Advanced Methods in Biostatistics I

Lecture 11

Martin Lindquist

October 3, 2017

Multivariate normality

- In order to perform inference on the linear model, we typically assume the response variable follows a multivariate normal distribution.
- Hence, it is important to understand the properties of this distribution.
- However, before discussing the multivariate normal distribution, let's briefly review the univariate normal distribution.

The univariate normal distribution

- A random variable z follows a standard normal distribution if its density is

$$f_z(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2) \qquad -\infty < z < \infty$$

- We say $z \sim N(0, 1)$,
- A standard normal random variable has mean 0 and variance 1.

The univariate normal distribution

- To obtain a normal random variable with arbitrary mean μ and and variance σ^2 we use the transformation:

$$y = \mu + \sigma z.$$

- Here $E[y] = \mu$ and $Var(y) = \sigma^2$.

The univariate normal distribution

- Conversely,

$$z = (y - \mu)/\sigma$$

is standard normal if y is a normal random variable with mean μ and standard deviation σ.

The univariate normal distribution

- The density for a normal random variable with mean μ and standard deviation σ is given by

$$f_y(y) = f_z\left(\frac{y - \mu}{\sigma}\right)/\sigma.$$

- We write that $y \sim N(\mu, \sigma^2)$.

Moment generating function

Theorem

If $y \sim N(\mu, \sigma^2)$ the moment generating function (m.g.f.) of y is given by

$$M_y(t) \equiv E[e^{ty}] = \exp\{\mu t + t^2\sigma^2/2\}.$$

The multivariate normal distribution

- Suppose $z_1, z_2, \ldots z_p$ are independent identically distributed (i.i.d.) standard normal random variables.
- The joint density of $\mathbf{z} = (z_1, z_2, \ldots z_p)'$ is then given by

$$
\begin{aligned}
f_{\mathbf{z}}(\mathbf{z}) &= \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}} \exp(-z_i^2/2) \\
&= (2\pi)^{-p/2} \exp(-\mathbf{z}'\mathbf{z}/2)
\end{aligned}
$$

The multivariate normal distribution

- This is the multivariate standard normal distribution for a random vector $\mathbf{z}$ with mean $\mathbf{0}$ and variance $\mathbf{I}$.
- We write this as $\mathbf{z} \sim N_p(\mathbf{0}, \mathbf{I})$.
- Here p corresponds to the number of elements in y.

The multivariate normal distribution

- To obtain a multivariate normal random variable with arbitrary mean $\boldsymbol{\mu}$ and variance-covariance matrix Σ we use the transformation:

$$\mathbf{y} = \mu + \Sigma^{1/2}\mathbf{z}.$$

  where $\Sigma^{1/2}$ is the symmetric square root of Σ.

- Here $E[\mathbf{y}] = \mu$ and $Var(\mathbf{y}) = \Sigma^{1/2}\Sigma^{1/2} = \Sigma$, which is assumed to be positive definite.

The multivariate normal distribution

- Conversely,

$$\mathbf{z} = \Sigma^{-1/2}(\mathbf{y} - \mu).$$

  is a multivariate standard normal random variable if y is a multivariate normal random variable with mean μ and variance-covariance matrix Σ.

The multivariate normal distribution

- The density of the non-standard multivariate normal distribution is given by

$$(2\pi)^{-n/2}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{y}-\mu)'\Sigma^{-1}(\mathbf{y}-\mu)\right\}.$$

- In this setting, we say that $\mathbf{y} \sim N_p(\mu, \Sigma)$.

The multivariate normal distribution

- Note that all full row rank linear transformations of the normal are also normal.
- That is, $\mathbf{a} + \mathbf{A}\mathbf{y}$ is normal if $\mathbf{A}$ has full row rank.
- We will also show that all conditional and submarginal distributions of the multivariate normal are also normal.

