Stat 206: Linear Models

Lecture 5

October 12, 2015

Simple Linear Regression in Matrix Form

The regression equations:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, \cdots n$$

can be written in a compact matrix form:

$$\underset{n\times1}{\mathbf{Y}} = \underset{n\times2}{\mathbf{X}}\,\underset{2\times1}{\boldsymbol{\beta}} + \underset{n\times1}{\boldsymbol{\epsilon}}.$$

- **Response vector Y** and **error vector** : $n \times 1$ column vectors

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_i \\ \vdots \\ Y_n \end{bmatrix}, \qquad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_n \end{bmatrix}$$

- **Design matrix**: an $n \times 2$ matrix:

$$\mathbf{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_i \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}.$$

- **Coefficient vector**: a 2×1 column vector:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}.$$

- Model assumptions:

$$E(\epsilon_i) = 0, \quad Var(\epsilon_i) = \sigma^2, \text{ for all } i = 1, \cdots, n$$

$$Cov(\epsilon_i, \epsilon_j) = 0, \quad \text{for all } i \neq j.$$

- Matrix form:

$$\mathbf{E}\{\boldsymbol{\epsilon}\} = \mathbf{0}_n, \quad \boldsymbol{\sigma}^2\{\boldsymbol{\epsilon}\} = \sigma^2 \mathbf{I}_n.$$

- In terms of the response vector:

$$\mathbf{E}\{\mathbf{Y}\} = \mathbf{X}\boldsymbol{\beta}, \quad \boldsymbol{\sigma}^2\{\mathbf{Y}\} = \sigma^2 \mathbf{I}_n.$$

- $\mathbf{0}_n$ is the $n \times 1$ zero vector, $\mathbf{I}_n$ is the $n \times n$ identity matrix.
- Mean of the error vector:

$$\mathbf{E}\{\boldsymbol{\epsilon}\} := \begin{bmatrix} E(\epsilon_1) \\ E(\epsilon_2) \\ \vdots \\ E(\epsilon_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}_n.$$

Variance-covariance matrix of the error vector:

$$
\begin{aligned}
\boldsymbol{\sigma}^{\mathbf{2}}\{\boldsymbol{\epsilon}\}: \quad &= \begin{bmatrix} Var(\epsilon_1) & Cov(\epsilon_1, \epsilon_2) & \cdots & Cov(\epsilon_1, \epsilon_n) \\ Cov(\epsilon_2, \epsilon_1) & Var(\epsilon_2) & \cdots & Cov(\epsilon_2, \epsilon_n) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(\epsilon_n, \epsilon_1) & Cov(\epsilon_n, \epsilon_2) & \cdots & Var(\epsilon_n) \end{bmatrix} \\
&= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_n.
\end{aligned}
$$

Mean response vector: an $n \times 1$ column vector

$$\mathbf{E}\{\mathbf{Y}\} = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_i) \\ \vdots \\ E(Y_n) \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_i \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_i \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \mathbf{X}\boldsymbol{\beta}.$$

Simple Linear Regression in Matrix Form

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times 2}{\mathbf{X}} \underset{2\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\epsilon}}.$$

- $\boldsymbol{\epsilon}$ is a random vector with $\mathbf{E}\{\boldsymbol{\epsilon}\} = \mathbf{0}_n,\ \boldsymbol{\sigma}^{\mathbf{2}}\{\boldsymbol{\epsilon}\} = \sigma^2\mathbf{I}_n$.
- Normal error model: $\boldsymbol{\epsilon} \sim \mathrm{Normal}(\mathbf{0}_n, \sigma^2\mathbf{I}_n)$.

Least Squares Estimation in Matrix Form

- Least squares criterion:

$$Q(b_0, b_1) = \sum_{i=1}^{n} (Y_i - (b_0 + b_1 X_i))^2 .$$

- Matrix form :

$$Q(\mathbf{b}) = (\mathbf{Y} - \mathbf{X}\mathbf{b})' (\mathbf{Y} - \mathbf{X}\mathbf{b}) = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\mathbf{b} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}.$$

- Differentiate Q with respect to $\mathbf{b}$:

$$\frac{\partial}{\partial \mathbf{b}} Q = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\mathbf{b}.$$

- Set the gradient to zero $\Longrightarrow$ normal equation:

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}. \tag{1}$$

Least-square estimators are the solutions of equation (1).

