Advanced Methods in Biostatistics I

Lecture 9

Martin Lindquist

September 26, 2017

Linear Models

- Today we will continue working with the linear model:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon.$$

- However, we will now add a couple of additional assumptions to the model.
- Let us begin by assuming that $E(\varepsilon) = \mathbf{0}$ and $\mathrm{var}(\varepsilon) = \sigma^2\mathbf{I}$.

Linear Models

- These assumptions can equivalently be expressed as as $E(\mathbf{y}) = \mathbf{X}\beta$ and $\text{var}(\mathbf{y}) = \sigma^2\mathbf{I}$.
- Under this formulation the term $\mathbf{X}\beta$ expresses how the expected value of the random vector $\mathbf{y}$ changes as a function of the explanatory variables contained in $\mathbf{X}$.
- It also implies that the observations y_i and y_j are uncorrelated for $i \neq j$.

Least Squares Estimator

- It is important to note that the least squares estimator

  $$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

  was derived without making these assumptions.

- Therefore, even if $E(\mathbf{y}) \neq \mathbf{X}\beta$, the linear model can still be used to fit to the data.

- However, the resulting estimate may have poor properties.

Least Squares Estimator

- In contrast, we will show that under these assumptions the estimates of β have some very favorable properties.
- To begin exploring these properties we begin by noting that the least-squares estimator is a random vector, and thus it has an expected value and variance.

Expected Value

Theorem

If $\mathbf{X}$ is of full rank, then the least squares estimate is unbiased, i.e., $E[\hat{\beta}] = \beta$.

Variance-covariance

Theorem

If $\mathbf{X}$ is of full rank, then the variance-covariance matrix of the least squares estimate is $\mathrm{var}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

Example

- Let us illustrate this result in the context of simple linear regression:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad \text{for } \; i = 1, \ldots n$$

or

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

Example

- Suppose we are interested in computing the variance-covariance matrix of $\hat{\beta}$.
- Recall that:

$$(\mathbf{X}'\mathbf{X}) = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_i x_i^2 \end{bmatrix}.$$

- Hence:

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

Example

- Using the matrix $(\mathbf{X}'\mathbf{X})^{-1}$ we obtain:

$$
\begin{aligned}
\mathrm{var}(\hat{\beta}) &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \\
&= \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2} \begin{bmatrix} \sum_i x_i^2/n & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}.
\end{aligned}
$$

Example

- Thus, it holds that

$$\mathrm{var}(\hat{\beta}_0) = \frac{\sigma^2 \sum_i x_i^2 / n}{\sum_i (x_i - \bar{x})^2},$$

$$\mathrm{var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2},$$

and

$$\mathrm{cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\sigma^2 \bar{x}}{\sum_i (x_i - \bar{x})^2}$$

Example

- Studying $\mathrm{var}(\hat{\beta}_1)$, we note three aspects that affect the variance of the regression slope:
  (i) the spread around the regression line;
  (ii) the spread of the **x** values; and
  (iii) the sample size n.

- Thus, with less scatter around the line the slope will be more consistent from sample to sample, a large variance of **x** provides a more stable regression, and having a larger sample size provides more consistent estimates.

BLUE

- It is important to note that one can estimate β in the linear model using other loss functions than the least-squares criteria.
- Why do we focus on the least-squares estimator?
- Because it is the best linear unbiased estimator (BLUE).

BLUE

- Note that here ‘best’ implies minimum variance, and ‘linear’ that the estimators are linear functions of **y**.
- Remarkably, the results holds for any distribution of **y**.
- The only assumption needed is $E(\varepsilon) = \mathbf{0}$ and $\mathrm{var}(\varepsilon) = \sigma^2\mathbf{I}$.

BLUE

- It is important to only consider unbiased estimators, since we could always minimize the variance by defining an estimator to be constant (hence variance 0).
- If one removes the restriction of unbiasedness, then minimum variance cannot be the definition of 'best'.
- Often one then looks to the mean squared error, the squared bias plus the variance, instead.

Gauss-Markov Theorem

Theorem

If $E(\mathbf{y}) = \mathbf{X}\beta$ and $\mathrm{var}(\mathbf{y}) = \sigma^2\mathbf{I}$, then the least squares estimator $\hat{\beta}$ is the best linear unbiased estimators (BLUE).

Gauss-Markov Theorem - Proof

- Note $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is a linear estimator.
- Consider an alternative linear estimator of β: $\mathbf{b} = \mathbf{A}\mathbf{y}$.
- As it is a linear estimator we can express $\mathbf{A}$ as follows:

$$\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{D}$$

  where $\mathbf{D}$ is a non-zero matrix.

Gauss-Markov Theorem - Proof

- For $\mathbf{b} = \mathbf{Ay}$ to be an unbiased estimator of β, $E(\mathbf{b}) = \beta$.
- The expected value can be written:

$$\begin{aligned} E(\mathbf{b}) &= \mathbf{A}E(\mathbf{y}) \\ &= ((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{D})E(\mathbf{y}) \\ &= ((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{D})\mathbf{X}\beta \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + \mathbf{DX}\beta \\ &= \beta + \mathbf{DX}\beta \end{aligned}$$

- This provides a condition for $\mathbf{b}$ to be an unbiased estimator: $\mathbf{DX} = \mathbf{0}$.

