

The data set (and description) can be downloaded here:

<http://lib.stat.cmu.edu/datasets/biomed.data>

### **Description:**

February 23, 1982

The 1982 annual meetings of the American Statistical Association (ASA) will be held August 16-19, 1982 in Cincinnati. At that meeting, the ASA Committee on Statistical Graphics plans to sponsor an "Exposition of Statistical Graphics Technology." The purpose of this activity is to more fully inform the ASA membership about the capabilities and uses of computer graphics in statistical work. This letter is to invite you to participate in the Exposition.

Attached is a set of biomedical data containing 209 observations (134 for "normals" and 75 for "carriers"). Each vendor or provider of statistical graphics software participating in the Exposition is to analyze these data using their software and to prepare tabular, graphical and text output illustrating the use of graphics in these analyses and summarizing their conclusions. The tabular and graphical materials must be direct computer output from the statistical graphics software; the textual descriptions and summaries need not be. The total display space available to each participant at the meeting will be a standard poster-board (approximately 4' x 2 1/2'). All entries will be displayed in one location at the meetings, together with brief written commentary by the committee summarizing the results of this activity.

### Reference

Exposition of Statistical Graphics Technology,  
L. H. Cox, M. M. Johnson, K. Kafadar,  
ASA Proc Stat. Comp Section, 1982, pp 55-56.  
Enclosures

### THE DATA

The following data arose in a study to develop screening methods to identify carriers of a rare genetic disorder. Four measurements m1, m2, m3, m4 were made on blood samples. One of these, m1, has been used before.

Because the disease is rare, there are only a few carriers of the disease from whom data are available. The data come in two files, one for normals and one for carriers of the disease. A description of the files is provided. The data have been stripped of the names and other identifiers. Otherwise the data are as received by the analyst.

### PURPOSE OF THE ANALYSIS

The purpose of the analysis is to develop a screening procedure to

detect carriers and to describe its effectiveness. Experts in the field have noted that young people tend to have higher measurements. The laboratory which prepared the measurements is worried that there may be a systematic drift over time in their measurement process. These effects should be considered in the analysis. Can graphical displays show the differences between the distributions of carriers and normals?

## FILE DESCRIPTION

### Column Content

1 Observation number (sequence number per patient)  
    Note that there are several samples per patient  
    for some patients.  
2-8 Blank  
9-12 Hospital identification number for blood sample  
13-18 Blank  
19-20 Age of patient  
21-26 Blank  
27-32 Date that blood sample was taken (mmddy)  
    Note that all day entries are 00.  
33-39 Blank  
40-43 m1 (measurement 1) sss.s  
44-50 Blank  
51-54 m2 (measurement 2) xxx.x Eight missing data points.  
55-61 Blank  
62-65 m3 (measurement 3) xxx.x  
66-72 Blank  
73-75 m4 (measurement 4) xxx Seven missing data points.

### Descriptive statistics:

Dataset= biomed : n= 194 , d= 4

Class1: n= 67

### Covariance matrix:

	[,1]	[,2]	[,3]	[,4]
[1,]	51498.5920	-305.2888	2716.0509	7973.548
[2,]	-305.2888	97.1740	-20.0220	-74.839
[3,]	2716.0509	-20.0220	296.2756	600.102
[4,]	7973.5479	-74.8390	600.1020	5247.087

### Correlation matrix:

	[,1]	[,2]	[,3]	[,4]
[1,]	1.0000	-0.1365	0.6953	0.4851
[2,]	-0.1365	1.0000	-0.1180	-0.1048
[3,]	0.6953	-0.1180	1.0000	0.4813
[4,]	0.4851	-0.1048	0.4813	1.0000

Median: 108.4149 91.6202 19.625 238.5618  
Mean: 185.791 92.9313 23.9328 250.9403  
MCD-estimated:  
MDC-0.975-Mean: 82.2917 93.4854 17.3 234.6667  
MDC-0.750-Mean: 82.2917 93.4854 17.3 234.6667  
MDC-0.500-Mean: 82.2917 93.4854 17.3 234.6667

Class2: n= 127

Covariance matrix:

	[,1]	[,2]	[,3]	[,4]
[1,]	345.8414	-75.8980	8.3771	152.9563
[2,]	-75.8980	149.7853	4.3577	92.4229
[3,]	8.3771	4.3577	19.3066	39.7680
[4,]	152.9563	92.4229	39.7680	1711.3575

Correlation matrix:

	[,1]	[,2]	[,3]	[,4]
[1,]	1.0000	-0.3335	0.1025	0.1988
[2,]	-0.3335	1.0000	0.0810	0.1825
[3,]	0.1025	0.0810	1.0000	0.2188
[4,]	0.1988	0.1825	0.2188	1.0000

Median: 36.9543 82.9641 11.8769 160.7384

Mean: 39.6165 82.3071 12.1457 164.5748

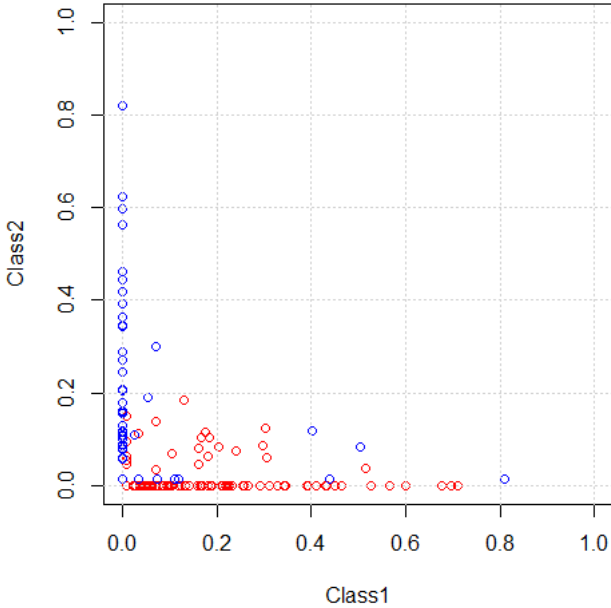
MCD-estimated:

MDC-0.975-Mean: 33.4415 83.8075 12.0377 160.5  
MDC-0.750-Mean: 33.4415 83.8075 12.0377 160.5  
MDC-0.500-Mean: 33.4415 83.8075 12.0377 160.5

Measures:

Mah.Dist: 1.9215  
Mah.Dist-MCD-0.975: 1.8959  
Mah.Dist-MCD-0.750: 1.8959  
Mah.Dist-MCD-0.500: 1.8959

**DD-Plot (zonoid): biomed**



**DD-Plot (random Tukey): biomed**

