Jimma University

College of Natural Science

Department of Statistics

Course Title: INTRODUCTION TO MULTIVARIATE METHODS

Course Code: STAT3063

Instructor: Yasin Negash (MSc.)

April, 2012

Jimma, Ethiopia

Chapter One

Introduction

Scientific inquiry is an iterative learning process. Objectives pertaining to the explanation of a social or physical phenomenon must be specified and then tested by gathering and analyzing data. In turn, an analysis of the data gathered by experimentation or observation will usually suggest a modified explanation of the phenomenon. Thus, the complexities of the most phenomena require an investigator to collect observations on many different variables.

It will become increasingly clear that many multivariate methods are based up on underlying probability model known as multivariate normal distribution. Other methods are ad hoc in nature and are justified by logical or common sense arguments.

The objectives of scientific investigation to which multivariate methods most naturally lend themselves include the following:

1. Data reduction or structural simplification: The phenomenon being studied is represented as simply as with out sacrificing valuable information.
2. Sorting and grouping: or Variables are created, based upon similarò o
   measured characteristics. Alternatively, rules of classifying objects in to well defined groups may be required.
3. Investigation of dependency among variables: The nature of the relationships among variables is of interest. Are all the variables mutually independent or are one or more variables dependent on the others? If so, how?
4. Prediction: Relationship between variables must be determined for the purpose of predicting the values of one or more variables on the basis of observations of on the other variables.
5. Hypothesis construction and testing: Specific statistical hypotheses formulated in terms of the parameters of multivariate populations, are tested. This may be done to validate assumptions or to reinforce prior convictions.

The organization of data

Arrays

Multivariate data arise whenever an investiga
to record. The value of these variables are all recorded for each distinct item, individual or experimental unit.

Notations

X_{jk} --- The k^{th} variable that is observed on the j^{th} item

Or

X_{jk} ----- Measurement of the k^{th} variable on the j^{th} item

Consequently, n measurements on p variables.

Variable / Items	Variable-1	Variable-2	.	.	.	Variable-k	.	.	.	Variable-p
Item-1	X_{11}	X_{12}	.	.	.	X_{1k}	.	.	.	X_{1p}
Item-2	X_{21}	X_{22}	.	.	.	X_{2k}	.	.	.	X_{2p}
.	.	.	.	.	.	.				.
.	.	.	.	.	.	.				.
.	.	.	.	.	.	.				.
Item-j	X_{j1}	X_{j2}	.	.	.	X_{jk}	.	.	.	X_{jp}
.	.									
.	.									
.	.									
Item-n	X_{n1}	X_{n2}	.	.	.	X_{nk}	.	.	.	X_{np}

We can display these data as rectangular array, called **X,** of n rows and p columns.

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \ldots\ldots & X_{1k} & \ldots\ldots & X_{1p} \\ X_{21} & X_{21} & \ldots\ldots & X_{2k} & \ldots\ldots & X_{2p} \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ X_{j1} & X_{j2} & \ldots\ldots & X_{jk} & \ldots\ldots & X_{jp} \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ X_{n1} & X_{n2} & \ldots\ldots\ldots & X_{nk} & \ldots\ldots & X_{np} \end{bmatrix}$$

Example (data array)

A selection of four receipts from a university bookstore was obtained in order to investigate the nature of book sales. Each receipt provided, among other things, the number of sold and the total amount of book sale. Let the first variable be the total dollar sales and the second variable be number of books sold. Then we can regard the corresponding numbers on the receipts as four measurements on two variables.

Variable-1 (Dollar sales): 42 52 48 58

Vraible-2 (No. Of books): 4 5 4 3

Using the notations just introduced:

$$\mathbf{X} = \begin{bmatrix} 42 & 4 \\ 52 & 5 \\ 48 & 4 \\ 58 & 3 \end{bmatrix}$$

Descriptive statistics

A large data is bulky, and its very mass passes a series of obstacles to any attempt to visually extract pertinent information. Much of the information contained in the data can be assessed by calculating certain summary numbers known as descriptive statistics. For example, the arithmetic average or sample mean is descriptive statistics that provides a measure of location that is a ñcentral valueò for a set of numbers. And ave the mean provides a measure of spread, or variation in the numbers.

We shall rely most heavily on descriptive statistics that measure location, spread and linear association.

Let X_{11}, X_{21} , éX$_{n1}$, be n- measurements on the first variable. Then the arithmetic average of these measurements is:

$$\bar{x}_1 = \frac{\sum_{j=1}^{n} x_{j1}}{n}$$ ------ Sample mean for the first variable

The sample mean can be computed from the n-measurements on each of the p-variables, so that, in general, there will be p-sample means.

$$\bar{x}_k = \frac{\sum_{j=1}^{n} x_{jk}}{n-1}, \qquad k = 1, 2, \ldots, p$$

A measure of spread is provided by the sample variance, defined for n- measurements on the first variable as:

$$S_1^2 = \frac{\sum_{j=1}^{n} (x_{j1} - \bar{x}_1)^2}{n-1}$$

In general for p-variables,

$$S_k^2 = \frac{\sum_{j=1}^{n} (x_{jk} - \bar{x}_k)^2}{n-1}, \qquad k = 1, 2, \ldots, p$$

In this situation it is convenient to use double subscripts on the variances in order to indicate their position in the array. Therefore, we introduce the notation S_{ii} to denote the same variance computed from measurements on the variable.

$$S_k^2 = S_{kk} = \frac{\sum_{j=1}^{n} (x_{jk} - \bar{x}_k)^2}{n-1}, \qquad k = 1, 2, \ldots, p$$

Consider n-pairs of measurement on each of the variables 1& 2.

$$\begin{matrix} X_{11} & X_{21} & & X_{n1} \\ X_{12} & X_{22} & \cdots & X_{n2} \end{matrix}$$

That is, X_{j1} and X_{j2} are observed on the j^{th} experimental item (j=1,2, …, n). A measure of linear association between the measurements of variables 1 & 2 is provided by the sample covariance:

$$S_{12} = \frac{\sum_{j=1}^{n} (x_{j1} - \bar{x}_1)(x_{j2} - \bar{x}_2)}{n-1}$$

$$S_{ik} = \frac{\sum_{j=1}^{n} (x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k)}{n-1}$$

Or the average product of deviations from their respective means, we note that the covariance reduces to the sample variance when i=k. More over, $S_{ik} = S_{ki}$ for all i & k.

Sample correlation coefficient:

$$r_{ik} = \frac{S_{ik}}{\sqrt{S_{ii}}\sqrt{S_{kk}}} = \frac{\sum_{j=1}^{n}(x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k)}{\sqrt{\sum_{j=1}^{n}(x_{ji} - \bar{x}_i)^2}\sqrt{\sum_{j=1}^{n}(x_{jk} - \bar{x}_k)^2}}$$

For i = 1 , 2 , é . , p

k = 1 , 2 , é . p

Note that, $r_{ik} = r_{ki}$

The sample correlation coefficient is a standardized version of the sample covariance, where the product of the square roots of the sample variances provide the standardization.

Arrays of basic descriptive statistics

Sample means:

$$\bar{X} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}$$

Sample variances & covariances:

$$S_n = \begin{bmatrix} S_{11} & S_{12} & \ldots\ldots & S_{1p} \\ S_{21} & S_{22} & \ldots\ldots & S_{2p} \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ S_{p1} & S_{p2} & \ldots\ldots & S_{pp} \end{bmatrix}$$

Sample correlation;

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix}$$

Consider the previous example of sales and number of books

$$X = \begin{bmatrix} 42 & 4 \\ 52 & 5 \\ 48 & 4 \\ 58 & 3 \end{bmatrix}$$

$$\bar{x}_1 = \frac{\sum_{j=1}^{4} x_{j1}}{4} = \frac{(42+52+48+58)}{4} = \frac{200}{4} = 50$$

$$\bar{x}_2 = \frac{\sum_{j=1}^{4} x_{j2}}{4} = \frac{(4+5+4+3)}{4} = \frac{16}{4} = 4$$

The sample variances and covriances:

$$S_{11} = \frac{\sum_{j=1}^{4} (x_{j1} - \bar{x}_1)^2}{3}$$

$$= \frac{1}{3}\left[(42-50)^2 + (52-50)^2 + (48-50)^2 + (58-50)^2\right]$$

$$= 25.33$$

$$S_{22} = \frac{\sum_{j=1}^{4} (x_{j2} - \bar{x}_2)^2}{3}$$

$$= \frac{1}{4}\left[(4-4)^2 + (5-4)^2 + (4-4)^2 + (3-4)^2\right]$$

$$= \frac{2}{3} = 0.667$$

$$S_{12} = \frac{1}{3}\sum_{j=1}^{4} (x_{j1} - \bar{x}_1)(x_{j2} - \bar{x}_2)$$

$$= \frac{1}{3}\left[(42-50)(4-4)+(52-50)(5-4)+(48-50)(4-4)+(58-50)(3-4)\right]$$

$$= \frac{-6}{3} = -2$$

$S_{12} = S_{21}$

$$S_n = \begin{pmatrix} 25.33 & -2 \\ -2 & 0.667 \end{pmatrix}$$

Sample correlation:

$$r_{12} = \frac{S_{12}}{\sqrt{S_{11}}\sqrt{S_{22}}} = \frac{-2}{\sqrt{25.33}\sqrt{0.667}} = -0.4865$$

