

Preliminaries

Notation for Scalars, Vectors, and Matrices

Lowercase letters \implies scalars: x , c , σ .

Boldface, lowercase letters \implies vectors: \mathbf{x} , \mathbf{y} , $\boldsymbol{\beta}$.

Boldface, uppercase letters \implies matrices: \mathbf{A} , \mathbf{X} , $\boldsymbol{\Sigma}$.

The Product of a Scalar and a Matrix

Suppose A is a matrix with m rows and n columns. Let $a_{ij} \in \mathbb{R}$ be the element or entry in the i th row and j th column of A . We convey all this information with the notation

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

For any $c \in \mathbb{R}$,

$$cA = Ac = [ca_{ij}];$$

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

Vector and Vector Transpose

In STAT 510, a vector is a matrix with one column:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

In STAT 510, we use \mathbf{x}' to denote the *transpose* of the vector \mathbf{x} :

$$\mathbf{x}' = [x_1, \dots, x_n];$$

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

Transpose of a Matrix

Suppose A is an $m \times n$ matrix. Then we may write A as $[a_1, \dots, a_n]$, where a_i is the i th column of A for each $i = 1, \dots, n$.

The transpose of the matrix A is

$$A' = [a_1, \dots, a_n]' = \begin{bmatrix} a'_1 \\ \vdots \\ a'_n \end{bmatrix}.$$

Matrix Multiplication

$$\text{Suppose } \underset{m \times n}{\mathbf{A}} = [a_{ij}] = [\mathbf{a}_1, \dots, \mathbf{a}_n] = \begin{bmatrix} \mathbf{a}'_{(1)} \\ \vdots \\ \mathbf{a}'_{(m)} \end{bmatrix}$$

$$\text{and } \underset{n \times k}{\mathbf{B}} = [b_{ij}] = [\mathbf{b}_1, \dots, \mathbf{b}_k] = \begin{bmatrix} \mathbf{b}'_{(1)} \\ \vdots \\ \mathbf{b}'_{(n)} \end{bmatrix}.$$

$$\begin{aligned} \text{Then } \underset{m \times n}{\mathbf{A}} \underset{n \times k}{\mathbf{B}} &= \underset{m \times k}{\mathbf{C}} = \left[c_{ij} = \sum_{l=1}^n a_{il} b_{lj} \right] = [c_{ij} = \mathbf{a}'_{(i)} \mathbf{b}_j] \\ &= [\mathbf{A} \mathbf{b}_1, \dots, \mathbf{A} \mathbf{b}_k] = \begin{bmatrix} \mathbf{a}'_{(1)} \mathbf{B} \\ \vdots \\ \mathbf{a}'_{(m)} \mathbf{B} \end{bmatrix} = \sum_{l=1}^n \mathbf{a}_l \mathbf{b}'_{(l)}. \end{aligned}$$

Transpose of a Matrix Product

The transpose of a matrix product is a product of the transposes in reverse order; i.e.,

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

Rank and Trace

The *rank* of a matrix A is written as $\text{rank}(A)$ and is the maximum number of linearly independent rows (or columns) of A

The *trace* of an $n \times n$ matrix A is written as $\text{trace}(A)$ and is the sum of the diagonal elements of A ; i.e.,

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

Idempotent Matrices

A matrix A is said to be *idempotent* if $AA = A$.

The rank of an idempotent matrix is equal to its trace; i.e.,

$$\text{rank}(A) = \text{trace}(A).$$

Linear Combinations and Column Spaces

- $A\mathbf{b}$ is a *linear combination* of the columns of an $m \times n$ matrix A :

$$A\mathbf{b} = [\mathbf{a}_1, \dots, \mathbf{a}_n] \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = b_1\mathbf{a}_1 + \dots + b_n\mathbf{a}_n.$$

- The set of all possible linear combinations of the columns of A is called the *column space* of A and is written as

$$\mathcal{C}(A) = \{A\mathbf{b} : \mathbf{b} \in \mathbb{R}^n\}.$$

- Note that $\mathcal{C}(A) \subseteq \mathbb{R}^m$.

