THE OPEN UNIVERSITY OF SRI LANKA
B.Sc./ B. Ed. DEGREE, CONTINUING EDUCATION PROGRAMME
CLOSED BOOK TEST 2006/2007
PSU 1182/PSE 3182/PSZ 4182 – BIO STATISTICS

DURATION: ONE AND HALF HOURS

Date: 29 – 08 – 2006	Time: 3.30pm – 5.00pm

ANSWER ALL QUESTIONS.

Non-programmable calculators are permitted.

1. A researcher is interested in finding out whether a particular chemical solution has any effect on the time taken for sunflower plants to flower. He randomly selected 171 sunflower plants for the study. For each of the 46 plants in the treatment group, 1ml of the chemical solution was added after two weeks of planting. The other 125 plants were used as the control and were kept in similar conditions as the treatment group except that the chemical solution was not added. The researcher recorded the group membership of each plant (as treatment or control) and the time taken for the plant to flower.

   The accompanying table summarises the data collected in the experiment.

Number of plants in each category	Time taken to flower (in weeks)				
	11 – 15	16-20	21- 25	26-30	total
treatment group	5	21	14	6	46
Control group	23	31	58	13	125

   i) Classify the data collected by the researcher as nominal, ordinal, interval or ratio.
   ii) Construct a suitable graph that highlights the effect of the chemical solution, if any, on the time taken for sunflower plants to flower.
   iii) Clearly describe the findings from the graph constructed in part ~~(i)~~ (ii)

2. A researcher is interested in studying the increase in fatty acids in mustard plants when they are genetically modified by crossing with canola plants. The following table presents the percentage increase in fatty acids measured on 30 plants that were crossed with the canola plants.

2	2	3	4	2	2	3	4	5	2
6	3	3	2	4	2	4	3	2	6
2	3	2	3	3	3	4	2	3	2

i) Construct a suitable graphical summary that highlights the variations in the percentage increase in fatty acids of mustard plants that are genetically modified. Your graph should also highlight extreme observations if any.
ii) Clearly state the findings from the graph constructed in part ~~(ii)~~. (i)

3. The following table summarises the lengths (in cm) of 40 fish in a stream.

21	13	14	14	11	21	20	14	17	17
15	15	19	11	16	14	14	12	20	19
12	15	15	16	16	17	13	14	18	14
15	11	14	14	13	12	15	19	20	20

i) Compute the sample mean of the data and describe what it measures.
ii) Estimate the quartiles of the data.
iii) Clearly describe what each of the quartiles estimate.

2006/2007
CBT

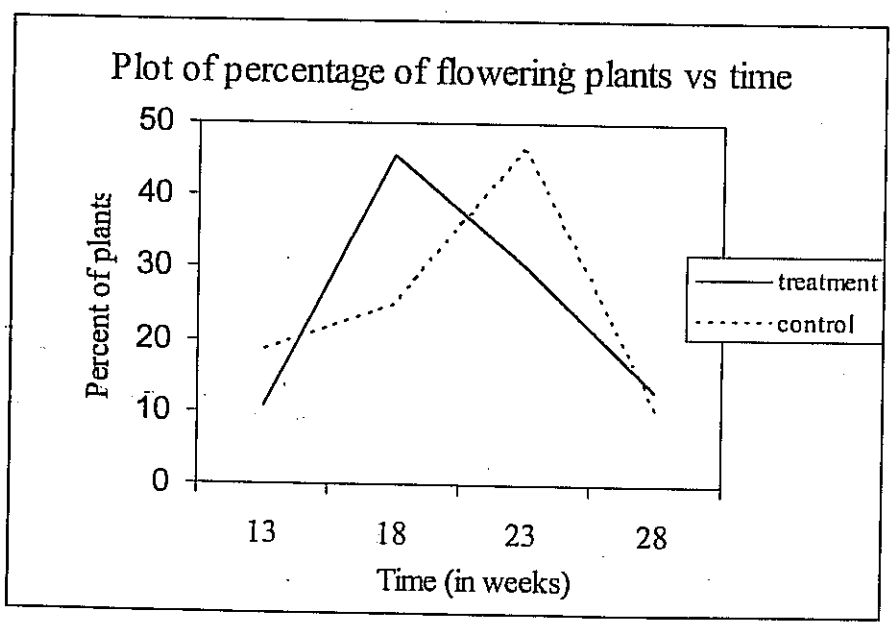
PSU 1182/ PSE 3182/ PSZ 4182 – Bio Statistics
Closed Book Test 2006/ 2007
Solutions

