International Institute for Technology and Management

November, 2008 **Unit: 76 – Management Mathematics**

This paper is not to be removed from the Examination Halls

Student Name :	
Student Number :	

TIME ALLOWED : 3 hours

Candidates should answer all of the **FIVE** questions. All questions carry equal marks.

Candidates are strongly advised to divide their time accordingly.

Graph paper is provided at the student request.

A hand held non-programmable calculator **may be used** when answering questions on this paper. The make and type of machine used must be stated clearly on the front cover of the answer book.

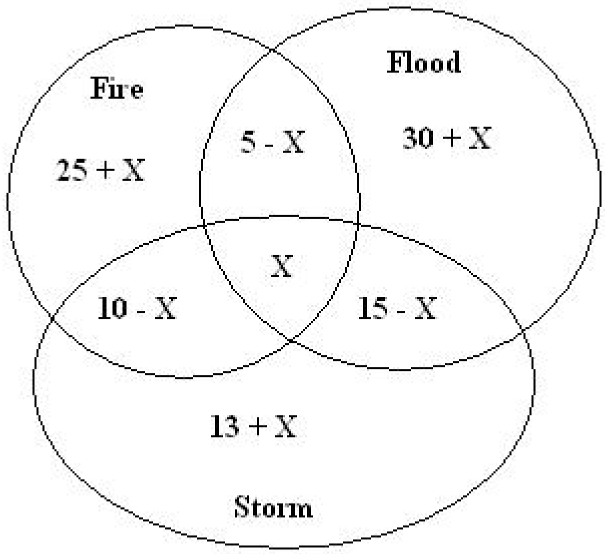
PLEASE TURN OVER

Answer **ALL** of the following **Five** questions:

1. An insurance company insures 20 000 businesses against the perils of fire, flood and storm damage. During a ten year period 99% of these businesses make no claim at all against the insurance company. No business claims for more than one of the perils at a time but of those businesses that have made one or more claims during the stated ten year period time:
40% have claimed for fire damage.
50% have claimed for flood damage.
38% have claimed for storm damage.
10% have claimed on different occasions for fire and storm damages.
15% have claimed on different occasions for flood and storm damages.
5% have claimed on different occasions for fire and flood damages.

a) If **X** is the number of businesses making claims for all the three perils, calculate the number of businesses (as a function of **X**) in each of the eight disjoint subsets which can be logically identified and produce an appropriate Venn diagram for percentage of businesses making claims. **(8 Marks)**

In total, the number of businesses making claims= 20000 /100 = 200.
The Venn diagram below is in "Percentage of businesses making claims":



b) How many businesses have claimed for all three types of damages (fire, flood and storm) on separate occasions? **(6 Marks)**

25+X+5-X+30+X+10-X+X+15-X+13+X = 100
98+X = 100 , X = 2% of 200 i.e. 4 businesses.

c) Assuming no business has claimed for the same type of damage more than once, how many claims in total have been made ? **(6 Marks)**

The total number of claims is:
2[27+32+15+2(3+13+8)+3(2)] = 256

2. Suppose the consumption this year is the average of last years consumption and this year's income, that is :

$$\mathbf{C_t = \tfrac{1}{2}\,(Y_t + C_{t-1})}$$

Suppose also that the relationship between next year's income and current investment is $\mathbf{Y_{t+1} = k\, I_t}$ for some positive constant k.

a.) Assuming the equilibrium condition : $Y_t = C_t + I_t$ holds, show that Y_t satisfies the following second order difference equation:

$$Y_t - \left(\frac{k+1}{2}\right)Y_{t-1} + \frac{k}{2}Y_{t-2} = 0$$

$Y_{t+1} = k\, I_t \;\Rightarrow\; I_t = \frac{1}{k}\, Y_{t+1}$ and $Y_t = C_t + I_t$

