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

GRADUATE DIPLOMA, 2000

Applied Statistics I

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* $\log_e$.

THERE IS A 10-PAGE APPENDIX FOR USE WITH THIS EXAMINATION PAPER

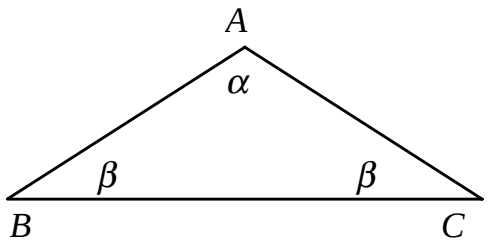
Each candidate must ensure (s)he has a copy of the Appendix.
The Appendix is to be handed in with the question paper and the answer book at the end of the examination. It must NOT be kept by the candidate.

This examination paper consists of 11 printed pages followed by 1 blank page (page 12). This front cover is page 1. The reverse of the front cover is intentionally left blank and is page 2. Question 1 starts on page 3.

There are 8 questions altogether in the paper.

In addition, the Appendix is now included as pages 13-22 of the examination paper.
In the actual examination, it was supplied to candidates separately.

1. In the triangle *ABC* it is known that sides *AB* and *AC* are equal in length. The sizes of the angles are unknown and denoted by α and β as in the diagram.



The cost of measurement dictates that only 12 measurements can be made. Two suggestions are being considered:

Method P: measure angles at *A*, *B* and *C* four times each;

Method Q: measure the angle at *A* six times, and those at *B* and *C* three times each.

You may assume that all measurements will be unbiased and uncorrelated, all with the same variance.

(i) Use the method of least squares to obtain estimators of α and β using method P and using method Q.

(ii) Compare the usefulness of the two methods.

(iii) Can you think of a way of determining any other method of taking 12 measurements which is better than either of these two methods? You should describe the characteristics of such a method and how it might be derived: there is no need to derive the method itself.

(20)

2. (a) Explain what are meant by the terms *stationarity* and *weak stationarity* in the context of a time series.

What is meant in practice when a series is described as being stationary?

(7)

(b) On **pages 3-5 of the Appendix**, the correlograms for five different series are presented. For three of these (series 2, 4 and 5), the sample partial autocorrelation function is also given. The numbers of data points in the various series are 144 in series 1 and 70 in each of the others. In the light of the information provided, answer parts (i) and (ii) following for each of the five series.

(i) Explain why you think the series is, or is not, stationary.

(ii) If you think the series is stationary, identify a possible model (or models) that you would consider fitting to the series. You should justify your choice of model on the basis of features exhibited by the sample autocorrelation function (and partial autocorrelation function where appropriate).

If you think the series is non-stationary, identify an initial operator you would apply in an attempt to achieve stationarity. Again, you should justify your answer.

(13)

3. (a) In the context of a multiple linear regression analysis:

(i) explain what is meant by the leverage of a data point;

(ii) define the "hat" matrix in terms of the design matrix (**X**) and briefly explain how the "hat" matrix is used in assessing leverage.

(3)

(b) Blood volume in a newborn can be calculated by injecting dye into the bloodstream and then dividing the amount injected by the concentration measured in the blood. The optical density (OD) of the dye is measured at wavelengths 620 (OD620) and 740 (OD740). It is known that the dye affects measurements at wavelength 620 but not at wavelength 740. The table shows the optical densities of blood for 36 newborns on phototherapy for jaundice.

Newborn	*OD620*	*OD740*	*Newborn*	*OD620*	*OD740*
1	28	14	19	26	9
2	14	7	20	36	17
3	37	12	21	48	20
4	84	40	22	54	30
5	28	11	23	56	31
6	38	16	24	135	74
7	98	54	25	40	16
8	21	9	26	21	8
9	44	22	27	48	19
10	118	74	28	30	10
11	42	18	29	22	11
12	60	31	30	50	30
13	106	48	31	18	8
14	62	42	32	35	16
15	49	22	33	241	124
16	38	18	34	73	29
17	26	9	35	40	11
18	46	23	36	42	20

A regression analysis of these data is given on **pages 6 and 7 of the Appendix**.

(i) Using the plot of OD620 against OD740, describe the apparent relationship between the two variables, pointing out any possible problems with the data.

(2)

(ii) Briefly comment on the form of the plot of standardised residuals against fitted values.

(2)

(iii) Discuss the reasons for repeating the analysis with observation 33 removed. Describe the effects of doing this, and the reason for each of the observed effects. Which of the two models would you choose and why?

(13)

4. An experiment was carried out to investigate the effect of length of exposure to sunlight on growth of plants. Six treatments (8, 12 and 16 hours of exposure in each of two glass-house conditions, low and high night temperatures) were assigned at random to pots with three pots per treatment, and four plants were taken at random from a large group and assigned to each pot. The growth in cm of the stems of plants was measured after one week, and gave the following data. (Treatments 1-3 are the three exposure times for low night temperature, and treatments 4-6 are the same three exposure times at high night temperature.) The data, together with some totals, are shown in the table.

