EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY
(formerly the Examinations of the Institute of Statisticians)

HIGHER CERTIFICATE IN STATISTICS, 1996

Paper II : Statistical Methods

Time Allowed: Three Hours

Candidates should answer **FIVE** *questions.*

All questions carry equal marks.

Graph paper and Official tables are provided.

Candidates may use silent, cordless, non-programmable electronic calculators.

Where a calculator is used the **method** *of calculation should be stated in full.*

Note that $\binom{n}{r}$ *is the same as* ${}^{n}C_{r}$ *and that* $\ln$ *stands for* $\log_e$.

1. (a) In tests of hypotheses relating to measures of location, there are situations in which (i) the Normal distribution, (ii) the *t* distribution, and (iii) neither the Normal nor the *t* distribution may be used. Explain this and illustrate your answer with simple examples.

(b) Two types of gas boiler, *X* and *Y*, used in central heating are being checked for the dust deposits left during operation. Random samples of each type of boiler were compared under identical operating conditions. The quantities of dust deposited in each boiler over the test period were as follows (grams):

X:	25,	31,	13,	39,	43,	37,	18,	36,	26,	32.
Y:	27,	32,	18,	22,	17,	8,	13,	23,	20.	

Using an appropriate statistical test, investigate the hypothesis that, under the given operating conditions, the deposits from boiler *X* are, on average, at least 5 gm more than those from *Y*.

State the assumptions necessary for your test to be valid. Find a 95% confidence interval for the difference between the mean deposits from the two boilers.

2. (a) A quality control department is concerned that a new method for filling cans with soup might increase the standard deviation of 'fill' above the target of $\sigma = 2$ grams. A simple random sample of $n = 10$ cans gave the following results in grams:

404, 407, 397, 403, 402, 410, 405, 396, 406, 400

Test the hypothesis H_0: $\sigma = 2$ grams against a suitable alternative and explain your results. (You may assume that the weights are independently Normally distributed.)

(b) A manufacturer of bottled water is assessing two filling machines to determine which produces the more regular fill. Each machine is set to produce an average one litre of water, and the actual quantities in millilitres produced in a number of experiments are as follows:

Machine A	998,	1000,	1004,	997,	1008,	1000,	1005,	1007,	1001,	1003,
	1004,	1015,	1002,	1006,	999.					
Machine B	998,	998,	1001,	1000,	1005,	993,	998,	993,	1004,	996,
	1002,	998,	997,	997,	1005,	1005,	1003,	998,	1002,	1000.

Carry out a statistical test and advise the manufacturer accordingly.

Turn over

3. A Machine Shop produces specialised machine tools. It is hypothesised that more faulty machine tools are produced later in the working day when operatives are tired. The number of faulty machine tools produced in the last month is given below, according to the hour of manufacture.

Working Hour:	1	2	3	4	5	6	7	8
Faulty Items:	2	4	3	3	3	7	8	10

Investigate the hypothesis that the number of faulty machine tools comes from a population having a uniform distribution, using:

(i) a chi-squared goodness-of-fit test;

(ii) a Kolmogorov-Smirnov test;

and comment on the comparison of your results.

4. (a) Write down the linear (additive) model used for a two-way analysis of variance with fixed effects. Explain what each term in the model represents, and state any assumptions required for the analysis to be valid.

(b) A supermarket chain is introducing a new hard cheese and has experimented with its price and the packaging in twelve comparable stores over the 4-week test period. The unit sales from the test are given below.

Unit sales during 4-week test

		Packaging			
		A	*B*	*C*	*D*
Price	Low	71	73	76	76
	Medium	62	68	65	71
	High	59	64	61	62

Carry out a suitable analysis of these data, and write a report on your findings for a manager who is untrained in Statistics.

5. (a) A brewery is investigating the quality of river water in two areas, *R* and *S*. Ten independent samples were taken, five from each area. The results of the analysis are given below, with the quality (1 = best, 10 = worst) stated in rank order.

Quality of water:	1	2	3	4	5	6	7	8	9	10
Area of sample:	*R*	*R*	*S*	*R*	*R*	*S*	*R*	*S*	*S*	*S*

Analyse these data using a suitable statistical test, justifying your choice, and state your conclusions in plain English.

(b) In a Health Club, two groups of women took part in a 'Fitness Programme', one using 'Aerobics', the other 'Circuit Training'. The fitness of each person was approximately the same at the outset and the allocation to the groups was random. After four weeks the fitness of each woman was measured. The improvement in fitness, given in coded units, after four weeks was as follows:

Aerobics:	16,	8,	19,	20,	14,	17,	13,	21,	12,	10
Circuits:	14,	12,	14,	13,	15,	11,	13,	14,	12,	12

Use a suitable non-parametric test to assess the results, stating any assumptions you make and reservations you may have about the validity of this analysis.

6. A random sample of trains leaving a station was taken, and the numbers of passengers on each train was noted as follows.

Number of passengers	Number of trains
0 – 9	28
10 – 29	20
30 – 49	37
50 – 74	62
75 – 99	34
100 and over	19

(a) Draw a histogram of these data and estimate appropriate measures of central tendency (average) and dispersion.

(b) What information do the data and your statistics show?

(c) Find 95% confidence limits for the percentage of trains which had 100 or more passengers, and state any assumptions made.

7. (a) In a trial with hay-fever sufferers, 40 patients were randomly allocated to one of two medicines with the following results.

Experience of symptoms

		Relief	*No relief*
Medicine	*A*	15	5
	B	7	13

Apply a chi-squared (χ^2) test to these data and explain your results. What is Yates's correction, and why is it sometimes used in such tests?

(b) A sample of 90 producers of non-alcoholic beverages is taken, and it is discovered whether they advertise on television and/or in the press. The results are as follows:

	Advertise on television	
Advertise in the press	Yes	No
Yes	36	10
No	28	16

Apply McNemar's test to the above results.

A conventional 2 x 2 χ^2 test of the above results, without using Yates's correction, gives a test statistic of 2.34. How does any difference between the outcomes of the two tests arise?

8. Use examples to explain and distinguish between the following pairs of concepts:

(a) Type 1 and Type 2 errors.

(b) Paired *t*-tests and unpaired *t*-tests.

(c) The 1% level of statistical significance and the 5% level.

(d) Rank correlation and product moment correlation.