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

HIGHER CERTIFICATE IN STATISTICS, 1999

CERTIFICATE IN OFFICIAL STATISTICS, 1999

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* ${}^nC_r$ *and that* ln *stands for* $\log_e$.

This examination paper consists of 10 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. A journalist on a serious newspaper wants to write an article on the effects of income on spending patterns, contrasting low, middle and high income households' expenditure. Write briefing notes to bring out the information in the following table. You should perform such simple calculations and/or draw such diagrams as you think suitable.

(20)

Weekly household expenditure by gross income decile group
United Kingdom 1997 - 98,
£

Commodity or service	Lowest ten percent	2nd decile group	3rd decile group	4th decile group	5th decile group	6th decile group	7th decile group	8th decile group	9th decile group	Highest ten percent	All house-holds
Housing (net)	14.30	22.00	27.10	37.70	45.70	50.40	61.20	69.50	82.10	105.30	51.50
Fuel and Power	8.70	10.30	10.70	11.40	12.50	12.90	13.10	14.40	14.70	18.00	12.70
Food and non-alcoholic drinks	22.40	30.30	36.40	44.70	49.90	57.70	62.50	73.60	80.70	101.00	55.90
Alcoholic drink	3.50	4.50	5.90	8.50	10.70	14.70	16.30	18.20	23.30	27.80	13.30
Tobacco	4.40	5.10	5.60	6.90	5.40	6.80	8.00	7.00	6.60	5.50	6.10
Clothing and footwear	3.80	7.60	9.10	13.20	14.20	17.90	20.10	27.70	32.50	48.50	20.00
Household goods	8.50	12.20	15.50	18.40	21.60	26.10	31.20	33.00	44.20	58.40	26.90
Household services	5.50	7.20	8.60	13.00	13.60	21.70	16.50	21.80	26.20	44.80	17.90
Personal goods and services	3.80	4.90	6.80	8.20	9.60	14.20	14.30	17.10	18.70	27.80	12.50
Motoring	5.60	9.80	13.50	29.40	34.40	46.50	56.50	70.70	84.70	115.20	46.60
Fares and other travel costs	2.60	3.40	4.10	6.50	4.80	8.30	8.60	8.50	11.60	22.90	8.10
Leisure goods	4.40	4.90	6.90	12.20	13.00	15.60	18.60	25.00	24.70	38.20	16.30
Leisure services	7.90	11.60	16.00	24.20	27.60	37.00	40.20	47.30	69.10	107.20	38.80
Miscell-aneous	0.30	1.00	0.80	1.00	1.80	2.10	2.40	3.10	3.70	3.90	2.00
All expenditure groups	95.6	134.8	167.2	235.3	264.8	331.7	369.3	436.9	527.7	724.5	328.8

Source : Family Spending, 1997-98.

Turn over

2. The gender, birth weights in grams and week of pregnancy in which delivery occurred in a random sample of 24 new born babies born in a particular maternity hospital are given in the following table.

	Gender	*Birth weight (grams)*	*Week of delivery*
1	Male	3279	41
2	Female	2951	37
3	Female	2967	40
4	Male	2886	39
5	Female	2738	37
6	Male	2764	35
7	Male	3462	44
8	Male	3386	41
9	Female	3203	42
10	Female	2861	36
11	Male	2765	38
12	Male	3199	40
13	Female	3294	41
14	Female	2412	35
15	Male	3153	40
16	Female	3473	44
17	Male	2803	38
18	Male	2779	36
19	Male	2661	39
20	Female	2614	39
21	Male	3067	41
22	Female	2962	40
23	Female	2952	42
24	Male	2596	39

(i) Draw a box and whisker plot of the birth weight data and comment on the distribution. Note that, for the birth weight data, median = 2951.5, lower quartile = 2761.5, upper quartile = 3201.0.

(6)

(ii) Calculate a 95% confidence interval for the mean of the birth weight of a new born baby stating any assumptions which you make.

(7)

(iii) A baby is said to be a full term if it is delivered during or after the 40th week of pregnancy. Using an appropriate statistical test investigate whether the mean birth weight of full term babies differs between males and females.

(7)

3. (a) Explain the meaning of and the association between the following statistical terms used in hypothesis tests.

(i) Type I error and level of significance.

(4)

(ii) Type II error and power.

(4)

(b) For each of the 65 minor accidents occurring in a particular factory over a one year period in which only full five day weeks were worked, the safety officer noted the day of the week on which it occurred. The safety officer then compiled the following table.

Day of week	*Number of accidents*
Monday	17
Tuesday	10
Wednesday	12
Thursday	11
Friday	15

(i) Investigate the hypothesis that an accident is equally likely to occur on any day of the working week using an appropriate test.

