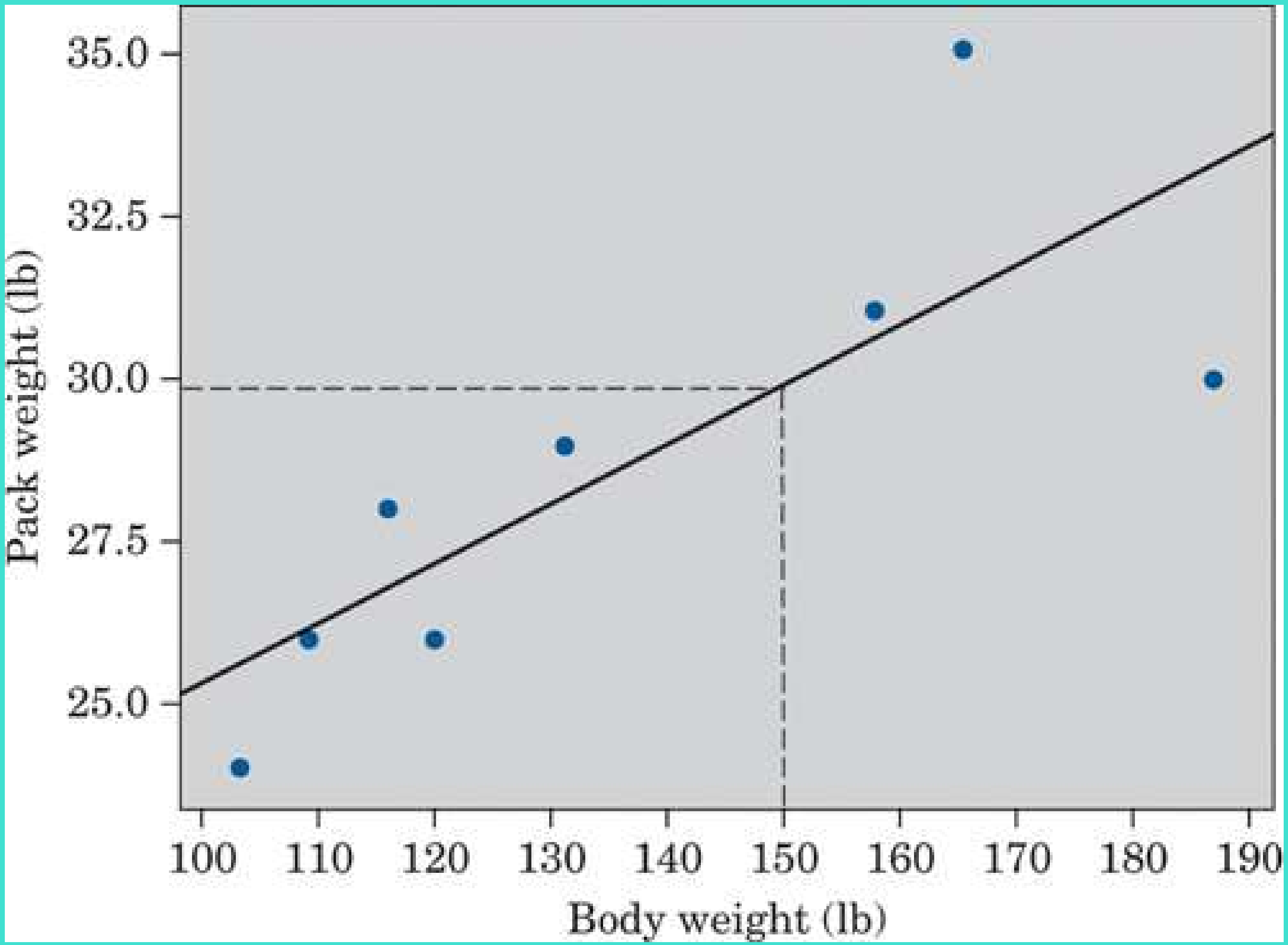



**Figure 4.17**
Using a straight-line pattern for prediction. The data are the body weights and backpack weights of 8 hikers.

