Inna Williams
##

1. Form the normal equations, and compute the least squares solution and 2-norm error for the following inconsistent systems:

$$\begin{bmatrix} 3 & -1 & 2 \\ 4 & 1 & 0 \\ -3 & 2 & 1 \\ 1 & 1 & 5 \\ -2 & 0 & 3 \end{bmatrix} \mathbf{x} \begin{bmatrix} x1 \\ x2 \\ x3 \end{bmatrix} = \begin{bmatrix} 10 \\ 10 \\ -5 \\ 15 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 3 & -1 & 2 \\ 4 & 1 & 0 \\ -3 & 2 & 1 \\ 1 & 1 & 5 \\ -2 & 0 & 3 \end{bmatrix}$$

$$\begin{bmatrix} x1 \\ x2 \\ x3 \end{bmatrix} = \begin{bmatrix} 10 \\ 10 \\ -5 \\ 15 \\ 0 \end{bmatrix} \qquad (1)$$

> $with(LinearAlgebra):$

> $A := Matrix([[3,-1,2],[4,1,0],[-3,2,1],[1,1,5],[-2,0,3]]);$

$$A := \begin{bmatrix} 3 & -1 & 2 \\ 4 & 1 & 0 \\ -3 & 2 & 1 \\ 1 & 1 & 5 \\ -2 & 0 & 3 \end{bmatrix} \qquad (2)$$

> $b := \langle 10, 10, -5, 15, 0 \rangle;$

(3)

$$b := \begin{bmatrix} 10 \\ 10 \\ -5 \\ 15 \\ 0 \end{bmatrix} \tag{3}$$

> *Transpose(A).A; Transpose(A).b*

$$\begin{bmatrix} 39 & -4 & 2 \\ -4 & 7 & 5 \\ 2 & 5 & 39 \end{bmatrix}$$

$$\begin{bmatrix} 100 \\ 5 \\ 90 \end{bmatrix} \tag{4}$$

> *sol := LinearSolve(Transpose(A).A, Transpose(A).b)*

$$sol := \begin{bmatrix} \frac{2257}{894} \\ \frac{1183}{1788} \\ \frac{3743}{1788} \end{bmatrix} \tag{5}$$

> *sol_ev := evalf(sol);*

$$sol\_ev := \begin{bmatrix} 2.524608501 \\ 0.6616331096 \\ 2.093400447 \end{bmatrix} \tag{6}$$

check the solution

> *ls := LeastSquares(A, b)*

$$ls := \begin{bmatrix} \frac{2257}{894} \\ \frac{1183}{1788} \\ \frac{3743}{1788} \end{bmatrix} \tag{7}$$

> *evalf(ls)*

$$\begin{bmatrix} 2.524608501 \\ 0.6616331096 \\ 2.093400447 \end{bmatrix} \tag{8}$$

> *norm2error := evalf(Norm(b−A.sol, 2))*

$$norm2error := 2.413492091 \quad (9)$$

> $RMSE := evalf\left(\frac{norm2error}{\text{sqrt}(3)}\right)$

$$RMSE := 1.393430309 \quad (10)$$

Answer:

$$\mathbf{x\_HAT} = \begin{bmatrix} 2.524608501 \\ 0.6616331096 \\ 2.093400447 \end{bmatrix}$$

