2.1 a) It is natural to think that the amount of time a student spends studying for a statistics exam affects the grade on the exam. Thus, we should probably view the amount of time spent studying as the explanatory variable and grade on the exam as the response variable.

b) We would expect weight and height to be associated but it is not obvious that one "causes" the other. Thus, it is reasonable to simply explore the relationship between the two variables.

c) One generally assumes that the amount of yearly rainfall affects the yield of a crop. Thus, we should probably view the amount of yearly rainfall as the explanatory variable and the yield of a crop as the response variable.

d) It is possible that an employee who takes many sick days is an irresponsible employee and that this will result in a lower salary. However, many factors affect salary and many factors affect the number of sick days one takes. Thus, it may be most reasonable to simply explore the relationship between these two variables.

e) It is not unreasonable to believe that the economic class of a father will have some effect on the economic class of the son. Thus, one should probably take the economic class of the father as the explanatory variable and the economic class of the son as the response variable.

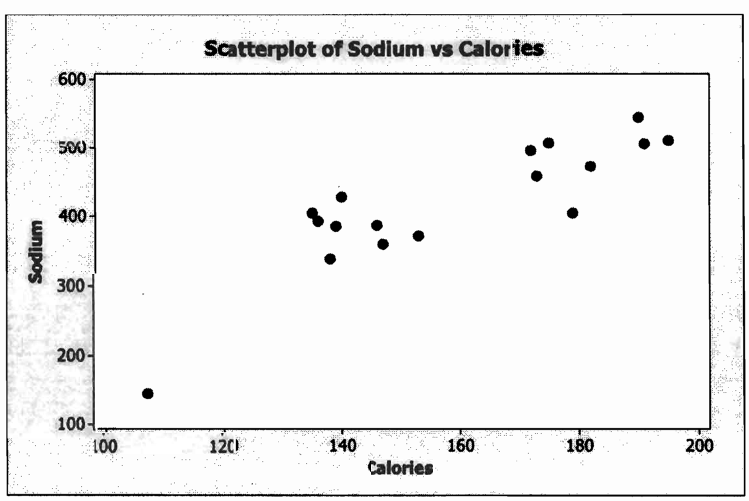
2.2 *Price at beginning of year* would be the explanatory variable, and *price at end of year* would be the response. These are quantitative variables.

2.3 The explanatory variable would be *type of hand wipe*. The response variable would be *level of skin irritation*. These are categorical variables.

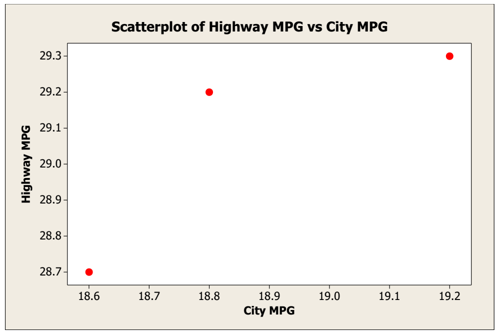
- Exercise 2.19

  (a) The overall pattern is linear, the association is positive, and the strength is moderately strong. Yes, hot dogs high in calories tend to be high in salt.

  (b) Since this brand markets itself as a diet brand, it's likely to be low in calories. It appears to be Brand 13, with 107 calories and 144 mg of sodium.



- Exercise 2.25

(a) 

(b)

$\bar{x} = 18.87$
$s_x = 0.3055$
$\bar{y} = 29.07$
$s_y = 0.3215$

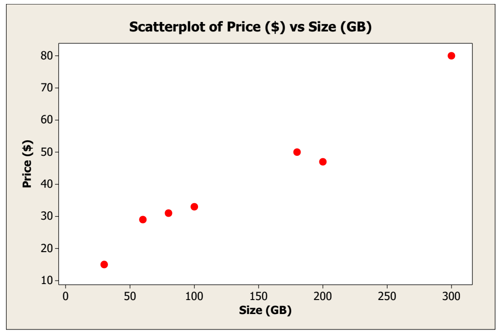
City MPG (x)	Highway MPG (y)	$\frac{x_i - \bar{x}}{s_x}$	$\frac{y_i - \bar{y}}{s_y}$	$\left(\frac{x_i - \bar{x}}{s_x}\right) \cdot \left(\frac{y_i - \bar{y}}{s_y}\right)$
18.6	28.7	-0.8838	-1.1509	1.0172
19.2	29.3	1.0802	0.7154	0.7728
18.8	29.2	-0.2291	0.4044	-0.0926
				1.6974

$$r = \frac{1.6974}{2} = \boxed{0.8487}$$

(c) The association is positive and fairly strong.

