

# STA 360/602L: MODULE 4.1

## MULTIVARIATE NORMAL MODEL I

DR. OLANREWAJU MICHAEL AKANDE

# MULTIVARIATE DATA

- So far we have only considered basic models with scalar/univariate outcomes,  $Y_1, \dots, Y_n$ .
- In practice however, outcomes of interest are actually often multivariate, e.g.,
  - Repeated measures of weight over time in a weight loss study
  - Measures of multiple disease markers
  - Tumor counts at different locations along the intestine
- Longitudinal data is just a special case of multivariate data.
- Interest then is often on how multiple outcomes are correlated, and on how that correlation may change across outcomes or time points.

# MULTIVARIATE NORMAL DISTRIBUTION

- The most common model for multivariate outcomes is the **multivariate normal distribution**.
- Let  $\mathbf{Y} = (Y_1, \dots, Y_p)^T$ , where  $p$  represents the dimension of the multivariate outcome variable for a single unit of observation.
- For multiple observations,  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ip})^T$  for  $i = 1, \dots, n$ .
- $\mathbf{Y}$  follows a multivariate normal distribution, that is,  $\mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ , if

$$p(\mathbf{y}|\boldsymbol{\mu}, \Sigma) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\},$$

where  $|\Sigma|$  denotes the determinant of  $\Sigma$ .

# MULTIVARIATE NORMAL DISTRIBUTION

If  $\mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ , then

- $\boldsymbol{\mu}$  is the  $p \times 1$  mean vector, that is,

$$\boldsymbol{\mu} = \mathbb{E}[\mathbf{Y}] = \{\mathbb{E}[Y_1], \dots, \mathbb{E}[Y_p]\} = (\mu_1, \dots, \mu_p)^T.$$

- $\Sigma$  is the  $p \times p$  **positive definite and symmetric** covariance matrix, that is,
  - $\Sigma = \{\sigma_{jk}\}$ , where  $\sigma_{jk}$  denotes the covariance between  $Y_j$  and  $Y_k$ .
- $Y_1, \dots, Y_p$  may be linearly dependent depending on the structure of  $\Sigma$ , which characterizes the association between them.
- For each  $j = 1, \dots, p$ ,  $Y_j \sim \mathcal{N}(\mu_j, \sigma_{jj})$ .

# BIVARIATE NORMAL DISTRIBUTION

- In the bivariate case, we have

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathcal{N}_2 \left[ \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} = \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22} = \sigma_2^2 \end{pmatrix} \right],$$

and  $\sigma_{12} = \sigma_{21} = \text{Cov}[Y_1, Y_2]$ .

- The correlation between  $Y_1$  and  $Y_2$  is defined as

$$\rho_{1,2} = \frac{\text{Cov}[Y_1, Y_2]}{\sqrt{\text{Var}[Y_1]}\sqrt{\text{Var}[Y_2]}} = \frac{\sigma_{12}}{\sigma_1\sigma_2}.$$

- $-1 \leq \rho_{1,2} \leq 1$ .
- Correlation coefficient is free of the measurement units.

# BACK TO THE MULTIVARIATE NORMAL

- There are many special properties of the multivariate normal as we will see as we continue to work with the distribution.
- First, dependence between any  $Y_j$  and  $Y_k$  does not depend on the other  $p - 2$  variables.
- Second, while generally, **independence implies zero covariance**, for the normal family, the converse is also true. That is, **zero covariance also implies independence**.
- Thus, the covariance  $\Sigma$  carries a lot of information about marginal relationships, especially **marginal independence**.
- If  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_p) \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$ , that is,  $\epsilon_1, \dots, \epsilon_p \stackrel{iid}{\sim} \mathcal{N}(0, 1)$ , then

$$\mathbf{Y} = \boldsymbol{\mu} + A\boldsymbol{\epsilon} \Rightarrow \mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$$

holds for any matrix square root  $A$  of  $\Sigma$ , that is,  $AA^T = \Sigma$  (see Cholesky decomposition).