At the last minute, another hiker is assigned to the group. This student weighs about 150 lbs. Can we predict this hiker's backpack weight?

# Regression Equations

Least-squares regression line

the line that makes the sum of the squares of the vertical distances of the data points from the line as small as possible.

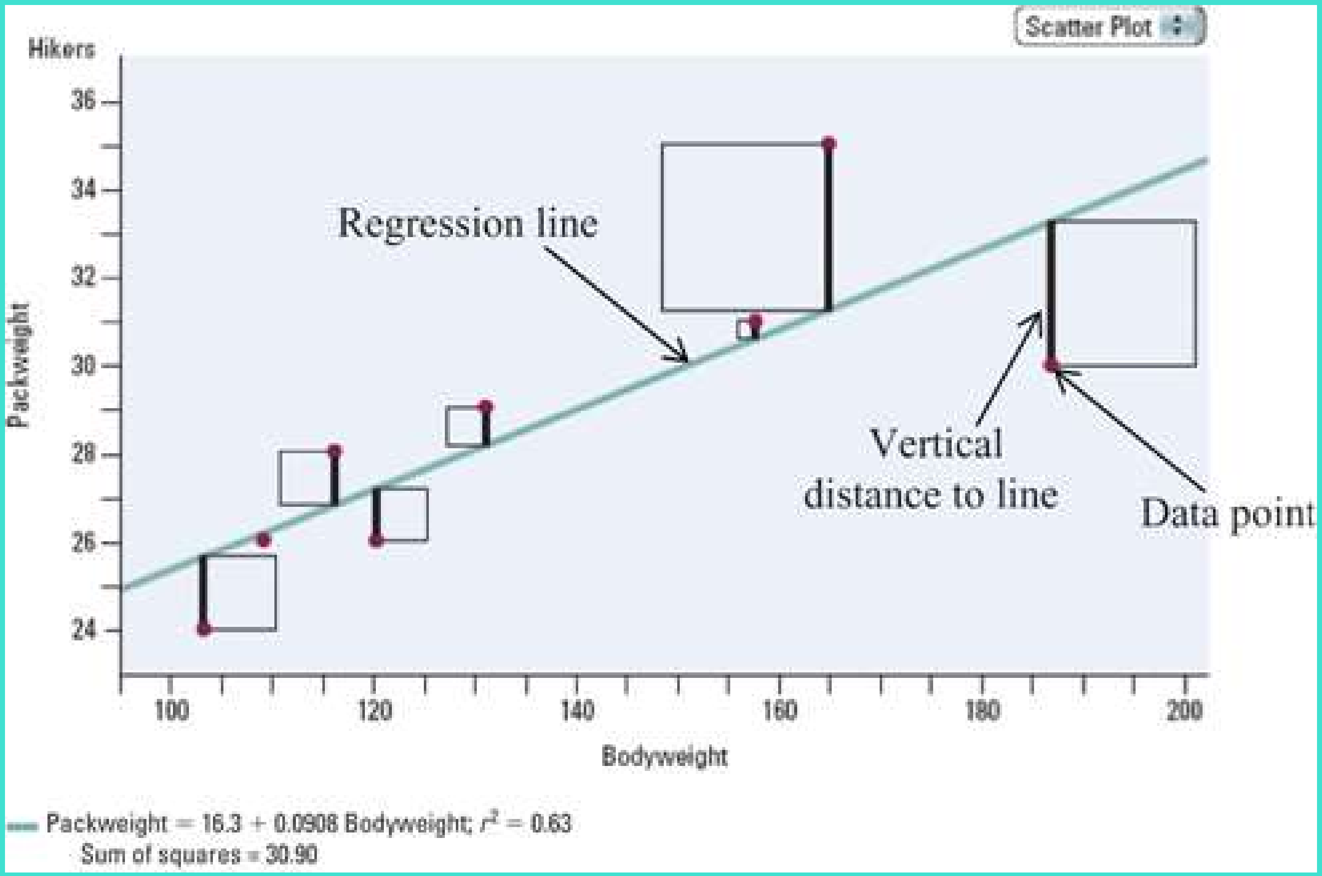


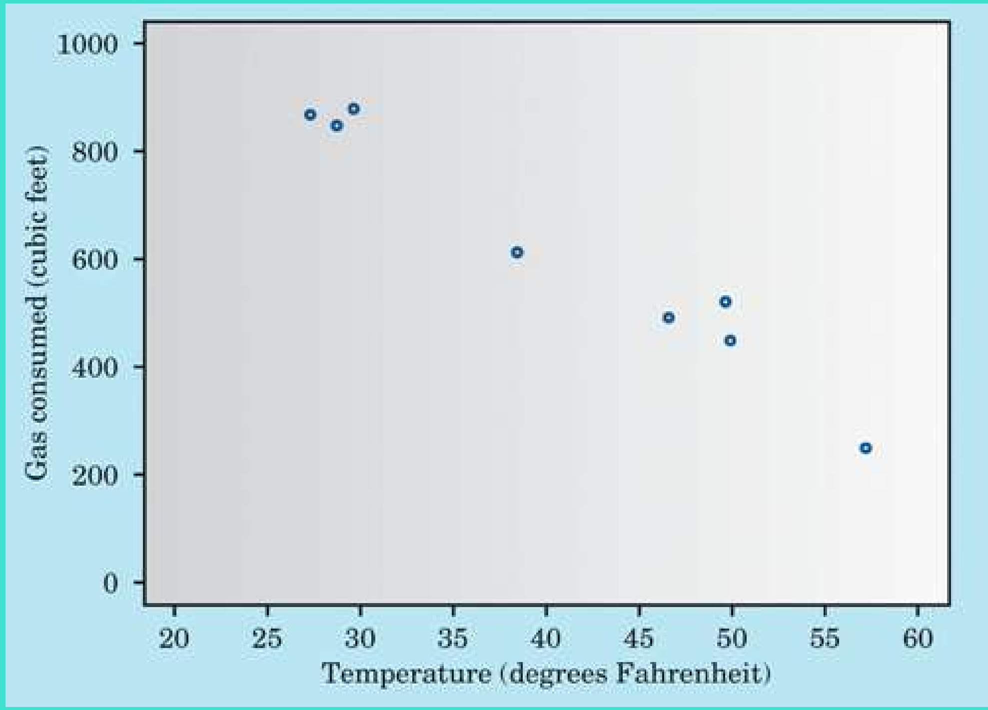

**Figure 4.19**
The least-squares regression line results in the smallest possible sum of squared distances of the points from the line.

# Using the regression equation

The equation of the least-squares line for the backpackers is

Pack weight = 16.3 + 0.09(body weight)

Slope = 0.09
Intercept = 16.3
For prediction, substitute the value of body weight and calculate pack weight



