

SMA2204

NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF APPLIED MATHEMATICS

B.Sc. PART II HONOURS EXAMINATIONS 2005

SMA 2204 – APPLIED STATISTICS

MAY 2005

3 HOURS (100 Marks)

This paper has 6 pages

Answer **FOUR** questions: Question 1 in **SECTION A** (28 Marks) and **THREE** from **SECTION B** (24 Marks each). Where a question contains subdivisions, the mark value of each subdivision is indicated in brackets.

Candidates are expected to spend not more than one hour on Question 1. Calculators may be used. Statistical Tables and graph paper are provided, however, Statistical Tables should not be marked or taken out of the examination room. **GOOD LUCK!**

SECTION A (COMPULSORY)

1. (a) In a study of the relationship between diabetes and birth defects among the Pima Indians of Arizona, Pima mothers were classified by diabetic status as being either **nondiabetic**, **prediabetic**, (i.e. not currently having the disease but showing indications of developing it later), or **diabetic**. Among 1207 children born to these mothers, the observed frequencies of birth defects are shown below:

MOTHER'S DIABETIC STATUS	CHILD'S BIRTH DEFECT STATUS	
	One or more defects	No defects
Nondiabetic	31	754
Prediabetic	13	362
Diabetic	9	38

Test at the 1% level of significance whether there is any relationship between the incidence of birth defects among the Pima Indians and the diabetic status of the mother. Explain clearly your findings. **(8 marks)**

- (b) (i) Five measurements of the tar content of a certain kind of cigarette yielded 14.5, 14.2, 14.4, 14.3, and 14.6 mg/cig. Show that at the 5% level of significance the null hypothesis $\mu = 14.0$ must be rejected in favour of the alternative hypothesis $\mu \neq 14.0$. Assume that the data are a random sample from a normal population.

(4 marks)

- (ii) Suppose that in (i) above the first measurement is recorded incorrectly as 16.0 instead of 14.5. Show that this will reverse the conclusion drawn in (i), and explain the apparent paradox that even though the difference between the sample mean and $\mu = 14.0$ has increased, it is no longer significant. (4 marks)
- (c) To find whether the inhabitants of two South Pacific islands may be regarded as having the same racial ancestry, an anthropologist determines the cephalic indices of six adult males from each island, getting $\bar{x}_1 = 77.4$, $\bar{x}_2 = 72.2$, and the corresponding standard deviations $s_1 = 3.3$ and $s_2 = 2.1$. Use the 0.01 level of significance to test whether the difference between the two sample means can reasonably be attributed to chance. Assume that the populations are normal. What assumption have you made concerning the population variances? Use the 0.05 level of significance to test whether this assumption holds or not. (8 marks)
- (d) If X and Y are random variables with a joint distribution such that $\text{Var}(X) = 4$, $\text{Var}(Y) = 9$, and $\text{Cov}(XY) = 2$; find
- (i) $\text{Var}(Z)$, where $Z = 3X - 4Y + 8$ (1 mark)
- (ii) $\text{Cov}(P, Q)$, where $P = X - 2Y$, and $Q = 3X + Y$. (3 marks)

SECTION B (Answer any THREE QUESTIONS)

2. (a) Three brands of batteries are under study. It is suspected that the life (in weeks) of the three brands is different. Five batteries of each brand are tested and the following results obtained:

Brand 1	Brand 2	Brand 3
100	76	108
96	80	100
92	75	96
96	84	98
92	82	100

- (i) Test at the 5% level of significance, whether there are any significant differences in the mean lives of the three brands of batteries. (8 marks)
- (ii) If over the years the traditional brand used has been **Brand 1**; compare the other two brands with it at the 5% level of significance. (6 marks)

- (b) A chemist wishes to test the effect of four chemical agents on the strength of a particular type of cloth. Because she suspects that there could be variability from one bolt to another she selects four bolts and applies all four chemicals in random order to each bolt. The following tensile strengths are obtained:

Chemical	Bolt of Cloth			
	1	2	3	4
A	73	68	74	71
B	73	67	75	72
C	75	68	78	73
D	73	71	75	75

Test the null hypothesis of no significant differences in tensile strength due to the four different chemical agents. Use the 0.01 level of significance. (10 marks)

3. (a) It is claimed that a new diet will reduce a person's weight by 4.5 kilograms on the average in a period of two weeks. The weights of seven women who followed this diet were recorded before and after a two-week period.

