Module 4: The Multivariate Normal Distribution

Rebecca C. Steorts

Hoff, Section 7.4

Agenda

- Motivational reading comprehension case study
- Introduction/Review of vectors, matrices
- Population means/covariance matrices
- General multivariate notation
- Background on linear algebra (with practice exercises)
- Determinants, traces, quadratic forms

Agenda

- The multivariate normal distribution (MVN)
- Exercise with the MVN
- Case study on reading comprehension

What you should learn

- You will learn background on linear algebra
- You will learn how to model multivariate data, where we consider an application to reading comprehension tests
- You will learn the notation for multivariate random variables
- You will learn about the multivariate density of the normal

Goal

The goal of this module is to be able **to understand how to work with multivariate distributions**, such as the multivariate normal distribution.

We also want to understand how univariate models that we have used in the past translate to the multivariate setting.

Before we can delve in, we must review background on matrices, vectors, and **multivariate notation**. We also must review background on **linear algebra**.

Example: Reading Comprehension

A sample of 22 children are given reading comprehension tests before and after receiving a particular instructional method.[1]

Each student i will then have two scores, $Y_{i,1}$ and $Y_{i,2}$ denoting the pre- and post-instructional scores respectively.

Denote each student's pair of scores by the vector $\boldsymbol{Y}_i$

$$\boldsymbol{Y}_i = \begin{pmatrix} Y_{i,1} \\ Y_{i,2} \end{pmatrix} = \begin{pmatrix} \text{score on first test} \\ \text{score on second test} \end{pmatrix}$$

where $i = 1, \ldots, n$ and $p = 2$.

[1] This example follows Hoff (Section 7.4, p. 112).

Example: Reading Comprehension

$$\boldsymbol{X}_{n\times p} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ x_{31} & x_{32} & \dots & x_{3p} \\ x_{i1} & x_{i2} & \dots & x_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}.$$

- A row of $\boldsymbol{X}_{n\times p}$ represents a covariate we might be interested in, such as age of a person.
- Denote x_i $(p \times 1)$ as the ith row vector of the $X_{n\times p}$ matrix.

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}$$

Example: Reading Comprehension

We may be interested in the population mean $\boldsymbol{\mu}_{p \times 1}$.

$$E[\boldsymbol{Y}] =: E[\boldsymbol{Y}_i] = \begin{pmatrix} Y_{i,1} \\ Y_{i,2} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \boldsymbol{\mu}$$

Example: Reading Comprehension

We also may be interested in the population covariance matrix, $\Sigma_{p\times p}$.

By definition:

$$\begin{aligned}
\Sigma &= Cov(\boldsymbol{Y}) && (1)\\
&= \begin{pmatrix} E[Y_1^2]-E[Y_1]^2 & E[Y_1Y_2]-E[Y_1]E[Y_2] \\ E[Y_1Y_2]-E[Y_1]E[Y_2] & E[Y_2^2]-E[Y_2]^2 \end{pmatrix} && (2)\\
&= \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix} && (3)
\end{aligned}$$

Remark: $Cov(Y_1) = Var(Y_1) = \sigma_1^2$. $\quad Cov(Y_1, Y_2) = \sigma_{1,2}$.

How do we expand this beyond our reading comprehension example

We introduced our notation based upon a specific example to reading comprehension.

How can we make this more general and applicable to general case studies and problems?

General Notation

Assume that $\boldsymbol{y}_{p\times 1} \sim (\mu_{p\times 1}, \Sigma_{p\times p})$.

$$\boldsymbol{y}_{p\times 1} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}.$$

$$\boldsymbol{\mu}_{p\times 1} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}$$

$$\Sigma_{p\times p} = Cov(\boldsymbol{y}) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_p^2 \end{pmatrix}.$$

Background

Before proceeding, we need to review some basic concepts from linear algebra:

1. Basic properties of matrices
2. Useful lemmas for working with matrices

The determinant of a matrix

Assume a matrix $A_{n\times n}$ is invertible. The

$$\det(A) = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in},$$

where Aij are the co-factors and are computed from

$$A_{ij} = (-1)^{i+j} det(M_{ij}).$$

M_{ij} is known as the minor matrix and is the matrix you get if you eliminate row i and column j from matrix A. You must apply this technique recursively.

We only use this technique when doing such calculations by hand or in proof-based approaches.

The determinant of a matrix

- How on earth do I use the complicated formula on the pervious slide.

Easy: Use the `det` command in `R` when faced with an application.

- You will also see a determinant in the definition of the multivariate normal distribution.

Important point: It's just a function and we typically do not need to evalute it in this course!