||e2|| = 2.413492091
RMSE **= 1.393430309**

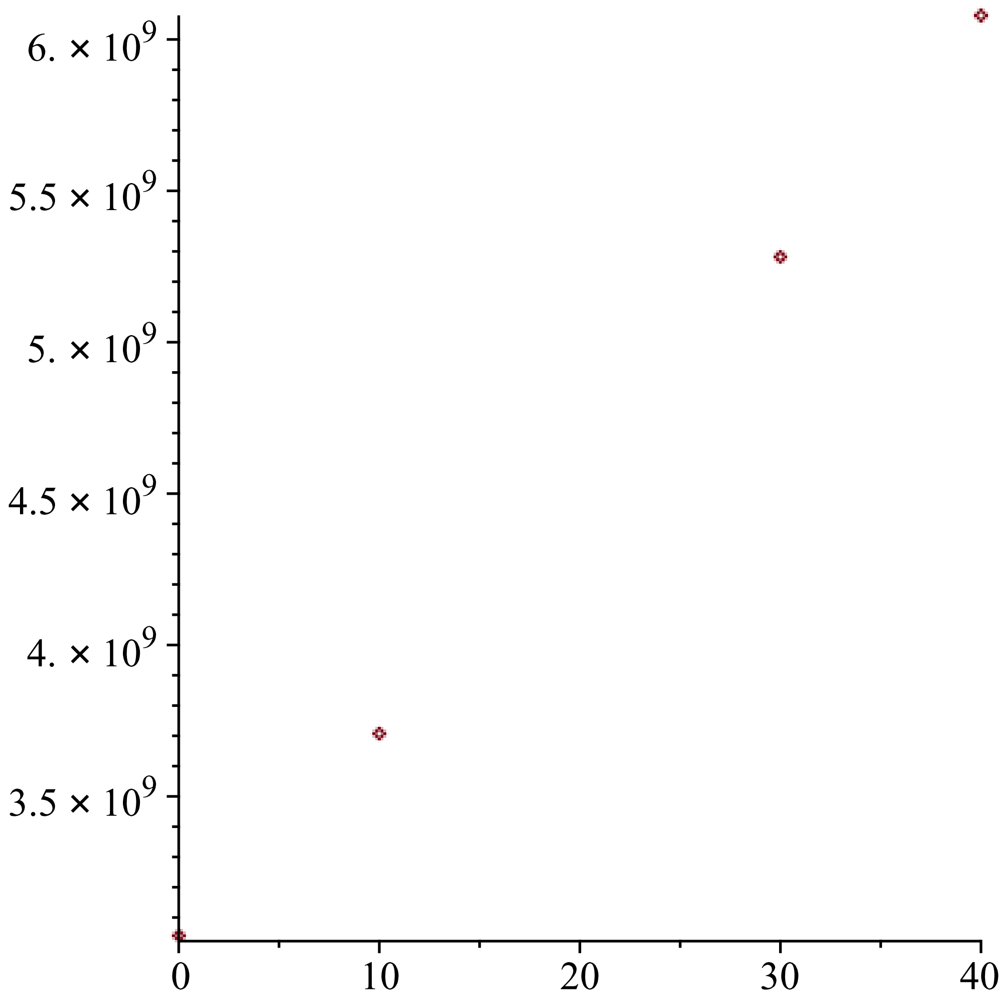
###

3. ***Consider the world population data of Computer Problem*** **3.1.1** ***. Find the best least squares***
(*a*) ***line through the data points*****, and** ***the RMSE of the fit. In each case, estimate the*** **1980** ***population. Which fit gives the best estimate?***

model : y=a+b(x-1960)

> $with(plots) : with(CurveFitting) : with(LinearAlgebra) :$

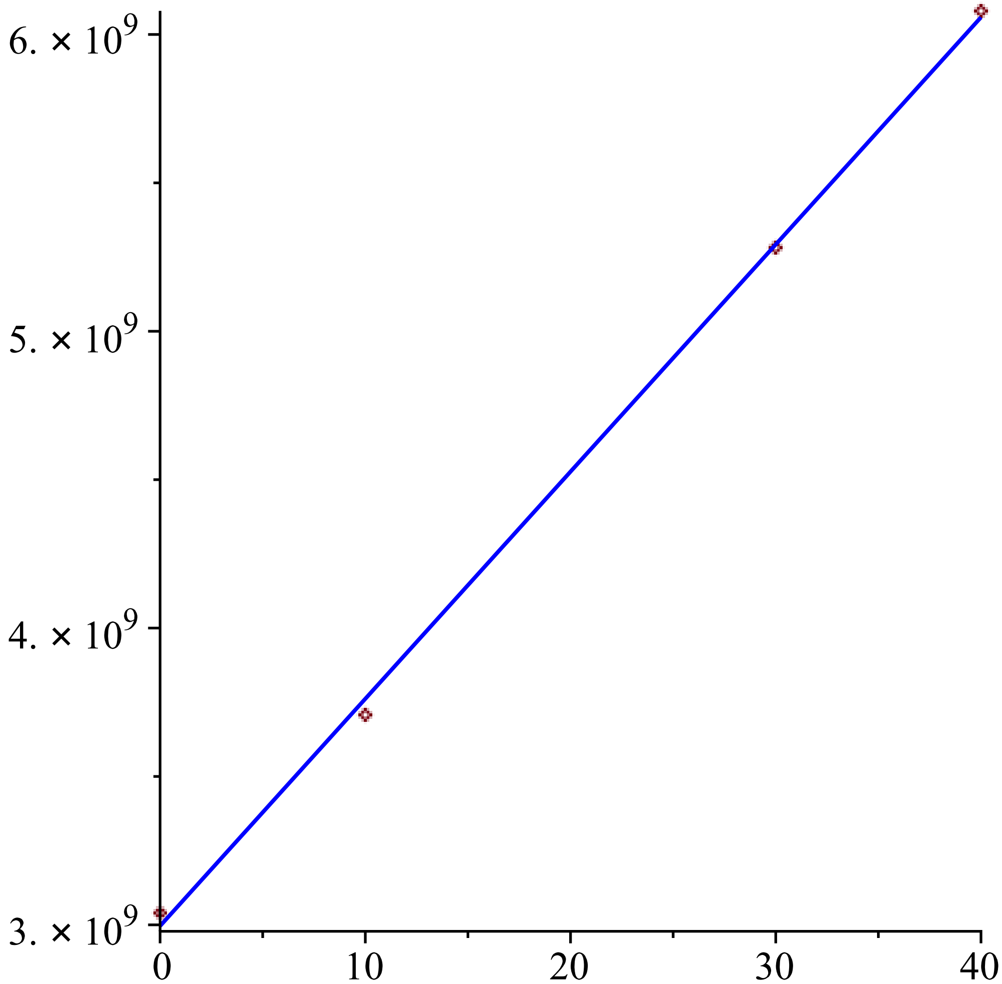
> $xvalues := \langle 0, 10, 30, 40\rangle : yvalues := \langle 3039585530, 3707475887, 5281653820, 6079603571\rangle :$
$p1 := plot(xvalues, yvalues, style = point) : display(p1)$



