Chapter 1: Some Multivariate Measures

Contents

1. Introduction

The majority of data sets collected by researchers are of multivariate type including several measurements, features, observations and units. Examples:

- Psychologists and other behavioral scientists often record the values of several different cognitive variables on a number of subjects.
- Educational researchers may be interested in the examination marks obtained by students for a variety of different subjects.
- Archaeologists may make a set of measurements on artefacts of interest.
- Environmentalists might assess pollution levels of a set of cities along with noting other characteristics of the cities related to climate and human ecology.

2. Representation

Mostly in a rectangular format

the elements of each row = variable values of a particular unit

the elements of the columns = values taken by a particular variable.

Unit	Variable 1	...	Variable q
1	x_{11}	...	x_{1q}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	x_{n1}	...	x_{nq}

n units, q variables recorded on each unit. x_{ij} value of the jth variable on ith unit.

A hypothetical Example

Table 1.1: hypo data. Hypothetical Set of Multivariate Data.

individual	sex	age	IQ	depression	health	weight
1	Male	21	120	Yes	Very good	150
2	Male	43	NA	No	Very good	160
3	Male	22	135	No	Average	135
4	Male	86	150	No	Very poor	140
5	Male	60	92	Yes	Good	110
6	Female	16	130	Yes	Good	110
7	Female	NA	150	Yes	Very good	120
8	Female	43	NA	Yes	Average	120
9	Female	22	84	No	Average	105
10	Female	80	70	No	Good	100

Here, the number of units (people in this case) is $n = 10$, with the number of variables being $q = 7$.

R Code to make Table 1.1:

```
gender<-c("Male","Female")
sex <- rep(gender, each=5)
age <- c(21, 43, 22, 86, 60, 16, NA, 43, 22, 80)
IQ <- c(120, NA, 135, 150, 92, 130, 150, NA, 84, 70)
depression <- c("Yes", "No", "No", "No", "Yes", "Yes", "Yes", "Yes", "No", "No")
health <- c("Very good", "Very good", "Average", "Very poor", "Good", "Good",
"Very good", "Average","Average","Good")
weight <- c(150, 160, 135, 140, 110, 110, 120, 120, 105, 100)
hypo <- data.frame(sex,age,IQ,depression,health,weight)
```

Extraction

```
> hypo [1:2, c("health", "weight")]
health weight
1 Very good 150
2 Very good 160
```

Level of measurements

Nominal: Unordered categorical variables. Examples include treatment allocation, the sex of the respondent, hair color, presence or absence of depression, and so on.
Ordinal: Where there is an ordering but no implication of equal distance between the different points of the scale. Examples include social class, self-perception of health (each coded from I to V, say), and educational level (no schooling, primary, secondary, or tertiary education).
Interval: Where there are equal differences between successive points on the scale but the position of zero is arbitrary. The classic example is the measurement of temperature using the Celsius or Fahrenheit scales.
Ratio: The highest level of measurement, where one can investigate the relative magnitudes of scores as well as the differences between them. The position of zero is fixed. The classic example is the absolute measure of temperature (in Kelvin, for example), but other common ones includes age (or any other time from a fixed event), weight, and length.

3. Covariances, Correlations, and Distances

Expectation

The mean or expectation of a random $q \times 1$ vector $\boldsymbol{X} = (X_1, \dots, X_q)'$ is defined to be the vector of expectations

$$\boldsymbol{\mu} = E(\boldsymbol{X}) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_q) \end{pmatrix}$$

Covariance

The covariance of two random variables is a measure of their linear dependence.
$\sigma_{ij}^2 = Cov(X_i, X_j) = E(X_i - \mu_i)(X_j - \mu_j), \qquad \mu_i = E(X_i), \quad \mu_j = E(X_j)$. Also
$\sigma_i^2 = Var(X_i) = Cov(X_i, X_i) = E(X_i - \mu_i)^2$.
In a multivariate data set with q observed variables, there are q variances and $q(q-1)/2$ covariances. These quantities can be conveniently arranged in a $q \times q$ symmetric matrix, $\boldsymbol{\Sigma}$ where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1q} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{q1} & \sigma_{q2} & \cdots & \sigma_q^2 \end{pmatrix}.$$

Note that $\sigma_{ij} = \sigma_{ji}$. This matrix is generally known as the variance-covariance matrix or simply the covariance matrix of the data.
$\boldsymbol{\Sigma} = Cov(\boldsymbol{X}) = E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'] = E(\boldsymbol{X}\boldsymbol{X}') - \boldsymbol{\mu}\boldsymbol{\mu}'$.

Obviously $\boldsymbol{\Sigma}$ is symmetric, i.e., $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}'$. Indeed, the class of covariance matrices coincides with the class of ***non-negative*** definite matrices.
Recall that an $q \times q$ symmetric matrix $\boldsymbol{A}$ is called non-negative (positive definite) definite if

$$\boldsymbol{\alpha}'\boldsymbol{A}\boldsymbol{\alpha} \geq (>) 0 \qquad for\ all \qquad \boldsymbol{\alpha} \in \mathbb{R}^q.$$

Lemma 3-1: The $\boldsymbol{q} \times \boldsymbol{q}$ matrix $\boldsymbol{\Sigma}$ is a covariance matrix iff (if and only if) it is non-negative definite.

Proof: Suppose $\boldsymbol{\Sigma}$ is the covariance matrix of a random vector $\boldsymbol{X}$, with $\boldsymbol{\mu} = E(\boldsymbol{X})$. Then for all $\boldsymbol{\alpha} \in \mathbb{R}^q$

$$Var(\boldsymbol{\alpha}'\boldsymbol{X}) = E[(\boldsymbol{\alpha}'\boldsymbol{X} - \boldsymbol{\alpha}'\boldsymbol{\mu})^2] = E[(\boldsymbol{\alpha}'(\boldsymbol{X} - \boldsymbol{\mu}))^2] = E[\boldsymbol{\alpha}'(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{\alpha}] = \boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha} \geq 0,$$

so that $\boldsymbol{\Sigma}$ is non-negative definite. Now suppose $\boldsymbol{\Sigma}$ is non-negative definite matrix of rank r, say ($r \leq q$). Write $\boldsymbol{\Sigma} = \boldsymbol{C}\boldsymbol{C}'$, where C is an $q \times r$ matrix of rank r. Let $\boldsymbol{Y}$ be an $r \times 1$ vector of independent random variables with mean $\mathbf{0}$ and $Cov(\boldsymbol{Y}) = \boldsymbol{I}_r$ and put $\boldsymbol{X} = \boldsymbol{C}\boldsymbol{Y}$. Then $E(\boldsymbol{X}) = \mathbf{0}$ and $Cov(\boldsymbol{X}) = E(\boldsymbol{X}\boldsymbol{X}') = E(\boldsymbol{C}\boldsymbol{Y}\boldsymbol{Y}'\boldsymbol{C}') = \boldsymbol{C}E(\boldsymbol{Y}\boldsymbol{Y}')\boldsymbol{C} = \boldsymbol{C}\boldsymbol{C}' = \boldsymbol{\Sigma}$, So that $\boldsymbol{\Sigma}$ is a covariance matrix. ■

Theorem 3-1: If $\boldsymbol{X}$ is an $\boldsymbol{q} \times \mathbf{1}$ random vector, then its distribution is uniquely determined by the distribution of linear function $\boldsymbol{\alpha}'\boldsymbol{X}$, for every $\boldsymbol{\alpha} \in \mathbb{R}^{\boldsymbol{q}}$.

Proof: The characteristic function of $\boldsymbol{X}$ is

$$\phi(t, \boldsymbol{\alpha}) = E\left[e^{it\boldsymbol{\alpha}'\boldsymbol{X}}\right],$$

So that

$$\phi(1, \boldsymbol{\alpha}) = E\left[e^{i\boldsymbol{\alpha}'\boldsymbol{X}}\right],$$

which, considered as a function of $\boldsymbol{\alpha}$, is the characteristic function of $\boldsymbol{X}$ (i.e., the joint characteristic function of the components of $\boldsymbol{X}$). The required result then follows by invoking the fact that a distribution in $\mathbb{R}^q$ is uniquely determined by its characteristic function.■
For a set of multivariate observations, perhaps sampled from some population, the matrix $\boldsymbol{\Sigma}$ is estimated by

$$\boldsymbol{S} = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})'$$

where $\boldsymbol{x}_i' = (x_{i1}, \ldots, x_{iq})$ is the vector of observations for the ith individual

$$\overline{\boldsymbol{x}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i$$

is the mean vector of the observations.
The diagonal of $\boldsymbol{S}$ contains the sample variances of each variable, which we shall denote as s_i^2. We have

$$\mathbf{S} = \begin{pmatrix} s_1^2 & s_{12} & \cdots & s_{1q} \\ s_{21} & s_2^2 & \cdots & s_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ s_{q1} & s_{q2} & \cdots & s_q^2 \end{pmatrix}$$

Consider a data set which consists of chest, waist, and hip measurements on a sample of men and women and the measurements for 20 individuals are shown in Table 1.2.

