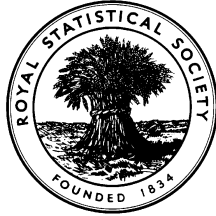**EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY**

(formerly the Examinations of the Institute of Statisticians)

HIGHER CERTIFICATE IN STATISTICS, 2002

CERTIFICATE IN OFFICIAL STATISTICS, 2002

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* $\log_{e}$.

1

1. (i) Compare the uses of the two-sample t test and the paired t test. Using examples to illustrate your answer, clearly explain why each method should be used in preference to the other in particular situations.

(6)

(ii) A farmer wishes to investigate whether the inclusion of a chicken food additive would affect the number of eggs laid by his chickens. To examine the impact of the food additive, the farmer selected a random sample of 24 chickens, all of similar age, and randomly allocated 12 to receive the normal food for three weeks and 12 to receive the normal food together with the food additive for three weeks. The number of eggs laid by each chicken during this period was recorded as follows.

Normal food only	Normal food and additive
9	18
16	17
13	15
11	14
14	17
18	16
14	14
12	11
12	13
15	18
14	16
12	15

Using an appropriate statistical test, investigate whether the inclusion of the food additive in the diet has any effect on the number of eggs laid by the chickens.

(14)

Turn over

2. (a) Explain the meaning of the following statistical terms used in hypothesis tests.

    (i)   Type I error.
                                                                                      (2)

    (ii) Type II error.
                                                                                      (2)

    (iii) Level of significance.
                                                                                      (2)

    (iv) Power.
                                                                                      (2)

    (b) According to medical experts, high sodium intake may be related to ulcers and stomach cancer. As a consequence, experts state that the human requirement for salt is only 220 milligrams per day. It is believed that this level of salt is exceeded by the salt in an average 20g packet of a popular snack food. A random sample of 20g packets of the snack food was selected, and each packet was analysed in the laboratory with the following results.

                  220, 218, 222, 215, 227, 229, 224, 217, 226.

    (i)   Perform a suitable statistical test to investigate whether, on average, the salt content of a single serving of this snack food exceeds the recommended daily intake.
                                                                                      (6)

    (ii) Suppose that the sample had consisted of 25 packets and the values of the sample mean and standard deviation had been found to be the same as the actual sample of 9 packets. Perform a second analysis, using this sample of 25 packets, and compare and comment on the conclusions of the two analyses.
                                                                                      (6)

3

3. **United Kingdom Trade in Goods and Services, £m, current prices, 1992-1999.**

	1992	1993	1994	1995	1996	1997	1998	1999
Exports								
To European Union	77633	82971	91983	107171	114650	115698	117804	121211
To North American FreeTrade Area	23552	27907	30495	31874	35339	38667	40552	44469
Other	42106	51200	56289	63367	70314	74961	67118	63969
Total	143291	162078	178767	202412	220303	229326	225474	229649
Imports								
From European Union	83563	89391	99283	112208	120135	120543	123727	130139
From North American Free Trade Area	21711	25397	27277	31293	35329	37911	39466	39812
Other	45393	53986	56770	61720	69028	70368	70236	74927
Total	150667	168774	183330	205221	224492	228822	233429	244878
Gross Domestic Product	608165	639356	677594	713980	756058	805402	851654	891106

(Source: United Kingdom Balance of Payments 2000 edn, table 9.3.
United Kingdom National Accounts 2000 edn, table 1.2.)

Note: Gross Domestic Product is the total final net output of all sectors of the economy, exclusive of property income from abroad and not taking account of the depreciation of capital assets.

Making use of some or all of the above data, write a report of appropriate length on some aspect or aspects of United Kingdom trade, 1992-1999. Your report should incorporate such diagrams and such statistics calculated from the table as you think appropriate.

(20)

4

Turn over

4. (a) State and explain a linear model that can be 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.

