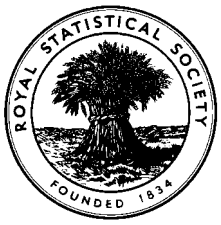
EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY
(formerly the Examinations of the Institute of Statisticians)



GRADUATE DIPLOMA IN STATISTICS, 1998

Options Paper

Time Allowed: Three Hours

This paper contains four questions from each of six option syllabuses. Each option syllabus is one Section.

Section	*A:*	*Statistics for Economics*
	B:	*Econometrics*
	C:	*Operational Research*
	D:	*Medical Statistics*
	E:	*Biometry*
	F:	*Statistics for Industry and Quality Improvement*

Candidates should answer **FIVE** *questions chosen from* **TWO SECTIONS ONLY.**

Do **NOT** *answer more than* **THREE** *questions from any* **ONE** *Section.*

ANSWER EACH SECTION IN A SEPARATE ANSWER BOOK.

Label each book clearly with its Section letter and name.

All questions carry equal 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.*

SECTION A - STATISTICS FOR ECONOMICS

A1. Statistics of the exports of goods and services, £m at 1990 prices for each year between 1975 and 1996 inclusive are collated from Table 1.3 of *United Kingdom National Accounts*, 1997 edition, and are denoted by X. Their natural logarithms, correct to two decimal places, are denoted by y, so that $y = \log_e X$. The variable t is defined as taking the values −10.5, −9.5, ...10.5 over the period, so there are 22 pairs of observations (t,y). It is found that $\sum t=0$, $\sum t^2=885.5$, $\sum y=256.07$, $\sum y^2=2981.6757$ and $\sum ty=31.165$.

(i) Estimate α and β in the model $y=\alpha+\beta t+u$, where u has the usual properties, by ordinary least squares. Obtain the coefficient of determination (r^2) and estimate σ (the standard deviation of u) and the standard errors of the ordinary least squares estimators of α and β.

(ii) Use your estimated model to predict y for 1997.

(iii) Two different standard errors are associated with such predictions, and hence two different confidence intervals. Explain what these intervals relate to, and calculate them for your prediction.

(iv) What assumptions were necessary in making your estimates of the standard errors of the coefficients? How could you examine their validity?

(v) In numerical terms, what model giving X as a function of t is equivalent to the model you fitted in part (i) of this question? What value of X for 1997 is predicted by your estimated model?

(vi) The model $X = a+bt+u$ may be estimated as

$$X = 116558 + 4158.3\,t, \quad r^2 = 0.928, \quad s = 7712$$
$$\qquad\quad (1644) \qquad (259.2)$$

and gives a prediction of 164378 for 1997. Compare the models and account for the difference between the predictions. (You are not expected to obtain standard errors etc of the prediction using the model $X = a+bt+u$.)

Turn over

A2. In order to study the operation of the labour market in the twelve Government Office Regions (GOR) into which the United Kingdom is divided, the following data relating to Spring 1996 are collected:

GOR	% of population in employment	Average weekly hours of work of f/t employees	Average weekly earnings (£)	ILO unemployment rate (%)	Claimant unemployment rate (%)
England					
North East	51.9	43.7	347.7	10.8	10.6
North West	55.2	43.5	367.8	8.4	8.0
Yorkshire and the Humber	56.5	44.0	350.7	8.1	8.0
East Midlands	59.4	44.3	352.9	7.4	6.8
West Midlands	56.9	44.0	360.1	9.2	7.4
Eastern	61.1	44.7	382.3	6.2	6.1
London	56.9	44.0	514.3	11.3	8.9
South East	61.7	44.6	412.7	6.0	5.4
South West	58.1	43.8	364.8	6.3	6.2
Wales	52.3	43.8	345.5	8.3	8.2
Scotland	56.0	43.3	363.6	8.7	7.9
Northern Ireland	53.3	42.4	337.4	9.7	10.9

Source: Regional Trends, 1997 edn

Note: ILO unemployment rates are those according with international practice, while claimant unemployment rates (i.e. the percentage of the labour force out of work and claiming benefit) are those customarily in use in the United Kingdom.

