

Data-based graphs: Multivariate graphs

Scatter Plot Matrix

The scatter plot matrix provides an organized way of simultaneously looking at bivariate plots of up to ten variables. Except for the diagonal, each frame contains a scatter plot. The diagonals contain the variable names and the minimum and maximum values of the corresponding variables. Each scatter plot is a visual summary of linearity, nonlinearity, and separated points (see, for example, Cook & Weisberg (1994)).

Two of the many factors that are known to have some influence or relevance to the condition of the human heart are physical fitness and blood cholesterol level. In a related research project, four homogeneous groups of adult males were considered. A number of plasma lipid parameters were measured on each of the 66 individuals and fitness parameters were also measured for three of the four groups. The four groups of patients are:

- Weightlifters (n = 17)
- Students (control group, n = 20)
- Marathon athletes (n = 20)
- Coronary patients (n = 9).

The triglyceride level is represented by the variable Trigl in these data, and Age represents the age of a patient in years. The cholesterol level is represented by the variable Cholest. The group identification of patients is given by the variable Group.

In the scatter plot matrix shown below, the first row of plots at the top of the matrix contains plots of Cholest against Age and Trigl respectively. The third plot in this row contains the variable name and the observed minimum and maximum values of Cholest. The first plot in the second row shows the scatter plot of Trigl against Age. The other scatter plots are essentially mirror images of the three discussed this far and are the plots for (Cholest, Trigl), (Trigl, Age) and (Cholest, Age) respectively. The diagonal elements of the matrix contain the names and the minimum and maximum values of each variable.

From the display, it is apparent that the (Age, Trigl) and (Age, Cholest) scatter plots of coronary patients, denoted by a "+" symbol, are clustered together, away from the main cluster of points formed by the other three groups. In the (Age, Age) segment the minimum and maximum values of the Age variable are given. In general, the coronary patients are older and have higher cholesterol and triglyceride levels than patients in the remaining groups.



Creating a scatter plot matrix

Select **fitchol.ss3** from the **Examples\Primer\Graphics** folder. A scatter plot matrix is obtained by selecting the **Data-based Graphs**, **Multivariate** option from the **File** menu to display the **Multivariate plot** dialog box shown below.

🚰 SuperMix - [fitchol.ss3]				<u>_ ×</u>
📝 File Edit Window Help				_ 8 ×
New Project	Ctrl+N			
Import Data File	Ctrl+I			
⊆lose		eight	(D)_Weight	(E)_%I≜
- New Martin Colum	est. u	179.20	107.10	
, New Model Setup	Ctrl+w	183.00	112.20	
Open Existing Model Setup	Ctrl+E	175.70	78.00	
New Syntax File		182.50	79.70	
Provident Provid		178.00	81.80	
Open Tayl File		169.80	78.00	
		178.60	81,10	
Data-based Graphs	۱.	<u>E</u> xpla	ratory 20	
Open Graph	Ctrl+G	Univa	ariate 80	
		Bivari	iate 30	
1 Save	Ctrl+S	Multiv	variate 80	
1 Save <u>A</u> s		1172.70	04. 10	
1 Evit		172.40	71.40	•
]		

The **Multivariate plot** dialog box is displayed. Select Age, Trigl and Cholest as the **Plot** variables and select Group as the **Mark** variable as shown below. Complete the dialog box as shown and click Plot when done to obtain the graph as described earlier.

tultivariate plot List of Variables Name Group Age Height Weight %Fat Strength Trigl Cholest		Mark		
Only one variable may be si marking. No more than 8 m	elected	for selected	•	
Plot	Cancel			