EXERCISE 2.3

(a) The line drawn for part (a) will depend on each student’s subjective choice about the position of the line. For this reason, it has been omitted.

(b) \( b_2 = -1.514286 \)

\( b_1 = 10.8 \)

(c) \( \bar{y} = 5.5 \)

\( \bar{x} = 3.5 \)

\( \hat{y} = 5.5 \)
Exercise 2.3 (Continued)

(d)  

<table>
<thead>
<tr>
<th>$\hat{e}_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.714286</td>
</tr>
<tr>
<td>0.228571</td>
</tr>
<tr>
<td>−1.257143</td>
</tr>
<tr>
<td>0.257143</td>
</tr>
<tr>
<td>−1.228571</td>
</tr>
<tr>
<td>1.285714</td>
</tr>
</tbody>
</table>

$\sum \hat{e}_i = 0$.

(e) $\sum x_i \hat{e}_i = 0$

EXERCISE 2.6

(a) The intercept estimate $b_0 = −240$ is an estimate of the number of sodas sold when the temperature is 0 degrees Fahrenheit. Clearly, it is impossible to sell −240 sodas and so this estimate should not be accepted as a sensible one.

The slope estimate $b_1 = 8$ is an estimate of the increase in sodas sold when temperature increases by 1 Fahrenheit degree. One would expect the number of sodas sold to increase as temperature increases.

(b) $\hat{y} = −240 + 8 \times 80 = 400$

(c) She predicts no sodas will be sold below 30°F.

(d) A graph of the estimated regression line:
EXERCISE 2.9

(a) The repair period comprises those months between the two vertical lines. The graphical evidence suggests that the damaged motel had the higher occupancy rate before and after the repair period. During the repair period, the damaged motel and the competitors had similar occupancy rates.

(b) A plot of $MOTEL\_PCT$ against $COMP\_PCT$ yields:

There appears to be a positive relationship the two variables. Such a relationship may exist as both the damaged motel and the competitor(s) face the same demand for motel rooms.
Exercise 2.9 (continued)

(c)  \[ MOTEL\_PCT = 21.40 + 0.8646 \times COMP\_PCT. \]

The competitors’ occupancy rates are positively related to motel occupancy rates, as expected. The regression indicates that for a one percentage point increase in competitor occupancy rate, the damaged motel’s occupancy rate is expected to increase by 0.8646 percentage points.

(d) 

![Figure xr2.9(d) Plot of residuals against time](image)

The residuals during the occupancy period are those between the two vertical lines. All except one are negative, indicating that the model has over-predicted the motel’s occupancy rate during the repair period.

(e) We would expect the slope coefficient of a linear regression of \( MOTEL\_PCT \) on \( RELPRICE \) to be negative, as the higher the relative price of the damaged motel’s rooms, the lower the demand will be for those rooms, holding other factors constant.

\[ MOTEL\_PCT = 166.66 - 122.12 \times RELPRICE \]

(f) From this equation, we have that:

\[ E(MOTEL\_PCT) = \beta_1 + \beta_2 \times REPAIR = \begin{cases} \beta_1 + \beta_2 & \text{if } REPAIR = 1 \\ \beta_1 & \text{if } REPAIR = 0 \end{cases} \]
**Exercise 2.9(f) (continued)**

The expected occupancy rate for the damaged motel is $\beta_1 + \beta_2$ during the repair period; it is $\beta_1$ outside of the repair period. Thus $\beta_2$ is the difference between the expected occupancy rates for the damaged motel during the repair and non-repair periods.

The estimated regression is:

$$MOTEL \_ PCT = 79.3500 - 13.2357 \times REPAIR$$

In the non-repair period, the damaged motel had an estimated occupancy rate of 79.35%. During the repair period, the estimated occupancy rate was 79.35$ - 13.24 = 66.11\%$. Thus, it appears the motel did suffer a loss of occupancy and profits during the repair period.

(g) From the earlier regression, we have

$$MOTEL_0 = b_1 = 79.35\%$$

$$MOTEL_1 = b_1 + b_2 = 79.35 - 13.24 = 66.11\%$$

For competitors, the estimated regression is:

$$COMP \_ PCT = 62.4889 + 0.8825 \times REPAIR$$

Thus,

$$COMP_0 = b_1 = 62.49\%$$

$$COMP_1 = b_1 + b_2 = 62.49 + 0.88 = 63.37\%$$

During the non-repair period, the difference between the average occupancies was:

$$MOTEL_0 - COMP_0 = 79.35 - 62.49 = 16.86\%$$

During the repair period it was

$$MOTEL_1 - COMP_1 = 66.11 - 63.37 = 2.74\%$$

This comparison supports the motel’s claim for lost profits during the repair period. When there were no repairs, their occupancy rate was 16.86% higher than that of their competitors; during the repairs it was only 2.74% higher.

(h) $MOTEL \_ PCT - COMP \_ PCT = 16.8611 - 14.1183 \times REPAIR$

The intercept estimate in this equation (16.86) is equal to the difference in average occupancies during the non-repair period, $MOTEL_0 - COMP_0$. The sum of the two coefficient estimates $16.86 + (-14.12) = 2.74$ is equal to the difference in average occupancies during the repair period, $MOTEL_1 - COMP_1$.

This relationship exists because averaging the difference between two series is the same as taking the difference between the averages of the two series.
EXERCISE 2.12

(a) The scatter plot in the figure below shows a positive relationship between selling price and house size.

