

DEGREE EXAMINATION

ST1505 Understanding Data

Wednesday 2 June 2004

(12noon to 2pm)

Only calculators approved by the Department of Mathematical Sciences may be used in this examination. Calculator memories must be clear at the start of the examination.

Marks may be deducted for answers that do not show clearly how the solution is reached.

Answer ALL questions.

1. Below is a back to back stem and leaf plot comparing reaction times (in hundredths of a second) for two groups of volunteers (denoted by A and B). There are 50 individuals in each group.

A	Group	B
1	4	111222457899
988443	5	00123345566666777899
9999886665422200	6	03445578
9976665432211100	7	000238
953322100	8	03
63	9	6
	10	6

Leaf Unit = 1.0

- (a) Calculate the median reaction time for both groups.
- (b) Without performing any further calculations, state which of the following statements is correct for group B :
- (i) the median is less than the mean;
 - (ii) the median is about equal to the mean;
 - (iii) the median is greater than the mean.
- (c) Without performing any further calculations, state which of the following numbers is approximately equal to the standard deviation of group A : 2.1, 3.4, 5.3, 6.8 or 10.7.
- (d) Calculate the quartiles for both groups and, using a suitable graphical method, sketch a plot to compare the quartiles (and some other features) of groups A and B .

[9]

2. An article in “The Press and Journal” (December 11th 1993) discussed the environmental problems caused by underground petrol tanks that leak. There are over 2 million tanks in Great Britain and several studies have suggested that roughly 25% of them leak.

Suppose a random sample of 30 tanks is selected from the population of all tanks. Let S denote the event that a tank leaks and suppose that $P(S) = 0.25$.

- (a) What assumptions must we make to model X , the number of leaking tanks out of the 30 tanks sampled, by a binomial distribution? What are the parameters of the distribution?
- (b) Assuming the binomial approximation is correct:
- What are the mean and variance of the number of leaking tanks?
 - What is the probability that exactly half of the tanks in the sample leak?
 - What is the probability that 15 or fewer of the sampled tanks leak?
 - What is the probability that more than half of the sampled tanks leak?

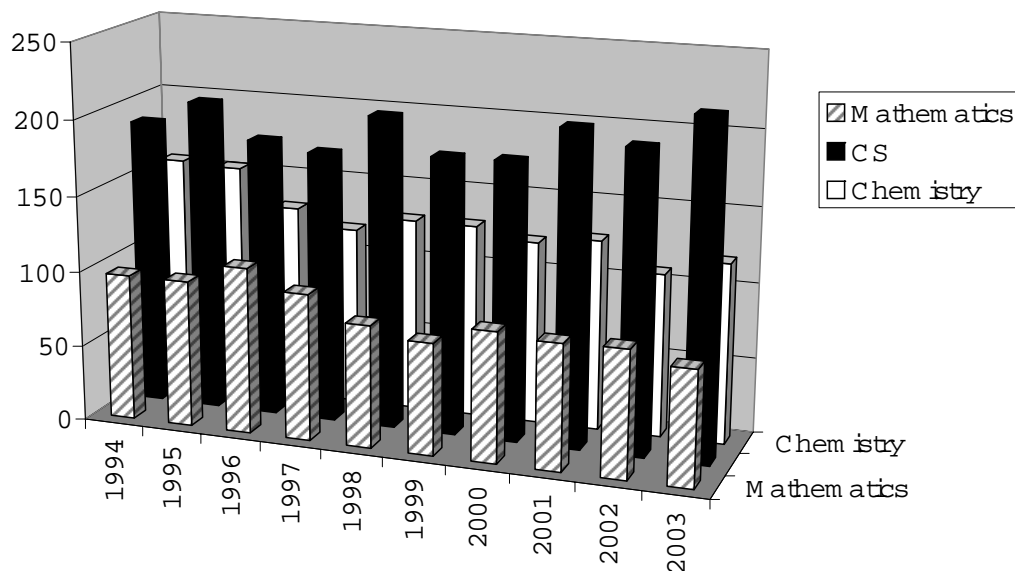
[8]

3. The ACME company produces widgets, 25% of its output by method A and the rest by method B . Method A produces 1 faulty widget in 20 and method B produces 1 faulty widget in 10 (on average).

- (a) What proportion of the company’s output is faulty?
- (b) Given that a randomly chosen widget is faulty, find the probability that it was produced by method A .

[6]

4. In a meeting, a Scottish Executive minister shows the following graph illustrating the numbers of students applying to study maths, computer science and chemistry over a 10 year period.



- (a) List two problems with this graph.
- (b) Sketch a graph to show how you would show these data in a more appropriate way.
- (c) Briefly comment on any trends seen in these data.

[6]

5. The resting heart rate of women in a certain population can be approximated by a normal distribution with mean 71 beats per minute and standard deviation 12 beats per minute.

(a) What is the probability that a randomly selected woman has a resting heart rate

- (i) greater than 80 beats per minute?
- (ii) between 68 and 80 beats per minute?

(b) The resting heart rates of 25 women sampled at random from this population were measured. What is the probability that their mean heart rate was less than 65 beats per minute?

(c) The resting heart rates of a random sample of 100 women from another population were measured, with results x_1, x_2, \dots, x_{100} . You are given that

$$\sum_{i=1}^{100} x_i = 6774 \cdot 8, \text{ and } \sum_{i=1}^{100} x_i^2 = 472283.$$

Calculate a 95% confidence interval for the mean resting heart rate of this population. [12]

6. In order to determine which of two varieties of pine gave the better yield of timber, six regions in Scotland were chosen and two stands of pine, one of each variety, planted in each region.

The following yields, in m^3 of timber, were obtained:

Region	1	2	3	4	5	6
Variety A	818	745	917	483	930	668
Variety B	766	700	905	495	850	642

(a) Use the data to test the null hypothesis that there is no difference in yield between the species.

(b) State the assumptions you have made in your test in (a). [8]

7. The following table summarises the results of a survey of men at a large company which investigated the association between blood pressure (in mm of mercury) and occupation.

		Type of Work		
		Sales	Managerial	Professional
Blood Pressure	≤ 80	34	38	17
	> 80	63	51	49

A chi-squared test of independence between work type and blood pressure was performed using MINITAB. Edited output from MINITAB is given below:

Chi-Square Test

Expected counts are printed below observed counts

	Sales	Manage	Prof	Total
1	34	38	17	89
	34.26	<a>	23.31	
2	63	51	49	163
	62.74	57.57	42.69	
Total	97	89	66	252

Chi-Sq = 0.0019 + + 1.71
+ 0.0011 + 0.75 + 0.93 = ?????
DF = <c>

(a) Using a suitable method, graphically compare the levels of blood pressure for the three types of occupation.

(b) Showing necessary working, find the missing values <a>, and <c>.

(c) Is there any evidence of an association between the type of work and blood pressure? Report the results of this test briefly in a non-technical way.

[12]

8. Explain, with an example, what is meant by each of the following terms:

- (a) a simple random sample;
- (b) a stratified sample;
- (c) a systematic sample.

[9]