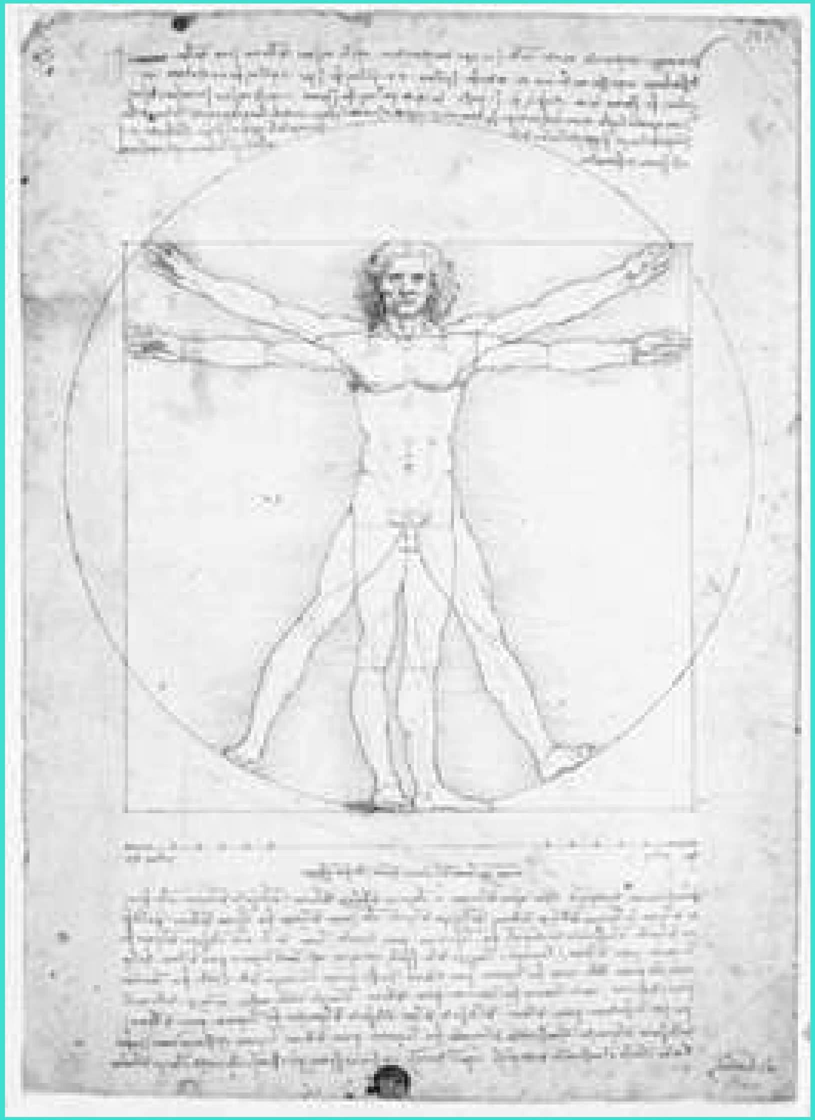

Enter the data from our previous example. We'll let the calculator find our regression line.

What is the correlation value?

| Month | Oct | Nov | Dec | Jan | Feb | Mar | Apr | May |
|---|---|---|---|---|---|---|---|---|
| Temp (x) | 49.4 | 38.2 | 27.2 | 28.6 | 29.5 | 46.4 | 49.7 | 57.1 |
| Gas consumed (y) | 520 | 610 | 870 | 850 | 880 | 490 | 450 | 250 |

# Correlation and Regression

- Describe only linear relationships
- Prediction outside the range of the available data is risky (interpolation vs. extrapolation)
- Both are strongly affected by outliers
- Regression line always passes through the point $(\bar{x},\bar{y})$
- Switching x and y has no effect on correlation, but it does affect regression



Vitruvius and the ideal man

According to the ancient architect Vitruvius, the measurements of various parts of the human body have been set by nature to follow certain ratios.  Leonardo da Vinci paid tribute to these claims in his painting entitled *Vitruvian Man*.

4-2 Activity  Vitrivius and the ideal man.docx

Activity (by row)

Measure the distance from the elbow to the tip of your middle finger.
Measure your height.
Record data.
Make a scatterplot of the measurements
Describe direction, form, and strength of the relationship

# Understanding prediction: residuals

A residual is the difference between an observed value of the response variable and the value predicted by the regression line.

residual = observed y – predicted y

# Understanding prediction: residuals

A residual plot should show no obvious pattern
The residuals should be relatively small in size

