Chapter 1

INTRODUCTION

Linear models have a dominant role in statistical theory and practice. Most standard statistical methods are special cases of the general linear model, and rely on the corresponding theory for justification.

The goal of this course is to develop the theoretical basis for analyses based on a linear model. We shall be concerned with laying the theoretical foundation for simple as well as complex data sets.

Linear model is one of the oldest topics in the statistics curriculum. The main role of linear model in statistical practice, however, has begun to undergo a fundamental change due in large measure to available computing. Balanced experiments were often required to make analysis possible. This has produced a fundamental change in the way we can think about linear models, as much less stress can be placed on the special cases where computations are easy and more can be placed on general ideas. Topics that might have been standard, such as the recovery of interblock information in an incomplete block experiment, is of much less interest when computers can be used to appropriately maximize functions.

However, standard results are so elegant, and so interesting, that they deserve study in their own right, and for that reason we will study the traditional body of material that makes up linear models, including many standard simple models as well as a general approach.

The goal of these notes is to develop a *coordinate-free approach* to linear models. Coordinates can often sever to make problems unnecessarily complex, and understanding the features of a problems that are not dependent on coordinates is extremely valuable. The problems introduced by parameters are more easily understood given the coordinate-free background.

Example 1.1 *Simple Linear Regression*

- *Modeling the relationship between two variables (x, y).*
- *We have observations (x_1, y_1), ..., (x_n, y_n), where*
  - *the x_i's are known fixed values (independent or predictor variables), and*
  - *the y_i's are dependent or response variables (random).*
- *We have the model:*

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ , \ \ i = 1, \ldots, n.$$

- *ε: random error representing random fluctuations, measurement errors, or the effect of factors outside of our control*
  - *y: annual melanoma mortality in a state, x: latitude of the center of the state (in degrees).*
  - *y: income, x: number of years of education*
  - *y: height, x: weight of people*
  - *y: response, x: dose of a drug*
- *Usual assumptions:*

– *The ε_i's are $i.i.d.$ from some distribution with*

$$E(\varepsilon_i) = 0 \text{ and } Var(\varepsilon_i) = \sigma^2.$$

- *With this model, we may want to*

  1. *Estimate β_0, β_1, and σ^2.*
  2. *Test hypotheses about β_1, confidence limits for β_1.*
  3. *Predict a future y at a given x.*

- *To do exact inference, we need ε_i to be normally distributed. Otherwise, we have to base inferences on large sample theory.*

- *Estimation and inferential procedures for the simple linear regression model are developed and illustrated later in detail.*

Example 1.2 *One Way Analysis-of-Varinace (ANOVA)*

- *Interested in comparing several populations or several conditions in a study*

- *Example: We want to examine the effect of NO_2 on the lungs. Consider mice which are i) not exposed ii) mildly exposed, iii) heavily exposed. ($k = 3$). Response variable is percent serum fluorescence. High readings indicate damage to lung tissues.*

- *Suppose we have observations y_{ij}, $j = 1, \ldots, n_i$, $i = 1, \ldots, k$. We have k populations with n_i observations in population i.*

pop1	*pop2*	...	*popk*
y_{11}	y_{21}	$\cdots$	y_{k1}
$\vdots$	$\vdots$		$\vdots$
y_{1n_1}	y_{2n_2}	$\cdots$	y_{kn_k}

- *Assume* $E(y_{ij}) = \mu_i$ *and* $Var(y_{ij}) = \sigma^2$.

- *We can write:*

$$y_{ij} \;=\; \mu_i + \varepsilon_{ij}\ ,\ \textit{where}\ j = 1, \ldots, n_i, i = 1, \ldots, k.$$

- *The* ε_{ij}*'s are* $i.i.d.$ *from a distribution with* $E(\varepsilon_{ij}) = 0$*, and* $Var(\varepsilon_{ij}) = \sigma^2$.

- *The model* $y_{ij} = \mu_i + \varepsilon_{ij}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, k$ *is often referred to as a* <u>*means*</u> *model.*

- *We may want to*

  1. *Estimate the* μ_i*'s and* σ^2.
  2. *Test hypotheses about the* μ_i*'s, e.g.*

     $$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$

     or

     $$H_0 : c_1\mu_1 + c_2\mu_2 + \ldots + c_k\mu_k = 0\ \ ,$$

     where $\sum_{i=1}^{k} c_i = 0$.
  3. *Decide which group mean is largest, multiple comparisons, etc ....*

- *If we reparameterize and write*

$$\mu_i = \mu + \alpha_i,$$

then the model becomes

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}\ , j = 1, \ldots, n_i, i = 1, \ldots, k.$$

- *ANOVA models: special case of regression models but will be treated separately later in the course.*

Example 1.3 *General Linear Model*

- *Both of the examples just presented and many others are special cases of the general linear model.*

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i,\ i = 1, \ldots, n,$$

  where

  - y_i*'s are observations;*
  - x_{ij}*'s are known (fixed) values;*
  - β_j*'s are unknown parameters;*
  - ε_i*'s are* unobservable *random variables with*

$$E(\varepsilon_i) = 0,\ \textit{Cov}(\varepsilon_i, \varepsilon_j) = \sigma_{ij},\ i, j = 1, \ldots, n.$$

- *This is the most general setup.*

- **Note***: If an intercept is included in the model, then*

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i.$$

- *Often, more restrictive assumptions are made about the* ε_i*'s.*

1. *Uncorrelated:* $Cov(\varepsilon_i, \varepsilon_j) = 0$ *for* $i \neq j$.
2. *Equal variances:* $Cov(\varepsilon_i, \varepsilon_i) = var(\varepsilon_i) = \sigma^2$, $i = 1, \ldots, n$.
3. *the* ε_i*'s are normally distributed.*

- *Example:*
  - y_i = *oxygen consumption*
  - x_1 = *treadmill duration*
  - x_2 = *heart rate*
  - x_3 = *age*
  - x_4 = *height*

- *We can write this model in matrix notation. Define*

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}_{n\times 1} \qquad X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}_{n\times p}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}_{p\times 1} \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}_{n\times 1}$$

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ & \sigma_{22} & & \vdots \\ & & \ddots & \vdots \\ & & & \sigma_{nn} \end{pmatrix}_{n\times n} .$$

- *We can now write the model as*

$$Y = X\beta + \varepsilon ,$$

$$E(\varepsilon) = 0, \quad \mathit{Cov}(\varepsilon) = \Sigma \,,$$

or often

$\mathit{Cov}(\varepsilon) = \sigma^2 I$, in the uncorrelated, equal variance errors case.

- **Note:** *In this formulation, the model says that*

$$\begin{aligned} \mu &= E(Y) = X\beta \\ &= [X_1, X_2, \ldots, X_p]\beta \\ &= \beta_1 X_1 + \ldots + \beta_p X_p \,, \end{aligned}$$

where

$$X_j = \begin{pmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{pmatrix}$$

is the jth column of X, $j = 1, \ldots, p$.

We see that μ is a <u>linear combination</u> of the columns of X.

- *In a more abstract setting,*

$$\mu \in \Omega,$$

where Ω is a linear subspace in an n dimensional vector space.

- *In the general linear model, we may be interested in*

1. **Prediction**: *overall influence of the x variables on y*
2. **Data Description** *or* **Explanation**: *use of the estimated model to summarize or describe the observed data*
3. **Parameter Estimation**: *estimates of the model parameters are essential for prediction and data description or explanation.*
4. **Variable Selection** *or* **Screening**: *determining the importance of each predictor variable in modeling the variation in y*
5. **Control of Output**: *Under a cause-and-effect relationship between x and y, use the estimated model to control the output of a process.*

- *Estimation and inferential procedures are discussed later in detail.*

- *The one–way ANOVA model introduced earlier:*

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij},\ j = 1, \ldots, n_i, i = 1, \ldots, k,$$

can now be written in matrix notation as

$$Y = X\beta + \varepsilon, \varepsilon \sim N_N(0, \sigma^2 I),$$

where $N = \sum_{i=1}^{k} n_i$,

$$Y = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{kn_k} \end{pmatrix}_{N\times 1}, \qquad \varepsilon = \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \\ \vdots \\ \varepsilon_{kn_k} \end{pmatrix}_{N\times 1}$$

$$X = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & 1 & 0 & \dots & \vdots \\ \vdots & 0 & 1 & \dots & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & 1 & \dots & 0 \\ \vdots & \vdots & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix}_{N\times(k+1)} \qquad \beta = \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix}_{(k+1)\times 1}$$

- *For example if* $k = 3$, $n_1 = 3$, $n_2 = 1$, $n_3 = 2$, $N = 3 + 1 + 2 = 6$

$$Y = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{31} \\ y_{32} \end{pmatrix}_{6\times 1} \qquad X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}_{6\times 4}$$

$$\beta = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}_{4\times 1} \quad \varepsilon = \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{31} \\ \varepsilon_{32} \end{pmatrix}_{6\times 1} \quad .$$

Chapter 2

MATRIX ALGEBRA

Useful resource: Schott (1997), Christensen (2011, Appendices A and B)

Matrix, Vector and Scalar

- *Matrix*:
  - A rectangular or square array of numbers of variables;
  - Real numbers or variables for all elements of matrices
  - $$\mathbf{A} = (a_{ij}) = \begin{pmatrix} 65 & 154 \\ 73 & 182 \\ 68 & 167 \end{pmatrix} = \begin{pmatrix} 65 & 154 \\ 73 & 182 \\ 68 & 167 \end{pmatrix} \tag{2.1}$$
    where the subscript i and j denote the row and column, respectively.
  - (2.1) has 3 rows and 2 columns; (the size of) $\mathbf{A}$ is 3×2
- *Vector*:
  - A matrix with a single row or column
  - Often use a single subscript to identify elements in a vector

- $$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},\ \mathbf{x}' = (x_1, x_2, x_3) = (x_1\ x_2\ x_3)$$

  where $\mathbf{x}$ is a column vector and $\mathbf{x}'$ is a row vector (will define the prime symbol (′), *transpose* shortly)

- Geometric interpretation: a vector with p elements can be thought as a point in a p-dimensional space

- *Scalar*:
  - A single real number
  - 2.5, -9, and 7.26
  - Technically distinct from 1×1 matrix
- **Matrix Equality**:
  - same size and elements in corresponding positions are equal
  - $$\begin{pmatrix} 3 & -2 & 4 \\ 1 & 3 & 7 \end{pmatrix} = \begin{pmatrix} 3 & -2 & 4 \\ 1 & 3 & 7 \end{pmatrix}.$$
- **Transpose**:
  - Interchanging the rows and columns of a matrix $\mathbf{A}$;
  - <u>Ex</u>:
    $$\mathbf{A} = \begin{pmatrix} 6 & -2 \\ 4 & 7 \\ 1 & 3 \end{pmatrix}, \mathbf{A}' = \begin{pmatrix} 6 & 4 & 1 \\ -2 & 7 & 3 \end{pmatrix}.$$
  - If $\mathbf{A} = (a_{ij})$ then $\mathbf{A}^{'} = (a_{ij})^{'} = (a_{ji})$.
  - What about the dimension of a matrix $\mathbf{A}$?
  - What if we transpose $\mathbf{A}$ twice?
- **Symmetric Matrix**:
  - If $\mathbf{A}^{'} = \mathbf{A}$;

– Ex:

$$\mathbf{A} = \begin{pmatrix} 3 & 2 & 6 \\ 2 & 10 & -7 \\ 6 & -7 & 9 \end{pmatrix}, \quad \text{symmetric?}$$

– All symmetric matrices are *square* (# of rows = # of columns);

- **Diagonal Matrix**:
  - The *diagonal* of a $p \times p$ square matrix $\mathbf{A} = (a_{ij})$: $a_{11}, a_{22}, \ldots, a_{pp}$;
  - If 0s for all off-diagonal positions;
  - Ex:

$$\mathbf{D} = \begin{pmatrix} 8 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}.$$

- **Identity Matrix**:
  - A diagonal matrix with a 1 in each diagonal position
  - Ex:

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

- **Upper Triangular Matrix**:
  - A square matrix with 0s below the diagonal
  - Ex:

$$\mathbf{T} = \begin{pmatrix} 7 & 2 & -3 & 5 \\ 0 & 0 & -2 & 6 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 8 \end{pmatrix}.$$

  - What is a **Lower Triangular Matrix**?