Table 1.2: measure data. Chest, waist, and hip measurements on 20 individuals (in inches).

chest	waist	hips	gender	chest	waist	hips	gender
34	30	32	male	36	24	35	female
37	32	37	male	36	25	37	female
38	30	36	male	34	24	37	female
36	33	39	male	33	22	34	female
38	29	33	male	36	26	38	female
43	32	38	male	37	26	37	female
40	33	42	male	34	25	38	female
38	30	40	male	36	26	37	female
40	30	37	male	38	28	40	female
41	32	39	male	35	23	35	female

R Code to make Table 1.2 as it is:

```
chest = c(34,37,38,36,38,43,40,38,40,41)
waist = c(30,32,30,33,29,32,33,30,30,32)
hips = c(32,37,36,39,33,38,42,40,37,39)
gender = rep("male",10)
f1= data.frame(chest,waist,hips,gender)
chest = c(36,36,34,33,36,37,34,36,38,35)
waist = c(24,25,24,22,26,26,25,26,28,23)
hips = c(35,37,37,34,38,37,38,37,40,35)
gender = rep("female",10)
f2= data.frame(chest,waist,hips,gender)
measure=cbind(f1,f2)
```

But we use the following format

```
chest = c(34,37,38,36,38,43,40,38,40,41,36,36,34,33,36,37,34,36,38,35)
waist = c(30,32,30,33,29,32,33,30,30,32,24,25,24,22,26,26,25,26,28,23)
hips = c(32,37,36,39,33,38,42,40,37,39,35,37,37,34,38,37,38,37,40,35)
gender = c("Male","Female")
sex = rep(gender, each=10)
measure=data.frame(chest,waist,hips,sex)
```

The covariance matrix for the data in Table 1.2 can be obtained using the var() function in R; however, we have to ``remove" the categorical variable gender from the measure data frame by subsetting on the numerical variables first:

```
cov(measure[, c("chest", "waist", "hips")])
chest waist hips
chest 6.632 6.368 3.000
waist 6.368 12.526 3.579
hips 3.000 3.579 5.945
```

If we require the separate covariance matrices of men and women, we can use

```
cov(measure[11:20, c("chest", "waist", "hips")])
chest waist hips
chest 2.278 2.167 1.556
waist 2.167 2.989 2.756
hips 1.556 2.756 3.067
```

```
cov(measure [1:10, c("chest", "waist", "hips")])
chest waist hips
chest 6.7222 0.9444 3.944
waist 0.9444 2.1000 3.078
hips 3.9444 3.0778 9.344
```

Correlation

The covariance is often difficult to interpret because it depends on the scales on which the two variables are measured; consequently, it is often standardized by dividing by the product of the standard deviations of the two variables to give a quantity called the correlation coefficient (Pearson linear correlation), ρ_{ij}, where

$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}, \qquad \sigma_i = \sqrt{\sigma_i^2}.$

The advantage of the correlation is that it is independent of the scales of the two variables. The correlation coefficient lies between -1 and $+1$ and gives a measure of the linear relationship of the variables X_i and X_j . It is positive if high values of X_i are associated with high values of X_j and negative if high values of X_i are associated with low values of X_j . If the relationship between two variables is non-linear, their correlation coefficient can be misleading.

With q variables there are $q(q-1)/2$ distinct correlations, which may be arranged in a $q \times q$ correlation matrix the diagonal elements of which are unity. For observed data, the correlation matrix contains the usual estimates of the ρs, and is generally denoted by $\boldsymbol{R}$. We have

$$\mathbf{R} = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1q} \\ r_{21} & 1 & \cdots & r_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ r_{q1} & r_{q2} & \cdots & 1 \end{pmatrix}$$

The $\boldsymbol{R}$ matrix may be written in terms of the sample covariance matrix $\boldsymbol{S}$

$$\boldsymbol{R} = \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{S} \boldsymbol{D}^{-\frac{1}{2}},$$

Where

$\boldsymbol{D}^{-\frac{1}{2}} = diag\left(\frac{1}{s_1}, \dots, \frac{1}{s_q}\right)$ and $s_i = \sqrt{s_i^2}$ is the sample standard deviation of variable i. (In most considered situations, we will be dealing with covariance and correlation matrices of full rank, q, so that both matrices will be non-singular, that is, invertible, to give matrices $\boldsymbol{S}^{-1}$ or $\boldsymbol{R}^{-1}$.)

The sample correlation matrix for the three variables in Table 1.2 is obtained by using the function cor() in R:

```
> cor(measure[, c("chest", "waist", "hips")])
chest waist hips
chest 1.0000 0.6987 0.4778
waist 0.6987 1.0000 0.4147
hips 0.4778 0.4147 1.0000
```

Distances

Given two data points $\boldsymbol{x} = (x_1, \dots, x_q)'$ and $= (y_1, \dots, y_q)'$, what serves as a measure of distance between them? The most common measure used is Euclidean distance, which is defined as

$$d(\boldsymbol{x}, \boldsymbol{y}) = [(\boldsymbol{x} - \boldsymbol{y})'(\boldsymbol{x} - \boldsymbol{y})]^{\frac{1}{2}} = \sqrt{(x_1 - y_1)^2 + \dots + \left(x_q - y_q\right)^2}.$$

Euclidean distance can be calculated using the dist() function in R.
Here we shall illustrate this on the body measurement data (Table 1.2) and divide each variable by its standard deviation using the function scale() before applying the dist() function

```
> dist(scale(measure[, c("chest", "waist", "hips")],center = FALSE))
```

There are some other distances such as City-block, Minkowski, Canberra, Bhattacharyya and etc. One other important distance measure is the Mahalanobis distance, which is defined as

$$d(\boldsymbol{x}, \boldsymbol{y}) = [\,(\boldsymbol{x} - \boldsymbol{y})'\boldsymbol{S}^{-1}(\boldsymbol{x} - \boldsymbol{y})]^{\frac{1}{2}},$$

where S is the covariance between x and y. Then the Mahalanobis angle θ between x and y, subtended at the origin, is defined by

$$\theta = arc\cos\frac{x'S^{-1}y}{d(x,0)d(y,0)}.$$

In the case of two populations, Given a sample of size n_i, with sample mean $\bar{x}_i$ and sample covariance matrix S_i, from the ith population ($i = 1,2$) we have the sample version

$$d(\bar{x}_1, \bar{x}_2) = \left[(\bar{x}_1 - \bar{x}_2)'S_p^{-1}(\bar{x}_1 - \bar{x}_2)\right]^{\frac{1}{2}},$$

where $S_p = \frac{1}{n_1+n_2-2}\,[(n_1 - 1)S_1 + (n_2 - 1)S_2]$ is a pooled sample covariance matrix.
In general, a distance measure $d(\boldsymbol{x}, \boldsymbol{y})$ must satisfy the following conditions

I. $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$,

II. $d(\boldsymbol{x},\boldsymbol{y}) > 0$ if $\boldsymbol{x} \neq \boldsymbol{y}$ and $d(\boldsymbol{x},\boldsymbol{y}) = 0$ if $\boldsymbol{x} = \boldsymbol{y}$,

III. $d(\boldsymbol{x},\boldsymbol{y}) \leq d(\boldsymbol{x},\boldsymbol{z}) + d(\boldsymbol{z},\boldsymbol{y})$ for three vectors $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ of the same size.

4. Multivariate Normal Distribution

Construction

Theorem 4-1: If $\boldsymbol{X}$ has an $\boldsymbol{q}$-variate normal distribution, then both $\boldsymbol{\mu} = \boldsymbol{E}(\boldsymbol{X})$ and $\boldsymbol{\Sigma} = \boldsymbol{Cov}(\boldsymbol{X})$ exist and the distribution of $\boldsymbol{X}$ is determined by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

The q-variate normal distribution of the random vector $\boldsymbol{X}$ of Theorem 2-2 will be denoted by $\boldsymbol{X} \sim N_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Theorem 4-2: If $\boldsymbol{X} \sim \boldsymbol{N_q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma}$ is positive definite, then the density of $\boldsymbol{X}$ is

$$f_X(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{q}{2}}(\det \boldsymbol{\Sigma})^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

(Here, and throughout the note, det denotes determinant).

Proof: Write $\boldsymbol{\Sigma} = \boldsymbol{CC}'$ where $\boldsymbol{C}$ is a nonsingular matrix $q \times q$ and put $\boldsymbol{X} = \boldsymbol{CU} + \boldsymbol{\mu}$ where $\boldsymbol{U}$ is an $q \times 1$ vector of independent $N(0,1)$ random variables, i.e., $\boldsymbol{U} \sim N_q(\boldsymbol{0}, \boldsymbol{I}_q)$. The joint density function of $U_1, \dots, U_q$ is

$$f(\boldsymbol{u}) = \prod_{i=1}^{q} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\, u_i^2\right) = \frac{1}{(2\pi)^{\frac{q}{2}}} \exp\left(-\frac{1}{2}\, \boldsymbol{u}'\boldsymbol{u}\right)$$

The inverse transform is $\boldsymbol{U} = \boldsymbol{B}(\boldsymbol{X} - \boldsymbol{\mu})$, with $\boldsymbol{B} = \boldsymbol{C}^{-1}$, and the Jacobian of this transformation is

$$\det \begin{bmatrix} \frac{\partial u_1}{\partial x_1} & \cdots & \frac{\partial u_1}{\partial x_q} \\ \vdots & \ddots & \vdots \\ \frac{\partial u_q}{\partial x_1} & \cdots & \frac{\partial u_q}{\partial x_q} \end{bmatrix} = \det \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{q1} & b_{q2} & \dots & b_{qq} \end{bmatrix}$$

$$= \det \boldsymbol{B} = \det \boldsymbol{C}^{-1} = (\det \boldsymbol{C})^{-1} = [\det(\boldsymbol{CC}')]^{-\frac{1}{2}} = (\det \boldsymbol{\Sigma})^{-\frac{1}{2}}$$

So that the density function of $\boldsymbol{X}$ is

$$f_X(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{q}{2}}(\det \boldsymbol{\Sigma})^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{C}^{-1'}\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right);$$

and since $\boldsymbol{\Sigma}^{-1} = \boldsymbol{C}^{-1'}\boldsymbol{C}^{-1}$, the proof is complete.

The density function of multivariate normal distribution is constant whenever the quadratic form in the exponent is, so that it is constant on the ellipsoid

$$(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = k$$

in $\mathbb{R}^q$, for every $k > 0$. This ellipsoid has center μ, and axes $\pm c\sqrt{\lambda_i}\boldsymbol{e}_i$ (where λ_i and e_i are the eigenvalues and eigenvectors of Σ), while Σ determines its shape and orientation. The above ellipsoid gives the contours of multivariate normal distribution for different values k.