		Treatment						
Pot	Plant	1	2	3	4	5	6	
1	1	3.5	5.0	5.0	8.5	6.0	7.0	
	2	4.0	5.5	4.5	6.0	5.5	9.0	
	3	3.0	4.0	5.0	9.0	3.5	8.5	
	4	4.5	3.5	4.5	8.5	7.0	8.5	
Totals		15.0	18.0	19.0	32.0	22.0	33.0	
2	1	2.5	3.5	5.5	6.5	6.0	6.0	
	2	4.5	3.5	6.0	7.0	8.5	7.0	
	3	5.5	3.0	5.0	8.0	4.5	7.0	
	4	5.0	4.0	5.0	6.5	7.5	7.0	
Totals		17.5	14.0	21.5	28.0	26.5	27.0	
3	1	3.0	4.5	5.5	7.0	6.5	11.0	
	2	3.0	4.0	4.5	7.0	6.5	7.0	
	3	2.5	4.0	6.5	7.0	8.5	9.0	
	4	3.0	5.0	5.5	7.0	7.5	8.0	
Totals		11.5	17.5	22.0	28.0	29.0	35.0	
	Totals	44.0	49.5	62.5	88.0	77.5	95.0	416.5

If y_{ijk} denotes the week's growth for plant k in pot j receiving treatment i then $\sum_{i,j,k} y_{ijk}^2 = 2665.25$.

(i) Explain why this is a nested design, and write down the form of the model, stating any assumptions.

(6)

(ii) Copy the ANOVA table given below into your answer book and complete it.

(7)

Source of variation	**sum of squares**	**degrees of freedom**	**mean square**
Between treatments			
Between pots within treatments			
Error			
Total	255.91	71	

(iii) Hence carry out suitable hypothesis tests on the data. You should state your hypotheses clearly, and comment on the implications of your conclusions.

(4)

(iv) What further analyses might you recommend on these data?

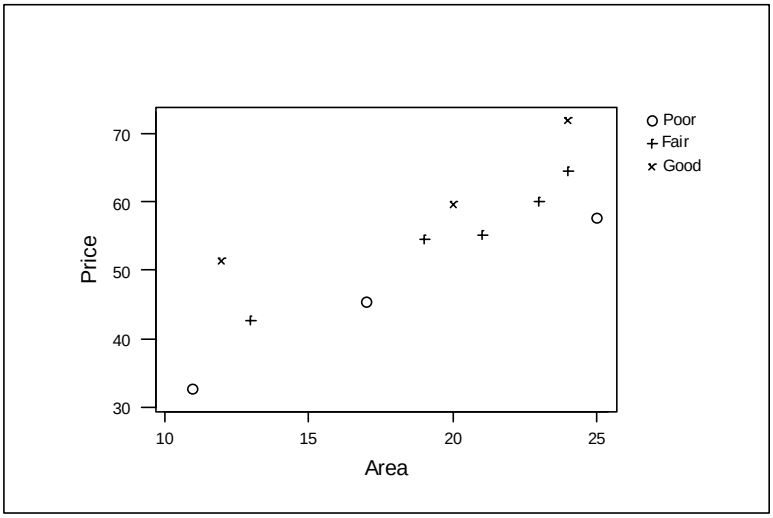
(3)

5. The following table contains data, collected by an estate agent, on the selling price (in thousands of pounds), floor area (in hundreds of square feet) and condition (poor=1, fair=2, good=3) of eleven recently sold semi-detached houses.

area	condition	price	area	condition	price
23	2	60.0	21	2	55.3
11	1	32.7	24	2	64.5
20	3	59.7	13	2	42.6
17	1	45.5	19	2	54.5
12	3	51.3	25	1	57.7
24	3	72.0			

(i) Interpret the diagram below, which is the plot of selling price against floor area, with symbols indicating house condition.

(4)



(ii) A model is considered in which, for each type of house condition, selling price has a linear regression on floor area. The regression lines are not restricted to being parallel, and the intercepts may not be the same. Write down a mathematical representation of this model, defining all the terms used.

(5)

(iii) Referring to the computer output on **pages 8 and 9 of the Appendix**, and using backward selection with a 5% significance level, justify the selection of the model including both main effects but no interaction. Briefly interpret this chosen model in the context of the study.

(6)

(iv) Specify the chosen model for each house condition, describing how you would determine the models completely.

(2)

(v) What else would you do with the data to confirm the validity of the selected model?

(3)

6. In many instances of linear modelling, a response variable y can be dependent on more than one variable x. Thus a set of variables x_i (i = 1, 2, ..., p) is used to predict y using the general linear model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon \quad .$$

(i) State clearly all assumptions made in fitting such a model.

(2)

(ii) Write down the equivalent matrix formulation of the model, and state (without proof) the matrix form of the least squares estimators for the parameters in the model.

(3)

(iii) Explain why highly correlated predictor variables can cause problems in fitting a multiple regression model. What methods can be used to overcome such problems?

(3)

(iv) Explain why an adjusted R^2 value may be preferred to R^2 when comparing models.

(2)

(v) What is meant by residual analysis? Why is it a necessary part of model-building? Describe various analyses which can be carried out on residuals, and their purpose.

(4)

(vi) Describe some of the diagnostics which can be obtained from statistical packages to detect influential observations.

(6)

7. (a) Four laboratories were requested to perform a certain routine chemical analysis in order to determine the percentage by weight of a particular compound present in a fertiliser product. Identical quantities of the product were sent to each laboratory to be split into individual samples and analysed. Results were as shown in the table.

