Section 1.5: Linear Regression

Regression is the method of finding the best fitting formula for a data set. Note: Best fitting does not always mean that the formula goes through each data point.

Example: The number of applications to medical schools in the United States increased rapidly from 1988 to 1994 as indicated by the data in the table. Let the years be represented by the last two digits (i.e., time starts in 1900) and applications are given in thousands.

<table>
<thead>
<tr>
<th>Year</th>
<th>88</th>
<th>89</th>
<th>90</th>
<th>91</th>
<th>92</th>
<th>93</th>
<th>94</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apps</td>
<td>22</td>
<td>27</td>
<td>30</td>
<td>33</td>
<td>37</td>
<td>40</td>
<td>45</td>
</tr>
</tbody>
</table>

B) Find the regression line (least squares line).

\[ y = 3.1714x - 254.5429 \]

C) Predict the number of applications in the year 1990.

\[ x = 90 \quad y = ? \]

\[ y = 3.1714 (90) - 254.5429 \]

\[ = 30.8931 \]

50,893 apps

D) In what year would the number of applications be 79,500?

\[ y = 79.5 \quad x = ? \]

\[ x = 105.324 \]

2005
Example: The following data shows the chlorine residual measured in ppm (parts per million) for a swimming pool at various times after it has been treated with chemicals.

<table>
<thead>
<tr>
<th>Number of hrs.</th>
<th>2</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlorine residual</td>
<td>1.8</td>
<td>1.95</td>
<td>1.5</td>
<td>1.4</td>
<td>1.25</td>
<td>1.1</td>
<td>1.1</td>
<td>0.9</td>
</tr>
</tbody>
</table>

A) Find the regression line.

\[ y = -0.0956x + 1.9727 \]

b) What is the concentration 24 hours after treatment?

\[ -0.3217 \]