$r_{12} = r_{21}$

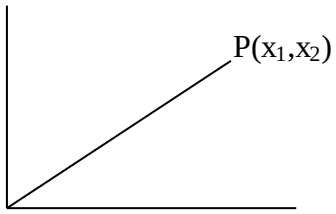
$$\mathbf{R} = \begin{pmatrix} 1.0 & -0.4865 \\ -0.4865 & 1.0 \end{pmatrix}$$

Exercise: calculate the descriptive statistics for the following data for a sample of observation:

Variable-1: (X1): 5 4 6 2 2 8 3

Vraible-2 (X2): 5 55 4 7 10 5 75

Distances



D(O,P)= $\sqrt{x_1^2 + x_2^2}$

If the point P has p-coordinates so that:

$P = (x_1, x_2, ..., x_p)$, the straight line distance from P to the origin O = (0, 0, é., 0) i

D(O,P)= $\sqrt{x_1^2 + x_2^2 + ... + x_p^2}$

The straight line distance between two arbitrary points P & Q with coordinates:

$P = (x_1, x_2, ..., x_p)$ and $Q = (y_1, y_2, ..., y_p)$ is :

$$d(P,Q) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_p - y_p)^2}$$

When the coordinates represent measurements that are subject to random fluctuations of differing magnitudes, it is often desirable to weight coordinates subject to a great deal of variability less heavily than those that are not highly variable. Our purpose is now to develop a ñ s t a t i s t i c a l ò d i s t a n c e t h a t a c c o u n t f o r d i f f e correlation. Because choice will depend up on the sample variances and covariances, at this point we use the term statistical distance to distinguish it from ordinary Euclidean distance. It is statistical distance that is fundamental to multivariate analysis.

Distances of standard coordinates

Suppose we have n- pairs of measurements on two variables say, x_1 and x_2:

The standard coordinates are:

$$x_1^* = \frac{x_1}{\sqrt{S_{11}}}, \quad x_2^* = \frac{x_2}{\sqrt{S_{22}}}$$

The quantities $\frac{1}{S_{11}}$ and $\frac{1}{S_{22}}$ are weights. Note that if the sample variances are the same, $\frac{1}{S_{11}}$ and $\frac{1}{S_{11}}$ then x_1^2 and x_2^2 will receive the same weight. In cases where the weights are the same, it is convenient to ignore the common divisor and use the usual Euclidean distance formula.

The expression can be generalized to accommodate the calculation of statistical distance from an arbitray point, $P=(x_1, x_2)$ to any fixed point $Q= (y_1,y_2)$,

$$d(P,Q)= \sqrt{\frac{(x_1 - y_1)^2}{S_{11}} + \frac{(x_2 - y_2)^2}{S_{22}}}$$

This can be extende to more than two dimensions.

A statistical distance of the point $P(x_1, x_2 , \text{é} \text{é}_p)$. to point $Q(y_1,y_2 , \text{é} \text{é}_p)$ can be computed from standardized coordinates is;

$$d(P,Q) = \sqrt{\frac{(x_1 - y_1)^2}{S_{11}} + \frac{(x_2 - y_2)^2}{S_{22}} + ... + \frac{(x_p - y_p)^2}{S_{pp}}}$$

$$= \sqrt{\sum_{i=1}^{p} \frac{(x_i - y_i)^2}{S_{ii}}}$$

If $S_{11} = S_{22} = \text{é} .S_{pp}$, Euclidean formula is appropriate.

Chapter-two

Matrix algebra and random vectors

A rectangular array of numbers, for instance, n- rows and p-columns is called a matrix of dimension n x p.

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \ldots\ldots & X_{1k} & \ldots\ldots & X_{1p} \\ X_{21} & X_{21} & \ldots\ldots & X_{2k} & \ldots\ldots & X_{2p} \\ \vdots & & & & & \\ X_{j1} & X_{j2} & \ldots\ldots & X_{jk} & \ldots\ldots & X_{jp} \\ \vdots & & & & & \\ X_{n1} & X_{n2} & \ldots\ldots & X_{nk} & \ldots\ldots & X_{np} \end{bmatrix}$$

Vectors

An array of **X** of n- real numbers $x_1, x_2, \ldots\ldots, x_n$ **is** called a vector, and it is written as

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \textbf{Or} \quad X' = \begin{bmatrix} x_1, x_2, \ldots\ldots, x_n \end{bmatrix}$$

A vector can be expanded or contracted by multiplying it by a constant **c.**

$$\mathbf{cX} = \begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{bmatrix}$$

A vector has both direction and magnitude. If n=2, then $\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

The length of **X** is written as Lx and defined as: $L_x = \sqrt{x_1^2 + x_1^2}$

The length of a vector $X' = [x_1, x_2, ..., x_n]$ with n- components is:

$$L_x = \sqrt{x_1^2 + x_1^2 + ... + x_n^2}$$

Multiplication of a vector by a scalar c changes the length:

$$L_{cx} = \sqrt{c^2 x_1^2 + ... + c^2 x_n^2}$$

$$= |c| \sqrt{x_1^2 + x_2^2 + ... + x_n^2}$$

The fundamental scalar relationship about the existence of an inverse of a number is a^{-1} such that $a^{-1}a = a a^{-1} = 1$ if $a \neq 0$ has the following matrix a such that:

$B*A_{(kxk)} = A*B_{(kxk)} = I_{(kxk)}$

Then B is called the inverse of A and is denoted by A^{-1}. The technical condition that an inverse exists is that the k-columns $a_1, a_2, ..., a_k$ are linearly independent. That is, the existence of A^{-1} is equivalent to:

$c_1 a_1 + c_2 a_2 + ...c_k a_k = 0$ only if $c_1 = c_2 = ... = c_k = 0$

Example

$$\mathbf{A} = \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix}$$

Let us see the existence of inverse of the matrix A:

$$c_1 \begin{pmatrix} 3 \\ 4 \end{pmatrix} + c_2 \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 3c_1 \\ 4c_1 \end{pmatrix} + \begin{pmatrix} 2c_2 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 3c_1 + 2c_2 \\ 4c_1 + c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\left.\begin{aligned} 3c_1 + 2c_2 &= 0 \\ 4c_1 + c_2 &= 0 \end{aligned}\right\}$$ Holds only if $c_1 = c_2 = 0$

This confirms the inverse of the matrix exists. The columns of A are linearly independent.

$$A^{-1} = \begin{pmatrix} -\frac{1}{5} & \frac{2}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{pmatrix}$$

Let A be a kxk square symmetric matrix. Then A has k- pairs of eigen values and eigen vectors namely,

$(\lambda_1, \mathbf{e_1}),\ (\lambda_2, \mathbf{e_2}),\ \ldots,\ (\lambda_k, \mathbf{e_k})$

The eigen values are chosen to satisfy: $\mathbf{e_1}'\mathbf{e_1} = \mathbf{e_2}'\mathbf{e_2} = \ldots = \mathbf{e_k}'\mathbf{e_k} = 1$ and be mutually perpendicular. The eigen values are unique unless two or more eigen values are equal.

$\mathbf{A} = \begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix}$ then, find the Eigen value and Eigen vector

$$|A - \lambda I| = 0$$

$$\left| \begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$$

$$\begin{vmatrix} 1-\lambda & -5 \\ -5 & 1-\lambda \end{vmatrix} = 0$$

$$\lambda^2 - 2\lambda - 24 = 0$$

$$\lambda_1 = 6, \quad \lambda_2 = -4$$

$\mathbf{A}^*\mathbf{X} = \lambda \mathbf{X}$

$$\begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 6 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$x_1 - 5x_2 = 6x_1$$

$$-5x_1 + x_2 = 6x_2$$

$$x_1 = -x_2$$

$$x_1 = 1, \quad x_2 = -1$$

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\mathbf{e_1} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$$

$\mathbf{A}^*\mathbf{X} = \lambda_2 \mathbf{X}$

$$\begin{pmatrix} 1 & -5 \\ -5 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = -4 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\left.\begin{aligned} x_1 - 5x_2 &= -4x_1 \\ -5x_1 + x_2 &= -4x_2 \end{aligned}\right\}$$

$$\left\{\begin{aligned} -5x_2 &= -5x_1 \quad \text{ie. } x_1 = x_2 = 1 \\ -5x_1 &= -5x_2 \end{aligned}\right.$$

$$\mathbf{X} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\mathbf{e_2} = \frac{X}{\sqrt{X'X}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

A spectral decomposition of a kxk symmetric matrix A is given by:

$$\mathbf{A} = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + ... + \lambda_k e_k e_k'$$

Where $\lambda_1, \lambda_2, ..., \lambda_k$ are the eigen values of A and $\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_k}$, are the associated normalized eigen vectors. Thus, $e_i' e_i = 1$ for all $e_i' e_j = 0$, $i \neq j$

