Page 697, exercise 10.6 For the following dataset, x is measured in average degree-days per day and y is daily gas consumption in hundreds of cubic feet.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>15.6</td>
<td>26.8</td>
<td>37.8</td>
<td>36.4</td>
<td>35.5</td>
<td>18.6</td>
<td>15.3</td>
<td>7.9</td>
<td>0.0</td>
</tr>
<tr>
<td>y</td>
<td>5.2</td>
<td>6.1</td>
<td>8.7</td>
<td>8.5</td>
<td>8.8</td>
<td>4.9</td>
<td>4.5</td>
<td>2.5</td>
<td>1.1</td>
</tr>
</tbody>
</table>

(a) Find the equation of the least-squares line.

Here is the relevant output from S+ (a different statistics program from Minitab):

Coefficients:

| Value  | Std. Error | t value | Pr(>|t|) |
|--------|------------|---------|---------|
| (Intercept) | 1.2324 | 0.2860 | 4.3084 | 0.0035 |
| ddays | 0.2022 | 0.0114 | 17.6627 | 0.0000 |

Residual standard error: 0.4345 on 7 degrees of freedom
Multiple R-Squared: 0.9781

Therefore, the equation for the least-squares line is

\[
\text{gas} = 1.2324 + 0.2022 \text{ deegedays}
\]

(b) Test the null hypothesis that the slope is zero and describe your conclusion.

The t statistic for the relevant test is given in the row above marked ddays. This statistic, 17.6627, is very large and the corresponding p-value is very close to zero. Therefore, we reject \( H_0 : \beta_1 = 0 \) and conclude that the true slope is nonzero.

(c) Give a 95% confidence interval for the slope.

The only thing not given in the table that we need for a confidence interval is the value of \( t^* \) for a 95% interval on 7 degrees of freedom. From Table D, we see that \( t^* = 2.365 \). Therefore, the confidence interval is \( 0.2022 \pm 2.365(0.0114) \), or \( 0.2022 \pm 0.0270 \).

(d) The parameter \( \beta_0 \) corresponds to natural gas consumption for cooking, hot water, and other uses when there is no demand for heating. Give a 95% confidence interval for this parameter.

We may use the same value of \( t^* \) here as in (c), but now we use the values from the Intercept row instead of the ddays row. Thus, the confidence interval is \( 1.2324 \pm 2.365(0.2860) \), or \( 1.2324 \pm 0.6764 \).
Page 697, exercise 10.8 The data below show the lean of the Leaning Tower of Pisa in tenths of a millimeter more than 2.9 meters for the years from 1975 to 1987.

<table>
<thead>
<tr>
<th>Year</th>
<th>75</th>
<th>76</th>
<th>77</th>
<th>78</th>
<th>79</th>
<th>80</th>
<th>81</th>
<th>82</th>
<th>83</th>
<th>84</th>
<th>85</th>
<th>86</th>
<th>87</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lean</td>
<td>642</td>
<td>644</td>
<td>656</td>
<td>667</td>
<td>673</td>
<td>688</td>
<td>696</td>
<td>713</td>
<td>717</td>
<td>725</td>
<td>742</td>
<td>757</td>
<td></td>
</tr>
</tbody>
</table>

(a) Plot the data. Does the trend in lean over time appear to be linear?

![Plot of data](image)

The plot above shows a pattern which looks very nearly linear.

(b) What is the equation of the least-squares line? What percentage of the variation in lean is explained by this line?

Here is the relevant output from S+:

Coefficients:

|            | Value | Std. Error | t value | Pr(>|t|) |
|------------|-------|------------|---------|----------|
| (Intercept)| -61.1209 | 25.1298 | -2.4322 | 0.0333   |
| year       | 9.3187  | 0.3099    | 30.0686 | 0.0000   |

Residual standard error: 4.181 on 11 degrees of freedom
Multiple R-Squared: 0.988

This gives the least-squares equation as

\[
\text{lean} = -61.121 + 9.319 \times \text{year}.
\]

Since the value of \(r\)-squared is 98.8\%, we conclude that 98.8\% of the variation in lean is explained by this line.
(c) Give a 95% confidence interval for the average rate of change (tenths of a millimeter per year) of the lean.

In this case, the average rate of change in tenths of a millimeter per year is nothing but the slope \( \beta_1 \) (do you see why?). The 95% confidence interval for \( \beta_1 \) is \( 9.319 \pm 2.201(0.310) \), or \( 9.319 \pm 0.682 \).

Page 703, exercise 10.24 The corn yields of Table 10.2 (page 696) and the soybean yields of Table 10.3 (page 703) both vary over time for similar reasons, including improved technology and weather conditions. Let's examine the relationship between the two yields.

(a) Plot the two yields with corn on the x axis and soybeans on the y axis. Describe the relationship.

The plot above shows an obvious positive association which appears mostly linear. There may be a slight upward curvature, but overall the impression is a linear relationship.

(b) Find the correlation. How well does it summarize the relation?

Here is the relevant output from S+:

Coefficients:

|            | Value  | Std. Error | t value | Pr(>|t|) |
|------------|--------|------------|---------|----------|
| Intercept  | 12.175 | 1.0766     | 11.3092 | 0.0000   |
| corn       | 0.1831 | 0.0114     | 16.0352 | 0.0000   |

Residual standard error: 1.695 on 38 degrees of freedom
Multiple R-Squared: 0.8712

We may find the correlation by taking the square root of the \( r \)-squared and then giving it the same sign as the slope. Doing this gives a correlation of \( r = 0.933 \). Alternatively,
we could have found this correlation directly using the correlation function. Because the relationship between the variables in question appears quite linear, the correlation describes that relationship quite well. However, there is some curvature to the relationship, as noted above, and the value of $r$ is unable to describe this fact at all.

(c) Use corn yield to predict soybean yield. Give the equation and the results of the significance test for the slope. This test also tests the null hypothesis that the two yields are uncorrelated.

Using the output above, the equation is

$$\text{soybeans} = 12.175 + 0.183\text{corn}.$$  

With a $t$ statistic giving a $p$-value of essentially zero, the test of the significance of the slope is unequivocal: We reject $H_0 : \beta_1 = 0$ and conclude that the true slope is nonzero.

(d) Obtain the residuals from the model in part (c) and plot them versus time. Describe the pattern.

There is a clear pattern in the plot above. The residuals in the middle tend to be negative, while the residuals from very early or very late years tend to be positive. This indicates that there is some pattern due to time which the model relating corn to soybeans has not accounted for. Perhaps corn technology improved faster than soybean technology early on but then soybean technology started to catch up. In any case, the residual plot points to an interesting feature of the data we may want to investigate.