Math142 Lecture Notes

1.8 - Regression and Mathematical Models

The following graphs are plots of various sets of data. For each plot, find the type of function that best fits the overall “shape” of the curve.

Factors to consider when choosing a regression model

1. General shape of the scatterplot.

2. The correlation coefficient ($r = \text{value}$).

3. Will the model be used for interpolation (analyze points between the highest and lowest data values) or extrapolation (analyze points beyond the highest and lowest data values)? If the model will be used for extrapolation, the shape of the curve beyond the data points needs to be considered.

4. Keep it as simple as possible.

Use the regression function in the calculator to determine the best fitting model for each data set and then use the model to answer the questions that follow.


<table>
<thead>
<tr>
<th>May</th>
<th>June</th>
<th>July</th>
<th>Aug</th>
<th>Sept</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>April</th>
<th>May</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>240</td>
<td>260</td>
<td>300</td>
<td>350</td>
<td>380</td>
<td>440</td>
<td>510</td>
<td>550</td>
<td>625</td>
<td>730</td>
<td>790</td>
<td>875</td>
</tr>
</tbody>
</table>

(a) Standardize the data and then determine the best mathematical model that gives sensible extrapolation both in the past and future.

(b) Use the model to determine the estimated commerce in May 1998.

(c) Use the model to find the average rate of change from July 1996 to Oct 1996 and compare it to the average rate of change from July 1997 to Oct 1997.
2. The following table shows the costs of producing a certain type of shoe. In the table $x$ represents the number of pairs of shoes produced and $y$ is the average cost of the pair of shoes in dollars.

<table>
<thead>
<tr>
<th>$x$</th>
<th>4</th>
<th>7</th>
<th>13</th>
<th>17</th>
<th>19</th>
<th>27</th>
<th>33</th>
<th>49</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>110</td>
<td>75</td>
<td>65</td>
<td>67</td>
<td>55</td>
<td>45</td>
<td>55</td>
<td>60</td>
</tr>
</tbody>
</table>

(a) Determine the best model to represent the average cost if the data is to be interpolated.

(b) Use the model to determine how many shoes should be produced to keep costs at a minimum.

(c) Use the model to find the average rate of change in the cost of production when $x = 5$ and $\Delta x = 10$

(d) If the company can spend no more than $75 on production, how many shoes can they produce?

3. The following table shows the profits earned for the same type of shoe in the example above. ($x$ and $y$ represent the same units as above)

<table>
<thead>
<tr>
<th>$x$</th>
<th>14</th>
<th>28</th>
<th>35</th>
<th>49</th>
<th>56</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>43</td>
<td>120</td>
<td>127</td>
<td>86</td>
<td>32</td>
</tr>
</tbody>
</table>

(a) Determine the best model to represent the profit if the data is to be interpolated.

(b) Use the model to determine how many shoes should be produced to maximize profits.

(c) Find the break-even point for producing this certain type of shoe.
**Important Reminders**

- Go to $\boxed{y=}$ to check that your STAT PLOT is turned on, and resize your window for each set of data.
- Standardize the data, and enter into your calculator in STAT, then EDIT
- For logarithmic curves, do not start the first x-value at **zero**.
- If the correlation coefficient ($r$ value) is not showing, go to $2^{nd}$ CATALOG, and then scroll down to DIAGNOSTICS ON.
- When answering the question, “Which model fits best”, your answer must show a correlation coefficient and be reasonable with the endline behavior if extrapolating the data.
- Keep it simple. A **cubic regression** always has a higher $R^2$ value than **quadratic regression**, but is the $R^2$ value improved enough to go with the more complicated formula? Likewise, a **quartic regression** is always a better fit than **cubic regression**, but discuss the various $R^2$ values to defend your choice for the best model.