- Exercise 2.26

  (a) There is a strong positive linear relationship.



  (b)

  $\bar{x} = 135.7$
  $s_x = 95.2$
  $\bar{y} = 40.71$
  $s_y = 20.90$

Size in GB (x)	Price in $ (y)	$\frac{x_i - \bar{x}}{s_x}$	$\frac{y_i - \bar{y}}{s_y}$	$\left(\frac{x_i - \bar{x}}{s_x}\right) \cdot \left(\frac{y_i - \bar{y}}{s_y}\right)$
80	31	-0.5851	-0.4646	0.2718
60	29	-0.7952	-0.5603	0.4456
180	50	0.4653	0.4445	0.2068
300	80	1.7258	1.8799	3.2443
200	47	0.6754	0.3010	0.2033
100	33	-0.3750	-0.3689	0.1383
30	15	-1.1103	-1.2301	1.3658
				5.8759

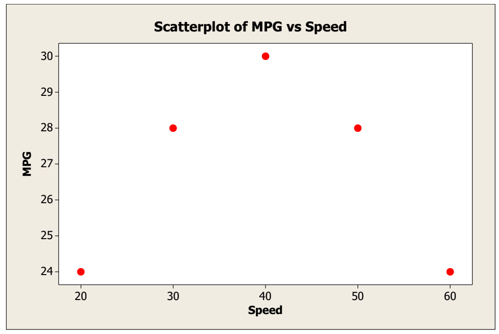
$$r = \frac{5.8759}{6} = \boxed{0.9793}$$

  (c) Entering the data into a calculator and using the calculator's *correlation* function provides the answer $r = \boxed{0.97929}$ (when rounded to 5 decimal places.)

  (d) The MINITAB answer is $r = \boxed{0.979}$

- Exercise 2.29

(a)



(b) $\bar{x} = 40$
$s_x = 15.81$
$\bar{y} = 26.80$
$s_y = 2.68$

Speed (x)	MPG (y)	$\frac{x_i - \bar{x}}{s_x}$	$\frac{y_i - \bar{y}}{s_y}$	$\left(\frac{x_i - \bar{x}}{s_x}\right) \cdot \left(\frac{y_i - \bar{y}}{s_y}\right)$
20	24	-1.265	-1.045	1.322
30	28	-0.633	0.448	-0.284
40	30	0.000	1.194	0.000
50	28	0.633	0.448	0.284
60	24	1.265	-1.045	-1.322
				0

$$r = \frac{0}{0} = \boxed{0}$$

(c) Speeds 20 and 50

(d) Speeds 30 and 60

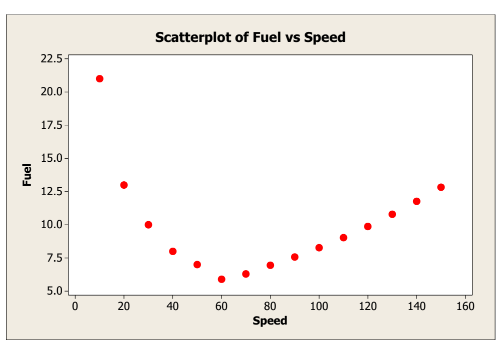
(e) Speed 40

(f) $r = 0$

(g) $r = 0$

- Exercise 2.7

(a) Speed is the predictor (explanatory) variable.



(b) The plot shows a curved relationship. Fuel use declines from very low speeds to about 60 km/hour, then begins to increase again for higher speeds. (At very low speeds and very high speeds, engines are inefficient and use more fuel. They are more efficient and use less fuel at moderate speeds.)

