Advanced Methods in Biostatistics I

Lecture 8

Martin Lindquist

September 21, 2017

Least squares

- Up to this point, our exploration of linear models has focused on least squares and projections.
- We now begin discussing the statistical properties of our estimators.
- But first we review some useful results from matrix algebra before turning our focus to multivariate expectations and variances.

Trace of a Matrix

Definition

The trace of an $n \times n$ square matrix $\mathbf{A}$ is defined as the sum of the diagonal elements, i.e.,

$$\mathrm{tr}(\mathbf{A}) \equiv \sum_i a_{ii} = a_{11} + a_{22} + \cdots + a_{nn}.$$

Trace of a Matrix

Theorem

For any scalar k and any $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$, the

- $\mathrm{tr}(k\mathbf{A}) = k\mathrm{tr}(\mathbf{A})$
- $\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$
- $\mathrm{tr}(\mathbf{A}') = \mathrm{tr}(\mathbf{A})$

Trace of a Matrix

Theorem

Let $\mathbf{A} = (a_{ij})$ represent an $m \times n$ matrix and $\mathbf{B} = (b_{ij})$ an $n \times m$ matrix. Then,

$$\mathrm{tr}(\mathbf{AB}) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ji}.$$

Trace of a Matrix

Theorem

If $\mathbf{A}$ and $\mathbf{B}$ are matrices such that $\mathbf{AB}$ is a square matrix, then:

$$\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA}) = \operatorname{tr}(\mathbf{B}'\mathbf{A}') = \operatorname{tr}(\mathbf{A}'\mathbf{B}').$$

Positive Semidefinite Matrix

Definition

A symmetric matrix $\mathbf{A}$ is positive semidefinite (p.s.d.) if $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x}$.

Positive Semidefinite Matrix

Theorem

If $\mathbf{A}$ is a p.s.d matrix, then

(a) The diagonal elements a_{ii} are all non-negative.

(b) $\text{tr}(\mathbf{A}) \geq 0$.

(c) All eigenvalues of $\mathbf{A}$ are nonnegative.

Positive Definite Matrix

Definition

A symmetric matrix $\mathbf{A}$ is called positive definite (p.d.) if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all non-zero $\mathbf{x}$.

Positive Definite Matrix

Theorem

If $\mathbf{A}$ is a p.d matrix, then

(a) All diagonal elements and all eigenvalues of $\mathbf{A}$ are positive.

(b) $\text{tr}(\mathbf{A}) > 0$.

(c) $|\mathbf{A}| > 0$.

(d) There is a nonsingular $\mathbf{R}$ such that $\mathbf{A} = \mathbf{R}\mathbf{R}'$ (necessary and sufficient for $\mathbf{A}$ to be p.d.).

(e) $\mathbf{A}^{-1}$ is p.d.

Random Vectors

- Often we will want to work with multiple random variables at the same time.
- A random vector or random matrix is a vector or matrix whose elements are random variables.

Random Vectors

Definition

A random vector is a vector of random variables

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}.$$

Random Matrices

Definition

A random matrix is a matrix of random variables

$$\mathbf{Z} = (Z_{ij}) = \begin{pmatrix} Z_{11} & \cdots & Z_{1n} \\ \vdots & & \vdots \\ Z_{m1} & \cdots & Z_{mn} \end{pmatrix}.$$

Expectation

Definition

The mean or expectation of a random vector **X** is defined as

$$E[\mathbf{X}] = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{pmatrix}.$$

Expectation

Definition

The mean or expectation of a random matrix $\mathbf{Z} = (Z_{ij})$ is defined as $E[\mathbf{Z}] = (E[Z_{ij}])$.

Expectation

- Thus, if $\mathbf{X}$ is any random vector (or matrix), then $E[\mathbf{X}]$ is simply the element-wise expected value.
- For example, we can compute the elements of a random vector $E[X_i]$ as follows:

$$E[X_i] = \int x_i f(x_i) dx_i$$

where $f(x_i)$ is the marginal density of x_i.