Laboratory	*Results*	*n*	Σy
1	58.7, 61.4, 60.9, 59.1, 58.2	5	298.3
2	62.7, 64.5, 63.1, 59.2, 60.3, 62.3	6	372.1
3	55.9, 56.1, 57.3, 55.2, 58.1	5	282.6
4	60.7, 60.3, 61.4, 60.9	4	243.3

If y_{ij} denotes measurement *j* from laboratory *i* then

$$\sum_{i,j} y_{ij}^{\,2} = 71676.99 \ .$$

(i) Carry out an analysis of variance of the data, assuming that a fixed effects model is appropriate. State your conclusions clearly.

(10)

(ii) Describe the circumstances under which a fixed effects model would be appropriate for these data.

(1)

(iii) Describe the circumstances under which a random effects model would be appropriate, and state how your analysis and/or conclusions might be different from those in (i).

(2)

(iv) Describe any further analyses you might validly carry out for each of the two types of model discussed in (ii) and (iii), stating the conditions under which such analyses would be appropriate.

(3)

(b) In a different experiment the response variable is a count, and there is strong evidence from the data that the variance is proportional to the mean. Discuss the problems this would cause for an analysis of variance, and suggest possible solutions with their relative merits.

(4)

8. The meaning of words changes with the course of history. However, the meaning of the numbers 1, 2, 3, … represents a clear exception. A first comparison of languages might be based on the numerals alone. Table 1.1 gives the first ten numbers in English, Polish, Hungarian and eight other modern European languages. (Only languages that use the Roman alphabet are included, and certain accent signs are omitted.)

English	Norweg-ian	Danish	Dutch	German	French	Spanish	Italian	Polish	Hungar-ian	Finnish
one	en	en	een	ein	un	uno	uno	jeden	egy	yksi
two	to	to	twee	zwei	deux	dos	due	dwa	ketto	kaksi
three	tre	tre	drie	drei	trois	tres	tre	trzy	harom	kolme
four	fire	fire	vier	vier	quatre	cuatro	quattro	cztery	negy	neua
five	fem	fem	vijf	funf	cinq	cinqo	cinque	piec	of	viisi
six	seks	seks	zes	sechs	six	seix	sei	szesc	hat	kuusi
seven	sju	syv	zeven	sieben	sept	siete	sette	siedem	het	seitseman
eight	atte	otte	acht	acht	huit	ocho	otto	osiem	nyolc	kahdeksan
nine	ni	ni	negen	neun	neuf	nueve	nove	dziewiec	kilenc	yhdeksan
ten	ti	ti	tien	zehn	dix	diez	dieci	dziesicc	tiz	kymmennen

Table 1.1

(i) From a cursory examination of the spellings of these numerals, suggest possible language groupings.

(3)

(ii) On a more systematic basis, languages might be compared by counting the frequencies of the same initial letter for the same number, for pairs of languages. Table 1.2 shows the frequencies of such "concordant" first letters.

	E	N	Da	Du	G	Fr	Sp	I	P	H	Fi
E	10										
N	8	10									
Da	8	9	10								
Du	3	5	4	10							
G	4	6	5	5	10						
Fr	4	4	4	1	3	10					
Sp	4	4	5	1	3	8	10				
I	4	4	5	1	3	9	9	10			
P	3	3	4	0	2	5	7	6	10		
H	1	2	2	2	1	0	0	0	0	10	
Fi	1	1	1	1	1	1	1	1	1	2	10

Table 1.2

Briefly explain why these concordances are not technically "similarities" and suggest a simple method for transforming these concordances to measures of similarity.

(4)

Question 8 continued on next page

(iii) Table 1.3 shows table 1.2 transformed by subtracting each concordance from 10. Explain why table 1.3 can be regarded as a matrix of "distances", and how this can be used to apply an agglomerative hierarchical clustering technique to the data.

	E	**N**	**Da**	**Du**	**G**	**Fr**	**Sp**	**I**	**P**	**H**	**Fi**
E	0										
N	2	0									
Da	2	1	0								
Du	7	5	6	0							
G	6	4	5	5	0						
Fr	6	6	6	9	7	0					
Sp	6	6	5	9	7	2	0				
I	6	6	5	9	7	1	1	0			
P	7	7	6	10	8	5	3	4	0		
H	9	8	8	8	9	10	10	10	10	0	
Fi	9	9	9	9	9	9	9	9	9	8	0

Table 1.3

(5)

(iv) Figures 1.1 to 1.3 on **page 10 of the Appendix** show the dendrograms from three agglomerative hierarchical clustering methods applied to the data. Critically discuss the possible interpretation and practical usefulness of these results, relating them to your answer in (i).

(8)

BLANK PAGE

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

GRADUATE DIPLOMA, 2000

Applied Statistics I

APPENDIX

Each candidate must have a copy of this Appendix.

The Appendix consists of TEN pages.

This front cover is page 1.

