We will present the basic elements of vector space theory needed for the development of material in the text. For a more in depth treatment, the reader is encouraged to turn to Halmos (1958) or Eaton (1983).

**Definition A.1.** A collection of vectors $V$ is a real vector space if the following conditions hold: for any pair $x$ and $y$ of vectors in $V$ there corresponds a vector $x + y$ and scalars $\alpha, \beta \in \mathbb{R}$ such that:

(a) $(x + y) + z = x + (y + z)$ (vector addition is commutative)
(b) $x + y = y + x$ (vector addition is associative)
(c) there exists a unique vector $0 \in V$ (the origin) such that $x + 0 = x$ for every vector $x$
(d) for every $x \in V$, there exists a unique vector $-x$ such that $x + (-x) = 0$
(e) $\alpha(\beta x) = (\alpha\beta)x$ (multiplication by scalars is associative)
(f) $1x = x$ for every $x$
(g) $(\alpha + \beta)x = \alpha x + \beta x$ (multiplication by scalars is distributive with respect to vector addition)
(h) $\alpha(x + y) = \alpha x + \beta y$ (multiplication of vectors is distributive with respect to scalar addition).

While the theory we will discuss applies more broadly, we will be restricting our attention to vector spaces in $\mathbb{R}^n$, where $x \in \mathbb{R}^n$ has the representation

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad x_i \in \mathbb{R}.$$

All vectors will be assumed to be column vectors. A row vector will be denoted as $x^T = (x_1, \ldots, x_n)$.

**Definition A.2.** A finite set of vectors $\{x_i\}$ is linearly dependent if there is a corresponding set of scalars $\{\alpha_i\}$, not all zero, such that $\sum \alpha_i x_i = 0$. 

---

A

Linear Algebra and Vector Space Theory
If, $\sum \alpha_i x_i = 0$ implies that $\alpha_i = 0$ for each $i$, then the set $\{x_i\}$ is linearly independent.

**Definition A.3.** A basis for a vector space $\mathbb{V}$ is a linearly independent set of vectors $B \equiv \{x_i\}$ such that every vector in $\mathbb{V}$ is a linear combination of the vectors in $B$. The vector space is **finite dimensional** if there is a finite number of vectors in the basis. The **dimension** of vector space $\mathbb{V}$, denoted $\dim(\mathbb{V})$, is the number of vectors in the basis for $\mathbb{V}$.

The standard or canonical basis for $\mathbb{R}^n$ is the collection of vectors $e_i$, $i = 1, \ldots, n$ of the form

$$e_i = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

where the one occurs in the $i$th position or coordinate of $e_i$. For any $x \in \mathbb{R}^n$, it is clear that $x$ may be expressed as $\sum_i x_i e_i$. The $x_i$ are called the **coordinates** of $x$ with respect to the (canonical) basis $\{e_i\}$. If we instead use an alternative collection of vectors $B = \{b_i\}$ as a basis, then any $x$ may be represented as $x = \sum_i \alpha_i b_i$ where the coordinates of $x$ with respect to the basis $B$ are now $\alpha_i$. The coordinates of $x$ will depend on the order of the vectors in the basis.

Other basic properties of vector spaces are as follows. If $\mathbb{V}$ is a finite dimensional vector space, then

(a) Any collection of independent vectors in $\mathbb{V}$ may be extended to form a basis for $\mathbb{V}$.
(b) All bases for $\mathbb{V}$ have the same number of vectors.
(c) Every set of $n + 1$ vectors in an $n$-dimensional vector space is linearly dependent. Such a collection is **overcomplete**.
(d) If $x_1, \ldots, x_n$ is a basis for $\mathbb{V}$, and $v \in \mathbb{V}$, then the representation $v = \sum \alpha_i x_i$ is unique.

Proofs of these may be found in Halmos (1958).

In linear models, the assumption that the mean vector $\mu$ is a linear combination of some predictor variables is equivalent to saying that the mean vector belongs to a subspace of $\mathbb{R}^n$.

