STAT 510

Prerequisite Knowledge

This is brief review of concepts that we will use regularly throughout the semester. Please note that the review is not exhaustive but focuses on the topics that will come up repeatedly.

1 Linear Algebra

Scalars, vectors, matrices. We will agree on the following notational conventions for the purpose of distinguishing between scalars, vectors and matrices.

- Lowercase letters $\Longrightarrow$ scalars: $x, \quad c, \quad \sigma.$
- Boldface, lowercase letters $\Longrightarrow$ vectors: $\boldsymbol{x}, \quad \boldsymbol{y}, \quad \boldsymbol{\beta}.$
- Boldface, uppercase letters $\Longrightarrow$ matrices: $\boldsymbol{A}, \quad \boldsymbol{X}, \quad \boldsymbol{\Sigma}.$

This implies that random vectors, e.g., $\boldsymbol{x}$ or $\boldsymbol{y}$, will be denoted using lowercase letters despite being random and contrary to what you might have learned in other statistics classes. For our purposes, it will become clear from the context whether $\boldsymbol{x}$ represents a random vector or a vector of observed values.

Vector and Vector Transpose. A vector is a matrix with one column:

$$\boldsymbol{x} = \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix}.$$

We use $\boldsymbol{x}^\top$ to denote the transpose of the vector $\boldsymbol{x}$:

$$\boldsymbol{x}^\top = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \end{bmatrix},$$

i.e., $\boldsymbol{x}$ is a matrix with one column and $\boldsymbol{x}^\top$ is the matrix with the same entries as $\boldsymbol{x}$ but written as a row rather than a column.

Notation for Dimensions and Elements of a Matrix. Suppose $\boldsymbol{A}$ is a matrix with m *rows* and n *columns*.

$$\underset{m\times n}{\boldsymbol{A}} = \underset{m\times n}{[a_{ij}]} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Then we say that $\boldsymbol{A}$ has <u>dimensions</u> $m \times n$.

Let $a_{ij} \in \mathbb{R}$ be the <u>element</u> in the i^{th} row and j^{th} column of $\boldsymbol{A}$. We convey all this information with the notation

$$\underset{m\times n}{\boldsymbol{A}} = [a_{ij}].$$

Elements of a Matrix. Assume $m = n$. Then,

- $a_{11}, a_{22}, \ldots, a_{nn}$ are called the <u>main diagonal</u> elements.
- all remaining elements are called <u>off-diagonal</u> elements.
- if all off-diagnonal elements are equal to zero, $\underset{n\times n}{\boldsymbol{A}}$ is called a diagonal matrix and can be written as follows:

$$\underset{n\times n}{\boldsymbol{A}} = \mathrm{diag}(a_{11}, a_{22}, \ldots, a_{nn}).$$

Zero and One Vectors and the Identity Matrix.

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \underset{m\times m}{\boldsymbol{I}} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

We will also use $\underset{m\times n}{\mathbf{0}}$ to denote a matrix whose entries are all zero.

Transpose of a Matrix. Suppose $\boldsymbol{A}$ is an $m \times n$ matrix. Then,

- we may write $\boldsymbol{A}$ as $[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$, where $\boldsymbol{a}_i$ is the i^{th} column of $\boldsymbol{A}$ for each $i = 1, \ldots, n$.

- The transpose of the matrix $\boldsymbol{A}$ is

$$\boldsymbol{A}^\top = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]^\top = \begin{bmatrix} \boldsymbol{a}_1^\top \\ \vdots \\ \boldsymbol{a}_n^\top \end{bmatrix}.$$

Sum of two matrices: Suppose

$$\underset{m\times n}{\boldsymbol{A}} = [a_{ij}] \quad \text{and} \quad \underset{m\times n}{\boldsymbol{B}} = [b_{ij}].$$

Then,

$$\underset{m\times n}{\boldsymbol{A}} + \underset{m\times n}{\boldsymbol{B}} = \underset{m\times n}{\boldsymbol{C}} = [c_{ij} = a_{ij} + b_{ij}]\,;$$

i.e., the sum of $m \times n$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ is an $m \times n$ matrix whose entry in the i^{th} row and j^{th} column is the sum of the entry in the i^{th} row and j^{th} column of $\boldsymbol{A}$ and the entry in the i^{th} row and j^{th} column of $\boldsymbol{B}$ ($i = 1, \ldots, m$ and $j = 1, \ldots, n$).