Moment generating function

Theorem

If $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \Sigma)$ the moment generating function (m.g.f.) of $\mathbf{y}$ is

$$M_{\mathbf{y}}(\mathbf{t}) \equiv E[e^{\mathbf{t}'\mathbf{y}}] = \exp\{\boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\}.$$

Moment generating function

Two important properties of moment generating functions:

- If two random vectors have the same moment generating function, they have the same density.
- Two random vectors are independent if and only if their joint moment generating function factors into the product of their two separate moment generating functions;

Properties

Properties

Let $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \Sigma)$, and let $\mathbf{a}$ be a $p \times 1$ vector, $\mathbf{b}$ a $k \times 1$ vector, and $\mathbf{C}$ a $k \times p$ matrix with rank$= k \leq p$, then

- $x = \mathbf{a}'\mathbf{y} \sim N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\Sigma\mathbf{a})$.
- $\mathbf{x} = \mathbf{C}\mathbf{y} + \mathbf{b} \sim N_p(\mathbf{C}\boldsymbol{\mu} + \mathbf{b}, \mathbf{C}\Sigma\mathbf{C}')$.

Example

- Let $\mathbf{Z} = (Z_1, Z_2)' \sim N_2(\mathbf{0}, \mathbf{I})$, and let $\mathbf{A}$ be the linear transformation matrix

$$\mathbf{A} = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}.$$

- Let $\mathbf{Y} = (Y_1, Y_2)'$ be the linear transformation

$$\mathbf{Y} = \mathbf{AZ} = \begin{pmatrix} (Z_1 - Z_2)/2 \\ (Z_2 - Z_1)/2 \end{pmatrix}.$$

- Now $\mathbf{Y} \sim N(\mathbf{0}, \Sigma)$ where $\Sigma = \mathbf{AA}'$.

Properties

Theorem

If $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \Sigma)$, then any $r \times 1$ subvector of $\mathbf{y}$ has a r-variate normal distribution.

Properties

- It follows directly from this result that if $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \Sigma)$ then $y_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, \ldots p$.
- Thus, joint normality implies marginal normality.
- The converse is not necessarily true.

Partitioning

- Let $\mathbf{y} \sim N_n(\boldsymbol{\mu}, \Sigma)$ be partitioned as follows:

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix},$$

where $\mathbf{y}_1$ is $p \times 1$ and $\mathbf{y}_2$ is $q \times 1$ with $p + q = n$.

Partitioning

- Then, the mean and covariance matrix are correspondingly partitioned as

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$$

and

$$\begin{aligned} \Sigma &= \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \\ &= \begin{pmatrix} \mathrm{var}(\mathbf{y}_1) & \mathrm{cov}(\mathbf{y}_1, \mathbf{y}_2) \\ \mathrm{cov}(\mathbf{y}_2, \mathbf{y}_1) & \mathrm{var}(\mathbf{y}_2) \end{pmatrix}. \end{aligned}$$

Marginal distribution

- The marginal distributions are $\mathbf{y}_1 \sim N_p(\boldsymbol{\mu}_1, \Sigma_{11})$ and $\mathbf{y}_2 \sim N_q(\boldsymbol{\mu}_2, \Sigma_{22})$.

Independence

Theorem

If

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}$$

is $N_{p+q}(\boldsymbol{\mu}, \Sigma)$, then $\mathbf{y}_1$ and $\mathbf{y}_2$ are independent if $\Sigma_{12} = \mathbf{0}$.

Independence

- However, if $\mathbf{y}_1 \sim N_p(\boldsymbol{\mu}_1, \Sigma_{11})$ and $\mathbf{y}_2 \sim N_q(\boldsymbol{\mu}_2, \Sigma_{22})$, and $\Sigma_{12} = \Sigma_{21}' = \mathbf{0}$, this does not necessarily mean that $\mathbf{y}_1$ and $\mathbf{y}_2$ are independent.
- We also need $\mathbf{y}$ to be jointly normal.