Simulation

```
library(MASS)

mu=c(1,2)

Sigma=matrix(c(10,3,3,2),2,2)

mvrnorm(n=5,mu,Sigma)
```

```
          [,1]        [,2
[1,]  2.8810714  0.03039017
[2,]  3.0003637  1.55234615
[3,] -1.1981544 -0.47694161
[4,]  1.9683900  1.15906644
[5,] -0.6981529  0.29307889
```

Hint:

```
var(mvrnorm(n=5,mu,Sigma))
```

```
         [,1]     [,2]
[1,] 4.740504 1.806905
[2,] 1.806905 1.726933
```

```
var(mvnorm(n=5,mu,Sigma,empirical=TRUE))
```

(If true, mu and Sigma specify the empirical not population mean and covariance matrix)

```
     [,1] [,2]
[1,]   10    3
[2,]    3    2
```

It is worthwhile looking explicitly at the bivariate normal distribution ($q = 2$). In this case

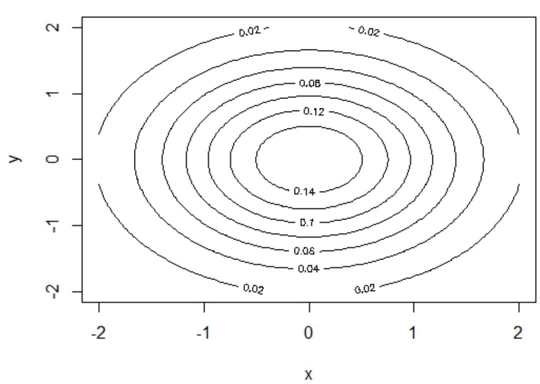
$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \qquad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \qquad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

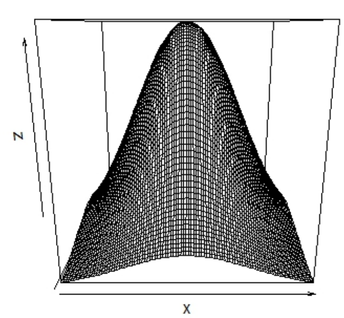
where $E(X_1) = \mu_1$, $E(X_2) = \mu_2$, $Var(X_1) = \sigma_1^2$, $Var(X_2) = \sigma_2^2$, and the correlation between X_1 and X_2 is ρ. For the distribution of X to be nonsingular normal we need $\sigma_1^2 > 0$, $\sigma_2^2 > 0$, and $\det\Sigma = \sigma_1^2\sigma_2^2(1-\rho^2) > 0$, so that $-1 < \rho < 1$. Then the joint density function of X_1 and X_2 is

$$f_X(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\,\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right\}.$$

Graph

```
require(MASS)
require(graphics)
require(mnormt)
x=seq(-2,2,len=100)
y=seq(-2,2,len=100)
mu=c(0,0)
Sigma=matrix(c(1,0,0,10),2,2)
g=function(x,y){ dmnorm(cbind(x,y),mu,Sigma)}
z=outer(x,y,g)
persp(x,y,z)
contour(x,y,z, drawlabels=FALSE, xlab="x", ylab="y")
```





Testing Multivariate Normality

For many multivariate methods, the assumption of multivariate normality is not critical to the results of the analysis, but there may be occasions when testing for multivariate normality may be of interest. A start can be made perhaps by assessing each variable separately for univariate normality using a probability plot. Such plots are commonly applied in univariate analysis and involve ordering the observations and then plotting them against the appropriate values of an assumed cumulative distribution function. There are two basic types of plots for comparing two probability distributions, the probability-probability (p-p) plot and the quantile-quantile (q-q) plot. The diagram in Figure 1.2 may be used for describing each type.

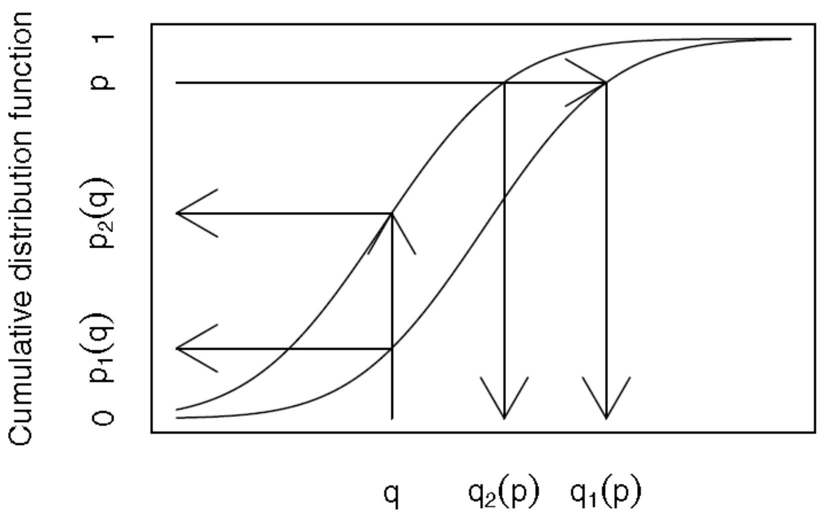


Fig. 1.2. Cumulative distribution functions and quantiles.

A plot of points whose coordinates are the cumulative probabilities $p_1(q) = P(X_1 \leq q)$ and $p_2(q) = P(X_2 \leq q)$ for diffeerent values of q for random variables X_1 and X_2 is a probability-probability plot, while a plot of the points whose coordinates are the quantiles $(q_1(p), q_2(p))$ for different values of p with

$q_1(p) = p_1^{-1}(p), \quad q_2(p) = p_2^{-1}(p),$

is a quantile-quantile plot. For example a q-q plot for investigating the assumption that a set of data is form a normal distribution would involve plotting the ordered sample values of variable 1 (i.e. $x_{(1)}, \dots, x_{(n)}$) against the quantiles of a standard normal distribution, $\Phi^{-1}(p(i))$, where usually $p_i = \frac{i-\frac{1}{2}}{n}$, and $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$.

If the two distributions being compared are similar, the points in the Q–Q plot will approximately lie on the line *y* = *x*. If the distributions are linearly related, the points in the Q–Q plot will approximately lie on a line, but not necessarily on the line *y* = *x*.

For multivariate data, normal probability plots may be used to examine each variable separately, although marginal normality does not necessarily imply that the variables follow a multivariate normal distribution.

Alternatively (or additionally), each multivariate observation might be converted to a single number in some way before plotting. For example, in the specific case of assessing a data set for multivariate normality, each q-dimensional observation, $\boldsymbol{x}_i$, could be converted into a generalized distance, d_i^2, giving a measure of the distance of the particular observation from the mean vector of the complete sample, $\overline{\boldsymbol{x}}$; d_i^2 is calculated as

$$d_i^2 = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})' \boldsymbol{S}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}),$$

where $\boldsymbol{S}$ is the sample covariance matrix. This distance measure takes into account the different variances of the variables and the covariances of pairs of variables. If the observations do arise from a multivariate normal distribution, then these distances have approximately a chi-squared distribution with q degrees of freedom, also denoted by the symbol χ_q^2 . So plotting the ordered distances against the corresponding quantiles of the appropriate chi-square distribution should lead to a straight line through the origin.

We will now assess the body measurements data in Table 1.2 for normality, although because there are only 20 observations in the sample there is really too little information to come to any convincing conclusion. Figure 1.3 shows separate probability plots for each measurement; there appears to be no evidence of any departures from linearity. The chi-square plot of the 20 generalized distances in Figure 1.4 does seem to deviate a little from linearity, but with so few observations it is hard to be certain. The plot is set up as follows. We first extract the relevant data

```
x <- measure [, c("chest", "waist", "hips")]
```

and estimate the means of all three variables (i.e., for each column of the data) and the covariance matrix

```
cm <- colMeans(x)
S <- cov(x)

qqnorm(measure[,"chest"], main = "chest"); qqline(measure[,"chest"])
qqnorm(measure[,"waist"], main = "waist"); qqline(measure[,"waist"])
qqnorm(measure[,"hips"], main = "hips"); qqline(measure[,"hips"])
```

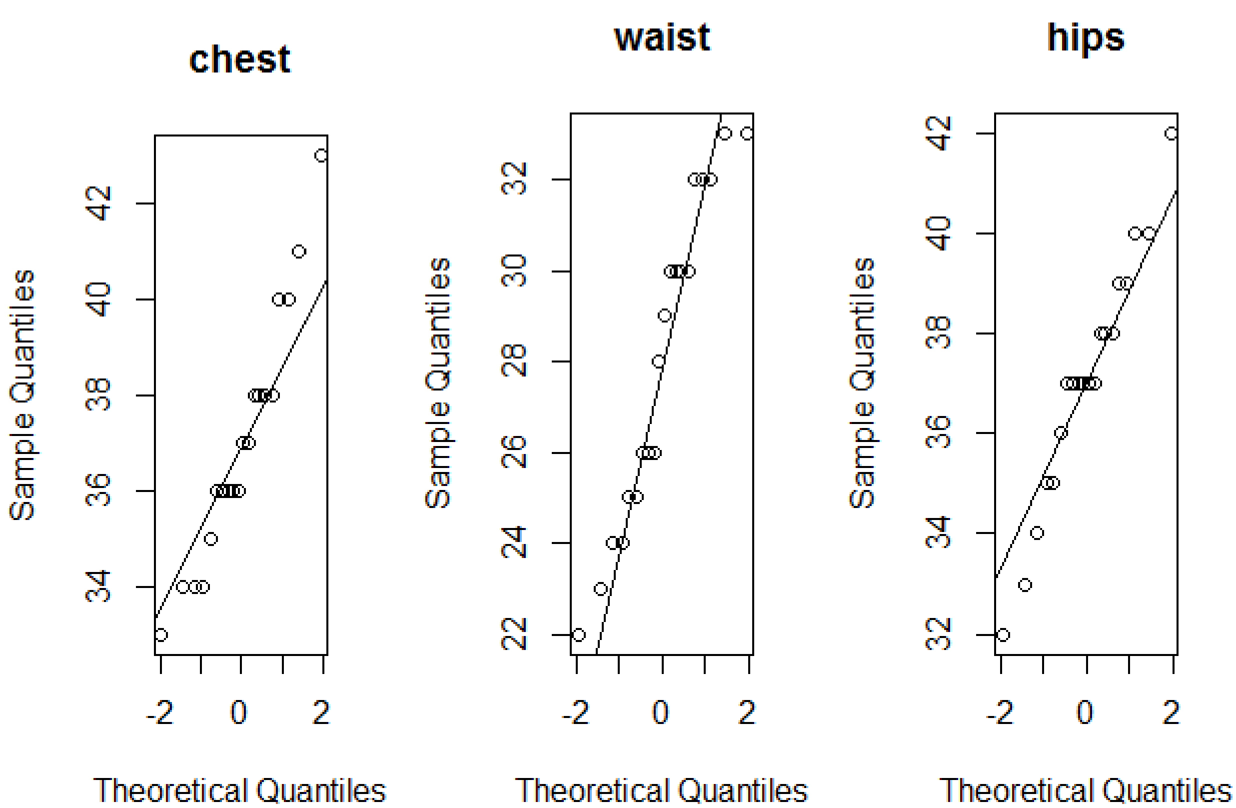


Fig. 1.3. Q-Q plots for Normality of chest, waist, and hip measurements.