$\Rightarrow Y_t = C_t + \frac{1}{k}\, Y_{t+1}$, we need to eliminate C_t

$C_t = Y_t - \frac{1}{k}\, Y_{t+1}$ replacing t by t-1 : $C_{t-1} = Y_{t-1} - \frac{1}{k}\, Y_t$

Substitute these two in : $C_t = \tfrac{1}{2}\,(Y_t + C_{t-1})$

$$Y_t - \frac{1}{k}\, Y_{t+1} = \tfrac{1}{2}\,(Y_t + Y_{t-1} - \frac{1}{k}\, Y_t)$$

$Y_{t+1} - \left(\frac{k+1}{2}\right)Y_t + \frac{k}{2}Y_{t-1} = 0$ replace t by t-1 : $Y_t - \left(\frac{k+1}{2}\right)Y_{t-1} + \frac{k}{2}Y_{t-2} = 0$

b) Suppose k = 3 and that the initial value Y_0 is positive.
Solve the equation :

$$Y_t - 2Y_{t-1} + \frac{3}{2}Y_{t-2} = 0$$

The auxiliary equation : $r^2 - 2r + 3/2 = 0$
has complex roots so the solution Y_t is :

$$Y_t = \left(\sqrt{\frac{3}{2}}\right)^t (A\cos\alpha\, t + B\sin\alpha\, t) \;;\; \alpha = \cos^{-1}\left(\frac{-a}{2\sqrt{b}}\right) = \cos^{-1}\left(\frac{-2}{2\sqrt{\frac{3}{2}}}\right) = \frac{\pi}{2}$$

$$Y_t = \left(\sqrt{\frac{3}{2}}\right)^t \left(A\cos\frac{\pi}{2}t + B\sin\frac{\pi}{2}t \right)$$

Show that Y_t oscillates with increasing magnitude.
complex roots means oscillating.

As $t \to \infty$; $\left(\sqrt{\frac{3}{2}}\right)^t \to \infty$ since $\left|\sqrt{\frac{3}{2}}\right| \succ 1$

$\therefore Y_t$ oscillates with increasing magnitude c.) Discuss according to the values of k the behaviour of Y_t the solution of the equation given in (a): $Y_t - \left(\frac{k+1}{2}\right)Y_{t-1} + \frac{k}{2}Y_{t-2} = 0$

The auxiliary equation : $r^2 - \left(\frac{k+1}{2}\right)r + \frac{k}{2} = 0$

The discriminant of the quadratic formula :

$b^2 - 4ac = \left(\frac{k+1}{2}\right)^2 - 4\frac{k}{2}$; Three cases occur :

1. If $\left(\frac{k+1}{2}\right)^2 - 4\frac{k}{2} < 0 \Rightarrow k^2 - 6k + 2 < 0$

then the roots are complex and the general

solution is of the form :

$$Y_t = \left(\sqrt{\frac{k}{2}}\right)^t (A\cos\alpha t + B\sin\alpha t)$$

2. If $\left(\frac{k+1}{2}\right)^2 - 4\frac{k}{2} > 0 \Rightarrow k^2 - 6k + 2 > 0$

then the roots are real distinct and the general solution is of the form:

$$Y_t = Ar_1^t + Br_2^t$$

3. If $\left(\frac{k+1}{2}\right)^2 - 4\frac{k}{2} = 0 \Rightarrow k = 3 \pm 2\sqrt{2}$

then the roots are real equal and the general solution is of the form:

$$Y_t = (A + Bt)r^t$$

3. Eight employees (A,B,C...H) of a company give answers (Y=Yes,N=No) to six different questions about themselves. The results are as follows:

	Employee							
Question	A	B	C	D	E	F	G	H
Do you have a Management degree?	Y	Y	N	N	N	Y	N	N
Are you older than 40?	N	Y	N	N	Y	Y	N	N
Are you male?	Y	Y	Y	Y	N	Y	N	N
Did you join our company in the last 2 years?	Y	N	N	N	N	Y	N	N
Are you married?	N	Y	N	N	N	Y	N	N
Have you worked abroad?	N	Y	N	N	N	Y	Y	Y

The company wants to use this data in order to form 3 groups of employees who are as similar as possible within a group but as different as possible between the groups.