Properties of Expectation

Properties

Let $\mathbf{a}$ be a constant vector and $\mathbf{A}, \mathbf{B}, \mathbf{C}$ constant matrices, then:

- $E[\mathbf{a}] = \mathbf{a}$ and $E[\mathbf{A}] = \mathbf{A}$.
- $E[\mathbf{X} + \mathbf{Y}] = E[\mathbf{X}] + E[\mathbf{Y}]$
- $E[\mathbf{AX}] = \mathbf{A}E[\mathbf{X}]$ i
- $E[\mathbf{AZB} + \mathbf{C}] = \mathbf{A}E[\mathbf{Z}]\mathbf{B} + \mathbf{C}$
- $E[\mathbf{X}'] = E[\mathbf{X}]'$
- $E[\text{tr}(\mathbf{X})] = \text{tr}(E[\mathbf{X}])$

Covariance Matrix

Definition

Let **X** and **Y** be random vectors of length m and n, respectively. Then:

$$
\begin{aligned}
\operatorname{cov}(\mathbf{X}, \mathbf{Y}) &= [\operatorname{cov}(X_i, Y_j)] \\
&\equiv \begin{pmatrix} \operatorname{cov}(X_1, Y_1) & \operatorname{cov}(X_1, Y_2) & \cdots & \operatorname{cov}(X_1, Y_n) \\ \operatorname{cov}(X_2, Y_1) & \operatorname{cov}(X_2, Y_2) & \cdots & \operatorname{cov}(X_2, Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}(X_m, Y_1) & \operatorname{cov}(X_m, Y_2) & \cdots & \operatorname{cov}(X_m, Y_n) \end{pmatrix}.
\end{aligned}
$$

Covariance Matrix

Alternatively we can write the covariance matrix as follows:

$$
\begin{aligned}
\mathrm{cov}(\mathbf{X}, \mathbf{Y}) &= E[(\mathbf{X} - E[\mathbf{X}])(\mathbf{Y} - E[\mathbf{Y}])'] \\
&= E\left[\begin{pmatrix} X_1 - E[X_1] \\ \vdots \\ X_m - E[X_m] \end{pmatrix} (Y_1 - E[Y_1], \cdots, Y_n - E[Y_n]) \right].
\end{aligned}
$$

Variance-covariance Matrix

- An important special case is when $\mathbf{X} = \mathbf{Y}$.
- In this case we write: $\mathrm{var}(\mathbf{X}) = \mathrm{cov}(\mathbf{X}, \mathbf{X})$.
- We call $\mathrm{var}(\mathbf{X})$ the variance-covariance matrix of $\mathbf{X}$.

Variance-covariance Matrix

Definition

If $\mathbf{X}$ is a random vector, the variance-covariance matrix of $\mathbf{X}$ is defined as

$$\begin{aligned} \mathrm{var}(\mathbf{X}) &\equiv [\mathrm{cov}(X_i, X_j)] \\ &\equiv \begin{pmatrix} \mathrm{var}(X_1) & \mathrm{cov}(X_1, X_2) & \cdots & \mathrm{cov}(X_1, X_n) \\ \mathrm{cov}(X_2, X_1) & \mathrm{var}(X_2) & \cdots & \mathrm{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(X_n, X_1) & \mathrm{cov}(X_n, X_2) & \cdots & \mathrm{var}(X_n) \end{pmatrix}. \end{aligned}$$

Variance-covariance Matrix

- We often write: $\sigma_{ij} = \mathrm{cov}(X_i, X_j)$.
- If X_i and X_j are independent, then $\sigma_{ij} = \mathrm{cov}(X_i, X_j) = 0$.

Variance-covariance Matrix

Alternatively we can write the variance-covariance matrix as follows:

$$
\begin{aligned}
\mathrm{var}(\mathbf{X}) &= E[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])'] \\
&= E\left[\begin{pmatrix} X_1 - E[X_1] \\ \vdots \\ X_n - E[X_n] \end{pmatrix} (X_1 - E[X_1], \cdots, X_n - E[X_n])\right].
\end{aligned}
$$

Example

- Assume $X_1, \ldots, X_n$ are independent random variables with common variance σ^2.
- Then the covariances are 0 and the variance-covariance matrix is equal to $\mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, or $\sigma^2 \mathbf{I}_n$.

Properties of variances and covariances

Properties

Let $\mathbf{A}$ and $\mathbf{B}$ be constant matrices and $\mathbf{a}$ and $\mathbf{b}$ be constant vectors, then

- $\text{cov}(\mathbf{X}, \mathbf{Y}) = \text{cov}(\mathbf{Y}, \mathbf{X})'$
- $\text{var}(\mathbf{X}) = [\text{var}(\mathbf{X})]'$
- $\text{cov}(\mathbf{X} + \mathbf{a}, \mathbf{Y} + \mathbf{b}) = \text{cov}(\mathbf{X}, \mathbf{Y})$
- $\text{var}(\mathbf{X} + \mathbf{a}) = \text{var}(\mathbf{X})$
- $\text{cov}(\mathbf{AX}, \mathbf{BY}) = \mathbf{A}\text{cov}(\mathbf{X}, \mathbf{Y})\mathbf{B}'$
- $\text{var}(\mathbf{AX}) = \mathbf{A}\text{var}(\mathbf{X})\mathbf{A}'$

Properties of variance

- We can write: $\mathrm{var}(\mathbf{X}) = E[\mathbf{X}\mathbf{X}'] - E[\mathbf{X}](E[\mathbf{X}])'$
- $\mathrm{var}(\mathbf{X})$ is always square and symmetric.
- $\mathrm{var}(\mathbf{X})$ is always positive semidefinite. It is positive definite unless one variable is a linear combination of the others.

