Chapter 5 Matrix Formulation

Random vectors

- Let y_1, y_2, and y_3 be random variables, construct a 3×1 random vector $\mathbf{Y}$ as

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

- The expected value of the random vector is the vector of expected values of the random variables. Write $E(y_i) = \mu_i$, we have

$$E(\mathbf{Y}) = \begin{bmatrix} E(y_1) \\ E(y_2) \\ E(y_3) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \boldsymbol{\mu}$$

- Let

$$\text{Var}(y_i) = E(y_i - \mu_i)^2 = \sigma_{ii}$$

$$\text{Cov}(y_i, y_j) = E[(y_i - \mu_i)(y_j - \mu_j)] = \sigma_{ij}$$

The covariance matrix of $\mathbf{Y}$ is

$$\begin{aligned}
\text{Cov}(\mathbf{Y}) &= E[(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})'] \\
&= E\begin{bmatrix} (y_1-\mu_1)^2 & (y_1-\mu_1)(y_2-\mu_2) & (y_1-\mu_1)(y_3-\mu_3) \\ (y_2-\mu_2)(y_1-\mu_1) & (y_2-\mu_2)^2 & (y_2-\mu_2)(y_3-\mu_3) \\ (y_3-\mu_3)(y_1-\mu_1) & (y_3-\mu_3)(y_2-\mu_2) & (y_3-\mu_3)^2 \end{bmatrix} \\
&= \begin{bmatrix} E[(y_1-\mu_1)^2] & E[(y_1-\mu_1)(y_2-\mu_2)] & E[(y_1-\mu_1)(y_3-\mu_3)] \\ E[(y_2-\mu_2)(y_1-\mu_1)] & E[(y_2-\mu_2)^2] & E[(y_2-\mu_2)(y_3-\mu_3)] \\ E[(y_3-\mu_3)(y_1-\mu_1)] & E[(y_3-\mu_3)(y_2-\mu_2)] & E[(y_3-\mu_3)^2] \end{bmatrix} \\
&= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}
\end{aligned}$$

Matrix formulation of regression models

1. Simple Linear Regression (SLR) in Scalar Form

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

$$y_1 = \beta_0 + \beta_1 x_1 + \varepsilon_1$$

$$y_2 = \beta_0 + \beta_1 x_2 + \varepsilon_2$$

$$\vdots$$

$$y_n = \beta_0 + \beta_1 x_n + \varepsilon_n$$

- The SLR model in Matrix Form

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_1 \\ \beta_0 + \beta_1 x_2 \\ \vdots \\ \beta_0 + \beta_1 x_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Let

- $\mathbf{X}$ be the design matrix
- $\boldsymbol{\beta}$ be the vector of parameters
- $\boldsymbol{\epsilon}$ be the error vector
- $\mathbf{Y}$ be the response vector

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \boldsymbol{\epsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

we have

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

Variance-Covariance Matrix

- Independence means uncorrelated
- Uncorrelated is not necessarily independent, uncorrelated means that their covariance is 0
- It is possible to have variables that are dependent but uncorrelated, because correlation only measures linear dependence
- For normally distributed random variables, uncorrelated is equivalent to independent
- Covariance matrix of $\boldsymbol{\epsilon}$

$$\text{Cov}\{\boldsymbol{\epsilon}\} = \text{Cov}\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \sigma^2 \mathbf{I}_{n\times n} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \vdots & \sigma^2 \end{bmatrix}$$

where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, $\mathbf{I}$ is the $n \times n$ identity matrix

- Covariance matrix of $\mathbf{Y}$

$$\sigma^2\{\mathbf{Y}\} = \text{Cov}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sigma^2\mathbf{I}_{n\times n}$$

Example. Height and weight data are given in the following Table for 12 individuals. Write the SLR model for regressing weights (y) on heights (x) in matrix form.

Ht.	Wt.	Ht.	Wt.
65	120	63	110
65	140	63	135
65	130	63	120
65	135	72	170
66	150	72	185
66	135	72	160

Multiple Regression

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{p-1} x_{i,p-1} + \epsilon_i, i = 1, \cdots, n,$$

where

$$E(\epsilon_i) = 0, \text{Var}(\epsilon_i) = \sigma^2, \text{Cov}(\epsilon_i, \epsilon_j) = 0, i \neq j.$$