- **Other Special Matrices**:
  - **J**: a square matrix of 1s;

    Ex:
$$\mathbf{J} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$
  - **O**: a square matrix of 0s;

    Ex:
$$\mathbf{O} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$
  - **j**: a vector of 1s, **0**: a vector of 0's;

Operations

- *Sum* of Two Matrices (or Two Vectors):
  - *conformal*: two matrices (or vectors) are of the same size for addition;
  - Adding corresponding elements;
  - If $\mathbf{A} = (a_{ij})$: $n \times p$ and $\mathbf{B} = (b_{ij})$: $n \times p$ then

    $\mathbf{C} = \mathbf{A} + \mathbf{B} = (c_{ij}) = (a_{ij} + b_{ij})$: $n \times p$
  - Ex:
$$\begin{pmatrix} 7 & -3 & 4 \\ 2 & 8 & -5 \end{pmatrix} + \begin{pmatrix} 11 & 5 & -6 \\ 3 & 4 & 2 \end{pmatrix} = ? \qquad (2.2)$$

– What about the difference between $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{D} = \mathbf{A} - \mathbf{B}$?

Theorem If $\mathbf{A}$ and $\mathbf{B}$ are both $n \times m$, then

(i) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$ (commutative).

(ii) $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$.

– proof?

- *Product* of a Scalar and a Matrix:

  – Any scalar can be multiplied by any matrix;

  – Ex:

$$c\mathbf{A} = (ca_{ij}) = \begin{pmatrix} ca_{11} & ca_{12} & \cdots & ca_{1m} \\ ca_{21} & ca_{22} & \cdots & ca_{2m} \\ \vdots & \vdots & & \vdots \\ ca_{n1} & ca_{n2} & \cdots & ca_{nm} \end{pmatrix}$$

Theorem The product of a scalar and a matrix is commutative:

$$c\mathbf{A} = \mathbf{A}c.$$

- *Product* of Two Matrices (or Two Vectors):

  – To define $\mathbf{AB}$, $\mathbf{A}$ and $\mathbf{B}$ should be *conformal for multiplication*;

# of column in $\mathbf{A}$ = # of rows in $\mathbf{B}$

  – The (ij)th element of $\mathbf{C} = \mathbf{AB}$ is defined as

$$c_{ij} = \sum_k a_{ik} b_{kj}$$

– We multiply every row of $\mathbf{A}$ by every column of B.

– If $\mathbf{A}$ is $n \times m$ and $\mathbf{B}$ is $m \times p$ then $\mathbf{C}$ is $n \times p$.

Vector Spaces

The type of vector spaces we consider are *finite dimensional real vector spaces*.

Defn:
A real vector space $\mathcal{M}$ is a set of elements (called *vectors*) with the following properties:

A. Addition axioms:

For $x, y \in \mathcal{M}$, there corresponds a vector $x + y \in \mathcal{M}$ called the sum of x and y such that:

A1. $x + y = y + x$ (commutative)

A2. $x + (y + z) = (x + y) + z$ (associative)

A3. There exists a unique vector 0, the null vector, such that for all $x \in \mathcal{M}$, $x + 0 = x$

A4. for all $x \in \mathcal{M}$, there exists a unique element $-x$ such that $x + (-x) = 0$

B. Scalar multiplication axioms:

For any real number α and for any $x \in \mathcal{M}$, there exists a member of $\mathcal{M}$, αx, called product of α and x such that

B1. $\alpha(x + y) = \alpha x + \alpha y$ (distributive)

B2. $(\alpha + \beta)x = \alpha x + \beta x$ (distributive)

B3. $\alpha(\beta x) = (\alpha\beta)x$ (associative)

B4. There exists a scalr ξ such that $\xi \cdot x = x$ ($\xi = 1$).

Examples of Vector Spaces

Example 1 $\mathcal{M}$ = (n dimensional Euclidean space), R^n.

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$x + y = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix} \quad \alpha x = \begin{pmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{pmatrix} \quad .$$

Example 2 – The space of all 2×2 matrices with real elements is a vector space. $\mathcal{M} =$ space of a 2×2 matrices.

x is a 2×2 matrix.

$$0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$x = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad y = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}$$

$$x + y = \begin{pmatrix} 6 & 8 \\ 10 & 12 \end{pmatrix}$$

and so on.

Example 3 – The polynomials of degree n with real coefficients constitutes a vector space. $\mathcal{M} =$ polynomials of degree n.

$x = \beta_0 + \beta_1 t + \ldots + \beta_n t^n$.

$y = \gamma_0 + \gamma_1 t + \ldots + \gamma_n t^n$.

Defn: A set of vectors $D = \{x_1, \ldots, x_r\}$ is called linearly dependent if there is a set of scalars $\alpha_1, \ldots, \alpha_r$, not all zero, such that

$$\sum_{i=1}^{r} \alpha_i x_i = 0.$$

If $\sum_{i=1}^{r} \alpha_i x_i = 0 \Rightarrow \alpha_i = 0,\ i = 1, \ldots, r$, then $D = \{x_1, \ldots, x_r\}$ are linearly independent.

If D is linearly independent, then $D_1 \subset D$ is linearly independent.

Example 1: $\mathcal{M} = R^3$.

$$x_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad x_2 = \begin{pmatrix} 16 \\ 12 \\ 3 \end{pmatrix} \quad x_3 = \begin{pmatrix} 0 \\ 28 \\ 3 \end{pmatrix}$$

$D = \{x_1, x_2, x_3\}$ is linearly dependent since $16x_1 - x_2 + x_3 = 0$.

Example 2: Consider the linear model $y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$, $\beta_i \neq 0$, $i = 1, 2, 3$, and the $x_i^{'}$s are given by Example 1. What is the meaning of linear dependence for this linear model?

Since, for the example given above $x_2 - x_3 = 16x_1$, or $x_1 = (x_2 - x_3)/16$, by substituting for x_1, we write:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

$$= x_2(\beta_2 + \beta_1/16) + x_3(\beta_3 - \beta_1/16) + \varepsilon$$

$$= \gamma_1 x_2 + \gamma_2 x_3 + \varepsilon$$

where the $\gamma_i^{'}$s are defined by the above equation. This shows that the model with three parameters is equivalent to a model with only two parameters.

Example 3: $\mathcal{M} = R^n$

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad \ldots \quad e_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

$D = \{e_1, \ldots, e_n\}$ is linearly independent.

Suppose $D = \{x_1, \ldots, x_r\}$ is linearly dependent. This means that

$\sum_{i=1}^{r} \alpha_i x_i = 0$, and not all $\alpha_i = 0$.

Let k be such that $\alpha_k \neq 0$. Then

$$x_k = \sum_{\substack{i=1 \\ i \neq k}}^{r} \left(-\frac{\alpha_i}{\alpha_k} \right) x_i.$$

We are led to the following theorem.

Theorem A set of vectors is linearly dependent if and only if some vector of the set can be written as a linear combination of the others. That is there exists a k such that

$$x_k = \sum_{\substack{i=1 \\ i \neq k}}^{r} \left(-\frac{\alpha_i}{\alpha_k} \right) x_i,$$

where $\alpha_k \neq 0$.

Proof: Exercise.

Defn:
A basis in a vector space $\mathcal{M}$ is a set of linearly independent vectors such that every $x \in \mathcal{M}$ is a linear combination of vectors in the set.

Example: $\mathcal{M} = R^3$

The vectors

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

are a basis for R^3.

Note:

1. Every vector in R^3 can be written as a linear combination of the vectors above.
2. Bases are not unique.

The vectors

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 3 \\ 3 \\ 0 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

are also a basis for R^3.

Note: The representation of a vector as a linear combination of basis elements is unique. The coefficients of the vectors are often called coordinates with respect to that basis.

To illustrate, suppose $B = \{x_1, \ldots, x_r\}$ is a basis and $z = \sum_{i=1}^r \alpha_i x_i$.
Then, $z = \sum_{i=1}^r \beta_i x_i \Rightarrow \sum_{i=1}^r (\alpha_i - \beta_i) x_i = 0 \Rightarrow \alpha_i = \beta_i$, $i = 1, \ldots, n$.
In this example $(\alpha_1, \ldots, \alpha_r)$ are the coordinates of z with respect to B.
Defn:
A vector space $\mathcal{M}$ is said to be finite dimensional if it has a basis with a finite

number of elements.

Examples

- R^n is finite dimensional
- P^n is finite dimensional (polynomials of degree n).

Let D be a set of vectors.

1. The set of all possible linear combinations of elements of D is a vector space, called the span of D, $\mathcal{S}(D)$.
2. A basis for $\mathcal{M}$ is a linearly independent set of elements of $\mathcal{M}$ whose span is $\mathcal{M}$.

Theorem Every basis of a vector space $\mathcal{M}$ contains the same number of elements.

This number is called the dimension of $\mathcal{M}$, dim$(\mathcal{M})$. This number is also called the rank of $\mathcal{M}$, $r(\mathcal{M})$.

Example: $\mathcal{M}$ = R^n. We have the following facts:

1. dim$(R^n) = n$, $r(R^n) = n$
2. $n + 1$ vectors in an n–dimensional space must be linearly dependent.
3. n vectors in an n–dimensional space form a basis for R^n if and only if they are linearly independent.
4. n vectors in an n–dimensional space form a basis for R^n if and only if they span the space.
5. $D = \{x_1, \ldots, x_r\}$, $r < n$, cannot be a basis for R^n.

Theorem If $\{x_1, \ldots, x_r\}$ is a linearly independent set of vectors in $\mathcal{M}$, and $\dim(\mathcal{M}) = n$, $r < n$, then there exists elements $x_{r+1}, \ldots, x_n$ such that $\{x_1, \ldots, x_n\}$ is a basis for $\mathcal{M}$.

This theorem says that any linearly independent set of vectors can be extended to a basis.

Subspaces

Defn:
Let $\mathcal{M}$ be a vector space and let N be a set with $N \subset \mathcal{M}$. Then N is a subspace of $\mathcal{M}$ if and only if N is a vector space.

Theorem Let $\mathcal{M}$ be a vector space and let N be a nonempty subset of $\mathcal{M}$. If N is closed under addition and scalar multiplication, then N is a subspace of $\mathcal{M}$.

Examples of Subspaces

a) Let $\mathcal{M}$ = R^3. Choose a vector $x_0 \in R^3$, where $x_0 \neq 0$. Consider all vectors of the form αx_0, $\alpha \in R^1$.

$$\mathcal{S}(x_0) = \left\{\alpha x_0 : \alpha \in R^1\right\} \ .$$

$\mathcal{S}(x_0)$ is a subspace of R^3.

b) Choose vectors x_0 and x_1 which are linearly independent. The set

$$\left\{\alpha x_0 + \beta x_1 : \alpha, \beta \in R^1\right\}$$

is a subspace of R^3. The set above equals $\mathcal{S}(x_0, x_1)$.

Example: Let D be any set of vectors in a vector space $\mathcal{M}$. Then $\mathcal{S}(D)$ is a subspace in $\mathcal{M}$.

Let H and K be two linear subspaces. Define the sum H and K as

$$H + K = \{x + y : x \in H, y \in K\} \ .$$

Moreover, define $H \cap K = \{x : x \in H, x \in K\}$. We are led to the following theorem.

Theorem
Both $H + K$ and $H \cap K$ are linear subspaces.

proof: Homework

Defn:
Two subspaces are disjoint if $H \cap K = \mathbf{0}$, where $\mathbf{0}$ is the null vector.

Theorem
If $H \cap K = 0$ and $z \in H + K$, then the decomposition $z = x + y$ with $x \in H$ and $y \in K$ is unique.

Proof
Suppose $z = x + y$ and $z = x' + y'$. Then $x - x' \in H$ and $y - y' \in K$. Therefore, we must have $x + y = x' + y'$, which implies $x - x' = y - y'$. This in turn requires that $x - x' = y - y' = \mathbf{0}$ since $\mathbf{0}$ is the only vector common to H and K. Thus $x = x'$ and $y = y'$, and this completes the proof.

Theorem
If $H \cap K = 0$, then $r(H + K) = r(H) + r(K)$. In general, we have

$$r(H + K) = r(H) + r(K) - r(H \cap K).$$

Defn:

If N and N^c are disjoint subspaces of $\mathcal{M}$ and $\mathcal{M} = N + N^c$, then N^c is called the complement of N.

Remark: The complement is not unique. In R^2, a subspace N of dimension 1 consists of a line through the origin. A complement of N is given by any other line $N^c \neq \alpha N$ through the origin, because any two such lines span R^2.

Defn:

Suppose $\mathcal{M}$ is a vector space in R^n. Let x and y be two vectors in $\mathcal{M}$. Then x and y are said to be orthogonal, written $x \perp y$, if $x'y = 0$, where $x'y$ is the inner product between x and y, and x' denotes the transpose of x.

Two subspaces N_1 and N_2 are said to be orthogonal if for every $x \in N_1$ and $y \in N_2$ implies $x'y = 0$.

Defn:

Suppose N is a subspace of R^n. Then $\{x_1, \ldots, x_r\}$ is an orthogonal basis for N if for every $i \neq j$, ${x_i}'x_j = 0$. $\{x_1, \ldots, x_r\}$ is an orthonormal basis if in addition, ${x_i}'x_i = 1$, for $i = 1, \ldots, r$.

