Preface

This solution manual was prepared as an aid for instructors who will benefit by having solutions available. In addition to providing detailed answers to most of the problems in the book, this manual can help the instructor determine which of the problems are most appropriate for the class.

The vast majority of the problems have been solved with the help of available computer software (SAS, S-Plus, Minitab). A few of the problems have been solved with hand calculators. The reader should keep in mind that round-off errors can occur—particularly in those problems involving long chains of arithmetic calculations.

We would like to take this opportunity to acknowledge the contribution of many students, whose homework formed the basis for many of the solutions. In particular, we would like to thank Jorge Achcar, Sebastiao Amorim, W. K. Cheang, S. S. Cho, S. G. Chow, Charles Fleming, Stu Janis, Richard Jones, Tim Kramer, Dennis Murphy, Rich Raubertas, David Steinberg, T. J. Tien, Steve Verrill, Paul Whitney and Mike Wincek. Dianne Hall compiled most of the material needed to make this current solutions manual consistent with the sixth edition of the book.

The solutions are numbered in the same manner as the exercises in the book. Thus, for example, 9.6 refers to the 6th exercise of chapter 9.

We hope this manual is a useful aid for adopters of our *Applied Multivariate Statistical Analysis*, 6th edition, text. The authors have taken a little more active role in the preparation of the current solutions manual. However, it is inevitable that an error or two has slipped through so please bring remaining errors to our attention. Also, comments and suggestions are always welcome.

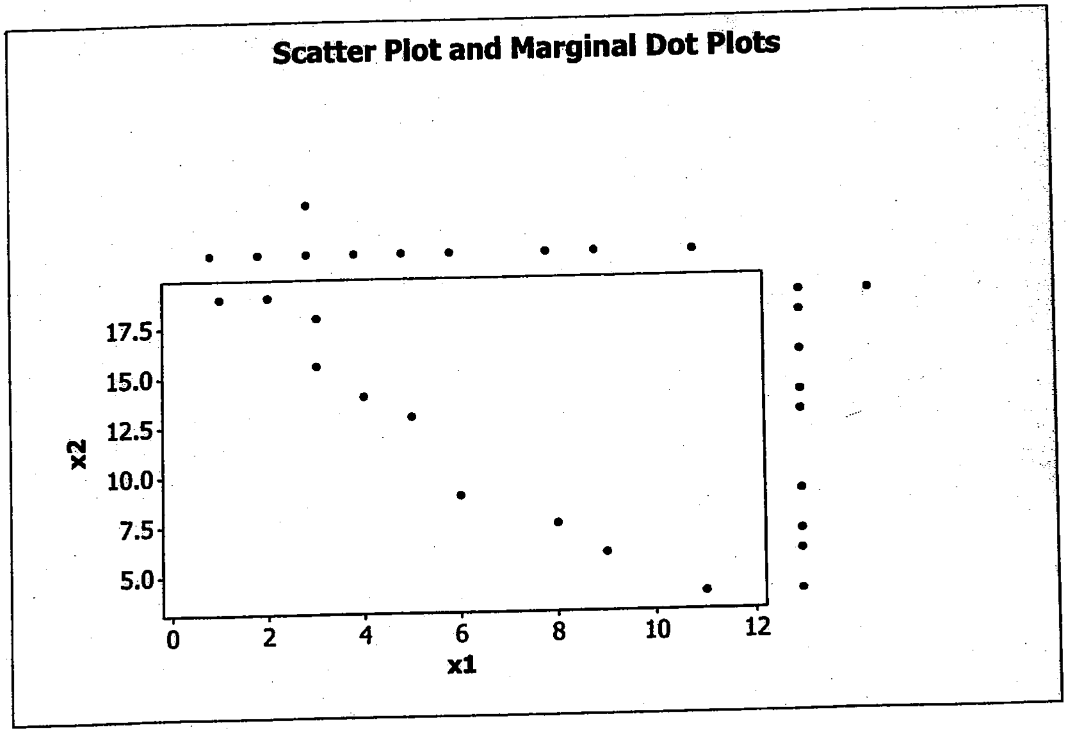
Richard A. Johnson
Dean W. Wichern

Chapter 1

1.1 $\bar{x}_1 = 4.29$ $\bar{x}_2 = 6.29$

$s_{11} = 4.20$ $s_{22} = 3.56$ $s_{12} = 3.70$

1.2 a)



b) s_{12} is negative

c)

$\bar{x}_1 = 5.20$ $\bar{x}_2 = 12.48$ $s_{11} = 3.09$ $s_{22} = 5.27$

$s_{12} = -15.94$ $r_{12} = -.98$

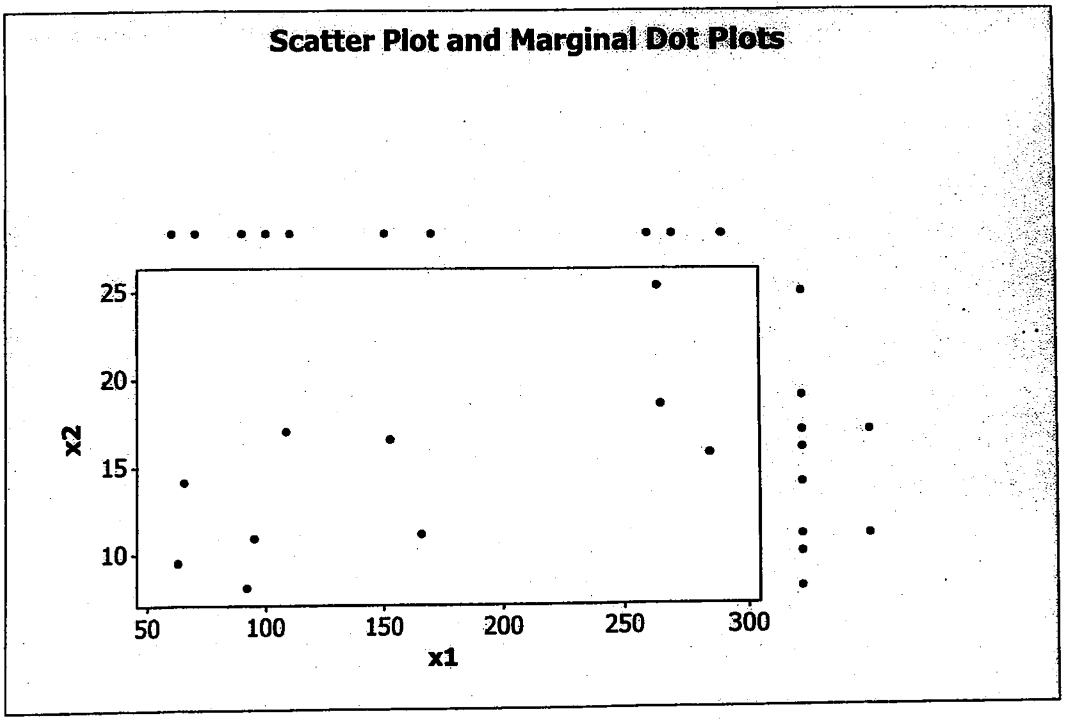
Large x_1 occurs with small x_2 and vice versa.

d)

$$\bar{x} = \begin{bmatrix} 5.20 \\ 12.48 \end{bmatrix} \quad S_n = \begin{bmatrix} 3.09 & -15.94 \\ -15.94 & 5.27 \end{bmatrix} \quad R = \begin{bmatrix} 1 & -.98 \\ -.98 & 1 \end{bmatrix}$$

1.3 $\bar{\underset{\sim}{x}} = \begin{bmatrix} 6 \\ 8 \\ 2 \end{bmatrix}$ $S_n = \begin{bmatrix} 6 & 4 & -1.4 \\ & 8 & 1.2 \\ \text{(symmetric)} & & 2 \end{bmatrix}$ $R = \begin{bmatrix} 1 & .577 & -.404 \\ & 1 & .300 \\ \text{(symmetric)} & & 1 \end{bmatrix}$

1.4 a) There is a positive correlation between x_1 and x_2. Since sample size is small, hard to be definitive about nature of marginal distributions. However, marginal distribution of x_1 appears to be skewed to the right. The marginal distribution of x_2 seems reasonably symmetric.



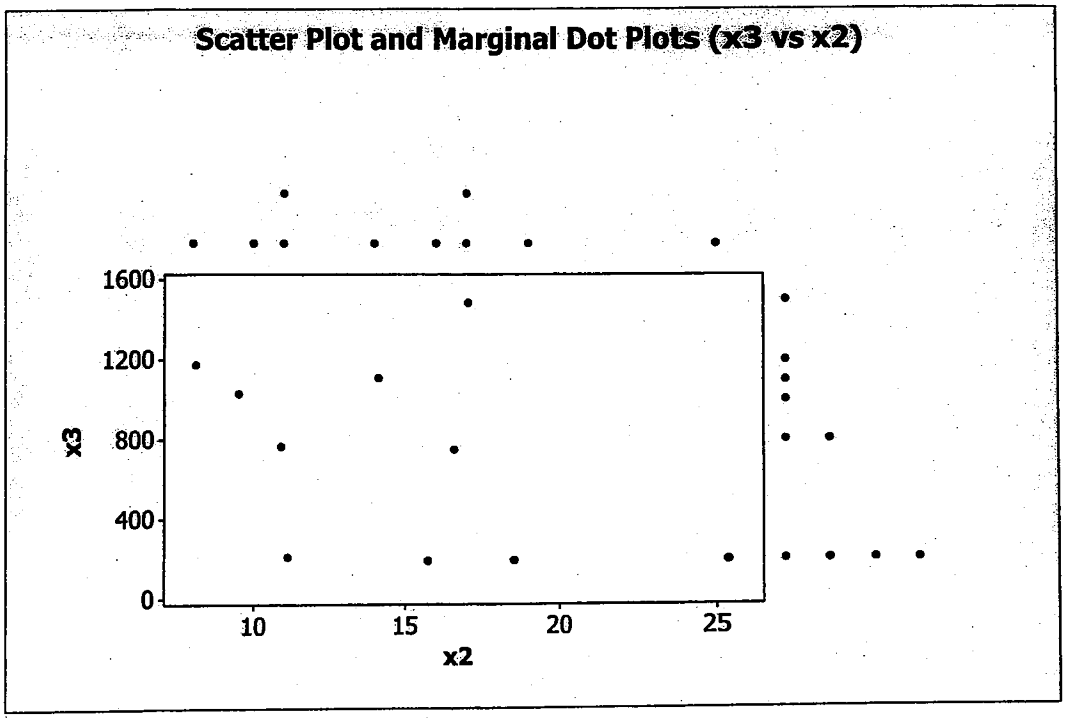
b)