The data are entered into a computer using the Minitab statistics package, into columns 1-5. They are analysed as shown at the end of this question.

(i) What two analyses have been carried out on the computer?

(ii) For some sets of data the two analyses in (i) would be likely to give rather different results. Explain, with the aid of diagrams, how this might happen.

Question continued on next page

(iii) Test the correlation between the two unemployment rates for statistical significance. Obtain and test for statistical significance the partial correlation coefficient between the two unemployment rates with respect to percent in employment. State any assumptions upon which your tests are based. Explain and discuss your results.

(iv) Draw a scatter diagram in which average weekly hours of work are plotted against percent in employment.

(v) Write an account at suitable length of the more notable features of the data as brought out by the computer analysis and your calculations, etc, for publication in a serious newspaper.

Turn over

A3. (a) (i) What is meant by stratified random sampling?

(ii) Explain verbally (rather than mathematically) why it is likely that a stratified sample with a uniform sampling fraction in each stratum (i.e. a proportional allocation of the sample among strata) will have a smaller standard error than a simple random sample of the same total size taken from the entire population. Illustrate your answer with practical examples, preferably from samples known to you.

(iii) Is it ever possible to reduce the standard error of a sample further, by using a non-proportional allocation? Why is this seldom done in practical social surveys, e.g. of family expenditure?

(b) A population of size N=10,000 is divided into three strata, of sizes 5,000, 3,000 and 2,000. A sample of n=2,000 is taken by simple random methods without replacement within each stratum with the results shown:

Stratum	N_i	n_i	Σx_i	Σx_i^2
A	5000	800	8016	83541
B	3000	600	8989	139740
C	2000	600	11924	246738
	10000	2000		

Σx_i and Σx_i^2 denote the sum of the sample observations and the sum of their squares in each stratum.

Estimate the mean in the population as a whole and assign a 95 per cent confidence interval to your estimate.

Comment on the allocation of the total sample among the strata.

A4. **Male population and deaths 1995**

	United Kingdom		**Northern Ireland**	
	Population (thousands)	*Deaths*	*Population (thousands)*	*Deaths*
Under 1 year	376	2595	12	93
1-4	1589	447	52	19
5-14	3861	702	135	33
15-24	3825	2925	130	135
25-34	4793	4748	125	142
35-44	3984	6754	105	176
45-64	6559	51712	160	1362
65-74	2330	85459	54	2040
75-84	1147	103324	26	2429
85 and over	263	52056	5	1053
	28727	310722	805	7482

(Source: Annual Abstract of Statistics, 1997 and Office of National Statistics.)

Note: The population data are estimated mid-year resident populations. Details may not add to totals because of rounding.

Find the crude death rates in (a) the United Kingdom and (b) Northern Ireland (which is part of the United Kingdom).

Explain what is meant in the context of area death rates by

(i) direct standardisation.
(ii) indirect standardisation.

Using the United Kingdom as the standard, calculate *either* the directly standardised death rate *or* the indirectly standardised death rate for Northern Ireland.

Use these data and your results to illustrate an account of the importance of age standardisation in relation to the Northern Ireland death rate.

Turn over

SECTION B - ECONOMETRICS

B1. Based on series of 26 annual observations from 1971 to 1996, the following model was estimated by least squares in an attempt to explain a country's volume of imports:

$$Log\ \hat{Y}_t = \underset{(0.440)}{0.504} + \underset{(0.11)}{0.26}\ Log\ X_{1t} - \underset{(0.045)}{0.108}\ LogX_{2t} + \underset{(0.122)}{0.626}\ Log\ Y_{t-1}$$

(Figures in brackets below parameter estimates are the associated estimated standard errors.) In this equation:

Y_t =Volume of imports into the country
X_{1t} =Real gross national product of the country
X_{2t} =Ratio of an index of import prices to an index of domestic prices in the country.

(i) Assuming that the model specification is adequate, write a report summarising what can be learned from this estimated model.

