STAT 714 fa 2025 Lec 01

Least squares estimation in linear models

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

1 Projection and idempotent matrices

2 Generalized inverses

3 Least-squares geometry

Idempotent matrix

A square matrix $\mathbf{A}$ is called *idempotent* if $\mathbf{A}^2 = \mathbf{A}$.

Exercise: Verify that these are idempotent matrices:

$$\begin{bmatrix} 3 & -2 \\ 3 & -2 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Projection matrix

A square matrix $\mathbf{P}$ is called a *projection matrix* onto the space V if

1. $\mathbf{P}$ is idempotent
2. for any $\mathbf{x}$, $\mathbf{Px} \in V$
3. for any $\mathbf{z} \in V$, $\mathbf{Pz} = \mathbf{z}$

Sometimes we call projection matrices simply "projections".

Theorem (Every idempotent matrix is a projection)

Every idempotent matrix is a projection onto its own column space.

Prove the result.

Exercise: Let

$$\mathbf{P} = \begin{bmatrix} 3 & -2 \\ 3 & -2 \end{bmatrix}, \quad \tilde{\mathbf{P}} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}, \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

1. Find $\mathbf{P}\mathbf{v}$, $\tilde{\mathbf{P}}\mathbf{v}$, $(\mathbf{I} - \mathbf{P})\mathbf{v}$, and $(\mathbf{I} - \tilde{\mathbf{P}})\mathbf{v}$.
2. Give the spaces onto which $\mathbf{P}$, $\tilde{\mathbf{P}}$, $(\mathbf{I} - \mathbf{P})$, and $(\mathbf{I} - \tilde{\mathbf{P}})$ project.
3. Draw pictures.

We like projections that let us orthogonally decompose any vector $\mathbf{x}$ as

$$\mathbf{x} = \mathbf{P}\mathbf{x} + (\mathbf{I} - \mathbf{P})\mathbf{x}, \quad \text{where} \quad \mathbf{P}\mathbf{x} \cdot (\mathbf{I} - \mathbf{P})\mathbf{x} = 0.$$

Orthogonal projection

Let $\mathbf{P}$ be a projection matrix onto a subspace V. The projection is an *orthogonal projection* if $(\mathbf{I} - \mathbf{P})$ is the projection matrix onto $V^{\perp}$.

Discuss: Which projection matrix corresponds to an orthogonal projection?

$$\mathbf{P} = \begin{bmatrix} 3 & -2 \\ 3 & -2 \end{bmatrix}, \quad \tilde{\mathbf{P}} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}.$$

Theorem (Symmetric, idempotent $\iff$ orthogonal projection)

A matrix $\mathbf{P}$ *is symmetric and idempotent iff it is an orthogonal projection matrix.*

Prove the result.

Theorem (Uniqueness of orthogonal projection matrices)

If $\mathbf{P}_1$ *and* $\mathbf{P}_2$ *are orthogonal projections onto the same subspace then* $\mathbf{P}_1 = \mathbf{P}_2$*.*

Prove the result.

1 Projection and idempotent matrices

2 Generalized inverses

3 Least-squares geometry

Generalized inverse of a matrix

A matrix $\mathbf{G}$ which satisfies $\mathbf{AGA} = \mathbf{A}$ is called a *generalized inverse* of $\mathbf{A}$.

Theorem (Generalized inverses for solving systems of equations)

Suppose $\mathbf{Ax} = \mathbf{b}$ *is consistent and let* $\mathbf{G}$ *be a generalized inverse of* $\mathbf{A}$. *Then*

1. $\mathbf{Gb}$ *is a solution to* $\mathbf{Ax} = \mathbf{b}$.
2. $\hat{\mathbf{x}}$ *is a solution to* $\mathbf{Ax} = \mathbf{b}$ *iff there exists* $\mathbf{z}$ *such that* $\hat{\mathbf{x}} = \mathbf{Gb} + (\mathbf{I} - \mathbf{GA})\mathbf{z}$.

See Res A.12 and A.13 of Monahan (2008).

Prove the results.

Theorem (Generalized inverse recipe using block structure)

Let $\mathbf{A}$ *be an* $m \times n$ *matrix with rank* r*. If we can partition* $\mathbf{A}$ *as*

$$\mathbf{A} = \begin{bmatrix} \mathbf{C} & \mathbf{D} \\ \mathbf{E} & \mathbf{F} \end{bmatrix}, \quad \text{with } \mathbf{C} \ r \times r \text{ invertible, then} \quad \mathbf{G} = \begin{bmatrix} \mathbf{C}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}$$

is a generalized inverse of $\mathbf{A}$*, where the* $\mathbf{O}$ *matrices have the necessary dimensions.*

Similarly, if $\mathbf{F}$ *is* $r \times r$ *invertible, then* $\mathbf{G} = \begin{bmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{F}^{-1} \end{bmatrix}$ *is a gen. inverse of* $\mathbf{A}$*.*

See Res A.10 and Cor A.3 of Monahan (2008).

Make gen. inv. of *any* matrix by permuting rows/columns to get such a partition.
See Res A.11 of Monahan (2008).

Prove the first result.

We sometimes denote the generalized inverse of a matrix $\mathbf{A}$ by $\mathbf{A}^-$.