The Product of a Scalar and a Matrix. Suppose $\underset{m\times n}{\boldsymbol{A}} = [a_{ij}]$. For any $c \in \mathbb{R}$,

$$c\underset{m\times n}{\boldsymbol{A}} = c\underset{m\times n}{[a_{ij}]} = \underset{m\times n}{[ca_{ij}]} = \begin{bmatrix} ca_{11} & ca_{12} & \cdots & ca_{1n} \\ ca_{21} & ca_{22} & \cdots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{m1} & ca_{m2} & \cdots & ca_{mn} \end{bmatrix},$$

i.e., the product of the scalar c and the matrix $\underset{m\times n}{\boldsymbol{A}} = [a_{ij}]$ is the matrix whose entry in the i^{th} row and j^{th} column is c times a_{ij} for each $i = 1, \ldots, m$ and $j = 1, \ldots, n$.

Matrix Multiplication. Consider matrices $\underset{m\times n}{\boldsymbol{A}}$ and $\underset{n\times k}{\boldsymbol{B}}$

$$\underset{m\times n}{\boldsymbol{A}} = [a_{ij}] = \begin{bmatrix} \boldsymbol{a}_{(1)}^\top \\ \vdots \\ \boldsymbol{a}_{(m)}^\top \end{bmatrix} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n], \quad \text{where } \boldsymbol{a}_{(1)}^\top = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{bmatrix}$$

and

$$\underset{n\times k}{\boldsymbol{B}} = [b_{ij}] = \begin{bmatrix} \boldsymbol{b}_{(1)}^\top \\ \vdots \\ \boldsymbol{b}_{(n)}^\top \end{bmatrix} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k], \quad \text{where } \boldsymbol{b}_{(1)}^\top = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \end{bmatrix}$$

Then,

$$\underset{m\times n}{\boldsymbol{A}}\ \underset{n\times k}{\boldsymbol{B}} = \underset{m\times k}{\boldsymbol{C}} = [c_{ij} = \sum_{l=1}^{n} a_{il}b_{lj}] = [c_{ij} = \boldsymbol{a}_{(i)}^\top \boldsymbol{b}_j]$$

$$= [\boldsymbol{A}\boldsymbol{b}_1, \ldots, \boldsymbol{A}\boldsymbol{b}_k] = \begin{bmatrix} \boldsymbol{a}_{(1)}^\top \boldsymbol{B} \\ \vdots \\ \boldsymbol{a}_{(m)}^\top \boldsymbol{B} \end{bmatrix} = \sum_{l=1}^{n} \boldsymbol{a}_l \boldsymbol{b}_{(l)}^\top.$$

Matrix Multiplication Special Cases. If

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \text{ and } \boldsymbol{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}, \text{ then } \boldsymbol{a}^\top \boldsymbol{b} = \sum_{i=1}^{n} a_i b_i.$$

Also,

$$\boldsymbol{a}^\top \boldsymbol{a} = \sum_{i=1}^{n} a_i^2 \equiv ||\boldsymbol{a}||^2.$$

$$||\boldsymbol{a}|| \equiv \sqrt{\boldsymbol{a}^\top \boldsymbol{a}} = \sqrt{\sum_{i=1}^{n} a_i^2} \text{ is known as the } \underline{\text{Euclidean norm}} \text{ of } \boldsymbol{a}.$$

Note that the <u>length</u> of a vector $\boldsymbol{a}$ corresponds to its Euclidean norm.

Linear Combinations. If $c_1, \ldots, c_n \in \mathbb{R}$ and $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in \mathbb{R}^m$, then

$$\sum_{i=1}^{n} c_i \boldsymbol{a}_i = c_1 \boldsymbol{a}_1 + \cdots + c_n \boldsymbol{a}_n$$

- is a <u>linear combination</u> (LC) of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$.
- $c_1, \ldots, c_n$ are called the <u>coefficients</u> of the LC.

Linear Independence and Linear Dependence.

- The vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ are <u>linearly independent</u> (LI) if and only if

$$\sum_{i=1}^{n} c_i \boldsymbol{a}_i = \boldsymbol{0} \Longrightarrow c_1 = \cdots = c_n = 0.$$

- The vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ are linearly dependent (LD) if and only if there exists a set of coefficients $c_1, \ldots, c_n$ with $c_i \neq 0$ for at least one i such that

$$\sum_{i=1}^{n} c_i \boldsymbol{a}_i = \boldsymbol{0}.$$

Orthogonality.

- The two vectors $\boldsymbol{x}, \boldsymbol{y}$ are orthogonal to each other if their inner product is zero, i.e.,

$$\boldsymbol{x}^\top \boldsymbol{y} = \boldsymbol{y}^\top \boldsymbol{x} = \sum_{i=1}^{n} x_i y_i = 0.$$

- If, in addition, $\boldsymbol{a}_i^\top \boldsymbol{a}_i = 1$ for all $i = 1, \ldots, n$, then the set is said to be orthonormal.
- Any orthogonal set of nonnull vectors is linearly independent.
- The vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are mutually orthogonal if

$$\boldsymbol{x}_i^\top \boldsymbol{x}_j = 0, \quad \forall\, i \neq j.$$

- The vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are mutually orthonormal if

$$\boldsymbol{x}_i^\top \boldsymbol{x}_j = 0 \quad \forall\, i \neq j, \text{ and } \|\boldsymbol{x}_i\| = 1 \quad \forall\, i = 1, \ldots, n.$$