The differences d_i have to be computed for all units in our data, so we iterate over the rows of $\boldsymbol{x}$ using the apply() function with argument MARGIN = 1 and, for each row, compute the distance d_i:

```
d <- apply(x, MARGIN = 1, function(x)  t(x - cm) %*% solve(S) %*% (x - cm))
```

The sorted distances can now be plotted against the appropriate quantiles of the χ^2_3 distribution obtained from qchisq(); see Figure 1.4.

```
plot(qchisq((1:nrow(x) - 1/2) / nrow(x), df = 3), sort(d),
        xlab = expression(paste(chi[3]^2, " Quantile")),
        ylab = "Ordered distances")
abline(a = 0, b = 1)
```

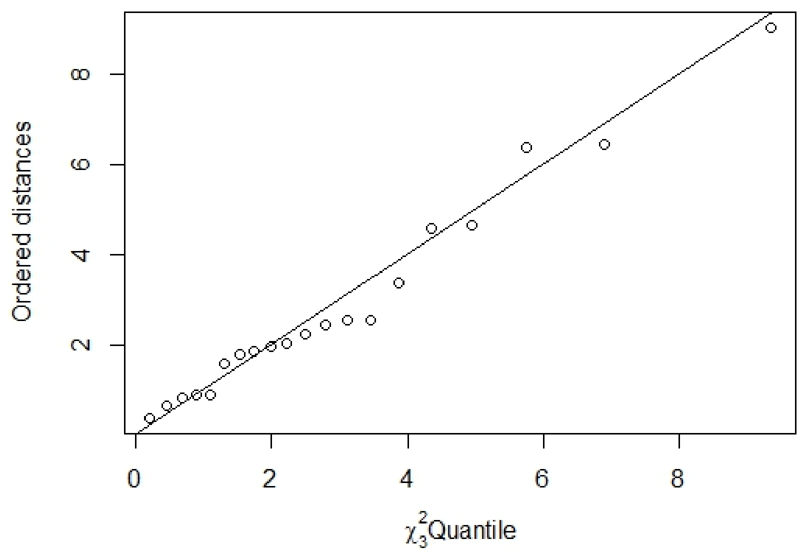


Fig. 1.4. Chi-square plot of generalised distances for body measurements data.

USairpollution data

Consider a data set for studying the air pollution in cities in the USA in 1981. The following variables were obtained for 41 US cities:

SO2: SO2 content of air in micrograms per cubic meter;
temp: average annual temperature in degrees Fahrenheit;
manu: number of manufacturing enterprises employing 20 or more workers;
popul: population size (1970 census) in thousands;
wind: average annual wind speed in miles per hour;

precip: average annual precipitation in inches;
predays: average number of days with precipitation per year.

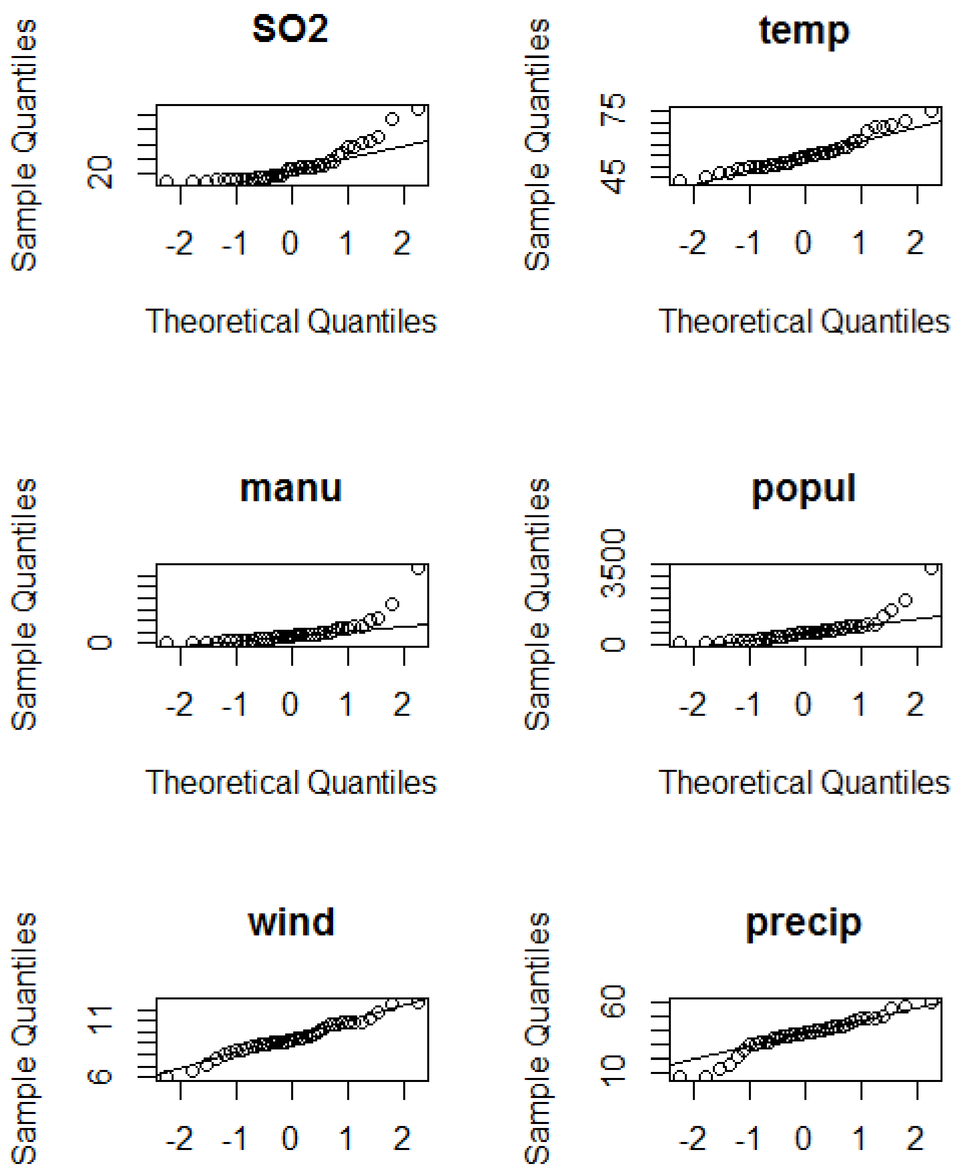
The data are shown in Table 1.5. (It is not here). R code for generating Table 1.5

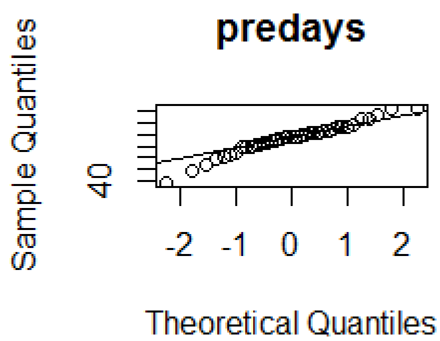
```
USairpollution=matrix(c(46,11,24,47,11,31,110,23,65,26,9,17,17,35,56,10,28,14,14,13,30,10,10,16,29,18,9,31,14,69,10,61,94,26,28,12,29,56,29,8,36,47.6,56.8,61.5,55.0,47.1,55.2,50.6,54.0,49.7,51.5,66.2,51.9,49.0,49.9,49.1,68.9,52.3,68.4,54.5,61.0,55.6,61.6,75.5,45.7,43.5,59.4,68.3,59.3,51.5,54.6,70.3,50.4,50.0,57.8,51.0,56.7,51.1,55.9,57.3,56.6,54.0,44,46,368,652,391,35,3344,462,1007,266,641,454,104,1064,412,721,361,136,381,91,291,337,207,569,699,275,204,96,181,1692,213,347,343,197,137,453,379,775,434,125,80,116,244,497,905,463,71,3369,453,751,540,844,515,201,1513,158,1233,746,529,507,132,593,624,335,717,744,448,361,308,347,1950,582,520,179,299,176,716,531,622,757,277,80,8.8,8.9,9.1,9.6,12.4,6.5,10.4,7.1,10.9,8.6,10.9,9.0,11.2,10.1,9.0,10.8,9.7,8.8,10.0,8.2,8.3,9.2,9.0,11.8,10.6,7.9,8.4,10.6,10.9,9.6,6.0,9.4,10.6,7.6,8.7,8.7,9.4,9.5,9.3,12.7,9.0,33.36,7.77,48.34,41.31,36.11,40.75,34.44,39.04,34.99,37.01,35.94,12.95,30.85,30.96,43.37,48.19,38.74,54.47,37.00,48.52,43.11,49.10,59.80,29.07,25.94,46.00,56.77,44.68,30.18,39.93,7.05,36.22,42.75,42.59,15.17,20.66,38.79,35.89,38.89,30.58,40.25,135,58,115,111,166,148,122,132,155,134,78,86,103,129,127,103,121,116,99,100,123,105,128,123,137,119,113,116,98,115,36,147,125,115,89,67,164,105,111,82,114),41)

cities = c("Albany", "Albuquerque","Atlanta","Baltimore","Buffalo","Charleston","Chicago", "Cincinnati","Cleveland","Columbus","Dallas","Denver","DesMoines","Detroit","Hartford","Houston","Indianapolis","Jacksonville","Kansas City","Little Rock","Louisville","Memphis","Miami", "Milwaukee","Minneapolis","Nashville","New Orleans","Norfolk","Omaha","Philadelphia", "Phoenix","Pittsburgh","Providence","Richmond","Salt Lake City","San Francisco","Seattle","St. Louis","Washington","Wichita","Wilmington")
variables = c("SO2","temp","manu","popul","wind", "precip","predays")
colnames(USairpollution) = variables
rownames(USairpollution)= cities
```