The trace of a matrix

Assume a matrix $H_{p \times p}$.

$$\text{trace}(H) = \sum_i h_{ii},$$

where h_{ii} are the diagonal elements of H.

The trace of a matrix

$$H = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

What is $\mathrm{tr}(H)$?

(Take 1 minute to complete this.)

Linear Algebra Tricks

Suppose that A is $n \times n$ matrix and suppose that B is a $n \times n$ matrix.

Lemma 1:

$$tr(AB) = tr(BA)$$

Proof: Exercise. (You can find the proof at the end of the slides to check your work).

Linear Algebra Tricks

Lemma 2:

Suppose $\boldsymbol{x}_{\boldsymbol{p}\times\mathbf{1}}$ is a vector and A is a $p \times p$ dimensional matrix.

Then $\boldsymbol{x}^T_{\mathbf{1}\times\boldsymbol{p}} A_{p\times p} \boldsymbol{x}_{\boldsymbol{p}\times\mathbf{1}}$ is called a **quadratic form**.

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

Proof: Exercise.

Linear Algebra Tricks

Proof of Lemma 2:

$$
\begin{aligned}
tr(\boldsymbol{x}^T A\boldsymbol{x}) &= \sum_i (x^t A x)_{ii} && (4)\\
&= (\boldsymbol{x}^T(A\boldsymbol{x})) && (5)\\
&= tr(A\boldsymbol{x}\boldsymbol{x}^T) \text{ (by Lemma 1)} && (6)
\end{aligned}
$$

$$
\begin{aligned}
tr(\boldsymbol{x}^T A\boldsymbol{x}) &= \sum_i (x^t A x)_{ii} && (7)\\
&= ((\boldsymbol{x}^T A)\boldsymbol{x}) && (8)\\
&= tr(\boldsymbol{x}\boldsymbol{x}^T A) \text{ (by Lemma 1)} && (9)
\end{aligned}
$$

Notation

- MVN is generalization of univariate normal.
- For the MVN, we write $\boldsymbol{y} \sim \mathcal{MVN}(\boldsymbol{\mu}, \Sigma)$.
- The $(i, j)^{\text{th}}$ component of Σ is the covariance between Y_i and Y_j (so the diagonal of Σ gives the component variances).

Example: $Cov(Y_1, Y_2)$ is just one element of the matrix Σ.

Multivariate Normal

Just as the probability density of a scalar normal is

$$p(x) = \left(2\pi\sigma^2\right)^{-1/2} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}, \tag{10}$$

the probability density of the multivariate normal is

$$p(\boldsymbol{x}) = (2\pi)^{-p/2}(\det\Sigma)^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}. \tag{11}$$

Univariate normal is special case of the multivariate normal with a one-dimensional mean "vector" and a one-by-one variance "matrix."

Standard Multivariate Normal Distribution

Lemma 3.

Consider

$$Z_1, \ldots, Z_n \overset{iid}{\sim} N(0, 1).$$

Show that

$$Z_1, \ldots, Z_n \sim MVN(\mathbf{0}, I_{n \times n}).$$

Proof of Lemma 3

Proof:

$$f_z(z) = \prod_{i=1}^{n} (2\pi)^{-1/2} e^{-z_i^2/2} \tag{12}$$

$$= (2\pi)^{-n/2} e^{\sum_i -z_i^2/2} \tag{13}$$

$$= (2\pi)^{-n/2} e^{-z^T z/2}. \tag{14}$$

The last line is follows since $\sum_i -z_i^2 = -z^T z$.

Thus, $Z_1, \ldots, Z_n \sim \mathrm{MVN}(\mathbf{0}, I)$.

Likelihood

$$p(\boldsymbol{y} \mid \boldsymbol{\theta}, \Sigma) = \prod_{i=1}^{n} (2\pi)^{-p/2} \det \Sigma^{-1/2} \exp\left\{-\frac{1}{2}(y_i - \boldsymbol{\theta})^T \Sigma^{-1} (y_i - \boldsymbol{\theta})\right\} \tag{15}$$

$$\propto \exp -\frac{1}{2}\left\{\sum_i y_i^T \Sigma^{-1} y_i - 2\sum_i \boldsymbol{\theta}^T \Sigma^{-1} y_i + \sum_i \boldsymbol{\theta}^T \Sigma^{-1} \boldsymbol{\theta}\right\} \tag{16}$$

$$\propto \exp -\frac{1}{2}\left\{-2\boldsymbol{\theta}^T \Sigma^{-1} n\bar{y} + n\boldsymbol{\theta}^T \Sigma^{-1} \boldsymbol{\theta}\right\} \tag{17}$$

$$\propto \exp -\frac{1}{2}\left\{-2\boldsymbol{\theta}^T b_1 + \boldsymbol{\theta}^T A_1 \boldsymbol{\theta}\right\}, \tag{18}$$

where

$$b_1 = \Sigma^{-1} n\bar{y}, \quad A_1 = n\Sigma^{-1}$$

and

$$\bar{y} := \left(\frac{1}{n}\sum_i y_{i1}, \ldots, \frac{1}{n}\sum_i y_{ip}\right)^T.$$

