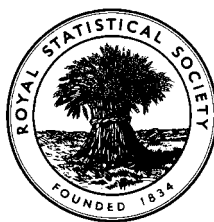**EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY**

(formerly the Examinations of the Institute of Statisticians)

HIGHER CERTIFICATE IN STATISTICS, 1999

Paper III : Statistical Applications and Practice

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* $^{n}C_{r}$ *and that* ln *stands for* $\log_{e}$.

There is a worksheet for use with question 4.

If you answer question 4, you must hand the worksheet in with your answer book. ENSURE YOU HAVE WRITTEN YOUR CANDIDATE'S NUMBER ON THE WORKSHEET.

This worksheet is included as the past page of this Web version of the examination paper

---

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 examination paper.

In addition, the worksheet for use with question 4 is supplied as the last page.

1.   The activity of the enzyme acid phosphatase was measured at four different pH values. Five replicate measurements were made at each pH value.

	Enzyme activity μM / min at pH		
3	*5*	*7*	*9*
11.1	12.0	11.2	5.6
10.0	15.3	9.1	7.2
13.5	15.1	9.6	6.4
10.5	15.0	10.0	5.9
11.3	13.2	9.8	6.3
mean 11.28	14.12	9.94	6.28

Plot the enzyme activities against pH.   (3)

A one-way analysis of variance of the enzyme activities using pH as the classification factor gave the following analysis of variance table:

Source of variation	Sum of Squares	Degrees of freedom
Between pH values		3
Error		
Total	178.57	

Complete the analysis of variance table by filling in the missing values and adding appropriate mean squares and a variance ratio. Carry out an appropriate statistical test on the variance ratio and state clearly in non-technical language what the analysis reveals. What assumptions have you made?   (10)

A simple linear model relating enzyme activity to pH was also fitted to the data and gave the following results.

The regression equation is
Enzyme activity = 16.2 – 0.959 (pH value)

Analysis of Variance for linear regression

Source of variation	Sum of squares	Degrees of freedom	Mean squares	Variance ratio
Linear regression	91.968	1	91.968	19.12
Error	86.601	18	4.811	
Total	178.570	19		

Draw the fitted line on your plot and comment on whether the fitted straight line adequately describes the relation between enzyme activity and pH.   (3)

The investigator wishes to carry out another series of experiments at four new pH values to determine more precisely the pH at which enzyme activity is at its greatest. What values would you suggest and why would you suggest them?   (4)

3

Turn over

2. The results of two surveys of public opinion relating to genetically engineered food were published in a popular science magazine. The magazine article began by claiming that opposition to genetically engineered food was growing in Britain and it also contained a quote from the director of Genewatch who was reported to have said "Opinion has hardened against genetic engineering quite significantly".

The article contained a table which summarised the results of two opinion polls, the first conducted in 1996 and the second in 1998. Both polls were carried out by a reputable market research organisation which asked 1000 adults the same question on both occasions. You may assume that the polls were properly conducted and the people questioned on each occasion were randomly sampled from the population. Note that the people questioned in 1996 were not the same as those questioned in 1998.

The table contained in the article is given below.

What do you think of genetically engineered food?		
	1996 (%)	*1998 (%)*
Support it to a great extent	6.0	6.0
Support it slightly	25.0	16.0
Neither support nor oppose it	16.0	15.0
Oppose it slightly	24.0	21.0
Oppose it to a great extent	27.0	37.0
Don't know	2.0	5.0

Given the above information carry out a chi-squared test to determine whether the claims made by the magazine are supported by the poll results. Comment on the limitations of this test.

(12)

Determine a 95% confidence interval for the change between 1996 and 1998 in the proportion of the population who oppose genetically engineered food slightly or to a great extent.

(8)

4

3. The infant mortality rate per 1000 live births and the percentage adult female literacy rate for a sample of countries are given in the table below.

Country	Infant mortality rate per 1000 live births	Percentage adult female literacy rate
A	160	71
B	116	79
C	94	47
D	87	70
E	84	71
F	83	84
G	83	80
H	60	88
J	50	86
K	43	90
L	35	95
M	31	88
N	27	93
P	25	96
R	22	93
Sum	1000	1231
Sum of squares	88268	103411
Sum of products	77068	

(i) Make a scatter plot of these data using the vertical axis for infant mortality rate.

(4)

(ii) Use the additional information contained in the table to fit a linear relation which could be used to predict infant mortality rate from adult female literacy rate.

(4)

(iii) Plot the fitted line on the scatter plot.

(2)

(iv) Modify your calculations to exclude country C from the fitted relation.

(4)

(v) Plot the new fitted line on the scatter plot.

(2)

(vi) What do you conclude about the effect of country C? What would be your recommendations about which equation should be used to predict infant mortality rate?

(4)

4. The quarterly deliveries of phosphate fertilizers to UK agriculture are shown on the worksheet for this question. A plot of the data is shown below.