Vector Spaces. Suppose that $\mathcal{S}$ is a nonempty set of vectors satisfying the following properties:

- if $\boldsymbol{x}_i \in \mathcal{S}, \boldsymbol{x}_j \in \mathcal{S}$, then $\boldsymbol{x}_i + \boldsymbol{x}_j \in \mathcal{S}$ (closed under addition),
- if $c \in \mathbb{R}, \boldsymbol{x}_i \in \mathcal{S}$, then $c\boldsymbol{x}_i \in \mathcal{S}$ (closed under scalar multiplication),

  in other words,

$$c_i \boldsymbol{x}_i + c_j \boldsymbol{x}_j \in \mathcal{S} \quad \forall\, c_i, c_j \in \mathbb{R};\ \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{S}.$$

A nonempty set of n vectors $\mathcal{V}$ is a subspace of a vector space $\mathcal{S}$ if $\mathcal{V}$ is a vector space and $\mathcal{V}$ is a subset of $\mathcal{S}$.

Span, Basis, Dimension of a Set of Vectors.

- A vector space $\mathcal{S}$ is said to be generated by a set of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ if for

$$\boldsymbol{x} \in \mathcal{S} \Rightarrow \boldsymbol{x} = \sum_{i=1}^{n} c_i \boldsymbol{x}_i \ \text{ for some } \ c_1, \ldots, c_n \in \mathbb{R}.$$

- $\text{span}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} = \mathcal{S}$ is the vector space generated by $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$.
- If a vector space $\mathcal{S}$ is generated by LI vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, then $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ form a basis for $\mathcal{S}$.
- A basis for a vector space is not unique, but the number of vectors in the basis, known as dimension of the vector space, is unique.

Column and Row spaces. Let $\boldsymbol{A}$ denote a $m \times n$ matrix (m – number of rows & n – number of columns). Then,

- The *row space* is the subspace of $\mathbb{R}^n$ spanned by the m row vectors of $\boldsymbol{A}$.
- The *column space* is the subspace of $\mathbb{R}^m$ spanned by the n column vectors of $\boldsymbol{A}$.

Note, that following the definition of vector spaces the following is true

- The set of all possible linear combinations of the columns of $\boldsymbol{A}$ is called the column space of $\boldsymbol{A}$ and is written as

$$\mathcal{C}(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^m : \boldsymbol{x} = \boldsymbol{A}\boldsymbol{c} \text{ for some } \boldsymbol{c} \in \mathbb{R}^n\}.$$

  Any member of the set $\mathcal{C}(\boldsymbol{A})$ is an $m \times 1$ vector, so $\mathcal{C}(\boldsymbol{A})$ is a subset of $\mathbb{R}^m$.

- The set of all possible linear combinations of the rows of $\boldsymbol{A}$ is called the row space of $\boldsymbol{A}$ and is written as

$$\mathcal{R}(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} = \boldsymbol{A}^\top \boldsymbol{d} : \boldsymbol{d} \in \mathbb{R}^m\}.$$

  Any member of the set $\mathcal{R}(\boldsymbol{A})$ is an $n \times 1$ vector, so $\mathcal{R}(\boldsymbol{A})$ is a subset of $\mathbb{R}^n$. Additionally, observe that $\mathcal{R}(\boldsymbol{A}) = \mathcal{C}(\boldsymbol{A}^\top)$.

- Note that $\mathcal{C}(\boldsymbol{A}) \subseteq \mathbb{R}^m$ and $\mathcal{R}(\boldsymbol{A}) \subseteq \mathbb{R}^n$.
- If $\boldsymbol{A}$ denote a $m \times n$ matrix, the row space and column space of $\boldsymbol{A}$ have the same dimension.
- The dimension of the row (or column) space of $\boldsymbol{A}$ is called *rank* of $\boldsymbol{A}$ denoted as $\text{rank}(\boldsymbol{A})$. (see more below)

Null Space. Let $\boldsymbol{A}$ denote a $m \times n$ matrix. The set

$$\mathcal{N}(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}$$

is called the <u>null space</u> of $\boldsymbol{A}$. Any member of the set $\mathcal{N}(\boldsymbol{A})$ is an $n \times 1$ vector, so $\mathcal{N}(\boldsymbol{A})$ is a subset of $\mathbb{R}^n$.

