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 $y = \beta_0 + \beta_1 x$?

$$\mathbf{X} = \begin{bmatrix}1 & 1\\ 1 & 3\\ 1 & 7\\ 1 & 9\end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix}-16\\ 10\\ 8\\ 10\end{bmatrix}$$

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-quares line that best fits the four given points. *Show all your work!*

we need to solve $X^T X \vec{\beta} = X^T \vec{y}$

$$\begin{bmatrix}4 & 20\\ 20 & 140\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix} = \begin{bmatrix}12\\ 160\end{bmatrix}$$

now, $\left[\begin{array}{cc|c}4 & 20 & 12\\ 20 & 140 & 160\end{array}\right] \sim \left[\begin{array}{cc|c}1 & 0 & -9.5\\ 0 & 1 & 2.5\end{array}\right]$

$\Rightarrow \beta_0 = -9.5$ & $\beta_1 = 2.5$

$\beta_0 = -9.5 = \frac{-19}{2}$ $\qquad$ $\beta_1 = 2.5 = \frac{5}{2}$

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*.

$$\begin{bmatrix}1 & 1\\ 1 & 3\\ 1 & 7\\ 1 & 9\end{bmatrix}\begin{bmatrix}-9.5\\ 2.5\end{bmatrix} = \begin{bmatrix}-7\\ -2\\ 8\\ 13\end{bmatrix} \qquad \mathbf{p} = \begin{bmatrix}-7\\ -2\\ 8\\ 13\end{bmatrix}$$

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

Four Residuals: $\vec{y} - \vec{p} = \begin{bmatrix}-9\\ 12\\ 0\\ -3\end{bmatrix}$ SOS: $81+144+0+9 = 234$

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: $\vec{y} - \begin{bmatrix}-16\\ -9.5\\ 3.5\\ 10\end{bmatrix} = \begin{bmatrix}-16\\ 10\\ 8\\ 10\end{bmatrix} - \begin{bmatrix}-16\\ -9.5\\ 3.5\\ 10\end{bmatrix} = \begin{bmatrix}0\\ 19.5\\ 4.5\\ 0\end{bmatrix}$ SOS: $380.25+20.25 = 400.5$

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

The 1ST line, since the SOS is smaller, this means the distance from $\begin{bmatrix}-7\\ -2\\ 8\\ 13\end{bmatrix}$ to $\begin{bmatrix}-16\\ 10\\ 8\\ 10\end{bmatrix}$ is smaller than the distance from $\begin{bmatrix}-16\\ -9.5\\ 3.5\\ 10\end{bmatrix}$ to $\begin{bmatrix}-16\\ 10\\ 8\\ 10\end{bmatrix}$