

Chapter 4: Multivariate Analysis and Computations in Calculus

I. Statistics

4.1. Graphical methods and variance-covariance matrix

Example:

The famous iris data were collected by Anderson (1935) and given and analyzed by Fisher (1936). There are 150 flowers. They belong to 3 species, Iris setosa, Iris virginica, and Iris versicolor. There are 50 of each of the 3 species. Each flower has 4 measurements (variables): **sepal length (Z1)**, **sepal width (Z2)**, **petal length (Z3)** and **petal width (Z4)**.

Data:

$$X_i = \begin{bmatrix} X_{i1} \\ X_{i2} \\ X_{i3} \\ X_{i4} \end{bmatrix} \equiv \begin{bmatrix} \text{sepal length} \\ \text{sepal width} \\ \text{petal length} \\ \text{petal width} \end{bmatrix}, i = 1, \dots, 150,$$

and the data matrix

$$X = \begin{bmatrix} X_1^t \\ X_2^t \\ \vdots \\ X_{150}^t \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & X_{13} & X_{14} \\ X_{21} & X_{22} & X_{23} & X_{24} \\ \vdots & \vdots & \vdots & \vdots \\ X_{(150)1} & X_{(150)2} & X_{(150)3} & X_{(150)4} \end{bmatrix}.$$

Example (Splus):

```
iris  
help(iris)  
attributes(iris)  
ir=rbind(iris[,1],iris[,2],iris[,3])  
### graphical methods  
brush(ir)  
pairs(ir)  
  
### basic statistics  
colMeans(ir)      ### column means  
var(ir)           ### variance-covariance matrix  
cor(ir)           ### correlation matrix
```

Note:

The plots of sepal length versus petal length, sepal length versus petal width, and petal length versus petal width indicate the two variables in these plots might be highly correlated.