# CONDITIONAL DISTRIBUTIONS

- Partition  $\mathbf{Y} = (Y_1, \dots, Y_p)^T$  as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \sim \mathcal{N}_p \left[ \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right],$$

where

- $\mathbf{Y}_1$  and  $\boldsymbol{\mu}_1$  are  $q \times 1$ ,
  - $\mathbf{Y}_2$  and  $\boldsymbol{\mu}_2$  are  $(p - q) \times 1$ ,
  - $\boldsymbol{\Sigma}_{11}$  is  $q \times q$ , and
  - $\boldsymbol{\Sigma}_{22}$  is  $(p - q) \times (p - q)$ , with  $\boldsymbol{\Sigma}_{22} > 0$ .
- Then,

$$\mathbf{Y}_1 | \mathbf{Y}_2 = \mathbf{y}_2 \sim \mathcal{N}_q \left( \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{y}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \right).$$

- Marginal distributions are once again normal, that is,

$$\mathbf{Y}_1 \sim \mathcal{N}_q (\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}); \quad \mathbf{Y}_2 \sim \mathcal{N}_{p-q} (\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}).$$

# CONDITIONAL DISTRIBUTIONS

- In the bivariate normal case with

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathcal{N}_2 \left[ \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} = \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22} = \sigma_2^2 \end{pmatrix} \right],$$

we have

$$Y_1 | Y_2 = y_2 \sim \mathcal{N} \left( \mu_1 + \frac{\sigma_{12}}{\sigma_2^2} (y_2 - \mu_2), \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2} \right).$$

which can also be written as

$$Y_1 | Y_2 = y_2 \sim \mathcal{N} \left( \mu_1 + \frac{\sigma_1}{\sigma_2} \rho (y_2 - \mu_2), (1 - \rho^2) \sigma_1^2 \right).$$



# MULTIVARIATE NORMAL LIKELIHOOD

- Suppose  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ip})^T \sim \mathcal{N}_p(\boldsymbol{\theta}, \Sigma)$ ,  $i = 1, \dots, n$ .
- Write  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)^T$ . The resulting likelihood can then be written as

$$p(\mathbf{Y}|\boldsymbol{\theta}, \Sigma) = \prod_{i=1}^n (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y}_i - \boldsymbol{\theta})^T \Sigma^{-1} (\mathbf{y}_i - \boldsymbol{\theta}) \right\}$$
$$\propto |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\theta})^T \Sigma^{-1} (\mathbf{y}_i - \boldsymbol{\theta}) \right\}.$$

- It will be super useful to be able to write the likelihood in two different formulations depending on whether we care about the posterior of  $\boldsymbol{\theta}$  or  $\Sigma$ .

# MULTIVARIATE NORMAL LIKELIHOOD

- For inference on  $\boldsymbol{\theta}$ , it is convenient to write  $p(\mathbf{Y}|\boldsymbol{\theta}, \Sigma)$  as

