Lecture 1 Review:

Linear models have the form (in matrix notation)

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{Y}^{n\times 1}$ response vector and $\mathbf{X}^{n\times p}$ is the model matrix (or “design matrix”) with one row for every subject and one column for every regression parameter. $\boldsymbol{\beta}^{p\times 1}$ is the vector of unknown regression parameters and $\boldsymbol{\varepsilon}^{n\times 1}$ is mean zero random error vector. We have

$$E[\mathbf{Y}|\mathbf{X}] = \sum_{j=0}^{p-1} \beta_j \mathbf{x}_j.$$

- $E[\mathbf{Y}|\mathbf{X}]$ is a linear combination of the $\mathbf{x}_j$ $(j = 0, \ldots, p-1)$,
- $E[\mathbf{Y}|\mathbf{X}] \in span(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{p-1}) \equiv \Omega$,

This class of models includes:

1. Regression models,
2. Anova models,
3. Ancova models.

The theory of linear models requires understanding basic facts and results in linear algebra and matrix analysis. We review these in this lecture. There may be bits and pieces you are not familiar with, which is fine, learn it now. If this is all new to you, you do not have the prerequisite for this class.

Notation and Elementary Properties:

1. Matrix: an $m \times n$ matrix with elements a_{ij} is denoted $\mathbf{A} = (a_{ij})_{m\times n}$.
2. Vector: a vector of length n is denoted $\mathbf{a} = (a_i)_n$. If all elements equal 1 it is denoted $\mathbf{1}_n$. We will stick to the convention that a vector is a column vector.
3. Diagonal Matrix:

$$\mathrm{diag}(a_1, \ldots, a_n) \equiv \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & a_n \end{pmatrix}.$$

4. Identity Matrix: $\mathbf{I}_{n\times n} \equiv \mathrm{diag}(\mathbf{1}_n)$.
5. Matrix Transpose: if $\mathbf{A} = (a_{ij})_{m\times n}$, then $\mathbf{A}'$ is an $n \times m$ matrix where $a'_{ij} = a_{ji}$.
6. If $\mathbf{A} = \mathbf{A}'$ then $\mathbf{A}$ is *symmetric*.
7. Matrix Sum: if $\mathbf{A} = (a_{ij})_{m\times n}$ and $\mathbf{B} = (b_{ij})_{m\times n}$,

$$\mathbf{A} + \mathbf{B} = (a_{ij} + b_{ij})_{m\times n}.$$

   Matrix sums satisfy

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

8. Matrix Product: if $\mathbf{A} = (a_{ij})_{m \times n}$ and $\mathbf{B} = (b_{ij})_{n \times p}$, then

$$\mathbf{AB} = (c_{ij})_{m \times p}, \quad c_{ij} = \sum_k a_{ik} b_{kj}.$$

   Matrix products satisfy

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

   not $\mathbf{A}'\mathbf{B}'$!

   Proof: (just think about how it works)

$$\left\{ \begin{pmatrix} & \vdots & \\ a_{i1} & \cdots & a_{in} \\ & \vdots & \end{pmatrix} \begin{pmatrix} & b_{1j} & \\ \cdots & \vdots & \cdots \\ & b_{nj} & \end{pmatrix} \right\}' = \begin{pmatrix} & \vdots & \\ b_{j1} & \cdots & b_{jn} \\ & \vdots & \end{pmatrix} \begin{pmatrix} & a_{1i} & \\ \cdots & \vdots & \cdots \\ & a_{1i} & \end{pmatrix}.$$

9. Matrix Trace: Let $\mathbf{A} = (a_{ij})_{m \times n}$. The trace of $\mathbf{A}$ is the sum of the diagonal elements,

$$\mathrm{tr}(\mathbf{A}) \equiv \sum_i a_{ii},$$

   where the sum is over $i \leq \min(m, n)$. If $\mathbf{A} = (a_{ij})_{m \times n}$ and $\mathbf{B} = (b_{ij})_{m \times n}$, then

$$\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B}).$$

   If $\mathbf{A}$ and $\mathbf{B}$ are square matrices, then

$$\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA}).$$

   Even though $\mathbf{AB} \neq \mathbf{BA}$.

SOME COMMENTS

When matrices are the right size, we can add and subtract them. As long as matrices have the same size, this works pretty much like adding and subtracting real numbers.

When matrices have the right size, we can multiply them. Sometimes this is like multiplying numbers, but sometimes it is different.

Like real numbers, multiplication has the associative property: $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$. Because of this, it is o.k. to just write $\mathbf{ABC}$.

Multiplication is not commutative. If you have $\mathbf{AB}$, this is not the same as $\mathbf{BA}$, which may not even be defined.

