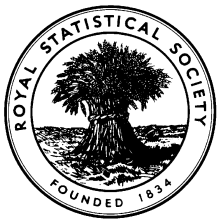
EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY
(formerly the Examinations of the Institute of Statisticians)



HIGHER CERTIFICATE IN STATISTICS, 2001

CERTIFICATE IN OFFICIAL STATISTICS, 2001

Paper II : Statistical Methods

Time Allowed: Three Hours

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

All questions carry equal marks.
The number of marks allotted for each part-question is shown in brackets.

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* $\log_e$.

This examination paper consists of 8 printed pages. This front cover is page 1. The reverse of the front cover, which is intentionally left blank, is page 2. Question 1 starts on page 3.

There are 8 questions altogether in the paper.

1. (i) State the *central limit theorem* and briefly explain its practical importance.

(6)

(ii) A fruit grower wishes to test a new spray that a manufacturer claims will reduce the amount of fruit lost due to damage by a certain insect. To test the claim, the grower sprays 100 trees with the new spray and 100 other trees with his standard spray. The yield of fruit was measured, in kg, for each tree. Summary statistics were as follows.

	New spray	*Standard spray*
Sample yield per tree	249	237
Sample variance	490	410

Construct a 95% confidence interval for the difference between the mean yields for the two sprays and interpret your findings.

(7)

(iii) The manufacturer of the new spray also claims that it can be used to prevent the loss due to insect damage of tender seedlings. To test this claim, the grower sprays 50 tomato seedlings with the new spray and his remaining 100 tomato seedlings with his standard spray. After six weeks, the fruit grower counts the number of healthy plants with the following results.

	New spray	*Standard spray*
Number of seedlings sprayed	50	100
Number of healthy plants at six weeks	40	70

Construct an approximate 95% confidence interval for the difference in the proportion of healthy plants six weeks after spraying between the two groups.

(7)

2. At a recent fund-raising event, a game was devised in which contestants were invited to predict the number of sixes that would be obtained from the simultaneous rolling of five fair six-sided dice. The organiser kept a record of the results from all 200 games played during the event and compiled the following table.

Number of sixes	0	1	2	3	4	5
Number of games	90	78	26	4	1	1

(i) Explain why the number of sixes obtained might be expected to follow a binomial distribution with parameters $n = 5$ and $p = 1/6$.

(5)

(ii) Test the hypothesis that the distribution of the number of sixes obtained is binomial with parameters $n = 5$ and $p = 1/6$, carefully explaining your conclusions.

(15)

3. A pharmaceutical company needs to determine whether a new drug successfully lowers blood cholesterol levels for patients with heart disease. To investigate this, 26 patients suffering from heart disease were randomised to receive the new drug or a placebo for a period of 6 weeks. At the end of this period, the patients had their blood cholesterol levels measured in suitable units with the following results.

Placebo 251 242 281 246 270 292 285 255 266 294 299

Drug 282 230 271 282 233 227 257 240 225 250 271 263 275 262 280

(i) Draw a dot-plot of these data and comment on the distribution of the observations in each group.

(4)

(ii) Using a suitable one-tailed non-parametric test, investigate whether the drug is more successful than the placebo in reducing blood cholesterol level in patients with heart disease.

(7)

(iii) It is suggested that a parametric test would be more appropriate to analyse these data. Repeat your analysis using a suitable parametric test, stating any assumptions necessary for this analysis to be valid.

(7)

(iv) Comment on the comparison of the results obtained in (ii) and (iii).

(2)

4. (i) A manufacturer of candles claims to be able to control the variability in the length of life of the candles so that the standard deviation σ (in minutes) is no greater than 25. Wishing to check this claim, a wholesaler takes a random sample of 8 candles from one day's large output and tests them in the laboratory, giving the following results in minutes.

725 741 706 711 735 697 745 752

Test the hypothesis H_0: $\sigma = 25$ against the alternative $\sigma > 25$. Explain your results and state any assumptions you made.