Matrix Product Transpose and Symmetric Matrices. If $\boldsymbol{A}$ is any matrix, and $\boldsymbol{A}^\top$ its transpose, then

- $(\boldsymbol{A}^\top)^\top = \boldsymbol{A}$
- if $\boldsymbol{A}$ and $\boldsymbol{B}$ have the same dimensions and c and d are arbitrary scalars, then $(c\boldsymbol{A} + d\boldsymbol{B})^\top = c\boldsymbol{A}^\top + d\boldsymbol{B}^\top$
- if $\boldsymbol{A}$ is conformal for post-multiplication by $\boldsymbol{B}$, then $(\boldsymbol{A}\boldsymbol{B})^\top = \boldsymbol{B}^\top\boldsymbol{A}^\top$.

Square Matrices.

- A matrix $\underset{m\times n}{\boldsymbol{A}}$ is said to be <u>square</u> if and only if $m = n$.
- A square matrix $\underset{n\times n}{\boldsymbol{A}}$ is <u>symmetric</u> if $\boldsymbol{A} = \boldsymbol{A}^\top$.
- If $\boldsymbol{A}$ is an $m \times n$ matrix, then $\boldsymbol{A}^\top\boldsymbol{A}$ is an $n \times n$ matrix.
- Thus, $\boldsymbol{A}^\top\boldsymbol{A}$ is a square matrix for any matrix $\boldsymbol{A}$.
- A square matrix $\underset{n\times n}{\boldsymbol{A}}$ is <u>diagonal</u> if $a_{ij} = 0, \forall\, i \neq j$.
- A square matrix $\underset{n\times n}{\boldsymbol{A}}$ is <u>upper triangular</u> if $a_{ij} = 0, \forall\, i > j$.
- A square matrix $\underset{n\times n}{\boldsymbol{A}}$ is <u>lower triangular</u> if $a_{ij} = 0, \forall\, i < j$.

Rank & Trace of a Square Matrix, Idempotent Matrices

- The <u>rank</u> of a matrix $\boldsymbol{A}$ is written as $\mathrm{rank}(\boldsymbol{A})$ and is the maximum number of linearly independent rows (or columns) of $\boldsymbol{A}$.
- The <u>trace</u> of an $n \times n$ matrix $\boldsymbol{A}$ is written as $\mathrm{trace}(\boldsymbol{A})$ and is the sum of the diagonal elements of $\boldsymbol{A}$; i.e.,

$$\mathrm{trace}(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ii}.$$

- A matrix $\boldsymbol{A}$ is said to be idempotent if and only if $\boldsymbol{A}\boldsymbol{A} = \boldsymbol{A}$.The rank of an idempotent matrix is equal to its trace; i.e.,

$$\operatorname{rank}(\boldsymbol{A}) = \operatorname{trace}(\boldsymbol{A}).$$

Orthogonal Matrices.

- A square matrix, $\underset{m\times m}{\boldsymbol{A}}$, with m mutually orthonormal columns is called an orthogonal matrix, i.e. $\boldsymbol{A}^\top \boldsymbol{A} = \underset{m\times m}{\boldsymbol{I}}$.
- In $\mathbb{R}^n$ an orthogonal matrix $\underset{n\times n}{\boldsymbol{Q}}$ with determinant 1 is sometimes called a rotation matrix because $\underset{n\times n}{\boldsymbol{Q}}$ can be used to rotate any vector $\underset{n\times 1}{\boldsymbol{x}}$ counterclockwise by an angle θ into a new position $\underset{n\times 1}{\boldsymbol{x}^*}$ with respect to a fixed reference frame, i.e., $\underset{n\times 1}{\boldsymbol{x}^*} = \boldsymbol{Q}\boldsymbol{x}$.

Inverse of a Matrix, Singular, Nonsingular Matrices. Consider an $n \times n$ matrix $\boldsymbol{A}$.

- If $\operatorname{rank}(\boldsymbol{A}) = n$, there exists a matrix $\underset{n\times n}{\boldsymbol{B}}$ such that $\underset{n\times n}{\boldsymbol{A}}\,\underset{n\times n}{\boldsymbol{B}} = \underset{n\times n}{\boldsymbol{I}}$.
- Such a matrix $\boldsymbol{B}$ is called the inverse of $\boldsymbol{A}$ and is denoted $\boldsymbol{A}^{-1}$.
- If $\operatorname{rank}(\boldsymbol{A}) = n$, then $\underset{n\times n}{\boldsymbol{A}}$ is said to be nonsingular.
- If $\operatorname{rank}(\boldsymbol{A}) < n$, then $\underset{n\times n}{\boldsymbol{A}}$ is said to be singular.

Inverse of a Matrix.

- An $n \times n$ square matrix $\boldsymbol{A}$ is nonsingular or invertible if and only if there exists an $n \times n$ square matrix $\boldsymbol{B}$ such that $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{I}$.
- If $\boldsymbol{A}$ is nonsingular and $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{I}$, then $\boldsymbol{B}$ is the unique inverse of $\boldsymbol{A}$ and is written as $\boldsymbol{A}^{-1}$.
- For a nonsingular matrix $\boldsymbol{A}$, we have $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$. (It is also true that $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$.)
- A square matrix without an inverse is called singular.
- An $n \times n$ matrix $\boldsymbol{A}$ is singular if and only if $\operatorname{rank}(\boldsymbol{A}) < n$.