> $line := evalf(CurveFitting[LeastSquares](xvalues, yvalues, x))$

$$line := 2.996236899\ 10^9 + 7.654214015\ 10^7\ x \quad \textbf{(11)}$$

> $p2 := plot(line, x = 0..40, color = blue) : display(p1, p2)$



We compute the y-outputs by mapping the x-values via the regression curve

> $yout := evalf\left(map\left(x \to 2.996236899\ 10^9 + 7.654214015\ 10^7\ x, xvalues\right)\right)$

$$yout := \begin{bmatrix} 2.996236899\ 10^9 \\ 3.761658300\ 10^9 \\ 5.292501103\ 10^9 \\ 6.057922505\ 10^9 \end{bmatrix} \quad \textbf{(12)}$$

Now we compute the RMSE value using the appropriate formula.

> $RMSE := evalf\left(\frac{Norm(yvalues - yout, 2)}{\sqrt{Dimension(xvalues)}}\right)$

$$RMSE := 3.675108794\ 10^7 \quad \textbf{(13)}$$

> $evalf(subs(x = 1980 - 1960, line))$

$$4.527079702\ 10^9 \quad \textbf{(14)}$$

Answer : for line

$$\boldsymbol{line := 2.996236899\ 10^9 + 7.654214015\ 10^7\ x}$$

$$\boldsymbol{RMSE := 3.675108794\ 10^7}$$

1980 population estimate = $\mathbf{4.527079702\ 10^9}$

****

(b) *parabola through the data points*, and *the RMSE of the fit. In each case, estimate the* 1980 *population. Which fit gives the best estimate?*

model : y = a + b(x-1960) + c(x-1960)^2

$$line := 2.996236899\ 10^9 + 7.654214015\ 10^7\, x$$

$$RMSE := 3.675108794\ 10^7$$

$$4.527079702\ 10^9 \tag{15}$$

```
> unassign('a','b','c','x') :
```

```
> parabola := CurveFitting[LeastSquares](xvalues, yvalues, x, curve = a + b·x + c·x^2)
```