(a) Construct a similarity matrix for the employees. (4 marks)

	A	B	C	D	E	F	G	H
A	-	2	4	4	2	3	2	2
B		-	2	2	2	5	2	2
C			-	6	4	1	4	4
D				-	4	1	4	4
E					-	1	4	4
F						-	1	1
G							-	6
H								-

(b) Using a single linkage hierarchical clustering approach, determine the constituent employees of the three groups and produce a suitable diagram to show the clustering process. (12 marks)
The highest coefficient is **6** , combine C and D (alternatively G and H) the updated matrix is (use max distances) :

	A	B	CD	E	F	G	H
A	-	2	4	2	3	2	2
B		-	2	2	5	2	2
CD			-	4	1	4	4
E				-	1	4	4
F					-	1	1
G						-	6
H							-

Next combine G and H :

	A	B	CD	E	F	GH
A	-	2	4	2	3	2
B		-	2	2	5	2
CD			-	4	1	4
E				-	1	4
F					-	1
GH						-

Next combine B and F :

	A	BF	CD	E	GH
A	-	3	4	2	2
BF		-	2	2	2
CD			-	4	4
E				-	4
F					1
GH					-

Next combine CD and GH :

	CDGH	A	BF	E
CDGH	-	4	2	4
A		-	3	2
BF			-	2
E				-

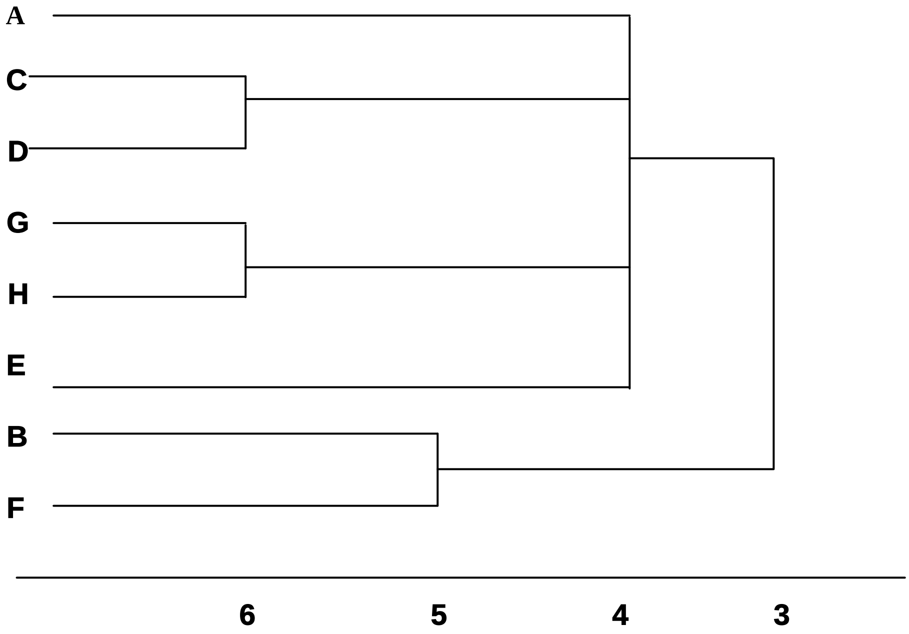
Next combine A and CDGH

	ACDGH	BF	E
ACDGH	-	3	4
BF		-	2
E			-

Next combine ACDGH and E

	ACDGHE	BF
ACDGHE	-	3
BF		-

Dendogram:



Note that we have 3 clusters (6 – 5 – 4);the three groups are : CDGH , BF and AE.

