Regression Analysis

We selected Q4.2.24 (p.205) as an example of using StatCrunch to perform a regression analysis.

Attending Class  The following data represent the number of days absent, $x$, and the final grade, $y$, for a sample of college students in a general education course at a large state university.

<table>
<thead>
<tr>
<th>No. of absences, $x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final grade, $y$</td>
<td>89.2</td>
<td>86.4</td>
<td>83.5</td>
<td>81.1</td>
<td>78.2</td>
<td>73.9</td>
<td>64.3</td>
<td>71.8</td>
<td>65.5</td>
<td>66.2</td>
</tr>
</tbody>
</table>

*Source: College Teaching, 53(1), 2005.*

(a) Find the least-squares regression line treating number of absences as the explanatory variable and final grade as the response variable.

(b) Interpret the slope and $y$-intercept, if appropriate.

(c) Predict the final grade for a student who misses five class periods and compute the residual. Is the final grade above or below average for this number of absences?

(d) Draw the least-squares regression line on the scatter diagram of the data.

(e) Would it be reasonable to use the least-squares regression line to predict the final grade for a student who has missed 15 class periods? Why or why not?

(a) Find the least-squares regression line treating number of absences as the explanatory variable and final grade as the response variable.

Step 1: 1) Download the data.

2) Click Stat → Regression → Simple Linear.
Step 2: 1) Choose **Absences** for X variable, and **Final grade** for Y variable.  
2) Click **Compute!**

The least-squares regression line is calculated.

\[ y = 88.7327 - 2.82727 \times x \], where \( y \) represents the predicted final grade and \( x \) represents # of absences. 

Round both the slope and \( y \)-intercept to 4 decimal places and rearrange the terms for the preferred format of a regression line equation, \( y = -2.8273x + 88.7327 \).

(Note: We should make sure there is a correlation between final grade and # of absences by comparing R (correlation coefficient) with \( r \) obtained from Table II prior to accepting the regression line equation calculated by StatCrunch. If there is no correlation between the two variables, the regression line equation is meaningless and should not be used.)
(b) Interpret the slope and $y$-intercept, if appropriate.

Slope $\approx -2.8273$  (rise/run where rise is grade change and run is one absence)

For every additional absence, the student's final grade drops 2.8273 points on average.

$y$ - intercept $\approx 88.7327$

The average final grade of a student who misses no class period is 88.7327.

(c) Predict the final grade for a student who misses five class periods and compute the residual. Is the final grade above or below average for this number of absences?

Since $y = -2.8273x + 88.7327$ and the given number of absence $x = 5$,

$y = -2.8273(5) + 88.7327$

$= 74.5962 \approx 74.60$.

From the paired data given in the table, when $x=5$, observed final grade $y=73.9$.

Residual = Observed final grade – Predicted final grade = $y - y = 73.9 - 74.60 = -0.7$.

Therefore, the final grade 73.9 is 0.7 below average for this number of absences.

(d) Draw the least-squares regression line on the scatter diagram of the data.

If the output window for part (a) is still on the display, you can simply click the error at the lower right corner of the window to view the least-squares regression line on the scatter diagram.

If you already closed the window, go to part (a) for the steps of regression analysis.

You can use Graph properties to customize the labeling of the graph if desired.

Step 1: Click **Options → Edit**.
Step 2: 1) Scroll down the dialogue box. Under the **Graph properties**: enter **Number of Absences** for X-axis label, **Final Grade** for Y-axis label, and **Grades vs Absences** for Title.  
2) Click **Compute!**

Step 3: Click the **arrow button** to see the scatter diagram with the regression line drawn.
e) Would it be reasonable to use the least-squares regression line to predict the final grade for a student who has missed 15 class periods? Why or why not?

Recall the given data set:

<table>
<thead>
<tr>
<th>No. of absences, x</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final grade, y</td>
<td>89.2</td>
<td>86.4</td>
<td>83.5</td>
<td>81.1</td>
<td>78.2</td>
<td>73.9</td>
<td>64.3</td>
<td>71.8</td>
<td>65.5</td>
<td>66.2</td>
</tr>
</tbody>
</table>

*Source: College Teaching, 53(1), 2005.*

The domain of $x$, No. of absences, is between 0 and 9 inclusive. Since 15 is not within the domain, the least-squares regression line using the given data set is not a good model for a student who has missed 15 class periods.