(c) The data have the shape of a parabola (curve), not a line.

(d) The relationship looks reasonably strong.

- Exercise 2.44

From MINITAB, the correlation is $r = \boxed{-0.172}$ The correlation is close to 0 since the relationship is a curve, not linear.

- Exercise 2.71

If the correlation has increased to 0.8, then there is a stronger relationship between American and European stocks. In particular, when American stocks have gone down, European stocks have tended to go down also. Thus, European stocks have not provided much protection against losses in American stocks since Europeans stocks tend to have losses at the same time.

- Exercise 2.72

No, this is not true. R^2 is the measure that the reporter needed. Price changes on Wall Street can explain only 64% of price changes in Europe.

- Exercise 2.15

(a) A positive association between decline and duration would mean that the longer the duration of a bear market, the greater the decline in stock prices.
Yes, the plot shows a positive association.

(b) The relationship is roughly linear but does not appear to be very strong.

(c) The bear market with the greatest decline is the point which is highest in the vertical direction on the graph. That bear market declined about 48% and had a duration of about 21 months.

- Exercise 2.73

(a)

$$b_1 = r \cdot \frac{s_y}{s_x} = (0.6285)\frac{11.20}{8.20} = \boxed{0.8584}$$

$$b_0 = \bar{y} - b_1\bar{x} = 24.67 - (0.8584)(10.73) = \boxed{15.46}$$

Therefore,

$$\widehat{y} = 15.46 + 0.8584\,x$$

OR

$$\text{Estimated Decline } = 15.46 + (0.8584)(\text{Duration})$$

(b)

$$R^2 = (0.6285)^2 = \boxed{39.50\%}$$

(c)

$$\widehat{y} = 15.46 + 0.8584\,x = 15.46 + (0.8584)(15) = \boxed{28.34\% \text{ decline}}$$

$$\text{residual} = \text{prediction error } = y - \widehat{y} = 14\% - 28.34\% = \boxed{-14.34\%}$$

- Exercise 2.49

  (a) 35.52 %

  (b)

$$\widehat{y} = 6.08\% + 1.707\,x$$

  (c) The point $(\bar{x}, \bar{y})$ is always on the regression line, as described in the Topic 8 notes. So the predicted change for the full year is

$$\widehat{y} = \bar{y} = 9.07\%$$

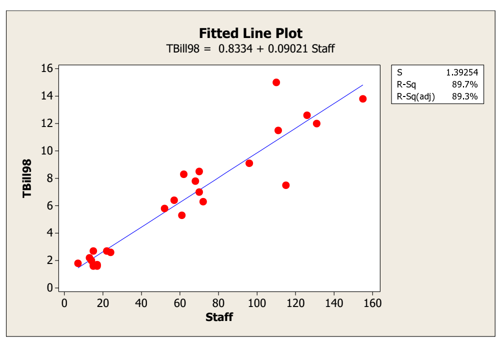
  if the January change is

$$x = \bar{x} = 1.75\%$$

  No calculation is necessary, although a calculation for $\widehat{y}$ with $x = 1.75$ will confirm the answer.

- Exercise 2.48

  (a) Note: The textbook's regression equation listed on page 123 appears to be incorrect.



$$\begin{aligned}\widehat{y} &= 0.8334 + 0.09021x \\ &= 0.8334 + (.09021)(111) \\ &= \boxed{10.8467}\end{aligned}$$

  Thus we predict \$10.8467 million in total billings for this firm.

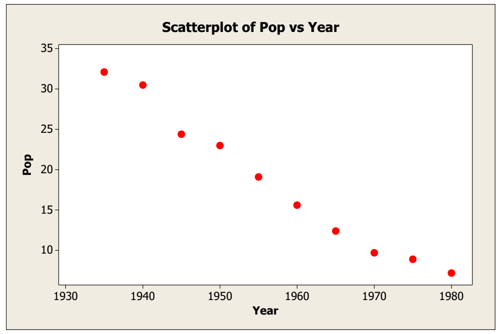
  (b) Schmidt Associates actually has 111 employees, with total billing $y = 11.5$. The prediction error is

$$\text{error } = y - \widehat{y} = 11.5 - 10.8467 = 0.6533 \quad \text{or} \quad \boxed{\$653{,}300}$$

- Exercise 2.88

  (a) From a calculator, $r = -0.9884$ and the regression equation is

  $$\widehat{y} = 1166.93 - 0.5868\,x$$

  (b) According to the regression, farm population declined by about 586,800 people each year, on average. 97.7% of the observed variation in population can be explained by time.