(ii) In 1985 the country instituted a set of reforms which loosened various barriers to free trade. Explain how the regression model can be modified to take account of the impact of these reforms.

(iii) Why is it important to test for autocorrelated errors in models such as this? Outline an appropriate test.

B2. Consider the regression model

$$y_i = \beta x_i + \varepsilon_i \quad (i = 1, ..., n)$$

where

$$E(\varepsilon_i) = 0 \quad E(\varepsilon_i^2) = \sigma^2 x_i^2 \quad E(\varepsilon_i\, \varepsilon_j) = 0 \text{ if } i \neq j\ .$$

(i) Derive the generalised least squares estimator, $\hat{\beta}$, of β, and find the variance of $\hat{\beta}$.

(ii) Let b denote the ordinary least squares estimator of β. Find the variance of b and prove that b is less efficient than $\hat{\beta}$.

(iii) Let

$$s^2 = (n-1)^{-1} \sum_{i=1}^{n} (y_i - bx_i)^2\ .$$

Find $E(s^2)$, and discuss the testing of hypotheses about β based on the usual ordinary least squares methodology.

B3. Write notes on three of the following, including a discussion of their relevance to practical econometric model building.

(i) Multicollinearity

(ii) Tests for heteroscedasticity in the errors of a linear regression model

(iii) Consistency of estimators of the parameters of an equation that is part of a set of simultaneous econometric equations.

(iv) Measurement errors in the independent variables of a linear regression equation.

(v) Duration models.

B4. A financial economist has 20 years of monthly data on spot prices and one month forward prices for the US dollar - Japanese yen exchange rate. The requirement is to construct a model of the evolution of this pair of related time series. It is suspected that the generating processes for the individual series each have a single unit autoregressive root. If this is the case, economic theory suggests the possibility of cointegration between the series. Given these factors, describe as fully as possible how you would carry out a practical analysis, and discuss the rationale behind any models and statistical tests you would employ.

Turn over

SECTION C - OPERATIONAL RESEARCH

C1. Consider the M | G | 1 queue where the arrival rate is λ, the mean and the variance of the service time distribution are $1/\mu$ and σ^2 respectively, and $\rho = \lambda/\mu$ is the *traffic intensity*.

(i) Write down the formula for the average queue length L_q .What is the name of this formula, and under what conditions does it hold?

(ii) Customers arrive at random at a rate of 25 an hour at an information desk in a railway station. The desk is staffed by a single server. Currently the mean service time is 2.2 minutes and the variance of the service time is 2.1 $(\text{minutes})^2$. Assuming that the times to deal with separate enquiries are independent, what is the expected queue length?

The station managers are not satisfied with this situation and have two choices for improving it. They could either retrain the server, which would reduce the mean service time to 1.7 minutes but would increase the variance to 2.5 $(\text{minutes})^2$, or introduce a computer system which would only reduce the mean service time to 1.9 minutes but would reduce the variance to 0.1 $(\text{minutes})^2$.

Which option would you recommend? What general point does this illustrate about improving the performance of queuing systems?

(iii) A factory employs one engineer to keep a very large number of machines in running order. The number of machines is so large that it may be treated as an infinite population from which calls for repair occur at random at a rate of 1 per hour. When a machine breaks down, it may require a minor repair (with probability 0.9) or a major repair (with probability 0.1). A minor repair takes half an hour and a major repair takes five hours.

Calculate the mean and variance of the repair time distribution and hence calculate the expected number of machines awaiting repair. Comment upon your result.

C2. Draw a network chart for the project whose activities and prerequisites and personnel requirements are given below. For each activity, find the earliest and latest start time, the earliest and latest completion time, and the free and total floats. Hence determine the critical path for the project.

Activity	*Prerequisites*	*Personnel*	*Duration*
A	-	3	5
B	-	5	10
C	-	4	1
D	B	1	9
E	D	7	3
F	B	2	8
G	A,F	4	7
H	B	3	10
I	D	9	4
J	C,E	1	5
K	A,F	0	3
L	G,H,I,J	0	8
M	C,E	0	4

Draw a Gantt chart for this project. By using resource levelling, keeping the critical path unchanged, calculate the maximum number of workers required.

