Formula analysis using linear regression

Brief Introduction:

We ran linear regression analysis for the 3 cases: Length, spelling error as well as pos tag pairs/grammar errors/tense. The code to run the linear regression on all the three cases is provided at the bottom.

We found the following equations while considering various parameters as follows (where score is the dependent variable)

1. In case of score(y) vs length(x):

$$y = 3.233412656970921 + 0.014286814870614017 * x$$

2. In case of score(y) vs spelling error(x):

$$y = 1.9987561855726985 - 0.10823077250492244 * x$$

3. In case of score(y) vs subject-verb-agreement(x):

$$y = 4.915 - 0.592*x$$

4. In case of score(y) vs tense-mistake-verb-mistake(x):

$$y = 5.051 - 1.558 \cdot x$$

Inference:

The equation found above clearly signifies how the score of essay is related to the length and the spelling error.

In case of the equation that has been given the formula is:
Final Score = $2 * a - b + c.i + c.ii + 2 * c.iii + 2 * d.i [+3 * d.ii]$

In the above equation it can be seen that in order to get the score from the length and the spelling error we have to add the length and deduct the spelling error.

Same is the case in equation 1. and 2. that have been found through regression analysis, we add length and deduct spelling error. Now the question is the value of coefficient.

Based on the regression analysis we the following formulas that worked well:
1.) 2*a - 2*b + 2*c1 + c2
2.) 2*a - 2*b + 2*c1 + 0.5*c2