Write the multiple regression model in matrix form.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{p-1} x_{i,p-1} + \epsilon_i, i = 1, \cdots, n,$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1,p-1} \\ 1 & x_{21} & x_{22} & \cdots & x_{2,p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{n,p-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\mathbf{Y}_{n\times 1} = \mathbf{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} + \boldsymbol{\epsilon}_{n\times 1}, E(\boldsymbol{\epsilon}) = \mathbf{0}, \text{Cov}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$$

Multiplying and adding the right-hand side gives

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_{p-1} x_{1,p-1} + \epsilon_1 \\ \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_{p-1} x_{2,p-1} + \epsilon_2 \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_{p-1} x_{n,p-1} + \epsilon_n \end{bmatrix}$$

Least Squares estimation of regression parameters

- For SLR, the least squares estimates are the values of β_0 and β_1 that minimize

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2.$$

- For multiple regression, the least squares estimates of the β_js minimize

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_{p-1} x_{i,p-1})^2.$$

- In matrix form, want to minimize the sum of squared residuals:

$$\sum_{i=1}^{i=n} e_i^2 = \begin{bmatrix} e_1 & e_2 & \cdots & e_n \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b})$$

Proposition. If rank$(\mathbf{X}) = p$, then

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

is the least squares estimate of $\boldsymbol{\beta}$.

- Normal equation

$$\mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{X}\mathbf{b},$$

solving for $\mathbf{b}$ gives the least squares solution for $\mathbf{b}$

-

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

Example. Simple linear regression LS estimates.

Proposition. Let $\mathbf{A}$ be a fixed $r \times n$ matrix, let c be a fixed $r \times 1$ vector, and let $\mathbf{Y}$ be an $n \times 1$ random vector, then

1. $E(\mathbf{AY} + \mathbf{c}) = \mathbf{A}E(\mathbf{Y}) + \mathbf{c}$
2. Cov$(\mathbf{AY} + \mathbf{c}) = \mathbf{A}Cov(\mathbf{Y})\mathbf{A}'$

Exercise: For multiple regression

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, E(\boldsymbol{\epsilon}) = \mathbf{0}, \text{Cov}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I},$$

Show that

- $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$, and Cov$(\mathbf{Y}) = \sigma^2\mathbf{I}$
- $\hat{\boldsymbol{\beta}}$ is unbiased, i.e., $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ and Cov$(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$

Review

Multivariate Theorem: Suppose $\mathbf{U} \sim N(\mathbf{u}, \sum)$, $\mathbf{V} = \mathbf{c} + \mathbf{D}\mathbf{U}$, a linear transformation of $\mathbf{U}$ where $\mathbf{c}$ is a vector and $\mathbf{D}$ is a matrix. Then $\mathbf{V} \sim N(\mathbf{c} + \mathbf{D}\mathbf{u}, \mathbf{D} \sum \mathbf{D}')$.

Consider

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{n \times n})$$

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$$

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

$$\mathbf{b} \sim N((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}, \sigma^2[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{I}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']')$$

$$\sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

Residuals, fitted values and leverage

Let $\mathbf{x}_i' = (1, x_{i1}, \cdots, x_{i,p-1})$ be the ith row of $\mathbf{X}$

- Fitted value

$$\hat{y}_i = b_0 + b_1 x_{i1} + \cdots + b_{p-1} x_{i,p-1} = \mathbf{x}_i' \mathbf{b}$$

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1' \hat{\boldsymbol{\beta}} \\ \vdots \\ \mathbf{x}_n' \hat{\boldsymbol{\beta}} \end{bmatrix} = \mathbf{X}\mathbf{b}$$

- Hat Matrix

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

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is called hat matrix. The leverage of the ith case is defined as the diagonal element h_{ii} (usually call h_i) of the hat matrix. The leverage can be interpreted as a measure of how unusual $\mathbf{x}_i$ is relative to the other rows of the $\mathbf{X}$ matrix.

- Residual is $\hat{\epsilon}_i = y_i - \hat{y}_i = y_i - \mathbf{x}_i'\mathbf{b}$

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

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

  SSE $= [\mathbf{Y} - \mathbf{X}\mathbf{b}]'[\mathbf{Y} - \mathbf{X}\mathbf{b}]$

  Cov$(\mathbf{e}) = \sigma^2(\mathbf{I} - \mathbf{H})$

  Var$(e_i) = \sigma^2(1 - h_{ii})$

  the standard error of e_i is $SE(e_i) = \sqrt{MSE(1 - h_{ii})}$

  the ith standardized residual is defined as

$$r_i = \frac{e_i}{\sqrt{MSE(1 - h_{ii})}}$$

  .