Gauss-Markov Theorem - Proof

- Now we can express the variance-covariance matrix as follows:

$$
\begin{aligned}
\text{var}(\mathbf{b}) &= \sigma^2 \mathbf{A}\mathbf{A}' \\
&= \sigma^2 [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{D}][(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{D}]' \\
&= \sigma^2 [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}' \\
&\quad + \mathbf{D}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{D}\mathbf{D}'] \\
&= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} + \sigma^2 \mathbf{D}\mathbf{D}' \\
&= \text{var}(\hat{\beta}) + \sigma^2 \mathbf{D}\mathbf{D}'
\end{aligned}
$$

- Since $\mathbf{DD}'$ is positive definite, the variance of $\text{var}(\mathbf{b})$ exceeds that of $\text{var}(\hat{\beta})$.

Gauss-Markov Theorem

- We can extend these results to hold for linear contrasts of $\boldsymbol{\beta}$.
- Here $\mathbf{q}'\hat{\beta}$ is the *best* estimator of $\mathbf{q}\beta$ in the sense of minimizing the variance among linear (in $\mathbf{Y}$) unbiased estimators.

Gauss-Markov Theorem

Theorem

If $E(\mathbf{y}) = \mathbf{X}\beta$ and $\mathrm{var}(\mathbf{y}) = \sigma^2\mathbf{I}$, then the best linear unbiased estimators of $\mathbf{q}'\beta$ is $\mathbf{q}'\hat{\beta}$, where $\hat{\beta}$ is the least-squares estimator $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

Estimating σ^2

- We have previously focused our attention on estimating $\boldsymbol{\beta}$.
- However, using the latest model formulation we also need to estimate the parameter σ^2.

Estimating σ^2

- Since $E(\varepsilon'\varepsilon) = n\sigma^2$ we can use the residuals to provide an estimate of σ^2.
- Let us define

$$s^2 = (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})/(n - r).$$

Estimating σ^2

Theorem

s^2 is an unbiased estimate of σ^2.

Estimating σ^2

Theorem

An unbiased estimate of $\text{var}(\hat{\beta})$ is given by

$$\widehat{\text{var}}(\hat{\beta}) = s^2(\mathbf{X}'\mathbf{X})^{-1}.$$

Model Misspecification

- It is important to realize that any linear model is only as good as the specified design matrix.
- Incorrect specification can lead to bias and model misfit, resulting in power loss and an inflated false positive rate.
- Problems can arise if either irrelevant explanatory variables are included, or relevant variables are omitted.

Model Misspecification

- Assume that the correctly specified linear model is

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \epsilon.$$

- Here $E(\mathbf{y}) = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2$.

Model Misspecification

- Further, suppose we instead use the model:

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \epsilon.$$

- This implies we are under-fitting the model.
- Then the least-squares estimator is given by

$$\hat{\beta}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}.$$

Model Misspecification

- Computing the expectation, we see that

$$
\begin{aligned}
E(\hat{\beta}_1) &= E((\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}) \\
&= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'E(\mathbf{y}) \\
&= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2) \\
&= \beta_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\beta_2
\end{aligned}
$$

- In addition,

$$
\mathrm{var}(\hat{\beta}_1) = \sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}.
$$

Model Misspecification

- Thus, the estimate of β_1 is biased.
- Note that the bias disappears if either $\beta_2 = 0$ or $\mathbf{X}_1'\mathbf{X}_2 = 0$.

Model Misspecification

- The estimate of σ^2 will also be biased, with

$$E(s^2) = \sigma^2 + \frac{1}{n-p}\beta_2'\mathbf{X}_2'(\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')\mathbf{X}_2\beta_2.$$

- This can be seen by noting that:

$$\begin{aligned} E(\mathbf{y}'(\mathbf{I} - \mathbf{P}_{\mathbf{X}_1})\mathbf{y}) &= (\mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2)'(\mathbf{I} - \mathbf{P}_{\mathbf{X}_1})(\mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2) \\ &\quad + \mathrm{tr}[(\mathbf{I} - \mathbf{P}_{\mathbf{X}_1})\sigma^2\mathbf{I})] \\ &= (\mathbf{X}_2\beta_2)'(\mathbf{I} - \mathbf{P}_{\mathbf{X}_1})(\mathbf{X}_2\beta_2) + (n-p)\sigma^2 \end{aligned}$$

Model Misspecification

- Now, assume the correctly specified model is

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \epsilon.$$

- However, suppose we instead use the model:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \epsilon \\ &= \mathbf{X}\boldsymbol{\beta} + \epsilon \end{aligned}$$

  where $\mathbf{X} = [\mathbf{X}_1 \;\; \mathbf{X}_2]$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}_1 \;\; \boldsymbol{\beta}_2]'$.

- This implies we are over-fitting the model.

Model Misspecification

- Now, one can show:

$$\begin{aligned}
E(\hat{\beta}_1) &= \beta_1 \\
E(s^2) &= \sigma^2 \\
\mathrm{var}(\hat{\beta}_1) &= \sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1} \\
+ & \quad \sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}(\mathbf{X}_1'\mathbf{X}_2)[\mathbf{X}_2'(\mathbf{I} - \mathbf{P}_{\mathbf{X}_1})\mathbf{X}_2]^{-1}\mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1')^{-1}
\end{aligned}$$

Model Misspecification

- Thus, if irrelevant variables are included, the parameters β_1 and s^2 will remain unbiased.
- However, the variance-covariance matrix of β_1 will be inflated effecting subsequent inference.