Q1

Least-Squares Lines

A common task in science and engineering is to analyze and understand relationships among several quantities that vary. This section describes a variety of situations in which data are used to build or verify a formula that predicts the value of one variable as a function of other variables. In each case, the problem will amount to solving a least-squares problem.

For easy application of the discussion to real problems that you may encounter later in your career, we choose notation that is commonly used in the statistical analysis of scientific and engineering data. Instead of $A\mathbf{x} = \mathbf{b}$, we write $X\boldsymbol{\beta} = \mathbf{y}$ and refer to X as the **design matrix**, $\boldsymbol{\beta}$ as the **parameter vector**, and $\mathbf{y}$ as the **observation vector**.

The simplest relation between two variables x and y is the linear equation $\boxed{y = \beta_0 + \beta_1 x.}$[1] Experimental data often produce points $(x_1, y_1), \ldots, (x_n, y_n)$ that, when graphed, seem to lie close to a line. We want to determine the parameters β_0 and β_1 that make the line as "close" to the points as possible.

Suppose β_0 and β_1 are fixed, and consider the line $y = \beta_0 + \beta_1 x$ in Figure 1. Corresponding to each data point (x_j, y_j) there is a point $(x_j, \beta_0 + \beta_1 x_j)$ on the line with the same x-coordinate. We call y_j the *observed* value of y and $\beta_0 + \beta_1 x_j$ the *predicted* y-value (determined by the line). The difference between an observed y-value and a predicted y-value is called a *residual.*

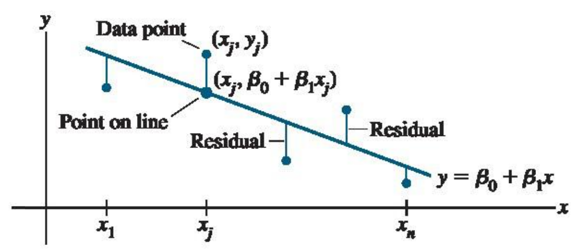

FIGURE 1 Fitting a line to experimental data.

There are several ways to measure how "close" the line is to the data. The usual choice (primarily because the mathematical calculations are simple) is to add the squares of the residuals. The **least-squares line** is the line $y = \beta_0 + \beta_1 x$ that minimizes the sum of the squares of the residuals. This line is also called a **line of regression of y on x**, because any errors in the data are assumed to be only in the y-coordinates. The coefficients β_0, β_1 of the line are called (linear) **regression coefficients.**[2]

If the data points were on the line, the parameters β_0 and β_1 would satisfy the equations

Predicted y-value		**Observed y-value**
$\beta_0 + \beta_1 x_1$	=	y_1
$\beta_0 + \beta_1 x_2$	=	y_2
$\vdots$		$\vdots$
$\beta_0 + \beta_1 x_n$	=	y_n

[1] This notation is commonly used for least-squares lines instead of $y = mx + b$.

[2] If the measurement errors are in x instead of y, simply interchange the coordinates of the data (x_j, y_j) before plotting the points and computing the regression line. If both coordinates are subject to possible error, then you might choose the line that minimizes the sum of the squares of the *orthogonal* (perpendicular) distances from the points to the line.

We can write this system as

$$X\boldsymbol{\beta} = \mathbf{y}, \quad \text{where } X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{1}$$

Of course, if the data points don't lie on a line, then there are no parameters β_0, β_1 for which the predicted y-values in $X\boldsymbol{\beta}$ equal the observed y-values in $\mathbf{y}$, and $X\boldsymbol{\beta} = \mathbf{y}$ has no solution. This is a least-squares problem, $A\mathbf{x} = \mathbf{b}$, with different notation!

The square of the distance between the vectors $X\boldsymbol{\beta}$ and $\mathbf{y}$ is precisely the sum of the squares of the residuals. The $\boldsymbol{\beta}$ that minimizes this sum also minimizes the distance between $X\boldsymbol{\beta}$ and $\mathbf{y}$. *Computing the least-squares solution of $X\boldsymbol{\beta} = \mathbf{y}$ is equivalent to finding the $\boldsymbol{\beta}$ that determines the least-squares line in Figure 1.*

EXAMPLE 1 Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points $(2, 1)$, $(5, 2)$, $(7, 3)$, and $(8, 3)$.

