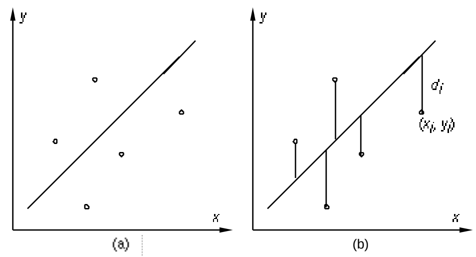
MATH 261, Spring 2001 **Due Date:**____________ **Name(s):** ________________________

Honors Project 20: Least Squares

In many science and engineering applications, one is often given a set of data points $\{(x_i, y_i), i = 1, ..., N\}$ in R^2, and interested in finding the line which "best fits this data" (see Fig. (a) below). One solution to this problem is provided by *classical least squares*: finding the line for which the sum of the squares of the distances d_i from the data points to the line in the direction of the dependent axis is a minimum (see Fig. (b) below).



As an application of optimization of functions of two variables — specifically, finding the m and b that minimize the error term

$$\sum)i = 1^N(y_i - (mx_i + b))^2,$$

— it follows that the line $y = mx + b$ which best approximates the data in the sense of classical least squares is given by

$$m = \frac{N\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{N\sum x_i^2 - \left(\sum x_i\right)^2} \quad \text{and} \quad b = \frac{\left(\sum y_i\right)\left(\sum x_i^2\right) - \left(\sum x_i\right)\left(\sum x_i y_i\right)}{N\sum x_i^2 - \left(\sum x_i\right)^2}.$$

Tasks

1. Verify the above formulas.
2. Devise a procedure for finding the plane Π whose equation is of the form $z = ax + by + c$ which "best fits" the data $\{(x_i, y_i, z_i), i = 1, ..., N\}$ in R^3. In the end you should arrive at a system of linear equations that need to be solved. You do not have to solve this system here, however. Indeed, to do Task 3 below you can use Maples `solve` command.
3. Apply the procedure you developed above to the following data set:

x_i	0.9	1.1	1.2	1.4	2.3	2.9	3.5	3.6	4.1	4.8
y_i	1.2	3.6	3.5	2.4	2.2	0.6	1.6	4.3	1.5	2.7
z_i	10.8	19.2	18.9	15.4	15.8	11.1	15.2	24.5	15.6	20.5

4. Without going through any analysis, what system of linear equations would you *guess* you would have to solve to find the hyperplane whose equations is $w = ax + by + cz + d$ which "best fits" the data $\{(x_i, y_i, z_i, w_i), i = 1, ..., N\}$ in R^4.