Turn over

C3. (a) Solve the following linear programming problem using the simplex method.

$$\textit{Maximise} \quad 6x_1 + 3x_2 + 4x_3$$

$$\begin{aligned} \textit{subject to} \quad & 2x_1 + 2x_2 \qquad \geq 6 \\ & 3x_1 \qquad + x_3 = 2 \\ & x_1 + x_2 + x_3 \leq 15 \end{aligned}$$

$$x_1, x_2, x_3 \geq 0$$

(b) A computer manufacturer produces two models of lap-top computer, Standard and Special. Demand for the next two months is shown in the following table:

Model	*Month 1*	*Month 2*
Standard	150	200
Special	125	150

For each computer, the production cost, the time required by the labour force for manufacture and the time required by the labour force for assembly are shown in the following table. The current inventory levels (at the start of Month 1) are also given.

Model	*Production cost (£)*	*Time for manufacture (hours)*	*Time for assembly (hours)*	*Current inventory*
Standard	45	10	3	25
Special	60	12	4	15

Last month, the company used a total of 4000 hours of labour. The company's labour relations policy will not allow the combined total hours of labour (manufacture plus assembly) to increase or decrease by more than 500 hours from month to month.

There are end-of-month inventory holding costs. For each computer in stock at the end of a month, the holding cost is 3% of its production cost. The company requires at least 25 computers of each model to be in stock at the end of the second month.

Write down a linear programming formulation, but **do not attempt to solve it**, for the problem of planning production so that demand is satisfied at minimum cost.

C4. (a) Juliet owns a ladies' fashion store. One of her best-selling lines is a range of Indian velvet evening trousers, which sell at a rate of 500 pairs every three months. The annual holding cost rate is 20%. Juliet's current buying policy is to order 500 pairs each time she places an order. It costs her £30 to place an order. The manufacturer offers several quantity discounts, as follows:

Order quantity	*Price per pair*
0-99	£36
100-199	£32
200-299	£30
300 or more	£28

What is the minimum cost order quantity for the trousers? What are the annual savings of this policy over Juliet's current policy?

(b) During December, Jim's market stall sells giant packs of assorted Christmas wrapping paper for £10 per pack. Jim can place a single order for the wrapping paper during November at a cost of £8 per pack. After Christmas, he can sell off all unsold packs for £5 per pack. Jim knows from experience that demand for the packs will be uniformly distributed between 200 and 800.

(i) What is the recommended order quantity?

(ii) What is the probability that Jim will not have ordered enough, i.e. that at least one potential customer will be disappointed after all the packs have been sold (a *stock out*)?

(iii) Jim is anxious to keep his customers happy and wants to avoid the situation in part (ii) above. If he is prepared to tolerate a 0.15 probability of a stock out, what would your recommended order quantity be?

Turn over

SECTION D – MEDICAL STATISTICS

D1. (a) Fifteen control subjects had their sensitivity to change in temperature assessed by a test in which their right index finger was placed on a probe, initially at 30°C but thereafter at a temperature which gradually increased over time. The subjects pressed a button when they detected a change from the baseline temperature, and the current temperature of the probe was recorded. For safety reasons, the test was stopped if the probe reached 55°C. For two subjects the equipment malfunctioned before any change had been detected, and it was only possible to record the temperature at malfunction. The fifteen temperature recordings (°C) were:

33, 39, 44, 35 (malfunction), 36, 55 (test abandoned), 38, 35, 42 (malfunction), 33, 35, 37, 36, 48, 55 (test abandoned).

Calculate the Kaplan-Meier estimate of the survival function. Use Greenwood's method to find the standard error of the estimate of survival probability at 36°C.

(b) A further fifteen subjects, who had experienced blanching of their fingers following long-term use of vibrating machinery at work, also went through this test.

What statistical procedures would you use to determine whether the second group had less sensitivity to change in temperature?