SOLUTION Use the x-coordinates of the data to build the design matrix X in (1) and the y-coordinates to build the observation vector $\mathbf{y}$:

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$$

For the least-squares solution of $X\boldsymbol{\beta} = \mathbf{y}$, obtain the normal equations (with the new notation):

$$\boxed{X^T X\boldsymbol{\beta} = X^T\mathbf{y}}$$

That is, compute

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

The normal equations are

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

Hence

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 142 & -22 \\ -22 & 4 \end{bmatrix} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 24 \\ 30 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$$

Thus the least-squares line has the equation

$$y = \frac{2}{7} + \frac{5}{14}x$$

See Figure 2. ■

$(\boxed{2}, 1) \quad (\boxed{5}, 2) \quad (\boxed{7}, 3) \quad (\boxed{8}, 3)$

(x, y) النقطة

$XB = y$

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}_{4\times 2} \qquad y = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$$

$$\boxed{X^T X B = X^T y}$$

$X^T_{2\times 4}$ خلي الصف عامود :

$$X^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$2 \times \boxed{4 \quad \checkmark \quad 4} \times 2$ $\qquad 2 \times 2$

$4 + 25 + 49 + 64$

$$X^T y = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

$2 \times \boxed{4 \quad \checkmark \quad 4} \times 1$ $\qquad 2 \times 1$

$2 + 10 + 21 + 24$

$$\overset{A}{\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}}^{-1} \begin{bmatrix} B_0 \\ B_1 \end{bmatrix} = \overset{C}{\begin{bmatrix} 9 \\ 57 \end{bmatrix}}$$

Inverse

$$\cancel{A^{-1}} \cancel{A} B = A^{-1} C \Rightarrow B = A^{-1} C$$

$$B = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

$$B = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix} \begin{matrix} \rightarrow \beta_0 \\ \rightarrow \beta_1 \end{matrix}$$

$$y = \beta_0 + \beta_1 x$$

$$\boxed{y = \frac{2}{7} + \frac{5}{14} x}$$

Pre-Final Review

Problem 1

6

Use the least squares method to find the equation that best fits the dataset

$$\mathcal{D} = \{(0, 2.1), (0.5, 4.4), (1, 6.9), (1.5, 10.5), (2, 12.2)\}$$

Solution

Let the equation of the best-fit line be

$$y = \beta_0 + \beta_1 x$$

Thus

$$\vec{y} = X\vec{\beta} \Rightarrow \begin{bmatrix} 2.0 \\ 4.5 \\ 7.0 \\ 9.5 \\ 12.0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0.5 \\ 1 & 1 \\ 1 & 1.5 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

Using the normal equation of the least squares method, we get

$$X^T X \hat{\beta} = X^T \vec{y}$$

To solve this equation, we need to show that $(X^T X)^{-1}$ exists.

First, we compute this matrix

$$X^T X = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0.5 & 1 & 1.5 & 2 \end{bmatrix}}_{2\times 5} \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 0.5 \\ 1 & 1 \\ 1 & 1.5 \\ 1 & 2 \end{bmatrix}}_{5\times 2} = \underbrace{\begin{bmatrix} 5 & 5 \\ 5 & 7.5 \end{bmatrix}}_{2\times 2}$$

Now, we compute its determinant

$$|X^T X| = 5\times 7.5 - 5\times 5 = 12.5 \neq 0$$

Since $|X^T X| \neq 0$, we conclude that $(X^T X)^{-1}$ exists.

Now, we can solve the normal equation

$$X^T X \hat{\beta} = X^T \vec{y} \Rightarrow (X^T X)^{-1}(X^T X)\hat{\beta} = (X^T X)^{-1} X^T \vec{y}$$

$$\Rightarrow I_2 \hat{\beta} = (X^T X)^{-1} X^T \vec{y}$$

$$\Rightarrow \hat{\beta} = (X^T X)^{-1} X^T \vec{y}$$

$$(X^T X)^{-1} = \begin{bmatrix} 5 & 5 \\ 5 & 7.5 \end{bmatrix}^{-1} = \frac{1}{12.5} \begin{bmatrix} 7.5 & -5 \\ -5 & 5 \end{bmatrix} = \begin{bmatrix} 0.6 & -0.4 \\ -0.4 & 0.4 \end{bmatrix}$$

$$X^T \vec{y} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0.5 & 1 & 1.5 & 2 \end{bmatrix} \begin{bmatrix} 2.0 \\ 4.5 \\ 7.0 \\ 9.5 \\ 12.0 \end{bmatrix} = \begin{bmatrix} 35 \\ 47.5 \end{bmatrix}$$

$(0, 2.1)$ $(0.5, 4.4)$ $(1, 6.9)$
$(1.5, 10.5)$ $(2, 12.2)$ =

$y = B_0 + B_1 x$ (x, y)

$$\boxed{X^T X B = X^T y}$$

??