Example: The spectral decomposition of a matrix A:

$$A = \begin{pmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{pmatrix}$$

The characteristic equation of is :

$$|A - \lambda I| = 0$$

$$\begin{pmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = 0 \quad \Rightarrow \quad \begin{vmatrix} 13-\lambda & -4 & 2 \\ -4 & 13-\lambda & -2 \\ 2 & -2 & 10-\lambda \end{vmatrix} = 0$$

$$(13-\lambda)\begin{pmatrix}13-\lambda & -2\\ -2 & 10-\lambda\end{pmatrix}-(-4)\begin{pmatrix}-4 & -2\\ 2 & 10-\lambda\end{pmatrix}+2\begin{pmatrix}-4 & 13-\lambda\\ 2 & -2\end{pmatrix}=0$$

$$\lambda_1 = 9,\ \lambda_2 = 9,\ \lambda_3 = 18$$

$$A = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \lambda_3 e_3 e_3'$$

$$\begin{pmatrix}13 & -4 & 2\\ -4 & 13 & -2\\ 2 & -2 & 10\end{pmatrix} = 9\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\\ 0\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\end{pmatrix} + 9\begin{pmatrix}\frac{1}{\sqrt{18}}\\ \frac{-1}{\sqrt{18}}\\ \frac{-4}{\sqrt{18}}\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{18}} & \frac{-1}{\sqrt{18}} & \frac{-4}{\sqrt{18}}\end{pmatrix} + 8\begin{pmatrix}\frac{2}{3}\\ \frac{-2}{3}\\ \frac{1}{3}\end{pmatrix}\begin{pmatrix}\frac{2}{3} & \frac{-2}{3} & \frac{1}{3}\end{pmatrix}$$

$$= \begin{pmatrix}\frac{9}{2} & \frac{9}{2} & 0\\ \frac{9}{2} & \frac{9}{2} & 0\\ 0 & 0 & 0\end{pmatrix} + \begin{pmatrix}\frac{1}{2} & \frac{-1}{2} & -2\\ \frac{-1}{2} & \frac{1}{2} & 2\\ -2 & 2 & 8\end{pmatrix} + \begin{pmatrix}8 & -8 & 4\\ -8 & 8 & -4\\ 4 & -4 & 2\end{pmatrix} = \begin{pmatrix}13 & -4 & 2\\ -4 & 13 & -2\\ 2 & -2 & 10\end{pmatrix}$$

Mean vector and covariance matrices

Suppose X= $(X_1, X_2, \ldots . X_p)$ is a px1 random vector. Then each element of X is a random variable with its own marginal probability distribution. The marginal means μ_i and variances σ_i^2 are defined as:

$$\mu_i = \begin{cases} \int_{-\infty}^{\infty} X_i f_i(x_i)\,dx_i \ldots\ldots\ldots \text{For continuous random variable Xi} \\ \sum_{i=1} X_i p_i(x_i) \text{ é é . For discrete random variable Xi} \end{cases}$$

$$\begin{cases} \int_{-\infty}^{\infty} (X_i - \mu_i)^2 f_i(x_i)\,dx_i \ldots\ldots\ldots \text{For continuous random variable Xi} \end{cases}$$

$\sigma_i^2 = \sum_{i=1}^{} (X_i - \mu_i)^2 p_i(x_i)$ é é . For discrete random variable

$\sigma_i^2 = \sigma_{ii}$ (marginal variances)

For random variables X_i & X_k,

$\sigma_{ik} = E(X_i - \mu_i)(X_k - \mu_k)$

$$= \begin{cases} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(X_i - \mu_i)(X_k - \mu_k) f_{ik}(x_i, x_k)dx_i d_k \text{..... For continuous random variable Xi} \\ \sum_{i=1}^{} \sum (X_i - \mu_i)(X_k - \mu_k)\, p_{ik}(x_i, x_k) \text{ é é . For discrete random var} \end{cases}$$

Where μ_i & μ_k are the marginal means for i, k = 1, 2, é., marginal variance.

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \\ \\ \\ \\ \\ X_p \end{bmatrix}, \qquad E(X) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \\ \\ \\ \\ \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \\ \\ \\ \\ \\ \mu_p \end{bmatrix}$$

$$\Sigma = E(X - \mu)(X - \mu)'$$

$$= E\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \\ \\ \\ \\ \\ X_p - \mu_p \end{bmatrix} [X_1 - \mu_1][X_2 - \mu_2].....[X_p - \mu_p]$$

=

COV (**X**) =

Example: Compute the covariance matrix of the following table

X2 X1	0	1	$P_1(x_1)$
-1	0.24	0.06	0.3
0	0.16	0.14	0.3
1	0.4	0.00	0.4
$P_2(x_2)$	0.8	0.2	1.0

= = =0.69

= =

=-0.08=

= = = = 0.16

= = =

The correlation coefficient is defined in terms of the covariance and the variances and as:

=

The correlation coefficient measures the amount of linear association between the random variables and .

The correlation matrix be the pxp symmetric matrix:

= =

The pxp standard deviation matrix is given by:

=

=

Note that = . The covariance matrix of is , and that of is and that of elements from and is or

The mean vector and covariances for linear combination of random variables

Recall that if single random variable, such as is multiplied by a constant C, then

= C = C

= =

If is a second random variable and a and b are constants, then using additional properties of expectation,

= =

= =

Finally, for the linear combinations ,

= =

=

$= E[aX_1 - a\mu_1 + bX_2 - b\mu_2]^2$

$= E[a(X_1 - \mu_1) + b(X_2 - \mu_2)]^2$

$= E[a^2(X_1 - \mu_2)^2 + b^2(X_1 - \mu_2)^2 + ab(X_1 - \mu_1)(X_2 - \mu_2)]$

$= a^2E(X_1 - \mu_1)^2 + b^2E(X_2 - \mu_2)^2 + abE(X_1 - \mu_1)(X_2 - \mu_2)$

$= a^2\sigma_{11} + b^2\sigma_{22} + ab\sigma_{12}$

Let $C' = [a, b]$, can be written as:

$$[a,b]\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = C'X$$

Z = =

The linear combinations:

= Have the following expectation and variance-covariance:

Where, and are the mean vector and variance ïcovariance matrix of X, respectively.

Example: Let = be a random vector with mean vector = and variance covariance matrix:

=

Find the mean vector and variance matrix for the linear combinations:

Or =

$$m_Z = Cm_X = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} m_1 - m_2 \\ m_1 + m_2 \end{bmatrix}$$

$$\ddot{a}_Z = C\ddot{a}_X C' = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

$$= \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} s_{11} - s_{12} & s_{12} + s_{11} \\ s_{21} - s_{22} & s_{22} + s_{21} \end{pmatrix}$$

$$= \begin{pmatrix} s_{11} - s_{12} - (s_{21} - s_{22}) & s_{11} + s_{12} - (s_{21} + s_{22}) \\ s_{11} - s_{12} + s_{21} - s_{22} & s_{11} + s_{12} + s_{21} + s_{22} \end{pmatrix}$$

$$= \begin{pmatrix} s_{11} - 2s_{12} - s_{22} & s_{11} - s_{22} \\ s_{11} - s_{22} & s_{11} + 2s_{21} + s_{22} \end{pmatrix}$$

Note that if $s_{11} = s_{22}$, that is if X_1 *and* X_2 have equal variances, the off diagonal terms in Σ_Z vanishes. This demonstrates the well-known result the sum and difference of two random variables with identical variances are uncorrelated.

Partitioning the sample mean vector and covariance matrices

Let $\bar{X}' = \left[\bar{X}_1, \bar{X}_2, ..., \bar{X}_p\right]$ be the vector of sample averages constructed from n- observations on p-variables $X_1, X_2, ..., X_p$ and

Let, $$S_n = \begin{pmatrix} S_{11} & S_{12} & .....S_{1p} \\ S_{21} & S_{22} & .....S_{2p} \\ & & \\ & & \\ & & \\ & & \\ S_{p1} & S_{p2} & .....S_{pp} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{n}\sum_{j=1}^{n}(x_{j1} - \bar{x}_1)^2 & \frac{1}{n}\sum_{j=1}^{n}(x_{j1} - \bar{x}_1)(x_{j2} - \bar{x}_2) & ..... & \frac{1}{n}\sum_{j=1}^{n}(x_{j1} - \bar{x}_1)(x_{jp} - \bar{x}_p) \\ \frac{1}{n}\sum_{j=1}^{n}(x_{j2} - \bar{x}_2)(x_{j1} - \bar{x}_1) & \frac{1}{n}\sum_{j=1}^{n}(x_{j2} - \bar{x}_2)^2 & ..... & \frac{1}{n}\sum_{j=1}^{n}(x_{j2} - \bar{x}_2)(x_{jp} - \bar{x}_p) \\ & & & \\ & & & \\ & & & \\ \frac{1}{n}\sum(x_{jp} - \bar{x}_p)(x_{j1} - \bar{x}_1) & \frac{1}{n}\sum(x_{jp} - \bar{x}_p)(x_{j2} - \bar{x}_2) & ..... & \frac{1}{n}\sum(x_{jp} - \bar{x}_p)^2 \end{pmatrix}$$

be the corresponding sample variance-covariance matrix.

The sample mean vector and covariance matrix can be partitioned in order to distinguish quantities corresponding to groups of variables. Thus,

$$\bar{X} = \begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_q \\ \bar{X}_{q+1} \\ \vdots \\ \bar{X}_p \end{pmatrix} = \begin{pmatrix} \bar{X}^{(1)} \\ \bar{X}^{(2)} \end{pmatrix}$$

$$S_n = \begin{pmatrix} S & S & S_q & S_{q+} & S_p \\ S & S & S_q & S_{q+} & S_p \\ S_q & S_q & S_{qq} & S_{q\,q+} & S_{q\,p} \\ S_{q+} & S_{q+} & S_{q+\,q} & S_{q+\,q+} & S_{q+\,p} \\ S_p & S_p & S_{p,q} & S_{p,q+1} & S_{p,p} \end{pmatrix} =$$

Where, are the sample mean vectors constructed from observations respectively. is the sample covariance matrix computed from observations , is the sample covariance matrix computed from ; $= S_{21}$ is the sample covariance matrix for elements of and elements of .