Note: Two orthogonal vectors are necessarily linearly independent.

Theorem (Gram-Schmidt)

Let N be a subspace of R^n with basis $\{x_1, \ldots, x_r\}$. Then there exists an orthonormal basis for N, $\{y_1, \ldots, y_r\}$ with $y_s \in \mathcal{S}(x_1, \ldots, x_s)$, $s = 1, \ldots, r$.
Explicitly, the ${y_s}'$s are given by

$$
\begin{aligned}
y_1 &= (x_1'x_1)^{-1/2}x_1 \\
w_s &= x_s - \sum_{i=1}^{s-1}({x_s}'y_i)y_i\,, \quad s = 2, \ldots, r \\
y_s &= ({w_s}'w_s)^{-1/2}w_s\,, \quad s = 2, \ldots, r
\end{aligned}
$$

Defn: (Orthogonal Complement)
Let N be a subspace of a vector space $\mathcal{M} \subset R^n$. Define

$$N^{\perp} = \{y \in \mathcal{M} : y \perp N\} .$$

$N^{\perp}$ is called the orthogonal complement of N with respect to $\mathcal{M}$. If $\mathcal{M} = R^n$, then $N^{\perp}$ is referred to as the orthogonal complement of N.

Theorem
Let $\mathcal{M}$ be a vector space and let $N^{\perp}$ be the orthogonal complement of N with respect to $\mathcal{M}$. Then $N^{\perp}$ is a subspace of $\mathcal{M}$, and if $x \in \mathcal{M}$, x can be written uniquely as $x = x_0 + x_1$, with $x_0 \in N$ and $x_1 \in N^{\perp}$. The ranks of these subspaces satisfy

$$r(\mathcal{M}) = r(N) + r(N^{\perp}) .$$

Also

$$\mathcal{M} = N + N^{\perp} = \left\{x : x = x_0 + x_1, x_0 \in N, x_1 \in N^{\perp}\right\} .$$

Matrices

A matrix can be defined as a linear transformation on vector space.

Defn:
Suppose $\mathcal{M}$ is an arbitrary vector space. A linear transformation A on a vector space $\mathcal{M}$ is a function mapping $\mathcal{M} \to \mathcal{M}$ such that

$$A(\alpha x + \beta y) = \alpha A(x) + \beta A(y)$$

For all $\alpha, \beta \in R^1$ and $x, y \in \mathcal{M}$.

Suppose A is an $n \times p$ matrix. Then each column of A is a vector in R^n. We can write

$$A = (x_1, \ldots, x_p)$$

where each $x_i \in R^n$, $i = 1, \ldots, p$.

The space spanned by the columns of A is called the column space of A, $C(A)$. That is $\mathcal{S}(A) = C(A)$. Also $r(A)$ will denote the rank of A.

Example

Suppose

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix},$$

Then

$$\begin{aligned} C(A) &= \mathcal{S}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \right\} \\ &= \left\{ \alpha \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} : \alpha, \beta \in R^1 \right\} \end{aligned}$$

Here $r(A) = 2$, since the two vectors are linearly independent.

Theorem
$r(A) = r(A')$.

Defn:
Suppose A is an $n \times n$ square matrix with ijth element a_{ij}. The trace of A is defined as

$$tr(A) = \sum_{i=1}^{n} a_{ii}$$

Theorem

$$tr(A + B) = tr(A) + tr(B).$$

Note: The trace is invariant under cyclic permutations. Suppose A, B, C are $n \times n$ square matrices. Then

$$tr(ABC) = tr(BCA) = tr(CAB)\ .$$

Defn:
Suppose A is an $n \times n$ square matrix. Then A is said to be nonsingular if there exists a matrix A^{-1} such that $A^{-1}A = AA^{-1} = I$ where I is an *identity matrix*. If no such matrix exists, then A is said to be singular.

Theorem
An $n \times n$ matrix A is nonsingular if and only if $r(A) = n$, i.e., columns of A form a basis for R^n. Thus A is nonsingular if and only if all of its columns are linearly independent.

If A is singular, then there exists a nonzero vector x such that $Ax = 0$, $x \in R^n$.

Defn:
The set of all x such that $Ax = 0$ is a vector space and is called the null space of A, written $\mathcal{N}(A)$.

Theorem
Suppose A is $n \times n$. If $r(A) = r$, then $r(\mathcal{N}(A)) = n - r$.

Defn:

Suppose A is an $n \times n$ square matrix. An eigenvector of A is any nonzero vector x satisfying

$$Ax = \lambda x \ , \quad \lambda \in R^1 \ .$$

λ is called an eigenvalue of A.

Note: Eigenvectors are not unique. To see this, note that $A(cx) = cAx = c\lambda x = \lambda(cx)$, so that cx is an eigenvector of A corresponding to λ.

Theorem

If x_1 and x_2 are eigenvectors with the same eigenvalue, then any nonzero linear combination of x_1 and x_2 is also an eigenvector with the same eigenvalue.

Theorem

λ is an eigenvalue of A if and only if $A - \lambda I$ is singular.

Some facts about eigenvalues and eigenvectors

1. If λ is a non-zero eigenvalue of A, then the eigenvectors corresponding to λ form a subspace of $C(A)$.

2. If A is symmetric, and λ and γ are distinct eigenvalues, then the eigenvectors corresponding to λ and γ are orthogonal. These eigenvectors forms a basis for a subspace of $C(A)$.

3. A is nonsingular if and only if all of its eigenvalues are nonzero.

4. **Theorem**

   If A is a symmetric matrix, then there exists a basis for $C(A)$ consisting of eigenvectors of nonzero eigenvalues. (The eigenvectors corresponding to the nonzero eigenvalues of A are a basis for $C(A)$.)

5. If A is $n \times n$ and symmetric, and all of its eigenvalues are nonzero, then $C(A) = R^n$, and the eigenvectors of A are a basis for R^n.

6. If A is $n \times n$ and symmetric, and some of its eigenvalues are 0, then the eigenvectors corresponding to the nonzero eigenvalues are a basis for $C(A) \subset R^n$. Thus, $r(A) =$ number of nonzero eigenvalues of A.

7. If A is $n \times n$ and symmetric, then the eigenvectors corresponding to the 0 eigenvalues (if any) are a basis for $\mathcal{N}(A)$.

8. If A is symmetric, then $\mathcal{N}(A) = C(A)^{\perp}$. That is, the null space of A corresponds to the orthogonal complement of A.

9. **Theorem**

   Suppose A is $n \times n$ and symmetric. Then there exists eigenvectors of A that are an orthogonal basis for $C(A)$. If A is nonsingular, then they are an orthogonal basis for R^n. If we normalize these eigenvectors, they are an orthonormal basis.

Theorem
Suppose A is $n \times n$ symmetric of rank $r \leq n$. Then

i) $\mathcal{N}(A) = C(A)^{\perp}$.

ii) $C(A) \cap \mathcal{N}(A) = 0$.

iii) $C(A) + \mathcal{N}(A) = R^n$.

iv) $r(A) = r, \;\; r(\mathcal{N}(A)) = n - r$.

The eigenvalues of a matrix A are found by finding the solutions of the equation for λ:

$$\det(A - \lambda I) = 0$$

where $\det(A)$ denotes the determinant of A.

Determinants:

- A: $n \times n$ matrix
- A scalar function of A defined as the sum of all $n!$ possible products of n elements s.t.

$$\begin{aligned} \det(A) = |A| &= \sum (-1)^{f(i_1,\ldots,i_n)} a_{1i_1} a_{2i_2} \cdots a_{ni_n} \\ &= \sum (-1)^{f(i_1,\ldots,i_n)} a_{i_1 1} a_{i_2 2} \cdots a_{i_n n} \end{aligned}$$

  where the summation is taken over all permutations, $(i_1, \ldots, i_n)$ of the set of integers $(1, \ldots, n)$, and $f(i_1, \ldots, i_n)$ equals the number of transpositions necessary to change $(i_1, \ldots, i_n)$ to $(1, \ldots, n)$.

- Ex:

$$A = \begin{pmatrix} 1 & -2 \\ -3 & 5 \end{pmatrix}. \text{ Then, } \det(A)?$$

Suppose A is $n \times n$ with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then

1. $\det(A) = \prod_{i=1}^{n} \lambda_i$.
2. If A is singular, then $\det(A) = 0$.
3. If A is nonsingular, (i.e., all eigenvalues nonzero), then A^{-1} exists and the eigenvalues are given by $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$.
4. The eigenvalues of A' are the same as those of A.
5. $tr(A) = \sum_{i=1}^{n} \lambda_i$ and $tr(A^{-1}) = \sum_{i=1}^{n} \lambda_i^{-1}$.
6. If A is symmetric, then $tr(A^r) = \sum_{i=1}^{n} \lambda_i^r$ for any integer r.

Defn:
A square matrix P is said to be orthogonal if $P' = P^{-1}$. If P is an orthogonal matrix, so is P'. Thus a square matrix is orthogonal if $PP' = P'P = I$.

Theorem
The product of two orthogonal matrices is orthogonal.

Proof:
If P_1 and P_2 are orthogonal matrices, then

$$(P_1P_2)(P_1P_2)' = P_1P_2P_2^{'}P_1{}' = P_1IP_1{}' = P_1P_1{}' = I \ .$$

Theorem
An $n \times n$ matrix P is orthogonal if and only if the columns of P form an orthonormal basis for R^n.

Defn:
Suppose A is an $n \times p$ matrix. Then

$$C(A) = \{z : Ax = z, x \in R^p\}$$

Theorem
Suppose A is an $n \times p$ matrix. Then

$$\mathcal{N}(A) = C(A')^{\perp} \ .$$

Proof

1) $\Rightarrow$

We first show that $\mathcal{N}(A) \subset C(A')^{\perp}$. Let $x \in \mathcal{N}(A)$. Then $Ax = 0$. $Ax = 0 \Rightarrow x'A^{'} = 0'$. Now write $A' = (a_1, \ldots, a_n)$, where a_j is $p \times 1$ and is the jth column of A'. Also write $0' = (0, \ldots, 0)$, so that $x'A' = (x'a_1, \ldots, x'a_n) = (0, \ldots, 0)$. This implies that $x'a_j = 0$ for each $j = 1, \ldots, n$. Now if $z \in C(A')$, then we can write $z = \alpha_1 a_1 + \ldots + \alpha_n a_n$. Therefore, $x'z = x'(\alpha_1 a_1 + \ldots + \alpha_n a_n) = \alpha_1 x'a_1 + \ldots + \alpha_n x'a_n = 0 + \ldots + 0 = 0$. Thus $x \in C(A')^{\perp}$.

2) $\Leftarrow$

Now we must show $C(A')^\perp \subset \mathcal{N}(A)$.

Let $x \in C(A')^\perp$. Then for any vector $z \in C(A')$, $x'z = 0$. Now $C(A') = \mathcal{S}(a_1, \ldots, a_n) = \{z : z = \alpha_1 a_1 + \ldots + \alpha_n a_n\}$. Therefore, $x'z = 0 \Rightarrow \alpha_1 x' a_1 + \ldots + \alpha_n x' a_n = 0$ for all $(\alpha_1, \ldots, \alpha_n)$, and $z \in C(A')$. This implies that $x' a_j = 0$ for all $j = 1, \ldots, n, \Rightarrow x'(a_1, \ldots, a_n) = (0, \ldots, 0), \Rightarrow Ax = 0$, and thus $x \in \mathcal{N}(A)$. Thus $C(A')^\perp \subset \mathcal{N}(A)$. This completes the proof.

<u>Miscellaneous Results</u>

1. Suppose A is an $n \times p$ matrix of rank r. Then $r \leq \min(n, p)$.

2. Suppose A is an $n \times p$ matrix of rank r and B is a $p \times s$ matrix with $r(B) \geq r$ and $AB \neq 0$. Then $C(A) = C(AB)$, and thus $r(AB) = r(A) = r$. If $r(B) < r$, then $C(AB) \subset C(A)$, and therefore $r(AB) \leq r(A)$.

3. If B is a square nonsingular matrix, then $C(A) = C(AB)$.

4. In general, $C(AB) \subset C(A)$ and $\mathcal{N}(B) \subset \mathcal{N}(AB)$.

Matrix Decompositions

Dealing with matrices is generally made easier by decomposing the matrix into a product of matrices, each of which is relatively easy to work with, and has some special structure of interest.

<u>Spectral Decomposition</u>

The spectral decomposition allows the representation of any <u>symmetric</u> matrix in terms of an orthogonal matrix and a diagonal matrix of eigenvalues.

