Solutions: Homework 3, Required Problems

2.15. The plot shows a fairly steady rate of improvement until the mid-1980s, with much slower progress after that (the record has only been broken once since 1986).

2.17. (a) Both men (plusses) and women (open circles) show fairly steady improvement. Women have made more rapid progress, but their progress seems to have slowed, while men’s records may be dropping more rapidly in recent years. (b) The data supports the first claim, but does not seem to support the second.
2.21. (a) Price is explanatory (and so is on the horizontal axis). The plot shows a positive linear association. (b) $\bar{x} = 50$ cents/lb and $s_x = 16.3248$ cents/lb; $\bar{y} = 1.738\%$ and $s_y = 0.9278\%$. The standardized values are below;! The correlation is $r = 3.8206/4 = 0.955$. (c) Obviously, the calculator value should be the same.

<table>
<thead>
<tr>
<th>$z_x$</th>
<th>$z_y$</th>
<th>$z_xz_y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.2864</td>
<td>-1.3451</td>
<td>1.7303</td>
</tr>
<tr>
<td>-0.6126</td>
<td>-0.1595</td>
<td>0.0977</td>
</tr>
<tr>
<td>0.2450</td>
<td>-0.0517</td>
<td>-0.0127</td>
</tr>
<tr>
<td>0.3063</td>
<td>0.0884</td>
<td>0.0271</td>
</tr>
<tr>
<td>1.3476</td>
<td>1.4679</td>
<td>1.9783</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.8206</td>
</tr>
</tbody>
</table>

2.40. (a) The new speed and fuel consumption (respectively) values are $x^* = x \times 1.609$ and $y^* = y \times 1.609 \div 100 \div 3.785 \div 0.004251 y$. (The factor of $1/100$ is needed since we were measuring fuel consumption in liters/100 km.) The transformed data has the same correlation as the original—$r = -0.172$ (computed in the solution to Exercise 2.32)—since a linear transformation does not alter the correlation. The scatterplot of the transformed data is not shown here; it resembles (except for scale) the plot shown in the solution to Exercise 2.14. (b) The new correlation is $r^* = -0.043$; the new plot is even less linear than the first.
2.58. (a) The slope is \( b = r s_y / s_x = (0.6)(8)/(30) = 0.16 \), and the intercept is \( a = \bar{y} - b \bar{x} = 30.2 \).

(b) Julie’s predicted score is \( \hat{y} = 78.2 = 30.2 + 0.16 \times 300 \). (c) \( r^2 = 0.36 \); only 36% of the variability in \( y \) is accounted for by the regression, so the estimate \( \hat{y} = 78.2 \) could be quite different from the real score.

**Calculus Supplement Problem:**

2.2 Using the reasoning in (S2.3) and following, to fit the line
\[ \hat{y}_i = b x_i, \]
the mathematical problem is to find the value of \( b \) that minimizes
\[ S(b) = \sum (y_i - b x_i)^2. \]
Since this is a function of only one variable, \( b \), we can use ordinary derivatives:
\[ \frac{dS(b)}{db} = -2 \sum (y_i - bx_i) x_i = -2 \sum (y_i x_i - bx_i^2) = -2[\sum y_i x_i - b \sum x_i^2] = 0 \]
Thus,
\[ b = \frac{\sum x_i y_i}{\sum x_i^2} \]
Notice that it is the formula you would get from the least squares analysis with both parameters \( a \) and \( b \) in the model if \( \bar{x} \) happens to be 0.

Finally, we should verify that the \( b \) we obtained gives a minimum for \( S(b) \), which we will do by checking that the second derivative is positive:
\[ \frac{d^2S(b)}{db^2} = \frac{d}{db} \left[ -2[\sum y_i x_i - b \sum x_i^2] \right] = 2 \sum x_i^2 > 0 \]