Advanced Methods in Biostatistics I

Lecture 10

Martin Lindquist

September 28, 2017

Variance-covariance Matrices

- In todays lecture we will explore new assumptions on the variance-covariance matrix in the linear model.
- In doing so we will need to take the square root of a symmetric nonnegative definite matrix.
- We therefore begin with a brief review of eigenvalues.

Eigenvalues

Definition

If $\mathbf{Ax} = \lambda\mathbf{x}$ where $\mathbf{x} \neq 0$, then λ is an eigenvalue of $\mathbf{A}$ and $\mathbf{x}$ is a corresponding eigenvector.

Eigenvalues

Theorem

If **A** is a symmetric matrix, then the eigenvalues of **A** corresponding to distinct eigenvalues are orthogonal.

Spectral Decomposition

Theorem

For any symmetric matrix **A** with eigenvalues $\lambda_1, \ldots, \lambda_n$, there exists an orthogonal matrix **Q** such that:

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

Linear Model

- Last time we worked with the model:

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

  where $E(\epsilon) = \mathbf{0}$ and $\mathrm{var}(\epsilon) = \sigma^2\mathbf{I}$.

- Today we will discuss ways of relaxing the assumption that $\mathrm{var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$.

Example - Clustered data

- Suppose we are dealing with clustered data.
- Let
$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_K \end{pmatrix},$$
where $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})'$ is a vector of responses on the i^{th} cluster (patient, household, school, etc).

Example - Clustered data

- Assuming the clusters are independent, we can write:

$$\mathrm{var}(\mathbf{y}) = \begin{pmatrix} \mathbf{V}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{V}_K \end{pmatrix}.$$

Here we might assume a common variance σ^2 and common pairwise correlation ρ within a cluster.

Example - Clustered data

- This corresponds to an exchangeable correlation structure:

$$\mathrm{var}(\mathbf{y}_i) = \sigma^2 \mathbf{V}_i = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}_{n_i \times n_i}$$

Example - Autocorrelated data

- Another example is when we are dealing with time series data.
- Assume $\mathbf{y} = (y_1, \ldots, y_T)'$ are a set of observations measured sequentially over time.
- Here there may be reason to believe that the error in adjacent time points are correlated with one another, and that this correlation decays as the time between observations increases.

Example - Autocorrelated data

- This could be modeled as follows:

$$\mathrm{var}(\mathbf{y}) = \sigma^2 \mathbf{V} = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^n \\ \rho & 1 & \rho & \cdots & \vdots \\ \rho^2 & \rho & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \rho \\ \rho^n & \rho^{n-1} & \cdots & \rho & 1 \end{pmatrix}$$

- This is an example of an AR(1) process.

Variance-covariance Matrices

- What happens in the linear model setting if we relax the assumption that $\mathrm{var}(\mathbf{y}) = \sigma^2\mathbf{I}$?
- Recall that $\hat{\beta}$ was derived without making any assumptions about the mean and variance of **y**.
- Thus, the least squares estimate $\hat{\beta}$ is uneffected.
- However, the properties of the estimator and any subsequent inference will be effected.

Variance-covariance Matrices

- To illustrate, assume $\mathrm{var}(\mathbf{y}) = \sigma^2\mathbf{V}$, where the matrix $\mathbf{V}$ is assumed to be known.
- Note, in practice, we will typically also have to estimate $\mathbf{V}$.
- However, we will wait to discuss this at a later time.

Variance-covariance Matrices

- In this setting, $\hat{\beta}$ is still unbiased.
- However, the variance-covariance matrix is

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

- In addition, there is no longer any guarantee that the estimator is the BLUE of β, as the G-M theorem assumed $\mathrm{var}(\mathbf{y}) = \mathbf{I}\sigma^2$.

Generalized Least Squares

- To circumvent this issue we now introduce the method of generalized least squares (GLS) to improve upon estimation efficiency for the case when $\mathrm{cov}(\mathbf{Y}) \neq \sigma^2\mathbf{I}$
- We seek to transform the model to a new set of observations that satisfy the constant variance assumption.
- Thereafter one can use the ordinary least squares to estimate the model parameters.