We will now look at using the chi-square plot on a set of data, namely the air pollution in US cities (see Table 1.5). The probability plots for each separate variable are shown in Figure 1.5. Here, we also iterate over all variables, this time using a special function, sapply(), that loops over the variable names:

```
layout(matrix(1:8, nc = 2))
qqnorm(USairpollution[,"SO2"], main = "SO2"); qqline(USairpollution[,"SO2"])
qqnorm(USairpollution[,"temp"], main = "temp"); qqline(USairpollution[,"temp"])
qqnorm(USairpollution[,"manu"], main = "manu"); qqline(USairpollution[,"manu"])
qqnorm(USairpollution[,"popul"],main ="popul");qqline(USairpollution[,"popul"])
qqnorm(USairpollution[,"wind"],main="wind"); qqline(USairpollution[,"wind"])
qqnorm(USairpollution[,"precip"],main="precip");qqline(USairpollution[,"precip"])
qqnorm(USairpollution[,"predays"],main = "predays ");qqline(USairpollution[,"pre days"])
```


SO2
Sample Quantiles
20
-2 -1 0 1 2
Theoretical Quantiles
temp
Sample Quantiles
45 75
-2 -1 0 1 2
Theoretical Quantiles
manu
Sample Quantiles
0
-2 -1 0 1 2
Theoretical Quantiles
popul
Sample Quantiles
0 3500
-2 -1 0 1 2
Theoretical Quantiles
wind
Sample Quantiles
6 11
-2 -1 0 1 2
Theoretical Quantiles
precip
Sample Quantiles
10 60
-2 -1 0 1 2
Theoretical Quantiles



The resulting seven plots are arranged on one page by a call to the layout matrix; see Figure 1.5. The plots for SO2 concentration and precipitation both deviate considerably from linearity, and the plots for manufacturing and population show evidence of a number of outliers. But of more importance is the chi-square plot for the data, which is given in Figure 1.6; the R code is identical to the code used to produce the chi-square plot for the body measurement data. In addition, the two most extreme points in the plot have been labeled with the city names to which they correspond using text().

```
x <- USairpollution
cm <- colMeans(x)
S <- cov(x)
d <- apply(x, 1, function(x) t(x - cm) %*% solve(S) %*% (x - cm))
plot(qc <- qchisq((1:nrow(x) - 1/2) / nrow(x), df = 6), sd <- sort(d),
        xlab = expression(paste(chi[6]^2, " Quantile")),
        ylab = "Ordered distances", xlim = range(qc) * c(1, 1.1))
oups <- which(rank(abs(qc - sd), ties = "random") > nrow(x) - 3)
text(qc[oups], sd[oups] - 1.5, names(oups))
abline(a = 0, b = 1)
```

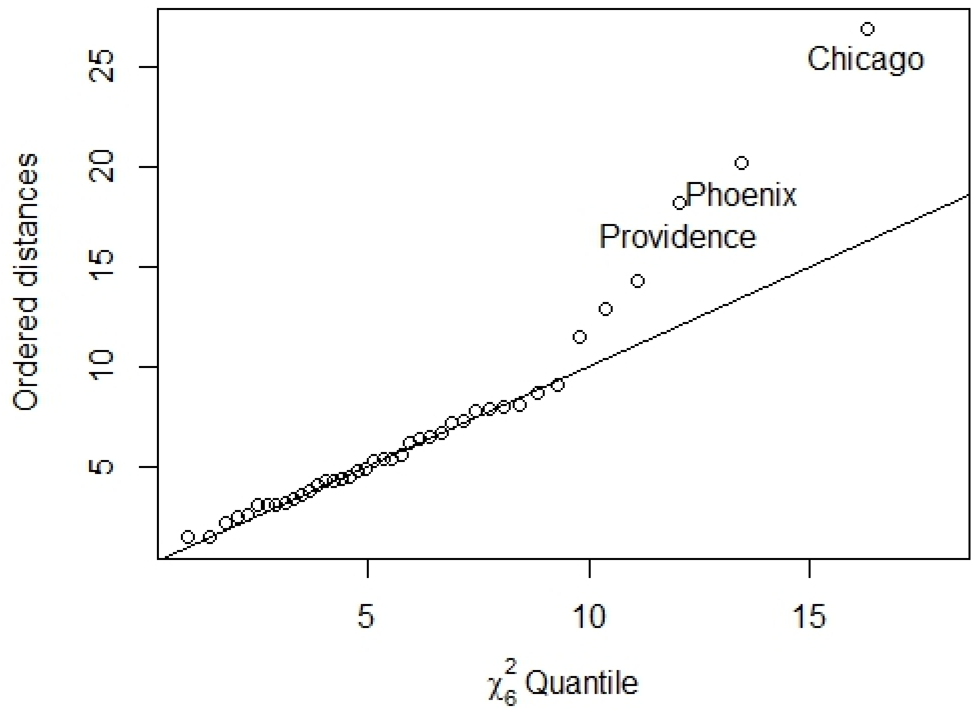


Fig. 1.6. χ^2 plot of generalised distances for USairpollution data.

This example illustrates that the chi-square plot might also be useful for detecting possible outliers (to avoid misleading effects) in multivariate data, where informally outliers are “abnormal” in the sense of deviating from the natural data variability.

Some Theorems

Theorem 4-3: If $\boldsymbol{X} \sim \boldsymbol{N_q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the characteristic function of $\boldsymbol{X}$ is

$$\boldsymbol{\phi_X(t) = exp\left(i\mu' t - \frac{1}{2}\, t'\Sigma t\right)}.$$

Recall that for a known matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $Cov(\boldsymbol{BX}) = \boldsymbol{B}Cov(\boldsymbol{X})\boldsymbol{B}'$ and

$$Cov(\boldsymbol{AX}, \boldsymbol{BX}) = \boldsymbol{A}Cov(\boldsymbol{X})\boldsymbol{B}'.$$

Theorem 4-4: If $\boldsymbol{X} \sim \boldsymbol{N_q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{B}$ is a $\boldsymbol{k} \times \boldsymbol{q}$, $\boldsymbol{b}$ is a $\boldsymbol{k} \times \boldsymbol{1}$, then

$$\boldsymbol{Y = BX + b \sim N_k(B\mu + b, B\Sigma B')}.$$

Theorem 4-5: If $\boldsymbol{X} \sim \boldsymbol{N_q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the marginal distribution of any subset of $\boldsymbol{k}(< \boldsymbol{q})$ components of $\boldsymbol{X}$ is k-variate normal.

For example, partition $\boldsymbol{X}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ as $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$, $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$, where $\boldsymbol{X}_1$ and $\boldsymbol{\mu}_1$ are $k \times 1$ and $\boldsymbol{\Sigma}_{11}$ is $k \times k$. Putting $\boldsymbol{B} = [\boldsymbol{I}_k \quad \boldsymbol{0}]$ $(k \times q)$, $\boldsymbol{b} = \boldsymbol{0}$ in Theorem 4-4, immediately shows that $\boldsymbol{X}_1 \sim N_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$.

Counterexample:

Suppose U_1, U_2, U_3 are independent $N(0,1)$ random variables and Z is an arbitrary random variable, independent of U_1, U_2, and U_3. Define X_1 and X_2 by

$$X_1 = \frac{U_1 + ZU_3}{\sqrt{1+Z^2}}, \qquad X_2 = \frac{U_2 + ZU_3}{\sqrt{1+Z^2}}.$$

Conditional on Z, $X_1 \sim N(0,1)$, and since this distribution does not depend on Z, it is the unconditional distribution of X_1. Similarly $X_2 \sim N(0,1)$. Again conditional on Z, the joint distribution of X_1 and X_2 is bivariate normal but the unconditional distribution clearly need not be.

Theorem 4-6: If $\boldsymbol{X} \sim N_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$and $\boldsymbol{X}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ are partitioned as $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$, $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$, where $\boldsymbol{X}_1$ and $\boldsymbol{\mu}_1$ are $k \times 1$ and $\boldsymbol{\Sigma}_{11}$ is $k \times k$,

- then the sub vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent iff $\boldsymbol{\Sigma}_{12} = 0$.
- The conditional distribution of $\boldsymbol{X}_2$ given $\boldsymbol{X}_1$ is $N_{q-k}(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22.1})$,

where $\boldsymbol{\Sigma}_{22.1}$ is the Schur complement of $\boldsymbol{\Sigma}_{11}$ given by $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$.

Central $chi - squared$ Distribution

If $\boldsymbol{X} \sim \boldsymbol{N_q}(\boldsymbol{\mu}, \boldsymbol{I_q})$ then the random variable $\boldsymbol{Z} = (\boldsymbol{X} - \boldsymbol{\mu})'(\boldsymbol{X} - \boldsymbol{\mu})$ has the density function

$$\boldsymbol{f}(\boldsymbol{z}) = \frac{\boldsymbol{1}}{\boldsymbol{2}^{\frac{q}{2}}\boldsymbol{\Gamma}\left(\frac{\boldsymbol{1}}{\boldsymbol{2}}\boldsymbol{q}\right)} \boldsymbol{e}^{-\frac{z}{2}} \boldsymbol{z}^{\frac{q}{2}-\boldsymbol{1}}, \qquad \boldsymbol{z} > \boldsymbol{0},$$

$\boldsymbol{Z}$ is said to have the central $\boldsymbol{\chi^2}$ distribution with $\boldsymbol{q}$ df, to be written as $\boldsymbol{\chi_q^2}(.)$.