(c) It is planned to repeat the above type of process in other divisions of the company. However it is felt than some of the questions are asking similar types of question to each other. Suggest two questions you might consider dropping from the questionnaire, and explain your reasoning. (4 marks)

The similarity matrix of the questions:

	1	2	3	4	5	6
1	-	6	6	7	7	5
2		-	4	5	7	7
3			-	5	5	3
4				-	6	4
5					-	6
6						-

The highest similarities are : 1-4 ,1-5 and 2-5 , 2-6
Questions 1 and 2 should be omitted.

4. (a) Explain the difference of Single Linkage and Complete Linkage in cluster analysis. Under what circumstances would you use one of these techniques in preference to the other? **(6 Marks)**

1. Single linkage : the simplest and most straightforward clustering method to understand is single linkage (nearest neighbour). Single linkage clustering provides a fairly accurate picture of the relationships between pairs of objects. As a consequence of chaining, results of Single Linkage clustering are fairly sensitive to noise in the data, because noise changes the similarity values and may thus easily modify the order in which objects cluster.

Summary: simple; contraction of space (chaining); combinatorial method. Good complement to ordination.

2. Complete linkage: agglomeration (farthest neighbour analysis) is essentially opposite in approach to the single linkage analysis. In this method, the fusion of two clusters relies on the most distant pair of objects instead of the closest. Thus, an object joins a cluster only when it is linked to all the objects already in the cluster. Two clusters can fuse only when all objects of the first cluster are linked to all objects of the second cluster. Complete Linkage clustering produces maximally linked and spherical clusters (instead of the chained clusters of Single Linkage).
Summary: space expansion; many objects cluster at low similarity; arbitrary rules to resolve conflicts; combinatorial method; Increases contrast among clusters.

(b) Consider the distances between pairs of five objects as follows:

$$D=(d_{ij})=\begin{array}{c} \\ 1\\2\\3\\4\\5\end{array}\begin{array}{c}\begin{array}{ccccc}\mathbf{1}&\mathbf{2}&\mathbf{3}&\mathbf{4}&\mathbf{5}\end{array}\\ \begin{bmatrix} - & & & & \\ 9 & - & & & \\ 3 & 7 & - & & \\ 6 & 5 & 9 & - & \\ 11 & 10 & 2 & 8 & - \end{bmatrix}\end{array}$$

Produce a Dendogram of clustering these objects using single and complete linkages hierarchical technique.

(14 Marks)

Single linkage

Remember : When you need to decide which data points you should **combine**, look for the **least** number in the similarity matrix. *Use this strategy regardless of which method you use single or complete*.
Looking at the similarity matrix ,the least is **2** ,so we start by combining **3 & 5** (computing new distances by taking the minimum):

$$\begin{array}{c} \\ 3-5\\1\\2\\4\end{array}\begin{array}{c}\begin{array}{cccc}\text{3-5}&\text{1}&\text{2}&\text{4}\end{array}\\ \begin{bmatrix} - & & & \\ 3 & - & & \\ 7 & 9 & - & \\ 6 & 6 & 5 & - \end{bmatrix}\end{array}$$

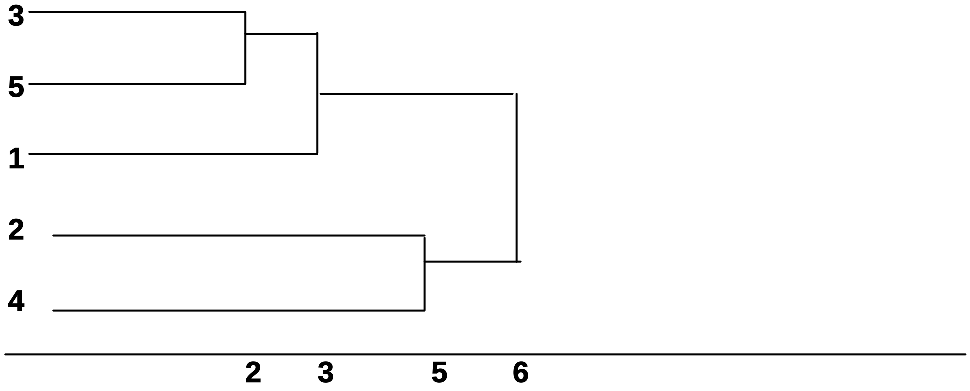
The least number now is **3** , we need to combine **3-5 & 1** :

