A. Using Population Regression Lines

- Exercise 10.3

  (a) $\beta_0 = 4.6$

  The mean overseas market return is 4.6% when the U.S. market is flat.

  (b) $\beta_1 = 0.67$

  For each 1% increase in return in the U.S. market, the mean overseas market increases by 0.67%.

  (c) There are two different (but equivalent) ways to write the population regression line, depending on whether you wish to emphasize mean response μ_y or a particular response y:

$$\mu_y = \beta_0 + \beta_1 x \tag{1}$$
$$y = \beta_0 + \beta_1 x + \epsilon \tag{2}$$

  Putting these two equations together shows that

$$y = \mu_y + \epsilon$$
$$\text{so the “error” } \epsilon = y - \mu_y$$

  represents the difference between a particular response y and the mean response μ_y.

  In this exercise,

$$\text{Mean Overseas Return } = 4.6 + 0.67 \times (\text{U.S. Return}) \tag{3}$$
$$\text{Overseas Return } = 4.6 + 0.67 \times (\text{U.S. Return}) + \epsilon \tag{4}$$

  so

$$\epsilon = \text{ Overseas Return} - \text{Mean Overseas Return}$$

  allows overseas returns to vary in different years, even if U.S. returns remain the same.

- Exercise 10.4

  (a) β_0

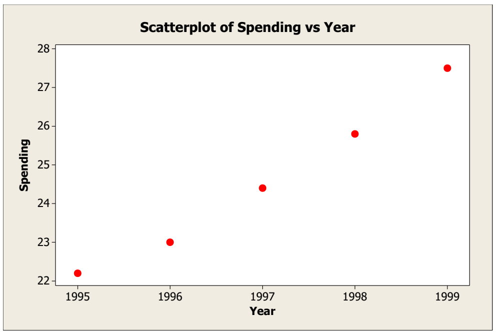
  (b) β_1. We expect $\beta_1 > 0$ since cost of production increases as more items are produced.

  (c) ϵ

B. Research and Development Spending

- Exercise 10.5

(a) There is a strong positive linear relationship between Year and Spending.



(b) Entering data into a calculator and using the calculator's *regression* function:

$$\widehat{\text{Spending}} = -2651.40 + 1.34 \text{ Year}$$

(c)

x = Year	y = Spending	$\widehat{y}$ = Predicted Spending	Prediction Error (Residual) $e = y - \widehat{y}$	$(y - \widehat{y})^2$
1995	22.2	21.9	0.30	.0900
1996	23.0	23.24	−0.24	.0576
1997	24.4	24.58	−0.18	.0324
1998	25.8	25.92	−0.12	.0144
1999	27.5	27.26	0.24	.0576
			0	0.252

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{n-2} = \frac{0.252}{5-2} = \frac{0.252}{3} = 0.084$$

$$s = \sqrt{s^2} = \sqrt{0.084} = \boxed{0.2898}$$

(d) Regression model:

$$\text{Spending} = \beta_0 + \beta_1 \times (\text{Year}) + \epsilon$$

$\widehat{\beta}_0 = b_0 = \boxed{-2651.40}$

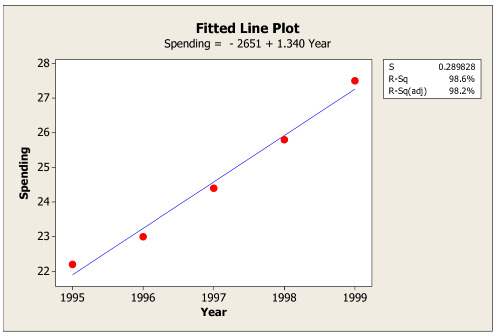
$\widehat{\beta}_1 = b_1 = \boxed{1.34}$

(e) $\widehat{y} = 29.94$

$$\text{Prediction error} = y - \widehat{y} = 32.7 - 29.94 = \$2.76 \text{ billion}$$

This prediction is dangerous because it represents an extrapolation from the original data for Years into the future. The trend might not stay the same after 1999.

(f) s = $\boxed{0.289828}$



(g) $s_y = \boxed{2.134}$ Yes, the reduction from 2.134 to 0.2898 is substantial!

