16.1. Classical Linear Model Assumptions

A. $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$

B. $E[\boldsymbol{\varepsilon}] = \mathbf{0}$

C. $\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$

D. $\boldsymbol{\varepsilon} \sim MVN$

We will consider the effects on inference when a model is fit with erroneous assumptions:

1. Underfitting the regression model.
2. Overfitting the regression model.
3. Mis-specifying the covariance matrix.
4. Non-normality.

Our goal is to understand the *effects* of erroneous assumptions.

16.2. Bias Due to Underfitting

Suppose the true model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \qquad E[\boldsymbol{\varepsilon}] = \mathbf{0},\ \operatorname{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I},$$

but we fit the smaller model:

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

We can assume that the columns of $\mathbf{Z}$ are linearly independent of the columns of $\mathbf{X}$ (linearly dependent columns add nothing new to the model). Also assume a full rank model, i.e., $\operatorname{rank}(\mathbf{X}_{n\times p}) = p$.

How will our estimates be affected?

Naive argument: I'm only interested in the parameters $\boldsymbol{\beta}$, so why bother estimating $\boldsymbol{\eta}$?

Consider: if the smaller model is used,

$$\begin{aligned} E[\hat{\boldsymbol{\beta}}] &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{Y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\eta}) \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\boldsymbol{\eta} \end{aligned}$$

Therefore,

$$\text{BIAS} \equiv E[\hat{\boldsymbol{\beta}}] - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\boldsymbol{\eta}.$$

Response to naive argument: estimates of your parameters of interest $\boldsymbol{\beta}$ will be biased if you ignore $\boldsymbol{\eta}$.

Unless?

Unless the columns of $\mathbf{X}$ are orthogonal to the columns of $\mathbf{Z}$.

The fitted values are also biased because they are based on the projection of $\mathbf{Y}$ onto the column space of $\mathbf{X}$ instead of the column space of $(\mathbf{X}, \mathbf{Z})$. Let $\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ be the projection operator onto the column space of $\mathbf{X}$.

$$E[\hat{\mathbf{Y}}] = E[\mathbf{P}_\mathbf{X}\mathbf{Y}] = \mathbf{P}_\mathbf{X}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\eta}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{P}_\mathbf{X}\mathbf{Z}\boldsymbol{\eta},$$

which is not the same as what we want, $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\eta}$.

Example: Fit

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

when the true model is

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon.$$

Then

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{\sum (x_i - \bar{x})^2} \begin{pmatrix} \sum x_i^2/n & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

and

$$\mathbf{X}'\mathbf{Z} = \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} x_1^2 \\ \vdots \\ x_n^2 \end{pmatrix} = \begin{pmatrix} \sum x_i^2 \\ \sum x_i^3 \end{pmatrix}.$$

The bias in $\hat{\boldsymbol{\beta}}$ is

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\beta_2 = \frac{\beta_2}{\sum (x_i - \bar{x})^2} \begin{pmatrix} (\sum x_i^2)^2/n - \bar{x} \sum x_i^3 \\ -\bar{x} \sum x_i^2 + \sum x_i^3 \end{pmatrix}.$$

Note that $\hat{\beta}_1$ is unbiased if $\bar{x} = 0$ and $\sum x_i^3 = 0$.

Note also that the bias depends on β_2. If β_2 is small the bias will be small.

Example: Fit

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad (i = 1, 2; j = 1, \ldots, n_i),$$

when the true model is

$$Y_{ij} = \mu_i + \eta z_{ij} + \varepsilon_{ij},$$

i.e., we compare two groups, ignoring the covariate z.

In matrix form the true model is $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\eta} + \boldsymbol{\varepsilon}$, or,

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} z_{11} \\ \cdots \\ z_{1n_1} \\ z_{21} \\ \cdots \\ z_{2n_2} \end{pmatrix} \eta + \begin{pmatrix} \varepsilon_{11} \\ \cdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \cdots \\ \varepsilon_{2n_2} \end{pmatrix}.$$

Then the bias in $(\hat{\mu}_1, \hat{\mu}_2)$ is

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\eta = \begin{pmatrix} \bar{z}_1 \\ \bar{z}_2 \end{pmatrix} \eta.$$

Assuming $\eta \neq 0$, the group comparison $\hat{\mu}_1 - \hat{\mu}_2$ is unbiased if and only if $\bar{z}_1 = \bar{z}_2$. i.e., the mean value of the covariate is the same in the two groups.

Randomization: Suppose we randomly assign experimental units to the two groups. Then $\bar{z}_1 \approx \bar{z}_2$ for any covariate z, as long as groups are fairly large. Thus, randomization eliminates bias due to unfitted covariates.

16.3. Effects of Underfitting on the Error Variance Estimate

The usual form for the covariance matrix of $\hat{\boldsymbol{\beta}}$ is still valid (why?):

$$\mathrm{cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

This follows simply by calculating $\mathrm{cov}(\hat{\boldsymbol{\beta}}) = \mathrm{cov}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y})$

The problem is the estimate of the error variance σ^2 is biased.

$$\begin{aligned} E[RSS] &= E[\mathbf{Y}'(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{Y}] \\ &= \mathrm{tr}\{(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\sigma^2\mathbf{I}\} + (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\eta})'(\mathbf{I} - \mathbf{P}_{\mathbf{X}})(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\eta}) \\ &= \sigma^2(n-p) + (\mathbf{Z}\boldsymbol{\eta})'(\mathbf{I} - \mathbf{P}_{\mathbf{X}})(\mathbf{Z}\boldsymbol{\eta}), \end{aligned}$$

because $(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{X} = \mathbf{0}$. Therefore,

$$E[S^2] = \frac{E[RSS]}{n-p} = \sigma^2 + \frac{\boldsymbol{\eta}'\mathbf{Z}'(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{Z}\boldsymbol{\eta}}{n-p} > \sigma^2$$

The lesson is that underfitting leads to overestimation of the error variance.