(c) Suppose that further data were available on all subjects concerning their age, sex, smoking and drinking habits, any of which might influence sensitivity to temperature change. Explain carefully how you would make use of this information. Say what assumptions are being made in the approach you use.

D2. For nearly 50 years the British Government has organised a register which shows the general practitioner (i.e. doctor) each person is registered with for primary care. The register keeps track of transfers, emigrations and deaths, storing all death certificates.

Given that an asbestos-processing factory has kept complete personnel records of its workforce, including job titles, explain how it might be possible to use the GP link register to assess whether exposure to asbestos increases the risk of death from lung cancer. Which statistical methods would be used? What difficulties would there be in interpretation? Might it be possible to assess a dose-response effect? You may assume that national age, sex and time-specific lung cancer mortality rates are available for comparison.

D3. Gart (Biometrics, 1962) reports the results of some of the earliest case-control studies of the association between cigarette smoking and lung cancer. Six are reproduced below.

	Lung cancer cases		*Controls*	
Study No.	*Smokers*	*Non-smokers*	*Smokers*	*Non-smokers*
1	83	3	72	14
2	90	3	227	43
3	129	7	81	19
4	412	32	299	131
5	1350	7	1296	61
6	60	3	106	27

(i) Use the Mantel-Haenezel procedure to estimate the common odds ratio and calculate a corresponding 95% confidence interval.

(ii) Discuss the issues which need to be borne in mind when pooling results from different studies.

D4. Write notes on the following features of phase III clinical trial design.

(i) The importance of randomisation

(ii) Blinding subjects and investigators

(iii) Methods of randomisation

(iv) Interim analysis

(v) Surrogate endpoints

Turn over

SECTION E - BIOMETRY

E1. (a) Explain briefly the importance of randomization in designing experiments.

(b) An agricultural experimenter asks for your advice on how to analyse data from an experiment she has conducted to investigate the effect of irrigation on the yield from two varieties of wheat, either in the presence or absence of clover.

Her experiment was conducted in a rectangular field. The west half contained clover and the east half did not. The north half contained one variety of wheat and the south half contained the other variety. She explains that each quarter of the field contained four replicates of two plots. Within each replicate, one of the plots was selected at random to be irrigated and the other was not irrigated.

(i) Produce a sketch of the layout used.

The experimenter has already produced an analysis of variance suitable for a completely randomized design, but asks if this is correct.

(ii) Explain why the analysis of variance produced is not correct and outline the correct analysis of variance for this experiment (i.e. give the degrees of freedom corresponding to each source of variation), explaining why it is appropriate.

(iii) Explain, clearly and with reasons, which main effects and interactions can be tested and which cannot.

E2. In an experiment to compare three different eye salves (A, B and C) eighteen patients (each with two eyes) were available for the experiment. For each patient, each eye was treated with one salve, the salves being applied to the two eyes of a patient being different. Six patients had their eyes treated with Salves A and B, six more with A and C and six more with B and C. The patients were allocated randomly to treatment groups. The pain relief scores, on a scale from 0 to 9 were:

Group 1			**Group 2**			**Group 3**		
Patient	*A*	*B*	*Patient*	*A*	*C*	*Patient*	*B*	*C*
1	4	6	7	5	8	13	4	3
2	4	4	8	2	3	14	6	8
3	3	5	9	6	9	15	2	4
4	5	4	10	5	7	16	7	6
5	2	3	11	4	8	17	4	6
6	5	7	12	6	7	18	5	5

Using this experiment and data as an example, give explanations to the medical experimenter for

(i) why blocking is necessary.

(ii) why randomization is essential.

(iii) how randomization should be applied for incomplete block designs.

(iv) how information from both “between blocks” and “within blocks” can be used to provide estimates of treatment differences. (No calculations are required.)

Turn over

E3. (a) Define *overdispersion* in a generalized linear model, with particular reference to logistic regression. State two possible causes of overdispersion and what can be done to counter them.