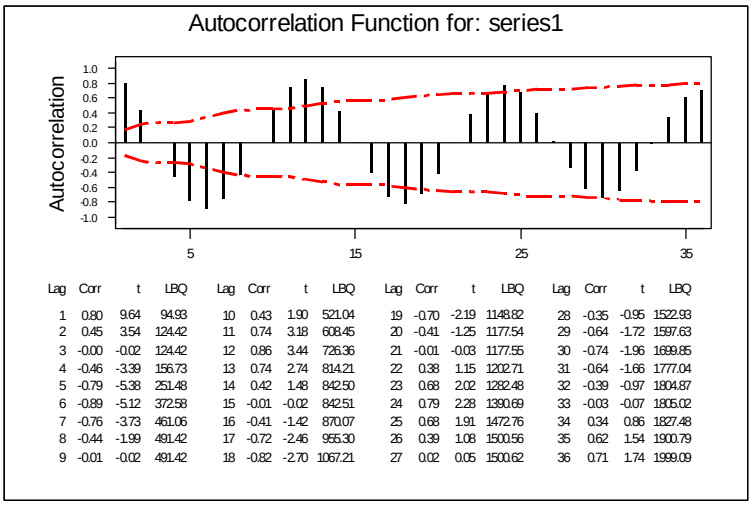
The reverse of the front cover, which is intentionally left blank, is page 2.

The text of the Appendix starts on page 3.

The Appendix is to be handed in with the candidate's examination paper and answer book at the end of the examination. It is NOT to be removed by the candidate.

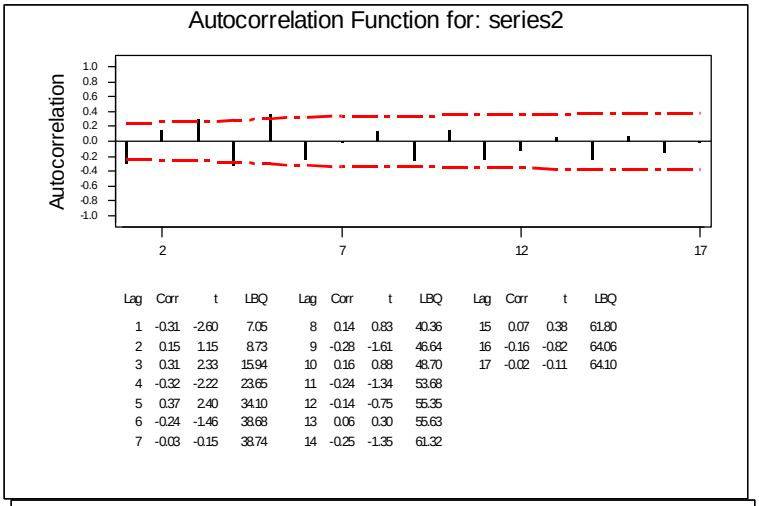
Correlograms for question 2. These are printed on this page and the next two pages

SERIES 1

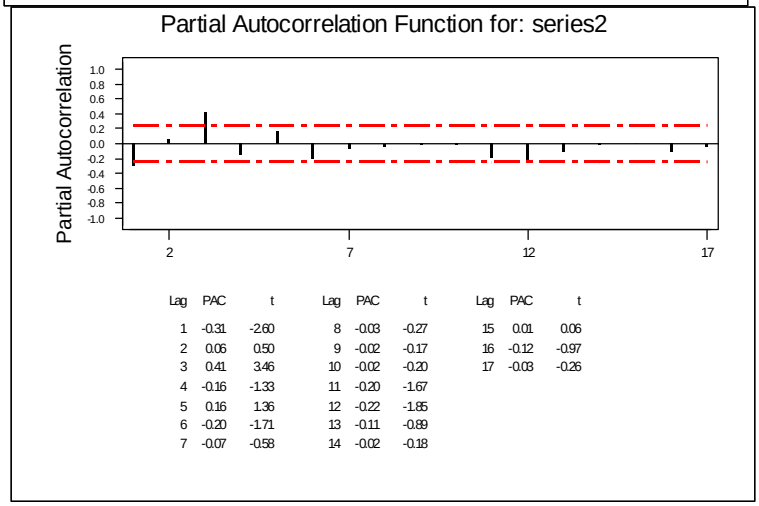


Lag	Corr	t	LBQ	Lag	Corr	t	LBQ	Lag	Corr	t	LBQ	Lag	Corr	t	LBQ
1	0.80	9.64	94.93	10	0.43	1.90	521.04	19	-0.70	-2.19	1148.82	28	-0.35	-0.95	1522.93
2	0.45	3.54	124.42	11	0.74	3.18	608.45	20	-0.41	-1.25	1177.54	29	-0.64	-1.72	1597.63
3	-0.00	-0.02	124.42	12	0.86	3.44	726.36	21	-0.01	-0.03	1177.55	30	-0.74	-1.96	1699.85
4	-0.46	-3.39	156.73	13	0.74	2.74	814.21	22	0.38	1.15	1202.71	31	-0.64	-1.66	1777.04
5	-0.79	-5.38	251.48	14	0.42	1.48	842.50	23	0.68	2.02	1282.48	32	-0.39	-0.97	1804.87
6	-0.89	-5.12	372.58	15	-0.01	-0.02	842.51	24	0.79	2.28	1390.69	33	-0.03	-0.07	1805.02
7	-0.76	-3.73	461.06	16	-0.41	-1.42	870.07	25	0.68	1.91	1472.76	34	0.34	0.86	1827.48
8	-0.44	-1.99	491.42	17	-0.72	-2.46	955.30	26	0.39	1.08	1500.56	35	0.62	1.54	1900.79
9	-0.01	-0.02	491.42	18	-0.82	-2.70	1067.21	27	0.02	0.05	1500.62	36	0.71	1.74	1999.09

