UNIVERSITY OF ABERDEEN **SESSION 2004–05**

DEGREE EXAMINATION

MX4035 Non-Parametric Tests and Other Topics

Wednesday 19 January 2005 (9am to 11am)

Only calculators approved by the Department of Mathematical Sciences may be used in this examination. Calculator memories must be clear at the start of the examination.

Marks may be deducted for answers that do not show clearly how the solution is reached.

Answer THREE questions. All questions carry equal weight.

1. Twelve sets of identical twins are given psychological tests to measure (in some sense) the amount of aggressiveness in their personalities, higher scores indicting greater aggressiveness.

The results were as follows:

Pair No.	1	2	3	4	5	6	7	8	9	10	11	12
Firstborn Twin	86	71	77	68	91	74	77	91	70	71	88	87
Second Twin	88	77	76	64	96	72	65	90	65	80	81	72

(a) Conduct a suitable rank test, in the Wilcoxon family, of the hypothesis that firstborn and second twins are, on average, equally aggressive.

(b) Use a rank method to find an approximate 99% confidence interval for the median of the difference in aggressiveness scores (of firstborn twin minus that of second twin), explaining briefly how your confidence interval is obtained.

2. **(a)** (i) Define the one-sample Kolmogorov-Smirnov test statistic for assessing the goodness-of-fit of a set of data $x_1, x_2, \ldots, x_n$ to a distribution with continuous distribution function $F(x)$ $(-\infty < x < \infty)$.

(ii) According to a certain scientific theory, the probability density function of the time X (in seconds) of a chemical process is

$$f(x) = \begin{cases} 0.02x & \text{for } 0 \leq x \leq 10 \\ 0 & \text{otherwise.} \end{cases}$$

In her experiments, a scientist has obtained the following ten values of X (in non-decreasing order):

$$1.7,\ 5.3,\ 7.6,\ 8.9,\ 9.0,\ 9.1,\ 9.3,\ 9.6,\ 9.8,\ 9.9$$

Use a one-sample Kolmogorov-Smirnov test, at the 5% significance level, to test the above theory. (Graph paper is provided, if required.)

(b) (i) Define the one-sample Cramér-von Mises statistic for the goodness-of-fit problem of **(a)**(i) above.

(ii) Apply this procedure to test the fit of the data of **(a)**(ii) above, using the 5% significance level. You are given the following Table of upper percentage points of the one-sample Cramér-von Mises statistic:

Per cent	10	5	1	0.1
Upper percentage point	0.3473	0.4614	0.7435	1.1679

3. **(a)** You are given that the Asymptotic Relative Efficiency (A.R.E.) of the median test relative to the Mann-Whitney test when the underlying variable has density $f(x)$ and median μ is

$$\frac{[f(\mu)]^2}{3\left\{\int_{-\infty}^{\infty} [f(x)]^2\, dx\right\}^2}.$$

Show that, if the underlying variable has a mode at μ, the A.R.E. cannot fall below $\frac{1}{3}$.

(b) Calculate the A.R.E. described in **(a)** when the underlying density has the following forms:

(i) normal;

(ii) logistic.

(c) (i) Describe briefly how you would conduct tests of normality for a large set of observations using the sample skewness (g_1) and the sample kurtosis (g_2).

(ii) Apply these tests to a set of $n = 200$ observations for which $g_1 = 0.153$ and $g_2 = 1.409$.

4. **(a)** In an investigation into the seeding of rainfall in the Snowy Mountains, the following data were collected, where x_i = years of seeding, Y_i = rainfall factor (in suitable units):

x_i	1	2	3	4	5
Y_i	1.26	1.27	1.12	1.16	1.03

Using the Theil-Kendall method, find a point estimate of the slope β_1 of the regression line $y = \beta_0 + \beta_1 x$, and an approximate 95% confidence interval for this slope.

(b) Two large firms in the computer industry, A and B, were asked to rank 7 brands of software in order of preference (rank 1 indicating the best software, and so on). The results were:

		Brand of software						
		1	2	3	4	5	6	7
Company	A	4	5	1	2	6	3	7
	B	3	4	2	1	6	5	7

Using a 5% significance level, test the hypothesis that there is no association between the preferences of the two firms. If you reject this hypothesis, state whether this is due to a tendency for the firms to agree or a tendency for them to disagree.

(c) Random samples of three different brands of an electrical appliance were tested to see how long they lasted in continuous use. The following data give the times to failure of the appliances (in days):

Brand A	84	80	73	82
Brand B	81	79	71	
Brand C	67	64	78	

Is there any evidence of significant differences between the brands?