Inverse of a Matrix

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

Generalized Inverses

- G is a *generalized inverse* of an $m \times n$ matrix A if $AGA = A$.
- We usually denote a generalized inverse of A by A^- .
- If A is nonsingular, i.e., if A^{-1} exists, then A^{-1} is the one and only generalized inverse of A .

$$AA^{-1}A = IA = AI = A$$

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

Finding a Generalized Inverse of a Matrix A

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

Positive and Non-Negative Definite Matrices

$\mathbf{x}'\mathbf{A}\mathbf{x}$ is known as a *quadratic form*.

We say that an $n \times n$ matrix \mathbf{A} is *positive definite (PD)* iff

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

We say that an $n \times n$ matrix \mathbf{A} is *non-negative definite (NND)* iff

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

Positive and Non-Negative Definite Matrices

A matrix that is positive definite is nonsingular; i.e.,

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

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

Random Vectors

A *random vector* is a vector whose components are random variables.

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

Expected Value of a Random Vector

The *expected value*, or *mean*, of a random vector \mathbf{y} is the vector of expected values of the components of \mathbf{y} .

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \implies E(\mathbf{y}) = \begin{bmatrix} E(y_1) \\ E(y_2) \\ \vdots \\ E(y_n) \end{bmatrix}$$

Variance of a Random Vector

The *variance* of a random vector $\mathbf{y} = [y_1, y_2, \dots, y_n]'$ is the matrix whose i, j th element is $\text{Cov}(y_i, y_j)$ ($i, j \in \{1, \dots, n\}$).

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

Variance of a Random Vector

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

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

Covariance Between Two Random Vectors

The *covariance* between random vectors $\mathbf{u} = [u_1, \dots, u_m]'$ and $\mathbf{v} = [v_1, \dots, v_n]'$ is the matrix whose i, j th element is $\text{Cov}(u_i, v_j)$ ($i \in \{1, \dots, m\}$, $j \in \{1, \dots, n\}$).

$$\begin{aligned}\text{Cov}(\mathbf{u}, \mathbf{v}) &= \begin{bmatrix} \text{Cov}(u_1, v_1) & \text{Cov}(u_1, v_2) & \cdots & \text{Cov}(u_1, v_n) \\ \text{Cov}(u_2, v_1) & \text{Cov}(u_2, v_2) & \cdots & \text{Cov}(u_2, v_n) \\ \vdots & \vdots & & \vdots \\ \text{Cov}(u_m, v_1) & \text{Cov}(u_m, v_2) & \cdots & \text{Cov}(u_m, v_n) \end{bmatrix} \\ &= E(\mathbf{u}\mathbf{v}') - E(\mathbf{u})E(\mathbf{v}').\end{aligned}$$

Linear Transformation of a Random Vector

If \mathbf{y} is an $n \times 1$ random vector, \mathbf{A} is an $m \times n$ matrix of constants, and \mathbf{b} is an $m \times 1$ vector of constants, then

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

is a *linear transformation* of the random vector \mathbf{y} .

Mean, Variance, and Covariance of Linear Transformations of a Random Vector \mathbf{y}

$$E(\mathbf{A}\mathbf{y} + \mathbf{b}) = \mathbf{A}E(\mathbf{y}) + \mathbf{b}$$

$$\text{Var}(\mathbf{A}\mathbf{y} + \mathbf{b}) = \mathbf{A}\text{Var}(\mathbf{y})\mathbf{A}'$$

$$\text{Cov}(\mathbf{A}\mathbf{y} + \mathbf{b}, \mathbf{C}\mathbf{y} + \mathbf{d}) = \mathbf{A}\text{Var}(\mathbf{y})\mathbf{C}'$$

Standard Multivariate Normal Distributions

If $z_1, \dots, z_n \stackrel{iid}{\sim} N(0, 1)$, then

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

has a *standard multivariate normal distribution*: $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I})$.

Multivariate Normal Distributions

Suppose \mathbf{z} is an $n \times 1$ standard multivariate normal random vector, i.e., $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}_{n \times n})$.