- Multiply both sides of equation (1) by $(\mathbf{X}'\mathbf{X})^{-1}$:

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

- The left hand side becomes

$$\mathbf{I}_2\mathbf{b} = \mathbf{b}$$

- **LS estimators**:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \tag{2}$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{bmatrix}, \quad \mathbf{X}'\mathbf{Y} = \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{bmatrix}.$$

When

$$D := n\sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2 = n\sum_{i=1}^n (X_i - \overline{X})^2 \neq 0$$

$$\begin{aligned}
(\mathbf{X}'\mathbf{X})^{-1} &= \begin{bmatrix} \frac{\sum_{i=1}^n X_i^2}{n\sum_{i=1}^n (X_i - \overline{X})^2} & -\frac{\sum_{i=1}^n X_i}{n\sum_{i=1}^n (X_i - \overline{X})^2} \\ -\frac{\sum_{i=1}^n X_i}{n\sum_{i=1}^n (X_i - \overline{X})^2} & \frac{n}{n\sum_{i=1}^n (X_i - \overline{X})^2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^n (X_i - \overline{X})^2} & -\frac{\overline{X}}{\sum_{i=1}^n (X_i - \overline{X})^2} \\ -\frac{\overline{X}}{\sum_{i=1}^n (X_i - \overline{X})^2} & \frac{1}{\sum_{i=1}^n (X_i - \overline{X})^2} \end{bmatrix}.
\end{aligned}$$

Note that $\mathbf{X}'\mathbf{X}$ and $(\mathbf{X}'\mathbf{X})^{-1}$ are symmetric positive definite matrices.

- LS estimators:

$$\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix}\overline{Y} - \hat{\beta}_1\overline{X}\\ \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sum_{i=1}^{n}(X_i-\overline{X})^2}\end{bmatrix},$$

  provided that X_is are not all equal.
- $n \times 1$ vector of fitted values:

$$\widehat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y},$$

  where $\mathbf{H} := \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is called the **hat matrix**.
- $n \times 1$ vector of residuals:

$$\mathbf{e} = \mathbf{Y} - \widehat{\mathbf{Y}} = (\mathbf{I}_n - \mathbf{H})\mathbf{Y}.$$

- Fitted values vector $\widehat{\mathbf{Y}}$ and residuals vector $\mathbf{e}$ are linear transformations of the observations vector $\mathbf{Y}$.

Hat Matrix

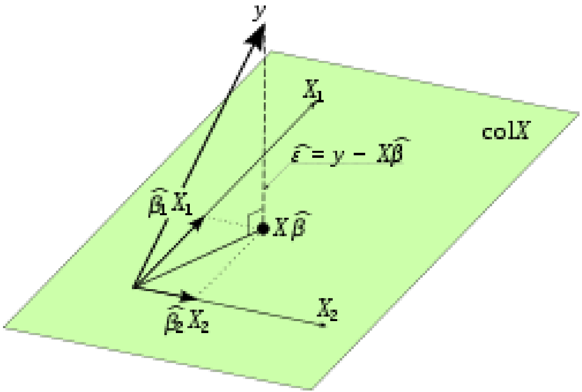
$$\underset{n\times n}{\mathbf{H}} := \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

- $\mathbf{H}$ and $\mathbf{I}_n - \mathbf{H}$ are **projection matrices**.
  - Symmetric:
  $$\mathbf{H}' = \mathbf{H}, \quad (\mathbf{I}_n - \mathbf{H})' = \mathbf{I}_n - \mathbf{H}$$
  - **Idempotent**:
  $$\mathbf{H}^2 := \mathbf{H}\mathbf{H} = \mathbf{H}, \quad (\mathbf{I}_n - \mathbf{H})^2 = \mathbf{I}_n - \mathbf{H}.$$
- Properties of projection matrices.
  - They have eigen-decomposition of the form: $Q\Lambda Q'$, where Q is an orthogonal matrix of eigenvectors and Λ is a diagonal matrix of eigenvalues.
  - Their eigenvalues are either 1 or 0. The number of nonzero eigenvalues equals to the rank.
  - $rank(\mathbf{H}) = 2$, $rank(\mathbf{I}_n - \mathbf{H}) = n - 2$.