Sample values of linear combinations of variables

Consider the constant vector C and the random variable vector X.

$$C = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_p \end{bmatrix}, \quad X' = [X_1, X_2, \ldots\ldots X_p]$$

=

Whose observed value on the j[th] trial is given by:

= , j = 1 , 2 é n

The n-observations has sample mean:

Sample mean =

= =

Sample variance =

=

Sample variance of $C'X = C'SC$

Suppose $b'X = b_1X_2 + b_2X_2 + \ldots\ldots + b_pX_p$, whose observed value on the j[th] trial is:

$b'X_j = b_1X_{j2} + b_2X_{j2} + \ldots\ldots + b_pX_{jp}$, j = 1 , 2 , é . . , n

Sample mean of $b'X = b'\overline{X}$

Sample variance of $b'X = \boldsymbol{b'S\ b}$

The sample covariance computed form pairs of observations on $b'X$ and $C'X$ is:

$$b'SC = \frac{(b'X_1 - b'\bar{X})(C'X_1 - C'\bar{X}).....(b'X_n - b'\bar{X})(C'X_n - C'\bar{X})}{n}$$

=

Example: Means and covariance for linear combinations

=

Consider the two linear combinations:

$b'X$ = =

= =

$= 2X_1 + 2X_2 - X_3$

=2(1) +2(2) ï5=1

$b'X_2 = 2X_1 + 2X_2 - X_3 = 2(4) + 2(1)$ ï6=4

$b'X_3 = 2X_1 + 2X_2 - X_3 = 2(4) + 2(0)$ ï4 =4

Sample mean = $\frac{1+4+4}{3}$ =3

Sample variance = $\frac{(1-3)^2 + (4-3)^2 + (4-3)^2}{3} = 2$

In similar manner, n=3 observations are:

= =

$C'X_1 = X_{11} + X_{12} + X_{13}$ =1 ï2 + 3(5)=14

$C'X_2$ =1(4) ï1 + 3(6)= 21

$C^{'}X_3$ =1(4) ï1(0) + 3(4)= 16

Sample mean = $\frac{14+21+16}{3}$ =17

Sample variance = $\frac{(14-17)^2+(21-17)^2+(16-17)^2}{3}=\frac{26}{3}$

More over the sample covariance computed from the pairs of observations $(b^{'}X_1,\ C^{'}X_1)$, and

Sample covariance:

=

=

Alternatively, we use sample mean vector and sample covariance matrix S derived from the original data X to calculate the sample means, variances and covariances for the linear combinations. Thus, if only the descriptive statistics are of interest, we do not even need to calculate the observations and .

Sample mean of $= b^{'}\bar{X}$

$$= \begin{bmatrix}2 & 2 & -1\end{bmatrix}\begin{bmatrix}3\\1\\5\end{bmatrix} = 3$$

Sample mean of $C^{'}X = C^{'}\bar{X}$

$$= \begin{bmatrix}1 & -1 & 3\end{bmatrix}\begin{bmatrix}3\\1\\5\end{bmatrix} = 17$$

Sample variance of $b^{'}X = b^{'}S\,b$

$$= \begin{bmatrix} 2 & 2 & -1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 0 \\ -1 & \frac{2}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 & 2 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = 2$$

Sample variance of $C'X$ =

= =

Sample covariance of $b'X$ and $C'X$ is

= = =

Excercises

Chapter-Three

Multivariate normal distribution

A generalization of familiar bell-shaped normal density to several dimensions; play a fundamental role in multivariate analysis. In fact, most of the techniques encountered in this subject are based on the assumption that the data were generated from a multivariate normal distribution. While real data were never exactly multivariate normal, the normal density is often a useful approximation to the "true" population distribution.

The normal distributions are useful in practice for two reasons. First, the normal distribution serves as a bona fide population model in some instances. Second, the sampling distributions of many multivariate statistics are approximately normal, regardless of the form of parent distribution, because of a central-limit effect.

The multivariate normal density is a general dimensions. Recall that the univariate normal distribution with mean m and variance s^2 has the probability density function:

, -Đ < X < Đ

The notation: can be extended to multivariate case. A Px1 vector of X of observation on several variables can be written as:

The P X 1 vector represents the expected value of the random vector X and a P X P matrix is the variance ïcovariance matrix of X. A p-dimensional normal density for the random vector X = has the form:

Where, -Đ ≤ < XĐ , i = 1 , 2 . . . p

X ~

Example: Bivariate normal density, P=2,

This 2-dimensional normal density function is denoted by X ~ $N_2(\mu, \ddot{a})$

$\mu_1 = E(\bar{X}_1)$, , ,

=

= =

Therefore, can be written as the following form:

=

=

=

We have, =

Then we have:

=

Additional properties of multivariate normal distribution

If a random vector X has a multivariate normal distribution then,

1. Linear combinations of the components of X are normally distributed.
2. All subsets of the components of X have a multivariate normal distribution.
3. Zero covariance implies that the corresponding components are independently distributed.
4. The conditional distributions of the components are also multivariate normal.

If a vector X is distributed as Np ~ $(\mu, \ddot{a})$, then any linear combination of the variables:

$a'X$ = is distributed as . Also if $a'X$ is distributed as

for every a, then X must be .

Proof

We have X~

=

E ($a'X$) = = =

= =

Consider the linear combination of $a'X$ of multivariate normal vector determined by choice:

= :

$a'X$ = =

$a'\mu$ = $\begin{pmatrix} \mu_1 \\ \mu_2 \\ \\ \\ \\ \\ \\ \mu_p \end{pmatrix}$ = μ_1

$a' \ddot{a}\ a =$ =

Therefore, the distribution of is . More generally, the marginal distribution of any component of is .

If X is distributed as , the q-linear combinations:

=

are distributed as $N_q(Am, A\ddot{a}\ A')$

Example:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$$

X~ $N_3(m, \ddot{a}\ )$, find the distribution of AX:

Solution

$$A\,X = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} X_1 - X_2 \\ X_2 - X_3 \end{pmatrix}$$

=

= , =

=

= , $X_2 = \begin{pmatrix} X_1 \\ X_3 \\ X_5 \end{pmatrix}$

$\sim N_2(\mu_1, \Sigma_{11})$

It is clear that from this example that the normal distribution for any subset can be expressed by simply selecting the appropriate means and covariances from the original and . In general, all subsets of X are normally distributed. If we respectively partition X, its mean and covariance as:

= , =

$\sim N_q(\mu_1, \Sigma_{11}), \quad X_2 \sim N_{p-q}(\mu_2, \Sigma_{22})$

=

= and > 0. Then the conditional distribution of given $X_2 = x_2$

$X_1 - \mu_1 - \ddot{a}_{12}\ddot{a}_{22}^{-1}(X_2 - \mu_2)$ when X_2 has the particular value x_2. Equivalently, given that $X_2 = x_2$, X_1 is distributed as;

Let X be distributed as with >0, then

Let $X_1, X_2, ..., X_n$ be mutually independent with distributed as . (Assume each has the same covariance matrix). Then,

= is distributed as:

. Moreover, and = are jointly multivariate normal with covariance matrix.

Consequently, and are independent if =0

Proof

Consider the np component vector,

= ~1xnp is a multivariate normal. In particular, X is distributed as , where

and =

The choice of the matrix A will take the form:

, where I is pxp identity matrix given as:

$$= \begin{bmatrix} \sum_{j=1}^{n} C_j X_j \\ \sum_{j=1}^{n} b_j X_j \end{bmatrix} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

is normal

The block diagonal terms have:

=

And

=

The off-diagonal term if of the form:

= .... is the covariance matrix for and . Consequently, when

=0, so that = 0, and are independent.

Example

Let be independent and identically distributed 3x1 random vectors with:

= and =

We first consider a linear combination of the three components of . This is a random variable with mean:

= = and variance,

=

Covariance matrix:

$$\left[b^2{}_1 + b^2{}_2 + b^2{}_3 + b^2{}_4\right] = 12x = \begin{bmatrix} 36 & -12 & 12 \\ -12 & 12 & 0 \\ 12 & 0 & 24 \end{bmatrix}$$

Finally the covariance matrix for the two linear combinations of random vectors is:

$$\left[C_1 b_1 + C_2 b_2 + C_3 b_3 + C_4 b_4\right] = \quad = \quad = 0x \quad = 0$$

Every component of the fist linear combination of random vectors have zero covariance with each component of the second combination of random vectors.

Sampling from multivariate normal distribution

The multivariate normal likelihood:

Let us assume that the px1 vectors represent a random from a multivariate normal population with mean vector and covariance matrix . Since are independent and each has distribution , the joint density function of all the observations is the product of the marginal normal densities.

=

When the numerical values of the observations become available, they may be substituted for the X_j . The resulting expression now considered as a function of μ and Σ for the fixed set of observations is called the likelihood function.

Let be random sample from a normal population with mean μ and covariance Σ . Then,

and = = are maximum likelihood estimator of m and $ä$ respectively. These observed values, and are called maximum likelihood estimate of m and $ä$.

The maximum likelihood estimator of = ,

Where = is the maximum likelihood estimator of .

=

=

=

= constant

The generalized variance determines the "peakedness" of the likelihood function and consequently is a natural measure of variability when the parent population is multivariate.

The sampling distribution of

The assumption that constitutes a random sample from a normal population with mean vector and covariance matrix completely determines the sampling distribution of .

= is distributed as times a chi-square variable having (n-1) degrees of freedom. In turn, this chi-square is the distribution of a sum of squares of independent standard normal random variables. is distributed as

=

The individual ~

The sampling distribution for the sample covariance matrix is called Wishart distribution.

