EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2009

MODULE 6 : Further applications of statistics

Time allowed: One and a half hours

Candidates should answer **THREE** *questions.*

Each question carries 20 marks.
The number of marks allotted for each part-question is shown in brackets.

Graph paper and Official tables are provided.

Candidates may use calculators in accordance with the regulations published in the Society's "Guide to Examinations" (document Ex1).

The notation log *denotes logarithm to base* ***e***.
Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

Note also that $\binom{n}{r}$ *is the same as* ${}^nC_r$.

This examination paper consists of 4 printed pages **each printed on one side only**.
This front cover is page 1.
Question 1 starts on page 2.

There are 4 questions altogether in the paper.

1. (i) Explain the purpose of *blocking* in designed experiments, and illustrate your answer with examples of situations where (a) blocks would be required and (b) blocks would not be necessary.

(5)

(ii) Two chemicals, A and B, are being assessed as potential constituents of a liquid for sterilising equipment used in food manufacture. An experiment is carried out with each chemical used at two levels ("high" and "low"). The four treatment combinations, given in the usual notation as (1), *a*, *b*, and *ab*, are run once on each of five days. A randomised block design is used in which the days form the blocks. At the end of each run, sterility is assessed on a scale 0 – 20, 0 being very bad and 20 excellent. The results *y* are given in the following table.

		Treatment combination				
		(1)	*a*	*b*	*ab*	*Block total*
Block	I	13	16	17	10	56
	II	12	12	13	11	48
	III	10	14	12	11	47
	IV	11	14	15	14	54
	V	12	16	18	14	60
	Total	58	72	75	60	265

$\Sigma y^2 = 3611$

(a) Copy and complete the following Analysis of Variance table.

(9)

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARE	*F* value
Blocks				
A				
B				
A $\times$ B				
Residual				
TOTAL				

(b) Report on the results, giving appropriate tables and/or diagrams, and suggest what further work should be undertaken to study the effects of the two chemicals A and B.

(6)

2. A sample of n items is drawn from a batch production process, and the batch is accepted if there are no faulty items in the sample. It is rejected if there are two or more faulty items. A further sample of n is taken if there is just one faulty item. The same decision procedure is used as often as necessary (and it may continue indefinitely).

(i) Suppose that $P(0)$ and $P(1)$ are the probabilities of 0 and 1, respectively, faulty items out of n. Find the probability that the batch is accepted after k steps of the decision process (for $k = 1, 2, \ldots$), and hence show that a batch is eventually accepted with probability $P(0)/\{1 - P(1)\}$. Find also the average number of batches inspected until the first one is rejected.

(13)

(ii) Now suppose that the proportion of faulty items in a batch is p, that the batch size is much greater than n and that items are drawn independently of one another. Calculate the acceptance probability and average number of batches specified in part (i) for the case $n = 100$, $p = 0.02$. Comment briefly on the results.

(7)

3. (a) In quality control of a continuous industrial process, CUSUM (cumulative sum) charts may be used. Describe how these charts are constructed, and discuss the advantages and disadvantages they may have compared with Shewhart charts for means.

(10)

(b) (i) Explain the purpose of *randomisation* in the design of experiments.

(6)

(ii) In randomised trials in medicine, *blinding* is also often used. Explain what this means and why it can be important.

(4)

Turn over

4. (a) Several graphical methods may be used to check the goodness of fit of (simple or multiple) linear regression models to a set of data in which $y_1, y_2, \ldots, y_n$ are the values of a response variable Y. Explain briefly what information can be deduced from each of the following five diagnostic plots.

(i) A dot plot of all the residuals. (2)

(ii) A plot of residuals against fitted values $\hat{y}_i$ ($i = 1, 2, \ldots, n$). (3)

(iii) A Normal probability plot of the residuals. (1)

(iv) A plot of residuals against one of the predictor variables used in the model. (2)

(v) A plot of the ith residual against the $(i - 1)$th residual in a set of data collected in the order $i = 1, 2, 3, \ldots, n$. (2)

(b) x_1, x_2, x_3, x_4 are four possible predictor variables to use in a model for explaining a response variable Y. Suppose there are 13 observations of (Y, x_1, x_2, x_3, x_4) available, with the total (corrected) sum of squares for Y equal to 2715.76. Some of the results of carrying out all possible regressions are shown in the following table.

Use backwards elimination to select a satisfactory model, giving reasons for your choice.

(10)

Variables included in model	*Residual sum of squares*
x_1 x_2 x_3 x_4	47.86
x_1 x_2 x_3	48.11
x_1 x_2 x_4	47.97
x_1 x_3 x_4	50.84
x_2 x_3 x_4	73.82
x_1 x_2	57.90
x_1 x_3	1227.07
x_1 x_4	74.76
x_2 x_3	415.44
x_2 x_4	868.88
x_3 x_4	175.74
x_1	1265.69
x_2	906.34
x_3	1939.40
x_4	883.87