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 0.5 \\ 1 & 1 \\ 1 & 1.5 \\ 1 & 2 \end{bmatrix} \qquad y = \begin{bmatrix} 2 \\ 4.5 \\ 7 \\ 9.5 \\ 12 \end{bmatrix}$$

+2.5
+2.5
2.5 ??
2.5

$$X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0.5 & 1 & 1.5 & 2 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0.5 & 1 & 1.5 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0.5 \\ 1 & 1 \\ 1 & 1.5 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 5 \\ 5 & 7.5 \end{bmatrix}$$

2×5 5×2 2×2

$0 + 0.25 + 1 + 2.25 + 4 =$

$$X^T y = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0.5 & 1 & 1.5 & 2 \end{bmatrix}}_{2\times 5} \begin{bmatrix} 2 \\ 4.5 \\ 7 \\ 9.5 \\ 12 \end{bmatrix} = \underbrace{\begin{bmatrix} 35 \\ 47.5 \end{bmatrix}}_{2\times 1}$$

$$0 + 2.25 + 7 + 14.25 + 24 \quad 5\times 1$$

$$\begin{bmatrix} 5 & 5 \\ 5 & 7.5 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 35 \\ 47.5 \end{bmatrix}$$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \underbrace{\begin{bmatrix} 5 & 5 \\ 5 & 7.5 \end{bmatrix}}_{2\times 2}^{-1} \begin{bmatrix} 35 \\ 47.5 \end{bmatrix}$$

$$\begin{bmatrix} 5 & 5 \\ 5 & 7.5 \end{bmatrix}^{-1} = \frac{1}{5\times 7.5 - 5\times 5} \begin{bmatrix} 7.5 & -5 \\ -5 & 5 \end{bmatrix}$$

12.5

$\det \neq 0 \Rightarrow$ inverse

$$= \begin{bmatrix} \frac{7.5}{12.5} & \frac{-5}{12.5} \\ \frac{-5}{12.5} & \frac{5}{12.5} \end{bmatrix} = \begin{bmatrix} 0.6 & -0.4 \\ -0.4 & 0.4 \end{bmatrix}$$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 0.6 & -0.4 \\ -0.4 & 0.4 \end{bmatrix} \begin{bmatrix} 35 \\ 47.5 \end{bmatrix}$$

$$\boxed{2 \times 2 \quad \checkmark \quad 2 \times 1}$$

$$= \begin{bmatrix} 2 \\ 5 \end{bmatrix} \begin{array}{l} \to \beta_0 \\ \to \beta_1 \end{array}$$

$$2 \times 1$$

$$0.6 \times 35 + -0.4 \times 47.5$$

$$-0.4 \times 35 + 0.4 \times 47.5$$

$$y = \beta_0 + \beta_1 x$$

$$\boxed{y = 2 + 5x}$$

Thus

$$\hat{\beta} = (X^T X)^{-1}(X^T \vec{y}) = \begin{bmatrix} 0.6 & -0.4 \\ -0.4 & 0.4 \end{bmatrix} \begin{bmatrix} 35 \\ 47.5 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

Finally, the equation of the best-fit line is

$$\boxed{y = 2 + 5x}$$

Problem 2

Let $\vec{a}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ $\vec{a}_2 = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}$ $\vec{a}_3 = \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix}$

1) Show that these vectors are linearly independent
2) Apply the Gram-Schmidt process to construct an orthonormal basis
3) Find a QR factorization of the matrix A.
4) What are the eigenvalues of the R matrix.

Solution

1) To show that these vectors are linearly independent we need to show that

$$c_1\vec{a}_1 + c_2\vec{a}_2 + c_3\vec{a}_3 = \vec{0} \Rightarrow c_1 = c_2 = c_3 = 0$$

Equivalently

$$\underbrace{\begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \vec{a}_3 \end{bmatrix}}_{A} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow c_1 = c_2 = c_3 = 0$$

where

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 1 \end{bmatrix} \Rightarrow |A| = 1\begin{vmatrix} 1 & 4 \\ 4 & 1 \end{vmatrix} - 2\begin{vmatrix} 2 & 4 \\ 3 & 1 \end{vmatrix} + 3\begin{vmatrix} 2 & 1 \\ 3 & 4 \end{vmatrix} = 20$$

