UNIVERSITY OF ABERDEEN **SESSION 2005–2006**

DEGREE EXAMINATION

ST1505 Understanding Data

Monday 22 May 2006 (12 noon–2 pm)

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.

Attempt ALL THREE of the questions in SECTION A and TWO of the questions in SECTION B. Each question in section A is worth 10 marks and each question in section B is worth 20 marks.

SECTION A

1. The following table gives the age distribution of the population of a town in Moldavia.

Age	Frequency
0–19	9071
20–39	9560
40–59	8432
60–79	4947
80–99	1152
100 & over	3

(i) Using the class mid-points and frequencies, estimate the mean age of the population.

(ii) Estimate the median age of the population.

(iii) Comment briefly on the age distribution of the population.

2. **(a)** Sampling and testing over a long period has shown that 10% of the components made on a particular production line are faulty. Write down the probability that a randomly chosen component is faulty. A sample of 10 components is chosen at random. Assuming that the number of faulty components is binomially distributed, find the probability that:

(i) none of the components in the sample are faulty;

(ii) exactly 1 of the components is faulty;

(iii) at least 2 of the components are faulty.

(b) The number of fish a man catches in an hour follows a Poisson distribution. On average he catches 3 fish per hour. Find the probability that:

(i) in the next hour he catches 3 fish;

(ii) in the next 2 hours he catches 6 fish or less.

3. The height of trees in a plantation is normally distributed with mean μ. Ten trees are chosen at random and their heights (in centimetres) are:

$$299,\ 292,\ 300,\ 292,\ 300,\ 297,\ 301,\ 296,\ 297,\ 300$$

(the data can be summarised by: $\sum x = 2974, \quad \sum x^2 = 884564$).

Calculate the sample mean, median and standard deviation.

Calculate a 90% confidence interval for the mean μ.

SECTION B

4. A food company produces cans of beans. The net weight of a can of beans is normally distributed with mean 524 gms and standard deviation 3 gms. A can of beans is randomly selected. Calculate the probability that the net weight is:

(i) more than 526 gms;

(ii) less than 523 gms;

(iii) between 523 gms and 526 gms.

A random sample of two cans of beans is chosen. What is the probability that the total net weight of the two cans is more than 1052 gms?

A machine fills the cans (assumed to be of constant weight) with beans. The machine breaks down and is repaired. In the next batch of production there is a suspicion that the mean net weight of the cans of beans has increased. A random sample of 50 cans of beans is taken and the mean net weight of the sample is 524.9 gms. Do these data provide evidence that the machine is over filling the cans? Formulate the hypotheses and calculate the p-value of the test used. State conclusions.

5. **(a)** An urn contains 7 red balls and 9 blue balls. Two balls are chosen at random from the urn *without* replacement. Let A be the event that the first ball chosen is red and let B be the event that the second ball chosen is blue.

(i) Write down the probability $P(A)$ and the conditional probability $P(B \mid A)$,

(ii) Using a tree diagram, or otherwise, find the probability $P(B)$ and the conditional probability $P(A \mid B)$.

(iii) What is the probability that the second ball is red?

If the two balls are chosen from the urn *with* replacement, write down $P(B \mid A)$ and $P(A \mid B)$.

(b) The mean intelligence quotient (I.Q.) of 14 randomly selected students from University A was 112 with a standard deviation of 8, while the mean I.Q. of 16 randomly selected students from University B was 107 with a standard deviation of 10. Assuming that the I.Qs are normally distributed at both universities, is there evidence to conclude that the mean I.Q. of students at University A is different from the mean I.Q. of students at University B at a 5% level of significance?

6. **(a)** A discrete random variable Y has the probability distribution:

y	-1	0	1	2
$P(Y = y)$	0.25	0.10	0.45	0.20

Calculate:

(i) the expectation $E(Y)$ of Y;

(ii) the variance $\mathrm{Var}(Y)$ of Y;

(iii) the expectation $E(Z)$ of Z where $Z = 3Y - 4$.

(b) A veterinary surgeon tested 100 cattle for an infection. There were 3 breeds of cattle involved. The table below summarises the data obtained.

	Breed of cow		
Test result	Friesian	Ayrshire	Jersey
Infected	10	20	5
Not infected	30	20	15

(i) Assuming independence between infection and the breed of a cow, draw up a table for the expected values.

(ii) Is there evidence, at the 5% significance level, of an association between the infection and the breed of cow?