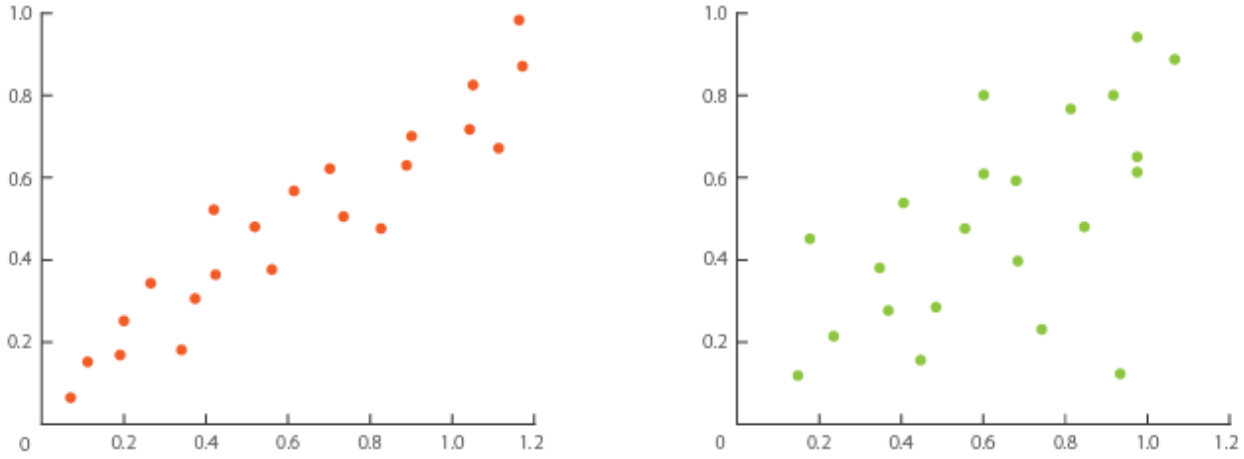
Regression (HW #5)

This material is in Section 4.3 of the 5th edition of your textbook.

How do private colleges such as Harvard and Stanford decide who to enroll?

My conjecture is they try to maximize dollars donated by alumni. How can they predict future donated dollars? They look at the admission records of all the old alumni and create a test score meant to predict amount of dollars donated. Then they make a scatterplot of scores vs. actual dollars donated.



The school refines the test until they get a scatterplot more like the one on the left than the one on the right. Then they find the best fitting line called the regression line that minimizes the sum of the square errors. The Least Squares Regression line (LSR line) allows them to predict the number of dollars an applicant will bring the college after graduation.

Solving a projection problem will find the LSR line. Our regression line is labeled $b = C + Dt$ and then we find the best solutions for $\hat{C}$ and $\hat{D}$ for the points (t_i, b_i) on the scatterplot.

Each point creates the equation $b_i = C + Dt_i$, so the system of equations becomes

$$\begin{bmatrix} \vec{1} & \vec{t} \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \vec{b}.$$

We then find the projection by multiplying by $\begin{bmatrix} \vec{1} & \vec{t} \end{bmatrix}^T$ and solving for $\begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix}$.

Find the LSR line if $b = 1, 3, 2$ when $t = 1, 2, 3$ respectively.

The Fourier coefficients in $\hat{x}$ give us the LSR line. Where are the components of the projection $A\hat{x}$ and the error vector $\vec{e} = \vec{b} - A\hat{x}$ on the following graph? Why does the FTLA guarantee the errors add up to 0?