(b) An agricultural experimenter has collected data which are to be used to compare the seeds from two cultivars of a particular crop plant to see how the storage time affects the probability of germination. For each cultivar he has stored 100 seeds for each of 10 storage times, giving the following data.

Storage time (days)	0	2	4	6	8	10	12	14	16	18
Cultivar 1 germinating (R_1)	92	94	94	86	77	69	52	38	15	4
Cultivar 2 germinating (R_2)	91	92	94	88	83	76	63	48	25	12
$\log\left(\frac{R_1}{100-R_1}\right)$	2.44	2.75	2.75	1.82	1.21	0.80	0.08	–0.49	–1.73	–3.18
$\log\left(\frac{R_2}{100-R_2}\right)$	2.31	2.44	2.75	1.99	1.59	1.15	0.53	–0.08	–1.10	–1.99

A number of logistic regression models have been fitted to the data. The following are parts of the output from a computer package.

Fitted terms: Constant + Time

	d.f.	*deviance*
Regression	1	826.97
Residual	18	61.39
Total	19	888.36

	estimate	s.e.
Constant	3.7380	0.1640
Time	–0.3038	0.0136

Fitted terms: Constant + Time + Cultivar

	d.f.	*deviance*
Regression	2	835.05
Residual	17	53.31
Total	19	888.36

	estimate	s.e.
Constant	3.5890	0.1710
Time	–0.3054	0.0137
Cultivar 2	0.3370	0.1190

Question continued on next page

Fitted terms: Constant + Time + Cultivar + Time.Cultivar

	d.f.	*deviance*
Regression	3	836.51
Residual	16	51.85
Total	19	888.36

	estimate	s.e.
Constant	3.7710	0.2330
Time	–0.3221	0.0198
Cultivar 2	–0.0360	0.3300
Time.Cultivar 2	0.0331	0.0274

(i) Describe the three models that have been fitted.

(ii) Say, with reasons, which of the three models seems most appropriate. On the basis of this model, estimate the odds ratio of germination (Cultivar 1 : Cultivar 2). What effect does storage time have on germination for each of the cultivars?

(iii) Using the given logits, draw a plot to check the assumption of the model chosen in part (ii) and on the basis of your plot suggest a further model that might be fitted.

Turn over

E4. (a) (i) Sketch the general form of the curve,

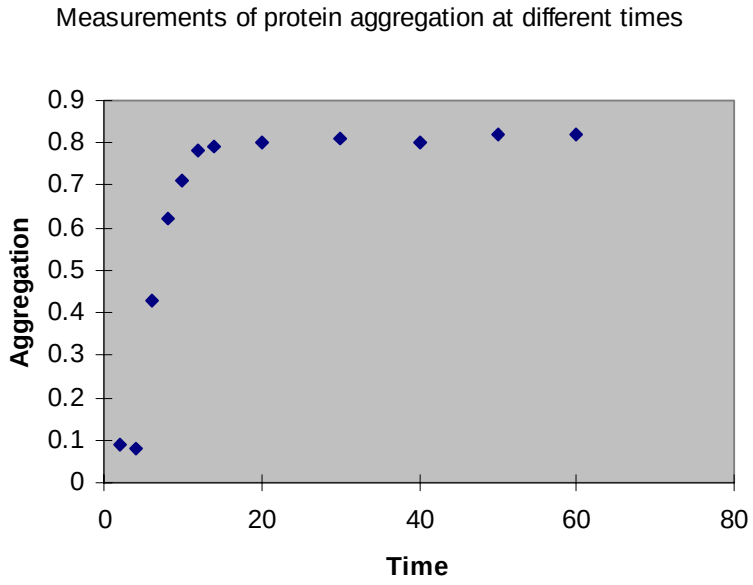
$$y = \frac{1}{\alpha + \beta e^{-(\gamma t + \delta t^2)}},$$

where $\alpha > 0$, $\beta > 0$, $\gamma > 0$ and $\delta < 0$ are unknown parameters.

(ii) Sketch the form of the curve in the case where $\alpha > 0$, $\beta > 0$, $\gamma > 0$ and $\delta = 0$.

