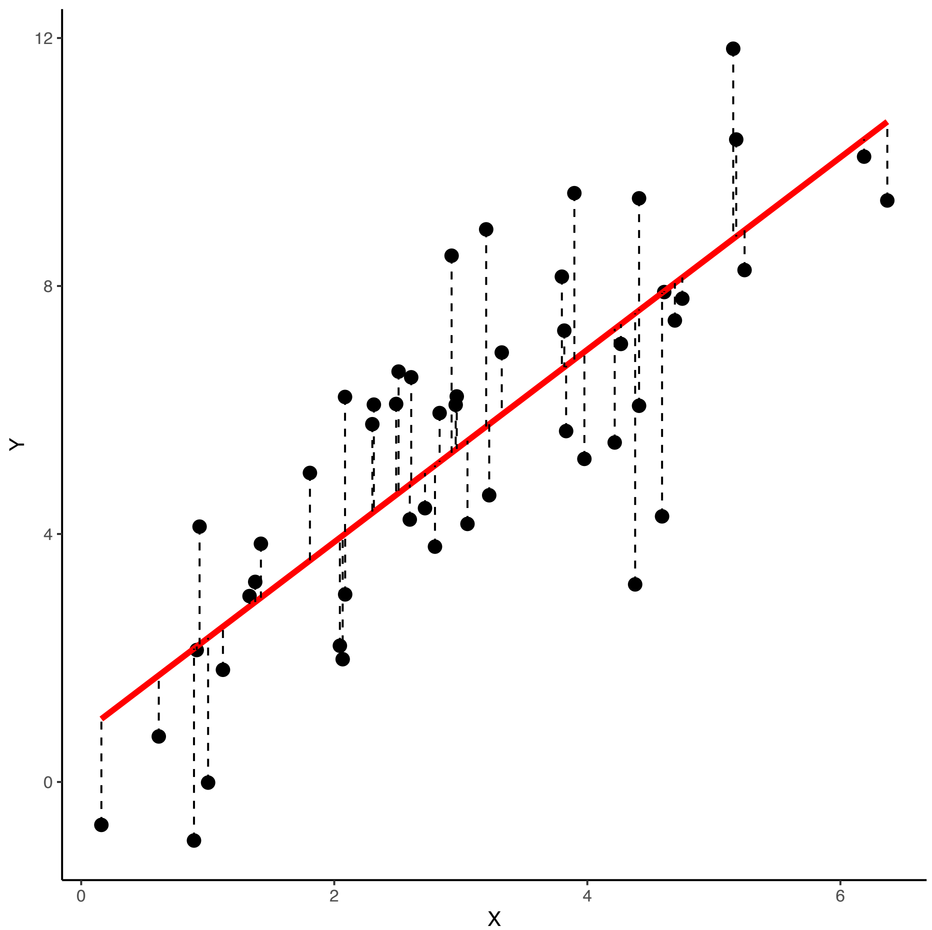
Bivariate Regression and Introduction to Matrices