Working with Multivariate Normal Distribution

The `R` package, `mvtnorm`, contains functions for evaluating and simulating from a multivariate normal density.

```r
library(ggplot2)
library(MASS)
library(mvtnorm)
library(car)
```

```
## Loading required package: carData
```

Simulating Data

Simulate a single multivariate normal random vector using the `rmvnorm` function.

```r
# Each row corresponds to a sample
# Here we have one sample (one row)
rmvnorm(n = 1, mean = rep(0, 2), sigma = diag(2))
```

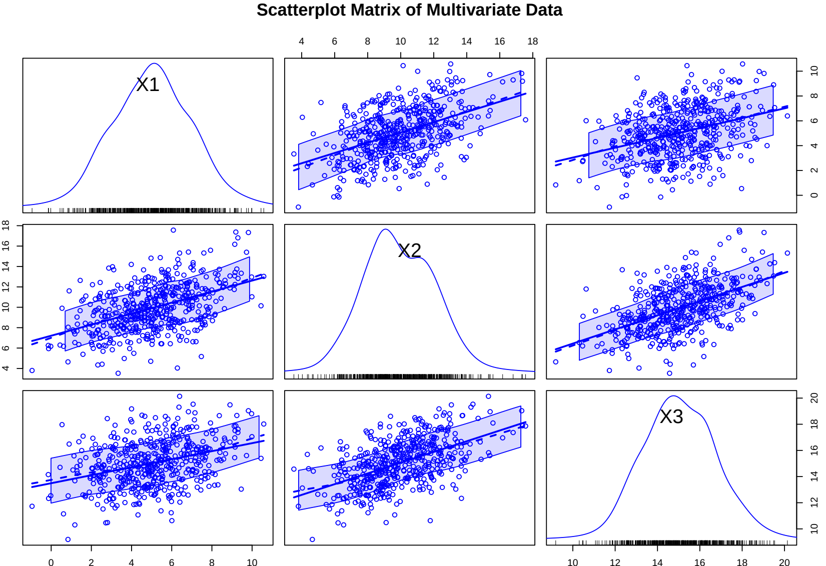
```
##           [,1]       [,2]
## [1,] 0.4423064 0.05889527
```

Simulation Study and Investigation

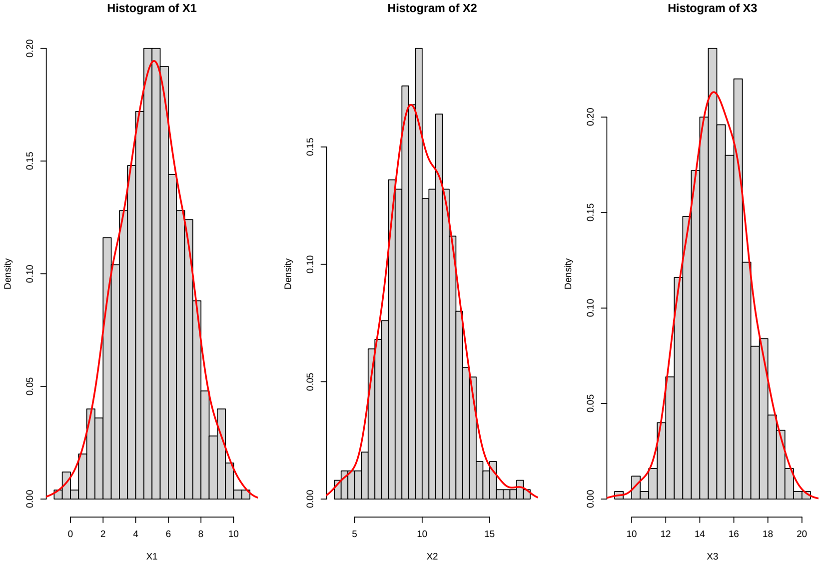
```r
n <- 500  # Number of samples
mu <- c(5, 10, 15)  # Mean vector
Sigma <- matrix(c(4, 2, 1,
                  2, 5, 2,
                  1, 2, 3),
                ncol = 3)  # Covariance matrix
data <- mvrnorm(n = n, mu = mu, Sigma = Sigma)
colnames(data) <- c("X1", "X2", "X3")
data <- as.data.frame(data)
```