$$\begin{aligned} p(\mathbf{Y}|\boldsymbol{\theta}, \Sigma) &\propto \underbrace{|\Sigma|^{-\frac{n}{2}}}_{\text{does not involve } \boldsymbol{\theta}} \cdot \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\theta})^T \Sigma^{-1} (\mathbf{y}_i - \boldsymbol{\theta}) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i^T - \boldsymbol{\theta}^T) \Sigma^{-1} (\mathbf{y}_i - \boldsymbol{\theta}) \right\} \\ &= \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left[ \underbrace{\mathbf{y}_i^T \Sigma^{-1} \mathbf{y}_i}_{\text{does not involve } \boldsymbol{\theta}} - \underbrace{\mathbf{y}_i^T \Sigma^{-1} \boldsymbol{\theta} - \boldsymbol{\theta}^T \Sigma^{-1} \mathbf{y}_i}_{\text{same term}} + \boldsymbol{\theta}^T \Sigma^{-1} \boldsymbol{\theta} \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^n [\boldsymbol{\theta}^T \Sigma^{-1} \boldsymbol{\theta} - 2\boldsymbol{\theta}^T \Sigma^{-1} \mathbf{y}_i] \right\} \\ &= \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \boldsymbol{\theta}^T \Sigma^{-1} \boldsymbol{\theta} - \frac{1}{2} \sum_{i=1}^n (-2)\boldsymbol{\theta}^T \Sigma^{-1} \mathbf{y}_i \right\} \\ &= \exp \left\{ -\frac{1}{2} n\boldsymbol{\theta}^T \Sigma^{-1} \boldsymbol{\theta} + \boldsymbol{\theta}^T \Sigma^{-1} \sum_{i=1}^n \mathbf{y}_i \right\} \\ &= \exp \left\{ -\frac{1}{2} \boldsymbol{\theta}^T (n\Sigma^{-1}) \boldsymbol{\theta} + \boldsymbol{\theta}^T (n\Sigma^{-1} \bar{\mathbf{y}}) \right\}, \end{aligned}$$

where  $\bar{\mathbf{y}} = (\bar{y}_1, \dots, \bar{y}_p)^T$ .

# MULTIVARIATE NORMAL LIKELIHOOD

- For inference on  $\Sigma$ , we need to rewrite the likelihood a bit.
- First a few results from matrix algebra:
  1.  $\text{tr}(\mathbf{A}) = \sum_{j=1}^p a_{jj}$ , where  $a_{jj}$  is the  $j$ th diagonal element of a square  $p \times p$  matrix  $\mathbf{A}$ , where  $\text{tr}(\cdot)$  is the **trace function** (sum of diagonal elements).
  2. Cyclic property:

$$\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{BCA}) = \text{tr}(\mathbf{CAB}),$$

given that the product  $\mathbf{ABC}$  is a square matrix.

3. If  $\mathbf{A}$  is a  $p \times p$  matrix, then for a  $p \times 1$  vector  $\mathbf{x}$ ,

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \text{tr}(\mathbf{x}^T \mathbf{A} \mathbf{x})$$

holds by (1), since  $\mathbf{x}^T \mathbf{A} \mathbf{x}$  is a scalar.

4.  $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$ .

# MULTIVARIATE NORMAL LIKELIHOOD

- It is convenient to rewrite  $p(\mathbf{Y}|\boldsymbol{\theta}, \Sigma)$  as

$$\begin{aligned} p(\mathbf{Y}|\boldsymbol{\theta}, \Sigma) &\propto |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\theta})^T \Sigma^{-1} (\mathbf{y}_i - \boldsymbol{\theta}) \right\} \\ &\quad \underbrace{\hspace{10em}}_{\text{no algebra/change yet}} \\ &= |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \underbrace{\text{tr} [(\mathbf{y}_i - \boldsymbol{\theta})^T \Sigma^{-1} (\mathbf{y}_i - \boldsymbol{\theta})]}_{\text{by result 3}} \right\} \\ &= |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \underbrace{\text{tr} [(\mathbf{y}_i - \boldsymbol{\theta})(\mathbf{y}_i - \boldsymbol{\theta})^T \Sigma^{-1}]}_{\text{by cyclic property}} \right\} \\ &= |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \underbrace{\sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\theta})(\mathbf{y}_i - \boldsymbol{\theta})^T \Sigma^{-1}}_{\text{by result 4}} \right] \right\} \\ &= |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\mathbf{S}_\theta \Sigma^{-1}] \right\}, \end{aligned}$$

where  $\mathbf{S}_\theta = \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\theta})(\mathbf{y}_i - \boldsymbol{\theta})^T$  is the residual sum of squares matrix.

# WHAT'S NEXT?

MOVE ON TO THE READINGS FOR THE NEXT MODULE!