So far we have been looking at ways to describe a list of values. There are times when we have two lists of values and when we believe that there is an underlying linear relationship between those values. For example, I know that as I exercise my heart rate increases. In fact, I have taken the following measurements of time and my heart rate at the start of my exercise routine:

Time in Minutes	0	1	3	5	6	8
Heart rate	55	63	77	81	102	120

It would seem that my heart rate is a function of time. That is, if you tell me a time then I should be able to predict my heart rate, or at least get quite close to it. As a linear function we are looking to find an equation of the form y=ax+b where x is the time and y is the predicted heart rate. Finding suitable values of a and b, values that make the best model, is termed finding a linear regression model. The computations for this model, for finding the values of a and b from the data in the two lists, are messy enough that we do not want to do them by hand. If we were to perform those computations on a TI-84 calculator we would find that for y=ax+b, and for the data in the table above, the line of best fit, the linear regression line, will have a=7.73 and b=53.37, rounding the computed values to 2 decimal places. This means that our equation becomes:

Heart rate = 7.73*time + 53.37

Now, if we want to find the predicted heart rate at 4 minutes then we simply put 4 into the equation for the "time" variable. This gives heart rate=7.73*4+53.37=30.92+53.37=84.29. We could use our model to find the predicted heart rate at 5 minutes by evaluating 7.73*5+53.37=38.65+53.37=92.02. Note that 92.02 is the "expected" value but, from the table, 81 was the observed value. We could even find the expected value at 30 seconds by evaluating 7.73*0.5+53.37=3.865+53.37=57.235.

We have observed values ranging from 0 minutes to 8 minutes. The regression equation that we have, heart rate=7.73*time+53.37, does not make any distinction about the values of time that we put into the function, that is, the domain of the function. Regression equations are quite good at finding expected values within the range of the observed values. In our case, it makes good sense to use the regression equation for values between 0 and 8 minutes. Doing so is called "interpolation". You can be fairly confident doing interpolation. Using the regression equation for values outside of the range of observed values can be done and it is called "extrapolation". Extrapolation can be quite dangerous, producing values that may be absurd. For example, our equation predicts that my heart rate at time 0 will be 53.37, not far off from the observed 55, and a reasonable value. But if we were to use the equation to predict my heart rate 6 minutes before exercise then the equation would be heart rate =7.73*(-6)+53.37=-

46.38+53.37=6.99, an absurd value indeed. The same is true if we were to use the equation to predict my heart rate at the 25 minute mark, giving heart

rate=7.73*25+53.37=193.25+53.37=246.62.

Linear regression, taking data from two lists and producing the values of a and b in the equation y=ax+b, gives us a way to predict values for the second variable, y, from the first variable

x. Once we have found the values of a and b, we can substitute any value we want for the variable x and we can compute the expected value of y. The table gives us the observed values, the equation gives us the expected values. Evaluating a linear regression equation for x-values within the range of the observed x values is fairly safe. Evaluating a linear regression equation for x values outside of the range of observed x values is dangerous and should be looked at with some skepticism.