Complete the calculation of the centred four point moving averages and the estimated seasonal effects on the worksheet provided. Use the estimated seasonal effects and the moving average trend values to complete the calculation of the residuals. Plot the residuals against the quarter number (i.e. 1, 2, 3 or 4).
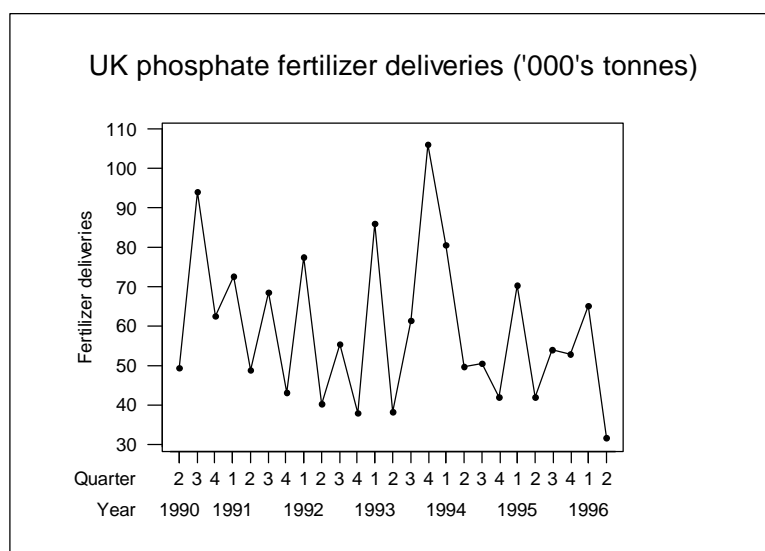
(14)

Comment on this residual plot.

(3)

Discuss whether deliveries for any quarters seem abnormal compared with the general pattern of deliveries.

(3)

UK phosphate fertilizer deliveries ('000's tonnes)

The worksheet for use with this question is supplied as a separate sheet of paper.

If you answer this question, you must hand the worksheet in with your answer book.

ENSURE YOU HAVE WRITTEN YOUR CANDIDATE'S NUMBER ON THE WORKSHEET.

5.   In a study of whether two forms of iron (Fe^{2+} and Fe^{3+}) are retained in different amounts within the body, 18 mice were randomly selected from a group of 36 to receive Fe^{2+} at a 0.3 millimolar concentration as an addition to their food and the remaining 18 mice received Fe^{3+} at the same concentration as an addition to their food. The percentage of each type of iron retained was measured for each mouse. The data are listed in the table below.

Fe3+	Fe2+
2.25	4.71
3.93	5.43
5.08	6.38
5.82	6.74
5.84	8.32
6.89	9.04
8.50	9.56
8.56	10.01
9.44	10.08
10.52	10.62
13.46	13.80
13.57	14.35
14.76	14.90
18.41	15.25
26.96	17.32
27.56	19.87
32.82	31.60
39.13	37.25

On a single diagram, draw boxplots for the two samples. Using this diagram, comment on

(i)   whether there appears to be a difference in the general level of retention of the two forms of iron

and

(ii)   whether the distributions of values appear symmetrical.

(10)

The investigators decided to perform a nonparametric test rather than the standard *t* test for comparison of means. Explain why this is an appropriate choice.

(3)

Carry out the appropriate test and state clearly any conclusions you reach.

(7)

6. A drug intended to reduce blood pressure was tested in the early stages of a clinical trial on 15 patients who had high blood pressure. For each patient both the systolic blood pressure and the diastolic blood pressure were measured before they were given the drug. They were then given the drug and both systolic and diastolic blood pressure were measured one hour later. Systolic blood pressure is measured as the heart contracts to pump blood around the body. Diastolic blood pressure is measured as the heart relaxes to allow blood to flow into it before the next contraction. Systolic blood pressure is always higher than diastolic blood pressure and the two are positively correlated.

The measured values and the differences between the values before and after taking the drug are given below.

Patient number	Systolic b.p. before	Systolic b.p. after	Systolic b.p. difference	Diastolic b.p. before	Diastolic b.p. after	Diastolic b.p. difference
1	210	201	9	130	125	5
2	169	165	4	122	121	1
3	187	166	21	124	121	3
4	160	157	3	104	106	-2
5	167	147	20	112	101	11
6	176	145	31	101	85	16
7	185	168	17	121	98	23
8	206	180	26	124	105	19
9	173	147	26	115	103	12
10	146	136	10	102	98	4
11	174	151	23	98	90	8
12	201	168	33	119	98	21
13	198	179	19	106	110	-4
14	148	129	19	107	103	4
15	154	131	23	100	82	18
Sum	2654	2370	284	1685	1546	139
Sum of squares	4775502	380062	6518	190817	161548	2327

Perform an appropriate test to determine whether the investigation provides evidence that the drug is effective in reducing the mean systolic blood pressure. Determine a 95% confidence interval for the mean change in systolic blood pressure.

(10)

Investigate whether the drug is equally effective in reducing diastolic blood pressure.