  (c) For $x = 1990$,

  $$\widehat{y} = 1166.93 - 0.5868(1990) = -0.802$$

  This would imply that the farm population in 1990 is $-802,000$, which is impossible. The danger comes from extrapolating beyond the range of the data for the predictor variable (time.)

- Exercise 2.91

  That reasoning assumes that the correlation between number of firefighters at a fire and the amount of damage done is due to causation, and that a greater number of firefighters causes more damage. Correlation does not imply causation.

  It is more likely that a lurking variable, namely the size of the fire, is behind the correlation. Larger fires cause more damage and require more firefighters to combat them.

- Exercise 2.93

  No. Correlation does not imply causation. I.e., the existence of a correlation between between the size of a hospital and the median number of days that patients remain in the hospital does not imply that the size of hospital causes the length of stay.

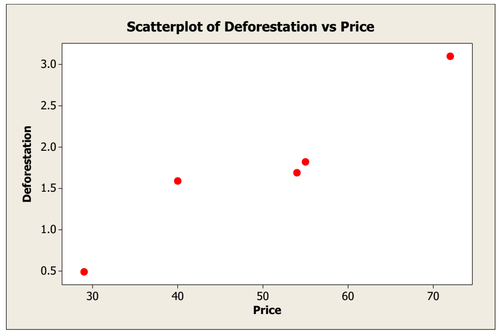
  It is possible that a lurking variable may cause the correlation. For instance, larger hospitals often have better facilities and equipment which enables them to care for patients with serious illness or injuries. The most serious patients may therefore be sent to larger hospitals rather than smaller hospitals, and the seriousness of the injury or illness may also require longer stays.

- Exercise 2.100

  A more plausible explanation is that heavier people tend to be on diets and try to reduce calories from sugar. Therefore they use artificial sweeteners.

- Exercise 2.31

  (a) The predictor (explanatory) variable is Price. The plot shows a strong positive linear pattern.



  (b)

  $\bar{x} = 50.0$
  $s_x = 16.32$
  $\bar{y} = 1.738$
  $s_y = 0.928$

Price (x)	Deforestation (y)	$\frac{x_i - \bar{x}}{s_x}$	$\frac{y_i - \bar{y}}{s_y}$	$\left(\frac{x_i - \bar{x}}{s_x}\right) \cdot \left(\frac{y_i - \bar{y}}{s_y}\right)$
29	0.49	−1.287	−1.345	1.731
40	1.59	−0.613	−0.159	0.097
54	1.69	0.245	−0.052	−0.013
55	1.82	0.306	0.088	0.027
72	3.10	1.348	1.468	1.979
				3.821

$$r = \frac{3.821}{4} = \boxed{0.95525}$$

  (c) Entering the data into a calculator and using the calculator's *correlation* function provides the answer $r = \boxed{0.95516}$ (when rounded to 5 decimal places.)

- Exercise 2.32

  No, measurement changes do not affect correlation.

  (b)

$$b_1 = r \cdot \frac{s_y}{s_x} = (0.9553)\frac{0.928}{16.32} = \boxed{0.0543}$$

$$b_0 = \bar{y} - b_1\bar{x} = 1.738 - (0.0543)(50) = \boxed{-0.977}$$

  Therefore,

$$\widehat{y} = -0.977 + 0.0543\,x$$

OR

Estimated Deforestation $= -0.977 + (0.0543)(\text{Price})$

(c) Interpret slope:

Deforestation increases by 0.0543% per year, on average, for every one-cent per pound increase in the price of coffee.

(d) Interpret intercept:

If the market for coffee collapses (i.e., the price drops to 0), then the forest will grow back at a rate of 0.977% per year.