<u>Theorem</u> (Spectral Theorem)
Suppose A is an $n \times n$ symmetric matrix. Then there exists an orthogonal matrix P such that

$$A = P \Lambda P' ,$$

where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ is an $n \times n$ diagonal matrix of the eigenvalues of A with $\lambda_1 \leq \lambda_2 \ldots \leq \lambda_n$. Here, P is the orthogonal matrix of orthonormal eigenvectors corresponding to the eigenvalues of A.

Theorem
The rank of a symmetric matrix is the number of nonzero eigenvalues.

Defn:
A symmetric $n \times n$ matrix A is positive definite if $x'Ax > 0$ for all $x \neq 0$, $x \in R^n$.

Defn:
A symmetric $n \times n$ matrix A is positive semidefinite if $x'Ax \geq 0$ for all $x \neq 0$, $x \in R^n$.

Theorem
The eigenvalues of a positive definite matrix are all positive, and the eigenvalues of a positive semidefinite matrix are all nonnegative.

Theorem
A is positive semidefinite if and only if there exists a matrix Q such that $r(A) = r(Q)$ and $A = QQ'$.

Corollary
A is positive definite if and only if there exists a nonsingular matrix Q such that $A = QQ'$.

We construct Q as follows: We know that $A = P\Lambda P'$, and define $\Lambda^{1/2} = \text{diag}(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2})$. Therefore, $A = P\Lambda^{1/2}\Lambda^{1/2}P' = (P\Lambda^{1/2})(P\Lambda^{1/2})' = QQ'$. Clearly Q is nonsingular since P is orthogonal and Λ is nonsingular.

Note: Covariance matrices are positive semidefinite.

Theorem
If A is positive definite and $A = P\Lambda P'$, then

$$A^{-1} = P\Lambda^{-1}P'$$

where $\Lambda^{-1} = \text{diag}(\lambda_1^{-1}, \ldots, \lambda_n^{-1})$.

Theorem
Suppose A is $m \times n$. Then $A'A$ and AA' are positive semidefinite.

Theorem
Suppose A is an $n \times n$ symmetric matrix. Let a_{ii} denote the i^{th} diagonal element of A. Then A is positive definite if and only if

1. $a_{ii} > 0$, for all $i = 1, \ldots, n$.

2. The determinant of every leading submatrix is positive. That is

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} > 0 \,, \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} > 0 \,, \ldots \,,$$

$$\det(A) > 0.$$

3. A is positive semidefinite if we replace > 0 in 1) and 2) above by ≥ 0.

Example

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

A is positive definite since

1. $a_{ii} > 0$, $i = 1, \ldots, 4$.

2.
$$\det \begin{pmatrix} 2 & -1 \\ -1 & 4 \end{pmatrix} = 8 - 1 = 7 > 0 \, ,$$
$$\det \begin{pmatrix} 2 & -1 & 1 \\ -1 & 4 & 0 \\ 1 & 0 & 1 \end{pmatrix} > 0 \, , \det(A) > 0.$$

Theorem
If A is a symmetric matrix, then there exists a basis for $C(A)$ consisting of eigenvectors of nonzero eigenvalues.

Theorem
If A is an $n \times n$ positive semidefinite matrix with non-zero eigenvalues $\lambda_1, \ldots, \lambda_r$, then there exists an $n \times r$ matrix $Q = Q_1 Q_2^{-1}$ such that Q_1 is an $n \times r$ matrix with orthonormal columns and $C(A) = C(Q_1)$ and Q_2 is a diagonal matrix with positive diagonal elements, and $Q'AQ = I$.

Proof
Let $v_1, \ldots, v_r$ be an orthonormal basis of eigenvectors corresponding to $\lambda_1, \ldots, \lambda_r$. Let $Q_1 = (v_1, \ldots, v_r)$. Then Q_1 has orthonormal columns and the columns of Q_1 are an orthonormal basis for $C(A)$. Therefore, $C(A) = C(Q_1)$. Furthermore, $Q_1' A Q_1 = \Lambda$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_r)$. Therefore $A = Q_1 \Lambda Q_1'$.
Now take $Q_2 = \Lambda^{1/2}$, $\Lambda^{1/2} = \text{diag}(\lambda_1^{1/2}, \ldots, \lambda_r^{1/2})$ and $Q = Q_1 Q_2^{-1}$. Clearly, Q_2 is diagonal with positive elements. Moreover,

$$\begin{aligned} Q'AQ &= (Q_1 Q_2^{-1})' A (Q_1 Q_2^{-1}) \\ &= (Q_1 \Lambda^{-1/2})' (Q_1 \Lambda Q_1') (Q_1 \Lambda^{-1/2}) \\ &= \Lambda^{-1/2} Q_1' Q_1 \Lambda Q_1' Q_1 \Lambda^{-1/2} \\ &= \Lambda^{-1/2} I \Lambda I \Lambda^{-1/2} = \Lambda^{1/2} \Lambda^{-1/2} = I \end{aligned}$$

Partitioned Matrices

Example: Let the 4×5 matrix A be partitioned as

$$A = \left(\begin{array}{rrr|rr} 7 & 2 & 5 & 8 & 4 \\ -3 & 4 & 0 & 2 & 7 \\ \hline 9 & 3 & 6 & 5 & -2 \\ 3 & 1 & 2 & 1 & 6 \end{array}\right) = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where

$$A_{11} = \begin{pmatrix} 7 & 2 & 5 \\ -3 & 4 & 5 \end{pmatrix}, A_{12} = \begin{pmatrix} 8 & 4 \\ 2 & 7 \end{pmatrix},$$
$$A_{21} = \begin{pmatrix} 9 & 3 & 6 \\ 3 & 1 & 2 \end{pmatrix}, A_{22} = \begin{pmatrix} 5 & -2 \\ 1 & 6 \end{pmatrix}.$$

Product of two conformal matrices:

If A and B are partitioned so that the sub matrices are conformal,

$$\begin{aligned} AB &= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \\ &= \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}. \end{aligned}$$

If B is replaced by a vector b,

$$Ab = (A_1, A_2) \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = A_1 b_1 + A_2 b_2.$$

This can be extended to individual columns of A and individual elements of b:

$$Ab = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_p) \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix} = b_1\mathbf{a}_1 + b_2\mathbf{a}_2 + \cdots + b_p\mathbf{a}_p.$$

Note: Ab is expressible as a linear combination of the columns of A and the coefficients are elements of b.

<u>Example</u>: Let

$$A = \begin{pmatrix} 6 & -2 & 3 \\ 2 & 1 & 0 \\ 4 & 3 & 2 \end{pmatrix}, b = \begin{pmatrix} 4 \\ 2 \\ -1 \end{pmatrix}.$$

Then

$$Ab = \begin{pmatrix} 17 \\ 10 \\ 20 \end{pmatrix}.$$

Using a linear combination of columns of A,

$$\begin{aligned} Ab &= b_1\mathbf{a}_1 + b_2\mathbf{a}_2 + b_3\mathbf{a}_3 \\ &= 4\begin{pmatrix} 6 \\ 2 \\ 4 \end{pmatrix} + 2\begin{pmatrix} -2 \\ 1 \\ 3 \end{pmatrix} - \begin{pmatrix} 3 \\ 0 \\ 2 \end{pmatrix} \\ &= \begin{pmatrix} 17 \\ 10 \\ 20 \end{pmatrix}. \end{aligned}$$

Note:

1. The columns of AB are linear combinations of the columns of A.
2. The coefficients for the jth column of AB are the elements of the jth column of B.
3. The rows of AB are linear combinations of the rows of B.
4. The coefficients for the ith row of AB are the elements of the ith row of A.

Singular Value Decomposition (SVD)

Theorem
Suppose A is an $n \times p$ matrix of rank r, ($r \leq \min(n, p)$). There exists orthogonal matrices $U_{p\times p}$ and $V_{n\times n}$ such that

$$V'AU = \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix}$$

where $\Delta = \text{diag}(\delta_1, \ldots, \delta_r)$ is an $r \times r$ diagonal matrix with $\delta_1 \geq \delta_2 \ldots \geq \delta_r > 0$. The δ_i's are called the singular values of A.

Implications of SVD

1) $A = VDU'$, where $D = \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix}$.

2) Split $V = (V_1, V_2)$ and $U = (U_1, U_2)$, where V_1 is $n \times r$, V_2 is $n \times (n - r)$, U_1 is $p \times r$ and U_2 is $p \times (p - r)$. Then

$$\begin{aligned} A = VDU' &= (V_1, V_2) \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} U_1' \\ U_2' \end{pmatrix} \\ &= V_1 \Delta U_1' + V_2 0 U_2' \\ &= V_1 \Delta U_1' \end{aligned}$$

This implies that

$$\begin{aligned} C(A) &= \{z : z = Ax = V_1 \Delta U_1' x, x \in R^p\} \\ &= \{z : z = V_1 x^*, x^* \in R^r\} \\ &= C(V_1) \end{aligned}$$

Note that since Δ and U_1' has rank r, we can set $x^* = \Delta U_1' x$ and get the equality above. Thus the columns of V_1 span the same space as the columns of A. Since V is an orthonormal basis for R^n, the columns of V_1 are an orthonormal basis for $C(A)$.

3) Similar arguments show that

$$C(A') = C(U_1)$$

4) Claim: $\mathcal{N}(A) = C(U_2) = C(A')^{\perp}$.

Proof

Since $A = V_1 \Delta U_1' + V_2 0 U_2'$, we have $AU_2 = V_1 \Delta U_1' U_2 + V_2 0 U_2' U_2$, and thus $AU_2 = 0 + V_2 0 = 0$. Therefore $AU_2 = 0$ which implies that $\mathcal{N}(A) = C(U_2)$.

5) Similar arguments show that $\mathcal{N}(A') = C(V_2) = C(V_1)^{\perp}$.

We have the following summary:

Matrix	Column space	Null space
A	$C(V_1)$	$C(U_2)$
A'	$C(U_1)$	$C(V_2)$

6) The columns of V_1 are the orthonormal eigenvectors corresponding to the nonzero eigenvalues of AA', and the columns of U_1 are the orthonormal eigenvectors corresponding to the nonzero eigenvalues of $A'A$.

7) If $r(A) = r$, then $\delta_1^2, \ldots, \delta_r^2$ are the eigenvalues of $A'A$.

Q-R factorization

Suppose A is an $n \times p$ matrix with linearly independent columns. Then A can be written uniquely in the form:

$$A = QR$$

where $Q_{n\times p}$ has orthonormal columns and $R_{p\times p}$ is an upper triangular matrix with positive diagonal elements.

Proof

We construct Q and R by using the Gram-Schmidt orthogonalization process. Write $A = (a_1, \ldots, a_p)$, where a_j is the jth column of A, given by

$$a_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix}_{n\times 1}$$

Now apply Gram-Schmidt to yield an orthogonal basis $(u_1, \ldots, u_p)$ for $C(A)$, such that

$$\mathcal{S}(a_1, \ldots, a_k) = \mathcal{S}(u_1, \ldots, u_k), \;\; k \leq p \; .$$

Note that u_k is a linear combination of $(a_1, \ldots, a_k)$, $k = 1, \ldots, p$. Let $U_{n\times p} = (u_1, \ldots, u_p)$, where u_k is an $n \times 1$ vector corresponding to the kth column of U and is given by

$$u_k = a_k - \sum_{i=1}^{k-1} \left(\frac{a_k' u_i}{u_i' u_i} \right) u_i \; ,$$

based on Gram-Schmidt procedure. Since $u_k = \mathcal{S}(a_1, \ldots, a_k)$, this means we can write

$$u_k = s_{1k}a_1 + \ldots + s_{kk}a_k$$

for some set of scalars $s_{1k}, \ldots, s_{kk}$. Now define the upper triangular matrix S as

$$S = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1p} \\ 0 & s_{22} & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & s_{pp} \end{pmatrix} .$$

Now in matrix notation we can write the Gram-Schmidt representation of U as

$$\begin{aligned} U &= (u_1, \dots, u_p) \\ &= (s_{11}a_1, \dots, s_{1k}a_1 + \dots + s_{kk}a_k, \dots, s_{1p}a_1 + \dots + s_{pp}a_p) \\ &= AS \end{aligned}$$

By definition of Gram-Schmidt, the diagonal elements of S are all 1. Now we normalize the columns of U by postmultiplying U by D, where

$$D = \text{diag}((u_1'u_1)^{-1/2}, \dots, (u_p'u_p)^{-1/2}).$$

Thus UD now has orthonormal columns and therefore

$$ASD = UD = Q .$$

Since D and S are nonsingular,

$$ASD = Q \Rightarrow A = Q(SD)^{-1}.$$

Note $(SD)^{-1}$ is an upper triangular matrix with positive diagonal elements. Therefore, we define $R = (SD)^{-1}$, and thus $A = QR$, where Q has orthonormal columns and R is upper triangular with positive diagonal elements.