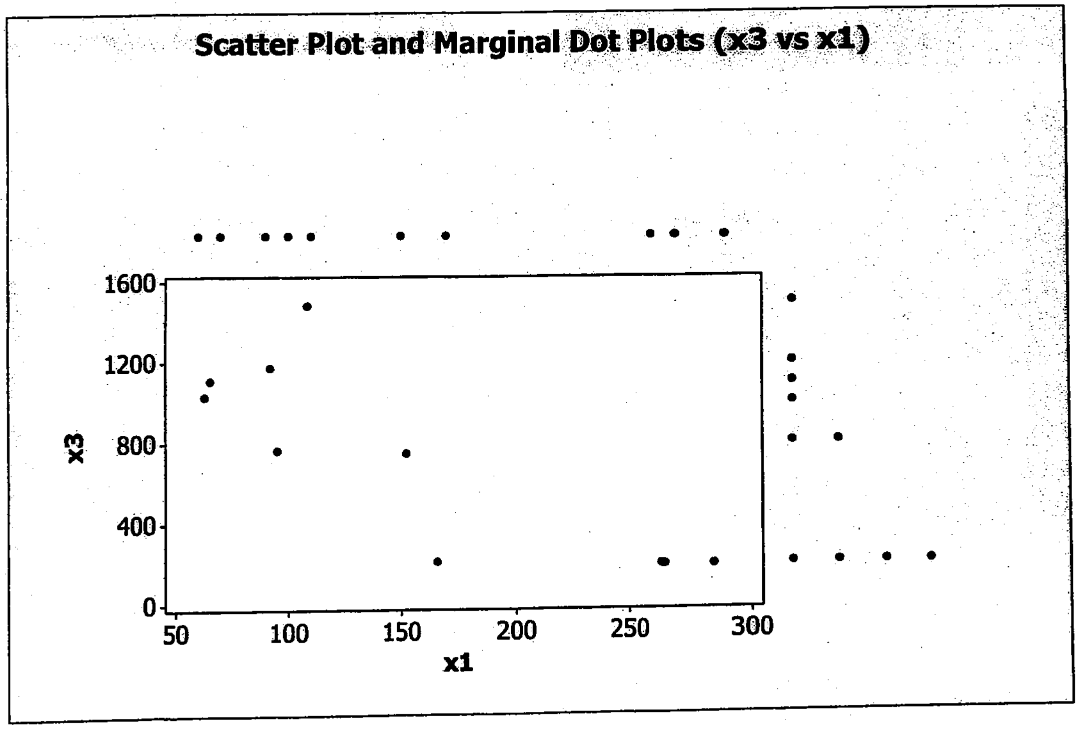
$\bar{x}_1 = 155.60$ $\bar{x}_2 = 14.70$ $s_{11} = 82.03$ $s_{22} = 4.85$

$s_{12} = 273.26$ $r_{12} = .69$

Large profits (x_2) tend to be associated with large sales (x_1); small profits with small sales.

1.5 a) There is negative correlation between x_2 and x_3 and negative correlation between x_1 and x_3. The marginal distribution of x_1 appears to be skewed to the right. The marginal distribution of x_2 seems reasonably symmetric. The marginal distribution of x_3 also appears to be skewed to the right.





1.5 b)

$$\bar{x} = \begin{bmatrix} 155.60 \\ 14.70 \\ 710.91 \end{bmatrix} \qquad S_n = \begin{bmatrix} 82.03 & 273.26 & -32018.36 \\ 273.26 & 4.85 & -948.45 \\ -32018.36 & -948.45 & 461.90 \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & .69 & -.85 \\ .69 & 1 & -.42 \\ -.85 & -.42 & 1 \end{bmatrix}$$

1.6 a) Histograms

X_1

```
MIDDLE OF    NUMBER OF
INTERVAL     OBSERVATIONS
     5.        5     *****
     6.        8     ********
     7.        7     *******
     8.       11     ***********
     9.        5     *****
    10.        6     ******
```

X_2

```
MIDDLE OF    NUMBER OF
INTERVAL     OBSERVATIONS
    30.        1    *
    40.        3    ***
    50.        2    **
    60.        3    ***
    70.       10    **********
    80.       12    ************
    90.        8    ********
   100.        2    **
   110.        1    *
```

X_3

```
MIDDLE OF    NUMBER OF
INTERVAL     OBSERVATIONS
     2.        1    *
     3.        5    *****
     4.       19    *******************
     5.        9    *********
     6.        3    ***
     7.        5    *****
```

X_4

```
MIDDLE OF    NUMBER OF
INTERVAL     OBSERVATIONS
     1.       13     *************
     2.       15     ***************
     3.        8     ********
     4.        5     *****
     5.        1     *
```

X_5

```
MIDDLE OF    NUMBER OF
INTERVAL     OBSERVATIONS
     5.        2     **
     6.        3     ***
     7.        5     *****
     8.        5     *****
     9.        6     ******
    10.        4     ****
    11.        4     ****
    12.        5     *****
    13.        4     ****
    14.        1     *
    15.        0
    16.        1     *
    17.        0
    18.        1     *
    19.        0
    20.        0
    21.        1     *
```

X_6

```
MIDDLE OF    NUMBER OF
INTERVAL     OBSERVATIONS
     2.        3     ***
     4.        4     ****
     6.        7     *******
     8.        7     *******
    10.        8     ********
    12.        5     *****
    14.        2     **
    16.        2     **
    18.        1     *
    20.        0
    22.        0
    24.        2     **
    26.        1     *
```

X_7

```
MIDDLE OF    NUMBER OF
INTERVAL     OBSERVATIONS
     2.        7     *******
     3.       25     *************************
     4.        9     *********
     5.        1     *
```

1.6 b)

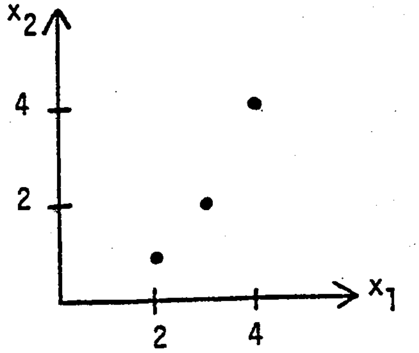
$$\bar{\mathbf{x}} = \begin{bmatrix} 7.5 \\ 73.857 \\ 4.548 \\ 2.191 \\ 10.048 \\ 9.405 \\ 3.095 \end{bmatrix} \quad S_n = \begin{bmatrix} 2.440 & -2.714 & -.369 & -.452 & -.571 & -2.179 & .167 \\ & 293.360 & 3.816 & -1.354 & 6.602 & 30.058 & .609 \\ & & 1.486 & .658 & 2.260 & 2.755 & .138 \\ & & & 1.154 & 1.062 & -.791 & .172 \\ & & & & 11.093 & 3.052 & 1.019 \\ & & & & & 30.241 & .580 \\ \text{(symmetric)} & & & & & & .467 \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & -.101 & -.194 & -.270 & -.110 & -.254 & .156 \\ & 1 & .183 & -.074 & .116 & .319 & .052 \\ & & 1 & .502 & .557 & .411 & .166 \\ & & & 1 & .297 & -.134 & .235 \\ & & & & 1 & .167 & .448 \\ & & & & & 1 & .154 \\ \text{(symmetric)} & & & & & & 1 \end{bmatrix}$$

The pair x_3, x_4 exhibits a small to moderate positive correlation and so does the pair x_3, x_5. Most of the entries are small.