(h) $R^2 = \boxed{98.6\%}$

C. Earnings for Female Bank Employees

- Exercise 10.1
  $x = 94$ months, $y = \$389$/week

$$\begin{aligned}\widehat{y} &= 349.4 + 0.5905x\\ \widehat{y} &= 349.40 + 0.5905(94)\\ \widehat{y} &= 349.40 + 55.51 = \boxed{\$404.91}\end{aligned}$$

  Prediction error = residual = $y - \widehat{y} = \$389 - \$404.91 = \boxed{-\$15.91}$

- Exercise 10.2
  $x = 1 \Longrightarrow \$349.99$
  $x = 100 \Longrightarrow \$408.45$
  $x = 200 \Longrightarrow \$467.50$
  $x = 400 \Longrightarrow \$585.60$

  The smallest LOS in the data is 7, and the largest is 228. Therefore predictions for 100 months and 200 months LOS are appropriate. But predictions for 1 month and 400 months LOS are extrapolations from the data and so should be considered a bit risky.

- Exercise 10.22
  90% confidence interval for β_0:

$$\begin{aligned}b_0 \pm t^*_{n-2} \cdot SE(b_0) &= b_0 \pm t^*_{57} \cdot SE(b_0)\\ &\approx b_0 \pm t^*_{50} \cdot SE(b_0)\\ &= 349.38 \pm (1.676)(18.10)\\ &= 349.38 \pm 30.34\\ &= \boxed{(319.04,\ 379.72)}\end{aligned}$$

  Interpretation:

  We are 90% confident that the average starting salary for female employees who have customer service jobs in Indiana banks is between \$319.04 and \$379.72 per week.

- Exercise 10.33
  (a)

$$\widehat{y} \pm t^* \cdot SE_{\hat{\mu}} = 423.2 \pm (2.009)(15.6) = \boxed{(\$391.86,\ \$454.54)}$$

  (b) (\$397.05, \$449.35)
  (c)
  1. The CI is red, the PI is green.
  2. About (\$390, \$420)
  3. About (\$275, \$650)

D. T-Bills and Inflation

- Exercise 10.6
  The scatterplot shows a fairly strong positive linear relationship.

$$\text{T-Bill Return } = 2.6662 + 0.6269 \times \text{Inflation}$$

- Exercise 10.7
  Four Steps:
  1.
  $$H_A\colon\ \beta_1 > 0$$
  $$H_0\colon\ \beta_1 \le 0$$
  2. $t = 6.32$, P-value$= (\frac{1}{2}) \times 0 = 0$
     Note: The P-value for this test is half the P-value given by MINITAB.
     (See Example 2 on pages 57–59 in the Notes.)
  3. Reject H_0 since P-value $= 0 < .05 = \alpha$.
  4. There is enough evidence to show that the rate of return on T-bills is positively related to the rate of inflation.

- Exercise 10.8

$$b_1 \pm t^*_{n-2} \cdot SE(b_1) = b_1 \pm t^*_{49} \cdot SE(b_1) \approx b_1 \pm t^*_{40} \cdot SE(b_1)$$
$$= 0.6269 \pm (2.021)(0.09924) = 0.6269 \pm 0.2006 = (0.4263, 0.8275)$$

  Interpretation:
  We are 95% confident that a 1% increase in the inflation rate is associated with between 0.4263% and 0.8275% increase in the rate of return on T-bills, on average.

- Exercise 10.23
  (a) β_0 is the mean rate of return on T-bills in the absence of inflation. We expect $\beta_0 > 0$ simply because no one will purchase T-bills from the federal government unless interest is paid. One can achieve a return of 0% simply by putting the money in a safe.
  (b) From MINITAB,
  $$\widehat{\beta}_0 = b_0 = 2.6662, \quad SE_{b_0} = 0.5038$$
  (c) Four Steps:
  1.
  $$H_A\colon\ \beta_0 > 0$$
  $$H_0\colon\ \beta_0 \le 0$$
  2. $t = 5.29$, P-value$= (\frac{1}{2}) \times 0 = 0$
  3. Reject H_0 since P-value $= 0 < .05 = \alpha$.
  4. There is enough evidence to show that the mean rate of return on T-bills in the absence of inflation exceeds 0%.

(continued)

(d)

$$b_0 \pm t^*_{n-2} \cdot SE(b_0) = \boxed{(1.6480,\ 3.6844)}$$

Interpretation:

We are 95% confident that the mean rate of return on T-bills in the absence of inflation is between 1.6480% and 3.6844%.

