**Linear Regression Practice**

The following figure shows three data points and the best fit line (linear regression line)

y = 3x + 2.

The x-coordinate, or "x", is our independent variable and the y-coordinate, or "y", is our dependent variable.

****

Answer the following questions using this figure.

1. What is the baseline prediction?
2. What is the Sum of Squared Errors (SSE)?
3. What is the Total Sum of Squares (SST)?
4. What is the R² of the model?

**ANSWERS**

1. The baseline prediction is the average value of the dependent variable. Our dependent variable takes values 2, 2, and 8 in our data set, the average is (2+2+8)/3 = 4.
2. The SSE is computed by summing the squared errors between the actual values and our predictions. For each value of the independent variable (x), our best fit line makes the following predictions:

If x = 0, y = 3(0) + 2 = 2,

If x = 1, y = 3(1) + 2 = 5.

Thus we make an error of 0 for the data point (0, 2), an error of 3 for the data point (1, 2), and an error of 3 for the data point (1, 8). So we have

SSE = 0² + 3² + 3² = 18.

1. The SST is computed by summing the squared errors between the actual values and the baseline prediction. From the first question, we computed the baseline prediction to be 4. Thus the SST is:

SST = (2 - 4)² + (2 - 4)² + (8 - 4)² = 24.

1. The R² formula is: R² = 1 - SSE/SST

Thus using our answers to the previous questions, we have that

R² = 1 - 18/24 = 0.25.