Preliminaries

Notation for 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: $\boldsymbol{A}, \quad \boldsymbol{X}, \quad \boldsymbol{\Sigma}.$

Transpose

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

The transpose of any matrix $\boldsymbol{A}$ is written as $\boldsymbol{A}'$.

Thus, $\boldsymbol{x}$ is a column vector and $\boldsymbol{x}'$ is a row vector.

Matrix Multiplication

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

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

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

Transpose of a Matrix Product

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

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

Rank and Trace

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

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

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

Idempotent Matrices

A matrix $\boldsymbol{A}$ is said to be *idempotent* if $\boldsymbol{AA} = \boldsymbol{A}$.

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

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

Linear Combinations and Column Spaces

- $\boldsymbol{Ab}$ is a *linear combination* of the columns of an $m \times n$ matrx $\boldsymbol{A}$:

$$\boldsymbol{Ab} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n] \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = b_1\boldsymbol{a}_1 + \cdots + b_n\boldsymbol{a}_n.$$

- 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

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

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

Inverse of a Matrix

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

Generalized Inverses

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

- 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 $\boldsymbol{A}$.

$$\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{I} = \boldsymbol{I}\boldsymbol{A} = \boldsymbol{A}$$

- 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 $\boldsymbol{A}$

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

Positive and Non-Negative Definite Matrices

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

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

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

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

- $\boldsymbol{A}$ is symmetric (i.e., $\boldsymbol{A} = \boldsymbol{A}'$), and
- $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$ 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.

Random Vectors

A *random vector* 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 *expected value*, or *mean*, of a random vector $\boldsymbol{y}$ is the vector of expected values of the components of $\boldsymbol{y}$.

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

Variance of a Random Vector

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

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

Variance of a Random Vector

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

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

Covariance Between Two Random Vectors

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

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

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 $\boldsymbol{y}$.

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

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

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

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

Standard Multivariate Normal Distributions

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

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

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

Multivariate Normal Distributions

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

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

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

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

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

Linear Transformations of Multivariate Normal Distributions are Multivariate Normal

$$
\begin{aligned}
\boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad &\Longrightarrow \quad \boldsymbol{y} \overset{d}{=} \boldsymbol{Az} + \boldsymbol{\mu},\ \boldsymbol{z} \sim N(\boldsymbol{0}, \boldsymbol{I}),\ \boldsymbol{AA}' = \boldsymbol{\Sigma} \\
&\Longrightarrow \quad \boldsymbol{Cy} + d \overset{d}{=} \boldsymbol{C}(\boldsymbol{Az} + \boldsymbol{\mu}) + \boldsymbol{d} \\
&\Longrightarrow \quad \boldsymbol{Cy} + d \overset{d}{=} \boldsymbol{CAz} + \boldsymbol{C\mu} + \boldsymbol{d} \\
&\Longrightarrow \quad \boldsymbol{Cy} + d \overset{d}{=} \boldsymbol{Mz} + \boldsymbol{u},\ \boldsymbol{M} \equiv \boldsymbol{CA},\ \boldsymbol{u} \equiv \boldsymbol{C\mu} + \boldsymbol{d} \\
&\Longrightarrow \quad \boldsymbol{Cy} + d \sim N(\boldsymbol{u}, \boldsymbol{MM}').
\end{aligned}
$$

Non-Central Chi-Square Distributions

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

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

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

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

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

Central Chi-Square Distributions

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

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

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

$$w \sim \chi^2_n.$$

A central chi-square distribution is a non-central chi-square 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 $\boldsymbol{A}$ is an $n \times n$ symmetric matrix of rank m such that $\boldsymbol{A}\boldsymbol{\Sigma}$ is idempotent (i.e., $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{\Sigma}$).

Then $\boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Longrightarrow \boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} \sim \chi^2_m(\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}/2)$.

Mean and Variance of Chi-Square Distributions

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

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

Non-Central t Distributions

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

Suppose $w \sim \chi^2_m$.

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 $w \sim \chi^2_m$.

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 $w_1 \sim \chi^2_{m_1}(\theta)$.

Suppose $w_2 \sim \chi^2_{m_2}$.

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 $w_1 \sim \chi^2_{m_1}$.

Suppose $w_2 \sim \chi^2_{m_2}$.

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 $\boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 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 $\boldsymbol{A}_1\boldsymbol{\Sigma}\boldsymbol{A}_2' = \boldsymbol{0} \implies \boldsymbol{A}_1\boldsymbol{y} \perp \boldsymbol{A}_2\boldsymbol{y}$.
- 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 $\boldsymbol{A}_1\boldsymbol{\Sigma}\boldsymbol{A}_2 = \boldsymbol{0} \implies \boldsymbol{A}_1\boldsymbol{y} \perp \boldsymbol{y}'\boldsymbol{A}_2\boldsymbol{y}$.
- If $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are $n \times n$ symmetric matrices of constants, then $\boldsymbol{A}_1\boldsymbol{\Sigma}\boldsymbol{A}_2 = \boldsymbol{0} \implies \boldsymbol{y}'\boldsymbol{A}_1\boldsymbol{y} \perp \boldsymbol{y}'\boldsymbol{A}_2\boldsymbol{y}$.