Correlation Matrix

Definition

The correlation matrix of **X** is defined as

$$\begin{aligned}
\mathrm{corr}(\mathbf{X}) &= [\mathrm{corr}(X_i, X_j)] \\
&= \begin{pmatrix} 1 & \mathrm{corr}(X_1, X_2) & \cdots & \mathrm{corr}(X_1, X_n) \\ \mathrm{corr}(X_2, X_1) & 1 & \cdots & \mathrm{corr}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{corr}(X_n, X_1) & \mathrm{corr}(X_n, X_2) & \cdots & 1 \end{pmatrix}.
\end{aligned}$$

Correlation Matrix

- If we denote $\mathrm{var}(\mathbf{X})$ by $\Sigma = (\sigma_{ij})$, then the correlation matrix and covariance matrix are related by

$$\mathrm{var}(\mathbf{X}) = \mathrm{diag}(\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{nn}}) \times \mathrm{corr}(\mathbf{X}) \times \mathrm{diag}(\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{nn}}).$$

- This is easily seen using $\mathrm{corr}(X_i, X_j) = \mathrm{cov}(X_i, X_j)/\sqrt{\sigma_{ii}\sigma_{jj}}$.

Examples

- If $X_1, \ldots, X_n$ are exchangeable, they have a constant variance σ^2 and a constant correlation ρ between any pair of variables.
- Thus

$$\mathrm{var}(\mathbf{X}) = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}.$$

- Sometimes called an exchangeable covariance matrix.

Partitioned Random Vectors

- Suppose the random vector $\mathbf{Z}$ is partitioned in two two subsets of variables, which we denote $\mathbf{X}$ and $\mathbf{Y}$:

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}.$$

- Then the mean and variance covariance can be written:

$$E(\mathbf{Z}) = E\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \begin{pmatrix} E(\mathbf{X}) \\ E(\mathbf{Y}) \end{pmatrix}$$

and

$$\mathrm{cov}(\mathbf{Z}) = \begin{pmatrix} \mathrm{var}(\mathbf{X}) & \mathrm{cov}(\mathbf{X},\mathbf{Y}) \\ \mathrm{cov}(\mathbf{Y},\mathbf{X}) & \mathrm{var}(\mathbf{Y}) \end{pmatrix}.$$

Properties

Theorem

Let $\mathbf{Z} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}$. Then

$$\mathrm{cov}(\mathbf{Z}) = \begin{pmatrix} \mathrm{var}(\mathbf{X}) & \mathrm{cov}(\mathbf{X}, \mathbf{Y}) \\ \mathrm{cov}(\mathbf{Y}, \mathbf{X}) & \mathrm{var}(\mathbf{Y}) \end{pmatrix}.$$

Quadratic Forms

Definition

A quadratic form is a function $f : \mathbb{R}^n \mapsto \mathbb{R}$ of the form:

$$f(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i,j} A_{ij} x_i x_j.$$

Quadratic Forms

- We are often interested in working with random vectors that are combined into quadratic forms.
- The result is a function of random variables which is a scalar, and itself a random variable.
- We will often need to compute its expected value.

Quadratic Forms

Theorem

Let $\mathbf{X}$ be a random vector with $E[\mathbf{X}] = \boldsymbol{\mu}$ and $\mathrm{cov}(\mathbf{X}) = \Sigma$, and let $\mathbf{A}$ be a constant symmetric matrix. Then

$$E[\mathbf{X}'\mathbf{A}\mathbf{X}] = \mathrm{tr}(\mathbf{A}\Sigma) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}.$$

Quadratic Forms

Theorem

Let $\mathbf{X}$ be a random vector with $E[\mathbf{X}] = \mu$ and $\operatorname{cov}(\mathbf{X}) = \Sigma$, and let $\mathbf{A}$ be a constant symmetric matrix. Then

$$E[(\mathbf{X} - \mu)'\mathbf{A}(\mathbf{X} - \mu)] = \operatorname{tr}(\mathbf{A}\Sigma).$$

Example

- Let $X_1, \ldots, X_n$ be independent random variables with common mean μ and variance σ^2.
- Then the sample variance $S^2 = \sum_i (X_i - \bar{X})^2/(n-1)$ is an unbiased estimate of σ^2.