Vector Spaces

The type of vector spaces we consider are *finite dimensional real vector spaces*.

Defn:
A real vector space $\mathcal{M}$ is a set of elements (called *vectors*) with the following properties:

A. Addition axioms:

For $x, y \in \mathcal{M}$, there corresponds a vector $x + y \in \mathcal{M}$ called the sum of x and y such that:

A1. $x + y = y + x$ (commutative)

A2. $x + (y + z) = (x + y) + z$ (associative)

A3. There exists a unique vector 0, the null vector, such that for all $x \in \mathcal{M}$, $x + 0 = x$

A4. for all $x \in \mathcal{M}$, there exists a unique element $-x$ such that $x + (-x) = 0$

B. Scalar multiplication axioms:

For any real number α and for any $x \in \mathcal{M}$, there exists a member of $\mathcal{M}$, αx, called product of α and x such that

B1. $\alpha(x + y) = \alpha x + \alpha y$ (distributive)

B2. $(\alpha + \beta)x = \alpha x + \beta x$ (distributive)

B3. $\alpha(\beta x) = (\alpha\beta)x$ (associative)

B4. There exists a scalr ξ such that $\xi \cdot x = x$ ($\xi = 1$).

Examples of Vector Spaces

Example 1 $\mathcal{M}$ = (n dimensional Euclidean space), R^n.

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$x + y = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix} \quad \alpha x = \begin{pmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{pmatrix} \quad .$$

Example 2 – The space of all 2×2 matrices with real elements is a vector space. $\mathcal{M} =$ space of a 2×2 matrices.

x is a 2×2 matrix.

$$0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$x = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad y = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}$$

$$x + y = \begin{pmatrix} 6 & 8 \\ 10 & 12 \end{pmatrix}$$

and so on.

Example 3 – The polynomials of degree n with real coefficients constitutes a vector space. $\mathcal{M} =$ polynomials of degree n.

$x = \beta_0 + \beta_1 t + \ldots + \beta_n t^n.$

$y = \gamma_0 + \gamma_1 t + \ldots + \gamma_n t^n.$

Defn: A set of vectors $D = \{x_1, \ldots, x_r\}$ is called linearly dependent if there is a set of scalars $\alpha_1, \ldots, \alpha_r$, not all zero, such that

$$\sum_{i=1}^{r} \alpha_i x_i = 0.$$

If $\sum_{i=1}^{r} \alpha_i x_i = 0 \Rightarrow \alpha_i = 0,\ i = 1, \ldots, r$, then $D = \{x_1, \ldots, x_r\}$ are linearly independent.

If D is linearly independent, then $D_1 \subset D$ is linearly independent.

Example 1: $\mathcal{M} = R^3$.

$$x_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad x_2 = \begin{pmatrix} 16 \\ 12 \\ 3 \end{pmatrix} \quad x_3 = \begin{pmatrix} 0 \\ 28 \\ 3 \end{pmatrix}$$

$D = \{x_1, x_2, x_3\}$ is linearly dependent since $16x_1 - x_2 + x_3 = 0$.

Example 2: Consider the linear model $y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$, $\beta_i \neq 0$, $i = 1, 2, 3$, and the $x_i^{'}$s are given by Example 1. What is the meaning of linear dependence for this linear model?

Since, for the example given above $x_2 - x_3 = 16x_1$, or $x_1 = (x_2 - x_3)/16$, by substituting for x_1, we write:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

$$= x_2(\beta_2 + \beta_1/16) + x_3(\beta_3 - \beta_1/16) + \varepsilon$$

$$= \gamma_1 x_2 + \gamma_2 x_3 + \varepsilon$$

where the $\gamma_i^{'}$s are defined by the above equation. This shows that the model with three parameters is equivalent to a model with only two parameters.

Example 3: $\mathcal{M} = R^n$

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad \dots \quad e_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

$D = \{e_1, \dots, e_n\}$ is linearly independent.

Suppose $D = \{x_1, \dots, x_r\}$ is linearly dependent. This means that

$\sum_{i=1}^{r} \alpha_i x_i = 0$, and not all $\alpha_i = 0$.

Let k be such that $\alpha_k \neq 0$. Then

$$x_k = \sum_{\substack{i=1 \\ i \neq k}}^{r} \left(-\frac{\alpha_i}{\alpha_k} \right) x_i.$$

We are led to the following theorem.

Theorem A set of vectors is linearly dependent if and only if some vector of the set can be written as a linear combination of the others. That is there exists a k such that

$$x_k = \sum_{\substack{i=1 \\ i \neq k}}^{r} \left(-\frac{\alpha_i}{\alpha_k} \right) x_i,$$

where $\alpha_k \neq 0$.

Proof: Exercise.

Defn:

A basis in a vector space $\mathcal{M}$ is a set of linearly independent vectors such that every $x \in \mathcal{M}$ is a linear combination of vectors in the set.

Example: $\mathcal{M} = R^3$

The vectors

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

are a basis for R^3.

Note:

1. Every vector in R^3 can be written as a linear combination of the vectors above.
2. Bases are not unique.

The vectors

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 3 \\ 3 \\ 0 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

are also a basis for R^3.

Note: The representation of a vector as a linear combination of basis elements is unique. The coefficients of the vectors are often called coordinates with respect to that basis.

To illustrate, suppose $B = \{x_1, \ldots, x_r\}$ is a basis and $z = \sum_{i=1}^{r} \alpha_i x_i$.

Then, $z = \sum_{i=1}^{r} \beta_i x_i \Rightarrow \sum_{i=1}^{r} (\alpha_i - \beta_i) x_i = 0 \Rightarrow \alpha_i = \beta_i$, $i = 1, \ldots, n$.

In this example $(\alpha_1, \ldots, \alpha_r)$ are the coordinates of z with respect to B.

Defn:

A vector space $\mathcal{M}$ is said to be finite dimensional if it has a basis with a finite number of elements.

Examples

- R^n is finite dimensional
- P^n is finite dimensional (polynomials of degree n).

Let D be a set of vectors.

1. The set of all possible linear combinations of elements of D is a vector space, called the span of D, $\mathcal{S}(D)$.
2. A basis for $\mathcal{M}$ is a linearly independent set of elements of $\mathcal{M}$ whose span is $\mathcal{M}$.

Theorem Every basis of a vector space $\mathcal{M}$ contains the same number of elements.

This number is called the dimension of $\mathcal{M}$, dim$(\mathcal{M})$. This number is also called the rank of $\mathcal{M}$, $r(\mathcal{M})$.

Example: $\mathcal{M}$ = R^n. We have the following facts:

1. dim$(R^n) = n$, $r(R^n) = n$
2. $n + 1$ vectors in an n–dimensional space must be linearly dependent.
3. n vectors in an n–dimensional space form a basis for R^n if and only if they are linearly independent.
4. n vectors in an n–dimensional space form a basis for R^n if and only if they span the space.
5. $D = \{x_1, \ldots, x_r\}$, $r < n$, cannot be a basis for R^n.

Theorem If $\{x_1, \ldots, x_r\}$ is a linearly independent set of vectors in $\mathcal{M}$, and $\dim(\mathcal{M}) = n$, $r < n$, then there exists elements $x_{r+1}, \ldots, x_n$ such that $\{x_1, \ldots, x_n\}$ is a basis for $\mathcal{M}$.

This theorem says that any linearly independent set of vectors can be extended to a basis.

Subspaces

Defn:
Let $\mathcal{M}$ be a vector space and let N be a set with $N \subset \mathcal{M}$. Then N is a subspace of $\mathcal{M}$ if and only if N is a vector space.

Theorem Let $\mathcal{M}$ be a vector space and let N be a nonempty subset of $\mathcal{M}$. If N is closed under addition and scalar multiplication, then N is a subspace of $\mathcal{M}$.

Examples of Subspaces

a) Let $\mathcal{M}$ = R^3. Choose a vector $x_0 \in R^3$, where $x_0 \neq 0$. Consider all vectors of the form αx_0, $\alpha \in R^1$.

$$\mathcal{S}(x_0) = \left\{\alpha x_0 : \alpha \in R^1\right\} \ .$$

$\mathcal{S}(x_0)$ is a subspace of R^3.

b) Choose vectors x_0 and x_1 which are linearly independent. The set

$$\left\{\alpha x_0 + \beta x_1 : \alpha, \beta \in R^1\right\}$$

is a subspace of R^3. The set above equals $\mathcal{S}(x_0, x_1)$.

<u>**Example:**</u> Let D be any set of vectors in a vector space $\mathcal{M}$. Then $\mathcal{S}(D)$ is a subspace in $\mathcal{M}$.

Let H and K be two linear subspaces. Define the sum H and K as

$$H + K = \{x + y : x \in H, y \in K\} \ .$$

Moreover, define $H \cap K = \{x : x \in H, x \in K\}$. We are led to the following theorem.

<u>**Theorem**</u>
Both $H + K$ and $H \cap K$ are linear subspaces.

proof: Homework

<u>**Defn:**</u>
Two subspaces are disjoint if $H \cap K = \mathbf{0}$, where $\mathbf{0}$ is the null vector.

<u>**Theorem**</u>
If $H \cap K = 0$ and $z \in H + K$, then the decomposition $z = x + y$ with $x \in H$ and $y \in K$ is unique.

<u>Proof</u>
Suppose $z = x + y$ and $z = x' + y'$. Then $x - x' \in H$ and $y - y' \in K$. Therefore, we must have $x + y = x' + y'$, which implies $x - x' = y - y'$. This in turn requires that $x - x' = y - y' = \mathbf{0}$ since $\mathbf{0}$ is the only vector common to H and K. Thus $x = x'$ and $y = y'$, and this completes the proof.

<u>**Theorem**</u>
If $H \cap K = 0$, then $r(H + K) = r(H) + r(K)$. In general, we have

$$r(H + K) = r(H) + r(K) - r(H \cap K).$$

Defn:
If N and N^c are disjoint subspaces of $\mathcal{M}$ and $\mathcal{M} = N + N^c$, then N^c is called the complement of N.

Remark: The complement is not unique. In R^2, a subspace N of dimension 1 consists of a line through the origin. A complement of N is given by any other line $N^c \neq \alpha N$ through the origin, because any two such lines span R^2.

Defn:
Suppose $\mathcal{M}$ is a vector space in R^n. Let x and y be two vectors in $\mathcal{M}$. Then x and y are said to be orthogonal, written $x \perp y$, if $x'y = 0$, where $x'y$ is the inner product between x and y, and x' denotes the transpose of x.

Two subspaces N_1 and N_2 are said to be orthogonal if for every $x \in N_1$ and $y \in N_2$ implies $x'y = 0$.

Defn:
Suppose N is a subspace of R^n. Then $\{x_1, \ldots, x_r\}$ is an orthogonal basis for N if for every $i \neq j$, ${x_i}'x_j = 0$. $\{x_1, \ldots, x_r\}$ is an orthonormal basis if in addition, ${x_i}'x_i = 1$, for $i = 1, \ldots, r$.

Note: Two orthogonal vectors are necessarily linearly independent.

Theorem (Gram-Schmidt)

Let N be a subspace of R^n with basis $\{x_1, \ldots, x_r\}$. Then there exists an orthonormal basis for N, $\{y_1, \ldots, y_r\}$ with $y_s \in \mathcal{S}(x_1, \ldots, x_s)$, $s = 1, \ldots, r$.
Explicitly, the ${y_s}'$s are given by

$$\begin{aligned}
y_1 &= (x_1'x_1)^{-1/2}x_1 \\
w_s &= x_s - \sum_{i=1}^{s-1}({x_s}'y_i)y_i\,, \quad s = 2, \ldots, r \\
y_s &= ({w_s}'w_s)^{-1/2}w_s\,, \quad s = 2, \ldots, r
\end{aligned}$$

Defn: (Orthogonal Complement)
Let N be a subspace of a vector space $\mathcal{M} \subset R^n$. Define

$$N^\perp = \{y \in \mathcal{M} : y \perp N\} .$$

$N^\perp$ is called the orthogonal complement of N with respect to $\mathcal{M}$. If $\mathcal{M} = R^n$, then $N^\perp$ is referred to as the orthogonal complement of N.

Theorem
Let $\mathcal{M}$ be a vector space and let $N^\perp$ be the orthogonal complement of N with respect to $\mathcal{M}$. Then $N^\perp$ is a subspace of $\mathcal{M}$, and if $x \in \mathcal{M}$, x can be written uniquely as $x = x_0 + x_1$, with $x_0 \in N$ and $x_1 \in N^\perp$. The ranks of these subspaces satisfy

$$r(\mathcal{M}) = r(N) + r(N^\perp) .$$

Also

$$\mathcal{M} = N + N^\perp = \left\{x : x = x_0 + x_1, x_0 \in N, x_1 \in N^\perp\right\} .$$

Matrices

A matrix can be defined as a linear transformation on vector space.

