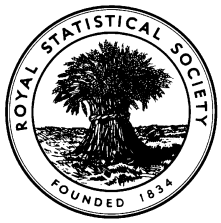
EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY
(formerly the Examinations of the Institute of Statisticians)



HIGHER CERTIFICATE IN STATISTICS, 2004

Paper II : Statistical Methods

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.*

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* ${}^{n}C_{r}$.

This examination paper consists of 8 printed pages **each printed on one side only**.
This front cover is page 1.
Question 1 starts on page 2.

There are 8 questions altogether in the paper.

1. (i) State an analysis of variance model suitable for a randomised block design. Explain clearly what each term in the model represents and state any assumptions required for the analysis based on this model to be valid.

(6)

(ii) An experiment was conducted to investigate the toxic effects of four different chemical compounds *A*, *B*, *C* and *D* on the skin. Four adjacent regions, each a square of side 3 cm, were marked on the left forearm of each of six subjects, and each chemical was applied to each subject, choosing the sites of application of the chemicals at random for each subject. After three hours, the skin was examined and scored from 0 to 10 depending on the degree of irritation, with 0 representing no irritation and 10 severe irritation. The data are given in the following table.

Subject 1	*Subject 2*	*Subject 3*	*Subject 4*	*Subject 5*	*Subject 6*
D 5	*A* 7	*B* 2	*B* 4	*A* 3	*D* 6
B 3	*C* 4	*A* 1	*A* 6	*B* 1	*B* 7
A 3	*D* 7	*D* 3	*C* 6	*D* 2	*C* 3
C 2	*B* 6	*C* 1	*D* 7	*C* 2	*A* 5

Carry out a suitable analysis of these data. Explain clearly your conclusions and comment on the differences, if any, between the toxic effects of the compounds. Discuss briefly whether all the necessary assumptions can safely be made.

(14)

2. A supermarket has a policy of only buying tomatoes from growers who can supply tomatoes that have a mean diameter of 3.0 cm and a standard deviation of no more than 0.5 cm. A representative of the supermarket goes to visit a potential new supplier and selects a random sample of 16 tomatoes from the tomato grower's greenhouse. The diameter of each tomato is measured and the data are as follows, recorded for convenience in ascending order.

2.2 2.3 2.5 2.6 2.6 2.7 2.9 3.0 3.2 3.3 3.4 3.6 3.6 3.8 3.8 3.9

By constructing suitable confidence intervals, analyse these data to establish whether the tomatoes provided by the grower will meet the supermarket's requirements, clearly stating any assumptions on which your analysis depends. Write a short report to the board of directors outlining your recommendations concerning whether or not to use this tomato grower to supply tomatoes for sale in the supermarket.

(16)

The supermarket representative suggests that the simplest way to select the sample would be to pick two tomato plants at random and select eight tomatoes at random from each. Comment on the suitability of this method.

(4)

Turn over

3. (i) Discuss the advantages and disadvantages of using non-parametric rather than parametric methods in statistical analyses.

(6)

(ii) Ten boys and ten girls were selected at random from a large group of 16-year-old school children. They were each asked to estimate how many hours each week they spent listening to music. The results are shown in the following table.

	Time in hours									
Boys	21	6	3	18	28	16	5	30	2	0
Girls	1	7	9	16	9	0	3	12	21	8

(a) Draw separate stem and leaf diagrams for the data obtained from the boys and girls and hence comment on the distribution of the measurements in each group. Why would a t test be unsuitable for analysing these data?

(6)

(b) Using a suitable non-parametric test, investigate whether there is sufficient evidence to indicate a difference between the distributions of the number of hours reported listening to music per week for boys and for girls.

(8)

4. The yields on the ordinary shares of 95 large United Kingdom companies at close of trading on 6 March 2003, as given in the *Financial Times*, were as follows.

Yield (%)	*Number of companies*
≥ 0 but < 1	5
≥ 1 but < 2	8
≥ 2 but < 3	13
≥ 3 but < 4	18
≥ 4 but < 5	19
≥ 5 but < 7.5	21
≥ 7.5 but < 10	8
≥ 10 but < 15	2
≥ 15 but < 20	1
Total	95

Source: "Financial Times", 6 March 2003.

You may treat the yields as constituting a random sample from a large population.

(i) Draw a histogram depicting the above data.

(7)

(ii) State the modal class interval and estimate the mean, median and standard deviation of the observations.

(6)

(iii) Why is it only possible to **estimate** the mean, median and standard deviation rather than **find** them?

(3)

(iv) Let p be the proportion of companies in the underlying population for which the shares yield 5 per cent or more. Construct a 95% confidence interval for p.

(4)

5. (i) Explain the meaning of the following terms used in hypothesis tests.

(a) Type I error.

(2)

(b) Type II error.

(2)

(c) Level of significance.

(2)

(d) Power.