Then $\boldsymbol{E}(\boldsymbol{Z}) = \boldsymbol{q}$ and $\boldsymbol{Var}(\boldsymbol{Z}) = \boldsymbol{2q}$.

Central t −Distribution

If $\boldsymbol{X} \sim \boldsymbol{N_1}(\boldsymbol{\mu}, \boldsymbol{\sigma^2})$ and $\boldsymbol{Z} \sim \boldsymbol{\chi^2_n}$ and $\boldsymbol{X}$ be independent of Z, then the random variable

$$\boldsymbol{t} = \frac{\frac{(\boldsymbol{X}-\boldsymbol{\mu})}{\boldsymbol{\sigma}}}{\sqrt{\frac{\boldsymbol{Z}}{\boldsymbol{n}}}}$$

has the density function

$$\boldsymbol{f(t)} = \frac{\Gamma\left[\frac{1}{2}(\boldsymbol{n}+\mathbf{1})\right]}{\sqrt{\mathbf{n\pi}}\,\Gamma\left(\frac{1}{2}\boldsymbol{n}\right)}\left(\mathbf{1}+\frac{\boldsymbol{t^2}}{\boldsymbol{n}}\right)^{-\frac{\boldsymbol{n}+\mathbf{1}}{\mathbf{2}}}, \qquad \boldsymbol{t} \in \mathbf{R},$$

$\boldsymbol{t}$ is said to have the central $\boldsymbol{t}$ −distribution with $\boldsymbol{n}$ df.

Then $E(t) = 0$ and $Var(t) = \frac{n}{n-2}$.

Central F Distribution

If $\boldsymbol{Z_1} \sim \boldsymbol{\chi^2_{n_1}}$ be independent of $\boldsymbol{Z_2} \sim \boldsymbol{\chi^2_{n_2}}$, then

$$\boldsymbol{F} = \frac{\frac{\boldsymbol{Z_1}}{\boldsymbol{n_1}}}{\frac{\boldsymbol{Z_2}}{\boldsymbol{n_2}}}$$

Has the density

$$\boldsymbol{g(f)} = \frac{\Gamma\left[\frac{1}{2}(\boldsymbol{n_1}+\boldsymbol{n_2})\right]}{\Gamma\left(\frac{1}{2}\boldsymbol{n_1}\right)\Gamma\left(\frac{1}{2}\boldsymbol{n_2}\right)}\left(\frac{\boldsymbol{n_1}}{\boldsymbol{n_2}}\right)^{\frac{\boldsymbol{n_1}}{2}}\boldsymbol{f}^{\frac{\boldsymbol{n_1}}{2}-\mathbf{1}}\left(\mathbf{1}+\frac{\boldsymbol{n_1}}{\boldsymbol{n_2}}\boldsymbol{f}\right)^{-\frac{\boldsymbol{n_1}+\boldsymbol{n_2}}{2}}, \qquad \boldsymbol{f} > \mathbf{0},$$

$\boldsymbol{F}$ is said to have the central $\boldsymbol{F}$ distribution with $\boldsymbol{n_1}$ and $\boldsymbol{n_2}$ df's, to be denoted by $\boldsymbol{F_{n_1,n_2}}$.

Then we have $E(F) = \frac{n_2(n_1)}{n_1(n_2-2)}$, $n_2 > 2$ and $Var(F) = 2\left(\frac{n_2}{n_1}\right)^2\left[\frac{(n_1)^2+n_1(n_2-2)}{(n_2-2)^2(n_2-4)}\right]$, $n_2 > 4$.

5. Wishart Distribution

Definitions

Definition 1: Let $\boldsymbol{W} = (w_{ij})$ be a $q \times q$ symmetric matrix of random variables that is positive definite with probability 1, and let $\boldsymbol{\Sigma}$ be a $q \times q$ positive definite matrix. If m is an integer such that $m \geq q$, then $\boldsymbol{W}$ is said to have a (nonsingular) Wishart distribution with m df if the joint density function of the $\frac{1}{2}q(q+1)$ distinct elements of $\boldsymbol{W}$ (in, say, the upper triangle) is

$$f\left(w_{11}, w_{12}, \dots, w_{qq}\right) = c^{-1} \det \boldsymbol{W}^{\frac{m-q}{2} - \frac{1}{2}} etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbf{W}\right),$$

where etr represents the operator e^{trace},

$c = 2^{\frac{mq}{2}} \det \boldsymbol{\Sigma}^{\frac{m}{2}} \Gamma_q\left(\frac{m}{2}\right)$,

where $\Gamma_q(a)$ is the multivariate gamma function given by

$$\Gamma_q(a) = \pi^{\frac{q(q-1)}{4}} \prod_{j=1}^{q} \Gamma\left[a + \frac{1}{2}(1-j)\right].$$

We shall write $\boldsymbol{W} \sim W_q(m, \boldsymbol{\Sigma})$.

Definition 2: Suppose that $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_m$ are independently and identically distributed (iid) as $N_q(0, \boldsymbol{\Sigma})$; then $\boldsymbol{W} = \sum_{i=1}^{m} \boldsymbol{x}_i \boldsymbol{x}_i{}'$ is said to have a Wishart distribution with m df.

Theorem 5-1: If $\boldsymbol{W} \sim \boldsymbol{W}_q(\boldsymbol{m}, \boldsymbol{\Sigma})$ and $\boldsymbol{C}$ is a $\boldsymbol{d} \times \boldsymbol{q}$ matrix of rank $\boldsymbol{d}$, then $\boldsymbol{CWC}' \sim \boldsymbol{W}_d(\boldsymbol{m}, \boldsymbol{C\Sigma C}')$.

Theorem 5-2: If $\boldsymbol{W} \sim \boldsymbol{W}_q(\boldsymbol{m}, \boldsymbol{\Sigma})$ and $\boldsymbol{l}$ is any nonzero $\boldsymbol{q} \times \boldsymbol{1}$ vector of constants, then $\boldsymbol{l}'\boldsymbol{W}\boldsymbol{l} \sim \boldsymbol{\sigma}_l^2 \boldsymbol{\chi}_m^2$, where $\boldsymbol{\sigma}_l^2 = \boldsymbol{l}'\boldsymbol{\Sigma}\boldsymbol{l} > \boldsymbol{0}$ (since $\boldsymbol{\Sigma} > \boldsymbol{0}$).

Theorem 5-3: Let $\boldsymbol{X}' = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_m)$, where $\boldsymbol{x}_i$ are iid $\boldsymbol{N}_q(\boldsymbol{0}, \boldsymbol{\Sigma})$, and let $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{l}$, where $\boldsymbol{l}\ (\neq \boldsymbol{0})$ is a $\boldsymbol{q} \times \boldsymbol{1}$ vector of constants. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $\boldsymbol{m} \times \boldsymbol{m}$ symmetric matrices of rank $\boldsymbol{r}$ and $\boldsymbol{s}$, respectively, and let $\boldsymbol{b}$ be an $\boldsymbol{m} \times \boldsymbol{1}$ vector of constants; then

- $\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} \sim \boldsymbol{W}_q(\boldsymbol{r}, \boldsymbol{\Sigma})$ iff $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} \sim \boldsymbol{\sigma}_l^2 \boldsymbol{\chi}_r^2$ for any $\boldsymbol{l}$, where $\boldsymbol{\sigma}_l^2 = \boldsymbol{l}'\boldsymbol{\Sigma}\mathbf{l}$.
- $\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$ and $\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}$ have independent Wishart distributions with $\boldsymbol{r}$ and $\boldsymbol{s}$ df, respectively, iff $\frac{\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}}{\boldsymbol{\sigma}_l^2}$ and $\frac{\boldsymbol{y}'\boldsymbol{B}\boldsymbol{y}}{\boldsymbol{\sigma}_l^2}$ are independently distributed as chi-square with $\boldsymbol{r}$ and $\boldsymbol{s}$ df, respectively, for any $\boldsymbol{l}$.
- $\boldsymbol{X}'\boldsymbol{b}$ and $\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$ are independently distributed as $\boldsymbol{N}_q$ and $\boldsymbol{W}_q(\boldsymbol{r}, \boldsymbol{\Sigma})$, respectively, iff $\boldsymbol{y}'\boldsymbol{b}$ and $\frac{\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}}{\boldsymbol{\sigma}_l^2}$ are independently distributed as normal and $\boldsymbol{\chi}_r^2$, respectively, for any $\boldsymbol{l}$.
- $\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} \sim \boldsymbol{W}_q(\boldsymbol{r}, \boldsymbol{\Sigma})$ iff $\boldsymbol{A}^2 = \boldsymbol{A}$.
- The Wishart variables $\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$ and $\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}$ are independent iff $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{0}$.

- $\boldsymbol{X'AX}$ and $\boldsymbol{X'b}$ are independently distributed as $\boldsymbol{W_q}$ and $\boldsymbol{N_q}$, respectively, iff $\boldsymbol{Ab}=\boldsymbol{0}$ and $\boldsymbol{A^2}=\boldsymbol{A}$.

Theorem 5-4: If $\boldsymbol{W \sim W_q(m,\Sigma)}$, then the characteristic function of $\boldsymbol{W}$ [joint characteristic function of the $\frac{1}{2}\boldsymbol{q(q+1)}$ variables $\boldsymbol{w_{ij}}$ $\boldsymbol{1 \le i \le j \le q}$] is

$$\boldsymbol{\phi(\Theta) = E\left[exp\left(i\sum_{j\le k}^{q}\theta_{jk}w_{jk}\right)\right] = det\left(I_q - i\Gamma\Sigma\right)^{-\frac{m}{2}}},$$

where $\boldsymbol{\Gamma=(\gamma_{ij})}$ for $\boldsymbol{i,j=1,\dots,q}$ with $\boldsymbol{\gamma_{ij}=\left(1+\delta_{ij}\right)\theta_{ij}}$, $\boldsymbol{\theta_{ji}=\theta_{ij}}$, and $\boldsymbol{\delta_{ij}}$ is the Kronecker delta given by

$$\boldsymbol{\delta_{ij}=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases}}.$$

Then $E(\boldsymbol{W}) = m\boldsymbol{\Sigma}$ and $Cov\left(w_{ij},w_{kl}\right) = m(\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk})$.