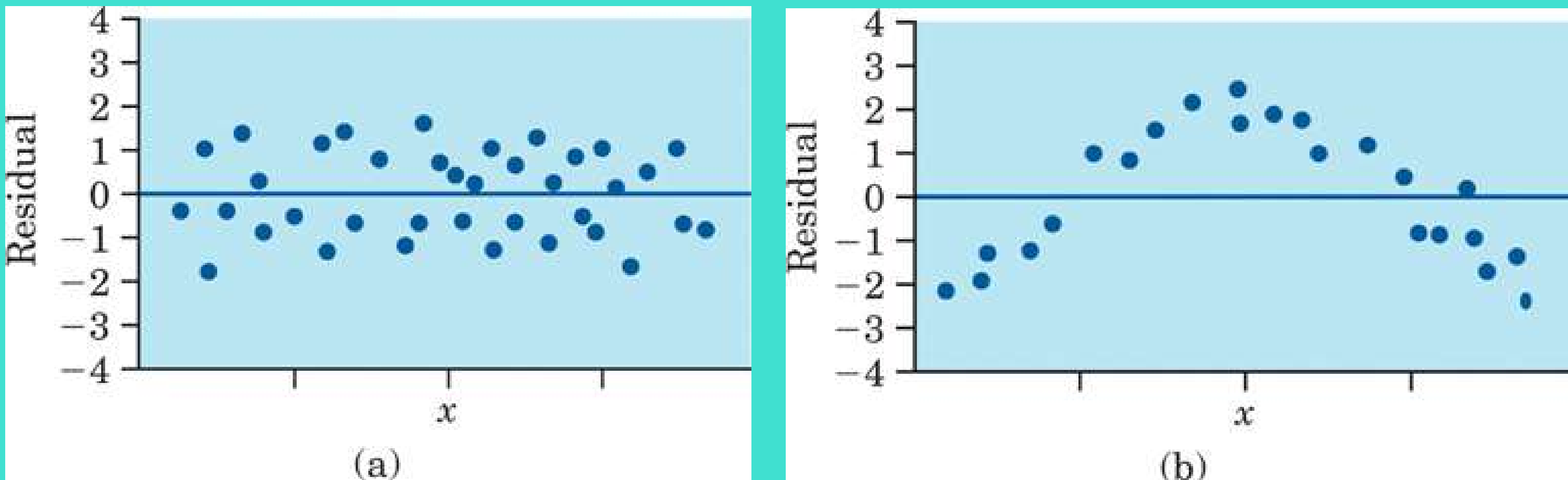



**Figure 4.23**
Two residual plots. In (a), the unstructured scatter of points indicates that the regression line fits the data well. In (b), the residuals have a curved pattern, so a straight line may not be the best model for the data.