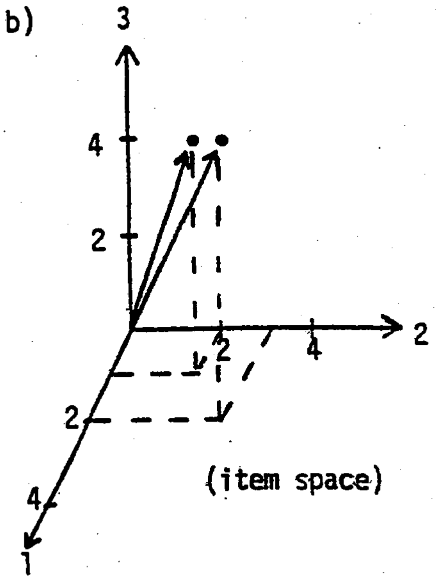
1.7

a)



Scatter plot
(variable space)

b)



(item space)

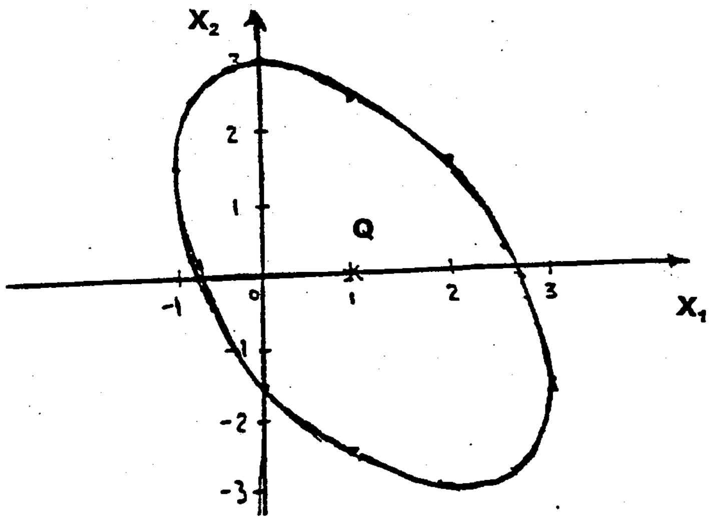
1.8 Using (1-12) $d(P,Q) = \sqrt{(-1-1)^2+(-1-0)^2} = \sqrt{5} = 2.236$

Using (1-20) $d(P,Q) = \sqrt{\frac{1}{3}(-1-1)^2+2(\frac{1}{9})(-1-1)(-1-0)+\frac{4}{27}(-1-0)^2} = \sqrt{\frac{52}{27}} = 1.388$

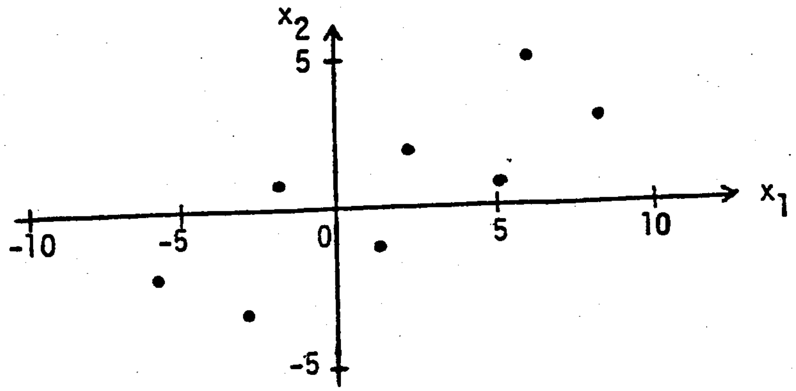
Using (1-20) the locus of points a constant squared distance 1 from Q = (1,0) is given by the expression $\frac{1}{3}(x_1-1)^2+\frac{2}{9}(x_1-1)x_2+\frac{4}{27}x_2^2 = 1$. To sketch the locus of points defined by this equation, we first obtain the coordinates of some points satisfying the equation:

(-1,1.5), (0,-1.5), (0,3), (1,-2.6), (1,2.6), (2,-3), (2,1.5), (3,-1.5)

The resulting ellipse is:



1.9 **a)** $s_{11} = 20.48$ $s_{22} = 6.19$ $s_{12} = 9.09$



1.9 b)

$\tilde{x}_1$	-6.20	-4.10	-1.23	.37	2.73	4.83	7.70	8.43
$\tilde{x}_2$	1.27	-1.10	1.87	-1.37	.73	-1.63	1.33	-1.40

c)

$\tilde{s}_{11} = 24.90$ $\tilde{s}_{22} = 1.77$ (Note $\tilde{s}_{12} = .00$)

d)

$(\tilde{x}_1,\tilde{x}_2) = (2.72, -3.55)$

$d(0,P) = 2.72$ using (1-17).

e) $d(0,P) = 2.72$ using (1-19).

1.10 a) This equation is of the form (1-19) with $a_{11} = 1$, $a_{12} = \frac{1}{2}$ and $a_{22} = 4$. Therefore this is a distance for correlated variables if it is non-negative for all values of x_1, x_2. But this follows easily if we write $x_1^2 + 4x_2^2 + x_1x_2 = (x_1 + \frac{1}{2}x_2)^2 + \frac{15}{4}x_2^2 \geq 0$.

b) In order for this expression to be a distance it has to be non-negative for all values x_1, x_2. Since, for $(x_1,x_2) = (0,1)$ we have $x_1^2-2x_2^2 = -2$, we conclude that this is not a valid distance function.

1.11

$$d(P,Q) = \sqrt{4(x_1-y_1)^2 + 2(-1)(x_1-y_1)(x_2-y_2) + (x_2-y_2)^2}$$

$$= \sqrt{4(y_1-x_1)^2 + 2(-1)(y_1-x_1)(y_2-x_2) + (x_2-y_2)^2} = d(Q,P)$$

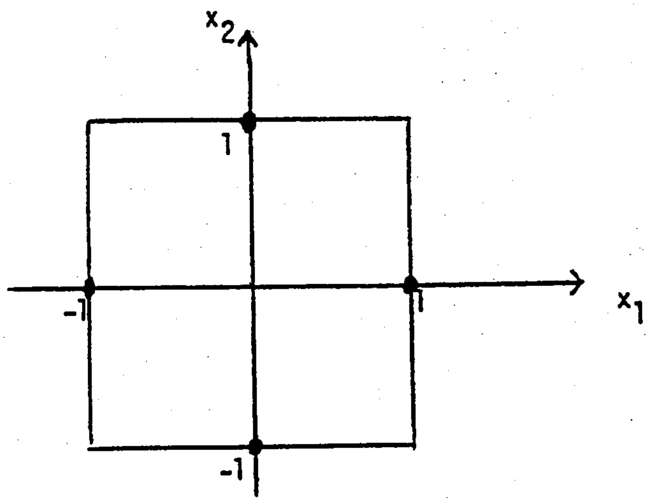
Next, $4(x_1-y_1)^2 - 2(x_1-y_1)(x_2-y_2) + (x_2-y_2)^2 =$

$$= (x_1-y_1-x_2+y_2)^2 + 3(x_1-y_1)^2 \geq 0 \text{ so } d(P,Q) \geq 0.$$

The second term is zero in this last expression only if $x_1 = y_1$ and then the first is zero only if $x_2 = y_2$.

1.12 a) If $P = (-3,4)$ then $d(0,P) = \max(|-3|,|4|) = 4$

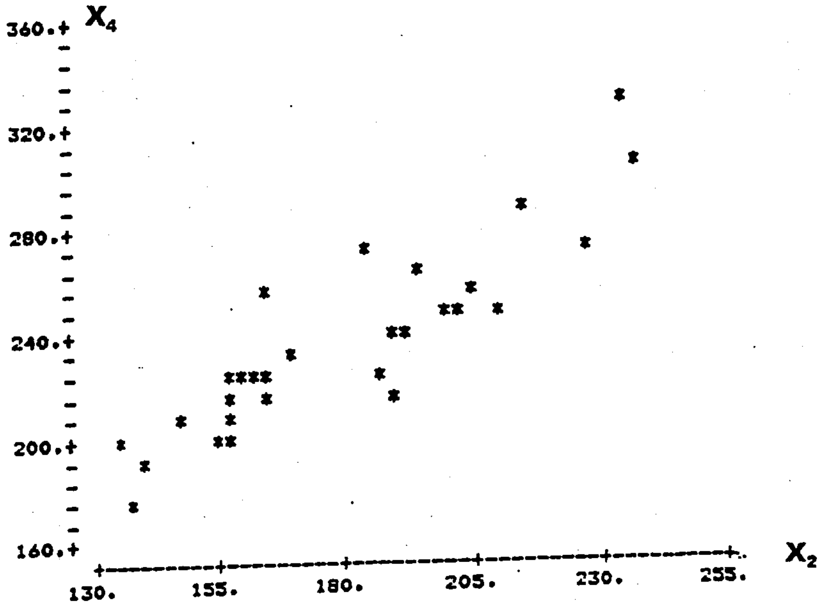
b) The locus of points whose squared distance from (0,0) is 1 is



c) The generalization to p-dimensions is given by $d(0,P) = \max(|x_1|,|x_2|,\ldots,|x_p|)$.

1.13 Place the facility at C-3.

1.14 a)



Strong positive correlation. No obvious "unusual" observations.

b) Multiple-sclerosis group.

$$\underset{\sim}{\bar{x}} = \begin{pmatrix} 42.07 \\ 179.64 \\ 12.31 \\ 236.62 \\ 13.16 \end{pmatrix}$$

$$S_n = \begin{pmatrix} 116.91 & 61.78 & -20.10 & 61.13 & -27.65 \\ & 812.72 & 218.35 & 865.32 & 90.48 \\ & & 305.94 & 221.93 & 286.60 \\ & & & 1146.38 & 82.53 \\ \text{(symmetric)} & & & & 337.80 \end{pmatrix}$$

$$R = \begin{pmatrix} 1 & .200 & -.106 & .167 & -.139 \\ & 1 & .438 & .896 & .173 \\ & & 1 & .375 & .892 \\ & & & 1 & .133 \\ \text{(symmetric)} & & & & 1 \end{pmatrix}$$