- $\mathbf{H}$ projects a vector to the column space $\langle X \rangle$ of the design matrix $\mathbf{X}$, i.e., for any $\mathbf{x} \in \mathbf{R}^n$
  - $\mathbf{Hx} \in \langle X \rangle$.
  - $\mathbf{x} - \mathbf{H}x \perp < X >$.
- In particular:
  - $\widehat{\mathbf{Y}} \in \langle X \rangle$: the fitted values vector is in the column space of $\mathbf{X}$.
  - $\mathbf{e} \perp \langle X \rangle$: the residuals vector is orthogonal to the column space of $\mathbf{X}$.

Geometric Interpretation of Linear Regression

Figure: Orthogonal projection of response vector **Y** onto the linear subspace of $\mathbb{R}^n$ generated by the columns of the design matrix **X**.



Expectations

- LS estimators are unbiased estimators :

$$\mathbf{E}\{\hat{\beta}\} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}\{\mathbf{Y}\} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta = \beta.$$

- Expectation of the fitted values:

$$\mathbf{E}\{\widehat{\mathbf{Y}}\} = \mathbf{E}\{\mathbf{X}\hat{\beta}\} = \mathbf{X}\mathbf{E}\{\hat{\beta}\} = \mathbf{X}\beta = \mathbf{E}\{\mathbf{Y}\}.$$

- Expectation of the residuals:

$$\mathbf{E}\{\mathbf{e}\} = \mathbf{E}\{\mathbf{Y} - \widehat{\mathbf{Y}}\} = \mathbf{E}\{\mathbf{Y}\} - \mathbf{E}\{\widehat{\mathbf{Y}}\} = \mathbf{0}_n.$$

Variance-covariance Matrices

- Variance-covariance of the LS estimators:

$$\begin{aligned}\boldsymbol{\sigma^2}\{\hat{\boldsymbol{\beta}}\} &= \boldsymbol{\sigma^2}\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\} = \left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\boldsymbol{\sigma^2}\{\mathbf{Y}\}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)' \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2\begin{bmatrix} \frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^n (X_i-\overline{X})^2} & -\frac{\overline{X}}{\sum_{i=1}^n (X_i-\overline{X})^2} \\ -\frac{\overline{X}}{\sum_{i=1}^n (X_i-\overline{X})^2} & \frac{1}{\sum_{i=1}^n (X_i-\overline{X})^2}\end{bmatrix}\end{aligned}$$

What is the covariance between $\hat{\beta}_0$ and $\hat{\beta}_1$? What happens if $\overline{X} = 0$?

- Variance-covariance of fitted values:

$$\boldsymbol{\sigma^2}\{\widehat{\mathbf{Y}}\} = \mathbf{H}\boldsymbol{\sigma^2}\{\mathbf{Y}\}\mathbf{H}' = \sigma^2\mathbf{H}.$$

- Variance-covariance of residuals:

$$\boldsymbol{\sigma^2}\{\mathbf{e}\} = (\mathbf{I}_n - \mathbf{H})\boldsymbol{\sigma^2}\{\mathbf{Y}\}(\mathbf{I}_n - \mathbf{H})' = \sigma^2(\mathbf{I}_n - \mathbf{H}).$$

Are residuals uncorrelated? Do they have the same variance?

Sum of Squares in Matrix Form

Error sum of squares:

$$SSE = \sum_{i=1}^{n} e_i^2.$$

- Matrix form:

$$SSE = \mathbf{e}'\mathbf{e} = \mathbf{Y}'(\mathbf{I}_n - \mathbf{H})'(\mathbf{I}_n - \mathbf{H})\mathbf{Y} = \mathbf{Y}'(\mathbf{I}_n - \mathbf{H})\mathbf{Y}.$$

- Recall that $\mathbf{I}_n - \mathbf{H}$ is a projection matrix.
- $df(SSE) = rank(\mathbf{I}_n - \mathbf{H}) = n - 2$.

Total sum of squares:

$$SSTO = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n(\overline{Y})^2.$$

- Matrix form:

$$SSTO = \mathbf{Y}'\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\mathbf{J}_n\mathbf{Y} = \mathbf{Y}'\left(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\right)\mathbf{Y}.$$

- $\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n$ is a projection matrix.

$$\mathbf{J}_n = \mathbf{1}_n\mathbf{1}_n' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

- $df(SSTO) = rank(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n) = n - 1.$

Regression sum of squares : $SSR = \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2$.

