Answer Key

Practice Problem: Regression

The following dataset contains datapoints for the number of hours a student studied and the grade they earned on the exam.

Hours of Study	Grade on Exam
2	50
3	71
4	87
5	95

1. Find the equation for the least-squares regression line for this dataset **by hand** and show your work.
   If you want, you may use excel to set up a table similar to this:

x	y	xy	x^2	y^2
1	30	30	1	900
2	45	90	4	2025
3	51	153	9	2601
4	57	228	16	3249
5	60	300	25	3600
6	65	390	36	4225
7	70	490	49	4900
8	71	568	64	5041
$\sum x$	$\sum y$	$\sum xy$	$\sum x^2$	$\sum y^2$
36	449	2249	204	26541

$\bar{x} = 2+3+4+5 / 4 = \frac{14}{4} = 3.5$

$\bar{y} = 50+71+87+95/4 = \frac{303}{4} = 75.75$

$\hat{\beta}_1 = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} = \frac{(2-3.5)(50-75.75)+(3-3.5)(71-75.75)+(4-3.5)(87-75.75)+(5-3.5)(95-75.75)}{(2-3.5)^2+(3-3.5)^2+(4-3.5)^2+(5-3.5)^2} = 15.1$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 75.75 - (15.1)(3.5) = 22.9$

$y = \hat{\beta}_0 + \hat{\beta}_1 x$ $\quad \therefore y = 22.9 + 15.1x$

2. Using your regression model, predict the exam grade for someone who studied for 30 minutes.

   30 min = 0.5 hr

   $y = 22.9 + 15.1(0.5) = \boxed{30.45}$

3. Calculate the R^2 of your model **by hand** and show your work.
   You may use a table in excel if you want.

1) $\bar{y} = 75.75$

2) $\text{var(mean)} = \frac{\sum_{i=1}^{n}(y-\bar{y})^2}{n} = \frac{(50-75.75)^2+(71-75.75)^2+(87-75.75)^2+(95-75.75)^2}{4} = 295.6875$

3) $\text{var(fit)} = \frac{\sum_{i=1}^{n}(y-\hat{y})^2}{n} = \frac{(50-53.1)^2+(71-68.2)^2+(87-83.3)^2+(95-98.4)^2}{4} = 10.675$

$\hat{y} = 22.9 + 15.1x$

x	2	3	4	5
$\hat{y}$	53.1	68.2	83.3	98.4

4) $R^2 = \frac{\text{var(mean)} - \text{var(fit)}}{\text{var(mean)}} = \frac{295.6875 - 10.675}{295.6875} = \boxed{0.964}$