SERIES 2

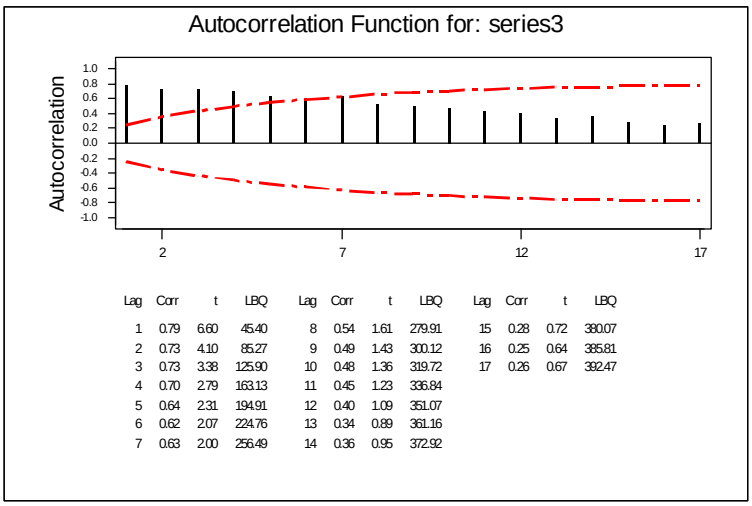


Lag	Corr	t	LBQ	Lag	Corr	t	LBQ	Lag	Corr	t	LBQ
1	-0.31	-2.60	7.05	8	0.14	0.83	40.36	15	0.07	0.38	61.80
2	0.15	1.15	8.73	9	-0.28	-1.61	46.64	16	-0.16	-0.82	64.06
3	0.31	2.33	15.94	10	0.16	0.88	48.70	17	-0.02	-0.11	64.10
4	-0.32	-2.22	23.65	11	-0.24	-1.34	53.68				
5	0.37	2.40	34.10	12	-0.14	-0.75	55.35				
6	-0.24	-1.46	38.68	13	0.06	0.30	55.63				
7	-0.03	-0.15	38.74	14	-0.25	-1.35	61.32				



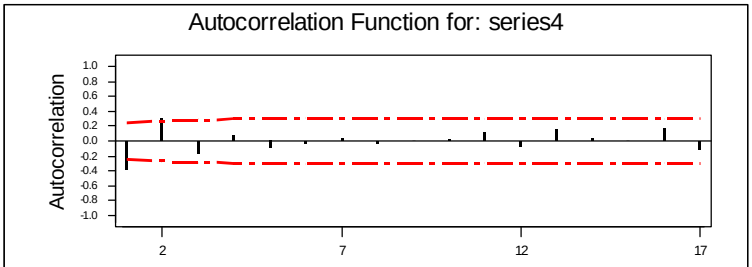
Lag	PAC	t	Lag	PAC	t	Lag	PAC	t
1	-0.31	-2.60	8	-0.03	-0.27	15	0.01	0.06
2	0.06	0.50	9	-0.02	-0.17	16	-0.12	-0.97
3	0.41	3.46	10	-0.02	-0.20	17	-0.03	-0.26
4	-0.16	-1.33	11	-0.20	-1.67			
5	0.16	1.36	12	-0.22	-1.85			
6	-0.20	-1.71	13	-0.11	-0.89			
7	-0.07	-0.58	14	-0.02	-0.18			

SERIES 3

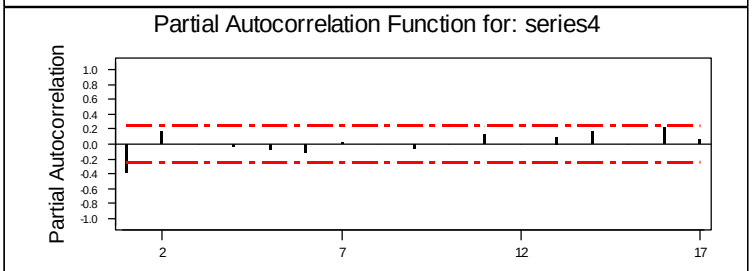


Lag	Corr	t	LBQ	Lag	Corr	t	LBQ	Lag	Corr	t	LBQ
1	0.79	6.60	45.40	8	0.54	1.61	279.91	15	0.28	0.72	380.07
2	0.73	4.10	85.27	9	0.49	1.43	300.12	16	0.25	0.64	385.81
3	0.73	3.38	125.90	10	0.48	1.36	319.72	17	0.26	0.67	392.47
4	0.70	2.79	163.13	11	0.45	1.23	336.84				
5	0.64	2.31	194.91	12	0.40	1.09	351.07				
6	0.62	2.07	224.76	13	0.34	0.89	361.16				
7	0.63	2.00	256.49	14	0.36	0.95	372.92				