(5)

Make a scatter plot which shows the relation between the change in diastolic blood pressure and the change in systolic blood pressure. What can be deduced from the scatter plot about the size of the change in diastolic blood pressure compared to the size of the change in systolic blood pressure? (No detailed calculations are required.)

(5)

7. The undergraduate population of Exminster University is made up as shown in the table below.

Faculty	Male	Female	Total
Medicine	772	535	1307
Science	1161	913	2074
Engineering	1031	404	1435
Social Sciences	773	947	1720
Arts	694	772	1466
Total	4431	3571	8002

Each student has an identification number which consists of the year of entry to the university (for example 1996), followed by a single digit Faculty code (1 = Medicine, 2 = Science, …, 5 = Arts), followed by a five digit serial number. Thus a student who is in the Faculty of Arts and entered the university in 1997 might have the identification number 1997500357. The university administration holds student records on a computer data base and could easily provide a list of all students, with each entry in the list containing the following information:

identification number, student's name, sex, faculty, year of entry, current address.

(i) Describe in detail how to take the following samples of students:

(a) a simple random sample of 800 students;

(b) a sample of 800 students stratified by faculty and sex with sample sizes in the strata to be proportionally allocated and a random sampling scheme used;

(c) a quota sample of 800 students based on appropriate faculty and sex quotas.

(10)

(ii) Compare the merits and disadvantages of methods A and B as described below in carrying out the survey. Comment briefly on which method you would recommend and why.

Method A: a printed questionnaire is to be distributed to the students selected to participate in the survey. They complete the questionnaire in their own time and they must return it to a collection point in the administration department.

Method B: students selected are interviewed personally by trained interviewers hired from a market research company.

(10)

Turn over

8. The amount of water absorbed by different types of resin used for dental fillings was investigated. Ten blocks of resin were prepared with glass fibre added to reinforce the resin. Ten blocks of resin were prepared without the addition of glass fibre. Five blocks of each type were selected at random and hardened by chemical treatment. The remaining blocks were hardened by heat treatment. All blocks were placed in water for the same length of time and then the amount of water absorbed by each block was measured, the results being expressed in micrograms of water per cubic millimetre of resin.

The experiment was thus a two factor experiment with five replicates at each factor level combination. The two factors are (1) glass fibre which is either present or absent and (2) method of hardening which is either chemical treatment or heat treatment.

The sample mean and standard error of the mean (s.e.m.) of the water absorption for each of the factor level combinations are given below.

		Hardening method			
		Heat		*Chemical*	
		mean	*(s.e.m.)*	*mean*	*(s.e.m.)*
	Not added to resin	17.590	(0.417)	16.970	(0.380)
Glass fibre					
	Added to resin	20.590	(0.793)	16.390	(0.252)

Complete the following analysis of variance table and write a brief report on the findings of the analysis. Your report should include an appropriate diagram to illustrate any interaction there is between the factors.

Analysis of Variance for water absorption by dental resin.

Source	D.F.	S.S.	M.S.	V.R.
Glass fibre	1	*	*	*
Hardening	1	*	*	*
Interaction	1	*	*	*
Error	*	20.21	*	
Total	*	72.59		

(20)

		Deliveries	M.Av.	Y−M.Av.	Seasonal	Predicted	Residual
1990	2	49.5	*	*	-14.94	*	*
	3	94.2	*	*		*	*
	4	62.6	69.6750	-7.0750	-2.1675	67.5075	-4.9075
1991	1	72.7	66.3875	6.3125			
	2	48.9	60.7250	-11.8250	-14.94	45.7850	3.1150
	3	68.5					
	4	43.0			-2.1675		
1992	1	77.6	55.6750	21.9250			
	2	40.2	53.4000	-13.2000	-14.94	38.4600	1.7400
	3	55.3	53.8250	1.4750			
	4	38.0	54.6125	-16.6125	-2.1675	52.4450	-14.4450
1993	1	86.0	55.1125	30.8875			
	2	38.1	64.3875	-26.2875	-14.94	49.4475	-11.3475
	3	61.4	72.2125	-10.8125			
	4	106.1	72.9750	33.1250	-2.1675	70.8075	35.2925
1994	1	80.5	73.0500	7.4500			
	2	49.7	63.6625	-13.9625	-14.94	48.7225	0.9775
	3	50.4	54.3750	-3.9750			
	4	42.0	52.1375	-10.1375	-2.1675	49.9700	-7.9700
1995	1	70.3	51.6375	18.6625			
	2	42.0			-14.94	38.5225	3.4775
	3	54.1					
	4	52.9			-2.1675	50.0700	2.8300
1996	1	65.1	*	*		*	*
	2	31.7	*	*		*	*

Calculation of seasonal effects

Q1	Q2	Q3	Q4	
6.3125	-11.8250		-7.0750	
21.9250	-13.2000	1.4750		
30.8875	-26.2875	-10.8125	-16.6125	
7.4500	-13.9625	-3.9750	33.1250	
18.6625			-10.1375	
*	*	*		
17.0475		-0.755		-1.63
	-14.94		-2.1675	