(2)

(ii) A manufacturer of coffee uses a machine to fill jars. The machine is calibrated so that the amount of coffee dispensed into each jar is Normally distributed with mean (μ) 200 grams and standard deviation (σ) 15 grams. Each hour, a random sample of 9 jars is taken from the previous hour's output and the sample mean amount ($\bar{x}$) is evaluated. If the sample mean lies in the interval $190 < \bar{x} < 210$, the previous hour's output is accepted, otherwise it is rejected and the machine is recalibrated before continuing.

(a) Calculate the probability of committing a type I error by rejecting the previous hour's output when $\mu = 200$ grams and $\sigma = 15$ grams.

(6)

(b) Calculate the probability that the previous hour's output will be accepted when $\mu = 216$ grams and $\sigma = 15$ grams.

(6)

6. (i) Using examples to illustrate your answers, discuss the uses of the following distributions in tests of hypotheses relating to measures of location.

(a) The Normal distribution.

(4)

(b) The *t* distribution.

(4)

(ii) A manufacturer of rechargeable mobile phone batteries claims that on average its new improved batteries will operate for 600 minutes longer before needing to be recharged than the standard batteries sold by a leading competitor. It would not be economic to change to the new batteries unless they do in fact last at least 600 minutes longer than the standard batteries. A random sample of 10 batteries of each type was obtained and the batteries were fitted into identical mobile phones that were then set into operational mode. The numbers of minutes that the various phones were able to operate before their batteries needed recharging were recorded and the data are given below.

Standard batteries, numbers of minutes:

2342, 2168, 2115, 1954, 2358, 2040, 1985, 2076, 2224, 2328.

New batteries, numbers of minutes:

2670, 2964, 3000, 2886, 2994, 2658, 2788, 2924, 2734, 2682.

Let μ_1 denote the population mean length of time, in minutes, that the standard batteries will operate before needing to be recharged, and let μ_2 denote the corresponding quantity for the new batteries. Using an appropriate statistical test, investigate the hypothesis that $\mu_2 = \mu_1 + 600$.

(10)

State the assumptions on which your statistical test is based.

(2)

Turn over

7. (i) A psychologist wished to examine the degree of association between intelligence and the ability to think laterally. An experiment was conducted in which each member of a random sample of 12 subjects was given both an intelligence test and a test of lateral thinking. The scores obtained by each subject on each test are given in the following table. The intelligence test has a maximum score of 150 while the lateral thinking test has a maximum score of 10.

Subject	1	2	3	4	5	6	7	8	9	10	11	12
Intelligence test score	121	148	108	137	141	124	131	115	118	110	132	127
Lateral thinking test score	3	9	6	7	8	2	5	5	6	4	8	7

The psychologist analyses these data using Spearman's rank correlation coefficient.

(a) Some subjects had the same lateral thinking scores. How, if at all, should this be allowed for in the analysis?

(b) Calculate Spearman's coefficient between these two test scores, and test it for statistical significance.

(c) What does your result indicate about the association between intelligence and the ability to think laterally?

(12)

(ii) The psychologist suspects that females have a greater ability to think laterally than males of equal intelligence. To test this hypothesis, each member of a random sample of 50 males was paired with a female who achieved a similar score on a standard intelligence test. Each subject was given five minutes to complete an identical lateral-thinking task. Once the time limit had elapsed, it was noted whether or not each subject had successfully completed the task. The data are summarised in the following table.

		Males	
		Successful	*Unsuccessful*
Females	*Successful*	16	15
	Unsuccessful	9	10

Carry out a suitable analysis of these data and comment on your results.

(8)

8. (a) A zoologist conducted an investigation into the nesting habits of a particular species of bird that builds its nests in medium sized shrubs. One question of particular interest was whether the birds have any directional preference when building their nests. To investigate this, pairs of birds, chosen at random, were monitored and the directional positions of their nests were recorded. The data are summarised in the following table.

Nest position	N	NE	E	SE	S	SW	W	NW
Number of nests	27	24	35	33	38	33	24	26

Test the hypothesis that the birds have no directional preference in positioning their nests, using a χ^2 goodness of fit test.

(8)

What limitations, if any, does your test have? Why would a Kolmogorov-Smirnov test have been inappropriate here?

(4)

(b) An art gallery is due to celebrate its 50th anniversary in 2005. As part of its celebrations, it wishes to commission a new sculpture to be displayed in the gallery. To find a suitable sculpture, it decided to run a competition in which it invited local artists to submit designs. A panel of experts selected a short-list of three designs for the gallery to choose from. To assist in the final decision, the gallery conducted a survey in which a random sample of local adults were sent copies of the three designs and asked to indicate their preference. The replies received from male and female adults are given in the following table.

	Preferred design		
	A	*B*	C
Males	129	24	47
Females	126	44	55

Carry out a suitable analysis of these data to investigate whether the opinion of adults concerning the preferred design is the same for males and females.

(8)