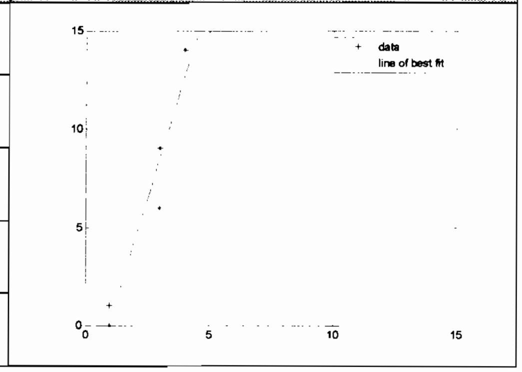
Problem 1:

Data was collected and listed in the table on the right. A line of best fit *y=mx+b* was determined using the using the normal equations for $A\begin{bmatrix} m \\ b \end{bmatrix} = \vec{y}$ (see Theorem 13 page 411). Determine the values for *m* and *b* using the normal equations. Calculate the correlation between the vector $\vec{x}$ and $\vec{y}$.

x	y	plot
1	0	
1	1	
3	6	
3	9	
4	14	

(a)

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 3 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 6 \\ 9 \\ 14 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 1 & 1 & 3 & 3 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 3 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 1 & 1 & 3 & 3 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 6 \\ 9 \\ 14 \end{bmatrix}$$

$$\begin{bmatrix} 36 & 12 \\ 12 & 5 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 102 \\ 30 \end{bmatrix}$$

25/6

$$\text{RREF} \begin{bmatrix} 36 & 12 & 102 \\ 12 & 5 & 30 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 4.1667 \\ 0 & 1 & -4.000 \end{bmatrix}$$

$$\therefore \boxed{y = \frac{25}{6}x - 4}$$

m b

(b)

Correlation: $\dfrac{\vec{x}\cdot\vec{y}}{\|\vec{x}\|\|\vec{y}\|} = \dfrac{1+18+27+56}{\sqrt{36}\ \sqrt{314}} =$

$$= \frac{102}{6\sqrt{314}} = \boxed{.9594}$$

a linear fit for this data has a strong correlation

Problem 2.

a. Using the data from problem one, construct the matrix A used in the normal form. Calculate the QR factorization for A. Use the QR factorization to determine the values of m and b in a line of best fit (See numerical note, page 415). Hint: Your answer should be the same as problem 1.

b. Starting with the normal equation $A\vec{x} = \vec{b}$ derive the equation $R\vec{x} = Q^T\vec{b}$.

a

$$\vec{V}_1 = \begin{bmatrix} 1 \\ 1 \\ 3 \\ 3 \\ 4 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 3 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}$$

$$\vec{V}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{[1\ 1\ 3\ 3\ 4] \cdot [1\ 1\ 1\ 1\ 1]}{\|[1\ 1\ 3\ 3\ 4]\|^2} \begin{bmatrix} 1 \\ 1 \\ 3 \\ 3 \\ 4 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \overset{1/3}{\frac{\cancel{12}}{\cancel{36}}} \begin{bmatrix} 1 \\ 1 \\ 3 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 2/3 \\ 0 \\ 0 \\ -1/3 \end{bmatrix} \Rightarrow \|V_2\| = 1$$

$$\vec{u}_1 = \frac{\vec{V}_1}{\|\vec{V}_1\|} = \begin{bmatrix} 1/6 \\ 1/6 \\ 1/2 \\ 1/2 \\ 2/3 \end{bmatrix} \qquad \vec{u}_2 = \frac{\vec{V}_2}{\|\vec{V}_2\|} = \begin{bmatrix} 2/3 \\ 2/3 \\ 0 \\ 0 \\ -1/3 \end{bmatrix}$$

$$Q = \begin{bmatrix} 1/6 & 2/3 \\ 1/6 & 2/3 \\ 1/2 & 0 \\ 1/2 & 0 \\ 2/3 & -1/3 \end{bmatrix} \qquad R = Q^T A$$

$$\Rightarrow R = \begin{bmatrix} 1/6 & 1/6 & 1/2 & 1/2 & 2/3 \\ 2/3 & 2/3 & 0 & 0 & -1/3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 3 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ 0 & 1 \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} 6 & 2 \\ 0 & 1 \end{bmatrix}}_{R}\begin{bmatrix} m \\ b \end{bmatrix} = Q^T\vec{b} = \begin{bmatrix} 17 \\ -4 \end{bmatrix}$$

$$\Rightarrow 6m + 2b = 17 \quad \Rightarrow \quad 6m = 25 \Rightarrow \boxed{m = \frac{25}{6}}$$

$$\boxed{b = -4}$$

$$\boxed{y = \frac{25}{6}x - 4}$$

(b) Let $A = QR$ where Q is a matrix w/orthonormal columns.