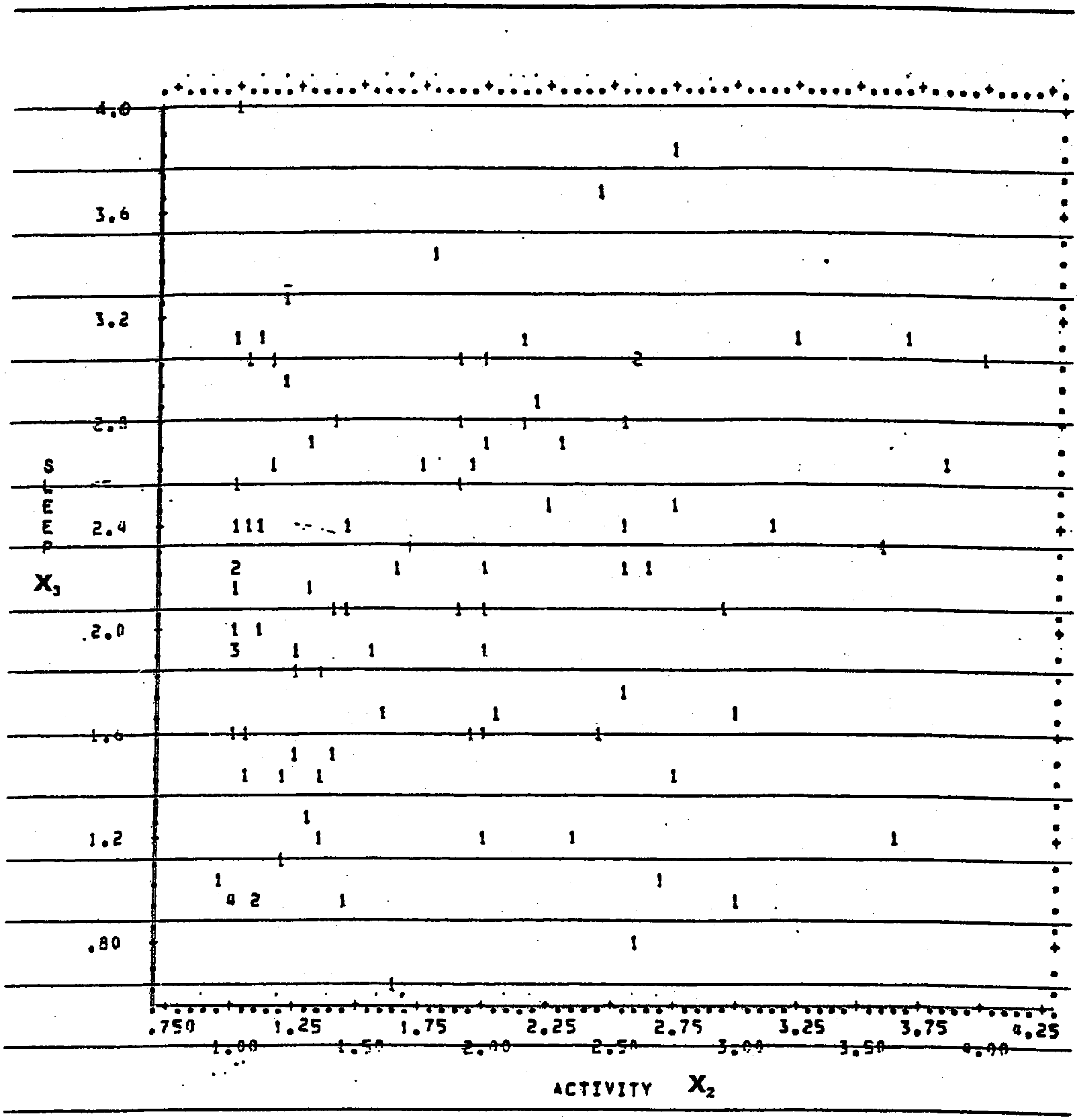
Non multiple-sclerosis group.

$$\underset{\sim}{\bar{x}} = \begin{pmatrix} 37.99 \\ 147.21 \\ 1.56 \\ 195.57 \\ 1.62 \end{pmatrix}$$

$$S_n = \begin{pmatrix} 273.61 & 95.08 & 5.28 & 101.67 & 3.20 \\ & 110.13 & 1.84 & 103.28 & 2.15 \\ & & 1.78 & 2.22 & .49 \\ & & & 183.04 & 2.35 \\ \text{(symmetric)} & & & & 2.32 \end{pmatrix}$$

$$R = \begin{pmatrix} 1 & .548 & .239 & .454 & .127 \\ & 1 & .132 & .727 & .134 \\ & & 1 & .123 & .244 \\ & & & 1 & .114 \\ \text{(symmetric)} & & & & 1 \end{pmatrix}$$

1.15 a) Scatterplot of x_2 and x_3.



b)

$$\bar{\underset{\sim}{x}} = \begin{pmatrix} 3.54 \\ 1.81 \\ 2.14 \\ 2.21 \\ 2.58 \\ 1.27 \end{pmatrix}$$

1.15

$$S_n = \begin{pmatrix} 4.61 & .92 & .58 & .27 & 1.06 & .15 \\ & .61 & .11 & .12 & .39 & -.02 \\ & & .57 & .09 & .34 & .11 \\ & & & .11 & .21 & .02 \\ & & & & .85 & -.01 \\ \text{(symmetric)} & & & & & .85 \end{pmatrix}$$

$$R = \begin{pmatrix} 1 & .551 & .362 & .386 & .537 & .077 \\ & 1 & .187 & .455 & .535 & -.035 \\ & & 1 & .346 & .496 & .156 \\ & & & 1 & .704 & .071 \\ & & & & 1 & -.010 \\ \text{(symmetric)} & & & & & 1 \end{pmatrix}$$

The largest correlation is between appetite and amount of food eaten. Both activity and appetite have moderate positive correlations with symptoms. Also, appetite and activity have a moderate positive correlation.

1.16

There are significant positive correlations among all variables. The lowest correlation is 0.4420 between Dominant humerus and Ulna, and the highest correlation is 0.89365 bewteen Dominant hemerus and Hemerus.

$$\bar{\mathbf{x}} = \begin{pmatrix} 0.8438 \\ 0.8183 \\ 1.7927 \\ 1.7348 \\ 0.7044 \\ 0.6938 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 1.00000 & 0.85181 & 0.69146 & 0.66826 & 0.74369 & 0.67789 \\ 0.85181 & 1.00000 & 0.61192 & 0.74909 & 0.74218 & 0.80980 \\ 0.69146 & 0.61192 & 1.00000 & 0.89365 & 0.55222 & 0.44020 \\ 0.66826 & 0.74909 & 0.89365 & 1.00000 & 0.62555 & 0.61882 \\ 0.74369 & 0.74218 & 0.55222 & 0.62555 & 1.00000 & 0.72889 \\ 0.67789 & 0.80980 & 0.44020 & 0.61882 & 0.72889 & 1.00000 \end{pmatrix},$$

$$\mathbf{S}_n = \begin{pmatrix} 0.0124815 & 0.0099633 & 0.0214560 & 0.0192822 & 0.0087559 & 0.0076395 \\ 0.0099633 & 0.0109612 & 0.0177938 & 0.0202555 & 0.0081886 & 0.0085522 \\ 0.0214560 & 0.0177938 & 0.0771429 & 0.0641052 & 0.0161635 & 0.0123332 \\ 0.0192822 & 0.0202555 & 0.0641052 & 0.0667051 & 0.0170261 & 0.0161219 \\ 0.0087559 & 0.0081886 & 0.0161635 & 0.0170261 & 0.0111057 & 0.0077483 \\ 0.0076395 & 0.0085522 & 0.0123332 & 0.0161219 & 0.0077483 & 0.0101752 \end{pmatrix}.$$

1.17

There are large positive correlations among all variables. Particularly large correlations occur between running events that are "similar", for example, the 100m and 200m dashes, and the 1500m and 3000m runs.

$$\mathbf{x} = \begin{bmatrix} 11.36 \\ 23.12 \\ 51.99 \\ 2.02 \\ 4.19 \\ 9.08 \\ 153.62 \end{bmatrix} \quad \mathbf{S}_n = \begin{bmatrix} .152 & .338 & .875 & .027 & .082 & .230 & 4.254 \\ .338 & .847 & 2.152 & .065 & .199 & .544 & 10.193 \\ .875 & 2.152 & 6.621 & .178 & .500 & 1.400 & 28.368 \\ .027 & .065 & .178 & .007 & .021 & .060 & 1.197 \\ .082 & .199 & .500 & .021 & .073 & .212 & 3.474 \\ .230 & .544 & 1.400 & .060 & .212 & .652 & 10.508 \\ 4.254 & 10.193 & 28.368 & 1.197 & 3.474 & 10.508 & 265.265 \end{bmatrix}$$

$$\mathbf{R} = \begin{bmatrix} 1.000 & .941 & .871 & .809 & .782 & .728 & .669 \\ .941 & 1.000 & .909 & .820 & .801 & .732 & .680 \\ .871 & .909 & 1.000 & .806 & .720 & .674 & .677 \\ .809 & .820 & .806 & 1.000 & .905 & .867 & .854 \\ .782 & .801 & .720 & .905 & 1.000 & .973 & .791 \\ .728 & .732 & .674 & .867 & .973 & 1.000 & .799 \\ .669 & .680 & .677 & .854 & .791 & .799 & 1.000 \end{bmatrix}$$