Suppose \mathbf{A} is an $m \times n$ matrix of constants and $\boldsymbol{\mu}$ is an $m \times 1$ vector of constants.

Then $\mathbf{Az} + \boldsymbol{\mu}$ has a *multivariate normal distribution* with mean $\boldsymbol{\mu}$ and variance \mathbf{AA}' :

$$\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}) \implies \mathbf{Az} + \boldsymbol{\mu} \sim N(\boldsymbol{\mu}, \mathbf{AA}').$$

Multivariate Normal Distributions

If $\boldsymbol{\mu}$ is an $m \times 1$ vector of constants and $\boldsymbol{\Sigma}$ is a $m \times m$ symmetric, non-negative definite (NND) matrix of rank n , then $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ signifies the multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$.

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

Linear Transformations of Multivariate Normal Distributions are Multivariate Normal

$$\begin{aligned} \mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &\implies \mathbf{y} \stackrel{d}{=} \mathbf{A}\mathbf{z} + \boldsymbol{\mu}, \mathbf{z} \sim N(\mathbf{0}, \mathbf{I}), \mathbf{A}\mathbf{A}' = \boldsymbol{\Sigma} \\ &\implies \mathbf{C}\mathbf{y} + d \stackrel{d}{=} \mathbf{C}(\mathbf{A}\mathbf{z} + \boldsymbol{\mu}) + d \\ &\implies \mathbf{C}\mathbf{y} + d \stackrel{d}{=} \mathbf{C}\mathbf{A}\mathbf{z} + \mathbf{C}\boldsymbol{\mu} + d \\ &\implies \mathbf{C}\mathbf{y} + d \stackrel{d}{=} \mathbf{M}\mathbf{z} + \mathbf{u}, \mathbf{M} \equiv \mathbf{C}\mathbf{A}, \mathbf{u} \equiv \mathbf{C}\boldsymbol{\mu} + d \\ &\implies \mathbf{C}\mathbf{y} + d \sim N(\mathbf{u}, \mathbf{M}\mathbf{M}'). \end{aligned}$$

Non-Central Chi-Square Distributions

If $\mathbf{y} \sim N(\boldsymbol{\mu}, \mathbf{I}_{n \times n})$, then

$$w \equiv \mathbf{y}'\mathbf{y} = \sum_{i=1}^n y_i^2$$

has a *non-central chi-square distribution* with n degrees of freedom and *non-centrality parameter* $\boldsymbol{\mu}'\boldsymbol{\mu}/2$:

$$w \sim \chi_n^2(\boldsymbol{\mu}'\boldsymbol{\mu}/2).$$

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

Central Chi-Square Distributions

If $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}_{n \times n})$, then

$$w \equiv \mathbf{z}'\mathbf{z} = \sum_{i=1}^n z_i^2$$

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

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

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

Important Distributional Result about Quadratic Forms

Suppose Σ is an $n \times n$ positive definite matrix.

Suppose A is an $n \times n$ symmetric matrix of rank m such that $A\Sigma$ is idempotent (i.e., $A\Sigma A\Sigma = A\Sigma$).

Then $\mathbf{y} \sim N(\boldsymbol{\mu}, \Sigma) \implies \mathbf{y}'A\mathbf{y} \sim \chi_m^2(\boldsymbol{\mu}'A\boldsymbol{\mu}/2)$.

Mean and Variance of Chi-Square Distributions

If $w \sim \chi_m^2(\theta)$, then

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

Non-Central t Distributions

Suppose $y \sim N(\delta, 1)$.

Suppose $w \sim \chi_m^2$.

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 $z \sim N(0, 1)$.

Suppose $w \sim \chi_m^2$.

Suppose 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 $w_1 \sim \chi_{m_1}^2(\theta)$.

Suppose $w_2 \sim \chi_{m_2}^2$.

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 $w_1 \sim \chi_{m_1}^2$.

Suppose $w_2 \sim \chi_{m_2}^2$.

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

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

Some Independence (\perp) Results

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

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