Since $|A| \neq 0$, A^{-1} exists and $\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = A^{-1}\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

Consequently, $\vec{a}_1, \vec{a}_2, \vec{a}_3$ are linearly independent

2) Gram-Schmidt Process

First, we construct a set of orthogonal vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$

$$\vec{v}_1 = \vec{a}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$\vec{v}_2 = \vec{a}_2 - \frac{\vec{a}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} - \frac{\begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \frac{1}{7}\begin{bmatrix} 6 \\ -9 \\ 4 \end{bmatrix}$$

PROBLEM 3

- (a) Use the least-squares method to find the equation of the line $L(x)$ that best fits the dataset:

$$D = \{(-4,-5),(-2,-1),(0,1),(2,3),(4,7)\}$$

- (b) The deviation of a data point (x_i, y_i) from the line $L(x)$ found in part (a) is defined by the error:

$$e_i = y_i - L(x_i)$$

To add new data points to the dataset without changing $L(x)$, the errors of the new points must satisfy:

$$\boxed{\sum_i e_i = 0} \quad \text{and} \quad \sum_i x_i e_i = 0$$

Using these conditions, show that if a single new data point (x_0, y_0) is added and the least-squares line remains $L(x)$, then the new point must lie on $L(x)$.

$$L(x) = y = B_0 + B_1 x$$

$(-4, -5)$ $(-2, -1)$ $(0, 1)$ $(2, 3)$ $(4, 7)$

(x, y)

$$\underbrace{X^T X} B = X^T y$$

$$X = \begin{bmatrix} 1 & -4 \\ 1 & -2 \\ 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix} \qquad y = \begin{bmatrix} -5 \\ -1 \\ 1 \\ 3 \\ 7 \end{bmatrix}$$

$$X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -4 & -2 & 0 & 2 & 4 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -4 & -2 & 0 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & -4 \\ 1 & -2 \\ 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 40 \end{bmatrix}$$

$16 + 4 + 0 + 4 + 16$

$$x^T y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -4 & -2 & 0 & 2 & 4 \end{bmatrix} \begin{bmatrix} -5 \\ -1 \\ 1 \\ 3 \\ 7 \end{bmatrix} \rightarrow \begin{bmatrix} 5 \\ 56 \end{bmatrix}$$

$$20 + 2 + 0 + 6 + 28$$

$$\begin{bmatrix} 5 & 0 \\ 0 & 40 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 5 \\ 56 \end{bmatrix}$$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 40 \end{bmatrix}^{-1} \begin{bmatrix} 5 \\ 56 \end{bmatrix}$$

$$\begin{bmatrix} 5 & 0 \\ 0 & 40 \end{bmatrix}^{-1} = \frac{1}{200} \begin{bmatrix} 40 & 0 \\ 0 & 5 \end{bmatrix}$$

$$\begin{bmatrix} 5 & 0 \\ 0 & 40 \end{bmatrix}^{-1} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.025 \end{bmatrix}$$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.025 \end{bmatrix} \begin{bmatrix} 5 \\ 56 \end{bmatrix} = \begin{bmatrix} 1 \\ 1.4 \end{bmatrix}$$

$$\boxed{y = 1 + 1.4x}$$

error : $y_{c'} - L(x_{c'})$

$(-4, \boxed{-5})$ $(-2, \boxed{-1})$ $(0, \boxed{1})$

$(2, \boxed{3})$ $(4, \boxed{7})$

$L(x_{c'} = -4) = 1 + 1.4(-4) = -4.6$

$e = -5 - (-4.6) = \boxed{-0.4}$

$L(x_{c'} = -2) = 1 + 1.4(-2) = -1.8$

$e = -1 - (-1.8) = \boxed{0.8}$

$L(x_{c'} = 0) = 1 + 1.4(0) = 1$

$e = 1 - 1 = \boxed{0}$

$L(x_{c'} = 2) = 1 + 1.4(2) = 3.8$

$e = 3 - 3.8 = \boxed{-0.8}$

$$\hat{y}(x_i = 4) = 1 + 1,4(4) = 6,6$$

$$e = 7 - 6,6 = \boxed{0,4}$$

$$\Sigma e = \cancel{-0,4} + \cancel{0,8} + 0 + \cancel{-0,8} + \cancel{0,4}$$

$$\boxed{\Sigma e = 0}$$