You cannot "cancel out" matrices like you can real numbers. If you have an equation $\mathbf{AB} = \mathbf{AC}$, you cannot in general conclude that $\mathbf{B} = \mathbf{C}$.

Linear Independence, Range, Rank, and Null Space:

1. **Linear Independence**: vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are *linearly independent* if $\sum_i c_i \mathbf{a}_i \neq 0$ unless $c_i = 0$ for all i.

2. **Range (Column Space)**: $\mathcal{R}(\mathbf{A}) \equiv$ the linear space spanned by the columns of $\mathbf{A}$.

3. **Rank**: $\text{rank}(\mathbf{A}) \equiv \text{r}(\mathbf{A}) \equiv$ the number of linearly independent columns of $\mathbf{A}$ (i.e., the dimension of $\mathcal{R}(\mathbf{A})$), or, equivalently, the number of linearly independent rows of $\mathbf{A}$.

   Examples:

   $$\text{rank}\begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} = 1,$$

   $$\mathcal{R}\begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} = \{c\begin{pmatrix} 1 \\ 1 \end{pmatrix} : -\infty < c < \infty\},$$

   $$\text{rank}\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = 2.$$

   COMMENTARY: Linear independence and related notions (e.g., rank) are deep and important concepts.

4. **Decreasing property of rank (Seber & Lee A2.1)**:

   $$\text{rank}(\mathbf{AB}) \leq \min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\}.$$

5. Null Space: $\mathcal{N}(\mathbf{A}) \equiv \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{0}\}$. The *nullity* of $\mathbf{A}$ is the dimension of $\mathcal{N}(\mathbf{A})$. We have

$$\text{rank}(\mathbf{A}) + \text{nullity}(\mathbf{A}) = n,$$

the number of columns of $\mathbf{A}$ (Seber & Lee A2.3).

6. $\text{r}(\mathbf{A}) = \text{r}(\mathbf{A}') = \text{r}(\mathbf{A}'\mathbf{A}) = \text{r}(\mathbf{A}\mathbf{A}')$ (Seber & Lee A2.4).

Matrix Inverse:

1. Definition: An $n \times n$ matrix $\mathbf{A}$ is *invertible* (or *non-singular*) if there is a matrix $\mathbf{A}^{-1}$ such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_{n \times n}.$$

$\mathbf{A}$ $(n \times n)$ is *invertible* if and only if $\text{rank}(\mathbf{A}) = n$.

2. Inverse of Product: $(\mathbf{AB})^{-\mathbf{1}} = \mathbf{B}^{-\mathbf{1}}\mathbf{A}^{-\mathbf{1}}$ if $\mathbf{A}$ and $\mathbf{B}$ are invertible. (Easy to prove: multiply $\mathbf{AB}$ by $\mathbf{B}^{-1}\mathbf{A}^{-1}$ on the left or right)

Generalized Inverses:

1. Definition: Let $\mathbf{A}$ be an $m \times n$ matrix. A *generalized inverse* of $\mathbf{A}$ is any matrix $\mathbf{G}$ such that $\mathbf{AGA} = \mathbf{A}$.

2. Generalized inverses always exist. Generalized inverses are not unique, except for square, non-singular matrices.

3. Notation: Write $\mathbf{A}^{-}$ for a generalized inverse of $\mathbf{A}$.

Vectors: Inner Product, Length, and Orthogonality:

1. Inner product: $\mathbf{a}'\mathbf{b} = \sum_i a_i b_i$, where $\mathbf{a} = (a_i)$, $\mathbf{b} = (b_i)$ are vectors with the same length.
2. Vector norm (length): $||\mathbf{a}|| = \sqrt{\mathbf{a}'\mathbf{a}}$.
3. Orthogonal vectors: $\mathbf{a} = (a_i)$ and $\mathbf{b} = (b_i)$ are orthogonal if $\mathbf{a}'\mathbf{b} = 0$.
4. Orthogonal matrix: $\mathbf{A}$ is orthogonal if its columns are orthogonal. If $\mathbf{A}$ is orthogonal, $\mathbf{A}'\mathbf{A}$ is diagonal. A square matrix $\mathbf{A}$ is orthonormal if its columns are orthogonal vectors of length 1, so that $\mathbf{A}^{-1} = \mathbf{A}'$. I may be sloppy and use "orthogonal" when I mean "orthonormal."

Determinants:

Definition (recursive): for a square matrix $\mathbf{A}$, $|\mathbf{A}| \equiv \sum_i a_{ij} A_{ij}$, where the cofactor $A_{ij} = (-1)^{i+j}|\mathbf{M}_{ij}|$, and $\mathbf{M}_{ij}$ is the matrix obtained by deleting the ith row and jth column from $\mathbf{A}$.