**Definition A.4.** Given a set of vectors $\{x_1, \ldots, x_r\} = \mathbb{X} \subset \mathbb{V}$, the span of $\mathbb{X}$, $S(\mathbb{X})$, is the subspace of $\mathbb{V}$ generated by the set of all linear combinations of the vectors in $\mathbb{X}$. If the set of vectors $\{x_1, \ldots, x_r\}$ are linearly independent, then the $\dim(\mathbb{X}) = r$ and the vectors $\{x_1, \ldots, x_r\}$ form a basis for the subspace $S(\mathbb{X})$. 
A Linear Algebra and Vector Space Theory  

$S(\mathcal{X})$ is often referred to as the subspace spanned by the set $\mathcal{X}$. When the set $\{x_i, i = 1, \ldots, r\} \subset \mathbb{R}^n$ is viewed as an $n \times r$ matrix $X$ with the vectors $x_i$ forming the columns of the matrix, we will use $C(X)$ to refer to the equivalent subspace spanned by the columns of the matrix $X$. The notation $R(X) \subset \mathbb{R}^r$ will be used to denote the subspace spanned by the rows of the matrix $X$.

**Definition A.5.** A collection of vectors $\{x_i\} = V_0 \subseteq \mathbb{V}$ is a subspace of $\mathbb{V}$ if and only if it is a vector space. That is, any nonempty set $V_0 \subseteq \mathbb{V}$ is a subspace, if it is closed under vector addition and scalar multiplication, so that for each $x, y \in V_0$ and $\alpha, \beta \in \mathbb{R}, \alpha x + \beta y$ belongs to $V_0$.

Using the fact that a subspace is a vector space, the following properties of subspaces are easy to show. If $V_0$ and $V_1$ are subspaces of $\mathbb{V}$, then

(a) The dimension of $V_0$ is less than or equal to the dimension of $\mathbb{V}$, $\dim(V_0) \leq \dim(\mathbb{V})$.

(b) If $\dim(V_0) = m$ and $\dim(V) = n$ with $m < n$, then there exists a basis $\{x_1, \ldots, x_m, \ldots, x_n\}$ for $\mathbb{V}$ such that $\{x_1, \ldots, x_m\}$ is a basis for $V_0$.

(c) The intersection of any two subspaces is again a subspace.

The span of $X$ may be viewed as the intersection of all subspaces that contain $X$, so that $S(X)$ is the smallest subspace that contains $X$. In addition to intersections, it is useful to consider unions of subspaces. If $X$ and $Y$ are two subspaces of $\mathbb{V}$, then the span of their union $Z = S(X \cup sY)$ is the set of all vectors of the form $z = x + y$ where $x \in X$ and $y \in Y$. The span of the union of $X$ and $Y$ may be represented as $X \cup Y$. This leads to the following decomposition of a vector space $\mathbb{V}$:

(a) Let $X$ and $Y$ denote subspaces of $\mathbb{V}$, with $\dim(X) = r$, $\dim(Y) = s$ and $\dim(X \cap Y) = q$. Then there exist vectors $\{z_1, \ldots, z_q\}$, $\{x_{q+1}, \ldots, x_r\}$, and $\{y_{q+1}, \ldots, y_s\}$ such that the vectors $\{z_1, \ldots, z_q\}$ form a basis for the intersection of the two subspaces $Z_i = z_1, \ldots, z_q, x_{q+1}, \ldots, x_r$ form a basis for $X$, and $\{z_1, \ldots, z_q, y_{q+1}, \ldots, y_s\}$ form a basis for $Y$. If $q = 0$ then the intersection is empty. The collection of $\{z_1, \ldots, z_q, x_{q+1}, \ldots, x_r, y_{q+1}, \ldots, y_s\}$ is a basis for the space $X \cup Y$.

(b) $\dim(X + Y) = \dim(X) + \dim(Y) - \dim(X \cap Y)$

(c) There exists a complementary subspace, $X^c$, of $\mathbb{V}$ that satisfies $X \cap X^c = \{0\}$ and $X + X^c = \mathbb{V}$.