SERIES 4

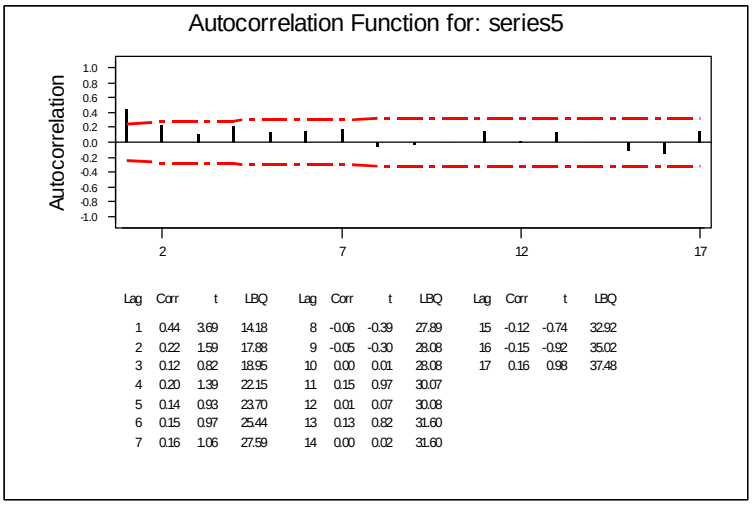


Lag	Corr	t	LBQ	Lag	Corr	t	LBQ	Lag	Corr	t	LBQ
1	-0.39	-3.26	11.10	8	-0.04	-0.29	21.57	15	-0.01	-0.04	25.09
2	0.30	2.23	17.97	9	-0.00	-0.03	21.57	16	0.17	1.12	27.89
3	-0.17	-1.14	20.03	10	0.01	0.10	21.59	17	-0.11	-0.71	29.06
4	0.07	0.48	20.41	11	0.11	0.73	22.62				
5	-0.10	-0.65	21.14	12	-0.07	-0.45	23.03				
6	-0.05	-0.31	21.32	13	0.15	0.97	24.97				
7	0.04	0.24	21.42	14	0.04	0.23	25.09				

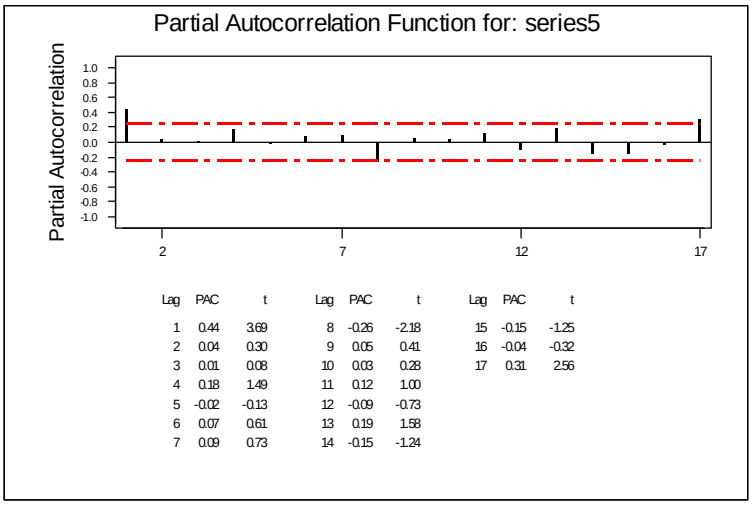


Lag	PAC	t	Lag	PAC	t	Lag	PAC	t
1	-0.39	-3.26	8	0.00	0.04	15	-0.00	-0.01
2	0.18	1.50	9	-0.06	-0.47	16	0.22	1.85
3	0.00	0.02	10	0.00	0.03	17	0.05	0.44
4	-0.04	-0.37	11	0.14	1.19			
5	-0.07	-0.58	12	-0.01	-0.08			
6	-0.12	-1.01	13	0.09	0.77			
7	0.02	0.16	14	0.17	1.40			

SERIES 5

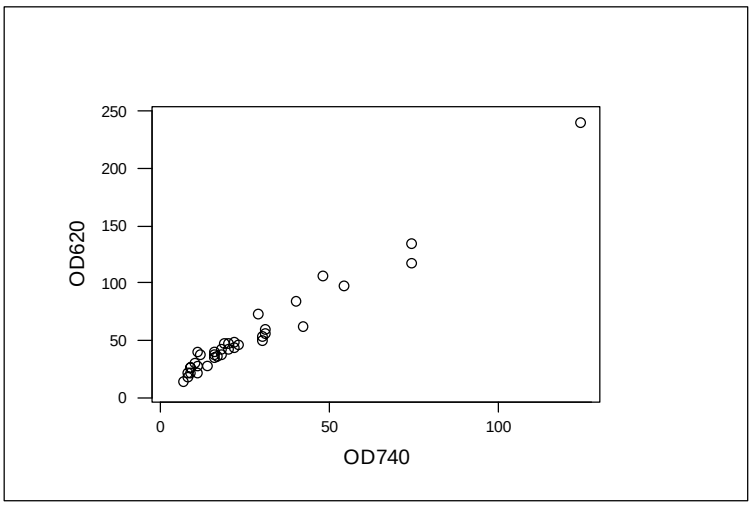


Lag	Corr	t	LBQ	Lag	Corr	t	LBQ	Lag	Corr	t	LBQ
1	0.44	3.69	14.18	8	-0.06	-0.39	27.89	15	-0.12	-0.74	32.92
2	0.22	1.59	17.88	9	-0.05	-0.30	28.08	16	-0.15	-0.92	35.02
3	0.12	0.82	18.95	10	0.00	0.01	28.08	17	0.16	0.98	37.48
4	0.20	1.39	22.15	11	0.15	0.97	30.07				
5	0.14	0.93	23.70	12	0.01	0.07	30.08				
6	0.15	0.97	25.44	13	0.13	0.82	31.60				
7	0.16	1.06	27.59	14	0.00	0.02	31.60				



