UNIVERSITY OF ABERDEEN **SESSION 2003-04**

Degree Examination

ST2003 Mathematical Statistics

Tuesday 20 January 2004 (3pm to 5pm)

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 FOUR questions. All questions carry equal weight.

1. **(a)** A random sample of 978 computer users was surveyed regarding their future purchasing intentions. Two questions asked were whether their next purchase would be a laptop or desktop machine, and on the timing of their next purchase. The results are summarised in the following table.

	Purchase within 3 months	Purchase between 3 months and 1 year	Purchase after 1 year	Total
Laptop	34	156	258	448
Desktop	56	346	128	530
Total	90	502	386	978

One person is chosen at random from the 978. Let

$L = \{\text{person's next purchase will be a laptop}\}$

$A = \{\text{person will purchase within 3 months}\}$.

Find

(i) $P(L)$;

(ii) $P(L \cap A)$;

(iii) $P(L|A)$;

(iv) $P(A|L)$.

Explain

(v) why $P(L|A) \neq P(A|L)$;

(vi) whether A and L are independent;

(vii) whether A and L are mutually exclusive.

(*continued on next page*)

(b) The profit, X, on an investment for each £100 invested has the probability distribution $p(x)$ given below.

x	5	12	14	20
$p(x)$	0.1	0.4	0.3	0.2

(i) Verify that $p(x)$ satisfies the conditions for a probability distribution.

(ii) Find the expected value and standard deviation of X.

(iii) Let X_1, X_2 be profits made on £100 invested on two separate occasions. If X_1, X_2 both have the probability distribution above and are independent, what is the probability that $X_1 + X_2$ is less than £20?

2. **(a)** Footballers in a particular league have a probability of 0.2 of being injured at any given time. Use the binomial distribution to calculate the probability that in a squad of 30 players

(i) none is currently injured;

(ii) four or more players are currently injured;

(iii) between three and five players, inclusive, are currently injured.

(b) What assumptions are necessary for the binomial distribution to be appropriate in **(a)**? Discuss whether they are likely to hold in this example.

(c) The moment generating function of a binomial random variable, X, with n trials and probability of success p is $M(t) = [(1-p) + pe^t]^n$.

Use $M(t)$ to show that $E(X) = np$, $\text{var}(X) = np(1-p)$.

(d) Suppose that X_i is a binomial random variable with n_i trials and probability of success p, for $i = 1, 2, 3$. Use the moment generating functions of X_1, X_2, X_3 to find the distribution of $X_1 + X_2 + X_3$.

3. **(a)** The number of murders in a city has a mean of 0.5 per month. Assuming that this random variable has a Poisson distribution, find the probability that

(i) there are no murders in a single month;

(ii) there are no murders in a 6-month period;

(iii) there are two or fewer murders in a single month;

(iv) there are between two and four murders, inclusive, in a 3-month period.

(b) In a one-year period, 12 murders occur. Discuss whether you believe that this provides evidence of an increase in the murder rate.

(c) What assumptions are necessary for the Poisson distribution to be valid in **(a)** and **(b)**?

(d) Suppose that $X_1, X_2, \ldots, X_n$ are independent $N(\mu, \sigma^2)$ random variables. Let $Z_i = \dfrac{(X_i - \mu)}{\sigma}$, $i = 1, 2, \ldots, n$; $Y = \displaystyle\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2}$; $\bar{X} = \displaystyle\frac{1}{n}\sum_{i=1}^{n} X_i$; $S^2 = \displaystyle\frac{1}{(n-1)}\sum_{i=1}^{n}(X_i - \bar{X})^2$. What is the distribution of

(i) Z_i ?

(ii) Y ?

(iii) $\bar{X}$?

(iv) $(n-1)S^2/\sigma^2$?

(e) Use the results of **(d)**, together with the fact that a χ^2 distribution with ν degrees of freedom has mean ν, to show that $E(S^2) = \sigma^2$.

4. **(a)** A random variable X has probability density function (p.d.f.)

$$f(x) = \begin{cases} \frac{3}{x^4} & x \geq 1 \\ 0 & \text{elsewhere.} \end{cases}$$

(i) Verify that $f(x)$ is a valid p.d.f.

(ii) Find $P[3 \leq X \leq 6]$.

(iii) Find $\text{var}(X)$.

(b) Two groups of students are taught by different methods.

Let X be the score of a student in an examination.

In group A, X can be assumed to be normally distributed with mean 60 and standard deviation 15; for group B, the scores can be assumed to be normally distributed with mean 65 and standard deviation 10.

One student is selected randomly from each group.

Find the probability that

(i) the group A student scores more than 50 marks;

(ii) the group B student scores fewer than 70 marks;

(iii) the group A student scores more marks than the group B student.

(c) For the group B students in **(b)**, a random sample of 4 students is selected. Find the probability that

(i) the mean score of the 4 students is less than 70;

(ii) the mean score of the 4 students is less than the score of a single randomly selected group A student.

5. **(a)** A company recorded the number of absences through sickness for each day of the week. The results were:

Day	Mon	Tue	Wed	Thu	Fri	Total
Number of absences	50	32	24	30	44	180

Use a χ^2 goodness-of-fit test to decide whether absences are equally likely on each day of the week.

(b) The company in **(a)** decides to investigate whether the pattern of absences is different in summer and winter. It classifies the data above as follows:

Day	Mon	Tue	Wed	Thur	Fri	Total
Summer	28	16	4	12	20	80
Winter	22	16	20	18	24	100
Total	50	32	24	30	44	180

Test whether the distribution of absences across days of the week is independent of the time of year.

(c) The diameters of washers made on a production line can be assumed to be normally distributed with standard deviation $\sigma = 0.2$mm. A new process for manufacturing washers is introduced. For a random sample of 11 washers from the new process, the standard deviation is $s = 0.15$mm. Test whether the underlying standard deviation, σ, for the new process is still 0.2mm or whether it has been reduced from this value.

6. **(a)** A company uses 120-minute video tapes in its training courses. It is concerned that the mean length of tapes may be less than the nominal 120 minutes. It tests a sample of 9 tapes and finds the following times (in minutes):

121 117 116 117 120 118 118 115 120

The sum and sum of squares of these data are $\Sigma x = 1062$, $\Sigma x^2 = 125348$.

(i) Find a (two-sided) 95% confidence interval for μ, the mean length of tapes in the population from which the sample is drawn.

(ii) What assumptions have you made in calculating the confidence interval in (i)?

(iii) Without doing any further calculations, discuss whether the company's concerns are well-founded.

(iv) Suppose that, from previous data, it can be assumed that the population standard deviation, σ, of lengths of tape is $\sigma = 2$ minutes. Find a 95% confidence interval for μ given this information, and compare it with the interval in (i).

(b) Suppose that $X_1, X_2, \ldots, X_n$ are independent $N(\mu, \sigma^2)$ and that σ^2 is known. Show that the maximum likelihood estimator of μ is $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.