- Exercise 10.34
  (a) Verify the answer.
  (b) (0.561%, 9.410%)

- Exercise 10.42
  (a)

$$\hat{y} \pm t^*_{n-2,90\%} \cdot SE_{\hat{\mu}}$$

$$t \text{ Table: } t^*_{n-2,90\%} = t^*_{49,90\%} \approx t^*_{40,90\%} = 1.684$$

$$\implies \quad 4.986 \pm (1.684)(0.307) = 4.986 \pm 0.517 = \boxed{(4.469\%,\ 5.503\%)}$$

  (b) Unfortunately, MINITAB does not provide $SE_{\hat{y}}$ as part of the regression output so there's no easy way to calculate the answer.

E. Stocks and Bonds

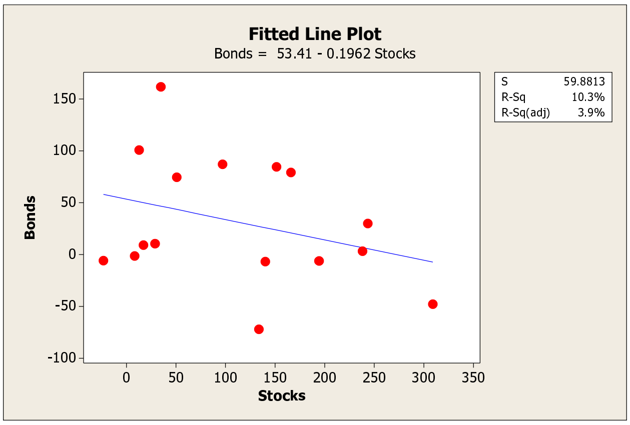
- Exercise 10.26
  (a) The fitted regression line is

$$\hat{y} = 53.41 - (0.1962)\,x$$

where

$$x = \text{Net cash flow into stock funds, in billions of \$}$$
$$y = \text{Net cash flow into bond funds, in billions of \$}$$

The fitted line plot suggests that there may be a negative relationship between net inflow into bond funds and net inflow into stock funds.



(b) Four Steps:

1. Regression model:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Hypotheses:

$$H_A\colon\ \beta_1 \neq 0$$
$$H_0\colon\ \beta_1 = 0$$

2. $t = \boxed{-1.27}$ $\quad P$-value$=\boxed{0.226}$

3. Fail to Reject H_0 since

$$P\text{-value} = 0.226 > .05 = \alpha$$

4. There is insufficient evidence to show that mean net cash flow into bond funds is linearly related to net cash flow into stock funds.

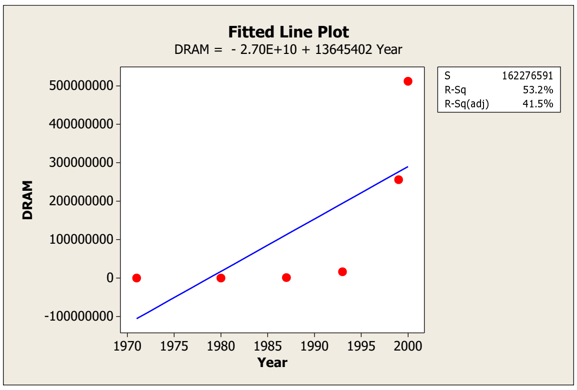
(c) The plot shows that the relationship between x and y is fairly weak (correlation $r = -0.321$.) There is *some* evidence of a negative relationship, but not *convincing* evidence.

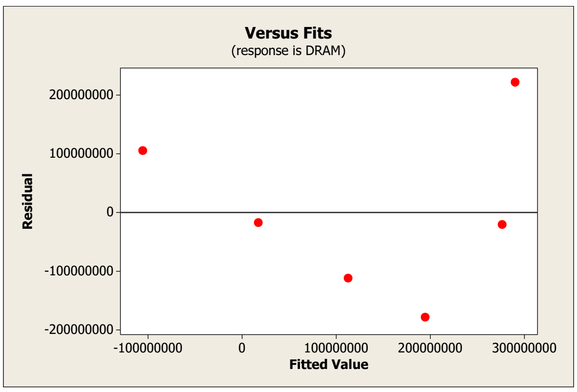
(d) Stop! It is not recommended to use this regression model for prediction or estimation since the predictor variable is not statistically significant.