Lag	PAC	t	Lag	PAC	t	Lag	PAC	t
1	0.44	3.69	8	-0.26	-2.18	15	-0.15	-1.25
2	0.04	0.30	9	0.05	0.41	16	-0.04	-0.32
3	0.01	0.08	10	0.03	0.28	17	0.31	2.56
4	0.18	1.49	11	0.12	1.00			
5	-0.02	-0.13	12	-0.09	-0.73			
6	0.07	0.61	13	0.19	1.58			
7	0.09	0.73	14	-0.15	-1.24			

Regression analysis for question 3. This is printed on this page and the next page



Regression Analysis

The regression equation is
OD620 = 7.66 + 1.76 OD740

Predictor	Coef	Stdev	t-ratio	p
Constant	7.661	2.004	3.82	0.001
OD740	1.76097	0.05663	31.09	0.000

s = 7.976 R-sq = 96.6% R-sq(adj) = 96.5%

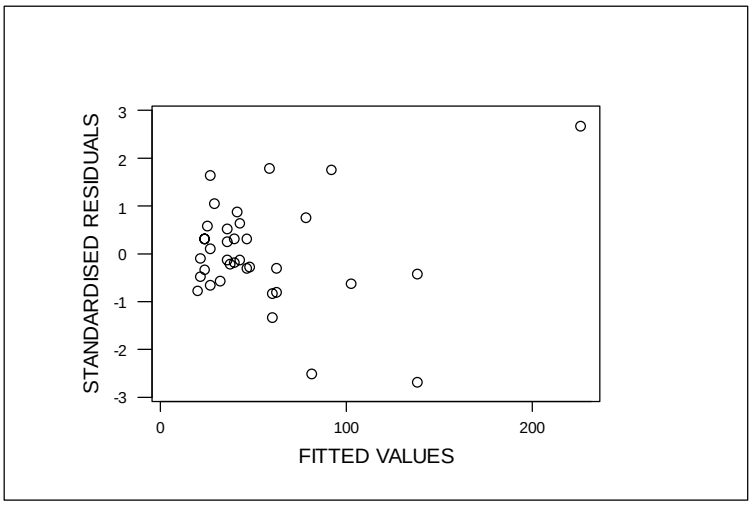
Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	61502	61502	966.82	0.000
Error	34	2163	64		
Total	35	63665			

Unusual Observations

Obs.	OD740	OD620	Fit	Stdev.Fit	Residual	St.Resid
10	74	118.00	137.97	3.00	-19.97	-2.70R
14	42	62.00	81.62	1.59	-19.62	-2.51R
33	124	241.00	226.02	5.68	14.98	2.68RX

R denotes an obs. with a large st. resid.
X denotes an obs. whose X value gives it large influence.


STANDARDISED RESIDUALS
3
2
1
0
-1
-2
-3
0
100
200
FITTED VALUES

Regression analysis after deletion of observation 33

The regression equation is
OD620 = 10.8 + 1.61 OD740

Predictor	Coef	Stdev	t-ratio	p
Constant	10.774	2.090	5.16	0.000
OD740	1.61145	0.07175	22.46	0.000

s = 7.193 R-sq = 93.9% R-sq(adj) = 93.7%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	26096	26096	504.38	0.000
Error	33	1707	52		
Total	34	27804			

Unusual Observations

Obs.	OD740	OD620	Fit	Stdev.Fit	Residual	St.Resid
10	74.0	118.00	130.02	3.81	-12.02	-1.97 X
13	48.0	106.00	88.12	2.13	17.88	2.60R
14	42.0	62.00	78.46	1.79	-16.46	-2.36R
24	74.0	135.00	130.02	3.81	4.98	0.82 X
33	29.0	73.00	57.51	1.27	15.49	2.19R

R denotes an obs. with a large st. resid.
X denotes an obs. whose X value gives it large influence.