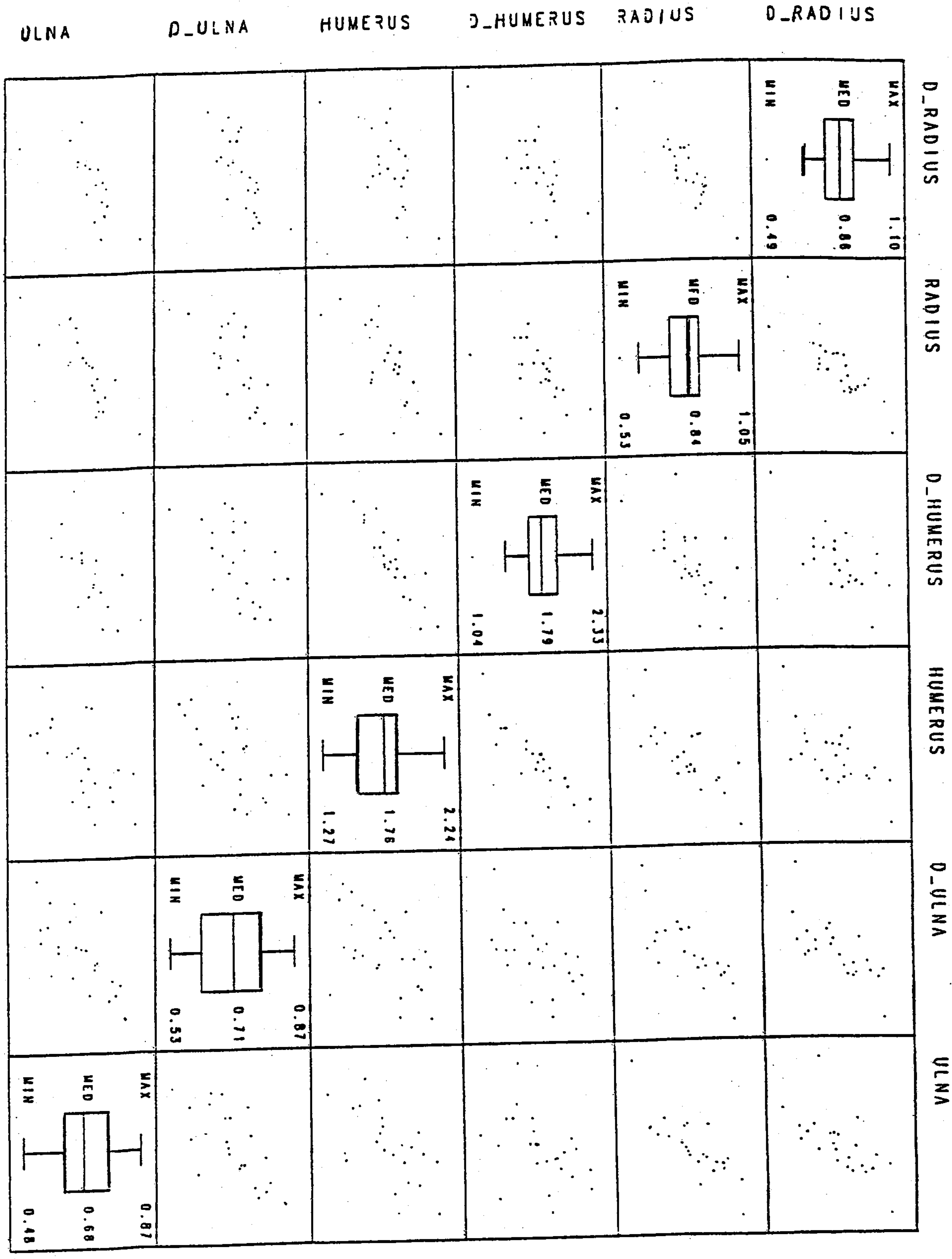
1.18

There are positive correlations among all variables. Notice the correlations decrease as the distances between pairs of running events increase (see the first column of the correlation matrix **R**). The correlation matrix for running events measured in meters per second is very similar to the correlation matrix for the running event times given in Exercise 1.17.

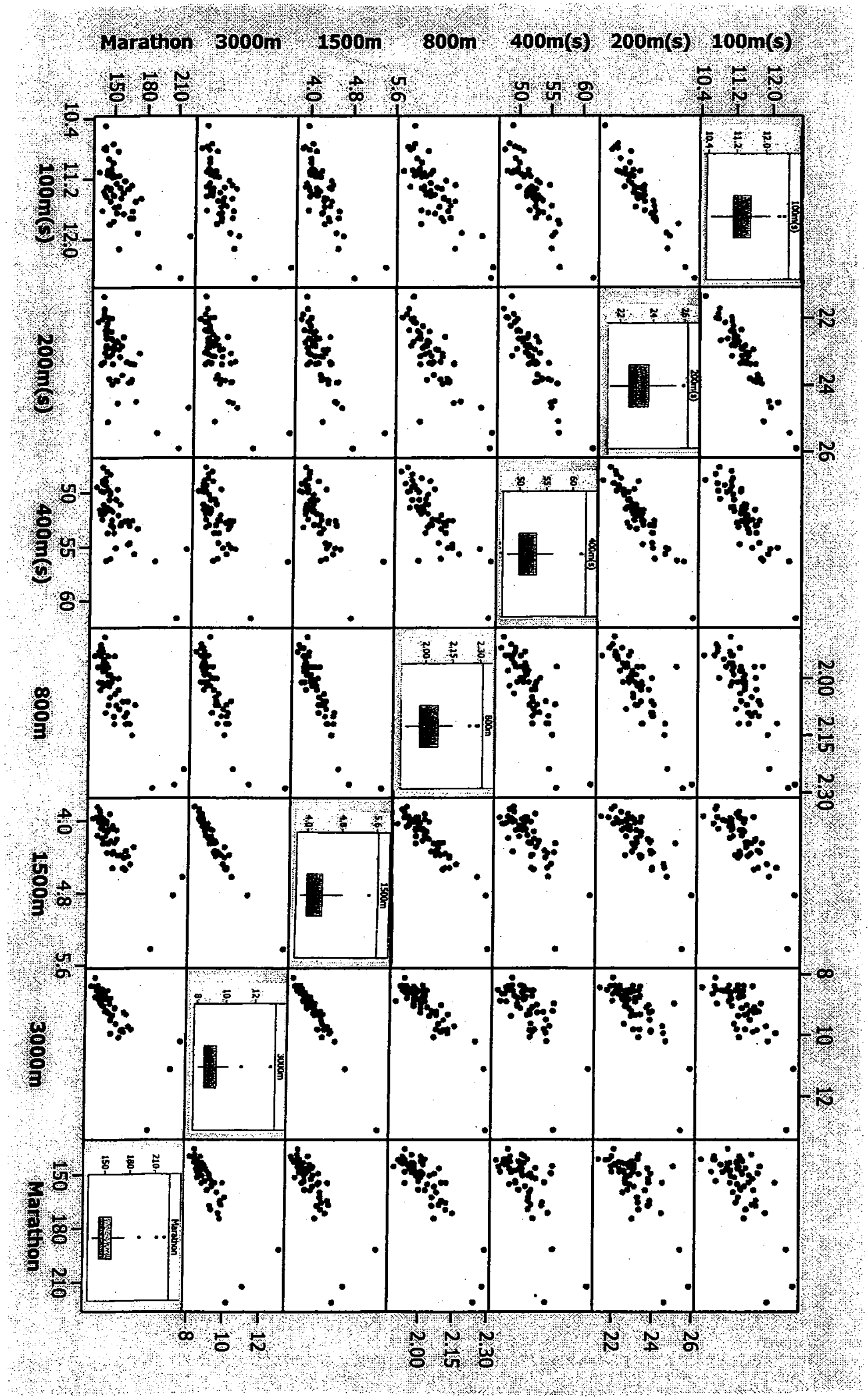
$$\bar{\mathbf{x}} = \begin{bmatrix} 8.81 \\ 8.66 \\ 7.71 \\ 6.60 \\ 5.99 \\ 5.54 \\ 4.62 \end{bmatrix} \qquad \mathbf{S}_n = \begin{bmatrix} .091 & .096 & .097 & .065 & .082 & .092 & .081 \\ .096 & .115 & .114 & .075 & .096 & .105 & .093 \\ .097 & .114 & .138 & .081 & .095 & .108 & .102 \\ .065 & .075 & .081 & .074 & .086 & .100 & .094 \\ .082 & .096 & .095 & .086 & .124 & .144 & .118 \\ .092 & .105 & .108 & .100 & .144 & .177 & .147 \\ .081 & .093 & .102 & .094 & .118 & .147 & .167 \end{bmatrix}$$

$$\mathbf{R} = \begin{bmatrix} 1.000 & .938 & .866 & .797 & .776 & .729 & .660 \\ .938 & 1.000 & .906 & .816 & .806 & .741 & .675 \\ .866 & .906 & 1.000 & .804 & .731 & .694 & .672 \\ .797 & .816 & .804 & 1.000 & .906 & .875 & .852 \\ .776 & .806 & .731 & .906 & 1.000 & .972 & .824 \\ .729 & .741 & .694 & .875 & .972 & 1.000 & .854 \\ .660 & .675 & .672 & .852 & .824 & .854 & 1.000 \end{bmatrix}$$