= Wishart distribition with m-degrees of freedom

= distribution of

Where 's are independently distributed as $N_p(0, \ddot{a})$

In summary:

Let be a random sample from p-variate normal distribution with mean and covariance matrix . Then

1. $\bar{X}$ is distributed as $N_p\left(\mu, \frac{1}{n}\ddot{a}\right)$

2. is distributed as a Wishart random matrix with (n-1) degrees of freedom

3. $\bar{X}$ and S^2 are independent

Let be independent observations from a population with mean and finite (non-singular) matrix . Then,

$\sqrt{n}(\bar{X} - \mu)$ is approximately $N_p(0, \ddot{a})$ and

$n(\bar{X} - \mu)' S^{-1}(\bar{X} - \mu)$ is approximately χ^2_p for (n-1) large.

Chapter Four

Inference about mean vector

$$T^2 = \left(\bar{X} - \mu_0\right)\left(\frac{1}{n} S\right)^{-1}\left(\bar{X} - \mu_0\right)$$

=

Where ,

=

The statistics T^2 is called Hotelling's T^2 in hounor of Harold Hotelling, a pioneer in multivariate analysis. Here is the estimated covariance matrix of .

If the observed statistical distance T^2 is too large, that is is "too far" from , the hypothesis is rejected.

$T^2 \sim$, where is an F- distribution with p and (n-p) degrees of freedom.

In general , let be a random sample from population. Then with

and

$$a = P\left[T^2 > \frac{(n-1)p}{(n-p)} F_{p,n-p}(a)\right] = P\left[n\left(\bar{X} - \mu\right)' S^{-1}\left(\bar{X} - \mu\right) > \frac{(n-1)p}{(n-p)} F_{p,n-p}(a)\right]$$

is the upper (100)th percentile of the distribution. This leads to the hypothesis; vs . At level of significance, we reject in favour of , if the observed

T^2 = >

Example: let the data matrix for a random sample of size n=3 from a bivariate normal population be given as follows:

X=

= =

T^2 = , = - =

$T^2 = (3)\begin{bmatrix} -1, 1 \end{bmatrix}$ $= (3)\begin{bmatrix} -1, 1 \end{bmatrix}\begin{bmatrix} -\frac{2}{9} \\ \frac{1}{27} \end{bmatrix} = (3)\left(\frac{2}{9} + \frac{1}{27}\right) = \frac{7}{9}$

Therefore,

$$S_y = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \bar{Y})(Y_j - \bar{Y})' = C\mu + d$$

Therefore, T^2 computed with the y's and a hypothesized value $\mu_{y.0} = C\mu_0 + d$ is:

$$T^2 = n(\bar{y} - \mu_{y,0})' S_y^{-1}(\bar{y} - \mu_{y,0})$$

=

=

=

=

=

The likelihood ratio tests

The maximum of the multivariate normal likelihood as and are varied over their possible values is given by:

=

Where, the estimate of = are the MLE 's

and $= \dfrac{\sum_{j=1}^{n} X_j}{n}$

Recall that and the estimator of are those choices for and that best explain the observed values the random sample. Under the hypothesis , the normal likelihood specializes to:

=

Here is fixed, but be varied to find the value that is "most likely" to the observed sample. The value is obtained by maximizing with respect to .

=

With estimate of $= \frac{1}{n}\sum_{j=1}^{n}(X_j - \mu_0)(X_j - \mu_0)'$

To determine whether is a plausible value of μ, the maximum of is compared with the unrestricted maximum of $L(\mu, \Sigma)$. The resulting ratio is called the likelihood ratio statistic.

Likelihood ratio= $\Lambda = \frac{\max_{\Sigma} L(\mu_0, \Sigma)}{\max_{\mu,\Sigma} L(\mu, \Sigma)} = \left(\frac{|\Sigma|}{|\Sigma_0|}\right)^{\frac{n}{2}}$

The equivalent statistic $\Lambda^{\frac{2}{n}} = \frac{|\Sigma|}{|\Sigma_0|}$ is called Wilk's Lamda . If the observed value of this likelihood ratio is too small, the hypothesis is unlikely to be true and is therefore rejected.

Let be a random sample from an $N(\mu, \Sigma)$ population. Then the test based on T^2 is equivalent to ratio test of versus .

$$\Lambda^{\frac{2}{n}} = \left(1 + \frac{T^2}{n-1}\right)^{-1}$$

Proof

Let (p+1)x(p+1) matrix:

$$A = \begin{bmatrix} \sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})' & \sqrt{n}(\bar{X} - \mu_0) \\ \sqrt{n}(\bar{X} - \mu_0)' & -1 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

$|A| =$

$=$

$=$

$=$

We can write:

$=$

$= \qquad + n(\bar{X} - \mu_0)(\bar{X} - \mu_0)'$

$$(-1)\left|\sum_{j=1}^{n}(X_j - \mu_0)(X_j - \mu_0)'\right| =$$

Or =

Thus, $\Lambda^{\frac{2}{n}} = \frac{|\Sigma|}{|\Sigma_0|} =$

Here is rejected for small value of $\Lambda^{\frac{2}{n}}$ or equivalently, for large values of T^2.

Confidence regions

The confidence region for the mean of a p-dimensional normal population is:

=

For a particular sample, can be computed, and the inequality Ò

$$\left(\frac{n-1}{n-p}\right)pF_{p,n-p}(\alpha)$$

Where , $S = \frac{\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})'}{n-1}$

To determine whether any falls with in the confidence region (is plausible value of μ), we need to compute the generalized squared distance and compare it with $\left(\frac{n-1}{n-p}\right)pF_{p,n-p}(\alpha)$. If the squared distance is larger than $\left(\frac{n-1}{n-p}\right)pF_{p,n-p}(\alpha)$, is not in the confidence region.

Example: Consider the bivariate case, i.e., p=2

n=42,

,

The 95 percent confidence ellipse for consists of all values satisfying:

$$\text{Ò } \left(\frac{n-1}{n-p}\right) pF_{p,n-p}(a)$$

=

= Ò

=2.05x3.23=6.6215

42(203.018) +42(200.228) - 84(163.391) Ò 6 . 6 2

To see whether =

Where $t_{n-1}\left(\frac{a}{2}\right)$ is the upper 100 percentile of a t-distribution with (n-1) degrees of freedom.

If , ... is the usual confidence interval for the univariate random variable.

Let be a random sample from an population covariance matrix . Then, simultaneously for all "a" the interval is:

will contain

with probability 1- .

Proof

$T^2 =$ Ò implies Ò for every a.

Ò Ò for every a. Choosing =

gives the interval that will contain for all a, with probability 1 - EMBED Equation.3 = EMBED Equation.3 .

Successive choices of , , and so on through EMBED Equation.3 for the interval allows us to conclude that:

Ò Ò EMBED Equati

$$\bar{X}_1 + \sqrt{\frac{p(n-1)}{n-p}F_{p,n-p}(a)}\sqrt{\frac{S_{11}}{n}}$$

X_{jk} ÒX_{jk} Ò$\bar{X}_2 + \sqrt{\frac{p(n-1)}{n-p}F_{p,n-p}(a)}\sqrt{\frac{S_{22}}{n}}$

.

$$\bar{X}_2 + \sqrt{\frac{p(n-1)}{n-p}} F_{p,n-p}(\alpha) \sqrt{\frac{S_{22}}{n}} \quad \text{Ò} \qquad \text{Ò}$$

all holds simultaneously with confidence coefficient 1- .

With out modifying the coefficient (1-), we can make statements about the differences corresponding to:

, where

, then,

Ò Ò

is the 100(1-) percent confidence interval for the

difference, .

Chapter Five

Comparisons of Several Multivariate Means

Paired comparison

Suppose we have p-responses, two treatments and n-experimental units.

X_{1j1} =variable -1 under treatment -1

X_{1j2} = variable-2 under treatment-1

.

.

.

X_{1jp} = variable-p under treatment-1

X_{2j1} =variable -1 under treatment -2

X_{2j2} = variable-2 under treatment-2

.

.

.

X_{2jp} = variable-p under treatment-2

The p-paired difference random variables become,

$D_{j1} = X_{1j1} - X_{2j1}$

$D_{j2} = X_{1j2} - X_{2j2}$

.

.

.

$= X_{1jp} - X_{2jp}$

Let = , j=1,2,...,p

= = , =

If, in addition, are independent random vectors, inferences about the vectors of mean differences can be used up on a T^2-statistic.

T^2 = , where and

This T^2-is distributed as . If n and n-p are both large, T^2 is approximately distributed as a random variable, regardless of the form of the underlying population of the differences. Given the observed differences: , j=1,2,.....,n corresponding to random variables, an **-level**

$\delta_i : \bar{d}_i \mp \sqrt{\frac{p(n-1)}{n-p} F_{p,n-p}(\alpha)} \sqrt{\frac{S_{di}^2}{n}}$, where the δ_i is the i[th] element of $\bar{d}$ and is the i[th] diagonal element of .

For (n-p) large , and normality need not be assumed.

Comparison of mean vectors from two different populations

Consider a random sample of size n_1 from population -1 and a sample of size n_2 from population-2. The observation on p-variables can be arranged as:

Sample

1. , ,

2. , , $S_2 = \frac{\sum_{j=1}^{n} (X_{2j} - \bar{X}_2)(X_{2j} - \bar{X}_2)'}{n_2 - 1}$

Assumptions regarding the structure of the data:

1. The sample is a random sample from a p-variate population with mean vector μ_1 and covariance matrix Σ_1.

2. The sample $X_{21}, X_{22}, \ldots . X_{2n_2}$ is a random sample from a p-variate population with mean vector μ_2 and covariance matrix Σ_2.

