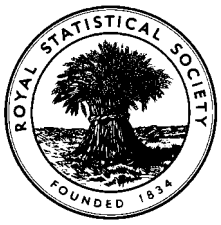
EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY
(formerly the Examinations of the Institute of Statisticians)



ORDINARY CERTIFICATE IN STATISTICS, 1998

Paper II

Time Allowed: Three Hours

There is no restriction on the number of questions that a candidate may attempt, nor on the order in which they are attempted. Candidates are not required to answer all the questions: they should answer as many as they can.

The number of marks allotted to each question is shown in brackets. The total for the whole paper is 100. A pass may be obtained by scoring at least 50 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.*

1. The following paragraph describes the number of punches thrown during the 1989 World Heavyweight Championship boxing match between Mike Tyson and Frank Bruno. Tyson won the fight at the end of the fifth round.

> The total number of punches thrown during the fight between Tyson and Bruno was 372. The greatest total number of punches thrown in a single round was 98 during the first round. The fewest was 55 during the fourth round. The greatest number of punches thrown in a single round by an individual boxer was 55. This occurred twice; with Bruno throwing that many punches in the first round, and Tyson throwing that number in the fifth round. Bruno threw the least number of punches in a round by either boxer, only managing 18 during round four. The least number of punches thrown in a single round by Tyson was in the third round, when he threw just four-ninths of all punches thrown in that round. The most equal round of all was the second round, where Bruno threw 42 of the total of 81 punches. Bruno threw an average of exactly 34 punches per round.
>
> *(Source of data:* The Independent *newspaper, 27 February, 1989)*

(i) From the information given, construct a frequency table (including marginal totals) of the number of punches thrown by each boxer in each round of their fight.

(9)

(ii) Use the table you constructed in (i) to answer the following questions:

(a) What was the average number of punches thrown per round by Tyson?

(b) What percentage of Bruno's punches were thrown during the first three rounds of the fight?

(c) What percentage of the last round's punches were thrown by Tyson?

(3)

Turn over

2. (a) An index number is a single figure which summarises a comparison between two sets of figures.

(i) Explain what is meant by a *chain based* index number.
(ii) State one advantage of using a chain based method rather than a non-chain based method in the calculation of a simple single-item price index.

(5)

(b) Table 1 shows the average prices of new dwellings (for which mortgages were approved) in the United Kingdom for each year from 1974 to 1982.

Table 1: The price of new dwellings in the United Kingdom, 1974-82.

Year	*Average price (£)*	*Year*	*Average price (£)*	*Year*	*Average price (£)*
1974	11300	1977	14800	1980	27200
1975	12400	1978	17700	1981	28000
1976	13400	1979	22700	1982	28500

(Source: The Open University, 1993)

(i) Use the graph paper provided to plot the rate of inflation in the prices of new dwellings in the United Kingdom for each year from 1975 to 1982.
(ii) Describe the change in the average price of new dwellings in the United Kingdom between 1974 and 1982.

(9)

3. (a) (i) Define the term p^{th} *percentile.*

(ii) A random sample of size 50 has been obtained from a continuous distribution; say how you would calculate the 10^{th} percentile.

(4)

(b) Table 2 shows the ages (in years) at which forty English kings and queens from William I to George VI died.

Table 2: Ages at Death of English Monarchs.

Monarch	*Age*	*Monarch*	*Age*	*Monarch*	*Age*	*Monarch*	*Age*
William I	60	Edward III	65	Edward VI	16	George I	67
William II	43	Richard II	34	Mary I	43	George II	77
Henry I	67	Henry IV	47	Elizabeth I	69	George III	81
Stephen	50	Henry V	34	James I	59	George IV	67
Henry II	56	Henry VI	49	Charles I	48	William IV	71
Richard I	42	Edward IV	41	Charles II	55	Victoria	81
John	50	Edward V	13	James II	68	Edward VII	68
Henry III	65	Richard III	35	William III	51	George V	70
Edward I	68	Henry VII	53	Mary II	33	Edward VIII	77
Edward II	43	Henry VIII	56	Anne	49	George VI	56

(Source: Gebski, Leung, McNeil and Lunn, 1992)

(i) Calculate the 25^{th}, 50^{th} and 75^{th} percentiles of the ages given in Table 2.