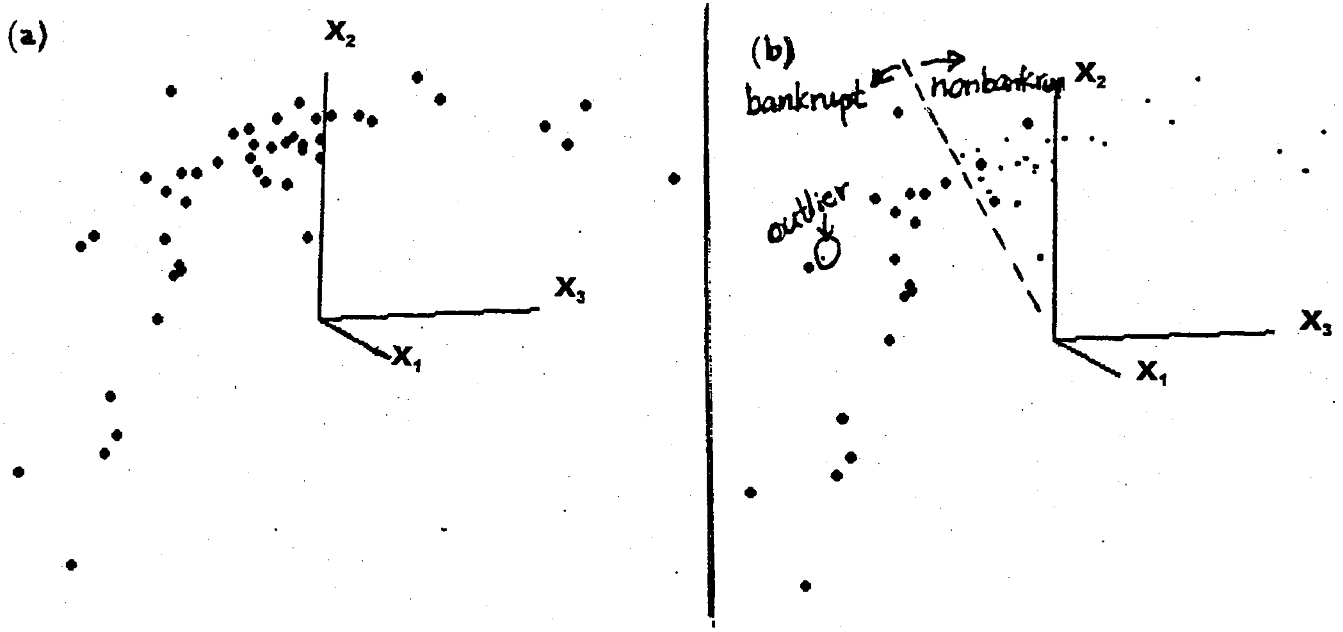
1.19 (a)


ULNA
D_ULNA
HUMERUS
D_HUMERUS
RADIUS
D_RADIUS
D_RADIUS
RADIUS
D_HUMERUS
HUMERUS
D_ULNA
ULNA
MIN
MED
MAX
0.49
0.86
1.10
MIN
MED
MAX
0.53
0.84
1.05
MIN
MED
MAX
1.04
1.79
2.33
MIN
MED
MAX
1.27
1.76
2.24
MIN
MED
MAX
0.53
0.71
0.87
MIN
MED
MAX
0.48
0.68
0.87

1.19 (b)


Marathon
3000m
1500m
800m
400m(s)
200m(s)
100m(s)
100m(s)
200m(s)
400m(s)
800m
1500m
3000m
Marathon
10.4
11.2
12.0
22
24
26
50
55
60
2.00
2.15
2.30
4.0
4.8
5.6
8
10
12
150
180
210

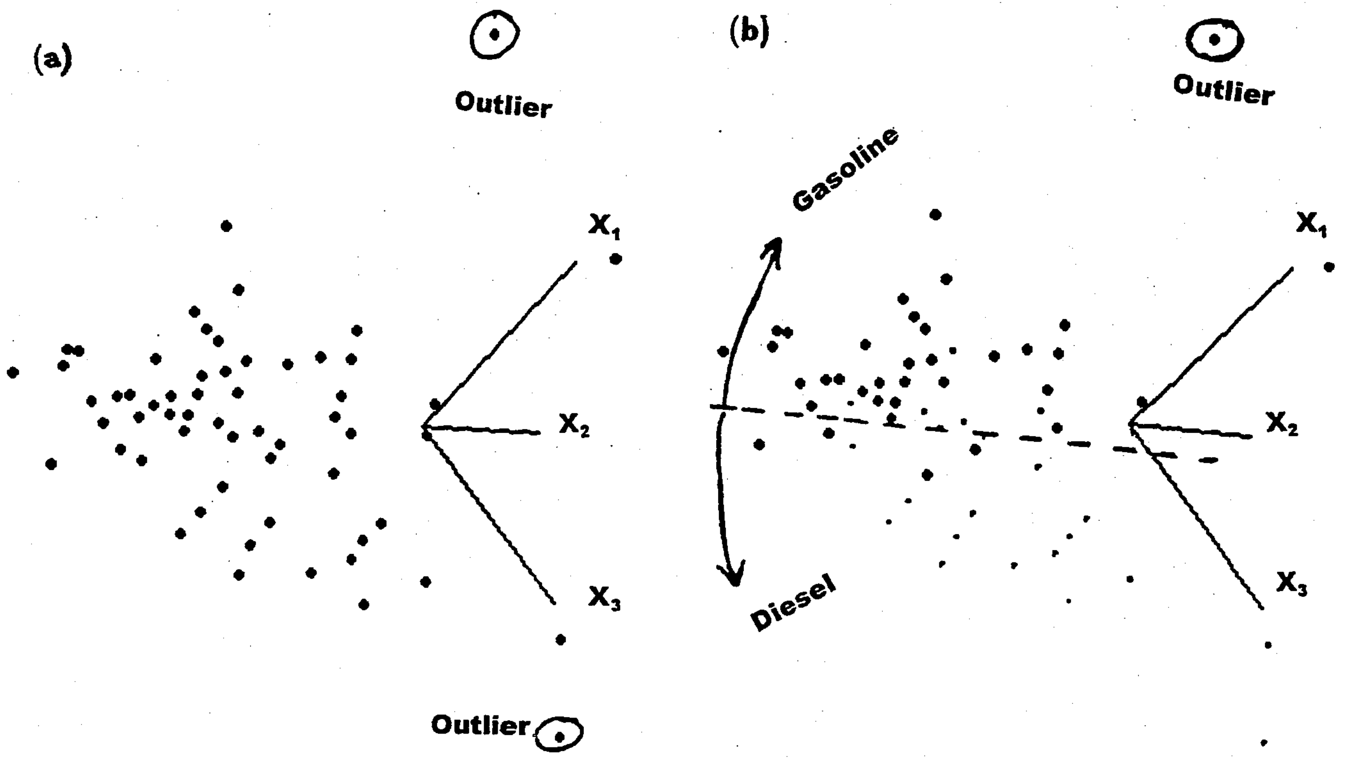
1.20



(a) The plot looks like a cigar shape, but bent. Some observations in the lower left hand part could be outliers. From the highlighted plot in (b) (actually non-bankrupt group not highlighted), there is one outlier in the nonbankrupt group, which is apparently located in the bankrupt group, besides the strung out pattern to the right.

(b) The dotted line in the plot would be an orientation for the classification.

