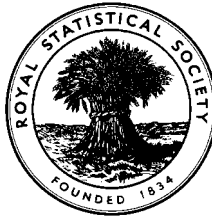# EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY
(formerly the Examinations of the Institute of Statisticians)

HIGHER CERTIFICATE IN STATISTICS, 1998

CERTIFICATE IN OFFICIAL STATISTICS, 1998

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) State the linear fixed effects additive model used for a two-way analysis of variance. Explain clearly what each term in the model represents and state any assumptions required for the analysis to be valid.

   (b) An experiment was performed to determine the relative effects of three different soil preparations on the first-year growth of pine seedlings. Each soil preparation was randomly applied to one of three plots at each of four separate locations. On each plot 20 seedlings were planted and the following table shows the average first-year growth (in cms) of the seedlings on each plot.

Soil preparation	1	2	3	4
A: no preparation	9	16	12	10
B: burning	12	18	16	13
C: fertilisation	10	13	14	11

   Carry out a suitable analysis of these data. Explain clearly your conclusions and comment upon the usefulness, or otherwise, of using a randomised block design for this experiment.

2. The times in minutes for a random sample of 70 factory workers to complete a standard task were summarised as follows:

Time in minutes	**Number of workers**
< 10	0
≥ 10 but <11	3
≥ 11 but <12	7
≥ 12 but <15	33
≥ 15 but <18	18
≥ 18 but <20	9
≥20	0

   (i) Construct a histogram of these data and find approximate summary statistics to describe the data. What do the data and your statistics reveal about the distribution of the number of minutes it takes to complete this task?

   (ii) Construct a 95% confidence interval for the mean number of minutes to complete the task and state any assumptions which you make.

Turn over

1

3. (a) A manufacturer of industrial light bulbs wishes to control the variability in the length of life of the bulbs so that its standard deviation σ is less than or equal to 150 hours. A random sample of 10 bulbs was taken and tested in the laboratory giving the following results in hours:

2100, 2302, 1951, 2415, 2067, 1911, 2149, 2489, 2083, 2124

Test the hypothesis H_0: $\sigma \leq 150$ hours against a suitable alternative. Explain your results and state any assumptions which you made.

(b) A chemical manufacturer using two production lines has made slight adjustments to the second in an attempt to reduce the variability in the levels of impurities in the chemical produced. Twelve randomly selected batches of chemical from each process were analysed and the level of impurities found to be as follows:

Process 1	2.12	2.45	2.43	2.51	2.52	2.44	2.56	2.51	2.41	2.46	2.43	2.38
Process 2	2.46	2.45	2.46	2.44	2.55	2.56	2.55	2.36	2.50	2.52	2.48	2.42

Using an appropriate statistical test, investigate whether the manufacturer has been successful in reducing the process variability for process 2. Explain your conclusions stating any assumptions which you made.

4. (a) A pharmaceutical company needs to determine whether a new drug alters blood pressure. Twelve male volunteers had their diastolic blood pressure measured, in appropriate units before and after receiving the drug, with the following results:

Patient	1	2	3	4	5	6	7	8	9	10	11	12
Before	120	124	129	118	141	128	140	133	126	130	136	127
After	125	126	137	117	144	128	146	131	127	135	136	131

Carry out a suitable analysis of these data stating any assumptions which you make.

(b) In a trial of anti-inflammatory drugs in the treatment of arthritis, 200 arthritic patients were allocated at random to receive one of two treatments. After one month each patient was asked to state whether their arthritis had improved. The results were as follows:

	Improved	Not improved
Treatment A	45	55
Treatment B	63	37

Apply a chi-squared test to these data and explain your results.

5. A random sample of 100 rats were each put into a maze until they were able to find the correct path. The number of attempts required by each rat was recorded as follows:

Number of attempts	1	2	3	4	5	6	≥ 7
Number of rats	56	27	13	3	0	1	0

(a) Explain why the number of attempts taken by each rat might follow a geometric distribution whose probability density function is $P(X=x) = (1-p)^{x-1} \cdot p$.

(b) Test the hypothesis that the distribution of the number of attempts needed for each rat is geometric, carefully explaining your conclusions.

Turn over

3

6. (a) Explain the use and importance of the *central limit theorem* in statistical inference.

(b) A study was conducted into the number and type of police emergency calls during standard 8 hour shifts in two districts of a large city. Random samples of shifts were selected from the police records for each district and the following descriptive statistics obtained.

	District 1	*District 2*
Total number of shifts	125	108
Mean number of calls	3.75	2.10
Standard deviation of numbers of calls	2.74	1.40
Number of shifts with a major emergency	26	15

(i) Construct a 95% confidence interval for the difference in the mean number of emergency calls per shift between the two districts and interpret your findings.

(ii) Test whether the proportion of shifts with major emergencies differs between the two districts.

7. One of the tasks undertaken in a particular laboratory is the measurement of the nitrogen content of various chemical preparations. It has been established that, when experienced workers repeatedly test the same preparation, the standard deviation of their measurements is 0.025 g%.

Two new laboratory workers A and B are given initial training. They are then each required to make ten repeat analyses of a test preparation, which is known to have an exact nitrogen content of 1.81 g%. The results (in g%) are as follows:

Worker A 1.73 1.75 1.80 1.83 1.79 1.88 1.85 1.79 1.78 1.80

Worker B 1.85 1.86 1.80 1.83 1.87 1.85 1.90 1.84 1.84 1.86

For each worker is there evidence that

(i) his/her experimental technique is more variable than the general laboratory standard?

(ii) his/her results are biased?

Using these results what comments would you make to the manager of the laboratory about the accuracy of measurements made by each worker?

8. (a) Discuss the advantages and disadvantages of using non-parametric rather than parametric methods in statistical analyses.

    (b) The weights of fish in two populations were compared by analysing the differences in the weights of a random sample of fish from each population matched by length. The weight differences in grams for 10 pairs of fish was as follows:

$$12, \ -13, \ -125, \ -120, \ -73, \ 2, \ 3, \ -147, \ -12, \ -4.$$

Explain why a parametric test would be unsuitable for comparing the weights of the fish from these two populations.

Analyse these data using a test which:

(i) ignores the magnitudes of the differences;

(ii) uses both the signs and magnitudes of the differences;

and comment on the comparison of your results.