# Better Than The Line of Best Fit

*Ali Astaneh, Vancouver, BC*

**(A) The line of best fit** :

Given a set of points $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ on the plane, consider the following function in variables $m$ and $b$, as the sum of squares of 'vertical distances' of the points from *the line of best fit* to be found

$$S(m,b) = \sum_{j=1}^{n}[y_j - (mx_j + b)]^2 ,$$

To obtain the pair $(m,b)$ that gives the line of best fit, just solve the system of two equations with the two unknowns $m$ and $b$, which is obtained by setting the partial derivatives of the function $S(m,b)$ equal to zero. A straightforward calculation will show that $m$ and $b$ are obtained from the following equations

$$m = [(\sum_{j=1}^{n} x_j)(\sum_{j=1}^{n} y_j) - n\sum_{j=1}^{n} x_j y_j] / [(\sum_{j=1}^{n} x_j)^2 - n\sum_{j=1}^{n} x_j^{\ 2}]$$

$$b = \frac{1}{n}\sum_{j=1}^{n}(y_j - mx_j).$$

**(B) Better Than The line of best fit** :

Given a set of points $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ on the plane, this time around in order to handle the expressions and equations involved in a more convent manner, first let us define for each $j = 1,\ 2,\ \ldots,\ n$;

$$X_j = x_j - \frac{1}{n}\sum_{i=1}^{n} x_i , \qquad Y_j = y_j - \frac{1}{n}\sum_{i=1}^{n} y_i .$$

(Here one can think of $X_j$ as the 'signed' deviation from $x_j$ to the mean of the first components $x_1,\ x_2,\ \ldots,\ x_n$, and similarly $Y_j$ as the 'signed' deviation from $y_j$ to the mean of those $y_1,\ y_2,\ \ldots,\ y_n$). Then, considering the function of $m$ and $b$ defined by

$$s(m,b) = \frac{1}{1+m^2}\sum_{j=1}^{n}[y_j - (mx_j + b)]^2$$

as the sum of squares of perpendicular distances of the points $(x_1,y_1),(x_2,y_2),\ \ldots,(x_n,y_n)$ from *the better than the line of best fit* to be found, again one can solve the system of two equations with the two unknowns $m$ and $b$, which is obtained by setting the partial derivatives of the function $s(m,b)$ equal to zero.

My calculations (hopefully without typos) show that this time around the following system will present the solutions fort $m$ and $b$,

$$(\sum_{j=1}^{n} X_j^{\ 2} - \sum_{j=1}^{n} x_j X_j)\, m^3 - (\sum_{j=1}^{n} x_j Y_j - 2\sum_{j=1}^{n} X_j Y_j)\, m^2 + (\sum_{j=1}^{n} Y_j^2 - \sum_{j=1}^{n} x_j X_j)\, m + \sum_{j=1}^{n} x_j Y_j = 0$$

$$b = \frac{1}{n}\sum_{j=1}^{n}(y_j - mx_j).$$