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 fiaal 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 ttal is recorded along with their final grade. The data follow in Table 1.

Table 1: Course grade versus the number of optional homework problems completed.

Problems	CourseGrade	Prb*Grd
51	62	3162
58	68	3944
62	66	4092
65	66	4290
68	67	4556
76	72	5472
77	73	5621
78	72	5616
78	78	6084
84	73	6132
85	76	6460
91	75	6825
873	848	62254
ΣPrb	ΣGrd	ΣPrb*Grd

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

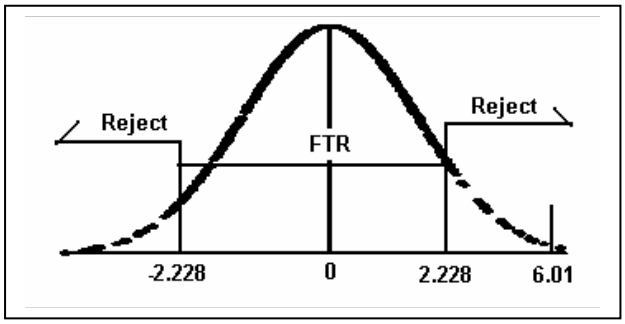
A) Ho: $\rho = 0$
Ha: $\rho \neq 0$

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

C)

$$t_{calc} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$
$$= \frac{0.885}{\sqrt{\frac{1-(0.885)^2}{10}}}$$
$$= 6.01$$

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

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Problems	12	72.75	76.50	73.10	11.99	3.46
CourseGr	12	70.67	72.00	70.80	4.81	1.39

Variable	Minimum	Maximum	Q1	Q3
Problems	51.00	91.00	62.75	82.50
CourseGr	62.00	78.00	66.25	74.50

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

Predictor	Coef	SE Coef	T	P
Constant	44.827	4.344	10.32	0.000
Problems	0.35519	0.05898	6.02	0.000

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

Figure 1: Regression plot for the grade versus homework study