Noncentral $chi-squared$ Distribution

Theorem 5-5: If $\boldsymbol{X \sim N_q(\mu, I_q)}$ then the random variable $\boldsymbol{Z = X'X}$ has the density function

$$\boldsymbol{f(z) = e^{-\frac{\delta}{2}}\,{}_0F_1\left(\frac{1}{2}q;\frac{1}{4}\delta z\right)\frac{1}{2^{\frac{q}{2}}\Gamma\left(\frac{1}{2}n\right)}e^{-\frac{z}{2}}z^{\frac{q}{2}-1} = \sum_{k=0}^{\infty}\frac{1}{k!\left(\frac{1}{2}q\right)_k}\frac{\left(\frac{1}{4}\delta z\right)^k e^{-\frac{\delta}{2}}}{2^{\frac{q}{2}}\Gamma\left(\frac{1}{2}n\right)}e^{-\frac{z}{2}}z^{\frac{q}{2}-1}},$$

$$\boldsymbol{z>0},$$

where $\boldsymbol{\delta=\mu'\mu}$, $\boldsymbol{(a)_k = a(a+1)\dots(a+k-1)}$. $\boldsymbol{Z}$ is said to have the noncentral $\boldsymbol{\chi^2}$ distribution with $\boldsymbol{q}$ df and noncentrality parameter $\boldsymbol{\delta}$, to be written as $\boldsymbol{\chi_q^2(\delta)}$.

Then $\boldsymbol{E(Z) = q+\delta}$ and $\boldsymbol{Var(Z) = 2q+4\delta}$.

Noncetral $t-$ Distribution

If $\boldsymbol{X \sim N_1(\mu,1)}$ and $\boldsymbol{Z \sim \chi_n^2}$ and $\boldsymbol{X}$ be independent of Z, then the random variable

$$\boldsymbol{t=\frac{X}{\sqrt{\frac{Z}{n}}}}$$

has the density function

$\boldsymbol{t}$ is said to have the noncentral $\boldsymbol{t}$ $-$distribution with $\boldsymbol{n}$ df and noncentrality parameter $\boldsymbol{\mu}$.

Then $$E(t) = \boldsymbol{\mu}\sqrt{n/2}\frac{\Gamma\left[\frac{1}{2}(\boldsymbol{n-1})\right]}{\Gamma\left(\frac{1}{2}\boldsymbol{n}\right)}, n > 1$$

and

$$Var(t) = \frac{n(1+\mu^2)}{n-2} - \frac{\mu^2 n}{2}\left(\frac{\Gamma\left[\frac{1}{2}(\boldsymbol{n-1})\right]}{\Gamma\left(\frac{1}{2}\boldsymbol{n}\right)}\right)^2, n > 2.$$

Noncentral F Distribution

Theorem 5-6: If $\boldsymbol{Z_1} \sim \boldsymbol{\chi^2_{n_1}(\delta)}$ be independent of $\boldsymbol{Z_2} \sim \boldsymbol{\chi^2_{n_2}}$, then

$$\boldsymbol{F} = \frac{\frac{\boldsymbol{Z_1}}{\boldsymbol{n_1}}}{\frac{\boldsymbol{Z_2}}{\boldsymbol{n_2}}}$$

Has the density

$$\boldsymbol{g(f)} = \sum_{\boldsymbol{k=0}}^{\boldsymbol{\infty}} \frac{\left(\frac{1}{2}(\boldsymbol{n_1+n_2})\right)_{\boldsymbol{k}}}{\boldsymbol{k!}\left(\frac{1}{2}\boldsymbol{n_1}\right)_{\boldsymbol{k}}}\left(\frac{\boldsymbol{n_1}}{\boldsymbol{n_2}}\right)^{\frac{\boldsymbol{n_1}}{2}+\boldsymbol{k}} \frac{\boldsymbol{\Gamma}\left[\frac{1}{2}(\boldsymbol{n_1+n_2})\right]}{\boldsymbol{\Gamma}\left(\frac{1}{2}\boldsymbol{n_1}\right)\boldsymbol{\Gamma}\left(\frac{1}{2}\boldsymbol{n_2}\right)} \frac{\boldsymbol{\delta^k e^{-\frac{\delta}{2}}}}{\boldsymbol{2^k}} \frac{\boldsymbol{f^{\frac{n_1}{2}+k-1}}}{\left(\boldsymbol{1+\frac{n_1}{n_2}f}\right)^{\frac{\boldsymbol{n_1+n_2}}{2}+\boldsymbol{k}}}, \qquad \boldsymbol{f > 0},$$

$\boldsymbol{F}$ is said to have the noncentral $\boldsymbol{F}$ distribution with $\boldsymbol{n_1}$ and $\boldsymbol{n_2}$ df and noncentrality parameter $\boldsymbol{\delta}$, to be denoted by $\boldsymbol{F_{n_1,n_2}(\delta)}$.

Then we have $E(F) = \frac{n_2(n_1+\delta)}{n_1(n_2-2)}$, $n_2 > 2$ and $Var(F) = 2\left(\frac{n_2}{n_1}\right)^2\left[\frac{(n_1+\delta)^2+(n_1+2\delta)(n_2-2)}{(n_2-2)^2(n_2-4)}\right]$, $n_2 > 4$.

Noncentral Wishart Distribution

Let $\boldsymbol{X}' = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m)$, where $\boldsymbol{x}_i$ are independently distributed as $N_q(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, $(i = 1, \ldots, m)$. Then $\boldsymbol{W} = \boldsymbol{X}'\boldsymbol{X}$ has noncentral distribution denoted by $\boldsymbol{W} \sim W_q(m, \boldsymbol{\Sigma}, \boldsymbol{\Phi})$, with noncentrality matrix parameter $\boldsymbol{\Phi}$ given by

$$\boldsymbol{\Phi} = \sum_{i=1}^{m}\left(\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}_i\right)\left(\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}_i\right)' = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{M}'\boldsymbol{M}\boldsymbol{\Sigma}^{-\frac{1}{2}},$$

where $\boldsymbol{M} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_m)'$.

Note that

$E(\boldsymbol{X}')E(\boldsymbol{X}) = \boldsymbol{M}'\boldsymbol{M} = \sum_{i=1}^{m}\boldsymbol{\mu}_i\boldsymbol{\mu}_i{}'$, $E(\boldsymbol{W}) = m\boldsymbol{\Sigma} + \boldsymbol{M}'\boldsymbol{M}$.

6. Hotelling's T^2 Distribution

Recall that if $\sim N(\mu, \sigma^2) \perp W \sim \sigma^2 \chi_m^2$; then $T = \frac{(X-\mu)/\sigma}{(W/m\sigma^2)^{\frac{1}{2}}} \sim t_m$, where t_m is the t-distribution with m df. Therefore, $T^2 = m(X-\mu)W^{-1}(X-\mu) \sim F_{1,m}$, since we have the identity $t_m^2 \equiv F_{1,m}$.

For the multivariate case $\boldsymbol{X} \sim N_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{W} \sim W_q(m, \boldsymbol{\Sigma})$, $\boldsymbol{X}$ is statistically independent of $\boldsymbol{W}$, the Hotelling's T^2 statistics is given by

$$T^2 = m(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{W}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}).$$

Theorem 6-1: Let $\boldsymbol{T^2} = \boldsymbol{my'W^{-1}y}$, where $\boldsymbol{y} \sim \boldsymbol{N_q(0, \Sigma)}$ is independent of $\boldsymbol{W} \sim \boldsymbol{W_q(m, \Sigma)}$. Then $\boldsymbol{\frac{m-q+1}{q}\frac{T^2}{m}} \sim \boldsymbol{F_{q,m-q+1}}$.

Theorem 6-2: If $\boldsymbol{X_1}, \dots, \boldsymbol{X_N}$ are independent $\boldsymbol{N_q(\mu, \Sigma)}$ random vectors and $\boldsymbol{N > q}$ then the maximum likelihood (ML) estimates $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are $\hat{\boldsymbol{\mu}} = \overline{\boldsymbol{X}}$ and $\widehat{\boldsymbol{\Sigma}} = \frac{\boldsymbol{n}}{\boldsymbol{N}}\boldsymbol{S}$, where $\boldsymbol{n = N - 1}$ and $\overline{\boldsymbol{X}}$ and $\boldsymbol{S}$ are given by $\overline{\boldsymbol{X}} = \frac{\boldsymbol{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{X_i}$, and $\boldsymbol{S} = \frac{\boldsymbol{1}}{\boldsymbol{n}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}(\boldsymbol{X_i} - \overline{\boldsymbol{X}})(\boldsymbol{X_i} - \overline{\boldsymbol{X}})'$.

Proof: Ignoring the constant, the likelihood functions is

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = det\, \boldsymbol{\Sigma}^{-\frac{N}{2}}\, etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right) exp\left[-\frac{1}{2}N(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})\right],$$

where $\boldsymbol{A} = \sum_{i=1}^{N}(\boldsymbol{X}_i - \overline{\boldsymbol{X}})(\boldsymbol{X}_i - \overline{\boldsymbol{X}})'$.

