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

HIGHER CERTIFICATE IN STATISTICS, 1997

CERTIFICATE IN OFFICIAL STATISTICS, 1997

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. The table below contains some data about hospital patients.

Descriptive statistics of 1383 hospital stays at Hospital del Mar, Barcelona, in 1988 and 1990		
Variable	*1988*	*1990*
Total number of stays	750	633
Mean age at admission *Standard deviation of age*	53.4 years 19.7 years	55.3 years 19.5 years
Males *Females*	349 401	321 312

(Source: S.J. Gange et al, Applied Statistics, No 3, 1996)

You may take these patients to be simple random samples of admissions to this hospital in 1988 and 1990. Use appropriate statistical tests to examine whether, between 1988 and 1990:

i) the mean age at admission increased;
ii) the sex ratio changed.

In each case, if you reject the null hypothesis of no change, estimate the change and provide a 95 per cent confidence interval for your estimate.

2. (a) State and explain *the linear, fixed-effects additive model for one-way analysis of variance.*

(b) The time taken for a medicinal tablet to dissolve is important to the pharmaceutical scientist. The data below show the effect of four different storage conditions on the time it takes, in seconds, for the first 50% of each tablet in a sample to dissolve in water.

					Sum	*Sum of squares*
Storage 1	19 22	28	21 19	19	128	2792
Storage 2	21.5	20.5	18.4	19.0	79.4	1582.06
Storage 3	22.0	24.9	22.7	21.1	90.7	2064.51
Storage 4	19.0	24.0	17.5		60.5	1243.25

(Source: M.J. Crowder, Applied Statistics, No 3, 1996)

(i) Carry out a one-way analysis of variance of these data, state the assumptions you have made and explain what you conclude as a result of your analysis.
(ii) How would your conclusions in this case be affected if you were now told that the measurements in sample 1 had been rounded?

Turn over

3. Twenty employees of a company were selected for computer training, and divided into two groups of ten. As numeracy was thought to be relevant, the employees were matched on this variable before the training course began and assigned to groups randomly. Different training methods were used for each group. After the training course was completed, an aptitude test was given, with the following results.

Matched Pair	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
First group	59	63	65	61	58	55	68	66	56	54
Second group	62	71	70	60	57	80	67	69	75	64

Test the null hypothesis that the training method does not affect aptitude test scores using (i) the sign test, and (ii) the Wilcoxon matched-pairs signed-ranks test. State and comment on your findings. Under what circumstances would a parametric test have been more appropriate?

4. The table below shows results from a clinical trial of antibiotics in controlling stomach ulcers, giving the number who did or did not recover within four weeks under each of two treatments. (Patients were allocated to Drug 1 or Drug 2 at random.) Patients who dropped out of treatment because of side-effects were excluded in (A), but included as 'not recovered' in (B).

	(A). Without drop-outs		**(B). With drop-outs**	
	Drug 1	*Drug 2*	*Drug 1*	*Drug 2*
Recovered	11	13	11	13
Not recovered	7	2	18	13

(Source: D.J Hand, Journal of the Royal Statistical Society, Series A, Part 3, 1994.)

(i) Use Fisher's exact test to analyse the data in Table (A), and a chi-square test for Table (B). Explain clearly in each case what you conclude from your test.

(ii) What is the effect of including or excluding the drop-outs from this analysis?

5. Explain the following statistical terms:

(i) *The Central Limit Theorem;*
(ii) *Two-way analysis of variance;*
(iii) *Degrees of freedom;*
(iv) *Confidence intervals.*

6. A machine is supposed to produce cartons with a mean net weight of at least 400 grams and a standard deviation not exceeding 8 grams. A simple random sample of the net weights (grams) of 16 cartons is as follows:

408	394	403	383	402	383	395	392	404	394
382	402	395	402	389	410				

Test whether the two production criteria are being met. Explain your conclusions in plain English, including any assumptions on which your analysis depends. Determine whether your conclusions would have been the same if the sample size had been four times larger, with the same sample mean and standard deviation?

7. A factory manager wishes to know whether or not the number of rejects produced by an industrial process within a set period follows a Poisson distribution. The previous month's data, given below, are thought to be typical:

Number of rejects	0	1	2	3	4	5	6 or more
Frequency	38	49	43	17	11	2	0

(i) For what reasons might the rejects follow a Poisson distribution?

(ii) Test the hypothesis that the distribution of rejects is Poisson, and explain your result to the factory manager.

8. (a) Using examples to illustrate your answer, discuss the uses made of *the F distribution* in statistical methods.

(b) New working practices have been introduced into a factory and its manager is concerned that, as a consequence, the variation in the time taken to complete a standard task (in minutes) might now differ between men and women. Times taken in random samples of their work are given below.

Men	21	23	12	18	20	27	23	25	26	10	14	19
Women	18	24	17	21	18	22	18	19	23	24		

Is there any statistical evidence that male times are more variable than female? Explain your conclusion, giving a confidence interval for the ratio of the variances, stating any assumptions which you made.