(6)

(b) A construction company makes concrete beams from cement mixed with gravel. The company wishes to compare the relative strengths of the concrete made from the different types of cement available. There are four different types of cement and three types of gravel. From each of the 12 different combinations of cement and gravel, a test beam was made and tested to destruction. The following table gives the breaking load, in coded units, for each test beam.

		Cement type			
		A	B	C	D
Gravel type	1	10	12	16	8
	2	14	15	18	10
	3	18	22	26	20

Carry out a suitable analysis of these data and write a report for the manager of the construction company who is not trained in statistics.

(14)

Turn over

5.   A random sample of 100 cherry trees in a particular orchard was selected and the yields in kilograms during the 2001 season were recorded for each tree. The data obtained are summarised in the following table.

Yield (kg)	Number of trees
≥ 10 but < 20	3
≥ 20 but < 25	10
≥ 25 but < 30	12
≥ 30 but < 35	14
≥ 35 but < 40	16
≥ 40 but < 45	18
≥ 45 but < 55	20
≥ 55 but < 65	6
≥ 65	1
Total	100

(i)   Draw a histogram depicting the above data.

(8)

(ii) State the modal class interval and estimate the mean, median and standard deviation of the yields (kg).

(8)

(iii) You are informed that the mean and median of the 100 individual values are really 37.5 kg and 38.0 kg respectively. Account for any differences between these figures and your answers in part (ii).

(4)

Turn over

6. A psychologist studying the behaviour of rats performed the following experiment. A random sample of 50 rats was taken. Each rat was placed in a box containing three doors, one leading to food the others leading to no food. The experiment was stopped as soon as the rat found the food. The number of doors tried, up to and including the one leading to the food, was recorded for each of the 50 rats. The results were as follows.

Number of doors tried	1	2	3	4	5	6	7	≥8
Number of rats	15	11	7	6	5	4	2	0

(i) Explain why the number of doors tried by each rat might follow a geometric distribution whose probability density function is

$$P(X = x) = (1-p)^{x-1} p$$

with parameter $p = 1/3$.

(6)

(ii) Test the hypothesis that the distribution of the number of doors tried by each rat is geometric with $p = 1/3$.

Explain your conclusions carefully.

How many degrees of freedom are there in the distribution you use in your test?

(14)

7

7. (i) Briefly discuss the advantages and disadvantages of using non-parametric rather than parametric methods in statistical analyses.

(6)

(ii) A dietician needs to examine the effectiveness of a new educational intervention programme. To study the intervention, a random sample of 14 overweight female patients was taken. Each patient was weighed immediately before and six months after receiving the intervention. The data obtained were as follows.

Patient	Weight before (kg)	Weight after (kg)
1	135	75
2	129	85
3	75	65
4	84	85
5	77	80
6	95	90
7	140	88
8	72	80
9	118	104
10	76	66
11	97	82
12	88	57
13	115	101
14	106	103

Investigate whether the new educational intervention programme has an effect on the weight of overweight women using

(a) a sign test,

(6)

(b) a Wilcoxon signed-rank test,

(6)

in each case with a significance level of 0.05, and comment on the comparison of your results.

(2)

8

Turn over

8. An estate agency wishes to investigate the selling prices of three-bedroomed houses in a popular suburb of a large city. To do this it randomly samples 20 sales of such houses from the previous month's records. The selling prices (in £s) were as follows.

| 68 000 | 85 950 | 78 000 | 95 000 | 99 950 | 85 000 | 120 000 | 122 950 | 97 500 | 74 950 |
| 105 000 | 75 000 | 95 000 | 115 000 | 119 950 | 82 950 | 79 950 | 95 000 | 85 000 | 108 000 |

(i) Draw a box and whisker plot of the selling prices and hence comment on the distribution.

(10)

(ii) The agency is surprised that the selling prices varied so widely. It takes a sample of 30 sales in an adjacent suburb and finds that the selling prices £x are summarised by

$$\Sigma x = 2\ 864\ 490 \quad \text{and} \quad \Sigma x^2 = 278\ 338\ 961\ 408 .$$

Test whether the house prices in the two suburbs were equally variable, stating any assumptions you make.

(10)