Defn:
Suppose $\mathcal{M}$ is an arbitrary vector space. A linear transformation A on a vector space $\mathcal{M}$ is a function mapping $\mathcal{M} \to \mathcal{M}$ such that

$$A(\alpha x + \beta y) = \alpha A(x) + \beta A(y)$$

For all $\alpha, \beta \in R^1$ and $x, y \in \mathcal{M}$.

Suppose A is an $n \times p$ matrix. Then each column of A is a vector in R^n. We can write

$$A = (x_1, \ldots, x_p)$$

where each $x_i \in R^n$, $i = 1, \ldots, p$.

The space spanned by the columns of A is called the column space of A, $C(A)$. That is $\mathcal{S}(A) = C(A)$. Also $r(A)$ will denote the rank of A.

Example

Suppose

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} ,$$

Then

$$\begin{aligned} C(A) &= \mathcal{S}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} , \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \right\} \\ &= \left\{ \alpha \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} : \alpha, \beta \in R^1 \right\} \end{aligned}$$

Here $r(A) = 2$, since the two vectors are linearly independent.

Theorem
$r(A) = r(A')$.

Defn:
Suppose A is an $n \times n$ square matrix with ijth element a_{ij}. The trace of A is defined as

$$tr(A) = \sum_{i=1}^{n} a_{ii}$$

Theorem

$$tr(A + B) = tr(A) + tr(B).$$

Note: The trace is invariant under cyclic permutations. Suppose A, B, C are $n \times n$ square matrices. Then

$$tr(ABC) = tr(BCA) = tr(CAB)\ .$$

Defn:
Suppose A is an $n \times n$ square matrix. Then A is said to be nonsingular if there exists a matrix A^{-1} such that $A^{-1}A = AA^{-1} = I$ where I is an *identity matrix*. If no such matrix exists, then A is said to be singular.

Theorem
An $n \times n$ matrix A is nonsingular if and only if $r(A) = n$, i.e., columns of A form a basis for R^n. Thus A is nonsingular if and only if all of its columns are linearly independent.

If A is singular, then there exists a nonzero vector x such that $Ax = 0$, $x \in R^n$.

Defn:
The set of all x such that $Ax = 0$ is a vector space and is called the null space of A, written $\mathcal{N}(A)$.

Theorem
Suppose A is $n \times n$. If $r(A) = r$, then $r(\mathcal{N}(A)) = n - r$.

Defn:
Suppose A is an $n \times n$ square matrix. An eigenvector of A is any nonzero vector x satisfying

$$Ax = \lambda x \ , \quad \lambda \in R^1 \ .$$

λ is called an eigenvalue of A.

Note: Eigenvectors are not unique. To see this, note that $A(cx) = cAx = c\lambda x = \lambda(cx)$, so that cx is an eigenvector of A corresponding to λ.

Theorem
If x_1 and x_2 are eigenvectors with the same eigenvalue, then any nonzero linear combination of x_1 and x_2 is also an eigenvector with the same eigenvalue.

Theorem
λ is an eigenvalue of A if and only if $A - \lambda I$ is singular.

Some facts about eigenvalues and eigenvectors

1. If λ is a non-zero eigenvalue of A, then the eigenvectors corresponding to λ form a subspace of $C(A)$.

2. If A is symmetric, and λ and γ are distinct eigenvalues, then the eigenvectors corresponding to λ and γ are orthogonal. These eigenvectors forms a basis for a subspace of $C(A)$.

3. A is nonsingular if and only if all of its eigenvalues are nonzero.

4. **Theorem**

   If A is a symmetric matrix, then there exists a basis for $C(A)$ consisting of eigenvectors of nonzero eigenvalues. (The eigenvectors corresponding to the nonzero eigenvalues of A are a basis for $C(A)$.)

5. If A is $n \times n$ and symmetric, and all of its eigenvalues are nonzero, then $C(A) = R^n$, and the eigenvectors of A are a basis for R^n.

6. If A is $n \times n$ and symmetric, and some of its eigenvalues are 0, then the eigenvectors corresponding to the nonzero eigenvalues are a basis for $C(A) \subset R^n$. Thus, $r(A) =$ number of nonzero eigenvalues of A.

7. If A is $n \times n$ and symmetric, then the eigenvectors corresponding to the 0 eigenvalues (if any) are a basis for $\mathcal{N}(A)$.

8. If A is symmetric, then $\mathcal{N}(A) = C(A)^{\perp}$. That is, the null space of A corresponds to the orthogonal complement of A.

9. **Theorem**

   Suppose A is $n \times n$ and symmetric. Then there exists eigenvectors of A that are an orthogonal basis for $C(A)$. If A is nonsingular, then they are an orthogonal basis for R^n. If we normalize these eigenvectors, they are an orthonormal basis.

Theorem
Suppose A is $n \times n$ symmetric of rank $r \leq n$. Then

i) $\mathcal{N}(A) = C(A)^{\perp}$.

ii) $C(A) \cap \mathcal{N}(A) = 0$.

iii) $C(A) + \mathcal{N}(A) = R^n$.

iv) $r(A) = r, \;\; r(\mathcal{N}(A)) = n - r$.

The eigenvalues of a matrix A are found by finding the solutions of the equation for λ:

$$\det(A - \lambda I) = 0$$

where $\det(A)$ denotes the determinant of A.

Determinants:

- A: $n \times n$ matrix
- A scalar function of A defined as the sum of all $n!$ possible products of n elements s.t.

$$\begin{aligned}\det(A) = |A| &= \sum (-1)^{f(i_1,\ldots,i_n)} a_{1i_1} a_{2i_2} \cdots a_{ni_n} \\ &= \sum (-1)^{f(i_1,\ldots,i_n)} a_{i_1 1} a_{i_2 2} \cdots a_{i_n n}\end{aligned}$$

  where the summation is taken over all permutations, $(i_1, \ldots, i_n)$ of the set of integers $(1, \ldots, n)$, and $f(i_1, \ldots, i_n)$ equals the number of transpositions necessary to change $(i_1, \ldots, i_n)$ to $(1, \ldots, n)$.

- Ex:

$$A = \begin{pmatrix} 1 & -2 \\ -3 & 5 \end{pmatrix}. \text{ Then, } \det(A)?$$

Suppose A is $n \times n$ with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then

1. $\det(A) = \prod_{i=1}^n \lambda_i$.
2. If A is singular, then $\det(A) = 0$.
3. If A is nonsingular, (i.e., all eigenvalues nonzero), then A^{-1} exists and the eigenvalues are given by $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$.
4. The eigenvalues of A' are the same as those of A.
5. $tr(A) = \sum_{i=1}^n \lambda_i$ and $tr(A^{-1}) = \sum_{i=1}^n \lambda_i^{-1}$.
6. If A is symmetric, then $tr(A^r) = \sum_{i=1}^n \lambda_i^r$ for any integer r.

Defn:
A square matrix P is said to be orthogonal if $P' = P^{-1}$. If P is an orthogonal matrix, so is P'. Thus a square matrix is orthogonal if $PP' = P'P = I$.

Theorem
The product of two orthogonal matrices is orthogonal.

Proof:
If P_1 and P_2 are orthogonal matrices, then

$$(P_1P_2)(P_1P_2)' = P_1P_2P_2^{'}P_1{}' = P_1IP_1{}' = P_1P_1{}' = I \; .$$

Theorem
An $n \times n$ matrix P is orthogonal if and only if the columns of P form an orthonormal basis for R^n.

Defn:
Suppose A is an $n \times p$ matrix. Then

$$C(A) = \{z : Ax = z, x \in R^p\}$$

Theorem
Suppose A is an $n \times p$ matrix. Then

$$\mathcal{N}(A) = C(A')^{\perp} \; .$$

Proof

1) $\Rightarrow$

We first show that $\mathcal{N}(A) \subset C(A')^{\perp}$. Let $x \in \mathcal{N}(A)$. Then $Ax = 0$. $Ax = 0 \Rightarrow x'A^{'} = 0'$. Now write $A' = (a_1, \ldots, a_n)$, where a_j is $p \times 1$ and is the jth column of A'. Also write $0' = (0, \ldots, 0)$, so that $x'A' = (x'a_1, \ldots, x'a_n) = (0, \ldots, 0)$. This implies that $x'a_j = 0$ for each $j = 1, \ldots, n$. Now if $z \in C(A')$, then we can write $z = \alpha_1 a_1 + \ldots + \alpha_n a_n$. Therefore, $x'z = x'(\alpha_1 a_1 + \ldots + \alpha_n a_n) = \alpha_1 x'a_1 + \ldots + \alpha_n x'a_n = 0 + \ldots + 0 = 0$. Thus $x \in C(A')^{\perp}$.

2) $\Leftarrow$

Now we must show $C(A')^{\perp} \subset \mathcal{N}(A)$.

Let $x \in C(A')^{\perp}$. Then for any vector $z \in C(A')$, $x'z = 0$. Now $C(A') = \mathcal{S}(a_1, \ldots, a_n) = \{z : z = \alpha_1 a_1 + \ldots + \alpha_n a_n\}$. Therefore, $x'z = 0 \Rightarrow \alpha_1 x' a_1 + \ldots + \alpha_n x' a_n = 0$ for all $(\alpha_1, \ldots, \alpha_n)$, and $z \in C(A')$. This implies that $x' a_j = 0$ for all $j = 1, \ldots, n$, $\Rightarrow x'(a_1, \ldots, a_n) = (0, \ldots, 0)$, $\Rightarrow Ax = 0$, and thus $x \in \mathcal{N}(A)$. Thus $C(A')^{\perp} \subset \mathcal{N}(A)$. This completes the proof.

Miscellaneous Results

1. Suppose A is an $n \times p$ matrix of rank r. Then $r \leq \min(n, p)$.

2. Suppose A is an $n \times p$ matrix of rank r and B is a $p \times s$ matrix with $r(B) \geq r$ and $AB \neq 0$. Then $C(A) = C(AB)$, and thus $r(AB) = r(A) = r$. If $r(B) < r$, then $C(AB) \subset C(A)$, and therefore $r(AB) \leq r(A)$.

3. If B is a square nonsingular matrix, then $C(A) = C(AB)$.

4. In general, $C(AB) \subset C(A)$ and $\mathcal{N}(B) \subset \mathcal{N}(AB)$.

Matrix Decompositions

Dealing with matrices is generally made easier by decomposing the matrix into a product of matrices, each of which is relatively easy to work with, and has some special structure of interest.

Spectral Decomposition

The spectral decomposition allows the representation of any symmetric matrix in terms of an orthogonal matrix and a diagonal matrix of eigenvalues.

Theorem (Spectral Theorem)
Suppose A is an $n \times n$ symmetric matrix. Then there exists an orthogonal matrix P such that

$$A = P \Lambda P' ,$$

where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ is an $n \times n$ diagonal matrix of the eigenvalues of A with $\lambda_1 \leq \lambda_2 \ldots \leq \lambda_n$. Here, P is the orthogonal matrix of orthonormal eigenvectors corresponding to the eigenvalues of A.

Theorem

The rank of a symmetric matrix is the number of nonzero eigenvalues.

Defn:

A symmetric $n \times n$ matrix A is positive definite if $x'Ax > 0$ for all $x \neq 0, x \in R^n$.

Defn:

A symmetric $n \times n$ matrix A is positive semidefinite if $x'Ax \geq 0$ for all $x \neq 0$, $x \in R^n$.

Theorem

The eigenvalues of a positive definite matrix are all positive, and the eigenvalues of a positive semidefinite matrix are all nonnegative.

Theorem

A is positive semidefinite if and only if there exists a matrix Q such that $r(A) = r(Q)$ and $A = QQ'$.

Corollary

A is positive definite if and only if there exists a nonsingular matrix Q such that $A = QQ'$.

We construct Q as follows: We know that $A = P\Lambda P'$, and define $\Lambda^{1/2} = \text{diag}(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2})$. Therefore, $A = P\Lambda^{1/2}\Lambda^{1/2}P' = (P\Lambda^{1/2})(P\Lambda^{1/2})' = QQ'$. Clearly Q is nonsingular since P is orthogonal and Λ is nonsingular.

Note: Covariance matrices are positive semidefinite.

Theorem
If A is positive definite and $A = P\Lambda P'$, then

$$A^{-1} = P\Lambda^{-1}P'$$

where $\Lambda^{-1} = \text{diag}(\lambda_1^{-1}, \ldots, \lambda_n^{-1})$.

Theorem
Suppose A is $m \times n$. Then $A'A$ and AA' are positive semidefinite.

Theorem
Suppose A is an $n \times n$ symmetric matrix. Let a_{ii} denote the i^{th} diagonal element of A. Then A is positive definite if and only if

1. $a_{ii} > 0$, for all $i = 1, \ldots, n$.

