Correlation and Regression Example solutions

A statistics instructor at a large western university would like to examine the relationship (if any) between the number of optional homework problems students do during the semester and their final course grade. She randomly selects 12 students for study and asks them to keep track of the number of these problems completed during the course of the semester. At the end of the class each student’s total is recorded along with their final grade. The data follow in Table 1.

1) For this setting identify the response variable.
   Course Grade

2) For this setting, identify the predictor variable.
   # of optional homework problems completed

3) Compute the linear correlation coefficient – r – for this data set
   See calculations on page 2

4) Classify the direction and strength of the correlation
   Moderate Positive

5) Test the hypothesis for a significant linear correlation. \( \alpha = 0.05 \)
   See calculations on page 2

6) What is the valid prediction range for this setting?
   The valid prediction range is the range of the “predictor” variable. In this case its from 51 - 91

7) Use the regression equation to predict a student’s final course grade if 75 optional homework assignments are done.
   Grade =44.8 + 0.355(75) = 71.4

8) Use the regression equation to compute the number of optional homework assignments that need to be completed if a student expects an 85.
   \( 85 = 44.8 + 0.355(x) \) \( \Rightarrow x \approx 113 \). This value is out of the prediction range so we have no confidence in it.
3) Calculations for problem 3

\[
r = \frac{n \Sigma(xy) - \Sigma x \cdot \Sigma y}{n(n-1)s_x s_y}
\]

\[
= \frac{12(62254) - (873)(848)}{12(11)(11.99)(4.81)}
\]

\[
= 0.885
\]

5) Hypothesis test for significant linear correlation

A) Ho: \(\rho = 0\)
Ha: \(\rho \neq 0\)

B) \(\alpha = 0.05;\) df = 10; \(t_{\text{crit}} = \pm 2.228\)

C)

\[
t_{\text{calc}} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}
\]

\[
= \frac{0.885}{\sqrt{\frac{1-(0.885)^2}{10}}}
\]

\[
= 6.01
\]

D) The decision graphic

E) Reject Ho

F) At a significance level of 0.05 we can conclude that there is a significant linear correlation between the number of homework assignments and a student’s final grade. Furthermore, we can conclude that this correlation is +
Output 1: Descriptive statistics for the grade versus homework study

Descriptive Statistics: Problems, CourseGrade

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>TrMean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problems</td>
<td>12</td>
<td>72.75</td>
<td>76.50</td>
<td>73.10</td>
<td>11.99</td>
<td>3.46</td>
</tr>
<tr>
<td>CourseGr</td>
<td>12</td>
<td>70.67</td>
<td>72.00</td>
<td>70.80</td>
<td>4.81</td>
<td>1.39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problems</td>
<td>51.00</td>
<td>91.00</td>
<td>62.75</td>
<td>82.50</td>
</tr>
<tr>
<td>CourseGr</td>
<td>62.00</td>
<td>78.00</td>
<td>66.25</td>
<td>74.50</td>
</tr>
</tbody>
</table>

Output 2: Regression output for the grade versus homework study

Regression Analysis: CourseGrade versus Problems

The regression equation is
CourseGrade = 44.8 + 0.355 Problems

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>44.827</td>
<td>4.344</td>
<td>10.32</td>
<td>0.000</td>
</tr>
<tr>
<td>Problems</td>
<td>0.35519</td>
<td>0.05898</td>
<td>6.02</td>
<td>0.000</td>
</tr>
</tbody>
</table>

S = 2.346    R-Sq = 78.4%  R-Sq(adj) = 76.2%

Figure 1: Regression plot for the grade versus homework study