$$\begin{array}{c} \\ 3-5-1\\2\\4\end{array}\begin{array}{c}\begin{array}{ccc}\text{3-5-1}&\text{2}&\text{4}\end{array}\\ \begin{bmatrix} - & & \\ 7 & - & \\ 6 & 5 & - \end{bmatrix}\end{array}$$

The least number now is **5** , we need to combine **2 & 4** :

$$\begin{array}{c} \\ 3-5-1\\2-4\end{array}\begin{array}{c}\begin{array}{cc}\text{3-5-1}&\text{2-4}\end{array}\\ \begin{bmatrix} - & \\ 6 & - \end{bmatrix}\end{array}$$

Dendogram



ii complete linkage

Looking at the similarity matrix, the least is **2** ,so we start by combining **3 & 5** (computing new distances by taking the maximum):

$$\begin{array}{c} \\ 3-5 \\ 1 \\ 2 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 3\text{-}5 & 1 & 2 & 4 \end{array} \\ \begin{bmatrix} - & & & \\ 11 & - & & \\ 10 & 9 & - & \\ 9 & 6 & 5 & - \end{bmatrix} \end{array}$$

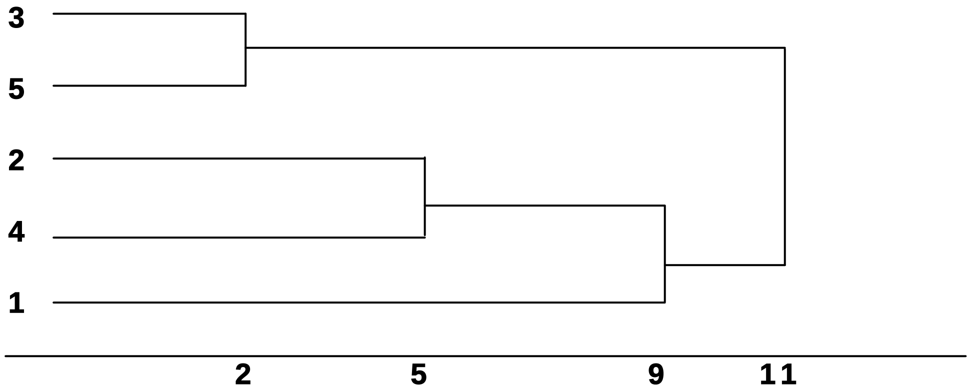
The least number now is **5** , we need to combine **2 & 4** :

$$\begin{array}{c} \\ 3-5 \\ 1 \\ 2-4 \end{array} \begin{array}{c} \begin{array}{ccc} 3\text{-}5 & 1 & 2\text{-}4 \end{array} \\ \begin{bmatrix} - & & \\ 11 & - & \\ 10 & 9 & - \end{bmatrix} \end{array}$$

The least number now is **9** , we need to combine **2-4 & 1** :

$$\begin{array}{c} \\ 3-5 \\ 2-4-1 \end{array} \begin{array}{c} \begin{array}{cc} 3\text{-}5 & 2\text{-}4\text{-}1 \end{array} \\ \begin{bmatrix} - & \\ 11 & - \end{bmatrix} \end{array}$$

Dendogram



5 (a). In multivariate data analysis what is a 'multiple scatter plot' and what are its uses? **(5 marks)**