2. The determinant of every leading submatrix is positive. That is

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} > 0\,, \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} > 0\,, \ldots\,,$$

$$\det(A) > 0.$$

3. A is positive semidefinite if we replace > 0 in 1) and 2) above by ≥ 0.

Example

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

A is positive definite since

1. $a_{ii} > 0$, $i = 1, \ldots, 4$.

2.
$$\det \begin{pmatrix} 2 & -1 \\ -1 & 4 \end{pmatrix} = 8 - 1 = 7 > 0 \,,$$

$$\det \begin{pmatrix} 2 & -1 & 1 \\ -1 & 4 & 0 \\ 1 & 0 & 1 \end{pmatrix} > 0 \,, \det(A) > 0.$$

Theorem
If A is a symmetric matrix, then there exists a basis for $C(A)$ consisting of eigenvectors of nonzero eigenvalues.

Theorem
If A is an $n \times n$ positive semidefinite matrix with non-zero eigenvalues $\lambda_1, \ldots, \lambda_r$, then there exists an $n \times r$ matrix $Q = Q_1 Q_2^{-1}$ such that Q_1 is an $n \times r$ matrix with orthonormal columns and $C(A) = C(Q_1)$ and Q_2 is a diagonal matrix with positive diagonal elements, and $Q'AQ = I$.

Proof
Let $v_1, \ldots, v_r$ be an orthonormal basis of eigenvectors corresponding to $\lambda_1, \ldots, \lambda_r$. Let $Q_1 = (v_1, \ldots, v_r)$. Then Q_1 has orthonormal columns and the columns of Q_1 are an orthonormal basis for $C(A)$. Therefore, $C(A) = C(Q_1)$. Furthermore, $Q_1' A Q_1 = \Lambda$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_r)$. Therefore $A = Q_1 \Lambda Q_1'$.
Now take $Q_2 = \Lambda^{1/2}$, $\Lambda^{1/2} = \text{diag}(\lambda_1^{1/2}, \ldots, \lambda_r^{1/2})$ and $Q = Q_1 Q_2^{-1}$. Clearly, Q_2 is diagonal with positive elements. Moreover,

$$\begin{aligned} Q'AQ &= (Q_1 Q_2^{-1})' A (Q_1 Q_2^{-1}) \\ &= (Q_1 \Lambda^{-1/2})' (Q_1 \Lambda Q_1') (Q_1 \Lambda^{-1/2}) \\ &= \Lambda^{-1/2} Q_1' Q_1 \Lambda Q_1' Q_1 \Lambda^{-1/2} \\ &= \Lambda^{-1/2} I \Lambda I \Lambda^{-1/2} = \Lambda^{1/2} \Lambda^{-1/2} = I \end{aligned}$$

Partitioned Matrices

Example: Let the 4×5 matrix A be partitioned as

$$A = \left(\begin{array}{rrr|rr} 7 & 2 & 5 & 8 & 4 \\ -3 & 4 & 0 & 2 & 7 \\ \hline 9 & 3 & 6 & 5 & -2 \\ 3 & 1 & 2 & 1 & 6 \end{array} \right) = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where

$$A_{11} = \begin{pmatrix} 7 & 2 & 5 \\ -3 & 4 & 5 \end{pmatrix}, A_{12} = \begin{pmatrix} 8 & 4 \\ 2 & 7 \end{pmatrix},$$
$$A_{21} = \begin{pmatrix} 9 & 3 & 6 \\ 3 & 1 & 2 \end{pmatrix}, A_{22} = \begin{pmatrix} 5 & -2 \\ 1 & 6 \end{pmatrix}.$$

Product of two conformal matrices:

If A and B are partitioned so that the sub matrices are conformal,

$$\begin{aligned} AB &= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \\ &= \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}. \end{aligned}$$

If B is replaced by a vector b,

$$Ab = (A_1, A_2) \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = A_1 b_1 + A_2 b_2.$$

This can be extended to individual columns of A and individual elements of b:

$$Ab = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_p) \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix} = b_1\mathbf{a}_1 + b_2\mathbf{a}_2 + \cdots + b_p\mathbf{a}_p.$$

Note: Ab is expressible as a linear combination of the columns of A and the coefficients are elements of b.

Example: Let

$$A = \begin{pmatrix} 6 & -2 & 3 \\ 2 & 1 & 0 \\ 4 & 3 & 2 \end{pmatrix}, b = \begin{pmatrix} 4 \\ 2 \\ -1 \end{pmatrix}.$$

Then

$$Ab = \begin{pmatrix} 17 \\ 10 \\ 20 \end{pmatrix}.$$

Using a linear combination of columns of A,

$$\begin{aligned} Ab &= b_1\mathbf{a}_1 + b_2\mathbf{a}_2 + b_3\mathbf{a}_3 \\ &= 4\begin{pmatrix} 6 \\ 2 \\ 4 \end{pmatrix} + 2\begin{pmatrix} -2 \\ 1 \\ 3 \end{pmatrix} - \begin{pmatrix} 3 \\ 0 \\ 2 \end{pmatrix} \\ &= \begin{pmatrix} 17 \\ 10 \\ 20 \end{pmatrix}. \end{aligned}$$

Note:

1. The columns of AB are linear combinations of the columns of A.
2. The coefficients for the jth column of AB are the elements of the jth column of B.
3. The rows of AB are linear combinations of the rows of B.
4. The coefficients for the ith row of AB are the elements of the ith row of A.

Singular Value Decomposition (SVD)

Theorem

Suppose A is an $n \times p$ matrix of rank r, ($r \leq \min(n, p)$). There exists orthogonal matrices $U_{p \times p}$ and $V_{n \times n}$ such that

$$V'AU = \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix}$$

where $\Delta = \text{diag}(\delta_1, \ldots, \delta_r)$ is an $r \times r$ diagonal matrix with $\delta_1 \geq \delta_2 \ldots \geq \delta_r > 0$. The δ_i's are called the singular values of A.

Implications of SVD

1) $A = VDU'$, where $D = \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix}$.

2) Split $V = (V_1, V_2)$ and $U = (U_1, U_2)$, where V_1 is $n \times r$, V_2 is $n \times (n - r)$, U_1 is $p \times r$ and U_2 is $p \times (p - r)$. Then

$$\begin{aligned} A = VDU' &= (V_1, V_2) \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} U_1' \\ U_2' \end{pmatrix} \\ &= V_1 \Delta U_1' + V_2 0 U_2' \\ &= V_1 \Delta U_1' \end{aligned}$$

This implies that

$$\begin{aligned} C(A) &= \{z : z = Ax = V_1 \Delta U_1' x, x \in R^p\} \\ &= \{z : z = V_1 x^*, x^* \in R^r\} \\ &= C(V_1) \end{aligned}$$

Note that since Δ and U_1' has rank r, we can set $x^* = \Delta U_1' x$ and get the equality above. Thus the columns of V_1 span the same space as the columns of A. Since V is an orthonormal basis for R^n, the columns of V_1 are an orthonormal basis for $C(A)$.

3) Similar arguments show that

$$C(A') = C(U_1)$$

4) Claim: $\mathcal{N}(A) = C(U_2) = C(A')^{\perp}$.

Proof

Since $A = V_1 \Delta U_1' + V_2 0 U_2'$, we have $AU_2 = V_1 \Delta U_1' U_2 + V_2 0 U_2' U_2$, and thus $AU_2 = 0 + V_2 0 = 0$. Therefore $AU_2 = 0$ which implies that $\mathcal{N}(A) = C(U_2)$.

5) Similar arguments show that $\mathcal{N}(A') = C(V_2) = C(V_1)^{\perp}$.

We have the following summary:

Matrix	Column space	Null space
A	$C(V_1)$	$C(U_2)$
A'	$C(U_1)$	$C(V_2)$

6) The columns of V_1 are the orthonormal eigenvectors corresponding to the nonzero eigenvalues of AA', and the columns of U_1 are the orthonormal eigenvectors corresponding to the nonzero eigenvalues of $A'A$.

7) If $r(A) = r$, then $\delta_1^2, \ldots, \delta_r^2$ are the eigenvalues of $A'A$.

Q-R factorization

Suppose A is an $n \times p$ matrix with linearly independent columns. Then A can be written uniquely in the form:

$$A = QR$$

where $Q_{n \times p}$ has orthonormal columns and $R_{p \times p}$ is an upper triangular matrix with positive diagonal elements.

Proof

We construct Q and R by using the Gram-Schmidt orthogonalization process. Write $A = (a_1, \ldots, a_p)$, where a_j is the jth column of A, given by

$$a_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix}_{n \times 1}$$

Now apply Gram-Schmidt to yield an orthogonal basis $(u_1, \ldots, u_p)$ for $C(A)$, such that

$$\mathcal{S}(a_1, \ldots, a_k) = \mathcal{S}(u_1, \ldots, u_k), \;\; k \leq p\ .$$

Note that u_k is a linear combination of $(a_1, \ldots, a_k)$, $k = 1, \ldots, p$. Let $U_{n \times p} = (u_1, \ldots, u_p)$, where u_k is an $n \times 1$ vector corresponding to the kth column of U and is given by

$$u_k = a_k - \sum_{i=1}^{k-1} \left(\frac{a_k' u_i}{u_i' u_i} \right) u_i\ ,$$

based on Gram-Schmidt procedure. Since $u_k = \mathcal{S}(a_1, \ldots, a_k)$, this means we can write

$$u_k = s_{1k} a_1 + \ldots + s_{kk} a_k$$

for some set of scalars $s_{1k}, \ldots, s_{kk}$. Now define the upper triangular matrix S as

$$S = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1p} \\ 0 & s_{22} & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & s_{pp} \end{pmatrix} .$$

Now in matrix notation we can write the Gram-Schmidt representation of U as

$$\begin{aligned} U &= (u_1, \dots, u_p) \\ &= (s_{11}a_1, \dots, s_{1k}a_1 + \dots + s_{kk}a_k, \dots, s_{1p}a_1 + \dots + s_{pp}a_p) \\ &= AS \end{aligned}$$

By definition of Gram-Schmidt, the diagonal elements of S are all 1. Now we normalize the columns of U by postmultiplying U by D, where

$$D = \text{diag}((u_1' u_1)^{-1/2}, \dots, (u_p' u_p)^{-1/2}).$$

Thus UD now has orthonormal columns and therefore

$$ASD = UD = Q .$$

Since D and S are nonsingular,

$$ASD = Q \Rightarrow A = Q(SD)^{-1}.$$

Note $(SD)^{-1}$ is an upper triangular matrix with positive diagonal elements. Therefore, we define $R = (SD)^{-1}$, and thus $A = QR$, where Q has orthonormal columns and R is upper triangular with positive diagonal elements.

Chapter 3

INTRODUCTION

Linear models have a dominant role in statistical theory and practice. Most standard statistical methods are special cases of the general linear model, and rely on the corresponding theory for justification.

The goal of this course is to develop the theoretical basis for analyses based on a linear model. We shall be concerned with laying the theoretical foundation for simple as well as complex data sets.

Linear model is one of the oldest topics in the statistics curriculum. The main role of linear model in statistical practice, however, has begun to undergo a fundamental change due in large measure to available computing. Balanced experiments were often required to make analysis possible. This has produced a fundamental change in the way we can think about linear models, as much less stress can be placed on the special cases where computations are easy and more can be placed on general ideas. Topics that might have been standard, such as the recovery of interblock information in an incomplete block experiment, is of much less interest when computers can be used to appropriately maximize functions.

However, standard results are so elegant, and so interesting, that they deserve study in their own right, and for that reason we will study the traditional body of material that makes up linear models, including many standard simple models as well as a general approach.

The goal of these notes is to develop a *coordinate-free approach* to linear models. Coordinates can often sever to make problems unnecessarily complex, and understanding the features of a problems that are not dependent on coordinates is extremely valuable. The problems introduced by parameters are more easily understood given the coordinate-free background.