F. Computer Memory

- Exercise 10.31

(a)



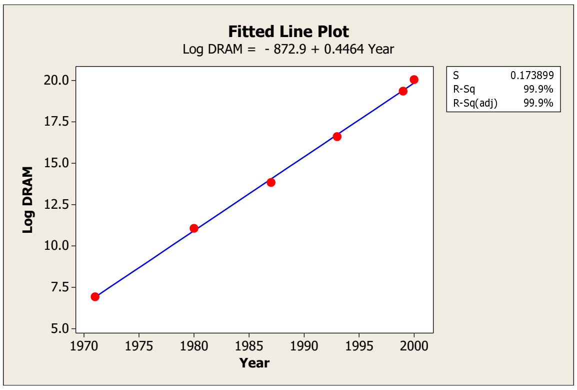


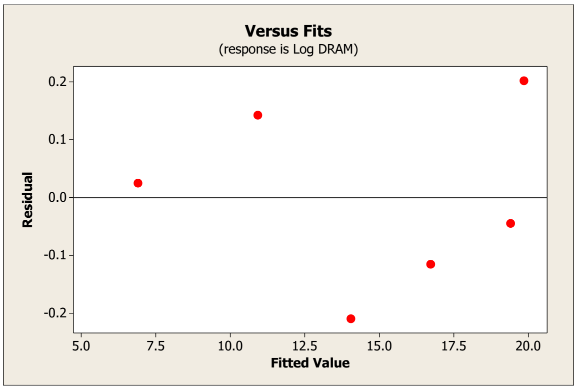
Questions to answer for (a):

1. $$\widehat{\text{DRAM}} = -27,000,000,000 + 13,645,402 \times \text{Year}$$

2. The residuals plot shows a clear pattern. The regression assumptions appear to be violated.

(b)





$$\log(\widehat{\text{DRAM}}) = -872.93 + 0.446390 \times \text{Year}$$

(c) 90% confidence interval for β_1:

$$\begin{aligned} b_1 \pm t_4^* \cdot SE(b_1) &= .44639 \pm (2.132)(.006856) \\ &= .44639 \pm .01462 \\ &= \boxed{(.43177\ ,\ .46101)} \end{aligned}$$

Interpretation:

We are 90% confident that log(DRAM) increases by between 0.43177 and 0.46101 every year, on average.

(d) A 90% prediction interval from MINITAB for log(DRAM) in 2002 is

$$(20.2971\ ,\ 21.1921)$$

so a 90% prediction interval for DRAM is

$$e^{20.2971} \text{ to } e^{21.1921} = \boxed{\text{(653,008,040 to 1,598,129,948) bits}}$$

G. Blood Alcohol Content

- Exercise 10.30

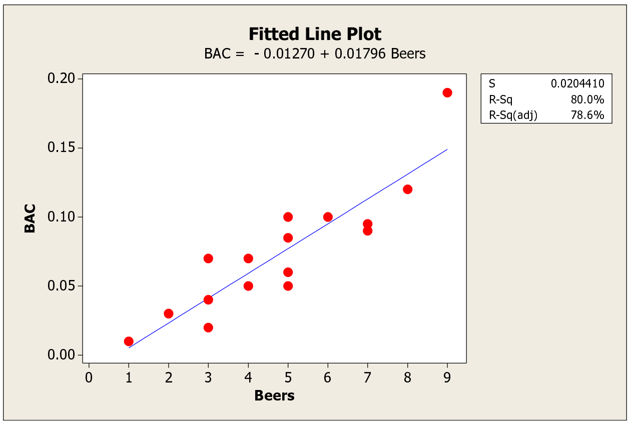
(a) The fitted regression line is

$$\widehat{y} = -0.01270 + (0.01796)\,x$$

where

$$x = \text{Number of beers consumed in 30 minutes}$$
$$y = \text{Blood alcohol content, in percent}$$

$R^2 = 80.0\%$. The plot shows a strong positive relationship.



(b) Hypothesis Test:

1.

$$H_A\colon\ \beta_1 > 0$$
$$H_0\colon\ \beta_1 \le 0$$

2. $t = \boxed{7.48}$
The P-value for this test is half the P-value given by MINITAB:

$$P\text{-value} = \left(\frac{1}{2}\right) \times 0 = \boxed{0}$$

3. Reject H_0 since P-value $= 0 < .05 = \alpha$.

4. There is sufficient evidence to show that mean BAC is positively related to the number of beers consumed in 30 minutes.

(c) 90% confidence interval for β_1:

$$\begin{aligned} b_1 \pm t^*_{n-2} \cdot SE(b_1) &= b_1 \pm t^*_{14} \cdot SE(b_1) \\ &= 0.017964 \pm (1.761)(.002402) \\ &= 0.017964 \pm .004229922 \\ &= \boxed{(0.0137\ ,\ 0.0222)} \end{aligned}$$

Interpretation:

We are 90% confident that each additional beer implies an increase of between 0.0137% and 0.0222% in mean BAC.