Generalized Inverses.

- $\boldsymbol{G}$ is a generalized inverse of an $m \times n$ matrix $\boldsymbol{A}$ if and only if $\boldsymbol{A}\boldsymbol{G}A = \boldsymbol{A}$.

- We usually denote a generalized inverse of $\boldsymbol{A}$ by $\boldsymbol{A}^{-}$.

- If $\boldsymbol{A}$ is nonsingular, i.e., if $\boldsymbol{A}^{-1}$ exists, then $\boldsymbol{A}^{-1}$ is the one and only generalized inverse of $\boldsymbol{A}$.
$$\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{I} = \boldsymbol{A}.$$

- If $\boldsymbol{A}$ is singular, i.e., if $\boldsymbol{A}^{-1}$ does not exist, then there are infinitely many generalized inverses of $\boldsymbol{A}$.

Finding a Generalized Inverse of a Matrix $\boldsymbol{A}$.

- Find any $n \times n$ nonsingular submatrix of $\boldsymbol{A}$ where $n = \text{rank}(\boldsymbol{A})$. Call this matrix $\boldsymbol{W}$.
- Invert and transpose $\boldsymbol{W}$, i.e., compute $(\boldsymbol{W}^{-1})^\top$.
- Replace each element of $\boldsymbol{W}$ in $\boldsymbol{A}$ with the corresponding element of $(\boldsymbol{W}^{-1})^\top$.
- Replace all other elements in $\boldsymbol{A}$ with zeros.
- Transpose the resulting matrix to obtain $\boldsymbol{G}$, a generalized inverse for $\boldsymbol{A}$.

Quadratic Forms. Let $\boldsymbol{x}$ be an $m \times 1$ vector and $\boldsymbol{y}$ an $n \times 1$ vector and $\boldsymbol{A}$ an $m \times n$ matrix. Then
$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y} = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j a_{ij}$$
is called a *bilinear form* in $\boldsymbol{x}$ and $\boldsymbol{y}$.
We will mostly be interested in the special case in which $m = n$, so that $\boldsymbol{A}$ is $m \times m$, and $\boldsymbol{x} = \boldsymbol{y}$ resulting in
$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = \sum_{i=1}^{m}\sum_{j=1}^{m} x_i x_j a_{ij}$$
which is called a *quadratic form* in $\boldsymbol{x}$. $\boldsymbol{A}$ is referred to as the matrix of the quadratic form. Moving forward we will assume for simplicity that $\boldsymbol{A}$ is symmetric (see Schott, 2016 for more detail).

Every symmetric matrix $\boldsymbol{A}$ and its associated quadratic form can be classified into one of the following categories:

- if $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}$, then $\boldsymbol{A}$ is <u>positive definite</u>.

- if $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}$, and $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = 0$ for some $\boldsymbol{x} \neq \boldsymbol{0}$, then $\boldsymbol{A}$ is <u>positive semidefinite</u>.

- if $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} < 0$ for all $\boldsymbol{x} \neq \mathbf{0}$, then $\boldsymbol{A}$ is negative definite.
- if $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \leq 0$ for all $\boldsymbol{x} \neq \mathbf{0}$, and $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = 0$ for some $\boldsymbol{x} \neq \mathbf{0}$, then $\boldsymbol{A}$ is negative semidefinite.
- if $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} < 0$ for some $\boldsymbol{x}$ and $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} > 0$ for some $\boldsymbol{x}$, then $\boldsymbol{A}$ is indefinite.

We will later see that quadratic forms play an important role in inferential statistics.

Positive and Non-Negative Definite Matrices. Note that authors do not always agree on the following.

Sometimes, the term nonnegative definite will be used to refer to a symmetric matrix that is either positive definite or positive semidefinite, e.g. Schott, 2016. Others reserve the term nonnegative definite strictly for positive semidefinite matrices. For this class we will assume the following: Let $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ be known as a quadratic form.

We say that an $n \times n$ matrix $\boldsymbol{A}$ is positive definite (PD) if and only if

- $\boldsymbol{A}$ is symmetric (i.e., $\boldsymbol{A} = \boldsymbol{A}^\top$), and
- $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} > 0$ for all $\boldsymbol{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$.

We say that an $n \times n$ matrix $\boldsymbol{A}$ is non-negative definite (NND) if and only if

- $\boldsymbol{A}$ is symmetric (i.e., $\boldsymbol{A} = \boldsymbol{A}^\top$), and
- $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$.

Positive and Non-Negative Definite Matrices. A matrix that is positive definite is nonsingular; i.e.,

$$\boldsymbol{A} \text{ positive definite } \implies \boldsymbol{A}^{-1} \text{ exists.}$$

A matrix that is non-negative definite but not positive definite is singular.