(ii) Use the graph paper provided to draw a boxplot *(box-and-whisker plot)* of these ages.

(11)

Turn over

4. Table 3 shows typical brain and body weights of twelve different animals.
The relationship between brain and body weight can be described approximately using the mathematical equation:

$$Brain\ Weight = A\,(Body\ Weight)^{B}$$

A and *B* can be estimated from data by least-squares linear regression, since by logarithmic transformation the above equation becomes:

$$\log(Brain\ Weight) = \log A + B\,\log(Body\ Weight)$$

Table 3: The Weights of Twelve Animals.

Animal	**Brain Weight (g)**		**Body Weight (kg)**	
	y	*log y*	*x*	*log x*
Mouse	0.023	-3.772	0.4	-0.916
Hamster	0.12	-2.120	1.0	0.000
Rat	0.28	-1.273	1.9	0.642
Guinea Pig	1.04	0.039	5.5	1.705
Rabbit	2.5	0.916	12.1	2.493
Cat	3.3	1.194	25.6	3.243
Goat	27.7	3.321	115	4.745
Sheep	55.5	4.016	175	5.165
Gorilla	207	5.333	406	6.006
Cow	465	6.142	423	6.047
Horse	521	6.256	655	6.485
Asian Elephant	2547	7.843	4603	8.434
Mean		2.325		3.671
Standard Deviation		3.714		2.927
Product-moment correlation between log y and log x = +0.994				

(Source: H.J. Jerison, 1973)

(i) Using the graph paper provided, draw a scatter-plot of *log y* (on the vertical axis) versus *log x*.

(6)

(ii) Using either the summary statistics given at the bottom of Table 3, or some other appropriate method, calculate the *least-squares linear regression* of log(Brain Weight) on log(Body Weight) to obtain estimates of *A* and *B*.

(7)

5. Applicants for a driving licence in the United Kingdom are required to perform both a *practical* test of their driving ability and a *written* test of their knowledge of driving theory. Each separate test is marked as either a 'pass' or a 'fail' according to the applicant's performance. An *overall pass* is obtained only if *both* tests are passed.

Mr Smith and Mr Jones both apply for a driving licence. The probabilities of obtaining a 'pass' in the practical test are 0.8 for Mr Smith and 0.7 for Mr Jones. It is estimated that Mr Jones has the same probability of failing the written test as he does of failing the practical test, but that Mr Smith has *twice* the probability of failing the written test as he does of failing the practical test. You may assume that passing the practical test is independent of passing the written test for both men.

(i) Construct a *Venn diagram* to show the events Mr Smith passes the practical test and Mr Smith passes the written test and indicate the probabilities of the four outcomes for Mr Smith (passes practical test only, passes written test only, passes both tests, passes neither test).

(6)

(ii) Who is more likely to obtain an overall pass, Mr Smith or Mr Jones ? (Full workings must be shown to achieve full marks for this question)

(7)

6. Table 4 shows the duration of 170 intervals between 171 successive eruptions of the Old Faithful geyser in Yellowstone National Park, USA. (Durations are measured in minutes, to the nearest minute.)

Table 4: Waiting times between eruptions of the Old Faithful geyser.

Length of Interval (minutes)	*Frequency*
40-49	9
50-59	34
60-69	19
70-79	53
80-89	46
90-99	8
100-109	1

(Source: Azzalini and Bowman, 1990)

Estimate the *mean* and *standard deviation* of these times (to 3 significant figures), showing your working.