Recall from your regression class: precision variables.

16.3. Effects of Overfitting

Suppose the true model is

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}, \qquad E[\boldsymbol{\varepsilon}] = \mathbf{0},\ \operatorname{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I},$$

but we fit the model

$$E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta} = (\mathbf{X}_1, \mathbf{X}_2)\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2,$$

i.e., we are fitting unnecessary terms in $\mathbf{X}_2$.

Consider: if the larger model is used,

$$\begin{aligned} E[\hat{\boldsymbol{\beta}}] &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{Y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}_1\boldsymbol{\beta}_1 \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}_1, \mathbf{X}_2)\begin{pmatrix}\boldsymbol{\beta}_1\\ \mathbf{0}\end{pmatrix} \\ &= \begin{pmatrix}\boldsymbol{\beta}_1\\ \mathbf{0}\end{pmatrix}. \end{aligned}$$

Therefore, $\hat{\boldsymbol{\beta}}$ is unbiased. In particular $\hat{\boldsymbol{\beta}}_1$ is unbiased for $\boldsymbol{\beta}_1$ and $\hat{\boldsymbol{\beta}}_2$ has mean $\mathbf{0}$, which is what we would hope.
Also, the fitted values $\hat{\mathbf{Y}}$ are unbiased because

$$E[\hat{\mathbf{Y}}] = \mathbf{X}E[\hat{\boldsymbol{\beta}}] = (\mathbf{X}_1, \mathbf{X}_2)\begin{pmatrix}\boldsymbol{\beta}_1\\ \mathbf{0}\end{pmatrix} = \mathbf{X}_1\boldsymbol{\beta}_1.$$

Effects of overfitting on $\text{cov}(\hat{\boldsymbol{\beta}})$:

How is $\text{cov}(\hat{\boldsymbol{\beta}}_1)$ affected by fitting the unnecessary $\hat{\boldsymbol{\beta}}_2$? We have

$$\begin{aligned} \text{cov}(\hat{\boldsymbol{\beta}}) &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2 \begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{pmatrix}^{-1} \\ &= \sigma^2 \begin{pmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}' & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}' & \mathbf{E}^{-1} \end{pmatrix} \end{aligned}$$

where

$$\mathbf{F} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2,$$

and

$$\begin{aligned} \mathbf{E} &= \mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2 \\ &= \mathbf{X}_2'(\mathbf{I} - \mathbf{P}_{\mathcal{R}(\mathbf{X}_1)})\mathbf{X}_2. \end{aligned}$$

Therefore,

$$\text{cov}(\hat{\boldsymbol{\beta}}_1) = \sigma^2[(\mathbf{X}_1'\mathbf{X}_1)^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}'],$$

compared with $\sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}$ which would result from fitting the true model $E[\mathbf{Y}] = \mathbf{X}_1\boldsymbol{\beta}_1$. It can be shown that $\mathbf{F}\mathbf{E}^{-1}\mathbf{F}'$ is positive definite unless $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$ (Seber & Lee p. 231). If $\mathbf{X}_1'\mathbf{X}_2 \neq \mathbf{0}$, the variance of parameter estimates will be inflated by overfitting. If $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$, then $\mathbf{F} = \mathbf{0}$, so $\text{cov}(\hat{\boldsymbol{\beta}}_1)$ is not affected.

So we haven't hurt ourselves here if we add unnecessary covariates that are orthogonal to the covariates of interest. Are we done? No, because σ^2 must be estimated. (Next page.)

Effects of overfitting on the error variance estimate:

S^2 remains unbiased:

$$\begin{aligned} E[RSS] &= E[\mathbf{Y}'(\mathbf{I}-\mathbf{P}_{\mathbf{X}})\mathbf{Y}] \\ &= \operatorname{tr}\{(\mathbf{I}-\mathbf{P}_{\mathbf{X}})\sigma^2\mathbf{I}\} + (E[\mathbf{Y}])'(\mathbf{I}-\mathbf{P}_{\mathbf{X}})(E[\mathbf{Y}]) \\ &= \operatorname{tr}\{(\mathbf{I}-\mathbf{P}_{\mathbf{X}})\sigma^2\mathbf{I}\} + (\mathbf{X}_1\boldsymbol{\beta}_1)'(\mathbf{I}-\mathbf{P}_{\mathbf{X}})(\mathbf{X}_1\boldsymbol{\beta}_1) \\ &= \sigma^2(n-p) \end{aligned}$$

because $(\mathbf{I}-\mathbf{P}_{\mathbf{X}})\mathbf{X}_1 = \mathbf{0}$.

The lesson is that overfitting does not introduce bias into regression coefficient estimates, but it does in general inflate their variances.

16.4. Summary of Effects of Underfitting and Overfitting

	Effect of Underfitting	Effect of Overfitting
$\hat{\boldsymbol{\beta}}$	biased	unbiased
$\hat{\mathbf{Y}}$	biased	unbiased
S^2	biased upward	unbiased
$\text{cov}(\hat{\boldsymbol{\beta}})$	still $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$	increased variance of estimable functions