Note that we are not interested in negative definite or negative semidefinite matrices as these are not the main focus in Linear Models.

2 Statistical Concepts

Random Vectors. A <u>random vector</u> is a vector whose components are random variables.

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Expected Value of a random vector. The <u>expected value</u>, or <u>mean</u>, of a random vector $\boldsymbol{y}$ is the vector of expected values of the components of $\boldsymbol{y}$.

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \Longrightarrow \mathrm{E}(\boldsymbol{y}) = \begin{bmatrix} \mathrm{E}(y_1) \\ \mathrm{E}(y_2) \\ \vdots \\ \mathrm{E}(y_n) \end{bmatrix}$$

Likewise, if $\boldsymbol{A} = [a_{ij}]$ is a matrix of random variables, then $\mathrm{E}(\boldsymbol{A}) = [\mathrm{E}(a_{ij})]$; i.e., the expected value of $\boldsymbol{A}$ is the matrix of expected values of the elements of $\boldsymbol{A}$.

Variance of a Random Vector. The <u>variance</u> of a random vector $\boldsymbol{y} = [y_1, y_2, \ldots, y_n]^\top$ is the matrix whose i^{th}, j^{th} element is $\mathrm{Cov}(y_i, y_j)$ $(i, j \in \{1, \ldots, n\})$.

$$\mathrm{Var}(\boldsymbol{y}) = \begin{bmatrix} \mathrm{Cov}(y_1, y_1) & \mathrm{Cov}(y_1, y_2) & \cdots & \mathrm{Cov}(y_1, y_n) \\ \mathrm{Cov}(y_2, y_1) & \mathrm{Cov}(y_2, y_2) & \cdots & \mathrm{Cov}(y_2, y_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(y_n, y_1) & \mathrm{Cov}(y_n, y_2) & \cdots & \mathrm{Cov}(y_n, y_n) \end{bmatrix}.$$

The covariance of a random variable with itself is the variance of that random variable. Thus,

$$\mathrm{Var}(\boldsymbol{y}) = \begin{bmatrix} \mathrm{Var}(y_1) & \mathrm{Cov}(y_1, y_2) & \cdots & \mathrm{Cov}(y_1, y_n) \\ \mathrm{Cov}(y_2, y_1) & \mathrm{Var}(y_2) & \cdots & \mathrm{Cov}(y_2, y_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(y_n, y_1) & \mathrm{Cov}(y_n, y_2) & \cdots & \mathrm{Var}(y_n) \end{bmatrix}.$$

Covariance Between Two Random Vectors. The <u>covariance</u> between random vectors $\boldsymbol{u} = [u_1, \ldots, u_m]^\top$ and $\boldsymbol{v} = [v_1, \ldots, v_n]^\top$ is the matrix whose i^{th}, j^{th} element is $\mathrm{Cov}(u_i, v_j)$ $(i \in \{1, \ldots, m\},\ j \in \{1, \ldots, n\})$.

$$
\begin{aligned}
\mathrm{Cov}(\boldsymbol{u}, \boldsymbol{v}) &= \begin{bmatrix} \mathrm{Cov}(u_1, v_1) & \mathrm{Cov}(u_1, v_2) & \cdots & \mathrm{Cov}(u_1, v_n) \\ \mathrm{Cov}(u_2, v_1) & \mathrm{Cov}(u_2, v_2) & \cdots & \mathrm{Cov}(u_2, v_n) \\ \vdots & \vdots & & \vdots \\ \mathrm{Cov}(u_m, v_1) & \mathrm{Cov}(u_m, v_2) & \cdots & \mathrm{Cov}(u_m, v_n) \end{bmatrix} \\
&= \mathrm{E}(\boldsymbol{u}\boldsymbol{v}^\top) - \mathrm{E}(\boldsymbol{u})\mathrm{E}(\boldsymbol{v}^\top).
\end{aligned}
$$

Linear Transformation of a Random Vector. If $\boldsymbol{y}$ is an $n \times 1$ random vector, $\boldsymbol{A}$ is an $m \times n$ matrix of constants, and $\boldsymbol{b}$ is an $m \times 1$ vector of constants, then

$$\boldsymbol{A}\boldsymbol{y} + \boldsymbol{b}$$

is a linear transformation of the random vector $\boldsymbol{y}$. Thus, $\boldsymbol{A}\boldsymbol{y} + \boldsymbol{b}$ is a random variable itself with

- $\mathrm{E}(\boldsymbol{A}\boldsymbol{y} + \boldsymbol{b}) = \boldsymbol{A}\mathrm{E}(\boldsymbol{y}) + \boldsymbol{b}$
- $\mathrm{Var}(\boldsymbol{A}\boldsymbol{y} + \boldsymbol{b}) = \boldsymbol{A}\mathrm{Var}(\boldsymbol{y})\boldsymbol{A}^\top$
- $\mathrm{Cov}(\boldsymbol{A}\boldsymbol{y} + \boldsymbol{b}, \boldsymbol{C}\boldsymbol{y} + \boldsymbol{d}) = \boldsymbol{A}\mathrm{Var}(\boldsymbol{y})\boldsymbol{C}^\top$

Multivariate and Standard Multivariate Normal Distributions.