Independence

Corollary

If $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \Sigma)$, then any two individual variables y_i and y_j are independent if $\sigma_{ij} = 0$.

Orthogonal transformations

Theorem

Let $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \sigma^2\mathbf{I})$, and let $\mathbf{T}$ be an orthogonal constant matrix. Then $\mathbf{Ty} \sim N_p(\mathbf{T}\boldsymbol{\mu}, \sigma^2\mathbf{I})$.

Orthogonal transformations

- The theorem states that mutually independent normal random variables with common variance remain mutually independent with common variance under orthogonal transformations.
- Orthogonal matrices correspond to rotations and reflections about the origin, i.e., they preserve the vector length:

$$||\mathbf{Ty}||^2 = (\mathbf{Ty})'(\mathbf{Ty}) = \mathbf{y}'\mathbf{T}'\mathbf{Ty} = \mathbf{y}'\mathbf{y} = ||\mathbf{y}||^2.$$

Conditional distributions

- Suppose that $\mathbf{y}_1$ and $\mathbf{y}_2$ are jointly multivariate normal with $\Sigma_{12} \neq 0$, i.e.

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right).$$

- Further assume that Σ_{11} is nonsingular.

Conditional distributions

- The conditional distribution of $\mathbf{y}_2 \mid \mathbf{y}_1$ is $N(\mu_{\mathbf{y}_2|\mathbf{y}_1}, \Sigma_{\mathbf{y}_2|\mathbf{y}_1})$, where

$$\begin{aligned} \mu_{\mathbf{y}_2|\mathbf{y}_1} &= \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{y}_1 - \mu_1) \\ \Sigma_{\mathbf{y}_2|\mathbf{y}_1} &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}'. \end{aligned}$$

Example

- Suppose that y and $\mathbf{x}$ are jointly multivariate normal with

$$\begin{pmatrix} y \\ \mathbf{X} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_y \\ \boldsymbol{\mu}_x \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \Sigma_{xy}' \\ \Sigma_{xy} & \Sigma_x \end{pmatrix}\right).$$

  where y is a scalar and $\mathbf{X}$ is a $p \times 1$ vector.

- Consider now predicting y given $\mathbf{X} = \mathbf{x}$.
- A good estimate for this would be $E[y \mid \mathbf{X} = \mathbf{x}]$.

Example

- Our results suggest that $\mathbf{y} \mid \mathbf{X} = \mathbf{x}$ is normal with mean:

$$\begin{aligned}
\boldsymbol{\mu}_{\mathbf{y}|\mathbf{x}} &= \mu_y + \Sigma_{xy}' \Sigma_x^{-1} (\mathbf{x} - \boldsymbol{\mu}_x) \\
&= \mu_y - \boldsymbol{\mu}_x' \Sigma_x^{-1} \Sigma_{xy} + \mathbf{x}' \Sigma_x^{-1} \Sigma_{xy} \\
&= \beta_0 + \mathbf{x}' \boldsymbol{\beta}
\end{aligned}$$

where

$$\beta_0 = \mu_y - \boldsymbol{\mu}_x' \Sigma_x^{-1} \Sigma_{xy}$$

and

$$\boldsymbol{\beta} = \Sigma_x^{-1} \Sigma_{xy}.$$

Example

- Consider the case of simple linear regression.
- Here,
$$\beta_1 = \frac{Cov(x, y)}{Var(x)} = \rho(x, y)\frac{\sqrt{Var(y)}}{\sqrt{Var(x)}}.$$
  and
$$\beta_0 = \bar{y} - \bar{x}\beta_1.$$

Example

- Hence, the conditional mean in this case mirrors the linear model.
- The slope is the inverse of the variance-covariance matrix of the predictors times the cross correlations between the predictors and the response.

Example

- This example provides motivation for the linear model in cases when the joint normality of the predictor and response is conceptually reasonable.
- Though we note that such joint normality is not always reasonable, such as when the predictors are binary, even though the linear model remains well justified.