Spectral Decomposition

- Because $\mathbf{V}$ is symmetric positive definite, we can write it as

$$\mathbf{V} = \mathbf{QDQ}',$$

where $\mathbf{Q}$ is orthogonal and $\mathbf{D}$ is the diagonal matrix consisting of $\lambda_1, \ldots, \lambda_n$, the eigenvalues of $\mathbf{V}$.

Spectral Decomposition

- Because $\mathbf{V}$ is p.d., we know that $\lambda_i > 0$, for $i = 1, \ldots n$.
- The symmetric square root of $\mathbf{V}$ can be written as:

  $$\mathbf{V}^{1/2} = \mathbf{QD}^{1/2}\mathbf{Q}',$$

  where $\mathbf{D}^{1/2} = \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$.
- Note that $\mathbf{V} = \mathbf{V}^{1/2}\mathbf{V}^{1/2}$.

Spectral Decomposition

- Also note that

$$\mathbf{V}^{-1} = \mathbf{V}^{-1/2}\mathbf{V}^{-1/2},$$

where

$$\mathbf{V}^{-1/2} = \mathbf{Q}\mathbf{D}^{-1/2}\mathbf{Q}'$$

and

$$\mathbf{D}^{-1/2} = \text{diag}(1/\sqrt{(}\lambda_1), \ldots, 1/\sqrt{\lambda_n}).$$

Generalized Least Squares

- Since $\sigma^2\mathbf{V}$ is a variance-covariance matrix, $\mathbf{V}$ is a symmetric non-singular matrix, and we can write:

$$\mathbf{V} = \mathbf{K}\mathbf{K},$$

  where $\mathbf{K}$ is the square root of $\mathbf{V}$.

- Using this matrix, let $\tilde{\mathbf{y}} = \mathbf{K}^{-1}\mathbf{y}$, $\tilde{\mathbf{X}} = \mathbf{K}^{-1}\mathbf{X}$, and $\tilde{\varepsilon} = \mathbf{K}^{-1}\varepsilon$.

Generalized Least Squares

- Then, it holds that
$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\beta + \tilde{\varepsilon},$$
where $E(\tilde{\varepsilon}) = \mathbf{0}$ and $\mathrm{var}(\tilde{\varepsilon}) = \sigma^2\mathbf{I}$.
- We are back to the standard assumptions of least squares.

Generalized Least Squares

- The least squares function can be expressed as follows:

$$\begin{aligned} f(\beta) &= ||\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\beta||^2 \\ &= (\mathbf{K}^{-1}\mathbf{y} - \mathbf{K}^{-1}\mathbf{X}\beta)'(\mathbf{K}^{-1}\mathbf{y} - \mathbf{K}^{-1}\mathbf{X}\beta) \\ &= (\mathbf{y} - \mathbf{X}\beta)'\mathbf{K}^{-1}\mathbf{K}^{-1}(\mathbf{y} - \mathbf{X}\beta) \\ &= (\mathbf{y} - \mathbf{X}\beta)'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta) \end{aligned}$$

- This is referred to as the generalized least-squares function.

Generalized Least Squares

- To minimize $f(\beta)$, begin by taking the derivative with respect to β and setting the results equal to 0.
- This gives the normal equations:

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

Generalized Least Squares

- Thus, least squares applied to the transformed $\mathbf{y}$ yields

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

- This is called the Generalized Least Squares (GLS) estimate.
- We refer to the solution

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

as the ordinary least squares (OLS) estimate.

Generalized Least Squares

Theorem

Properties of $\hat{\beta}_G$:

(a) $E[\hat{\beta}_G] = \beta$,

(b) $\operatorname{cov}(\hat{\beta}_G) = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$

Optimality of GLS estimates

Theorem

If $E[\mathbf{y}] = \mathbf{X}\beta$ and $\mathrm{cov}(\mathbf{y}) = \sigma^2\mathbf{V}$, then for any constant vector $\mathbf{q}$, $\mathbf{q}'\hat{\beta}_G$ is the BLUE of $\mathbf{q}'\beta$.