(iii) Data $(t_1, y_1), (t_2, y_2), \ldots, (t_n, y_n)$ have been collected and a graph of y against t plotted. Show how initial estimates of the parameters can be obtained for the cases given in parts (i) and (ii). Given that the model is to be fitted using nonlinear least squares, explain why initial estimates are useful.

(b) The following plot shows data from a study of a protein. The aggregation of the protein, *A*, is plotted against time, t.



Obtain suitable initial estimates of the parameters of the curve for this set of data.

SECTION F – STATISTICS FOR INDUSTRY AND QUALITY IMPROVEMENT

F1. A company prints two-coloured designs onto sheets of metal which are then made into tins. The speed of printing (*A*), and print roller pressure (*B*) can be adjusted. An experiment to investigate the effects of these adjustments has been carried out. The response was the diameter of dots on a test pattern. Larger diameters occur when the second colour does not print exactly over the first, and it follows that a small diameter is desirable. The results are summarised as follows.

Speed (coded units)	Pressure (coded units)	Diameter (mm)
–1	–1	15.1
+1	–1	17.7
–1	1	18.2
+1	1	19.8
0	0	17.4
0	0	17.8
0	0	17.2
0	0	17.2

(i) Give a suitable model, and state the corresponding assumptions. Calculate the main effects of speed, pressure, and their interaction. Show these results on a diagram. Explain how these estimates are related to the parameters of your model.

(ii) Calculate an estimate of the standard error of the effects. Is there evidence against a hypothesis of no interaction?

(iii) The company intends to search for a minimum diameter by changing the speed and pressure. What would the centre point for another such experiment be if you move 2 units in the direction of steepest descent?

(iv) What problems are there in estimating quadratic effects with the present design? Suggest a design with 9 points which would be an improvement in this respect. What advantage does the present design have, and could this be achieved by any other design with 8 points?

Turn over

F2. (a) A cooperative produces rolls of Harris tweed. Following customer complaints, the cooperative inspected every roll produced over a period of 20 days. The results follow.

day	1	2	3	4	5	6	7	8	9	10
Number of rolls manufactured	20	20	20	20	20	20	20	20	20	20
Number of blemishes	12	6	7	1	9	11	6	9	6	4

day	11	12	13	14	15	16	17	18	19	20
Number of rolls manufactured	25	25	25	25	25	25	25	25	25	25
Number of blemishes	12	10	9	12	5	11	16	10	6	9

(i) By calculating summary statistics for days 1-10 and 11-20 separately, comment on whether it is plausible that the number of blemishes follow a Poisson process.

(ii) The cooperative considered that the quality of production had been satisfactory over the 20 day period. Assume the number of blemishes does follow a Poisson process. Set up a control chart to analyse the following data which were obtained by inspecting a random sample of the rolls produced each week.

week	1	2	3	4	5	6	7
sample size	9	9	9	16	16	16	16
number of blemishes	6	4	9	7	11	8	15

(b) A company produces batches of a chemical used in the manufacture of printed circuit boards. The gold content of the chemical is measured for each batch. Discrepancies from specification for the past 30 batches are:

batch	1	2	3	4	5	6	7	8	9	10
discrepancy	2	5	–17	7	–1	–3	11	–11	1	20
batch	11	12	13	14	15	16	17	18	19	20
discrepancy	15	–5	–10	4	4	–11	–13	–8	9	–10
batch	21	22	23	24	25	26	27	28	29	30
discrepancy	–17	–12	–13	–27	–2	–13	3	0	0	–16

Question continued on next page

If the process is running satisfactorily the standard deviation of the discrepancies is 10 units.

(i) Draw a CUSUM chart. Show a suitable V-mask for batch 25, and explain the construction. What is the false alarm rate? Has there been any cause for concern over this period?

(ii) The discrepancies are based on the average of three analyses of samples from each batch. What else might be monitored in addition to the means? Give one reason why the company might argue that this is unnecessary.

F3. (i) What is a Markov process? Discuss the Markov property in relation to the exponential distribution.