- Matrix form: $\overline{\mathbf{Y}} = \frac{1}{n}\mathbf{J}_n\mathbf{Y}$

$$
\begin{aligned}
SSR &= \left(\widehat{\mathbf{Y}} - \overline{\mathbf{Y}}\right)'\left(\widehat{\mathbf{Y}} - \overline{\mathbf{Y}}\right) \\
&= \mathbf{Y}'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}_n\right)'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}_n\right)\mathbf{Y} \\
&= \mathbf{Y}'(\mathbf{H} - \frac{1}{n}\mathbf{J}_n)\mathbf{Y}.
\end{aligned}
$$

- $\mathbf{H} - \frac{1}{n}\mathbf{J}_n$ is a projection matrix.
- $df(SSR) = rank(\mathbf{H} - \frac{1}{n}\mathbf{J}_n) = 1$.

$E(SSE)$

The following three slides will be discussed on Wed.'s lab session.

$$
\begin{aligned}
E(SSE) &= E\left(\mathbf{Y}'(\mathbf{I}_n - \mathbf{H})\mathbf{Y}\right) = E\left(Tr((\mathbf{I}_n - \mathbf{H})\mathbf{Y}\mathbf{Y}')\right) \\
&= Tr\left((\mathbf{I}_n - \mathbf{H})E(\mathbf{Y}\mathbf{Y}')\right) \\
&= Tr\Big((\mathbf{I}_n - \mathbf{H})(\sigma^2\mathbf{I}_n + \mathbf{X}\beta\beta'\mathbf{X}')\Big) \\
&= \sigma^2 Tr\left(\mathbf{I}_n - \mathbf{H}\right) + Tr\left((\mathbf{I}_n - \mathbf{H})\mathbf{X}\beta\beta'\mathbf{X}'\right) \\
&= (n-2)\sigma^2.
\end{aligned}
$$

The last equality is because $Tr\left(\mathbf{I}_n - \mathbf{H}\right) = n - 2$ and $(\mathbf{I}_n - \mathbf{H})\mathbf{X} = \mathbf{0}$.

Properties of Projection Matrices

- They have eigen-decomposition of the form: $Q\Lambda Q'$, where Q is an orthogonal matrix of eigenvectors and Λ is a diagonal matrix of eigenvalues.
- Their eigenvalues are either 1 or 0. The number of nonzero eigenvalues equals to the rank.
- For simple linear regression: $rank(\mathbf{H}) = 2$, $rank(\mathbf{I}_n - \mathbf{H}) = n - 2$.

Sampling Distribution of *SSE* Under Normal Error Model

- $\mathbf{I}_n - \mathbf{H}$ is a projection matrix with rank $n-2 \Longrightarrow$

$$\mathbf{I}_n - \mathbf{H} = \mathbf{Q}^T \mathbf{\Lambda} \mathbf{Q},$$

  where $\mathbf{\Lambda} = \mathrm{diag}\{1, \cdots, 1, 0, 0\}$ and $\mathbf{Q}$ is an orthogonal matrix.

- $(\mathbf{I}_n - \mathbf{H})\mathbf{X} = \mathbf{0} \Longrightarrow$

$$\mathbf{e} = (\mathbf{I}_n - \mathbf{H})\mathbf{Y} = (\mathbf{I}_n - \mathbf{H})(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) = (\mathbf{I}_n - \mathbf{H})\boldsymbol{\epsilon}$$

- $SSE = \mathbf{e}^T\mathbf{e} = \boldsymbol{\epsilon}^t(\mathbf{I}_n - \mathbf{H})\boldsymbol{\epsilon} = (\mathbf{Q}\boldsymbol{\epsilon})^T\mathbf{\Lambda}(\mathbf{Q}\boldsymbol{\epsilon})$.
- Let $\mathbf{z} = \mathbf{Q}\boldsymbol{\epsilon}$, then $SSE = \sum_{i=1}^{n-2} z_i^2$ and

$$\mathbf{E}(\mathbf{z}) = \mathbf{Q}\mathbf{E}\{\boldsymbol{\epsilon}\} = \mathbf{0}, \quad \boldsymbol{\sigma}^2\{\mathbf{z}\} = \mathbf{Q}\boldsymbol{\sigma}^2\{\boldsymbol{\epsilon}\}\mathbf{Q}^T = \sigma^2\mathbf{Q}\mathbf{Q}^T = \sigma^2\mathbf{I}_n.$$

- So $E(SSE) = \sum_{i=1}^{n-2} E(z_i^2) = (n-2)\sigma^2$. Under Normal error model, z_is are i.i.d. $N(0, \sigma^2)$, so $SSE \sim \sigma^2\chi^2_{(n-2)}$.