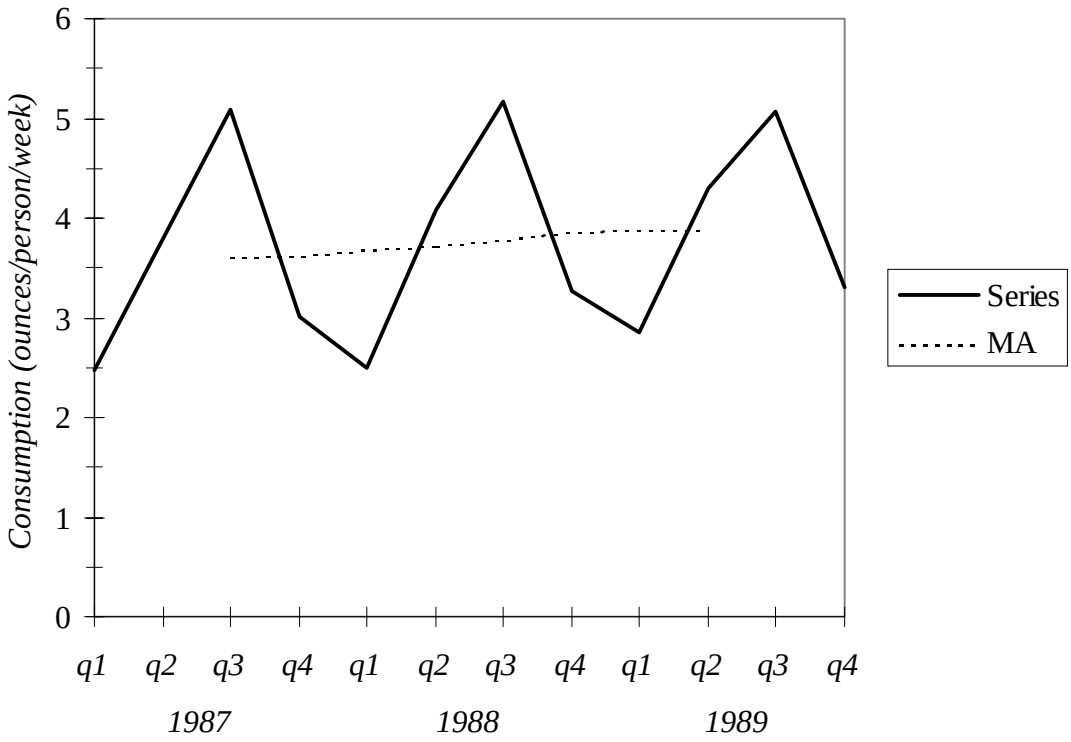
(9)

Turn over

7. (a) Describe the different circumstances in which it is appropriate to estimate the seasonal component of a time series using:
   (i) an *additive* model;
   (ii) a *multiplicative* model.

   Figure 1 and Table 5 show the consumption of fresh tomatoes in the United Kingdom between January 1987 and December 1989. Also included are *centred 4 point (simple) moving averages* of these consumption data.

   Figure 1: Tomato consumption in the United Kingdom, 1987-89.



(Question continued on next page)

Table 5: Tomato consumption in the United Kingdom, 1987-89.

Year	*Quarter*	*Consumption (ounces per person per week)*	*Centred 4 pt MA*
1987	1. Jan-Mar	2.47	-
	2. Apr-Jun	3.81	-
	3. Jul-Sep	5.09	3.60
	4. Oct-Dec	3.01	3.63
1988	1. Jan-Mar	2.49	3.67
	2. Apr-Jun	4.07	3.71
	3. Jul-Sep	5.16	3.79
	4. Oct-Dec	3.26	3.87
1989	1. Jan-Mar	2.86	3.88
	2. Apr-Jun	4.30	3.88
	3. Jul-Sep	5.06	-
	4. Oct-Dec	3.31	-

(Source: Ministry of Agriculture, Fisheries and Food)

Assuming an *additive* model for seasonality, estimate the seasonal component of tomato consumption.

(9)

(b) State two reasons why it may be more appropriate to calculate a *weighted*, rather than simple, moving average to describe the trend in a time series.

(4)

8. An investigator assesses two characteristics, Factor *A* and Factor *B* say, in each respondent in her sample survey. Factor *A* is a *nominal* variable and Factor *B* is an *ordinal* variable. Both variables can take *more than two* possible values. The investigator wishes to illustrate the relationship between Factor *A* and Factor *B* in a *single* graph.

(i) State the difference between *nominal* and *ordinal* variables. Give one example of each.

(4)

(ii) Describe the type of graph you would recommend the investigator to use. For example, give details of how this graph would be constructed, and which special features you would include in the graph in order to improve visual impact.
(Hint: If you wish, you may include a rough sketch in your answer to aid explanation.)

(7)