1.5 - Method of Least Squares

Idea: You have collected or been given a set of data and you want to find the line that best fits the data. A way to find this line is the Method of Least Squares.

Find the “best-fitting” line to the data below.

How do we know if we have the “best-fitting” line?

Finding the Least-Squares Line on the Calculator

1. Turn your “Diagnostics On”: Press \[2^{nd}\] [0] to bring up the catalog of functions in your calculator. Scroll down until you see “Diagnostic On” and press [ENTER] twice. (You only need to do this step once, unless your calculator has had all its batteries replaced or your calculator has been reset.)

2. Enter your data into lists: Press [STAT] Select 1:Edit and make sure your lists L1 and L2 are cleared (Cursor up to the name of the list and press [CLEAR] and [ENTER]). Once your lists are cleared, type your x-values into L1 and your y-values into L2. (If you do not see the lists L1 or L2, press [STAT], select 5:SetUpEditor, hit [ENTER] and then return to your lists.)

3. Quit to your home screen by pressing \[2^{nd}\] [MODE].

4. Find the least-squares line and store it in your calculator: Press [STAT]. Move right to CALC. Select 4:LinReg(ax+b). Press [VARS]. Go right to Y-VARS. Select 1:Function and then select the name of where your function is stored, say 1:Y1. Finish by pressing [ENTER].
Ex: Find the least-squares line for the following data set:
(1, 1), (2, 3), (3, 4), (4, 3), (5, 6)

Besides the coefficients needed to write the equation of the least-squares line of our data, the calculator also provides us with the value of $r$. This value is called the correlation coefficient. We use the value of $|r|$ to determine how good of a job a line does in predicting accurate values for our data. The closer $|r|$ is to 1, the better job the line does.

Ex: In our previous data set, is a line a good representation of our data? Why or why not?

Using a Linear Model to Make Predictions

Ex: Using our above data set and linear model, predict the value of $y$ when $x = 2.5$.

Ex: Using our above data set and linear model, predict the value of $x$ when $y = 5$. 
Ex: (From Tan) As online attacks persist, spending on information security software continues to rise. The following table gives the forecast for the worldwide sales (in billions of dollars) of information security software through 2007.

<table>
<thead>
<tr>
<th>Year</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spending</td>
<td>6.8</td>
<td>8.3</td>
<td>9.8</td>
<td>11.3</td>
<td>12.8</td>
<td>14.9</td>
</tr>
</tbody>
</table>

(a) By letting $x = 0$ represent the year 2002, find the least-squares line for this data. (Round all coefficients to 4 decimal places, if necessary.)

(b) Is this a good representation of the data? Why or why not?

(c) Use your unrounded results to estimate the spending on information security software in 2008, assuming the trend continues?

(d) Use your unrounded results to determine the first year $12,500,000,000$ was spent on information security software worldwide.

Ex: The following table gives the number of cell phones in a particular area (in thousands) in the given years.

<table>
<thead>
<tr>
<th>Year</th>
<th>1985</th>
<th>1990</th>
<th>1996</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Phones</td>
<td>10.2</td>
<td>15</td>
<td>21.23</td>
</tr>
</tbody>
</table>

(a) By letting $x$ represent the number of years since 1985, find the least-squares line for this data. (Round all coefficients to 4 decimal places, if necessary.)

(b) Use your unrounded results to estimate the number of cell phones that would have been in the area in 2000, assuming the trend continued?

(c) Use your unrounded results to determine the first year when 30,000 cell phones were in the area.