Paul M. Magwene

Bivariate Regression: Variable Space Representation

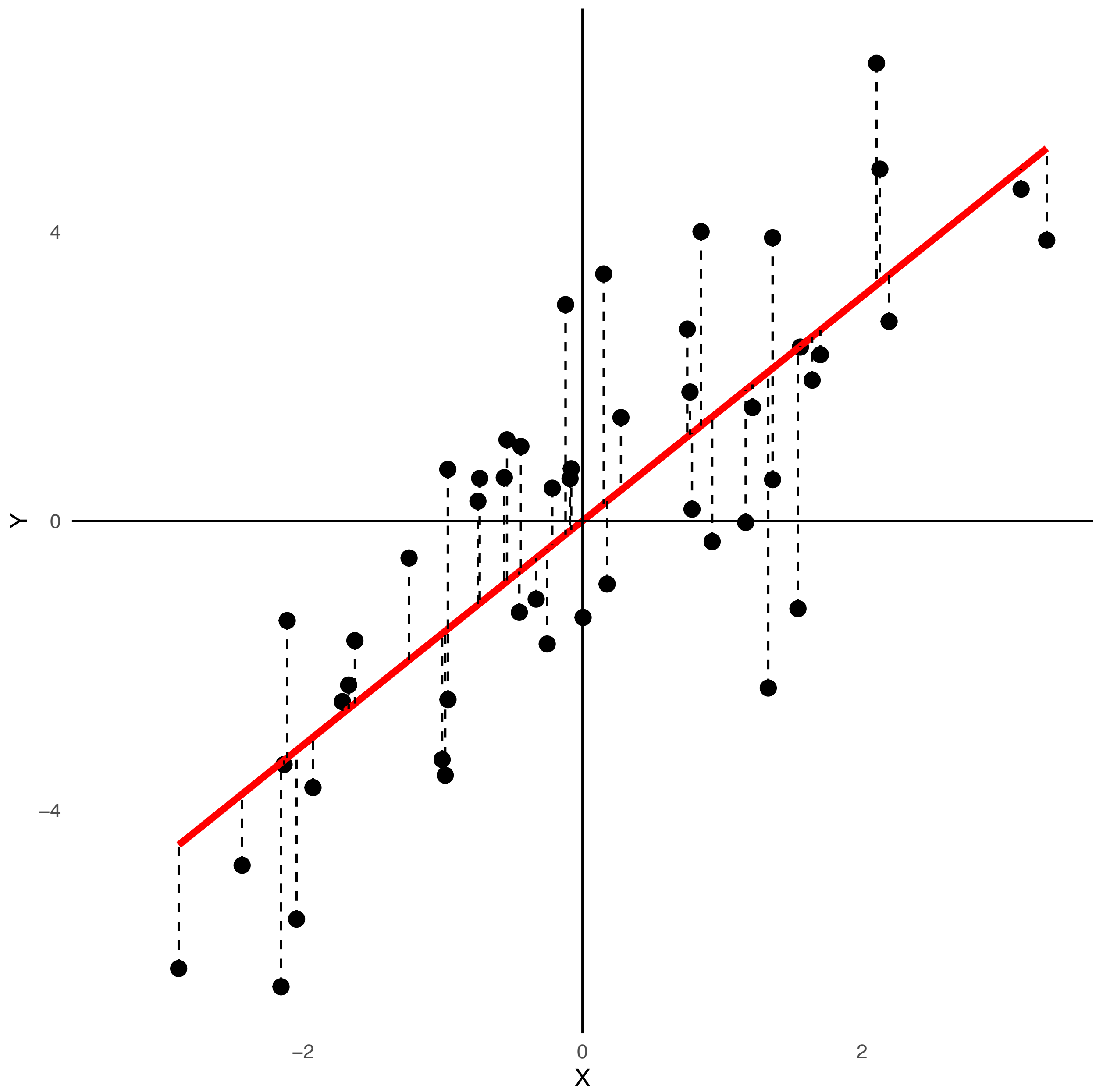


The standard bivariate regression equation relating one observed variable X (the predictor) to another observed variable of interest, Y (the outcome) is usually written as:

$$\widehat{Y} = a + bX.$$

where $\widehat{Y}$ is the predicted value of Y and a (intercept) and b (slope) are chosen to minimize $\sum(Y_i - \hat{Y}_i)^2$.

Bivariate regression, centered variables



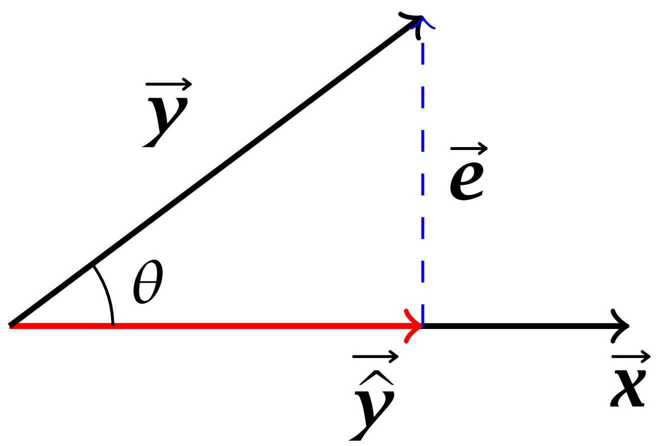
$$\hat{Y} = bX$$

Geometry of Bivariate Regression

Let's express this in vector terms, and work with mean-centered vectors so the equation becomes:

$$\vec{\hat{y}} = b\vec{x}$$

Geometric interpretation of regression as projection:



$$\vec{\hat{y}} = b\vec{x} \tag{1}$$

$$b = \frac{\vec{x} \cdot \vec{y}}{\vec{x} \cdot \vec{x}} \tag{2}$$

Bivariate regression: Alternate formulas for slope I

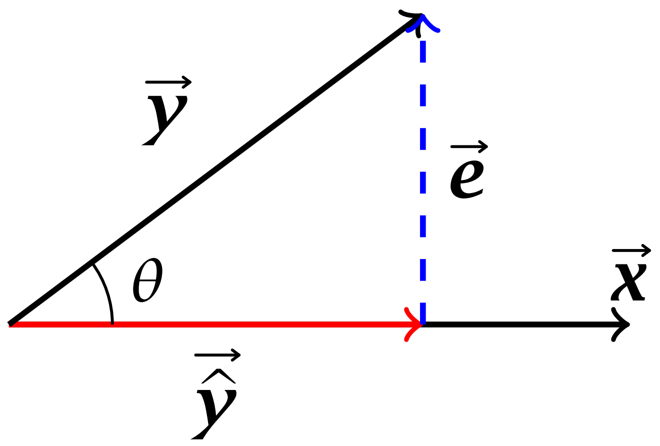
Regression equation for mean-centered vectors: $\vec{\hat{\boldsymbol{y}}} = b\vec{\boldsymbol{x}}$

There are multiple, equivalent ways to write the solution for b:

$$
\begin{aligned}
b &= \frac{\vec{\boldsymbol{x}} \cdot \vec{\boldsymbol{y}}}{(\vec{\boldsymbol{x}} \cdot \vec{\boldsymbol{x}})} \\
&= \frac{\vec{\boldsymbol{x}} \cdot \vec{\boldsymbol{y}}}{|\vec{\boldsymbol{x}}|^2} \\
&= \frac{|x||y| \cos\theta}{|x|^2} \\
&= \cos\theta \frac{|y|}{|x|} \\
&= r_{XY} \frac{|y|}{|x|}
\end{aligned}
$$

Geometry of Goodness of Fit

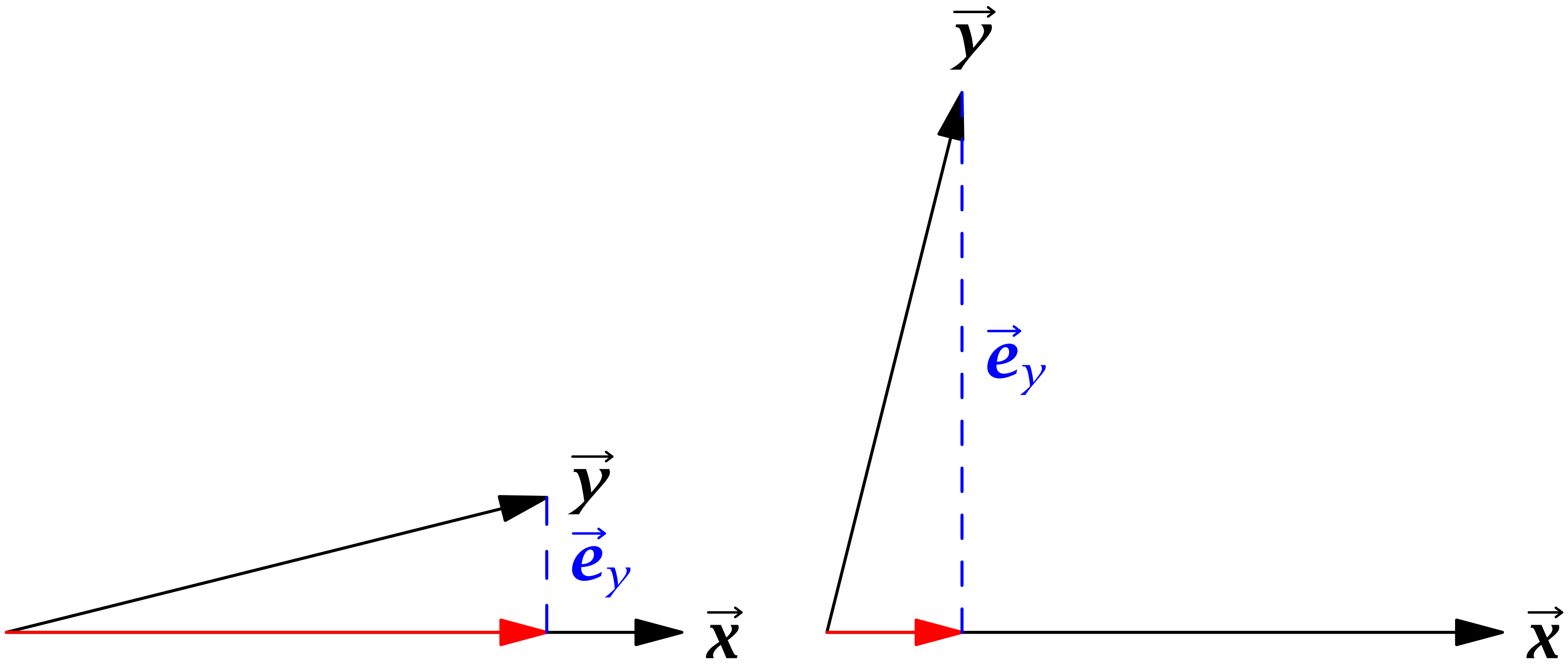
Geometric interpretation of regression goodness-of-fit:



$$|\vec{\hat{y}}|^2 + |\vec{e}|^2 = |\vec{y}|^2$$

The better the goodness-of-fit, the smaller the angle, $\cos\theta$, and the shorter residual vector, $\vec{e}$.

Geometry of Goodness of Fit



(a) Good fit

(b) Bad fit

Bivariate Regression, Goodness of Fit

How well does our prediction agree with our outcome?

- Measure the angle between $\vec{\hat{y}}$ and $\vec{y}$:

$$R = \cos\theta_{\vec{y},\vec{\hat{y}}} = \frac{|\vec{\hat{y}}|}{|\vec{y}|}$$

- In the single-predictor case $R = r_{XY}$, but this is not generally true when we have multiple predictors.
- Note that $|\vec{y}|$ can be expressed as follows:

$$\begin{aligned} |\vec{\hat{y}}|^2 + |\vec{e}|^2 &= |\vec{y}|^2 \\ SS_{\text{regression}} + SS_{\text{residual}} &= SS_{\text{total}} \end{aligned}$$

- With simple substitution we can show that:

$$\begin{aligned} SS_{\text{regression}} &= R^2 SS_{\text{total}} \\ SS_{\text{residual}} &= (1 - R^2) SS_{\text{total}} \end{aligned}$$

Vector representation of bivariate regression, non-centered case

$$\vec{\hat{\mathbf{y}}} = a\vec{\mathbf{1}} + b\vec{\mathbf{x}}$$

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = a \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + b \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

The predicted value of y is a linear combination of the 1-vector and x

Introduction to Matrices

Introduction to Matrices

- One way to think about a matrix is as a collection of vectors. This is, in essence, what a multivariate data set is.
- A matrix which has n rows and p columns will be referred to as a $n \times p$ matrix. $n \times p$ is the shape of the matrix.

$$\underset{(n\times p)}{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}$$

Scalar Multiplication of a Matrix

- Let k be a scalar and let A be the matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}$$

- then

$$kA = \begin{bmatrix} ka_{11} & ka_{12} & \cdots & ka_{1p} \\ ka_{21} & ka_{22} & \cdots & ka_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ ka_{n1} & ka_{n2} & \cdots & ka_{np} \end{bmatrix}$$

Addition and Subtraction of Matrices

- Let A and B be matrices that have the same shape, $n \times p$:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix}$$

- then

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1p} + b_{1p} \\ a_{21} + b_{11} & a_{22} + b_{22} & \cdots & a_{2p} + b_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \cdots & a_{np} + b_{np} \end{bmatrix}$$

$$A - B = A + (-B)$$

Multiplying a Matrix by a Vector

- Let A be a $n \times p$ matrix, and let $\boldsymbol{x}$ be a $p \times 1$ column vector

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

- then

$$A\boldsymbol{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{np}x_p \end{bmatrix}$$

Note that $A\boldsymbol{x}$ is a vector with shape $n \times 1$. The i-the element of $A\boldsymbol{x}$ is equivalent to the dot product of the i-th row vector of A with $\boldsymbol{x}$.

General Matrix Multiplication

- Let A be a $n \times p$ matrix and B be a $p \times q$ matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1q} \\ b_{21} & b_{22} & \cdots & b_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ b_{p1} & b_{n2} & \cdots & b_{nq} \end{bmatrix}$$

- The product AB is an $n \times q$ matrix whose (i, j)-entry is the dot product of the i-th row vector of A and the j-th column vector of B.
- A and B muse be *conformable* to calculate the product AB, i.e. the number of columns in A must be the same as the number of rows in B.