Example 3.1 **Example 1** – *Simple Linear Regression*

- *Modeling the relationship between two variables* (x, y).
- *We have observations* $(x_1, y_1), \ldots, (x_n, y_n)$, *where*
  - *the* x_i*'s are known fixed values (independent or predictor variables), and*
  - *the* y_i*'s are dependent or response variables (random).*
- *We have the model:*

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\ , \ \ i = 1, \ldots, n.$$

- ε*: random error representing random fluctuations, measurement errors, or the effect of factors outside of our control*
  - y*: annual melanoma mortality in a state,* x*: latitude of the center of the state (in degrees).*
  - y*: income,* x*: number of years of education*
  - y*: height,* x*: weight of people*
  - y*: response,* x*: dose of a drug*

- *Usual assumptions:*

  - *The ε_i's are $i.i.d.$ from some distribution with*

$$E(\varepsilon_i) = 0 \text{ and } Var(\varepsilon_i) = \sigma^2.$$

- *With this model, we may want to*

  1. *Estimate β_0, β_1, and σ^2.*
  2. *Test hypotheses about β_1, confidence limits for β_1.*
  3. *Predict a future y at a given x.*

- *To do exact inference, we need ε_i to be normally distributed. Otherwise, we have to base inferences on large sample theory.*

- *Estimation and inferential procedures for the simple linear regression model are developed and illustrated later in detail.*

Example 2 – One Way Analysis-of-Varinace (ANOVA)

- Interested in comparing several populations or several conditions in a study

- Example: We want to examine the effect of NO_2 on the lungs. Consider mice which are i) not exposed ii) mildly exposed, iii) heavily exposed. ($k = 3$). Response variable is percent serum fluorescence. High readings indicate damage to lung tissues.

- Suppose we have observations y_{ij}, $j = 1, \ldots, n_i$, $i = 1, \ldots, k$. We have k populations with n_i observations in population i.

pop1	pop2	...	popk
y_{11}	y_{21}	$\cdots$	y_{k1}
$\vdots$	$\vdots$		$\vdots$
y_{1n_1}	y_{2n_2}	$\cdots$	y_{kn_k}

- Assume $E(y_{ij}) = \mu_i$ and $\mathrm{Var}(y_{ij}) = \sigma^2$.

- We can write:

$$y_{ij} = \mu_i + \varepsilon_{ij} \text{ , where } j = 1, \ldots, n_i, i = 1, \ldots, k.$$

- The ε_{ij}'s are $i.i.d.$ from a distribution with $E(\varepsilon_{ij}) = 0$, and Var$(\varepsilon_{ij}) = \sigma^2$.

- The model $y_{ij} = \mu_i + \varepsilon_{ij}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, k$ is often referred to as a means model.

- We may want to

  1. Estimate the μ_i's and σ^2.
  2. Test hypotheses about the μ_i's, e.g.

     $$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$

     or

     $$H_0 : c_1\mu_1 + c_2\mu_2 + \ldots + c_k\mu_k = 0 \ ,$$

     where $\sum_{i=1}^{k} c_i = 0$.
  3. Decide which group mean is largest, multiple comparisons, etc . . ..

- If we reparameterize and write

$$\mu_i = \mu + \alpha_i,$$

then the model becomes

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \text{ , } j = 1, \ldots, n_i, i = 1, \ldots, k.$$

- ANOVA models: special case of regression models but will be treated separately later in the course.

Example 3 - General Linear Model

- Both of the examples just presented and many others are special cases of the general linear model.

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i,\ i = 1, \ldots, n,$$

  where

  - y_i's are observations;
  - x_{ij}'s are known (fixed) values;
  - β_j's are unknown parameters;
  - ε_i's are unobservable random variables with

$$E(\varepsilon_i) = 0,\ \text{Cov}(\varepsilon_i, \varepsilon_j) = \sigma_{ij},\ i, j = 1, \ldots, n.$$

- This is the most general setup.

- **Note**: If an intercept is included in the model, then

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i.$$

- Often, more restrictive assumptions are made about the ε_i's.

  1. Uncorrelated: Cov$(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$.
  2. Equal variances: Cov$(\varepsilon_i, \varepsilon_i) =$ var$(\varepsilon_i) = \sigma^2$, $i = 1, \ldots, n$.
  3. the ε_i's are normally distributed.

- Example:

  - y_i = oxygen consumption

- x_1 = treadmill duration
- x_2 = heart rate
- x_3 = age
- x_4 = height

- We can write this model in matrix notation. Define

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}_{n \times 1} \qquad X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}_{n \times p}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}_{p \times 1} \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}_{n \times 1}$$

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ & \sigma_{22} & & \vdots \\ & & \ddots & \vdots \\ & & & \sigma_{nn} \end{pmatrix}_{n \times n} .$$

- We can now write the model as

$$Y = X\beta + \varepsilon \ ,$$

$$E(\varepsilon) = 0, \ \ \textbf{Cov}(\varepsilon) = \Sigma \ ,$$

or often

$\text{Cov}(\varepsilon) = \sigma^2 I$, in the uncorrelated, equal variance errors case.

- **Note:** In this formulation, the model says that

$$\begin{aligned} \mu &= E(Y) = X\beta \\ &= [X_1, X_2, \ldots, X_p]\beta \\ &= \beta_1 X_1 + \ldots + \beta_p X_p \ , \end{aligned}$$

  where

$$X_j = \begin{pmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{pmatrix}$$

  is the jth column of X, $j = 1, \ldots, p$.

  We see that μ is a linear combination of the columns of X.

- In a more abstract setting,

$$\mu \in \Omega,$$

  where Ω is a linear subspace in an n dimensional vector space.

- In the general linear model, we may be interested in

  1. **Prediction**: overall influence of the x variables on y
  2. **Data Description** or **Explanation**: use of the estimated model to summarize or describe the observed data
  3. **Parameter Estimation**: estimates of the model parameters are essential for prediction and data description or explanation.

4. **Variable Selection** or **Screening**: determining the importance of each predictor variable in modeling the variation in y
5. **Control of Output**: Under a cause-and-effect relationship between x and y, use the estimated model to control the output of a process.

- Estimation and inferential procedures are discussed later in detail.

- The one–way ANOVA model introduced earlier:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij},\ j = 1, \ldots, n_i, i = 1, \ldots, k,$$

can now be written in matrix notation as

$$Y = X\beta + \varepsilon, \varepsilon \sim N_N(0, \sigma^2 I),$$

where $N = \sum_{i=1}^{k} n_i$,

$$Y = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{kn_k} \end{pmatrix}_{N \times 1}, \qquad \varepsilon = \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \\ \vdots \\ \varepsilon_{kn_k} \end{pmatrix}_{N \times 1}$$

$$X = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & 1 & 0 & \dots & \vdots \\ \vdots & 0 & 1 & \dots & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & 1 & \dots & 0 \\ \vdots & \vdots & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix}_{N\times(k+1)} \qquad \beta = \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix}_{(k+1)\times 1}$$

- For example if $k = 3$, $n_1 = 3$, $n_2 = 1$, $n_3 = 2$, $N = 3+1+2 = 6$

$$Y = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{31} \\ y_{32} \end{pmatrix}_{6\times 1} \qquad X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}_{6\times 4}$$

$$\beta = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}_{4\times 1} \qquad \varepsilon = \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{31} \\ \varepsilon_{32} \end{pmatrix}_{6\times 1} \qquad .$$

CHAPTER 2: MATRIX ALGEBRA

Useful resource: Schott (1997), Christensen (2011, Appendices A and B)

Matrix, Vector and Scalar

- *Matrix*:
  - A rectangular or square array of numbers of variables;
  - Real numbers or variables for all elements of matrices
  - $$\mathbf{A} = (a_{ij}) = \begin{pmatrix} 65 & 154 \\ 73 & 182 \\ 68 & 167 \end{pmatrix} = \begin{pmatrix} 65 & 154 \\ 73 & 182 \\ 68 & 167 \end{pmatrix} \qquad (3.1)$$

    where the subscript i and j denote the row and column, respectively.
  - (2.1) has 3 rows and 2 columns; (the size of) $\mathbf{A}$ is 3×2
- *Vector*:
  - A matrix with a single row or column
  - Often use a single subscript to identify elements in a vector
  - $$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \ \mathbf{x}' = (x_1, x_2, x_3) = (x_1 \ x_2 \ x_3)$$

    where $\mathbf{x}$ is a column vector and $\mathbf{x}'$ is a row vector (will define the prime symbol ($'$), *transpose* shortly)
  - Geometric interpretation: a vector with p elements can be thought as a point in a p-dimensional space

- *Scalar*:
  - A single real number
  - 2.5, -9, and 7.26
  - Technically distinct from 1×1 matrix
- **Matrix Equality**:
  - same size and elements in corresponding positions are equal
  - $$\begin{pmatrix} 3 & -2 & 4 \\ 1 & 3 & 7 \end{pmatrix} = \begin{pmatrix} 3 & -2 & 4 \\ 1 & 3 & 7 \end{pmatrix}.$$
- **Transpose**:
  - Interchanging the rows and columns of a matrix $\mathbf{A}$;
  - <u>Ex</u>:
    $$\mathbf{A} = \begin{pmatrix} 6 & -2 \\ 4 & 7 \\ 1 & 3 \end{pmatrix}, \mathbf{A}' = \begin{pmatrix} 6 & 4 & 1 \\ -2 & 7 & 3 \end{pmatrix}.$$
  - If $\mathbf{A} = (a_{ij})$ then $\mathbf{A}^{'} = (a_{ij})^{'} = (a_{ji})$.
  - What about the dimension of a matrix $\mathbf{A}$?
  - What if we transpose $\mathbf{A}$ twice?
- **Symmetric Matrix**:
  - If $\mathbf{A}^{'} = \mathbf{A}$;
  - <u>Ex</u>:
    $$\mathbf{A} = \begin{pmatrix} 3 & 2 & 6 \\ 2 & 10 & -7 \\ 6 & -7 & 9 \end{pmatrix}, \quad \text{symmetric?}$$
  - All symmetric matrices are *square* (# of rows = # of columns);

- **Diagonal Matrix**:
  - The *diagonal* of a $p \times p$ square matrix $\mathbf{A} = (a_{ij})$: $a_{11}, a_{22}, \ldots, a_{pp}$;
  - If 0s for all off-diagonal positions;
  - Ex:

$$\mathbf{D} = \begin{pmatrix} 8 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}.$$

- **Identity Matrix**:
  - A diagonal matrix with a 1 in each diagonal position
  - Ex:

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

- **Upper Triangular Matrix**:
  - A square matrix with 0s below the diagonal
  - Ex:

$$\mathbf{T} = \begin{pmatrix} 7 & 2 & -3 & 5 \\ 0 & 0 & -2 & 6 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 8 \end{pmatrix}.$$

  - What is a **Lower Triangular Matrix**?

- **Other Special Matrices**:
  - $\mathbf{J}$: a square matrix of 1s;

    Ex:
    $$\mathbf{J} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$
  - $\mathbf{O}$: a square matrix of 0s;

    Ex:
    $$\mathbf{O} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$
  - $\mathbf{j}$: a vector of 1s, $\mathbf{0}$: a vector of 0's;

Operations

- *Sum* of Two Matrices (or Two Vectors):
  - *conformal*: two matrices (or vectors) are of the same size for addition;
  - Adding corresponding elements;
  - If $\mathbf{A} = (a_{ij})$: $n \times p$ and $\mathbf{B} = (b_{ij})$: $n \times p$ then

    $\mathbf{C} = \mathbf{A} + \mathbf{B} = (c_{ij}) = (a_{ij} + b_{ij})$: $n \times p$
  - Ex:
    $$\begin{pmatrix} 7 & -3 & 4 \\ 2 & 8 & -5 \end{pmatrix} + \begin{pmatrix} 11 & 5 & -6 \\ 3 & 4 & 2 \end{pmatrix} = ? \qquad (3.2)$$
  - What about the difference between $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{D} = \mathbf{A} - \mathbf{B}$?

Theorem If $\mathbf{A}$ and $\mathbf{B}$ are both $n \times m$, then

(i) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$ (commutative).

(ii) $(\mathbf{A} + \mathbf{B})^{'} = \mathbf{A}^{'} + \mathbf{B}^{'}$.

– proof?

- *Product* of a Scalar and a Matrix:
  - Any scalar can be multiplied by any matrix;
  - Ex:

$$c\mathbf{A} = (ca_{ij}) = \begin{pmatrix} ca_{11} & ca_{12} & \cdots & ca_{1m} \\ ca_{21} & ca_{22} & \cdots & ca_{2m} \\ \vdots & \vdots & & \vdots \\ ca_{n1} & ca_{n2} & \cdots & ca_{nm} \end{pmatrix}$$

Theorem The product of a scalar and a matrix is commutative:

$$c\mathbf{A} = \mathbf{A}c.$$

- *Product* of Two Matrices (or Two Vectors):
  - To define $\mathbf{AB}$, $\mathbf{A}$ and $\mathbf{B}$ should be *conformal for multiplication*;

$$\text{\# of column in } \mathbf{A} = \text{\# of rows in } \mathbf{B}$$

  - The (ij)th element of $\mathbf{C} = \mathbf{AB}$ is defined as

$$c_{ij} = \sum_k a_{ik} b_{kj}$$

  - We multiply every row of $\mathbf{A}$ by every column of $\mathbf{B}$.
  - If $\mathbf{A}$ is $n \times m$ and $\mathbf{B}$ is $m \times p$ then $\mathbf{C}$ is $n \times p$.