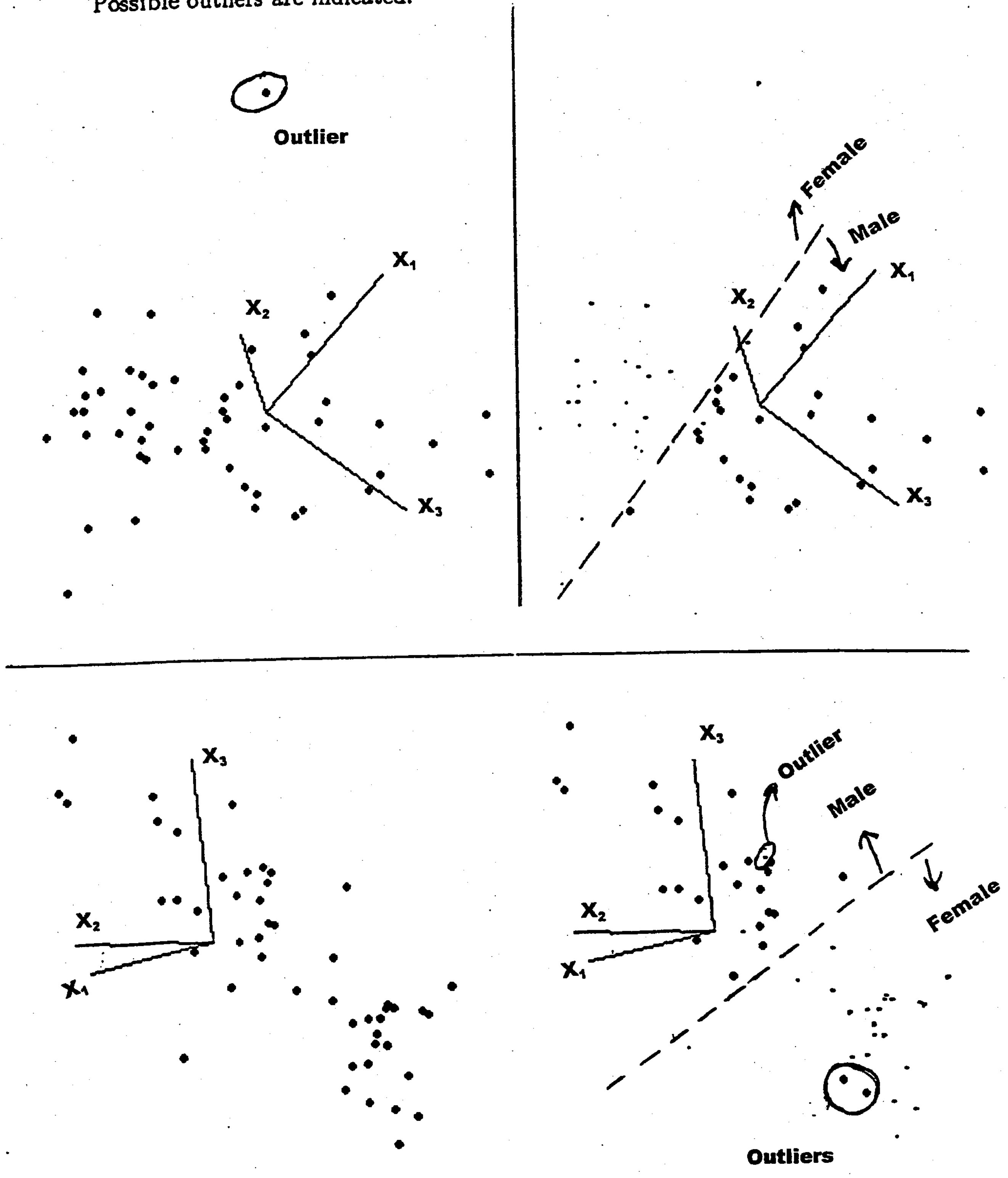
1.21



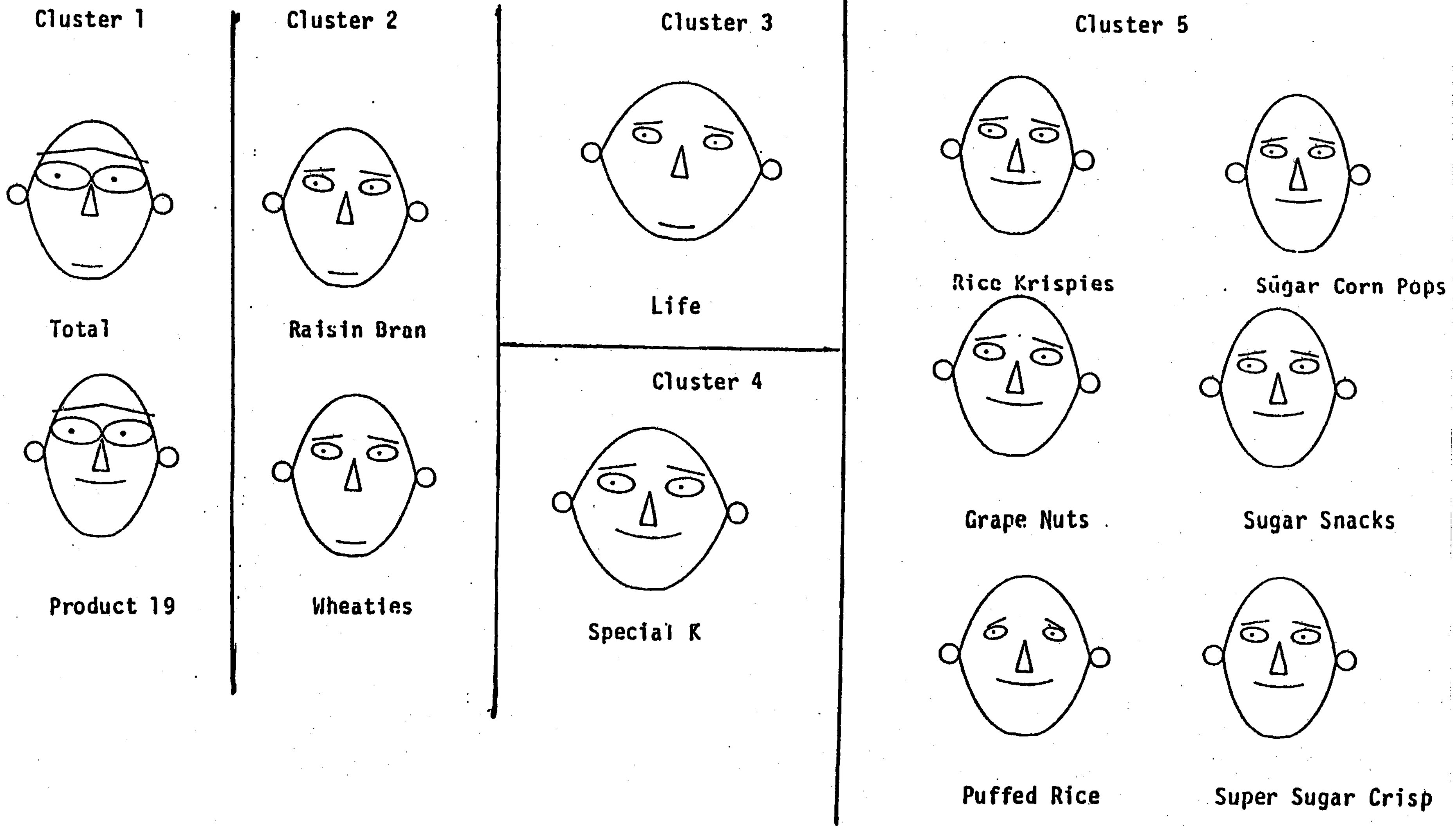
(a) There are two outliers in the upper right and lower right corners of the plot.

(b) Only the points in the gasoline group are highlighted. The observation in the upper right is the outlier. As indicated in the plot, there is an orientation to classify into two groups.

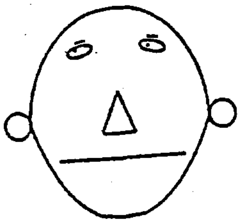
1.22 Possible outliers are indicated.



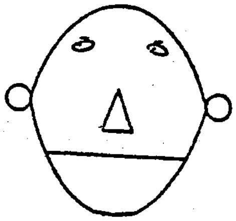
1.23 b) A visual clustering using Chernoff faces is given below.



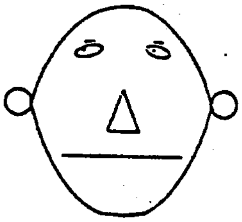
Cluster 1



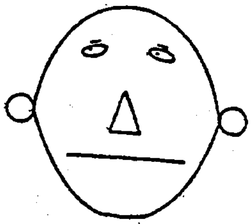
20



13

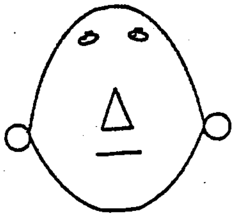


10

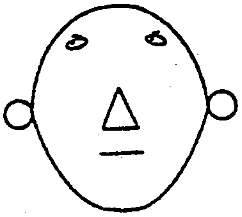


4

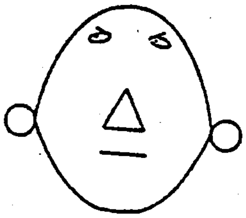
Cluster 2



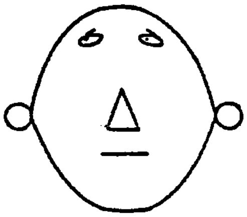
14



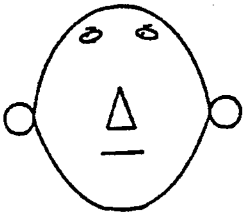
9



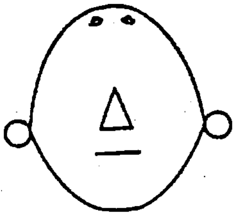
3



1

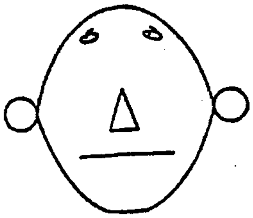


18

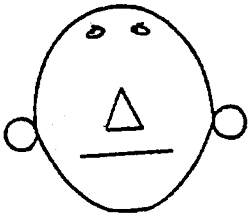


19

Cluster 3

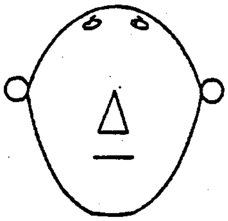


22

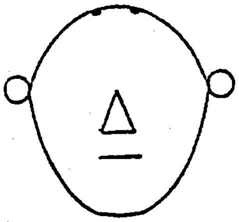


6

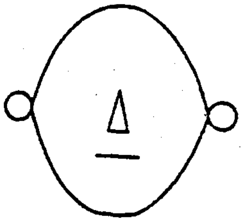
Cluster 4



8

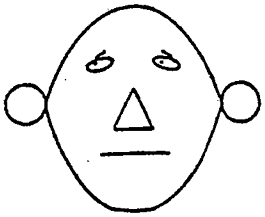


16

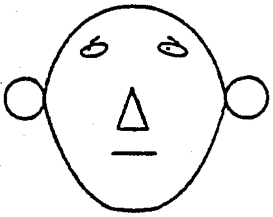


11

Cluster 5

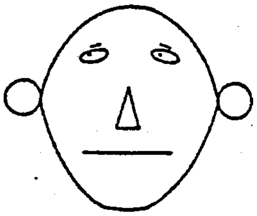


5

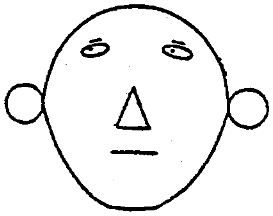


21

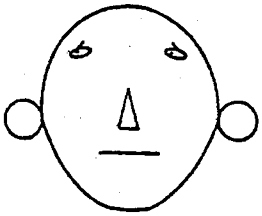
Cluster 6



2

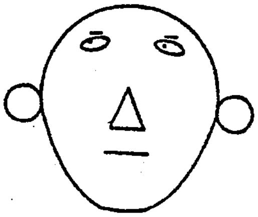


12

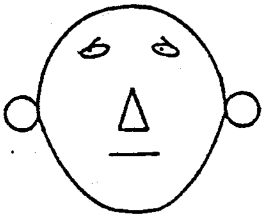


17

Cluster 7



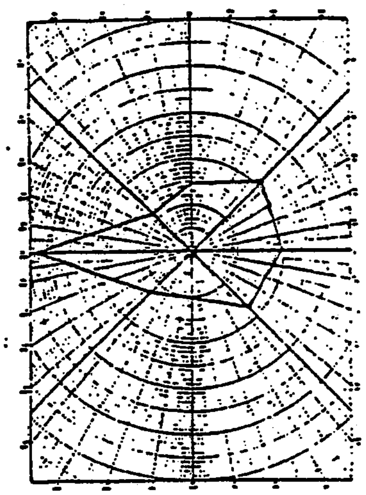
7



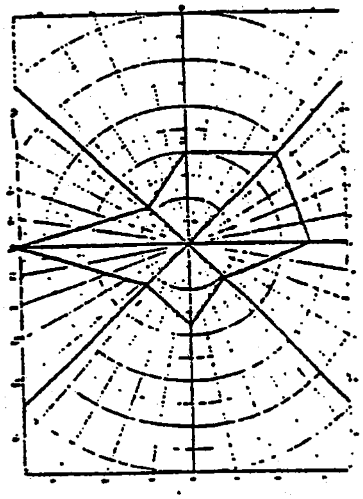
15

We have clustered these faces in the same manner as those in Example 1.12. Note, however, other groupings are equally plausible. For instance, utilities 9 and 18 might be switched from Cluster 2 to Cluster 3 and so forth.