Scatterplot Matrix

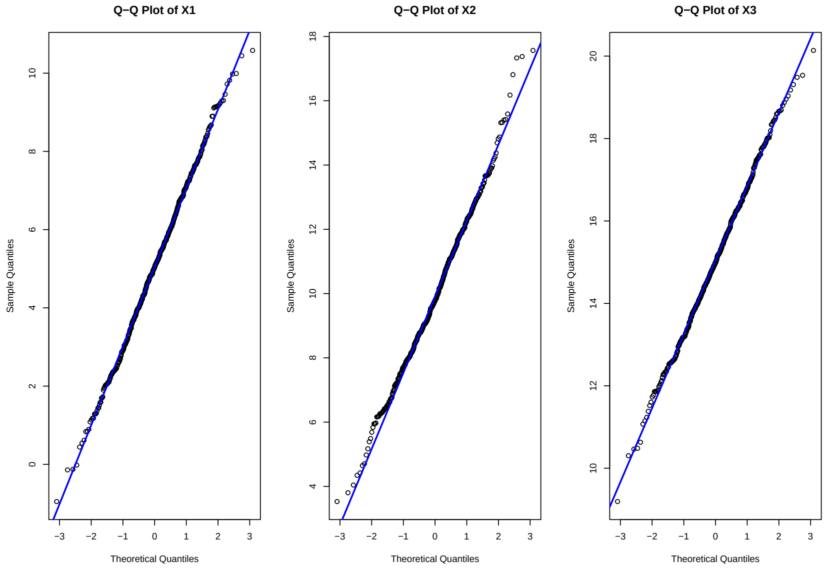
```
scatterplotMatrix(~ X1 + X2 + X3, data = data,
                  main = "Scatterplot Matrix of Multivariat
```



Marginal Histogram

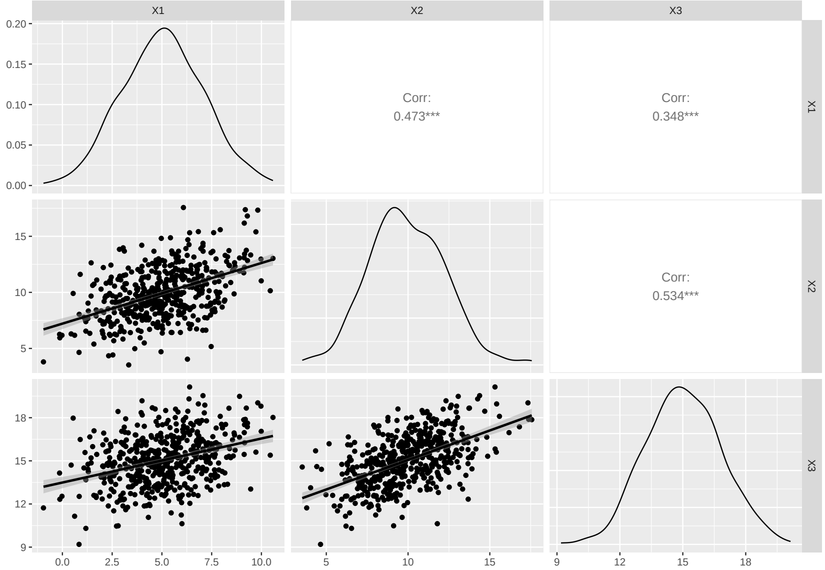


Q-Q Plots



Q−Q Plot of X1
Q−Q Plot of X2
Q−Q Plot of X3
Sample Quantiles
Theoretical Quantiles

Not sure



X1
X2
X3
Corr:
0.473***
Corr:
0.348***
Corr:
0.534***
0.20
0.15
0.10
0.05
0.00
15
10
5
18
15
12
9
0.0
2.5
5.0
7.5
10.0
5
10
15
9
12
15
18

Detailed Takeaways on Background

- Understanding vectors, matrices and notation
- Understanding how to write multivariate notation for a conceptual problem
- Understanding how to write general multivariate notation
- Background on linear algebra
- Determinants, traces, quadratic forms
- Knowing how to do simple proofs such as the exercises from class

Detailed Takeaways on Multivariate Normal Models

- The multivariate normal distribution (MVN)
- Exercise with the MVN

Proof of Lemma 1

$$tr(AB) = tr(BA)$$

Proof: Suppose that $A_{n\times n}$ and $B_{n\times n}$.

Recall that by definition $tr(A) = \sum_{i=1}^{n} a_{ii}$. By definition

$$tr(AB) = \sum_{i=1}^{n} (AB)_{ii} \tag{19}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ji} \tag{20}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} b_{ji} a_{ij} \tag{21}$$

$$= \sum_{i=1}^{n} (BA)_{ii} = tr(BA) \tag{22}$$