(ii) A small company has two injection moulding machines which should both operate continuously. The times between failures of a machine are independent and have an exponential distribution with a mean of $1/\lambda$ hours. There is one employee who can repair the machines, and the repair time has an exponential distribution with a mean of $1/\mu$ hours. For what proportion of the time will: neither machine be working; only one machine be working?

(iii) The company has expanded and now owns three machines, and employs two people to repair them. Only one person can work on the repair of one machine. The failure rate and mean repair time, per machine, are unchanged. Write down the equations you need to solve to find the proportion of time in each state. For what proportion of the time will all the machines be working?

(iv) Now suppose that two people can work on the repair of one machine. How would the equations in part (iii) change? State any assumptions you make.

Turn over

F4. (a) Twelve central heating pumps were installed 10 years ago. Two are still working. The times until failure of the other 10, in months, were:

18 33 41 42 46 52 56 68 105 106 .

Use a graphical method to estimate the parameters, α and β, in the Weibull distribution with cumulative distribution function:

$$F(x) = 1 - \exp\left(-(x/\beta)^{\alpha}\right) \ .$$

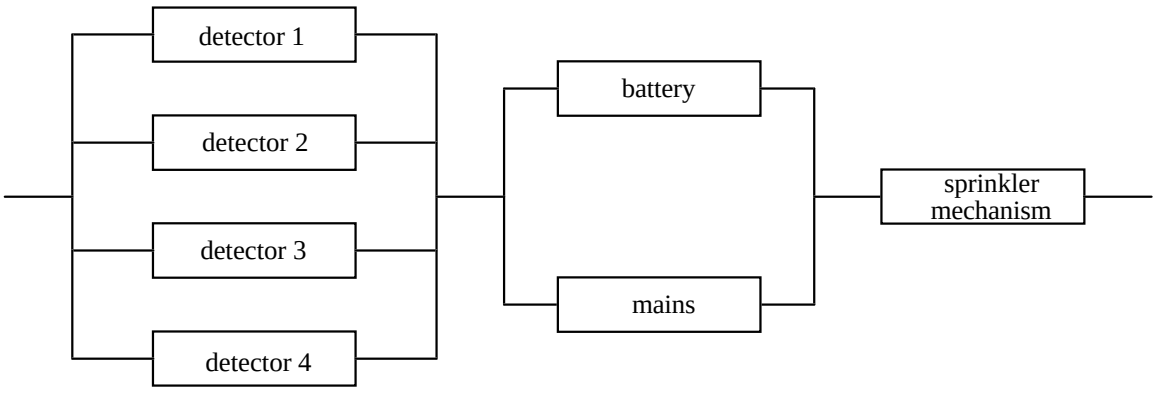
You may assume that:

$$F\left(x_{(i)}\right) = \frac{i - 0.4}{n + 0.2}$$

where $x_{(i)}$ is the i-th order statistic.

Hence estimate the probability that a pump will last longer than 12 years.

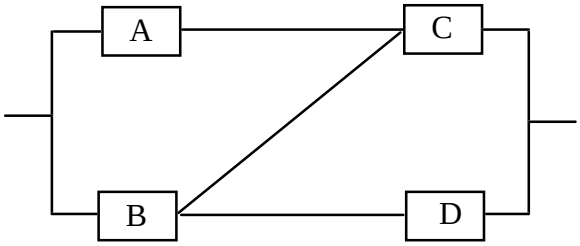
(b) The office sprinkler system, represented in the diagram below, will operate if at least one smoke detector works, the battery or mains supply works, and the sprinkler mechanism works. The detectors, battery, mains, and sprinkler mechanism have reliabilities R_d, R_b, R_m, and R_s respectively. Assuming that failures are independent, express the overall reliability in terms of R_d, R_b, R_m and R_s.



(c) How could you model a common cause of detector failure, such as not replacing the small internal batteries in the detectors in the sprinkler system of part (b) of this question?

Question continued on next page

(d) What is the overall reliability of the system:



if components have reliabilities R_A, R_B, R_C *and* R_D and failures are independent?