HW9 - Solutions

1) There are many reasons why this may not be a causal relationship:
- the relationship between the superbowl winner and the S&P may not be consistent
- there is no clear mechanism linking the two
- there may be a common cause for both (e.g., maybe manufacturing jobs are located near NFC teams, so when the manufacturing industry does better, the NFC teams get more money (more fans) and the S&P goes up).

2) (a) Cov(X,Y) = sqrt(25) * sqrt(16) * (-.5) = -10
(b) Cov(2X, 3Y) = 2 * 3 * Cov(X,Y) = -60
(c) Var(X + Y) = Var(X) + 2 * Cov(X,Y) + Var(Y) = 25 - 20 + 16 = 21
(d) Var(2X + 3Y) = 4 Var(X) + 12 Cov(X,Y) + 9 Var(Y) = 124

3) In all three cases, the mean is 20 * .04 + 30 * .10 = 3.8

The standard deviation is
sqrt(20^2 * (.02)^2 + 2 * 20 * 30 * (.02) * (.20) * rho + (30)^2 * (.20)^2)
= sqrt(36.16 + 4.8 * rho)

(a) rho = 0; sd = sqrt(36.16) = 6.01
(b) rho = .7; sd = sqrt(39.52) = 6.29
(c) rho = -.7; sd = sqrt(32.8) = 5.73

4)
10.7 A deterministic model does not allow for random error or variation, whereas a probabilistic model does. An example where a deterministic model would be appropriate is:

Let y = cost of a 2×4 piece of lumber and
x = length (in feet)

The model would be $y = \beta_1 x$. There should be no variation in price for the same length of wood.

An example where a probabilistic model would be appropriate is:

Let y = sales per month of a commodity and
x = amount of money spent advertising

The model would be $y = \beta_0 + \beta_1 x + \varepsilon$. The sales per month will probably vary even if the amount of money spent on advertising remains the same.

5)
10.9 No. The random error component, ε, allows the values of the variable to fall above or below the line.

6)

10.10 a.

x_i	y_i	x_i^2	x_iy_i
7	2	$7^2 = 49$	$7(2) = 14$
4	4	$4^2 = 16$	$4(4) = 16$
6	2	$6^2 = 36$	$6(2) = 12$
2	5	$2^2 = 4$	$2(5) = 10$
1	7	$1^2 = 1$	$1(7) = 7$
1	6	$1^2 = 1$	$1(6) = 6$
3	5	$3^2 = 9$	$3(5) = 15$

Totals:
$$\sum x_i = 7+4+6+2+1+1+3 = 24$$
$$\sum y_i = 2+4+2+5+7+6+5 = 31$$
$$\sum x_i^2 = 49+16+36+4+1+1+9 = 116$$
$$\sum x_i y_i = 14+16+12+10+7+6+15 = 80$$

b. $SS_{xy} = \sum x_i y_i - \dfrac{\left(\sum x_i\right)\left(\sum y_i\right)}{n} = 80 - \dfrac{(24)(31)}{7} = 80 - 106.2857143 = -26.2857143$

c. $SS_{xx} = \sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{7} = 116 - \dfrac{(24)^2}{7} = 116 - 82.28571429 = 33.71428571$

d. $\hat{\beta}_1 = \dfrac{SS_{xy}}{SS_{xx}} = \dfrac{-26.2857143}{33.71428571} = -.779661017 \approx -.7797$

e. $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{24}{7} = 3.428571429 \quad \bar{y} = \dfrac{\sum x_i}{n} = \dfrac{31}{7} = 4.428571429$

f. $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 4.428571429 - (-.779661017)(3.428571429)$
$= 4.428571429 - (-2.673123487) = 7.101694916 \approx 7.102$

g. The least squares line is $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 7.102 - .7797x$.

7)

2nd edition:

10.15 a. The straight-line model would be: $y = \beta_o + \beta_1 x + \varepsilon$

b. From the printout, the least squares line is: $\hat{y} = -146.856 + 1.144x$.

c. Since range of observed values for the 2000 SAT scores (x) does not include 0, the y-intercept has no meaning.

d. The slope of the line is β_1. In terms of this problem, β_1 is the change in the mean 2007 SAT score for each additional point increase in the 2000 SAT score. This interpretation is meaningful for values of x within the observed range. The observed range of x is 966 to 1,197.

3rd edition:

b. yhat = -97.414 + 1.188 x

8)

10.20 a. Some preliminary calculations are:

$$\sum x = 6167 \qquad \sum y = 135.8 \qquad n = 24$$
$$\sum x^2 = 1{,}641{,}115 \qquad \sum xy = 34{,}764.5$$

$$SS_{xy} = \sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n}$$
$$= 34{,}764.5 - \frac{(6167)(135.8)}{24} = -130.44167$$

$$SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n}$$
$$= 1{,}641{,}115 - \frac{(6167)^2}{24} = 56{,}452.95833$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{-130.44167}{56452.958} = -.002310625 \approx -.0023$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{135.8}{24} - (-.002310625)\left(\frac{6167}{24}\right) = 6.252067683 \approx 6.25$$

The least squares line is $\hat{y} = 6.25 - .0023x$

b. $\hat{\beta}_0 = 6.25$. Since $x = 0$ is not in the observed range, $\hat{\beta}_0$ has no interpretation other than being the y-intercept.

$\hat{\beta}_1 = -.0023$. For each additional increase of 1 part per million of pectin, the mean sweetness index is estimated to decrease by .0023.

c. $\hat{y} = 6.25 - .0023(300) = 5.56$