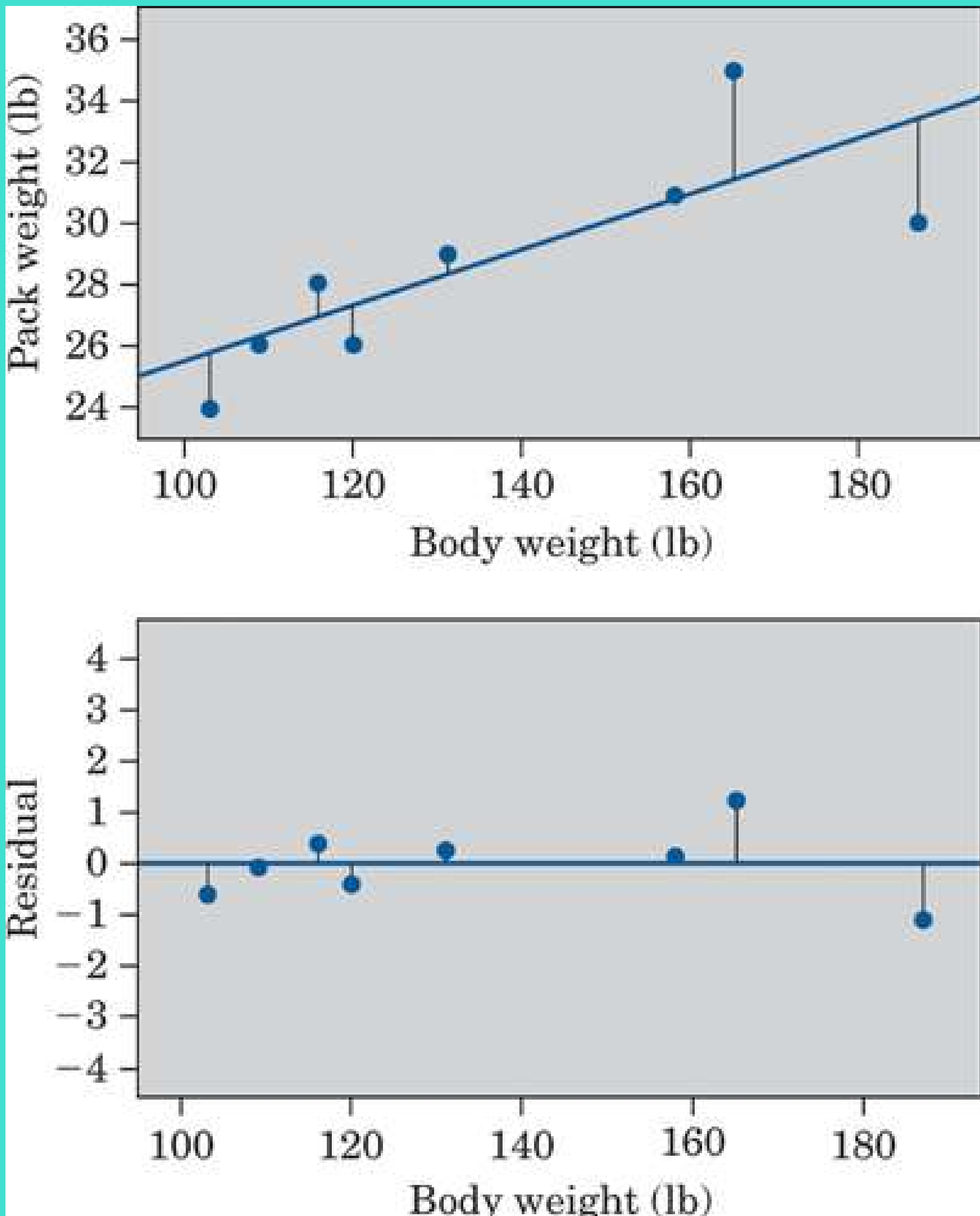



**Figure 4.22**
Scatterplot with least-squares line and a separate residual plot for the hiker data.

Weight (Grams) of a Bar of Soap Used to Shower

| Day | Weight | Day | Weight | Day | Weight |
|---|---|---|---|---|---|
| 1 | 124 | 8 | 84 | 16 | 27 |
| 2 | 121 | 9 | 78 | 18 | 16 |
| 5 | 103 | 10 | 71 | 19 | 12 |
| 6 | 96 | 12 | 58 | 20 | 8 |
| 7 | 90 | 13 | 50 | 21 | 6 |

*Source:* Rex Boggs.

Use the calculator to see our scatterplot with regression line and our residual plot.

To make a residual plot
2nd y= (STAT PLOT)
In plot 1, change Ylist to Resid
    2nd STAT (LIST)
    7. RESID (enter)
Zoom 9

# Understanding prediction: r2

The square of the correlation, r2, is the fraction of the variation in the values of y that is explained by the least-squares regression of y on x

A value of r2 = 0.632 means about 63% of the observed variation in pack weight is explained by the straight-line pattern.

This is a measure of how successful the regression was in explaining the response.

# The question of causation

There is a strong relationship between cigarette smoking and death rate from lung cancer.
Does smoking cigarettes cause lung cancer?

There is a strong association between the availability of handguns in a nation and that nation's homicide rate from guns.
Does easy access to handguns cause more murders?

In the 1990's, researchers measured the number of television sets per person x and the life expectancy y for the world's nations.  There was a high positive correlation: nations with many TV sets had higher life expectancies.