① $A\vec{x} = \vec{b} \Rightarrow QR\vec{x} = \vec{b}$

② $A^TA\vec{x} = A^T\vec{b} \Rightarrow (QR)^T QR\vec{x} = (QR)^T\vec{b}$

$$\Rightarrow R^T(Q^TQ)R\vec{x} = R^TQ^T\vec{b}$$

$$\Rightarrow R^T I R\vec{x} = R^TQ^T\vec{b}$$

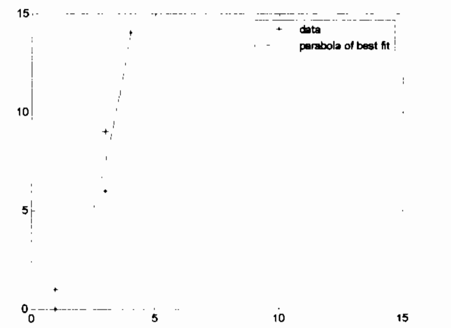
$$\Rightarrow R^TR\vec{x} = R^TQ^T\vec{b}$$

$$\Rightarrow \boxed{R\vec{x} = Q^T\vec{b}}$$

Problem 3.

Data was collected and listed in the table on the right. A parabola of best fit $y=ax^2+bx+c$ was determined using the using the normal equations for $A\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \vec{y}$ (see Theorem 13 page 411).

Determine the values for a, b, and c using the normal equations.

x	y	plot
1	0	
1	1	
3	6	
3	9	
4	14	

$y = ax^2 + bx + c$

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 9 & 3 & 1 \\ 9 & 3 & 1 \\ 16 & 4 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a \\ b \\ c \end{bmatrix}}_{\vec{x}} = \underbrace{\begin{bmatrix} 0 \\ 1 \\ 6 \\ 9 \\ 14 \end{bmatrix}}_{\vec{b}}$$

$$\Rightarrow \begin{bmatrix} 1 & 1 & 9 & 9 & 16 \\ 1 & 1 & 3 & 3 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 9 & 3 & 1 \\ 9 & 3 & 1 \\ 16 & 4 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 & 1 & 9 & 9 & 16 \\ 1 & 1 & 3 & 3 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 6 \\ 9 \\ 14 \end{bmatrix}$$

$$\begin{bmatrix} 420 & 120 & 36 \\ 120 & 36 & 12 \\ 36 & 12 & 5 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 360 \\ 102 \\ 30 \end{bmatrix}$$

$$\Rightarrow \text{RREF} \begin{bmatrix} 420 & 120 & 36 & 360 \\ 120 & 36 & 12 & 102 \\ 36 & 12 & 5 & 30 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -.5 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\boxed{\begin{matrix} a = 1 \\ b = -.5 \\ c = 0 \end{matrix}} \qquad \boxed{y = x^2 - \frac{1}{2}x}$$

Problem 4

Consider the meat consumption (in grams per day) and the incidence of colon cancer (per 100000 women per year) in various industrialized countries.

Country	Meat Consumption	Cancer Rate	Meat Consumption (Deviation from the Mean)	Cancer Rate (Deviation from the Mean)
Japan	26	7.5	-122	-10.7
Finland	101	9.8	-47	-8.4
Israel	124	16.5	-24	-1.7
Great Britain	205	23.3	57	5.1
United States	284	34	136	15.8
Mean	148	18.2		

Using the "Deviation from the Mean" data,

a. Determine the correlation between the "Meat Consumption" and "Cancer Rate" vectors.
b. Using least squares methods, propose a line of best fit for this data.

(a) $$r = \frac{\vec{x}\cdot\vec{y}}{\|\vec{x}\|\|\vec{y}\|} = \frac{4180.5}{(198.5)(21.5)} = \boxed{.9796}$$

$$\Rightarrow \begin{bmatrix} -122 & 1 \\ -47 & 1 \\ -24 & 1 \\ 57 & 1 \\ 136 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} -10.7 \\ -8.4 \\ -1.7 \\ 5.1 \\ 15.8 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} -122 & -47 & -24 & 57 & 136 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -122 & 1 \\ -47 & 1 \\ -24 & 1 \\ 57 & 1 \\ 136 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} =$$

$$\begin{bmatrix} -122 & -47 & -24 & 57 & 136 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -10.7 \\ -8.4 \\ -1.7 \\ 5.1 \\ +15.8 \end{bmatrix}$$

$$\begin{bmatrix} 39414 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 4186.5 \\ .1 \end{bmatrix}$$

$$\Rightarrow \quad m = \frac{4186.5}{39414} = .1061$$

$$b = .02$$

$$\Rightarrow \quad \hat{y} = .1061\vec{x} + .02$$

c) distance between the $\hat{y}$ vector predicted by the linear formula and the actual data

$$\hat{y} = [-12.92 \quad -4.97 \quad -2.53 \quad 6.07 \quad 14.45]$$

$$y = [-10.7 \quad -8.4 \quad -1.7 \quad 5.1 \quad 15.8]$$

$$\text{distance} = \|\hat{y} - \vec{y}\| = 4.49$$

$$\sqrt{(-12.92+10.7)^2 + (-4.97+8.4)^2 \ldots \text{etc}}$$