Regression II

(aka "Line of best fit", aka "Pearson correlation coefficient", aka "r")

Here is the same data as in the previous sheet except there is now an additional column; position.
This is now entered into L_3.

STAT 1

Goals scored	Points	Position
61	70	1
45	66	2
54	59	3
47	52	4
38	50	5
30	45	6
28	39	7
32	38	8
35	37	9
37	34	10
26	34	11
20	30	12
25	29	13
21	26	14
28	24	15
25	23	16

Now get it to perform the calculation

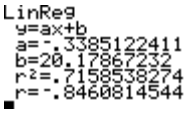
This time we are using goals as the x variable and position as the y variable.

STAT ▶ 4

Then check it has the correct lists for the x and y values.

Interpret the results

(If you do not see r^2 or r then you did not turn on the diagnostics!)



The r value tells us that the variables are strongly correlated, and that it is negative correlation. (See your textbook for more details on interpreting r.)
The a and b tell us that the calculated line of best fit is $y = -0.339x + 20.2$
ie gradient of -0.339 with a y-intercept of 20.2

In other words you expect to move up roughly a third of a position per goal – and if you failed to score any goals all season you would come in 20th place. Obviously this is impossible with only 16 teams and highlights the dangers of **extrapolation** (using your line of best it outside the values you actually have).

Another variation....

This time using points as the x variable and position as the y variable

STAT ▶ 4

Again it is negative correlation (ie the more points you score the lower your position) and the correlation is very strong.
The line of best fit is

$$y = -0.314x + 21.4$$

which, with an intercept of 21, shows that just because the correlation is very strong it doesn't mean you can extrapolate!