Matrix Arithmetic Rules

- i $A + B = B + A$
- ii $(A + B) + C = A + (B + C)$
- iii $k(A + B) = kA + kB$
- iv $(kA)B = k(AB)$
- v $(AB)C = A(BC)$ (associative)
- vi $A(B + C) = AB + AC$ (distributive)
- vii $(A + B)C = AC + BC$ (distributive)

Alert

*Matrix multiplication is **not** commutative, i.e. $AB \neq BA$ in general.*

Be careful when you expand expressions like $(A + B)(A + B)$.

Matrix Transpose

- We denote the transpose of a matrix as A^T
- If A is an $n \times p$ matrix, then A^T is a $p \times n$ matrix where $A^T_{ji} = A_{ij}$
- Transpose rules:
    - $(A^T)^T = A$
    - $(A + B)^T = A^T + B^T$
    - $(AB)^T = B^T A^T$
- Symmetric matrix– square matrix, A, where $A^T = A$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}$$

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1p} & a_{12} & \cdots & a_{np} \end{bmatrix}$$

Special Matrices

- Zero matrix

$$\mathbf{0} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

- Square matrix
  A matrix whose shape is is $n \times n$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

- Ones matrix

$$\mathbf{1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

- Diagonal matrix
  A square matrix where the off-diagonal elements are zero.

$$A = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

Identity matrix

An *identity matrix* is a $p \times p$ matrix with ones on the diagonal and zeros everywhere else.

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

- $IA = AI = A$ if I and A are $n \times p$ matrices
- $A = I\boldsymbol{x}$ is a diagonal matrix where $a_{ii} = x_i$ if I is an $n \times n$ matrix and $\boldsymbol{x}$ is a $n \times 1$ vector.

More Special Matrices

Symmetric matrix – square matrix, A, where $A^T = A$

Skew-symmetric matrix – square matrix, A, where $A^T = -A$

Orthogonal matrix – square matrix for which $A^T A = AA^T = I$.

Matrix Inverse

- If A is a *square matrix* and C is a matrix of the same size where $AC = I$ and $CA = I$ than C is the inverse of A and we denote it as A^{-1}.

$$AA^{-1} = A^{-1}A = I$$

- Rules for inverses:
  - Only square matrices are invertible; but not every square matrix can be inverted
  - A matrix for which we can find an inverse is called ***invertible*** (non-singular)
  - A matrix for which no inverse exists is ***singular*** (non-invertible)
  - Any diagonal matrix, A, where the a_{ii} are non-zero, is invertible
  - If A and B are both invertible $p \times p$ matrices than $(AB)^{-1} = B^{-1}A^{-1}$ (note change in order).

Descriptive statistics as matrix functions

Mean vector and mean matrix

Assume you have a data set represented as a $n \times p$ matrix, $\boldsymbol{X}$, with observations in rows and variables in columns.

Mean vector

To calculate a row vector of means, $\boldsymbol{m} = [\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_p]$

$$\boldsymbol{m} = \frac{1}{n}\mathbb{1}^T\boldsymbol{X}$$

where $\mathbb{1}$ is a $n \times 1$ column vector of ones.

Matrix of column means

$$\boldsymbol{M} = \mathbb{1}\boldsymbol{m}$$

Deviation matrix

To re-express each value as the deviation from the variable means (i.e. each column is a mean centered vector) we calculate a deviation matrix:

$$\boldsymbol{D} = \boldsymbol{X} - \boldsymbol{M}$$

Covariance and correlation matrices

Covariance matrix

$$\boldsymbol{S} = \frac{1}{n-1}\boldsymbol{D}^T\boldsymbol{D}$$

Correlation matrix

$$\boldsymbol{R} = \boldsymbol{V}\boldsymbol{S}\boldsymbol{V}$$

where $\boldsymbol{V}$ is a $p \times p$ diagonal matrix where $\boldsymbol{V}_{ii} = 1/\sqrt{\boldsymbol{S}_{ii}}$.

Matrix Concepts

Space Spanned by a List of Vectors

Definition

Let X be a finite list of n-vectors. The **space spanned** by X is the set of all vectors that can be written as linear combinations of the vectors in X.

Remember that a *linear combination* of vectors is an equation of the form $z = b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + \cdots + b_p\mathbf{x}_p$

Bivariate regression in matrix terms

Vector version

$$\vec{\hat{\mathbf{y}}} = a\vec{\mathbf{1}} + b\vec{\mathbf{x}}$$

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = a \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + b \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Matrix version

$$\vec{\hat{\mathbf{y}}} = \mathbf{X}\vec{\mathbf{b}}$$

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

The predictor is the vector in the
space spanned by the 1-vector and **x** that is closest to **y**