If $z_1, \ldots, z_n \overset{iid}{\sim} \mathcal{N}(0, 1)$, then

$$\boldsymbol{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$$

has a standard multivariate normal distribution: $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{n\times n})$.

Suppose $\boldsymbol{z}$ is an $n \times 1$ standard multivariate normal random vector, i.e., $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{n\times n})$ and suppose further $\boldsymbol{A}$ is an $m \times n$ matrix of constants and $\boldsymbol{\mu}$ is an $m \times 1$ vector of constants. Then, $\boldsymbol{A}\boldsymbol{z} + \boldsymbol{\mu}$ has a multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{A}\boldsymbol{A}^\top$:

$$\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{n\times n}) \Longrightarrow \boldsymbol{A}\boldsymbol{z} + \boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{A}\boldsymbol{A}^\top).$$

Multivariate Normal Distributions.

If $\boldsymbol{\mu}$ is an $m \times 1$ vector of constants and $\boldsymbol{\Sigma}$ is a $m \times m$ symmetric, positive definite (PD) matrix of rank n, then $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ signifies the multivariate normal distribution with mean $\boldsymbol{\mu}$

and variance $\boldsymbol{\Sigma}$.

If $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\boldsymbol{y} \stackrel{d}{=} \boldsymbol{A}\boldsymbol{z} + \boldsymbol{\mu}$, where $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{n\times n})$ and $\boldsymbol{A}$ is an $m \times n$ matrix of rank n such that $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{\Sigma}$.

Note that if $\boldsymbol{\Sigma}$ is nonnegative definite, i.e., positive semidefinite then $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is said to have a singular normal distribution implying that $\boldsymbol{\Sigma}^{-1}$ does not exist resulting in a degenerate density. However, the random vector $\boldsymbol{y}$ can still be expressed in terms of independent standard normal random variables.

Linear Transformations of Multivariate Normal Distributions are Multivariate Normal.

$$\begin{aligned}
\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad &\leftrightarrow \quad \boldsymbol{y} \stackrel{d}{=} \boldsymbol{A}\boldsymbol{z} + \boldsymbol{\mu},\ \boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}),\ \boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{\Sigma} \\
&\leftrightarrow \quad \boldsymbol{C}\boldsymbol{y} + \boldsymbol{d} \stackrel{d}{=} \boldsymbol{C}(\boldsymbol{A}\boldsymbol{z} + \boldsymbol{\mu}) + \boldsymbol{d} \\
&\leftrightarrow \quad \boldsymbol{C}\boldsymbol{y} + \boldsymbol{d} \stackrel{d}{=} \boldsymbol{C}\boldsymbol{A}\boldsymbol{z} + \boldsymbol{C}\boldsymbol{\mu} + \boldsymbol{d} \\
&\leftrightarrow \quad \boldsymbol{C}\boldsymbol{y} + \boldsymbol{d} \stackrel{d}{=} \boldsymbol{M}\boldsymbol{z} + \boldsymbol{u},\ \boldsymbol{M} \equiv \boldsymbol{C}\boldsymbol{A},\ \boldsymbol{u} \equiv \boldsymbol{C}\boldsymbol{\mu} + \boldsymbol{d} \\
&\leftrightarrow \quad \boldsymbol{C}\boldsymbol{y} + \boldsymbol{d} \sim \mathcal{N}(\boldsymbol{u}, \boldsymbol{M}\boldsymbol{M}^\top).
\end{aligned}$$

Non-Central Chi-Squared Distributions. If $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{I}_{n\times n})$, then

$$w \equiv \boldsymbol{y}^\top \boldsymbol{y} = \sum_{i=1}^{n} y_i^2$$

has a non-central chi-squared distribution with n degrees of freedom and non-centrality parameter $(\boldsymbol{\mu}^\top\boldsymbol{\mu})/2$:

$$w \sim \chi_n^2\left((\boldsymbol{\mu}^\top\boldsymbol{\mu})/2\right).$$

(Some define the non-centrality parameter as $\boldsymbol{\mu}^\top\boldsymbol{\mu}$ rather than $(\boldsymbol{\mu}^\top\boldsymbol{\mu})/2$.)