(continued)

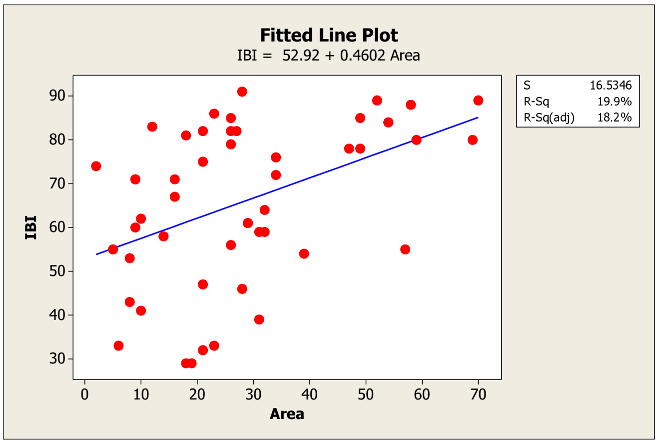
- Exercise 10.45

  We predict with 90% certainty that Steve's BAC is between 0.0400% and 0.1142%. This interval includes values larger than 0.08% so Steve can't be confident that he won't be arrested if he drives and is stopped.

H. Predicting Water Quality

- Exercise 10.13

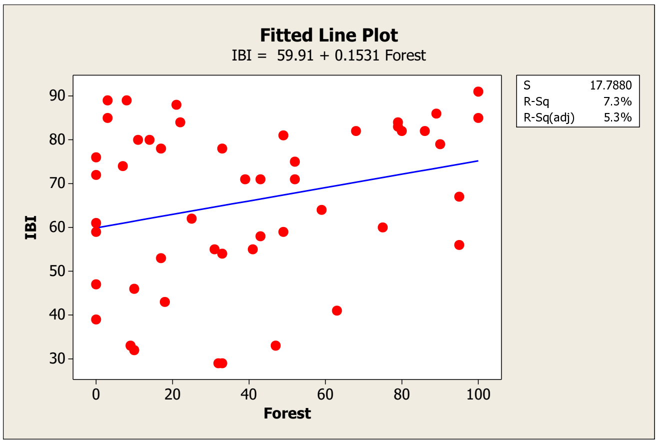
  (a)



$$\widehat{\text{IBI}} = 52.923 + 0.4602 \times \text{Area}$$

- Exercise 10.14

  (a)



$$\widehat{\text{IBI}} = 59.907 + 0.15313 \times \text{Forest}$$

(b) Either predictor variable can be used for regression since both are significant using $\alpha = 0.10$ (P-value for Area = 0.001, P-value for Forest = 0.061.)

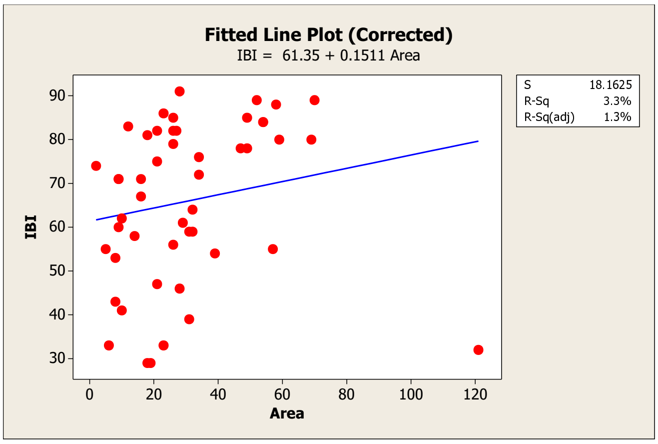
As a tie-breaker, choose Area as the preferred predictor on the basis of either of two equivalent comparisons:

(1) Regression standard deviation s for Area $= 16.53 < 17.79 = s$ for Forest.

(2) R^2 for Area $= 19.9\% > 7.3\% = R^2$ for Forest.

(c) Based on the model using Area, a 90% prediction interval for IBI is (47.48, 104.38).

(d)



$$\widehat{\text{IBI}} = 61.355 + 0.1511 \times \text{Area}$$

An outlier is a data point which is far away from most of the other data points in the scatterplot.

(e)

- After the data correction, Area is no longer a significant predictor at the 10% level (P-value= 0.209) and so is disqualified from consideration. However, since Forest is a significant predictor (P-value = 0.061), it can be used for regression.
- A 90% prediction interval for IBI using Forest is (34.33, 94.68).