	Woman						
	1	2	3	4	5	6	7
Weight before	58.5	60.3	61.7	69.0	64.0	62.6	56.7
Weight after	60.0	54.9	58.1	62.1	58.5	59.9	54.4

Assuming the distribution of weights to be approximately normal

- (i) Test the manufacturer's claim that the new diet will reduce a person's weight by 4.5kg on average. Use the 5% level of significance. (7 marks)
- (ii) Find the 95% confidence interval for the true mean difference in the weight. Draw your conclusion to the manufacturer's claim based on the 95% confidence interval. (3 marks)
- (b) Two protein diets were allocated randomly to male chicks of the same age and breed. The weights (in grams) of chicks after four weeks are as follows:

High Protein Diet: 150 136 144 168 152 140
Low Protein Diet: 121 109 115 106 99

- (i) Carry out a two-sample (pooled) t-test at the 5% level of significance to test whether there is a significant difference between the two protein diets. (6marks)

- (ii) Carry out a one-way analysis of variance (ANOVA) to test whether there is a significant difference between the two protein diets at the 5% level of significance. **(6marks)**
- (iii) Explain how the test in (i) is related to the test in (ii). **(2marks)**

4. Raw material used in the production of a synthetic fibre is stored in a place which has no humidity control. Measurements of the relative humidity and the moisture content of samples of the raw material (both in percentages) on 12 days yielded the following results:

Day	Relative Humidity (%)	Moisture Content (%)
1	46	12
2	53	14
3	37	11
4	42	13
5	34	10
6	29	8
7	60	17
8	44	12
9	41	10
10	48	15
11	33	9
12	40	13

- (a) Draw a scatter diagram of the response (dependent) variable, y , against the regressor (independent) variable, x . Fit the least squares straight line to these data and draw your fitted line on your scatter diagram. **(10 marks)**
- (b) Carry out an analysis of variance (ANOVA) to test at the 5% level of significance whether the slope is significantly different from zero. From the ANOVA table, compute the coefficient of determination, r^2 , and interpret it. **(8 marks)**
- (c) Predict the moisture content when the relative humidity is 38%. Compute the standard error of the predicted value and hence find the 95% confidence interval of the predicted value. **(6 marks)**

5. A bacteriologist is interested in the effect of three different culture mediums and three different times on the growth of a particular bacterium. She performs three replicates for each treatment combination and obtains the following results:

Time	Culture Medium		
	1	2	3
12 hr	52	74	63
	59	70	67
	63	76	61
18 hr	71	74	73
	74	73	75
	75	71	72
24 hr	76	76	72
	72	75	73
	74	78	70

- (a) Construct a two-way table of totals. (2 marks)
- (b) Carry out a preliminary ANOVA to test for the significance of treatment combinations at the 5% level of significance. (6 marks)
- (c) Carry out a complete ANOVA to test for the significance of the main effects and interaction at the 5% level of significance. (10 marks)
- (d) Give an appropriate table of means and appropriate standard errors for carrying out statistical comparisons. (6 marks)
6. The following data constitute final examination marks for **SMA 2204 – Applied Statistics** (taught by Mr Mlilo) over a three-year period (2002, 2003, and 2004):

2002: 88 49 31 86 41 26 52 39 46 40 37 58 97 43 90 34 54 28 48
40 89 22 32 35 45

2003: 54 48 36 53 45 31 49 42 46 44 41 51 63 45 59 39 50 33 47
43 57 27 37 40 46

2004: 59 51 0 58 45 0 53 41 50 44 38 56 68 47 64 32 55 0 50 42
62 0 0 36 49

Mr Mlilo was interested in determining whether there was any detectable trend in student performance over the three-year period. He asked his teaching assistant to find out by analysing the data. The teaching assistant calculated the sample means for the three sets of data and found that the means were **50.0** for 2002, **45.04** for 2003, and **40.0** for 2004. The teaching assistant understandably concluded that the student performance had deteriorated over the three-year period. However, Mr Mlilo was somewhat disturbed by this finding since he felt his teaching had improved over this time.

- (a) Critique the analysis done by Mr Mlilo's teaching assistant. **(6 marks)**
- (b) Being a devotee of the exploratory analysis method, Mr Mlilo asks you to construct for each set of data a **stem and leaf plot** and a **box plot**. What conclusion do you draw from these plots? **(18 marks)**

-----***** END OF QUESTION PAPER *****-----