Estimating σ^2

- The parameter σ^2 can be estimated in the usual way using $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{X}}$.

$$
\begin{aligned}
{s_G}^2 &= \frac{1}{N-p}(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\hat{\beta}_G)'(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\hat{\beta}_G) \\
&= \frac{1}{N-p}(\mathbf{K}^{-1}\mathbf{y} - \mathbf{K}^{-1}\mathbf{X}\hat{\beta}_G)'(\mathbf{K}^{-1}\mathbf{y} - \mathbf{K}^{-1}\mathbf{X}\hat{\beta}_G) \\
&= \frac{1}{N-p}(\mathbf{y} - \mathbf{X}\hat{\beta}_G)'\mathbf{K}^{-1}\mathbf{K}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}_G) \\
&= \frac{1}{N-p}(\mathbf{y} - \mathbf{X}\hat{\beta}_G)'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}_G)
\end{aligned}
$$

Estimating σ^2

Theorem

${s_G}^2$ is an unbiased estimate of σ^2.

Heteroscedasticity

- In certain situations the assumption of constant variance is violated.
- Instead we have the following variance-covariance matrix:

$$\mathrm{var}(\mathbf{y}_i) = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n^2 \end{pmatrix}$$

- This is referred to as heteroscedasticity.

Weighted Least Squares

- Let us consider the regression through the origin case with heteroscedastic error, where we have a single explanatory variable **x** and no intercept.
- Then, $E[\mathbf{y}] = \beta\mathbf{x}$, and $\mathrm{var}(\mathbf{y}) = \mathrm{diag}(\sigma_1^2, \ldots \sigma_n^2)$.
- Let us denote $w_i = 1/\sigma_i^2$.

Generalized Least Squares

- The GLS estimate of β is

$$\hat{\boldsymbol{\beta}}_G = \frac{\sum_{i=1}^n w_i x_i y_i}{\sum_{i=1}^n w_i x_i^2}.$$

- The OLS estimate is

$$\hat{\boldsymbol{\beta}} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

Weighted least squares

- The respective variances are given by:

$$\mathrm{var}(\hat{\beta}_G) = \frac{1}{\sum_{i=1}^n w_i x_i^2}$$

and

$$\mathrm{var}(\hat{\beta}) = \frac{\sum_{i=1}^n \frac{x_i^2}{w_i}}{(\sum_{i=1}^n x_i^2)^2}.$$

Cauchy-Schwarz inequality

- Recall from HW1:

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq ||\mathbf{u}|| \cdot ||\mathbf{v}||.$$

- This is Cauchy-Schwarz inequality and will come in handy.

Weighted least squares

- Let $x_i^2 = u_i v_i$, where $u_i = x_i\sqrt{w_i}$ and $v_i = x_i/\sqrt{w_i}$.
- Then,

$$
\begin{aligned}
(\sum_{i=1}^{n} x_i^2)^2 &= (\sum_{i=1}^{n} u_i v_i)^2 \\
&\leq \sum_{i=1}^{n} u_i^2 \sum_{i=1}^{n} v_i^2 \\
&= \sum_{i=1}^{n} w_i x_i^2 \sum_{i=1}^{n} x_i^2/w_i
\end{aligned}
$$

Weighted least squares

- Thus it holds that:

$$
\begin{aligned}
\mathrm{var}(\hat{\beta}_G) &= \frac{1}{\sum_{i=1}^{n} w_i x_i^2} \\
&\leq \frac{\sum_{i=1}^{n} \frac{x_i^2}{w_i}}{(\sum_{i=1}^{n} x_i^2)^2} \\
&= \mathrm{var}(\hat{\beta})
\end{aligned}
$$

Generalized Least Squares

Theorem

The GLS estimate and the OLS estimate are equal only when either one of the following conditions holds:

1. $\mathcal{R}(\mathbf{V}^{-1}\mathbf{X}) = \mathcal{R}(\mathbf{X})$.
2. $\mathcal{R}(\mathbf{V}\mathbf{X}) = \mathcal{R}(\mathbf{X})$.