Properties of Determinants:

1. $\left|\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right| = ad - bc.$
2. $|\mathbf{A}| = 0,$ if and only if $\mathbf{A}$ is singular.
3. $|\text{diag}(a_1, \ldots, a_n)| = \prod_i a_i.$
4. $|\mathbf{AB}| = |\mathbf{A}| \cdot |\mathbf{B}|.$
5. $\left|\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}\right| = |\mathbf{A}| \cdot |\mathbf{C}|.$

   Proof: by induction on the order of $\mathbf{A}$

Eigenvalues:

Definition: If $\mathbf{Ax} = \lambda\mathbf{x}$ where $\mathbf{x} \neq 0$, then λ is an *eigenvalue* of $\mathbf{A}$ and $\mathbf{x}$ is a corresponding *eigenvector*.

Properties: For any symmetric matrix $\mathbf{A}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$,

1. (Spectral Theorem–a.k.a. Principal Axis Theorem) For any symmetric matrix $\mathbf{A}$ there is an orthonormal matrix $\mathbf{T}$ such that:
$$\mathbf{T}'\mathbf{AT} = \mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n).$$
Terminology: $\mathbf{T}$ "diagonalizes" $\mathbf{A}$.

2. $\mathrm{r}(\mathbf{A}) =$ the number of non-zero λ_i
Proof: $\mathrm{r}(\mathbf{A}) = \mathrm{r}(\mathbf{ATT}') \leq \mathrm{r}(\mathbf{AT}) \leq \mathrm{r}(\mathbf{A})$.
Therefore, $\mathrm{r}(\mathbf{A}) = \mathrm{r}(\mathbf{AT})$.
Similarly, $\mathrm{r}(\mathbf{AT}) = \mathrm{r}(\mathbf{TT}'\mathbf{AT}) \leq \mathrm{r}(\mathbf{T}'\mathbf{AT}) \leq \mathrm{r}(\mathbf{AT})$.
So $\mathrm{r}(\mathbf{T}'\mathbf{AT}) = \mathrm{r}(\mathbf{AT}) = \mathrm{r}(\mathbf{A})$.
But $\mathrm{r}(\mathbf{T}'\mathbf{AT}) = \mathrm{r}(\mathbf{\Lambda})$.

3. $\mathrm{tr}(\mathbf{A}) = \sum_i \lambda_i$.
Proof: $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{ATT}') = \mathrm{tr}(\mathbf{T}'\mathbf{AT}) = \mathrm{tr}(\mathbf{\Lambda}) = \sum_i \lambda_i$.

4. $|\mathbf{A}| = \prod_i \lambda_i$.

Positive Definite and Semidefinite Matrices:

Definition: A symmetric matrix $\mathbf{A}$ is called *positive semidefinite* (p.s.d.) if $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for all non-zero $\mathbf{x}$.

Properties of a p.s.d matrix $\mathbf{A}$:

1. The diagonal elements a_{ii} are all non-negative.

   Proof: Let $\mathbf{x} = (1, 0, \ldots, 0)'$. Then $0 \leq \mathbf{x}'\mathbf{A}\mathbf{x} = a_{11}$. Similarly for the other a_{ii}.

2. All eigenvalues of $\mathbf{A}$ are nonnegative.

   Proof: Let $\mathbf{y} = (1, 0, \ldots, 0)'$ and $\mathbf{x} = \mathbf{T}\mathbf{y}$ (where $\mathbf{T}$ is the orthonormal matrix that diagonalizes $\mathbf{A}$). Then

   $$0 \leq \mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{y}'\mathbf{T}'\mathbf{A}\mathbf{T}\mathbf{y} = \mathbf{y}'\mathbf{\Lambda}\mathbf{y} = \lambda_1.$$

   Similarly for the other λ_i.

3. $\text{tr}(\mathbf{A}) \geq 0$. Follows from 1 above, since trace is the sum of the a_{ii}; or follows from 2 above because the trace is the sum of the eigenvalues.

Definition: A symmetric matrix $\mathbf{A}$ is called *positive definite* (p.d.) if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all non-zero $\mathbf{x}$.

Properties of a p.d matrix $\mathbf{A}$:

1. All diagonal elements and all eigenvalues of $\mathbf{A}$ are positive.
2. $\text{tr}(\mathbf{A}) > 0$.
3. $|\mathbf{A}| > 0$.
4. There is a nonsingular $\mathbf{R}$ such that $\mathbf{A} = \mathbf{R}\mathbf{R}'$ (necessary and sufficient for $\mathbf{A}$ to be p.d., Seber &Lee A4.2).
5. $\mathbf{A}^{-1}$ is p.d.
   Proof: $\mathbf{A}^{-1} = (\mathbf{R}\mathbf{R}')^{-1} = (\mathbf{R}')^{-1}\mathbf{R}^{-1} = \mathbf{S}\mathbf{S}'$, where $\mathbf{S} = (\mathbf{R}^{-1})'$ is non-singular.