(d) If $M$ and $E$ and are two complementary subspaces of $\mathbb{V}$, then each $y \in \mathbb{V}$ may be uniquely decomposed as $y = m + e$, where $m \in M$ and $e \in E$.

The decomposition of a vector space into two or more complementary subspaces is an integral part of model fitting, where $M$ represents the subspace that contains the mean vector, and in model selection, where $M_0$, a subspace of $M$, is the subspace that contains the mean under an alternative model specification. We will want to be able to decompose any vector into a unique part in subspace that contains the mean vector and a part that is in the residual space. This may be characterized via projections:
Definition A.6. Given complementary subspaces \( M \) and \( E \) (where \( M + E = V \) and \( M \cap E = \{0\} \)), if \( y = m + e \), with \( m \in M \) and \( e \in E \), then \( m \) is called the projection of \( y \) on \( M \) along \( E \) and \( e \) is called the projection of \( y \) on \( E \) along \( M \).

Given a basis for the complementary subspaces, one may always find \( m \) and \( e \) in terms of the coordinates of \( y \) with respect to the bases vectors for the two subspaces. It will be useful to represent the projection of \( y \) via a linear transformation of the vector \( y \). A key aspect of the coordinate free perspective is that inference does not depend on the basis that we have picked to represent \( M \) (the mean space). However, we will sometimes want to work with a particular basis or coordinate system for convenience, and need to be able to represent results with respect to some other basis, involving linear transformations of elements of the vector space.

Definition A.7. A function \( A \) defined on \( V \) which maps into \( W \) is called a linear transformation if \( A(\alpha x + \beta y) = \alpha A(x) + \beta A y \) for all \( x, y \in V \) and scalars \( \alpha, \beta \in \mathbb{R} \).

Of note, the space of all linear transformations from \( V \) to \( W \) is itself a vector space, and of course one may consider linear transformations of linear transformations! Such a vector space arises naturally when one considers random matrices, such as a matrix normal. For a more general development of multivariate models, please refer to Eaton (1983). As we will be concerned with linear transformation on \( \mathbb{R}^n \) to \( \mathbb{R}^q \), the linear transformation \( A \) may be represented as a \( n \times q \) matrix, and \( A(x) \) will be written as \( Ax \), corresponding to the usual matrix multiplication of a matrix times a vector.

Definition A.8. The \( q \times q \) matrix \( A \) is invertible or nonsingular if there exists a \( q \times q \) matrix \( A^{-1} \) such that \( A^{-1} A = I_q \), where \( I_q \) is the \( q \) dimensional identity matrix. \( A \) is invertible if and only if \( Ax = 0 \) implies that \( x \) equals 0.

Change of Basis: If \( \{x_1, \ldots, x_n\} \) is a basis for \( \mathbb{R}^n \), and \( A \) \((n \times n)\) is invertible, then \( \{Ax_1, \ldots, Ax_n\} \) is also a basis for \( V \).

As part of transforming spaces by linear transformations, we will need to understand their range and null spaces.

Definition A.9. The range or column space of \( A \) denoted as \( C(A) \) is defined as

\[
C(A) \equiv \{u | u = Av, \text{ for some } v \in V\}
\]

The null space of \( A \), \( N(A) \), is the set

\[
N(A) \equiv \{v | v \in V, Av = 0, \text{ for some } v \in V\} \subseteq V
\]

The rank of \( A \), \( r(A) \), is the dimension of the subspace \( C(A) \).
The column space of $A$ is the subspace spanned by the columns of the matrix $A$, so if the vectors $a_1, \ldots, a_q$ represent the columns of the $n \times q$ matrix $A$, then $C(A) = S(\{a_1, \ldots, a_q\})$.

We now characterize projections in terms of linear transformations.

**Definition A.10.** Let $M$ and $E$ be complementary subspaces of $V$ where $M \cap E = 0$, and $V$ may be written as the direct sum $M + E$. Then any $v \in V$ may be written uniquely as $v = m + e$, with $m \in M$ and $e \in E$. The projection on $M$ along $E$ is the unique linear transformation $P$ defined as $Pm = m$ and $Pe = 0$. $P$ may be seen as the identity transformation on $M$ and zero on $E$.