(10)

(ii) The candle manufacturer is considering whether making slight adjustments to the manufacturing process will reduce the variability in the length of life of the candles produced. Before making a decision, an experiment was conducted in which a number of candles were manufactured using each process and then tested in the laboratory, with the following results.

Process 1 724 743 705 711 736 699 745 752 740 705

Process 2 725 740 715 732 720 702 740 741 738 725

Using an appropriate statistical test, investigate whether the manufacturer has been successful in reducing the variability in the lifetime of the candles using process 2. Explain your conclusions, stating any assumptions that you made.

(10)

5. (a) A food tasting experiment was conducted in which a panel of judges were asked to taste two new mixtures, *A* and *B*, for a fruit flavoured soft drink and indicate which they preferred. The results were as follows.

Judge	1	2	3	4	5	6	7	8	9	10	11	12
Preference	*A*	*B*	*B*	*A*	*B*	*B*	*B*	*A*	*B*	*B*	*B*	*A*

Carry out a suitable analysis of these data to investigate which mixture would be more popular and comment on your results.

(8)

(b) In a psychological experiment to investigate the effects of stress on the ability to perform simple tasks, 90 volunteers were asked to perform a simple puzzle assembly task under normal conditions and under conditions of stress. Each subject was given three minutes to complete the task and on each occasion it was recorded whether or not they were successful. The order of the conditions under which each subject performed the task was determined at random. The results of the experiment are given in the following table.

		Normal conditions	
		Successful	*Unsuccessful*
Under stress	*Successful*	52	9
	Unsuccessful	20	9

(i) Apply McNemar's test to the above results.

(8)

(ii) A conventional 2×2 chi-squared test of the above results, without using Yates' correction, gives a test statistic of 3.26. How does any difference in the outcome of the two tests arise?

(4)

6. In a study of houseflies, a biologist measured the wing lengths of a random sample of 100 houseflies. The data obtained are given in the following table.

Wing length in millimetres	*Number of houseflies*
< 3.0	0
≥ 3.0 but < 4.0	14
≥ 4.0 but < 4.5	20
≥ 4.5 but < 5.0	32
≥ 5.0 but < 5.5	22
≥ 5.5 but < 6.5	12
≥ 6.5	0

(i) Draw a histogram of the above data and find approximate values of the median, mean and standard deviation of the data. What do the data and your statistics reveal about the distribution of the wing lengths of houseflies?

(14)

(ii) Construct a 95% confidence interval for the mean wing length of a housefly, stating any assumptions that you make.

(6)

7. (i) State and explain a linear model which can be used for a one-way analysis of variance. Explain clearly what each term in the model represents and state any assumptions required for the analysis to be valid.

(6)

(ii) A study was undertaken to investigate the water holding capacity of the soil in three different areas of woodland. In each area, a number of soil samples were collected randomly and sent to the same laboratory for analysis. The following table gives the water holding capacity (in millilitres per gram) of the soil samples collected in each area.

Woodland A	72 51 38 87 77 65 70 66 64 74
Woodland B	35 33 29 50 44 17 47 58
Woodland C	54 62 88 65 80 53

Carry out a suitable analysis of these data, stating the assumptions you have made and explaining what you conclude as a result of your analysis.

(14)

8. A specialist music school entered all its final year students for a national piano examination. The examination consisted of a written section and a practical section. The marks for each section together with the total mark achieved by each student are given in the following table.

Practical (out of 150)	*Written (out of 150)*	*Total*
106	107	213
127	110	237
100	97	197
125	120	245
108	115	223
124	114	238
111	106	217
96	104	200
115	105	220
134	100	234
145	145	290
107	103	210
140	133	273
105	98	203
110	96	206

(i) Draw a box and whisker plot for the total examination mark and hence comment on the distribution.

(8)

(ii) A scatter plot of the marks obtained in the practical and written sections is given below. The product-moment correlation coefficient between the two sets of marks is 0.747. Calculate the Spearman rank correlation coefficient between the two sets of marks. What do these coefficients indicate about the association between these two sets of marks?

(12)