

# Advanced Methods in Biostatistics I

## Lecture 8

Martin Lindquist

September 26, 2019

# Spectral decomposition

## Theorem

For any symmetric matrix  $\mathbf{A}$  with eigenvalues  $\lambda_1, \dots, \lambda_n$ , there exists an orthogonal matrix  $\mathbf{Q}$  such that:

$$\mathbf{Q}'\mathbf{A}\mathbf{Q} = \text{diag}(\lambda_1, \dots, \lambda_n).$$

# Properties of symmetric matrices

## Theorem

For any symmetric matrix  $\mathbf{A}$  with eigenvalues  $\lambda_1, \dots, \lambda_n$ ,

$$\text{tr}(\mathbf{A}) = \sum_i \lambda_i.$$

# Properties of a projection matrix

## Theorem

Let  $\mathbf{P}$  be a projection matrix, then  $\text{tr}(\mathbf{P}) = \text{rank}(\mathbf{P})$ .

# Properties of a projection matrix

## Theorem

If  $\mathbf{P}$  is an  $n \times n$  matrix and  $\text{rank}(\mathbf{P}) = r$ , then  $\mathbf{P}$  has  $r$  eigenvalues equal to 1 and  $n - r$  eigenvalues equal to 0.

# Properties of a projection matrix

## Theorem

Let  $\mathbf{P}$  be a projection matrix, then  $\mathbf{P}$  is positive semidefinite.

- Today we will continue working with the linear model:

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

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

# Linear Models

- These assumptions can equivalently be expressed 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  $\beta$  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.

## Theorem

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

## Theorem

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

# Example

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

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad \text{for } i = 1, \dots, n$$

or

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

# 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:

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

# Example

- Thus, it holds that

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

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

and

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



# Example

- Studying  $\text{var}(\hat{\beta}_1)$ , we note three aspects that affect the variance of the regression slope:
  - (i) the spread around the regression line;
  - (ii) the spread of the  $\mathbf{x}$  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  $\mathbf{x}$  provides a more stable regression, and having a larger sample size provides more consistent estimates.

- It is important to note that one can estimate  $\beta$  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).

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

- 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 is not a reasonable definition of 'best'.
- Often one then looks to the mean squared error, the squared bias plus the variance, instead.

# Gauss-Markov Theorem

## Theorem

If  $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$  and  $\text{var}(\mathbf{y}) = \sigma^2\mathbf{I}$ , then the least squares estimator  $\hat{\boldsymbol{\beta}}$  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  $\beta$ :  $\mathbf{b} = \mathbf{A}\mathbf{y}$ .
- 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  $\mathbf{b} = \mathbf{A}\mathbf{y}$  to be an unbiased estimator of  $\beta$ ,  $E(\mathbf{b}) = \beta$ .
- The expected value can be written:

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

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

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

# Estimating $\sigma^2$

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

# Estimating $\sigma^2$

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

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

# Estimating $\sigma^2$

## Theorem

$s^2$  is an unbiased estimate of  $\sigma^2$ .

## Theorem

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

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

# Fitted values and residuals

- Recall that the vector of fitted values is given by

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}.$$

- Similarly, the vector of residuals is given by

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\mathbf{y}.$$

# Properties of the fitted values and residuals

## Properties

- 1  $E(\mathbf{e}) = \mathbf{0}$
- 2  $Var(\mathbf{e}) = \sigma^2(\mathbf{I} - \mathbf{H})$
- 3  $Cov(\mathbf{e}, \mathbf{y}) = \sigma^2(\mathbf{I} - \mathbf{H})$
- 4  $Cov(\mathbf{e}, \hat{\mathbf{y}}) = \mathbf{0}$
- 5  $\sum_{i=1}^n e_i/n = 0$
- 6  $\mathbf{e}'\hat{\mathbf{y}} = 0$
- 7  $\mathbf{e}'\mathbf{X} = \mathbf{0}'$