**Theorem A.11.** A linear transformation $P$ is a projection on some subspace if and only if it is idempotent, $P = P^2$.

**Theorem A.12.** If $P$ is the projection on $M$ along $E$, then $M$ and $E$ are the sets of all solutions to the equations $Pv = v$ and $Pv = 0$, respectively. Then the column space of $P$ is $M$ and $E$ is the null space of $P$.

**Theorem A.13.** A linear transformation $P$ is a projection if and only if $I - P$ is a projection. If $P$ is the projection on $M$ along $E$ then $I - P$ is the projection on $E$ along $M$.

In order to talk about orthogonality of vectors, orthogonal projections and geometric aspects of linear models, we will need to introduce inner products and inner product spaces. A vector space with an inner product is called an inner product space. For the most part we will deal with the standard Euclidean inner product which for vectors $x, y \in \mathbb{R}^n$ is $\sum x_i y_i$ and may be represented as $x^T y$, where $x^T$ is the transpose of the vector $x$ (a $1 \times n$ matrix).

If $A$ is a $n \times n$ nonsingular symmetric matrix, then $x^T Ay$ is also an inner product.

**Definition A.14.** An inner product on $V$ is a real valued function on $V \times V$, $\langle \cdot, \cdot \rangle$ which satisfies the following:

(a) symmetry $\langle x, y \rangle = \langle y, x \rangle$

(b) non-negativity $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ only if $x = 0$.

(c) linearity $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$

It is easy to verify that $x^T Ay$ is an inner product.

**Definition A.15.** The norm or length of a vector $x$ is $||x|| \equiv \langle x, x \rangle^{1/2}$ and the distance between two vectors $x$ and $y$ is the length of their difference $||x - y||$.

In $\mathbb{R}^n$ with the usual inner product, when $x$ and $y$ are both not non-zero, the cosine of the angle between them is $x^T y / ||x|| ||y||$. Thus the angle between them is $\pi/2$ radians or $90^\circ$ if and only if $x^T y = 0$. 
Definition A.16. Two vectors $x, y$ in an inner product space $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ are orthogonal if $\langle x, y \rangle = 0$ and will be denoted as $x \perp y$. Two subspaces of $\mathcal{V}$ are said to be orthogonal, $X \perp Y$, if for all $x \in X$ and all $y \in Y$, $x \perp y$.

Definition A.17. A set of vectors $\{x_1, \ldots, x_k\}$ is called orthonormal if $\langle x_i, x_j \rangle = 1$ for $i = j$ and is zero otherwise. A set $\{x_1, \ldots, x_k\}$ is an orthonormal basis (ONB) for the subspace spanned by $\{x_1, \ldots, x_k\}$ if it is both a basis and is orthonormal.

It should be clear that the canonical basis $\{e_i\}$ for $\mathbb{R}^n$ with the standard inner product is a orthonormal basis for $\mathbb{R}^n$. However, starting with an arbitrary basis, one may find an orthonormal basis using the Gram-Schmidt procedure.

Theorem A.18 (Gram-Schmidt Orthogonalization). Given an inner product space $\mathcal{V}$, $(\langle \cdot, \cdot \rangle)$, let $X$ be the subspace spanned by the linearly independent set $\{x_1, \ldots, x_k\}$, $X \subseteq \mathcal{V}$. Then $\{u_1, \ldots, u_k\}$ defined sequentially by

$$u_1 = \frac{x_1}{\|x_1\|}$$

and

$$u_{i+1} = \frac{x_{i+1} - \sum_{j=1}^{i} \langle x_{i+1}, y_j \rangle y_j}{\|x_{i+1} - \sum_{j=1}^{i} \langle x_{i+1}, y_j \rangle y_j\|}$$

for $i = 1, \ldots, k-1$ define an ONB for $X$.