(6)

(ii) In the following year, data were collected from all the factories covered by the manufacturer with the following results.

Day of week	*Number of accidents*
Monday	170
Tuesday	100
Wednesday	120
Thursday	110
Friday	150

Perform a similar test on this second set of data and compare and comment on the results of the two tests.

(6)

Turn over

4. (i) An investigator suspects that one particular form of senile dementia is associated with a reduction in cerebral blood flow. To investigate this he identifies 120 people with this form of senile dementia and 120 people of the same age without dementia. The investigator then tests to see whether or not each person has reduced cerebral blood flow with the following results.

		Cerebral blood flow	
		Normal	*Reduced*
Dementia	*No*	92	28
	Yes	80	40

Apply an appropriate test to these data and interpret your results.

(10)

(ii) Suppose that in the study described in (i) each case with dementia had in fact been paired with a control of the same age and gender with the following results.

		Controls	
		Normal cerebral blood flow	*Reduced cerebral blood flow*
Cases	*Normal cerebral blood flow*	74	6
	Reduced cerebral blood flow	18	22

Carry out a suitable analysis of these data and comment on the usefulness or otherwise of matching cases with controls in this study.

(10)

5. (i) Describe and explain a linear model 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 farmer is considering adding diet supplements to the food currently given to his dairy cows in an attempt to increase milk yields. In order to assist his decision, he sets up an experiment using a random sample of 24 of his cows who are assigned randomly to one of the following 3 diet regimes: standard diet, standard diet plus supplement *A*, standard diet plus supplement *B*. At the end of the three month study period the average daily milk yield (in pints) is recorded for each cow with the following results.

1: Standard diet	*2: Standard diet plus supplement A*	*3: Standard diet plus supplement B*
16	24	25
18	20	21
19	29	23
21	25	19
24	27	24
21	26	22
25	23	26
17	21	20

Carry out an analysis of these data, and write a report on your findings for the farmer.

(14)

Turn over

6. (a) Using one or two illustrative examples in each case, discuss the uses of the following distributions in tests of hypotheses relating to measures of location:

(i) the Normal distribution,

(4)

(ii) the *t*-distribution.

(4)

(b) A company wishes to study whether the presence of a supervisor has an effect on the productivity of its work force. A random sample of 12 workers was selected and their work rate recorded on two separate occasions; once when a supervisor was present and on another occasion when a supervisor was absent. The order of the two occasions was determined randomly for each worker. The results obtained were as follows.

Worker	*Supervisor present*	*Supervisor absent*
1	23	28
2	35	38
3	29	29
4	32	35
5	43	42
6	32	30
7	30	24
8	29	32
9	33	33
10	34	37
11	43	42
12	30	33

Carry out a suitable analysis of these data and write a short report of your findings for a manager.

(12)

7. (a) Discuss the advantages and disadvantages of using nonparametric rather than parametric methods in statistical analyses.

(5)

(b) The time taken for an individual to respond to different stimuli is important to behavioural psychologists. Twenty randomly-chosen adults were used in a psychological word association experiment, ten being assigned randomly to each of two stimuli. The reaction times, in seconds, were as follows.

Stimulus 1	*Stimulus 2*
0.5	1.3
3.0	4.0
4.4	5.3
0.5	3.5
1.1	0.8
0.7	1.3
2.4	2.3
1.0	1.0
1.8	2.6
1.6	1.8

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

(4)

Why would a parametric test be unsuitable for comparing reactions to these stimuli?

(4)

Using a suitable nonparametric test, investigate whether there is sufficient evidence to indicate a difference in reaction times for the two stimuli.

(7)

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

(6)

(b) A manufacturer of pharmaceutical products purchases one particular material from two different suppliers. A mean level of impurities in the raw material is approximately the same for each supplier, but the manufacturer is concerned about the variability in the level of impurities from shipment to shipment. If the percentage of impurities varies excessively for one source of supply, it can affect the quality of the pharmaceutical product. To compare the variation in percentage impurities for the two suppliers, the manufacturer selects 16 shipments at random from each supplier and measures the percentage of impurities in the raw material for each shipment. The data obtained are given in the following table.

Percentage of impurities

Supplier 1	*Supplier 2*
2.2	1.6
1.3	1.7
2.3	2.1
1.6	2.0
1.5	2.1
2.3	1.6
2.2	1.7
2.1	1.7
1.6	1.9
1.7	1.8
1.6	2.1
2.4	2.1
1.4	2.0
2.1	1.8
1.5	1.9
2.3	2.2

Using a suitable statistical test, examine whether there is sufficient evidence to indicate a difference in the variability of the impurity levels between shipments of the raw material for the two suppliers. Based on the results of your test, what recommendation would you make to the pharmaceutical company?

(14)