Variance and estimated variance matrices of $\mathbf{b}$:

$$
\begin{aligned}
\text{var}(\mathbf{b}) &= \begin{bmatrix} \text{var}(b_0) & \text{cov}(b_0, b_1) & \cdots & \text{cov}(b_0, b_{p-1}) \\ \text{cov}(b_1, b_0) & \text{var}(b_1) & \cdots & \text{cov}(b_1, b_{p-1}) \\ \vdots & \vdots & & \vdots \\ \text{cov}(b_{p-1}, b_0) & \text{cov}(b_{p-1}, b_1) & \cdots & \text{var}(b_{p-1}) \end{bmatrix} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}
$$

Estimated variance covariance:

$$s^2(\mathbf{b}) = MSE(\mathbf{X}'\mathbf{X})^{-1}$$

Inference for individual regression coefficient

$$\mathbf{b} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

$$s^2(\mathbf{b}) = MSE(\mathbf{X}'\mathbf{X})^{-1}$$

$$s^2(b_k) = [s^2(\mathbf{b})]_{k,k}$$

the k-th diagonal element.

- Distribution of b_k: $\dfrac{b_k - \beta_k}{s(b_k)} \sim t(n-p) \quad k = 0, 1, \cdots, p-1$
- a $(1-\alpha)100\%$ confidence interval for β_k

$$b_k \pm t_{n-p}(1-\alpha/2)s\{b_k\}$$

- Significant test for β_k

$$H_0 : \beta_k = 0 \quad \text{v.s} \quad \beta_k \neq 0$$

$$t^* = \frac{b_k}{s\{b_k\}}$$

If H_0 is true, t^* has a t-distribution with $n - p$ degrees of freedom.

Alternative	Reject H_0 if
$H_\alpha : \beta_k > 0$	$t^* > t(1 - \alpha; n - p)$
$H_\alpha : \beta_k < 0$	$t^* < -t(1 - \alpha; n - p)$
$H_\alpha : \beta_k \neq 0$	$\lvert t^* \rvert > t(1 - \alpha/2; n - p)$

Simultaneous Confidence Intervals for $\beta_1, \cdots, \beta_{p-1}$

$$b_k \pm t(1 - \frac{\alpha}{2(p-1)}; n-p) s(b_k),$$

where $k = 1, 2, \cdots, p-1$.

Estimation of $E(Y_h)$

We want a point estimate and a confidence interval for the mean corresponding to the set of explanatory variables $\mathbf{x}_h$.

$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{p-1} x_{i,p-1} + \epsilon_i,$

$\mathbf{x}_h = (1, x_{h1}, x_{h2}, \cdots, x_{h,p-1})'$

$Y_h = \beta_0 + \beta_1 x_{h1} + \beta_2 x_{h2} + \cdots + \beta_{p-1} x_{h,p-1} + \epsilon_h$

$\mu_h = E(Y_h) = \beta_0 + \beta_1 x_{h1} + \beta_2 x_{h2} + \cdots + \beta_{p-1} x_{h,p-1} = \mathbf{x}_h' \boldsymbol{\beta}$

$\hat{\mu}_h = \widehat{E(Y_h)} = b_0 + b_1 x_{h1} + b_2 x_{h2} + \cdots + b_{p-1} x_{h,p-1} = \mathbf{x}_h' \mathbf{b}$

$s^2\{\hat{\mu}_h\} = \mathbf{x}_h' s^2\{\mathbf{b}\} \mathbf{x}_h = \text{MSE} \mathbf{x}_h' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_h$

95% CI $\hat{\mu}_h \pm s\{\hat{\mu}_h\} t_{n-p}(1 - \alpha/2)$

Prediction of $Y_{h(\textbf{new})}$

Predict a new observation Y_h at $\mathbf{x}_h$. We want a prediction of Y_h based on a set of predictor values with an interval that expresses the uncertainty in our prediction. As in SLR this interval is centered at Y_h and is wider than the interval for the mean.

$Y_h = \mathbf{x}_h' \boldsymbol{\beta} + \epsilon_h$
$\hat{Y}_h = \hat{u}_h = \mathbf{x}_h' \mathbf{b}$

$$\begin{aligned} s^2\{\text{pred}\} &= \text{var}(Y_{h(\text{new})} - \hat{Y}_h) \\ &= \text{var}(Y_{h(\text{new})}) + \text{var}(\hat{Y}_h) \\ &= \text{MSE}(1 + \mathbf{x}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_h) \end{aligned}$$

CI for $Y_{h(\text{new})}$: $\hat{Y}_h \pm s\{\text{pred}\} t_{n-p}(1-\alpha/2)$