CHAPTER 5: PRINCIPAL COMPONENTS ANALYSIS

Data used in Chapter 5 are provided in ASCII (*.txt) and, where the raw data (rather than a correlation matrix) are available, SPSS v.10 format (*.sav). The SPSS syntax for carrying out the analysis presented in the book is also given (*.sps) together with the output file (*.spo). SPSS output files (*.spo) can only be viewed in SPSS, but output files are also provided in pdf format.

Note that the first line of each syntax file will need to be edited to give the correct location of the data file.

subject.txt

The data are pairwise correlation coefficients between subject scores for a sample of 220 boys. The file *subject.txt* contains the lower triangle of the correlation matrix which is given in Table 5.2. Syntax for reading the correlation matrix into SPSS and carrying out a PCA is given in *subject.sps*.

The subject scores are in the following order: Gaelic, English, History, Arithmetic, Algebra, Geometry.

eurojob.txt / eurojob.sav

The data are the percentage of the workforce employed in each of 9 industries for 26 European countries in 1979.

The data file contains 10 variables:

COUNTRY	Country label (a string variable)
AGRIC	% employed in agriculture
MINING	% employed in mining
MANU	% employed in manufacturing
POWER	% employed in power supply industries
CONSTR	% employed in construction
SERVICE	% employed in service industries
FINANCE	% employed in finance
SOCIAL	% employed in social and personal services
TRANS	% employed in transport

econ.txt /econ.sav

The data are the values for five economic and demographic indicators for a sample of 25 countries. The data refer to 1990 and come from the United Nations Statistical Yearbook of 1997. The file contains 6 variables.

COUNTRY	Country
INCREASE	Annual percentage population growth rate
LIFE	Life expectancy in years
IMR	Infant mortality rate per 1000
TFR	Total fertility rate
GDP	Gross Domestic Product per capita in US dollars

socmob.txt

The data are based on information provided by 713 male or female married respondents to a survey carried out in 1949. The variables relate to the respondent, his or his spouse, father, father-in-law, and firstborn son. The file *socmob.txt* contains the full correlation matrix, shown in Table 5.14.

The 10 variables are coded as follows:

X1=Husband's father's occupational status
X2=Wife's father's occupational status
X3=Husband's further education
X4=Husband's qualifications
X5=Husband's occupational status
X6=Wife's further education
X7=Wife's qualifications
X8=Firstborn's further education
X9=Firstborn's qualifications
X10=Firstborn's occupational status

educ.txt

The data are 9 measurements made on secondary school girls in 1964 and 1968. The file *educ.txt* contains the full correlation matrix for these 9 variables, shown in Table 2.17.

The variables are coded as follows:

X1=Parental circumstances in 1964
X2=Details of class teacher in 1964
X3=School-parent interaction in 1964
X4=Girl's attitude in 1964
X5=Test score in 1964
X6=Type of school in 1968
X7=Parental circumstances in 1968
X8=School-parent interaction in 1968
X9=Test score in 1968

tv.txt

The data are correlations between liking to watch 10 different television programmes for a sample of 7000 UK adults in the 1970s. The file *tv.txt* contains the full correlation matrix for these 10 variables, shown in Table 5.16.

The 10 variables are coded as follows:

BOX	Boxing
THISWK	This Week
TODAY	Today
SPORT	World of Sport
GNDSTND	Grandstand
LINEUP	Line-Up
MATCH	Match of the Day
PANOR	Panorama
RUGBY	Rugby Special
HRS24	24 Hours