3. Also the two samples and $X_{21}, X_{22}, \ldots . X_{2n_2}$ are independent of each other.

For small n_1 *and* n_2, further assumptions are imposed.

- Both populations are multivariate normal
- Also $\Sigma_1 = \Sigma_2$ (The same covariance matrix)

=

=

Example: Fifty bars of soap are manufactured in each of two ways. Two characteristics, X_1 =lather and X_2 =mildness, are measured. The summary statistics for bars produced by method 1 and 2 are described below.

$$\bar{X}_1 = \begin{pmatrix} 8.3 \\ 4.1 \end{pmatrix}, \qquad S_1 = \begin{pmatrix} 2 & 1 \\ 1 & 6 \end{pmatrix}$$

= , =

Test the hypothesis of the equality of two mean vectors of the population i.e.,

or

or

=0.05

S_1 and are approximately equal, so it is reasonable to pool them.

= = $\begin{pmatrix} 2 & 1 \\ 1 & 6 \end{pmatrix} + \frac{49}{98}$

$$= \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 3 \end{pmatrix} + \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 2 \end{pmatrix}$$

$$= \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}$$

$$S_{pooled}^{-1} = \frac{1}{2}\begin{pmatrix} 5 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} \frac{5}{9} & -\frac{1}{9} \\ -\frac{1}{9} & \frac{2}{9} \end{pmatrix}$$

$$T^2 = \left(\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)\right)' \left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{pooled}^{-1} \left(\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)\right)$$

$$= \left(\bar{X}_1 - \bar{X}_2\right)' \left(\frac{n_1 n_2}{n_1 + n_2}\right) S_{pooled}^{-1} \left(\bar{X}_1 - \bar{X}_2\right)$$

$$\bar{X}_1 - \quad = \begin{bmatrix} 8.3 \\ 4.1 \end{bmatrix} - \qquad = \begin{bmatrix} -1.9 \\ -0.2 \end{bmatrix}$$

$$\frac{n_1 n_2}{n_1 + n_2} = \frac{50 * 50}{50 + 50} = \frac{2500}{100} = 25$$

$$T^2 = 25\left[-1.9, -.02\right] \begin{bmatrix} \frac{5}{9} & -\frac{1}{9} \\ -\frac{1}{9} & \frac{2}{9} \end{bmatrix} \begin{bmatrix} -1.9 \\ -0.2 \end{bmatrix}$$

$$= 25\left[-1.9, -.02\right] \begin{bmatrix} -\frac{9.3}{9} \\ -\frac{1.5}{9} \end{bmatrix} = 25\frac{17.67}{9} = 49.08$$

$$C^2 = \frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F_{p, n_1 + n_2 - p - 1} = \frac{98}{97} * 2 \; F_{2,97}(0.05) = 2.02 * 3.1 = 6.26$$

Since T^2 calculated larger than the critical value for decision, we reject and conclude that the mean vectors are significantly distinct.

Multivariate analysis of variance (MANOVA)

MANOVA model for comparing g-population mean vectors is given ass:

$X_{lj} = \mu + t_l + e_{lj}$, j = 1 , 2 , é é , n

l = 1 , 2 , é é , g

Where e_{lj} *are independent* $N_p(0, \Sigma_p)$ variables. Here the parameter vector μ is an over all mean (level) and τ_l represents the l[th] treatment effect with $\sum_{l=1}^{g} n_l \tau_l$

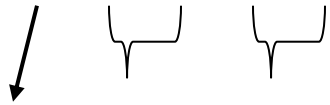
---response

μ----Overall mean

----treatment effect

---random error

A vector of observation may be decomposed as:



overall estimated residual

mean treatment effect

Consider:

=

=

Taking the sum over j , yields the cross product term vanish, because

Hence, $\sum_{l=1}^{g}\sum_{j=1}^{n_l}(X_{lj} - \bar{X})(X_{lj} - \bar{X})' = \sum_{l=1}^{g} n_l(\bar{X}_l - \bar{X})(\bar{X}_l - \bar{X})' + \sum_{l=1}^{g}\sum_{j=1}^{n_l}(X_{lj} - \bar{X}_l)(X_{lj} - \bar{X}_l)'$

Total Between (treatment) residual (with in)

sum of sum of sum of squares

sqaures squares

The hypothesis to be tested is:

(no treatment effect)

MANOVA table for comparing population mean vetcors

Sourse of variation	Sum of squares	degrees of freedom	
Treatment	B= $\sum_{l=1}^{g} n_l (\bar{X}_l - \bar{X})(\bar{X}_l - \bar{X})'$	g-1	
Residual (Error)	W=		

No of variables	No. of groups	sampling distribution for multivariate normal
p=1	g Ó 2	
p=2	g Ó 2	
p Ó 1	g=2	
p Ó 1	g=3	

If is large,

has

Chapter- Six

Principal component and Factor analysis

A principal component analysis is concerned with explaining the variance covariance structure of a set of variables through a few linear combinations of these variables. Its general objectives are:

(1) Data reduction &

(2) Interpretation

Although p-components are required to reduce, the total system variability, often this much of the variability can be accounted for by a small number of k- principal components. If so, there is (almost) as much information in the k-components as there is in the original p-variables. The k-principal components can then replace the initial p-variables, and the original data set, consisting of n- measurements on p-variables, is reduced to a data set consisting of n-measurements on k-principal components.

Analysis of principal components are more of a means to an end rather than an end in them selves, because they frequently serve as intermediate steps in much larger investigation.

Principal components depend solely on the covariance matrix (Or the correlation matrix r) of $X_1, X_2, ..., X_p$. Let the random vector $X' = [X_1, X_2, ..., X_p]$ have the covariance matrix with eigen values $/_1 \geq /_2 \geq ..... \geq /_p \geq 0$. Consider the linear combinations:

$$Y_1 = a_1' X = a_{11}X_1 + a_{12}X_2 + ... + a_{1p}X_p$$

$$Y_2 = a_2' X = a_{21}X_1 + a_{22}X_2 + ... + a_{2p}X_p$$

.

.

.

$$Y_p = a_p' X = a_{p1}X_1 + a_{p2}X_2 + ... + a_{pp}X_p$$

In general it can be written as:

$Y_i = a_i' X$, i=1,2,....,p

$Var(Y_i) = a_i' Var(X) a_i$

, i,k=1,2,....,p

The principal components are those uncorrelated linear combinations whose variances are as large as possible. The first principal component is the linear combination is with maximum variance. That is, it maximizes,

= . = , can be increased by multiplying any by some constant. To eliminate this indeterminacy, it is convenient to restrict attention to coefficient vectors of unit length.

First principal component = linear combination that maximizes subject to

Second principal component = linear combination that maximizes subject to

and =0

.

.

.

At ith step, ith principal component= linear combination that maximizes subject to

and =0 for i<k.

Let be the covariance matrix associated with the random vector $X' = [X_1, X_2, ..., X_p]$. Let have the eigen value-eigen vector pairs where $/_1 ² /_2 ² ..... ² /_p ² 0$. Then the ith principal component is given by:

= , i=1,2,....,p with the choices :

$Var(Y_i)$ = = , i=1,2,....,p

, for

If some 's are equal, the choices of coefficient vectors, and are not unique. Let $X' = [X_1, X_2, ..., X_p]$ have matrix , with eigen values ïeigen vectors where $/_1 ² /_2 ² ..... ² /_p ² 0$.

Let , ,....., be the principal component. Then,

= = =

Proof

=

=

We can write = , where is the diagonal matrix of eigen values and so that = = I

= = $= tr(L) =$

Thus = $= \sum_{i=1}^{p} Var(Y_i)$

$$\sum_{i=1}^{p} Var(Y_i) = \sum_{i=1}^{p} Var(e_i' X) = \sum_{i=1}^{p} e_i' \sum e_i = \sum_{i=1}^{p} /_i e_i e'_i = \sum_{i=}^{p} /_i$$

Total population variance: =

The proportion of total variance due to (explained by) the k^{th} principal component is:

The proportion of total variance due to

the k^{th} principal component =

$$= [0.383, \; -0.924, \; 0] \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

$$= 0.383X_1 - 0.924X_2$$

$$= [0, \; 0, \; 1] \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = X_3$$

$$Y_3 = e_3' X = [0.924, \; 0.383, \; 0] \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = 0.924X_1 + 0.383X_2$$

The variable X_3 is one of the principal components, because it uncorrelated with the other two variables.

$Var(Y_1) = \text{Var}\,(0.383X_1 - 0.924X_2)$

$= (0.383)^2 \text{var}(X_1) + (0.924)^2 Var(X_2) + 2(0.383)(-0.924)Cov(X_1, X_2)$

$= 0.147\,(1) + 0.854\,(5) - 0.708(-2) = 5.83 = /_1$

$Var(Y_2) = Var(X_3) = 2 = /_2$

$Cov(Y_1, Y_2) = Cov(0.383X_1 - 0.924X_2, X_3)$

$= 0.383Cov(X_1, X_2) - 0.924Cov(X_2, X_3)$

$= 0.383\,(0) - 0.924(0) = 0$

Total population variance: $= 1+5+2=8$

$$/_1 + /_2 + \ldots\ldots + /_3 = 5.83+2.00+0.17=8$$

The proportion of total variance accounted for by the first principal component is:

$$\frac{/_1}{/_1 + /_2 + /_3} = \frac{5.83}{8} = 0.73$$

Further, the first two account for a proportion of ——— =0.98 of the population variance. In this case, the components Y_1 and Y_2 could replace the original three variables with little loss of information.