1. i) The data collected on each plant are the group membership of the plant (treatment group/ control group) and the time taken for the plant to flower.
Group membership of the plant is nominal.
Time taken to flower is recorded as the class interval that the observation belongs to. Hence the data collected on time is in the ordinal scale. However, if the researcher had recorded the actual time taken by the plant to flower the resulting measurements are in the ratio scale.

ii)

Group label	Description	Class interval				Total
		11 – 15	16 – 20	21 – 25	26 – 30	
Treatment	Frequency	5	21	14	6	46
	Relative frequency	5/46	21/46	14/46	6/46	
	Percentage	10.87	45.65	30.43	13.04	
Control	Frequency	23	31	58	13	125
	Relative frequency	23/125	31/125	58/125	13/125	
	Percentage	18.40	24.80	46.40	10.40	

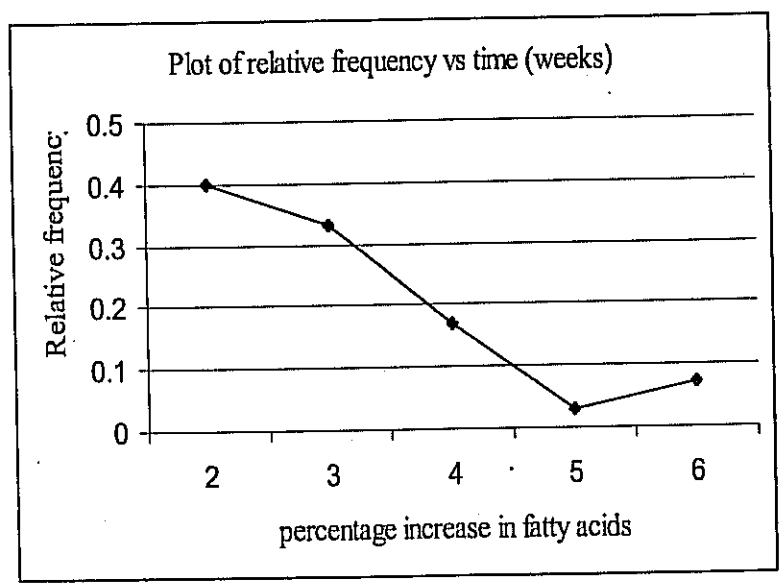
Note: The sample sizes are quite different. Our intention is to compare the two groups. Therefore, it is advisable to use relative frequencies or percentages in the comparison.



iii) The chemical solution has contributed for a shift in the peak towards the left. This means for majority of plants the time taken to flower has got reduced compared to the group that had no chemical solution. The random variation appears to be more or less similar for both groups. Also, none of the data show extreme observations.

2.

Time (in weeks)	Tally mark	Frequency	Relative frequency
2	𝍸 𝍸 ||	12	0.40
3	𝍸 𝍸	10	0.33
4	𝍸	5	0.17
5	|	1	0.03
6	||	2	0.07
Total		30	1.00



ii) Around 40% of the plants that were crossed with canola plants had an increased fatty acid level of 2 while around 30% had an increased fatty acid level of 3. The data show moderate variation in the percentage increase in fatty acids ranging from 2 % to 6%. The plot does not show any extreme observations.

3) i) Sample mean of the data $=\bar{x}=\frac{1}{40}\sum_{i=1}^{40}x_i$

$$=\frac{1}{40}(21+13+\cdots+20+20)=\frac{1}{40}\times 620$$
$$=15.50 \text{ cm}$$

The sample mean is the average length of fish in the sample. It estimates the average length of fish in the stream or the expected length of a randomly chosen fish.

ii) To compute the quartiles first we arrange the observations in ascending order. This gives the following.

11	11	11	12	12	12	13	13	13	14
14	14	14	14	14	14	14	14	15	15
15	15	15	15	16	16	16	17	17	17
18	19	19	19	20	20	20	20	21	21

There are 40 observations. Therefore, the second quartile or the median Q_2 is the average of 20^{th} and 21^{st} observations. This gives $Q_2=\frac{1}{2}(15+15)=15.$

The first quartile Q_1 is the average of 10^{th} and 11^{th} observations. This gives $Q_1=\frac{1}{2}(14+14)=14.$

The third quartile Q_3 is the average of 30^{th} and 31^{st} observations. This gives $Q_3=\frac{1}{2}(17+18)=17.5.$

iii) The first quartile estimates the expected largest length of the smallest 25 % of the fish. The second quartile estimates the expected highest length of the smallest 50% of the fish. The third quartile estimates the expected smallest length of the largest 25 % of the fish.