Central Chi-Squared Distributions. If $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{n\times n})$, then

$$w \equiv \boldsymbol{z}^\top \boldsymbol{z} = \sum_{i=1}^{n} z_i^2$$

has a central chi-squared distribution with n degrees of freedom:

$$w \sim \chi_n^2.$$

A central chi-squared distribution is a non-central chi-squared distribution with non-centrality parameter 0: $w \sim \chi^2_n(0)$.

Important Distributional Result about Quadratic Forms. Suppose $\boldsymbol{\Sigma}$ is an $n \times n$ positive definite matrix. Suppose further that $\boldsymbol{A}$ is an $n \times n$ symmetric matrix of rank m such that $\boldsymbol{A\Sigma}$ is idempotent (i.e., $\boldsymbol{A\Sigma A\Sigma} = \boldsymbol{A\Sigma}$).

Then, $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ implies

$$\boldsymbol{y}^\top \boldsymbol{A} \boldsymbol{y} \sim \chi^2_m \left((\boldsymbol{\mu}^\top \boldsymbol{A} \boldsymbol{\mu})/2\right)$$

.

Note that this result holds even when $\boldsymbol{\Sigma}$ is nonnegative definite (NND) as long as $\boldsymbol{\Sigma A \Sigma A \Sigma} = \boldsymbol{\Sigma A \Sigma}$ and trace$(\boldsymbol{A\Sigma} = m)$. For a proof see Schott, 2016.

Mean and Variance of Chi–Squared Distributions. If $w \sim \chi^2_m(\theta)$, then

$$\mathrm{E}(w) = m + 2\theta \qquad \text{and} \qquad \mathrm{Var}(w) = 2m + 8\theta.$$

Non-Central t–Distributions. Suppose $y \sim \mathcal{N}(\delta, 1)$, $w \sim \chi^2_m$ and suppose y and w are independent.

Then $y/\sqrt{w/m}$ has a non-central t-distribution with m degrees of freedom and non-centrality parameter δ:

$$\frac{y}{\sqrt{w/m}} \sim t_m(\delta).$$

Central t–Distributions. Suppose $z \sim \mathcal{N}(0, 1)$, $w \sim \chi^2_m$ and suppose that z and w are independent.

Then $z/\sqrt{w/m}$ has a central t-distribution with m degrees of freedom:

$$\frac{z}{\sqrt{w/m}} \sim t_m.$$

The distribution t_m is the same as $t_m(0)$.

Non-Central F–Distributions. Suppose $w_1 \sim \chi^2_{m_1}(\theta)$, $w_2 \sim \chi^2_{m_2}$, and suppose w_1 and w_2 are independent.

Then $(w_1/m_1)/(w_2/m_2)$ has a non-central F-distribution with m_1 numerator degrees of freedom, m_2 denominator degrees of freedom, and non-centrality parameter θ:

$$\frac{w_1/m_1}{w_2/m_2} \sim F_{m_1,m_2}(\theta).$$

Central F–Distributions. Suppose $w_1 \sim \chi^2_{m_1}$, $w_2 \sim \chi^2_{m_2}$ and suppose w_1 and w_2 are independent.

Then $(w_1/m_1)/(w_2/m_2)$ has a central F-distribution with m_1 numerator degrees of freedom and m_2 denominator degrees of freedom:

$$\frac{w_1/m_1}{w_2/m_2} \sim F_{m_1,m_2} \quad \text{(which is the same as the } F_{m_1,m_2}(0)\text{-distribution).}$$

Relationship between $t-$ and $F-$ Distributions.

$$\text{If } u \sim t_m(\delta), \text{ then } u^2 \sim F_{1,m}(\delta^2/2).$$

Some Independence Results. Suppose $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is an $n \times n$ PD matrix.

- If $\boldsymbol{A}_1$ is an $n_1 \times n$ matrix of constants and $\boldsymbol{A}_2$ is an $n_2 \times n$ matrix of constants, then $\boldsymbol{A}_1\boldsymbol{\Sigma}\boldsymbol{A}_2^\top = \mathbf{0} \implies \boldsymbol{A}_1\boldsymbol{y} \perp \boldsymbol{A}_2\boldsymbol{y}$.
- If $\boldsymbol{A}_1$ is an $n_1 \times n$ matrix of constants and $\boldsymbol{A}_2$ is an $n \times n$ symmetric matrix of constants, then $\boldsymbol{A}_1\boldsymbol{\Sigma}\boldsymbol{A}_2 = \mathbf{0} \implies \boldsymbol{A}_1\boldsymbol{y} \perp \boldsymbol{y}^\top\boldsymbol{A}_2\boldsymbol{y}$.
- If $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are $n \times n$ symmetric matrices of constants, then $\boldsymbol{A}_1\boldsymbol{\Sigma}\boldsymbol{A}_2 = \mathbf{0} \implies \boldsymbol{y}^\top\boldsymbol{A}_1\boldsymbol{y} \perp \boldsymbol{y}^\top\boldsymbol{A}_2\boldsymbol{y}$.