Does television extend life?

"In a new attack on third-world poverty, aid organizations today began delivery of 100,000 television sets."

# Lurking variable

A variable that has an important effect on the relationship among the variables in a study but is not one of the explanatory variables studied.

In the television example, national wealth is a lurking variable.
This is a common response: both the explanatory variable (# of television sets) and the response variable (life expectancy) are responding to the lurking variable (national wealth).

Even when direct causation is present, it is rarely a complete explanation of an association between two variables.

## Confounding

Two variables are confounded when their effects on a response variable cannot be distinguished from each other. The confounded variables may be either explanatory variables or lurking variables

# Direct cause-and-effect

When there is a real link: x causes y

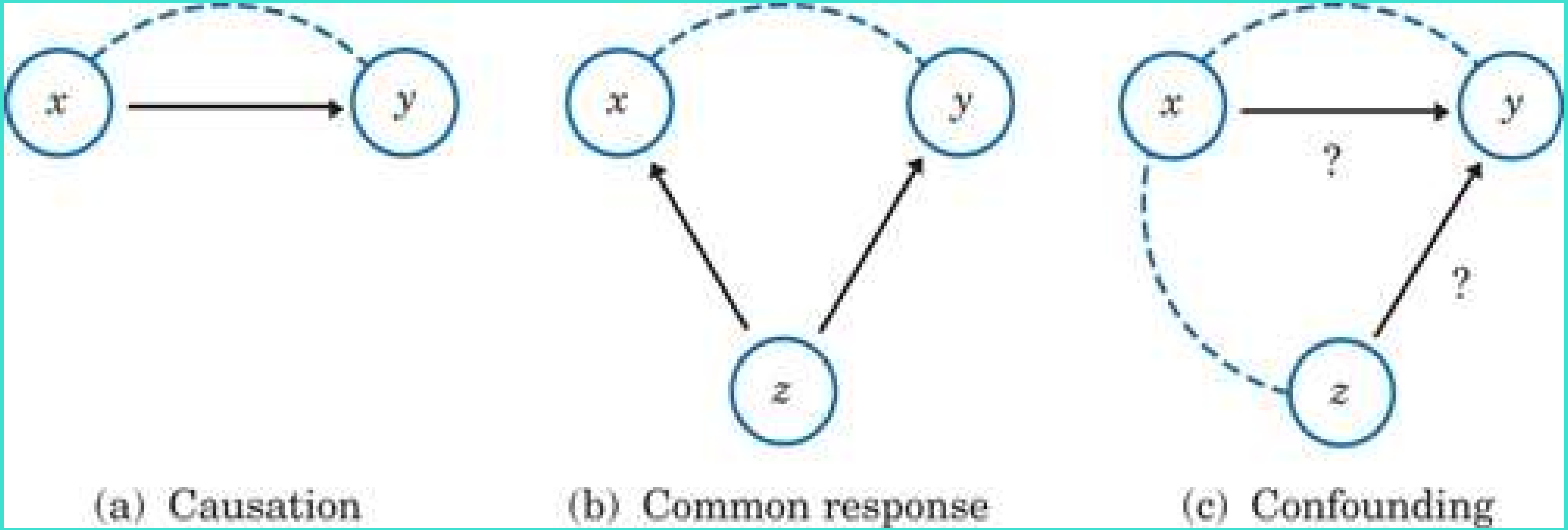



**Figure 4.26**
Some explanations for an observed association. A dashed line shows an association. An arrow shows a cause-and-effect link. Variable $x$ is explanatory, $y$ is a response variable, and $z$ is a lurking variable.

# Evidence for causation

*The best evidence for causation comes from well-designed experiments.*

Criteria for establishing causation:
The association is strong
The association is consistent
Higher doses are associated with stronger responses
The alleged cause precedes the effect in time
The alleged cause is plausible

Attachments

4-2 Activity  Vitrivius and the ideal man.docx