Now $L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \leq \det \boldsymbol{\Sigma}^{-\frac{N}{2}}\, etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbf{A}\right)$,

with equality iff $\boldsymbol{\mu} = \overline{\boldsymbol{X}}$, where we have used the fact that $(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu}) = 0$ iff $\boldsymbol{\mu} = \overline{\boldsymbol{X}}$. This shows that $\overline{\boldsymbol{X}}$ is the ML estimate of $\boldsymbol{\mu}$ for all $\boldsymbol{\Sigma}$. It remains to maximize the function

$$L(\overline{\boldsymbol{X}}, \boldsymbol{\Sigma}) = det\, \boldsymbol{\Sigma}^{-\frac{N}{2}}\, etr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right),$$

or equivalently the function

$$\begin{aligned}
\mathrm{g}(\boldsymbol{\Sigma}) = \log \mathrm{L}(\overline{\mathbf{X}}, \boldsymbol{\Sigma}) &= -\frac{1}{2}\mathrm{N}\log\det\boldsymbol{\Sigma} - \frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{A}) \\
&= \frac{1}{2}\mathrm{N}\log\det\boldsymbol{\Sigma}^{-1}\mathbf{A} - \frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{A}) - \frac{1}{2}\mathrm{N}\log\det\mathbf{A} \\
&= \frac{1}{2}\mathrm{N}\log\det\mathbf{A}^{\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\mathbf{A}^{\frac{1}{2}} - \frac{1}{2}\mathrm{tr}\left(\mathbf{A}^{\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\mathbf{A}^{\frac{1}{2}}\right) - \frac{1}{2}\mathrm{N}\log\det\mathbf{A} \\
&= \frac{1}{2}\sum_{\mathrm{i}=1}^{\mathrm{q}}(\mathrm{N}\log\lambda_{\mathrm{i}} - \lambda_{\mathrm{i}}) - \frac{1}{2}\mathrm{N}\log\det\mathbf{A}
\end{aligned}$$

where $\lambda_1, \ldots, \lambda_q$ are the latent roots of $\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\mathbf{A}^{\frac{1}{2}}$, i.e., of $\boldsymbol{\Sigma}^{-1}\mathbf{A}$. Since the function

$$f(\lambda) = N\log\lambda - \lambda$$

has a unique maximum at $\lambda = N$ of $N\log N - N$ it follows that

$g(\boldsymbol{\Sigma}) \le \frac{1}{2}Nq\log N - \frac{1}{2}qN - \frac{1}{2}N\log\det\boldsymbol{A},$

or $L(\overline{\boldsymbol{X}}, \boldsymbol{\Sigma}) \le N^{\frac{qN}{2}}e^{-\frac{qN}{2}}\det\boldsymbol{A}^{-\frac{N}{2}}$, with equality iff $\lambda_i = N$ $(i = 1, \ldots, q)$. This last condition is equivalent to $\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}^{\frac{1}{2}} = N\boldsymbol{I}_q$ and hence to $\boldsymbol{\Sigma} = \frac{1}{N}\boldsymbol{A}$. Therefore we conclude that

$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \le N^{\frac{Nq}{2}}e^{-\frac{qN}{2}}\det\boldsymbol{A}^{-\frac{N}{2}}$, with equality iff $\boldsymbol{\mu} = \overline{\boldsymbol{X}}$ and $\boldsymbol{\Sigma} = \frac{1}{N}\boldsymbol{A}$, and the proof is complete.■

Note that $\boldsymbol{S} \sim W_q\left(n, \frac{1}{n}\boldsymbol{\Sigma}\right), \quad n = N - 1.$

Theorem 6-3: If $\boldsymbol{X} \sim \boldsymbol{N_q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi_q^2$, and $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X} \sim \chi_{\mathrm{q}}^2(\boldsymbol{\delta})$, where $\boldsymbol{\delta} = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$.

Theorem 6-4: Let $\overline{\boldsymbol{X}}$ and $\boldsymbol{S}$ be the mean and covariance matrix formed from a random sample of size $\boldsymbol{N} = \boldsymbol{n} + \mathbf{1}$ from $\boldsymbol{N_q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution ($\boldsymbol{n} \ge \boldsymbol{q}$), and let $\boldsymbol{T}^2 = \boldsymbol{N}\overline{\boldsymbol{X}}'\boldsymbol{S}^{-1}\overline{\boldsymbol{X}}$. Then

$$\frac{\boldsymbol{n} - \boldsymbol{q} + \mathbf{1}}{\boldsymbol{q}}\frac{\boldsymbol{T}^2}{\boldsymbol{n}}$$

is $\boldsymbol{F_{q,n-q+1}}(\boldsymbol{\delta})$, $\boldsymbol{\delta} = \boldsymbol{N}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$.

7. Multivariate Beta Distribution

Univariate

Recall that if $X \sim \chi_n^2$ is independent of $Y \sim \chi_m^2$, then $U = \frac{X}{X+Y}$ has beta type 1 distribution denoted by $U \sim B1(n,m)$ with the following density

$$f(u) = \frac{u^{n-1}(1-u)^{m-1}}{B(n,m)}, \quad 0 < u < 1,$$

where $B(n,m) = \frac{\Gamma(n)\Gamma(m)}{\Gamma(n+m)}$. The beta type 1 distribution is well known in Bayesian methodology as a prior distribution on the success probability of a binomial distribution. Also $V = \frac{U}{1-U} = \frac{X}{Y}$ has beta type 2 or inverted beta distribution denoted by $V \sim B2(n,m)$ with the following density

$$f(v) = \frac{v^{n-1}(1+v)^{-(n+m)}}{B(n,m)}, \quad v > 0.$$

Matrix variate

Let $\boldsymbol{X} \sim W_q(n, \boldsymbol{\Sigma})$ be independent of $\boldsymbol{Y} \sim W_q(m, \boldsymbol{\Sigma})$, where $n, m \geq q$. By analogy with the above univariate approach, we could consider $\boldsymbol{X}(\boldsymbol{X}+\boldsymbol{Y})^{-1}$ and $\boldsymbol{X}\boldsymbol{Y}^{-1}$, but these matrices are not symmetric and do not lead to useful density functions. However, since $\boldsymbol{X}$ and $\boldsymbol{Y}$, and hence $\boldsymbol{X}+\boldsymbol{Y}$, are positive definite matrices with probability 1, we can obtain symmetry by defining the positive definite matrices

$$\boldsymbol{U} = (\boldsymbol{X}+\boldsymbol{Y})^{-\frac{1}{2}}\boldsymbol{X}(\boldsymbol{X}+\boldsymbol{Y})^{-\frac{1}{2}} \quad and \quad \boldsymbol{V} = \boldsymbol{Y}^{-\frac{1}{2}}\boldsymbol{X}\boldsymbol{Y}^{-\frac{1}{2}}$$

where $\boldsymbol{Y}^{-\frac{1}{2}}$ and $(\boldsymbol{X}+\boldsymbol{Y})^{-\frac{1}{2}}$ are the symmetric square roots of $\boldsymbol{Y}$ and $(\boldsymbol{X}+\boldsymbol{Y})$, respectively.

Theorem 7-1: $\boldsymbol{U}$ has matrix variate beta type I, denoted by $\boldsymbol{U} \sim \boldsymbol{B}_q^{\boldsymbol{I}}(\boldsymbol{n},\boldsymbol{m})$, and the joint density of the $\frac{1}{2}\boldsymbol{q}(\boldsymbol{q}+\mathbf{1})$ distinct elements of $\boldsymbol{U}$, namely, $\boldsymbol{g}(\boldsymbol{u}_{11}, \boldsymbol{u}_{12}, \dots, \boldsymbol{u}_{qq})$ [or $\boldsymbol{g}(\boldsymbol{U})$ for short] is given by

$$\boldsymbol{g}(\boldsymbol{U}) = \frac{\mathbf{1}}{\boldsymbol{B}_q(\boldsymbol{n},\boldsymbol{m})} \mathbf{det}\, \boldsymbol{U}^{\boldsymbol{n}-\frac{1}{2}(\boldsymbol{q}+\mathbf{1})} \mathbf{det}\big(\boldsymbol{I}_q - \boldsymbol{U}\big)^{\boldsymbol{m}-\frac{1}{2}(\boldsymbol{q}+\mathbf{1})}, \quad \mathbf{0} < \boldsymbol{U} < \boldsymbol{I}_q,$$

where $\boldsymbol{n}, \boldsymbol{m} > \frac{1}{2}(\boldsymbol{q}-\mathbf{1})$, and $\boldsymbol{B}_q(\boldsymbol{n},\boldsymbol{m})$ is the multivariate beta function given by

$$\boldsymbol{B}_q(\boldsymbol{n},\boldsymbol{m}) = \frac{\boldsymbol{\Gamma}_q(\boldsymbol{n})\boldsymbol{\Gamma}_q(\boldsymbol{m})}{\boldsymbol{\Gamma}_q(\boldsymbol{n}+\boldsymbol{m})}.$$

Note that under the meaning of partial lowner ordering, for two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\boldsymbol{A} > \boldsymbol{B}$, means $\boldsymbol{A} - \boldsymbol{B}$ is positive definite.

Theorem 7-2: $\boldsymbol{V}$ has matrix variate beta type II, denoted by $\boldsymbol{V} \sim \boldsymbol{B}_q^{\boldsymbol{II}}(\boldsymbol{n},\boldsymbol{m})$, and the joint density of the $\frac{1}{2}\boldsymbol{q}(\boldsymbol{q}+\mathbf{1})$ distinct elements of $\boldsymbol{V}$, namely, $\boldsymbol{h}(\boldsymbol{v}_{11}, \boldsymbol{v}_{12}, \dots, \boldsymbol{v}_{qq})$ [or $\boldsymbol{h}(\boldsymbol{V})$ for short] is given by

$$\boldsymbol{h}(\boldsymbol{V}) = \frac{\mathbf{1}}{\boldsymbol{B}_q(\boldsymbol{n},\boldsymbol{m})} \mathbf{det}\, \boldsymbol{V}^{\boldsymbol{n}-\frac{1}{2}(\boldsymbol{q}+\mathbf{1})} \mathbf{det}\big(\boldsymbol{I}_q + \boldsymbol{V}\big)^{-(\boldsymbol{n}+\boldsymbol{m})}, \quad \boldsymbol{V} > \mathbf{0},$$

where $\boldsymbol{n}, \boldsymbol{m} > \frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{q} - \mathbf{1})$.