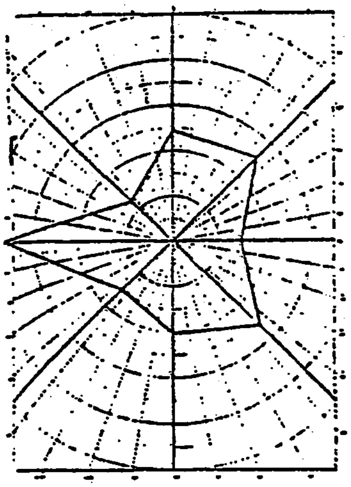
1.25 We illustrate one cluster of "stars". The remaining stars (not shown) can be grouped in 3 or 4 additional clusters.



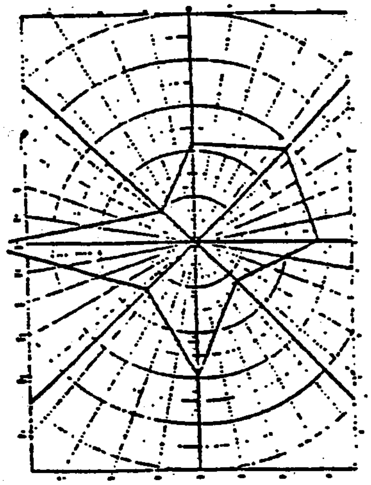
4



10



20

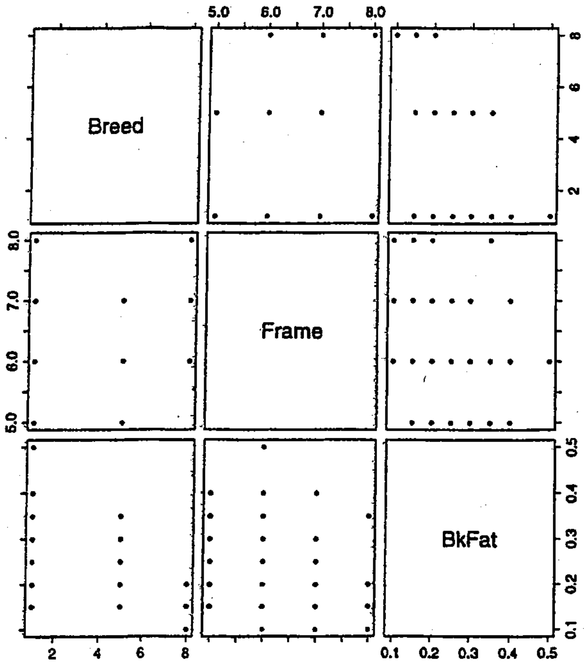


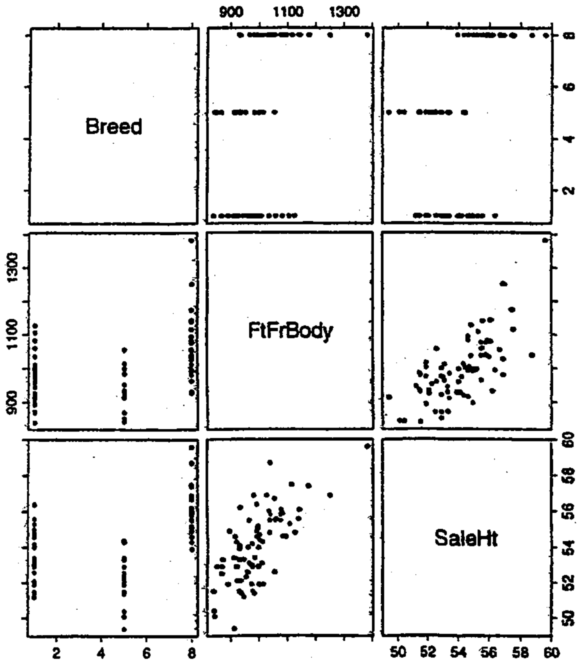
13

1.26 Bull data

(a)

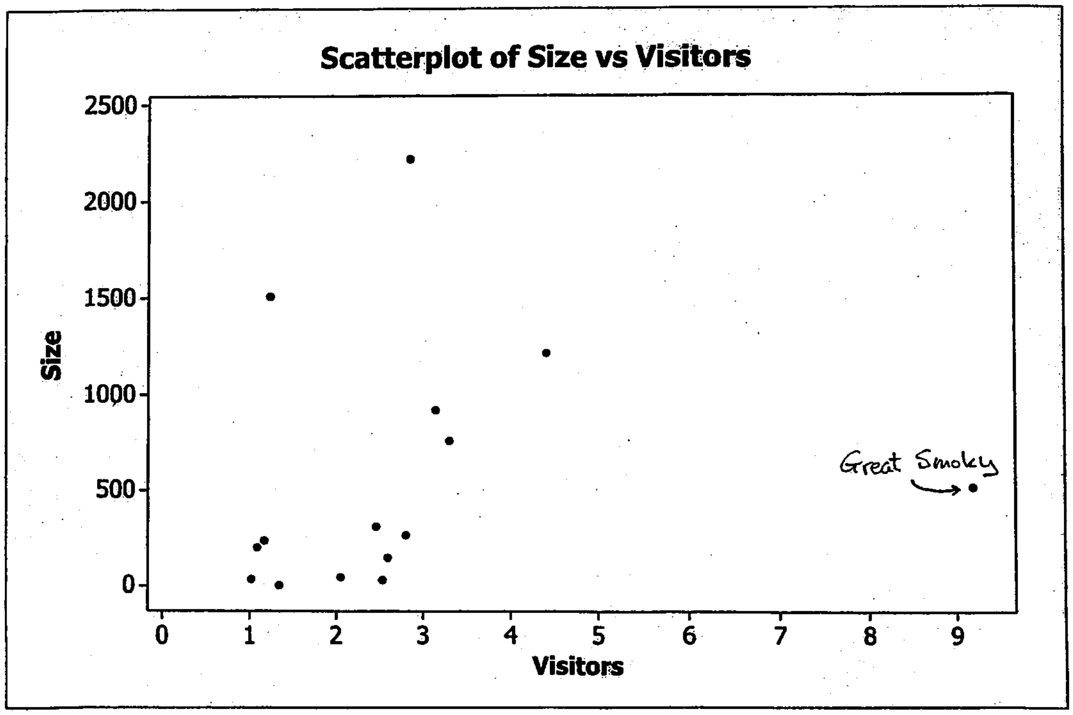
```
XBAR        R
             Breed  SalePr   YrHgt  FtFrBody  PrctFFB   Frame   BkFat  SaleHt  SaleWt
   4.3816    1.000  -0.224   0.525     0.409    0.472   0.434  -0.615   0.487   0.116
1742.4342   -0.224   1.000   0.423     0.102   -0.113   0.479   0.277   0.390   0.317
  50.5224    0.525   0.423   1.000     0.624    0.523   0.940  -0.344   0.860   0.368
 995.9474    0.409   0.102   0.624     1.000    0.691   0.605  -0.168   0.699   0.555
  70.8816    0.472  -0.113   0.523     0.691    1.000   0.482  -0.488   0.521   0.198
   6.3158    0.434   0.479   0.940     0.605    0.482   1.000  -0.260   0.801   0.368
   0.1967   -0.615   0.277  -0.344    -0.168   -0.488  -0.260   1.000  -0.282   0.208
  54.1263    0.487   0.390   0.860     0.699    0.521   0.801  -0.282   1.000   0.566
1555.2895    0.116   0.317   0.368     0.555    0.198   0.368   0.208   0.566   1.000

Sn
  Breed      SalePr   YrHgt FtFrBody  PrctFFB   Frame  BkFat  SaleHt    SaleWt
   9.55     -429.02    2.79   116.28     4.73    1.23  -0.17    3.00     46.32
-429.02   383026.64  450.47  5813.09  -226.46  272.78  15.24  480.56  25308.44
   2.79      450.47    2.96    98.81     2.92    1.49  -0.05    2.94     81.72
 116.28     5813.09   98.81  8481.26   206.75   51.27  -1.38  128.23   6592.41
   4.73     -226.46    2.92   206.75    10.55    1.44  -0.14    3.37     82.82
   1.23      272.78    1.49    51.27     1.44    0.85  -0.02    1.47     43.74
  -0.17       15.24   -0.05    -1.38    -0.14   -0.02   0.01   -0.05      2.38
   3.00      480.56    2.94   128.23     3.37    1.47  -0.05    3.97    145.35
  46.32    25308.44   81.72  6592.41    82.82   43.74   2.38  145.35  16628.94
```





1.27

(a) Correlation $r = .173$



(b) Great Smoky is unusual park. Correlation with this park removed is $r = .391$. This single point has reasonably large effect on correlation reducing the positive correlation by more than half when added to the national park data set.

(c) The correlation coefficient is a dimensionless measure of association. The correlation in (b) would not change if size were measured in square miles instead of acres.