= = = 0.925

= = = -0.998

Notice that the variable X_2 with coefficient -0.924, receives the greatest weight in the component Y_1. It also has the largest correlation (in absolute term) with Y_1.

The correlation between and Y_1 is 0.925, as large as that for X_2 , indicating that the variables are about equally important to the first principal component. The relative sizes of the coefficients of and suggest however, that contribute more to the determination of Y_1 than does . Since, in this case, both coefficients are reasonably large and they have opposite sign. Both variables help in the interpretation of Y_1.

= = 0, = = =1

The remaining correlation can be neglected, since the third component is unimportant.

Principal variables obtained from standard variables

Principal components may be obtained for the standard variables:

=

=

.

$$Z_p = \frac{(X_p - \mu_p)}{\sqrt{s_{pp}}}$$

In matrix notation:

= , where is the diagonal standard deviation matrix

=

The principal component of Z may be obtained from the eigen vectors of the correlation matrix of X.

The i^{th} principal component of the standaridized variables:

With Cov (Z) = is given by:

, i=1,2,.....,p

Moreover, = =p

In this case, are the eigen value ïeigen vector pairs for , with $\lambda_1 \geq \lambda_2 \geq ..... \geq \lambda_p \geq 0$.

The proportion of (standardized) population variance == , k=1,2,....,p

due to k^{th} principal component

Note that principal components obtained from covariance & correlation matrix are different.

Example:

$$= \begin{pmatrix} 1 & 4 \\ 4 & 100 \end{pmatrix}$$

The desired correlation matrix is :

$$= \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$$

Let work how the correlation matrix is obtained:

$$V = \begin{pmatrix} 1 & 0 \\ 0 & 100 \end{pmatrix},$$

=

=

$$= \begin{pmatrix} 1 & 4 \\ 4 & 100 \end{pmatrix} \quad = \quad =$$

Determining eigen value ïeigen vectors from are:

$/_1 = 100.16, \quad e_1^{'} =$

$/_2 = 0.84, \quad e_2^{'} =$

Eigen value ïeigen vectors from are:

$/_1 = 1+ \quad = 1.4, \quad e_1^{'} = [0.707, \ 0.707]$

$/_2 = 1- r_{21} = 0.6, \quad e_2^{'} = [0.707, \ -0.707]$

The respective principal components become:

: = $= 0.040X_1+0.999X_2$

= $= 0.999X_1 - 0.040X_2$

: = $= 0.707Z_1+0.707Z_2$

= +

=

$= [0.707, -0.707]$ $= 0.707Z_1 - 0.707Z_2$

= -

=

Because of its large variance, X_2 completely determines the first principal component deterimned from . Moreover, this first principal component explains a proportion:

= =0.992 of the total population variance.

When the X_1 and X_2 are standardized, however the resulting variables contribute equally to the principal components determined from .

$r_{Y_1,Z_1} = e_{11}\sqrt{l_1}$ $= 0.707\sqrt{1.4} = 0.837$

$r_{Y_1,Z_2} = e_{12}\sqrt{l_1}$

In this case, the first principal component explains a proportion:

= =0.7 of the total (standardized population varince).

Principal components for covariance matrices with special structurtes

There are certain patterned covariance and correlation matrices whose components can be expressed in simple forms. Suppose is a diagonal matrix:

=

Setting =

Also standardization has nothing to do with this special structure. Another patterned covariance matrix, which often describes the correspondence among certain biological variables such as the size of living things, has the general form:

ä =

The resulting correlation matrix is:

=

From the matrix we observe that the variables are equally correlated. When is positive, the largest eigen value is:

= 1+ with associated eigen vector:

=

The remaining (p-1) eigen values are:

=1 - r and the choice for eigen vector is:

$$e_2' = \left[\frac{1}{\sqrt{1x2}}, \frac{1}{\sqrt{1x2}}, 0, \ldots\ldots 0\right]$$

=

.

=

.

.

.

=

The first principal component is:

is the proportion of the original variables. It might be regarded as an index with

equal weights. This principal component explains a proportion:

$\frac{l_1}{p} = \frac{1+(p-1)r}{p} = r + \frac{1-r}{p}$ of the total population variation. We see that $\frac{l_1}{p} @$ for close to 1

or p large. For example, if =0.8 and p=5, thefirst component explains 84% of the total variance. When

is near 1, the last (p-1) components collectively contribute very little to the total variance can often be neglected.

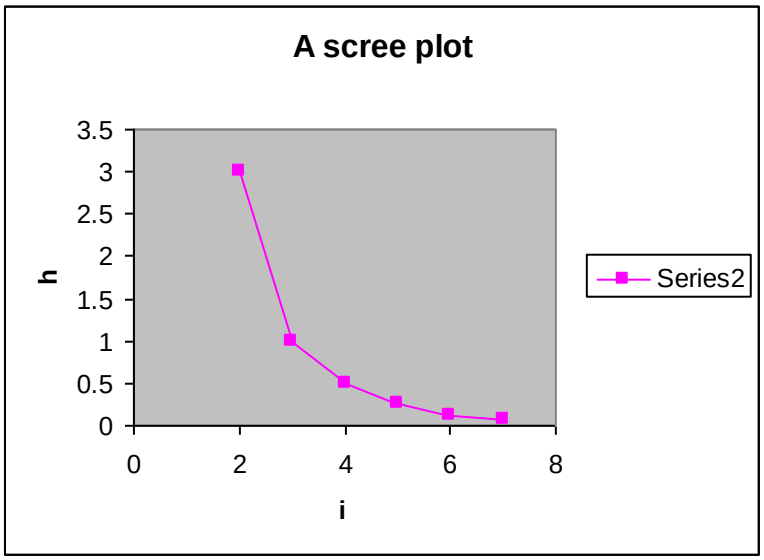
Summaarizing sample variation by principal components

Suppose the data $X_1, X_2, ..., X_n$ represent n- independent drawings from some p-dimensional population with mean vector $m$and covariance matrix ä . These data yield the sample mean vector $\overline{X}$, the sample covarioance matrix S and the sample correlation matrix R. Our efthe sample covarioance matrix S and

the sample. The uncorrelated combinations with the largest variances will be called the sample principal

I	$/\check{E}_i$
1	3
2	1
3	0.5
4	0.25
5	0.125
6	0.0675

An elbow occurs at about i=3, eigen values after $/\check{E}_2$ are relatively small and are about the same size.



A scree plot of $/\check{E}_i$

In this case, it appears with out any other evidence, that two (or perhaps three) sample principal components effectively summarize the total sample variance.

Standardizing the sample principal components

In general, the sample principal components are invariant with respect to changes in scale.

$$Z_j = D^{\frac{-1}{2}}\left(X_j - \overline{X}\right) = \begin{bmatrix} \frac{X_{j1} - \overline{X}_1}{\sqrt{S_{11}}} \\ \frac{X_{j2} - \overline{X}_2}{\sqrt{S_{22}}} \\ \\ \\ \\ \\ \frac{X_{jp} - \overline{X}_p}{\sqrt{S_{pp}}} \end{bmatrix} \quad j = 1, 2, \ldots, n$$

The nxp data matrix of standardized observations:

$$Z = \begin{bmatrix} Z_1' \\ Z_2' \\ \\ \\ \\ Z_n' \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} & \ldots & Z_{1p} \\ Z_{21} & Z_{22} & \ldots & Z_{2p} \\ & & & \\ & & & \\ & & & \\ Z_{n1} & Z_{n2} & \ldots & Z_{np} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{X_{11} - \overline{X}_1}{\sqrt{S_{11}}} & \frac{X_{12} - \overline{X}_1}{\sqrt{S_{11}}} & \ldots & \frac{X_{1p} - \overline{X}_1}{\sqrt{S_{11}}} \\ \frac{X_{21} - \overline{X}_2}{\sqrt{S_{22}}} & \frac{X_{22} - \overline{X}_2}{\sqrt{S_{22}}} & \ldots & \frac{X_{2p} - \overline{X}_2}{\sqrt{S_{22}}} \\ & & & \\ & & & \\ & & & \\ \frac{X_{n1} - \overline{X}_n}{\sqrt{S_{nn}}} & \frac{X_{n2} - \overline{X}_n}{\sqrt{S_{nn}}} & \ldots & \frac{X_{np} - \overline{X}_n}{\sqrt{S_{nn}}} \end{bmatrix}$$

Yields the sample mean vector:

$\bar{Z}$ = $\frac{1}{n}\left(1' Z\right)'$ = = == 0

The sample covariance matrix:

= =

=

$$R = \begin{pmatrix} 1 & r_{12} \ldots\ldots r_{1p} \\ r_{21} & 1 \ldots\ldots r_{2p} \\ & \\ & \\ & \\ & \\ & \\ r_{p1} & r_{p2} \ldots\ldots 1 \end{pmatrix}$$

If $Z_1, Z_2, ...Z_n$ are standardized observations with covariance matrix R, the i^{th} principal component is given as:

$$= e_{i1}Z_1 + e_{i2}Z_2 + ... + e_{ip}Z_p \ , \qquad i = 1 \ , \ 2 \ , \ é \ . \ , \ p$$

Where are the eigen value ïeigen vector pairs with .