Theorem (Projections constructed with a generalized inverse)

Let $\mathbf{A}^-$ be a generalized inverse of $\mathbf{A}$. Then

1. $\mathbf{A}\mathbf{A}^-$ *is a projection onto* $\operatorname{Col}\mathbf{A}$.
2. $(\mathbf{I} - \mathbf{A}^-\mathbf{A})$ *is a projection onto* $\operatorname{Nul}\mathbf{A}$.

See Res A.14 and A.15 of Monahan (2008).

Prove the result.

1 Projection and idempotent matrices

2 Generalized inverses

3 Least-squares geometry

For the remainder of the lecture, let $\mathbf{X}$ be an $n \times p$ matrix and $\mathbf{y}$ be a vector in $\mathbb{R}^n$.

We have in mind data coming from a model like

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

. . . but we are not thinking yet about the distribution of $\mathbf{e}$.

We consider least-squares "estimation" of $\mathbf{b}$, but no statistics yet—only geometry.

Least-squares solution

A *least-squares solution* to $\mathbf{X}\mathbf{b} = \mathbf{y}$ is a vector $\hat{\mathbf{b}} \in \mathbb{R}^p$ such that

$$\|\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}\| \leq \|\mathbf{y} - \mathbf{X}\mathbf{b}\| \quad \text{for all } \mathbf{b} \in \mathbb{R}^p.$$

Theorem (Least-squares solution iff solution to normal equations)

1. *The equation* $\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{y}$ *is consistent.*
2. $\|\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}\| \leq \|\mathbf{y} - \mathbf{X}\mathbf{b}\|$ *for all* $\mathbf{b} \in \mathbb{R}^p$ *if and only if* $\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}$.

See Cor 2.1 and Res 2.3 of Monahan (2008).

We call the set of equations $\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{y}$ the *normal equations*.

Prove the result.

Can also use calculus to obtain the normal equations...

For a real-valued function $Q(\mathbf{x})$ taking vectors in $\mathbb{R}^n$, define

$$\frac{\partial}{\partial \mathbf{x}} Q(\mathbf{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1} Q(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial x_n} Q(\mathbf{x}) \end{bmatrix}.$$

Theorem (Derivative of linear and quadratic forms)

For a vector $\mathbf{a}$ *and a matrix* $\mathbf{A}$, *we have*

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{a}^T \mathbf{x} = \mathbf{a} \quad \textit{and} \quad \frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{x} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x}.$$

A least-squares solution of $\mathbf{Xb} = \mathbf{y}$ is a minimizer of $Q(\mathbf{b}) = \|\mathbf{y} - \mathbf{Xb}\|^2$.

Exercise: Use fact that $\hat{\mathbf{b}}$ minimizes $Q(\mathbf{b})$ iff $\frac{\partial}{\partial \mathbf{b}} Q(\mathbf{b})\Big|_{\mathbf{b}=\hat{\mathbf{b}}} = \mathbf{0}$ to get normal eqs.

Theorem (Characterization of solutions to the normal equations)

The vector $\hat{\mathbf{b}}$ *is a solution to* $\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{y}$ *iff there exists a vector* $\mathbf{z}$ *such that*

$$\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y} + (\mathbf{I} - (\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{X})\mathbf{z}.$$

If $\mathbf{X}$ *has full-column rank, then* $\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ *is the unique solution.*

Prove the result.

Exercise: Characterize the set of solutions to the normal equations when

$$\mathbf{y} = \begin{bmatrix} 1 \\ -1 \\ 2 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

These are helper results for constructing the orthogonal projection onto Col $\mathbf{X}$.

Theorem ("Cool result" and generalized inverse of $\mathbf{X}$)

1. $\mathbf{X}^T\mathbf{X}\mathbf{A} = \mathbf{X}^T\mathbf{X}\mathbf{B} \iff \mathbf{X}\mathbf{A} = \mathbf{X}\mathbf{B}$.
2. $(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T$ *is a generalized inverse of* $\mathbf{X}$.

See Res 2.4 and 2.5 of Monahan (2008).

Prove the results.

Theorem (Orthogonal projection onto $\text{Col}\,\mathbf{X}$)

The matrix $\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T$ *and the matrix* $\mathbf{I} - \mathbf{P}_{\mathbf{X}}$ *are*

1. *projections onto* $\text{Col}\,\mathbf{X}$ *and* $\text{Nul}\,\mathbf{X}^T$, *respectively*
2. *invariant to the choice of generalized inverse*
3. *symmetric (therefore unique)*

See Thm 2.4 and Res 2.6 of Monahan (2008).

Prove the results.

Result

We have $\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{y}$ *if and only if* $\mathbf{X}\mathbf{b} = \mathbf{P}_{\mathbf{X}}\mathbf{y}$.

See Res 2.7 of Monahan (2008).

Prove the result.

Sums of squares

For $\hat{\mathbf{b}}$ satisfying $\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}$ we define the

1. *fitted values* as $\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}}$
2. *residuals* as $\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}}$
3. *total sum of squares (SST)* as $\|\mathbf{y}\|^2$
4. *regression sum of squares (SSR)* as $\|\hat{\mathbf{y}}\|^2$
5. *error sum of squares (SSE)* as $\|\hat{\mathbf{e}}\|^2$.

Theorem (Sum of squares decomposition)

We have SST $=$ SSR $+$ SSE, *or* $\|\mathbf{y}\|^2 = \|\hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{e}}\|^2$.

Prove the result.

Monahan, J. F. (2008). *A primer on linear models*. CRC Press.