(b) What is meant by the term 'outlier' and why are 'outliers' important? **(3 marks)**

(c) The following table shows bivariate observations on 21 subjects:

Subject	Value of X	Value of Y
1	1	1
2	2	2
3	2.5	1.5
4	2.5	3.5
5	3	2
6	3	3
7	3	3.5
8	3.5	2
9	3.5	2.5
10	4	3
11	4	3.5
12	4.5	2.5
13	5	4
14	5	5
15	5.5	5.5
16	6	3.5
17	6	4
18	6.5	1
19	6.5	5
20	7	11
21	11.5	6

You are asked to:

i. Draw A Box and Whisker diagram ('Box Plot') for each variable to identify any 'outliers' or 'extremes'. **(8 marks)**

ii. Produce a scatter diagram of X against Y and clearly mark the observation(s) that you have identified as possible 'outliers' or 'extremes'. Do any other subjects seem to produce 'outlying' results? **(4 marks)**

a) When carrying out large surveys for market research ,social research or scientific research ,it is always wise to carry out an initial simple analysis of the data:

i.) to detect outlying observation.
ii.) to check for typographical errors.
iii.) to compare the ranges of measurement of the variables.
iv.) to compare the dispersion of the variables

Many of these tasks (i) to (iv) can be carried out by graphical means like Scatter diagram which makes any outlying observation. If one is found ,it then can be checked for accuracy.

b) An outlier , extreme data values relative to others in a sample, an observation that lies an abnormal distance from other values in a random sample from a population. In a sense, this definition leaves it up to the analyst (or a consensus process) to decide what will be considered abnormal. Outlier identification is important in many applications of multivariate analysis. Either because there is some specific interest in finding anomalous observations or as a pre-processing task before the application of some multivariate method, in order to preserve the results from possible harmful effects of those observations. Because the existence of outliers in a data set may bias the results.

c) i. The box plot is a useful graphical display for describing the behavior of the data in the middle as well as at the ends of the distributions. The box plot uses the median and the lower and upper quartiles (defined as the 25th and 75th percentiles). If the lower quartile is Q1 and the upper quartile is Q2, then the difference (Q2 - Q1) is called the interquartile range or IQ.

Values of X : 1, 1, 2, 2.5, 2.5, 3, 3, 3, 3.5, 3.5, 4, 4, 4.5, 5, 5. 5, 6, 6, 6.5, 7, 11.5

The median : 4

Lower Quartile Q1 = 0.25(N+1) = ¼ (22) = 5.5th order = ½ (3+3) = 3

Upper Quartile Q2 = 0.75(N+1) = ¾ (22) = 16.5th order = ½ (6+6) = 6

Inter-Quartile range IQ = 6 – 3 = 3

Lower extreme limit = Q1 – 3IQ = 3 – 9 = - 6

Lower outlier limit = Q1 – 1.5IQ = 3 – 1.5(3) = - 1.5

Upper extreme limit = Q2 + 3IQ = 6 + 3(3) = 15

Upper outlier limit = Q2 + 1.5IQ = 6 + 1.5(3) = 10.5

These limits enable us to draw a Box plot and also to conclude that we should eliminate data of 11.5

Values of Y : 1, 1, 1.5 , 2, 2, 2, 2.5, 2.5, 3, 3, 3.5, 3.5, 3.5 , 3.5 , 4, 4, 5, 5 , 5.5, 6, 11

