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

$\hat{y}_i = \beta_0 + \beta_1 x_i$

$\beta_0 = Avg(y) - (\beta_1)(Avg(x))$

$\beta_1 = \frac{\sum (x_i - Avg(x))(y_i - Avg(y))}{\sum (x_i - x)^2}$

$R^2 = \frac{Var(mean) - Var(fit)}{Var(mean)}$

Avg = 3.5 Avg = 75.75

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

$\beta_0 = 75.75 - (\beta_1)(3.5)$; $\beta_1 = \frac{38.63 + 2.38 + 5.63 + 28.88}{2.25 + 0.25 + 0.25 + 2.25}$

$\beta_0 = 58.25$; $\beta_1 = 5$ → $\hat{y} = 58.25 + 5x$

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

$x = 0.5$

$\hat{y} = 58.25 + (5)(0.5) = 60.75$

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

$Var(fit) = \frac{333.06 + 5.06 + 76.56 + 138.06}{4} = 138.19$

$Var(mean) = \frac{663.06 + 22.56 + 126.56 + 370.56}{4} = 295.69$

$R^2 = \frac{295.69 - 138.19}{295.69}$

$R^2 = 0.533$