Definition A.19. If $\mathcal{M}$ is a subspace of $\mathcal{V}$, the orthogonal complement of $\mathcal{M}$, denoted by $\mathcal{M}^\perp$, consists of the the set of vectors that are orthogonal to all vectors in $\mathcal{M}$: $\mathcal{M}^\perp = \{e | e \perp m, \text{ for all } m \in \mathcal{M}\}$. Furthermore,

(a) $\mathcal{M} \cap \mathcal{M}^\perp = 0$

(b) $\mathcal{M} + \mathcal{M}^\perp = \mathcal{V}$

(c) $(\mathcal{M}^\perp)^\perp = \mathcal{M}$

When $\mathcal{V}$ is an inner product space, the projection on $\mathcal{M}$ along $\mathcal{M}^\perp$ is called the orthogonal projection onto $\mathcal{M}$. Recall that if $P$ is a projection on $\mathcal{M}$ along $E$ its null space is $E$. In the case of an orthogonal projection, its null space is the orthogonal complement of $\mathcal{M}$.

In the definition of a complementary space, each vector $v \in \mathcal{V}$ could be uniquely decomposed into two vectors $v = m + e$, $m \in \mathcal{M}$ and $e$ in the complement of $\mathcal{M}$. We may compute the inner product between $m$ and $e$ as

$$\langle e, m \rangle = \langle e, Pm \rangle = \langle P'e, m \rangle$$

where $P'$ is the unique linear trasformation called the adjoint or transpose of the transformation. Suppose now that $P' = P$; the transformation is self-adjoint or symmetric and

$$\langle P'e, m \rangle = \langle Pe, m \rangle.$$
But because $e$ is in the null space of the projection $P$ (since it is the complementary space is the null space of the projection), the inner product will be zero and the vectors $e$ and $m$ are orthogonal. This leads to the following characterization of orthogonal projections:

**Theorem A.20.** In the inner product space $(V, \langle \cdot, \cdot \rangle)$, the following are equivalent:

(a) $P$ is an orthogonal projection

(b) $P = P^2 = P'$

Symmetry and orthogonality are determined by the choice of inner product. In the usual Euclidean inner product, symmetry of $P$ corresponds to $P' = P^T$, the usual matrix transpose.

**Example A.21.** Let’s consider an alternative inner product defined as

$$\langle x, y \rangle_A \equiv x^T A y$$

where $A = A^T$ is a $n \times n$ nonsingular matrix. Then the matrix

$$P \equiv X(X^T A X)^{-1} X^T A$$

where $X$ is assumed to be of full column rank is an orthogonal projection in the inner product space $(\mathbb{R}^n, \langle x, y \rangle_A)$, but is not an orthogonal projection when using the usual Euclidean inner product.

Direct multiplication confirms

$$PP = X(X^T A X)^{-1} X^T A X(X^T A X)^{-1} X^T A = X(X^T A X)^{-1} X^T A = P$$

therefore $P$ is a projection on $C(X)$. To be an orthogonal projection, we must show that it is self-adjoint. As the adjoint, $P'$ is the unique matrix that satisfies, $\langle P' x, y \rangle = \langle x, Py \rangle$, for $P$ to be self-adjoint or symmetric, we must show that $\langle x, Py \rangle = \langle Px, y \rangle$. Direct substitution in the inner product $\langle x, Py \rangle$:

$$\langle x, Py \rangle = x^T A Py$$
$$= x^T A X (X^T A X)^{-1} X^T A y$$
$$= (X(X^T A X)^{-1} X^T A x)^T A y$$
$$= (Px)^T A y$$
$$= (Px, y)$$

confirms that $P = P'$ (symmetric), and therefore $P$ is an orthogonal projection onto $C(X)$.

Given a subspace $\mathbb{M}$ in an inner product space $(\mathbb{V}, \langle \cdot, \cdot \rangle)$, we will need to be able to characterize the orthogonal projection $P$ onto $\mathbb{M}$. 