The median : 3.5

Lower Quartile Q1 = 0.25(N+1) = ¼ (22) = 5.5th order = ½ (2+2) = 2

Upper Quartile Q2 = 0.75(N+1) = ¾ (22) = 16.5th order = ½ (4+5) = 4.5

Inter-Quartile range IQ = 4.5 – 2 = 2.5

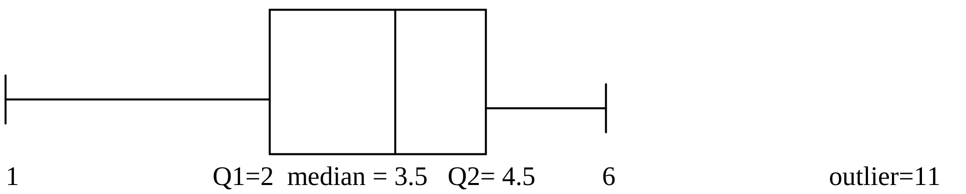
Lower extreme limit = Q1 – 3IQ = 2 – 7.5 = - 5.5

Lower outlier limit = Q1 – 1.5IQ = 2 – 1.5(2) = -1

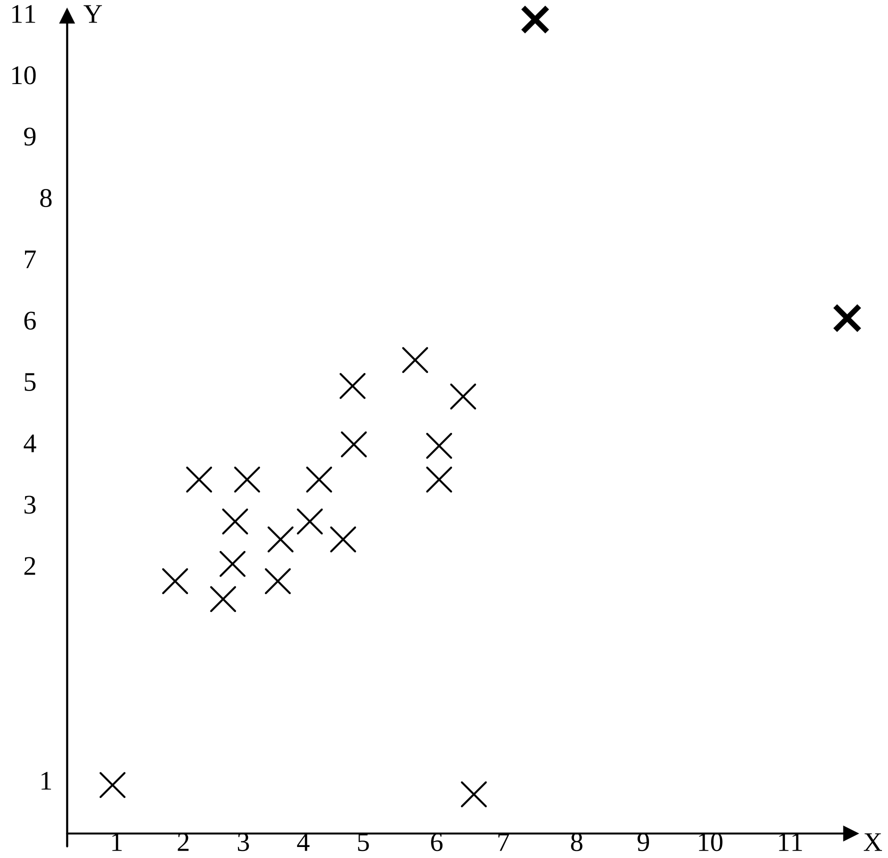
Upper extreme limit = Q2 + 3IQ = 4.5 + 3(2.5) = 12

Upper outlier limit = Q2 + 1.5IQ = 4.5 + 1.5(2.5) = 8.25

These limits enable us to draw a Box plot and also to conclude that we should eliminate data of 11



ii. Are useful when we wish to visualize the relationship between two measurement variables. To draw a scatter plot: we assign one variable to each axis and plot one point for each pair of measurements.



Outliers should be investigated carefully. Often they contain valuable information about the process under investigation or the data gathering and recording process. Before considering the possible elimination of these points from the data, one should try to understand why they appeared and whether it is likely similar values will continue to appear. Of course, outliers are often bad data points.

END OF PAPER