NOTE: We will sometimes note that $\mathbf{A}$ is p.s.d. by writing $\mathbf{A} \geq 0$ and note that $\mathbf{A}$ is p.d. by writing $\mathbf{A} > 0$.

Idempotent and Projection Matrices:

Definitions: A matrix $\mathbf{P}$ is *idempotent* if $\mathbf{P}^2 = \mathbf{P}$. A symmetric idempotent matrix is called a *projection matrix*.

Facts about projection matrices $\mathbf{P}$:

1. Let $\mathbf{P}$ be a symmetric matrix. $\mathbf{P}$ is idempotent and of rank r if and only if it has r eigenvalues equal to 1 and $n - r$ eigenvalues equal to zero. (Seber & Lee A6.1)

   Proof: ($\Rightarrow$) Suppose $\mathbf{P}^2 = \mathbf{P}$ and rank$(\mathbf{P}) = r$. Then $\mathbf{T}\Lambda\mathbf{T}' = \mathbf{P} = \mathbf{P}^2 = \mathbf{T}\Lambda\mathbf{T}'\mathbf{T}\Lambda\mathbf{T}' = \mathbf{T}\Lambda^2\mathbf{T}' \Rightarrow \Lambda = \Lambda^2$. Λ is a diagonal matrix $\Rightarrow \lambda_i = 0$ or 1 $(i = 1, 2, \ldots, n)$. By Seber & Lee `A.2.6`, rank(Λ)=rank$(\mathbf{P}) = r$.

   ($\Leftarrow$) Suppose $\lambda_1 = \lambda_2 = \cdots = \lambda_r = 1$ and $\lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0$. Then there exists an orthogonal matrix $\mathbf{T}$ such that $\mathbf{T}'\mathbf{P}\mathbf{T} = diag(\mathbf{1}_r, \mathbf{0}_{n-r}) = \Lambda$.
   $\mathbf{P} = \mathbf{T}\Lambda\mathbf{T}'$, $\mathbf{P}^2 = \mathbf{T}\Lambda\mathbf{T}'\mathbf{T}\Lambda\mathbf{T}' = \mathbf{T}\Lambda^2\mathbf{T}' = \mathbf{P}$, and rank$(\mathbf{P})$= rank$(\Lambda) = r$.

2. Projection matrices have tr$(\mathbf{P})$ = rank$(\mathbf{P})$.

3. Projection matrices are positive semidefinite.

   Proof: $\mathbf{x}'\mathbf{P}\mathbf{x} = \mathbf{x}'\mathbf{P}^2\mathbf{x} = (\mathbf{P}\mathbf{x})'(\mathbf{P}\mathbf{x}) = ||\mathbf{P}|| \geq 0.$

More Projections:

For two vectors $\mathbf{x}$ and $\mathbf{y}$, the *projection* of $\mathbf{y}$ onto $\mathbf{x}$ is

$$\text{Proj}_{\mathbf{x}}(\mathbf{y}) = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}}\mathbf{x}.$$

If V is a vector space and Ω is a subspace of V, then $\exists$ two vectors, $\mathbf{w}_1$, $\mathbf{w}_2 \in V$ such that

1. $\mathbf{y} = \mathbf{w}_1 + \mathbf{w}_2 \quad \forall\, \mathbf{y} \in V$,
2. $\mathbf{w}_1 \in \Omega$ and $\mathbf{w}_2 \in \Omega^{\perp}$.
3. If $\|\mathbf{y} - \mathbf{w}_1\| \leq \|\mathbf{y} - \mathbf{x}\|$ for any $\mathbf{x} \in \Omega$, then $\mathbf{w}_1$ is the *projection* of $\mathbf{y}$ onto Ω.

   $\diamond$ The transformation that takes $\mathbf{y}$ onto $\mathbf{w}_1$ is a linear transformation.

   $\diamond$ The matrix $\mathbf{P}$ that takes $\mathbf{y}$ onto $\mathbf{w}_1$ (i.e., $\mathbf{Py}$=$\mathbf{w}_1$) is a projection matrix. $\mathbf{P}$ projects $\mathbf{y}$ onto the space spanned by the column vectors of $\mathbf{P}$.

   $\diamond$ $\mathbf{I} - \mathbf{P}$ is a projection operator onto $\Omega^{\perp}$.