![Figure xr2.12(a) Scatter plot of selling price and living area](image)

(b) The estimated equation for all houses in the sample is

\[
\hat{\text{SPRICE}} = -30069 + 9181.7 \times \text{LIVAREA}
\]

The coefficient 9181.7 suggests that selling price increases by approximately $9182 for each additional 100 square foot in living area. The intercept, if taken literally, suggests a house with zero square feet would cost $-30,069, a meaningless value.
Exercise 2.12 (continued)

(c) The estimated quadratic equation for all houses in the sample is

$$SPRICE = 57728 + 212.611 \times LIVAREA^2$$

For a home with 1500 square feet of living space, the marginal effect is 6378.33:
That is, adding 100 square feet of living space to a house of 1500 square feet is estimated
To increase its expected price by approximately $6378.

(d) The quadratic model appears to fit the data better; it is better at capturing
The proportionally higher prices for large houses.

\[SSE\text{ of linear model, (b): } SSE = \sum \hat{e}_i^2 = 2.23 \times 10^{12}\]

\[SSE\text{ of quadratic model, (c): } SSE = \sum \hat{e}_i^2 = 2.03 \times 10^{12}\]

The \(SSE\) of the quadratic model is smaller, indicating that it is a better fit.

(e) The estimated equation for houses that are on large lots in the sample is:

$$SPRICE = 113279 + 193.83 \times LIVAREA^2$$

The estimated equation for houses that are on small lots in the sample is:

$$SPRICE = 62172 + 186.86 \times LIVAREA^2$$

The intercept can be interpreted as the expected price of the land – the selling price for a
house with no living area. The coefficient of \(LIVAREA\) has to be interpreted in the context
of the marginal effect of an extra 100 square feet of living area, which is \(2\beta_2 LIVAREA\).
Thus, we estimate that the mean price of large lots is $113,279 and the mean price of small
lots is $62,172. The marginal effect of living area on price is $387.66 \times LIVAREA for
houses on large lots and $373.72 \times LIVAREA for houses on small lots.
Exercise 2.12(e) (continued)

(f) The following figure contains the scatter diagram of \( \text{PRICE} \) and \( \text{AGE} \) as well as the estimated equation which is

\[
\hat{\text{PRICE}} = 137404 - 627.16 \times \text{AGE}
\]

We estimate that the expected selling price is $627 less for each additional year of age. The estimated intercept, if taken literally, suggests a house with zero age (i.e., a new house) would cost $137,404. The model residuals plotted below show an asymmetric pattern, with some very large positive values. For these observations the linear fitted model under predicts the selling price.
Exercise 2.12(f) (continued)

The following figure contains the scatter diagram of \( \ln(\text{PRICE}) \) and \( \text{AGE} \) as well as the estimated equation which is

\[
\ln(\text{SPRICE}) = 11.746 - 0.00476 \text{AGE}
\]

In this estimated model, each extra year of age reduces the selling price by 0.48%. To find an interpretation from the intercept, we set \( \text{AGE} = 0 \), and find an estimate of the price of a new home as

\[
\exp\left(\ln\left(\text{SPRICE}\right)\right) = \exp(11.74597) = $126,244
\]

The following residuals from the fitted regression of \( \ln(\text{SPRICE}) \) on \( \text{AGE} \) show much less of problem with under-prediction; the residuals are distributed more symmetrically around zero. Thus, based on the plots and visual fit of the estimated regression lines, the log-linear model is preferred.

(g) The estimated equation for all houses is:

\[
\text{SPRICE} = 115220 + 133797L\text{GELOT}
\]

The estimated expected selling price for a house on a large lot \( (L\text{GELOT} = 1) \) is \( 115220 + 133797 = $249017 \). The estimated expected selling price for a house not on a large lot \( (L\text{GELOT} = 0) \) is $115220.
EXERCISE 2.14

(a) There appears to be a positive association between \(VOTE\) and \(GROWTH\).

(b) The estimated equation for 1916 to 2008 is

\[
\hat{VOTE} = 50.848 + 0.88595GROWTH
\]

The coefficient 0.88595 suggests that for a 1 percentage point increase in the growth rate of \(GDP\) in the 3 quarters before the election there is an estimated increase in the share of votes of the incumbent party of 0.88595 percentage points.

We estimate, based on the fitted regression intercept, that that the incumbent party’s expected vote is 50.848% when the growth rate in \(GDP\) is zero. This suggests that when there is no real \(GDP\) growth, the incumbent party will still maintain the majority vote.

A graph of the fitted line and data is shown in the following figure.

(c) The estimated equation for 1916 - 2004 is

\[
\hat{VOTE} = 51.053 + 0.877982GROWTH
\]

The actual 2008 value for growth is 0.220. The predicted vote share for the incumbent party \(\hat{VOTE}_{2008} = 51.246\)
EXERCISE 2.14 (CONTINUED)

(d) The figure below shows a plot of \( VOTE \) against \( INFLATION \). There appears to be a negative association between the two variables.

The estimated equation (plotted in the figure below) is:

\[
\hat{VOTE} = 53.408 - 0.444312 \text{INFLATION}
\]

We estimate that a 1 percentage point increase in inflation during the incumbent party’s first 15 quarters reduces the share of incumbent party’s vote by 0.444 percentage points.

The estimated intercept suggests that when inflation is at 0% for that party’s first 15 quarters, the expected share of votes won by the incumbent party is 53.4%; the incumbent party is predicted to maintain the majority vote when inflation, during its first 15 quarters, is at 0%.