Computer output for question 5. This is printed on this page and the next page

```
General Linear Model


Factor   Levels Values
Cond          3     1     2     3


Analysis of Variance for Price

Source      DF     Seq SS     Adj SS     Adj MS       F      P
Area         1     855.77     751.37     751.37  124.63  0.000
Cond         2     293.44      33.23      16.61    2.76  0.156
Cond*Area    2       1.91       1.91       0.95    0.16  0.858
Error        5      30.14      30.14       6.03
Total       10    1181.27



Term          Coeff     Stdev  t-value      P
Constant     20.994     3.058     6.87  0.001
Area         1.7501    0.1568    11.16  0.000
Area*Cond
       1    0.0216    0.2120     0.10  0.923
       2    0.0999    0.2259     0.44  0.677



General Linear Model


Factor   Levels Values
Cond          3     1     2     3


Analysis of Variance for Price

Source     DF     Seq SS     Adj SS     Adj MS       F      P
Area        1     855.77     765.92     765.92  167.29  0.000
Cond        2     293.44     293.44     146.72   32.05  0.000
Error       7      32.05      32.05       4.58
Total      10    1181.27



Term         Coeff     Stdev  t-value      P
Constant    20.982     2.630     7.98  0.000
Area        1.7527    0.1355    12.93  0.000

Unusual Observations for Price

Obs.    Price       Fit Stdev.Fit  Residual   St.Resid
  3   59.7000   63.3369    1.2485   -3.6369     -2.09R

R denotes an obs. with a large st. resid.
```

General Linear Model

Analysis of Variance for Price

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Area	1	855.77	855.77	855.77	23.66	0.000
Error	9	325.49	325.49	36.17		
Total	10	1181.27				

Term	Coeff	Stdev	t-value	P
Constant	19.693	7.315	2.69	0.025
Area	1.8142	0.3730	4.86	0.000

General Linear Model

Factor	Levels	Values
Cond	3	1 2 3

Analysis of Variance for Price

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Cond	2	383.30	383.30	191.65	1.92	0.208
Error	8	797.97	797.97	99.75		
Total	10	1181.27				

Dendrograms for question 8

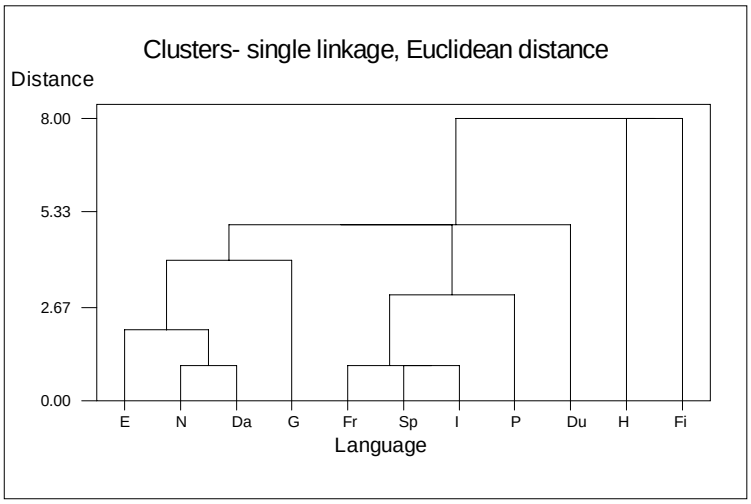


Figure 1.1

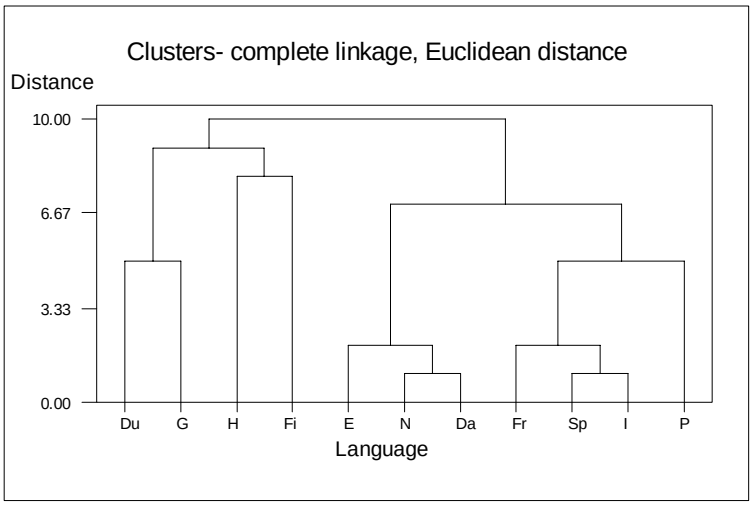


Figure 1.2

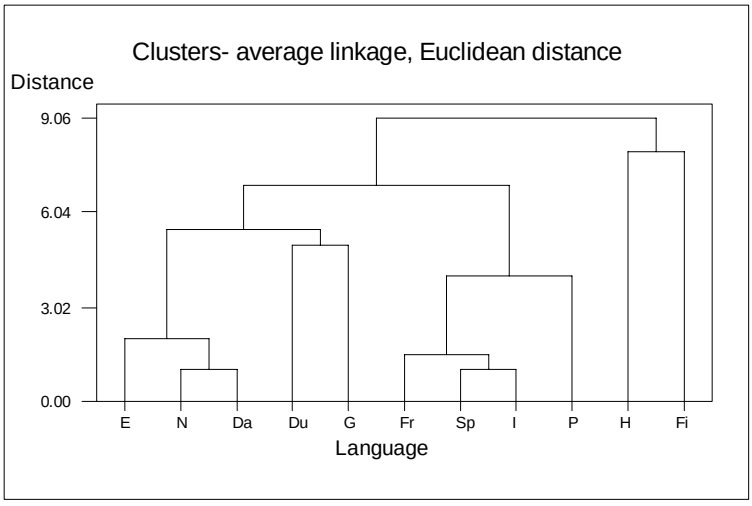


Figure 1.3