

Paper Reference(s)

6688

Edexcel GCE

Statistics S6

Advanced/Advanced Subsidiary

Wednesday 22 May 2003 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination

Answer Book (AB16)

Graph Paper (ASG2)

Mathematical Formulae (Lilac)

Items included with question papers

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S6), the paper reference (6688), your surname, other name and signature.

Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.

Full marks may be obtained for answers to ALL questions.

This paper has six questions.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. The model associated with a one-way analysis of variance can be written as

$$x_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

where x_{ij} denotes the j th observation on the i th treatment ($i = 1, 2, \dots, k$) and $\varepsilon_{ij} \sim N(0, \sigma^2)$.

(a) Write down the name of the experimental design used to collect data for a one-way analysis of variance. (1)

(b) Explain what you understand by the α_i and write down any condition associated with them. (2)

A one-way analysis of variance based on a total of 17 observations from 4 different treatments gave a within treatments sum of squares of 165.23.

(c) Find an unbiased estimate of σ^2 for this analysis. (3)

2. A university librarian selected a random sample of 7 students from the science faculty and recorded the number of books each student borrowed in a particular month. The results were as follows.

11, 13, 6, 8, 10, 17, 5

During the same month the librarian selected a random sample of 8 students from the arts faculty and the number of books each of these students borrowed that month was as follows.

19, 9, 20, 15, 16, 12, 14, 18

(a) Using the Wilcoxon rank sum test at the 5% level of significance, test whether or not there is any difference between the median number of books borrowed in these two faculties. (7)

(b) Explain briefly how your test would be modified if 50 students had been randomly selected from each faculty. (2)

3. Equal weights of eight different varieties of beans were purchased for a fertiliser experiment. One half by weight of each variety was treated by soaking in fertiliser and then planted. The other half was planted untreated. The total yields (in coded units) of the adult plants were as shown in the table.

	Variety							
	1	2	3	4	5	6	7	8
Treated plants	13.2	13.0	12.8	13.6	13.0	12.6	14.1	14.4
Untreated plants	12.0	12.4	13.2	12.2	12.0	12.4	13.4	12.8

- (a) Stating your hypotheses clearly and using a 5% level of significance, analyse these data using

- (i) the Sign test,
 (ii) the Wilcoxon matched pairs signed-ranks test.

(11)

- (b) Comment on your results.

(1)

An alternative test that might have been used to analyse these data is the paired t -test.

- (c) Write down an assumption that would have to be made in order to use this test.

(1)

4. An economist collected data relating to the percentage unemployment x , and the annual percentage increase in wages y over a period of 10 years. The data were summarised as follows.

$$\Sigma x = 18.9, \quad \Sigma y = 35.2, \quad \Sigma x^2 = 37.01, \quad \Sigma y^2 = 132.22, \quad \Sigma xy = 64.7$$

- (a) Calculate S_{xx} , S_{xy} and S_{yy} .

(3)

The product moment correlation coefficient for these data is -0.558 . It has been suggested that low unemployment and a low rate of wage increase cannot exist together.

- (b) Explain whether or not the value of the correlation coefficient for these data justifies this suggestion.

(2)

The equation of the regression line of y on x for these data is $y = 6.200 + bx$.

- (c) Find the value of b .

(2)

- (d) Test, at the 5% level of significance, whether or not the regression coefficient is less than -1.00 .

(7)

5. Three varieties of potato were compared for yield. An experiment was conducted by assigning each variety at random to 3 equal-size plots at each of 4 different locations. The yields, in coded units, from each plot were as shown in the table.

Location	Variety of potato			Total
	A	B	C	
1	18	13	12	43
2	20	23	21	64
3	14	12	9	35
4	11	17	10	38
Total	63	65	52	180

- (a) Write down the name of the experimental design that would have been used to obtain these results.

(1)

An analysis of these data found the total sum of squares to be 238.0 and the residual sum of squares to be 42.167 to 3 decimal places.

- (b) Complete the analysis of variance and test whether or not there are differences between

(i) variety of potato,

(ii) location.

State your hypotheses clearly and use a 5% level of significance.

(13)

- (c) Comment on your results.

(1)

6. A machine is set to produce pins such that their target length is 3.00 cm. In order to set up a control chart for the mean and the range, random samples of size 4 were taken at a time when the machine was assumed to be working satisfactorily. The length of each pin was measured and the results are summarised in the table.

	Sample							
	1	2	3	4	5	6	7	8
Mean	3.10	2.99	3.05	3.11	3.00	3.01	3.07	3.06
Range	0.13	0.04	0.14	0.13	0.17	0.10	0.22	0.16

- (a) Use the values of the range to show that an estimate of σ is 0.0662 to 3 significant figures. (3)
- (b) Draw a control chart for the mean using 95% warning limits and 99.8% action limits. (6)
- (c) Draw a control chart for the range. (3)
- (d) Plot the means and the ranges on your control charts. (2)
- (e) Comment on the state of the machine. (4)

END

**6688 STATISTICS S6, JUNE 2003
TABLE NEEDED FOR QUESTION 6**

CONTROL CHARTS VARIABILITY

For range charts multiply σ by the appropriate value of D.

For standard deviation charts multiply σ by the appropriate value of E.

To obtain an estimate of σ multiply the mean range by the appropriate value of b.

Normal distribution is assumed.

Sample size	D 0.999	D 0.975	D 0.025	D 0.001	E 0.999	E 0.975	E 0.025	E 0.001	b
2			3.170	4.654			2.24	3.29	0.8862
3	0.060	0.303	3.682	5.063	0.03	0.16	1.92	2.63	0.5908
4	0.199	0.595	3.984	5.309	0.09	0.27	1.76	2.33	0.4857
5	0.367	0.850	4.197	5.484	0.15	0.35	1.67	2.15	0.4299
6	0.535	1.066	4.361	5.619	0.20	0.41	1.60	2.03	0.3946
7	0.691	1.251	4.494	5.730	0.25	0.45	1.55	1.93	0.3698
8	0.835	1.410	4.605	5.823	0.29	0.49	1.51	1.86	0.3512
10	1.085	1.674	4.784	5.973	0.36	0.55	1.45	1.76	0.3249
12	1.293	1.884	4.925	6.096	0.41	0.59	1.41	1.69	0.3069