1. Consider the points $P_1 = (1, -16)$, $P_2 = (3, 10)$, $P_3 = (7, 8)$, and $P_4 = (9, 10)$, measurements made in a lab where the data was supposed lie on a single line. Obviously something went wrong; these four points can't possibly be on the same line.

1a. Explicitly, what are the *design matrix* $\mathbf{X}$ and *observation vector* $\mathbf{y}$ you would use in a matrix equation $\mathbf{X}\boldsymbol{\beta} = \mathbf{y}$ to find $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ such that the four points above are on the line $y = \beta_0 + \beta_1 x$?

$\mathbf{X} =$ $\mathbf{y} =$

1b. Since $\mathbf{X}\boldsymbol{\beta} = \mathbf{y}$ has no solution, we'll have to be content with the least-squares solution. Find it; ie, find β_0 and β_1 for which $y = \beta_0 + \beta_1 x$ is the least-squares line that best fits the four given points. *Show all your work!*

$\beta_0 =$ $\beta_1 =$

1c. Find the four y coordinates corresponding to $x = 1, 3, 7, 9$ on the best-fit line, and assemble them into a vector we'll call $\mathbf{p}$; $\mathbf{p}$ is our vector of *predicted values.*

$\mathbf{p} =$

1d. Find the four residuals, and then the sum of their squares (SOS).

Four Residuals: SOS:

1e. You might think that the line $y = -77/4 + 13/4\,x$ would be a better fit: At least it goes through P_1 and P_4. Find the four residuals for *this* line, and then the sum of *their* squares.

Four Residuals: SOS:

1f. Which line is a better fit? Explain.