LET $\vec{X} = (x_1, x_2, \dots, x_N)$ AND $\vec{Y} = (y_1, y_2, \dots, y_N)$.

THEN $M_x = \dfrac{x_1 + x_2 + \cdots + x_N}{N}$ AND $M_y = \dfrac{y_1 + y_2 + \cdots + y_N}{N}$

ARE THE MEANS OF $\vec{X}$ AND $\vec{Y}$, RESPECTIVELY.

$$SDx = \frac{1}{\sqrt{N}} \sqrt{(x_1 - M_x)^2 + (x_2 - M_x)^2 + \cdots + (x_N - M_x)^2}$$

$$= \frac{1}{\sqrt{N}} \| \vec{X} - M_x (1, 1, \dots, 1) \| \qquad \text{AND}$$

$$SDy = \frac{1}{\sqrt{N}} \sqrt{(y_1 - M_y)^2 + (y_2 - M_y)^2 + \cdots + (y_N - M_y)^2}$$

$$= \frac{1}{\sqrt{N}} \| \vec{Y} - M_y (1, 1, \dots, 1) \|$$

ARE THE STANDARD DEVIATIONS OF $\vec{X}$ AND $\vec{Y}$, RESPECTIVELY.

<u>RULE OF THUMB</u> - FOR MOST VECTORS $\vec{X}$ AT LEAST 68% OF THE ENTRIES WILL LIE BETWEEN $M_x - SDx$ AND $M_x + SDx$, WHILE AT LEAST 95% WILL LIE BETWEEN $M_x - 2SDx$ AND $M_x + 2SDx$.

THE LEAST SQUARES LINE (OFTEN CALLED THE BEST FIT LINE OR THE REGRESSION LINE) FOR $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ IS THE LINE WHICH MINIMIZES THE SUM OF THE SQUARES OF THE VERTICAL DISTANCES FROM THE POINTS TO THE LINE. IF $y = Bx + A$ IS THE LEAST SQUARES LINE, WE FIND A AND B