Sample variance: $= /_k$, k = 1 , 2 , é . , p

Sample covariance, $Cov(y_i, y_k) = 0$, for i¸ k

Total (standardized) sample variance =tr®= p= $/_1 + /_2 + ... + /_p$

$$r_{y_i, x_k} = e_{ik}\sqrt{/_i} \ , \qquad i \ , \ k = 1 \ , \ 2 \ , \ é \ , \ p$$

Total sample variance explained by principal component is:

$$\begin{pmatrix} \text{Proportion of sample variance} \\ \text{due to } i^{th} \text{ principal component} \end{pmatrix} = \frac{/_i}{p} \ , \qquad i = 1 \ , \ 2 \ , \ é \ , \ p$$

A rule of thumb suggests retaining only those components whose variances $/_i$ are greater than unity, or equivalently, only those components which individually explain at least a proportion $\frac{1}{p}$ of the totla variance. This rule does not have a great deal of theoretical support, however, and it should not be applied blindly. A scree plot is also useful for selecting the appropriate number of components.

Factor analysis

The essential purpose of factor analysis is to describe, if possible, the covariance relationship among many variables in terms of a few underlying, but unobservable, random quantities called factors.

Factor analysis can be considered an extension of principal component analysis. Both can be viewed as attepts to approximate the covariance matrix. However, the approximation based on the factor model is more elaborate.

The orthogonal factor model

The observable random vector X with p- components has mean μ and covariance matrix Σ . The factor model postulates that X is linearly dependent upon a few observable random variables $F_1, F_2, ..., F_m$, called common factors, and p-additional sources of variation $\varepsilon_1, \varepsilon_2, ..., \varepsilon_p$, called errors or some times called specific factors.

$$X_1 - \mu_1 = l_{11}F_1 + l_{12}F_2 + ... + l_{1m}F_m + \varepsilon_1$$

$$X_2 - \mu_2 = l_{21}F_1 + l_{22}F_2 + ... + l_{2m}F_m + \varepsilon_2$$

.

.

.

$$X_p - \mu_p = l_{p1}F_1 + l_{p2}F_2 + ... + l_{pm}F_m + \varepsilon_p$$

Or in matrix notation:

(mx1)

$$X - \mu = LF + \varepsilon \qquad \text{(Orthogonal factor model)}$$

(px1) (pxm) (px1)

The coefficient l_{ij} is called the loading of the i^{th} variable on the j^{th} factor, so the matrix L is the matrix of factor loadings. Note that the i^{th} specific factor ε_i is associated only with the i^{th} response X_i.

Assumptions

1. E[F]=0
2. $Cov(F) = E\lfloor FF' \rfloor = I$

3. $E(e) = 0$, $Cov(e, e') = \Psi_{pxp}$ diagonal matrix

4. F and are independent

5.

The model $X - \mu = LF + e$ with the above assumptions constitute the orthogonal factor model.

The covariance structure for the orthogonal factor model:

1. or

=

2. or

The model $X - \mu = LF + e$ is linear in common factors. The proportion of variance of the i^{th} variable contributed by the m-common factor is called the i^{th} communality. The portion of variance of Var(Xi)= due to the specific factor is often called the uniqueness or the specific variance, the i^{th} communality is denoted by .

$$= l_{i1}^2 + l_{i2}^2 + \ldots + l_{im}^2 + \Psi_i$$

$$= l_{i1}^2 + l_{i2}^2 + \ldots + l_{im}^2$$

The i^{th} communality is the sum of squares of the loadings of the i^{th} variable on the m-common factors.

Example:

$$= \begin{bmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{bmatrix}$$

= +

=

m=2,

= , = =

The communality of X1 is:

= $= 4^2 + 1^2 = 17$ and the variance of X1 can be decomposed as:

= = 17 + 2= 19

= $= 7^2 + 2^2 + 4 = 57$

A similar break down occurs for the other variables. The factor model assumes that the

variances and covariances for X reproduced from the p.m factor loadings l_{ij} and p- specific variances .

When m=p, any covariance can be reproduced exactly as , so

=

1=

0.9=

0.7=

1=

0.4=

1=

The pairs of equations:

0.7=

0.4=

0.9= 0.9= = = 1.255

Since the var(F1)= 1(by assumption) and var(X1)=1

= Cov(X1,F1)=Corr(X1,F1). Now a correlation coefficient can not be graeter than unity(in absolute value), so from this point of view = 1.25 is too large. Also the equation,

1= or which is un satisfactory, since it gives a negative value for $Var(e_1) = y_1$. Thus for this example with m=1, it is possible to get a unique numerical solutuion to the equation $= LL' + y$. However, the solution is not consistent, with the statistical interpretation of the coefficients, so it is not proper solution.

When m>1, there is always some inherent amguity associated with the factor model. To see this, let T be any mxm orthogonal matrix, so that , then

$X - m = LF + e =$ =

Where and

=0

It is impossible, on the basis of observations on X, to distinguish the loading L from the loadings . That the factors F and $F^* = T'F$ have the same statistical properties, eventhough the loadings are in general, different from the loadings L, they both generate the same covariance matrix . That is,

$= LL' + y \quad = LTT'L + y = L^*\left(L^*\right)' + y$

This ambiguity provides the rationale for "factor rotation", since orthogonal matrices correspond to rotations (and reflections) of the coordinate system for X.

Methods of estimation

The sample covariance matrix S is an estimator of the unknown population covariance matrix . If the off-diagonal elements of S are small or those of the sample correlation matrix R is essentially zero, the variables are not related and a factor rdinate system for X.

Let have eigen value ïeigen vector pairs with . Then

=

=

This fits

$$= LL' \qquad \text{(mxp)}$$

(pxm)

This representation assumes the specific factor, e's are of minor importance and can also be neglected in factoring of . If specific factors are included in the model, their variance may be taken to be the diagonal elements.

$$= LL' + \psi = \begin{bmatrix} \sqrt{l_1}e_1 & \sqrt{l_2}e_2 & \ldots\ldots & \sqrt{l_m}e_m \end{bmatrix} \begin{bmatrix} \sqrt{l_1}e_1 \\ \sqrt{l_2}e_2 \\ \\ \\ \\ \\ \\ \sqrt{l_m}e_m \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & \ldots\ldots & 0 \\ 0 & \psi_2 & 0 & \ldots\ldots 0 \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ 0 & 0 & \ldots\ldots & \psi_p \end{bmatrix}$$

Where $\psi_i = s_{ii} - \sum_{j=1}^{m} l_{ij}^2$, i = 1 , 2 , é . , p

To applyb this approach to a data set $X_1, X_2, \ldots . X_n$, it is customary first to center the observations by subtracting the sample mean $\bar{X}$. The centered observations are:

$$X_j - \bar{X} = \begin{bmatrix} X_{j1} \\ X_{j2} \\ \\ \\ \\ \\ \\ X_{jp} \end{bmatrix} - \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \\ \\ \\ \\ \\ \bar{X}_p \end{bmatrix} = \begin{bmatrix} X_{j1} - \bar{X}_1 \\ X_{j2} - \bar{X}_2 \\ \\ \\ \\ \\ \\ X_{jp} - \bar{X}_p \end{bmatrix} \qquad j = 1 , 2 , é . . , n$$

Have the same sample covariance as the original observations. In case where the units of variables are not commensurate, it is usually desirable to work with standardized variables.

$$Z_j = \begin{bmatrix} \frac{X_{j1} - \bar{X}_1}{\sqrt{S_{11}}} \\ \frac{X_{j2} - \bar{X}_2}{\sqrt{S_{22}}} \\ \\ \\ \\ \\ \\ \\ \frac{X_{jp} - \bar{X}_p}{\sqrt{S_{pp}}} \end{bmatrix} \quad j = 1, 2, \ldots, n$$

Whose sample covariance matrix is the sample correlation matrix R of the observations $X_1, X_2, \ldots . X_n$. Standardization avoids the problem of having one variable with large variance unduly influencing the determination factor loadings. The application of the above method to the sample covariance matrix or correlation matrix is known as p[rincipal component solution.

The principal componet factor analysis of the sample covariance matrix S is specified interms of its eigen value -eigen vector pairs $\left(\lambda_1, e_1\right), \left(\lambda_2, e_2\right), \ldots\ldots, \left(\lambda_p, e_p\right)$, where .

Let m < p be the number of common factors, then the matrix of estimated factor loadings $\left(\tilde{l}_{ij}\right)$ is given by:

$$\tilde{L} = \left[\sqrt{\lambda_1}e_1, \sqrt{\lambda_2}e_2, \ldots\ldots, \sqrt{\lambda_m}e_m\right]$$

The estimated specific variances are provided by the diagonal elements of the matrix S-LL', so:

$$\tilde{y} = \begin{pmatrix} \tilde{y}_1 & 0 & \ldots & 0 \\ 0 & \tilde{y}_2 & 0 & \ldots & 0 \\ & & & & \\ 0 & 0 & \ldots & 0 & \tilde{y}_p \end{pmatrix} \text{ with } \tilde{y}_i = S_{ii} - \sum_{j=1}^{m} l_{ij}^2$$

Communalities are estimated as:

$\tilde{h}_i^2 =$

The principal component factor analysis of the sample correlation matrix is obtained by starting with R in place of S. The residual matrix:

resulting from the approximation of S by the principal component solution. The diagonal elements are zero, and if the other elements are also small we may subjectively take the m-factor model be appropriate. Analyticallty,

sum of squares of entries .

Consequently, a small value for the sum of squares of the neglected eigen values implies a small value for the sum of squared errors of approximation.

Ideally, the contributions of the first few factors to the sample variances of the variables should be large. The contribution to sample variance Sii afrom the first common factor is .

The contribution to the total sample variance,

, from the first commen factor is then,

= =

Since the eigen vector has a unit length,