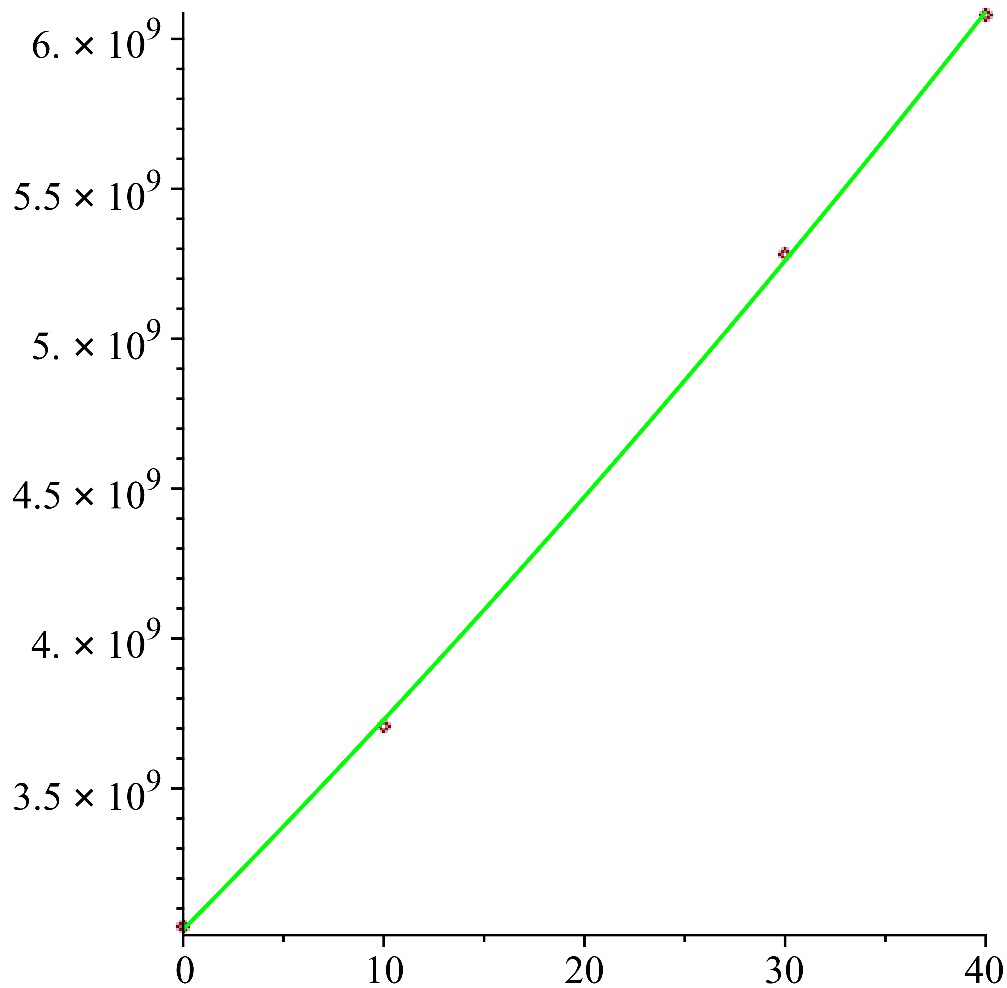
$$parabola := \frac{6057503495}{2} + \frac{4072290833}{60}\, x + \frac{65029697}{300}\, x^2 \tag{16}$$

```
> parabola := evalf(parabola)
```

$$parabola := 3.028751748\ 10^9 + 6.787151388\ 10^7\, x + 216765.6567\, x^2 \tag{17}$$

```
>
```

```
> p3 := plot(parabola, x = 0 ..40, color = green) : display(p1, p3)
```



We compute the y-outputs by mapping the x-values via the regression curve

> $yout := map\left(x \rightarrow 3.028751748\ 10^{9} + 6.787151388\ 10^{7}\ x + 216765.6567\ x^{2}, xvalues\right)$

$$yout := \begin{bmatrix} 3.028751748\ 10^{9} \\ 3.729143453\ 10^{9} \\ 5.259986255\ 10^{9} \\ 6.090437354\ 10^{9} \end{bmatrix} \tag{18}$$

> $yout := evalf(yout)$

$$yout := \begin{bmatrix} 3.028751748\ 10^{9} \\ 3.729143453\ 10^{9} \\ 5.259986255\ 10^{9} \\ 6.090437354\ 10^{9} \end{bmatrix} \tag{19}$$

>

Now we compute the new RMSE value using the appropriate formula.

> $RMSE := evalf\left(\dfrac{Norm(yvalues - yout, 2)}{\sqrt{Dimension(xvalues)}}\right)$

$$RMSE := 1.712971450\ 10^{7} \tag{20}$$

Notice that the new RMSE value is smaller than the one obtained for the line fit. This shows that the quadratic model is a better fit for the data.

> $evalf(subs(x = 1980 - 1960, parabola))$

$$4.472888289\ 10^{9} \tag{21}$$

Answer : for parabola

$\boldsymbol{parabola := 3.028751748\ 10^{9} + 6.787151388\ 10^{7}\ x + 216765.6567\ x^{2}}$

$\boldsymbol{RMSE := 3.675108794\ 10^{7} RMSE := 1.712971450\ 10^{7}}$

1980 population estimate = $\boldsymbol{4.472888289\ 10^{9}}$

Answer Best Estimate:

Line RMSE = $\boldsymbol{3.675108794\ 10^{7}}$

Parabola RMSE = $\boldsymbol{1.712971450\ 10^{7}}$ < Line **RMSE = $\boldsymbol{3.675108794\ 10^{7}}$**
Therefore the best estimate gives parabola.

>