Theorem A.22. Let \( \{u_1, u_r\} \) denote an ONB for \( \mathcal{M} \). Then the orthogonal projection \( P \) onto \( \mathcal{M} \) of \( v \in \mathcal{V} \) may be represented as
\[
Pv = \sum_{i=1}^{r} \langle u_i, v \rangle u_i
\]

Theorem A.23. For \( \mathcal{V} = \mathbb{R}^n \) with the usual inner product, let \( U \) denote the \( n \times r \) matrix with columns of \( U \) composed of the basis vectors \( u_1, \ldots, u_r \) where \( \{ u_1, \ldots, u_r \} \) form an ONB for \( \mathcal{M} \subset \mathcal{V} \). Then \( P =UU^T \) is an orthogonal projection onto the \( C(U) = \mathcal{M} \).

Using the characterization of projections in terms of inner products in Theorem A.22,
\[
Pv = \sum_{i=1}^{r} \langle u_i, v \rangle u_i
= \sum_{i} (u_i^T v) u_i
= \sum_{i} u_i u_i^T v
\text{ because the inner product is a scalar times } u_i
= UU^T v
\]
so that we may characterize the matrix of the linear transformation corresponding to \( P \) as \( P =UU^T \). If \( \mathcal{M} \) is generated by the span of a matrix \( X \), then we can always obtain \( U \) via Gram-Schmidt orthogonalization. Because the \( u_i \) have length one and their span generates a one dimensional subspace, it is easy to show that each of the matrices \( u_i u_i^T \) is a rank one orthogonal projection.

The above representation show that we may write an orthogonal projection onto a \( r \) dimensional space via the sum of rank one projections. In fitting linear models and model selection we will need to break spaces into different components. The question that then arises is if we take sum or differences of linear transformation, when is the result a projection? If it is a projection, then on what and along which spaces? This is addressed in the following theorem.

Theorem A.24. Let \( P_1 \) be a projection on \( \mathcal{M}_1 \) along \( \mathcal{M}_1^c \) and \( P_2 \) be a projection on \( \mathcal{M}_2 \) along \( \mathcal{M}_2^c \). Then
(a) \( P \equiv P_1 + P_2 \) is a projection if and only if \( P_1 P_2 = P_2 P_1 = 0 \). In the condition holds, then \( P \) is a projection on \( \mathcal{M} = \mathcal{M}_1 + \mathcal{M}_2 \) and the null space of \( P \) is the intersection of the null spaces of the two projections, \( \mathcal{M}_1^c \cap \mathcal{M}_2^c \)
(b) \( P \equiv P_1 - P_2 \) is a projection if and only if \( P_1 P_2 = P_2 P_1 = P_2 \). If so, then \( P \) is a projection on \( \mathcal{M}_1 \cap \mathcal{M}_2^c \) and along \( \mathcal{M}_1^c + \mathcal{M}_2 \).
(c) If \( P_1 P_2 = P_2 P_1 \equiv P \), then \( P \) is the projection on \( \mathcal{M}_1 \cap \mathcal{M}_2 \) along \( \mathcal{M}_1^c + \mathcal{M}_2^c \).

For the case of several projections we have the following result:
Theorem A.25. Let $P_1, \ldots, P_k$ denote (orthogonal) projections in an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then $P = \sum P_i$ is an orthogonal projection if and only if $P_i P_j = 0$ for all $i \neq j$.

For a given matrix $X$, one may use the Gram-Schmidt orthogonalization procedure to find an ONB for the column space of $X$ and define an orthogonal projection onto $C(X)$ as $P = UU^T$ where $U$ is the matrix with columns based on the vectors in the ONB. The example A.21 showed the construction of a projection directly in terms of the matrix $X$ and the inner product. To define the projection operator directly using the matrix $X$ a generalized inverse.

Definition A.26. A generalized inverse of a $n \times n$ matrix $A$ is any $n \times n$ matrix $G$ that satisfies $AGA = A$. A generalized inverse of $A$ is denoted as $A^{-}$.

Every matrix has at least one generalized inverse. To show this, it is useful to have the following decomposition of a square symmetric matrix.

Theorem A.27. Spectral Decomposition Let $A$ denote a $n \times n$ symmetric matrix $A = A^T$. Then there exists a $n \times n$ matrix $U$ such that $U^T U = UU^T = I$ and a diagonal matrix $D$ such that $A = UDU^T$. 