NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

FACULTY OF COMMERCE

DEPARTMENT OF INSURANCE AND ACTUARIAL SCIENCE

B. COM ACTUARIAL SCIENCE

ACTUARIAL STATISTICS II [CIN 2211]

APRIL/MAY 2006 SECOND SEMESTER EXAMINATION

TIME ALLOWED: THREE (3) HOURS

INSTRUCTIONS TO CANDIDATES

1. Answer ALL 8 questions
2. Write your student number on the answer booklet
3. Mark allocations are shown in brackets
4. Begin your answer to each question on a separate sheet
5. Credit will be awarded for clarity of answers
6. All numerical computations must be clearly shown

ADDITIONAL MATERIAL

- An electronic calculator
- Actuarial Examination Tables

Q1 Consider the displayed bivariate probability distribution for the discrete random variables X and Y (x = 0, 1, 2, 3 ; y = 0, 1, 2).

		X			
		0	1	2	3
	0	.01	.03	.05	.02
Y	1	.10	.12	.15	.08
	2	.05	.16	.20	.03

Calculate the conditional expectation E(Y/ X = 0) **[5]**

Q2 The number of claims, N, which arise under a group of policies in a year has a Poisson distribution with mean 200. The claim amounts are independent of one another and of N, and have a common distribution which is normal with mean \$250 and standard deviation \$40.

Let Y be the total claim amount arising from this group of policies in a year.

(i) Show that $E(Y|N) = 250N$ and $V(Y|N) = 1600N$. **[2]**

(ii) Hence, or otherwise, find the mean and standard deviation of Y. **[4]**

[Total 6]

Q3 A random sample of 50 pairs of observations is taken from a bivariate normal distribution with correlation coefficient ρ. The sample correlation coefficient is 0.58.

By using Fisher's transformation, calculate a 95% confidence interval for ρ. **[7]**

Q4 A market analyst is investigating the experience of two insurance companies, A and B, with regard to a particular class of insurance business (which is written by both companies). The analyst has access to a limited amount of information - a random sample of 10 recent claim amounts from company A and a random sample of 8 recent claim amounts from company B.

The data follow (amounts in units of \$100), with summaries:

A: 27.6 43.0 32.7 22.1 31.5 55.4 32.2 10.7 38.7 29.9

B: 31.2 26.8 11.2 22.8 36.5 28.9 13.2 25.1

Company A: n = 10, $\Sigma x = 323.8$, $\Sigma x^2 = 11772.90$
Company B: $n = 8$, $\Sigma x = 195.7$, $\Sigma x^2 = 5308.67$

(i) (a) Plot the two sets of data on a simple diagram for comparison purposes, and comment on the assumption that these samples may be regarded as coming from normal distributions.

(b) The analyst is quite happy to assume that the variation in claim amounts is the same for both companies.

Perform a test to justify this assumption. **[7]**

(ii) (a) Calculate a two-sided 95% confidence interval for the difference between the mean claim amounts for all such business for the two companies, and comment briefly on the result.

(b) Perform a test to investigate whether the mean claim amount for all such business for company A exceeds that for company B.

(State your approximate probability-value and your conclusion clearly). **[8]**

[Total 15]

Q5 Independent random samples of size n_1 and n_2 are taken from the normal populations $N(\mu_1, \delta_1^2)$ and $N(\mu_2, \delta_2^2)$. Let the sample means be $\overline{X}_1$ and $\overline{X}_2$ and the sample variances be S_1^2 and S_2^2.

You may assume that $\overline{X}_i$ and S_i^2 are independent and distributed as follows:

$$\overline{X}_i \sim \mathrm{N}(\mu_i, \frac{\delta_i^2}{n_i}) \text{ and } \frac{(n-1)S_i^2}{\delta_i^2} \sim \chi^2_{n_i-1} \quad i = 1, 2.$$

(i) It is required to construct a confidence interval for $(\mu_1 - \mu_2)$ the difference between the population means.

(a) Suppose that δ_1^2 and δ_2^2 are known. State the distribution of $(\overline{X}_1 - \overline{X}_2)$ and write down a suitable pivotal quantity together with its sampling distribution. Hence write down a 95% confidence interval for $(\mu_1 - \mu_2)$

(b) Suppose that δ_1^2 and δ_2^2 are unknown but are known to be equal. State the definition of a t_k variable in terms of independent $N(0, 1)$ and χ_k^2 variables and use it to develop a suitable pivotal quantity. Hence write down a 95% confidence interval for $(\mu_1 - \mu_2)$ **[7]**

(ii) It is required to construct a confidence interval for $\dfrac{\delta_1^2}{\delta_2^2}$ the ratio of the population variances.

State the definition of an $F_{k,l}$ variable in terms of independent $N(0, 1)$ and χ_k^2 variables and use it to develop a suitable pivotal quantity. Hence obtain a 90% confidence interval for $\dfrac{\delta_1^2}{\delta_2^2}$. **[5]**

(iii) A regional newspaper included a consumer rights article comparing the cost of shopping in “corner shops” and “supermarkets”. The researchers investigated the price of a standard “selection” of household goods in a sample of 10 corner shops selected at random from the region, and in a sample of 10 supermarkets selected at random from the region. The data yielded the following values:

	Sample mean	Sample s.d
Corner Shops	22.5	1.22
Supermarkets	19.72	0.96

(a) Use your result in part (i)(b) to calculate a 95% confidence interval for (μ_1 - μ_2) the difference between the population means (1 = corner shops, 2 = supermarkets).

(b) Use your result in part (ii) to calculate a 90% confidence interval for $\dfrac{\delta_1^2}{\delta_2^2}$, the ratio of the population variances.

Hence comment briefly on the assumption of equal variances required for the confidence interval in part (iii)(a). **[5]**
[Total 17]

Q6 For bivariate data (x_i, y_i) the normal simple linear regression model is

$$Y_i = \alpha + \beta x_i + \varepsilon_i \ : i=1,2,\ldots,n \qquad \text{model I}$$

where the Y_i’s are observations on a response variable Y,

the x_i's are fixed values of an explanatory variable, and the ε_i's are independent $N(0,\delta^2)$ error variables.

For algebraic convenience it is useful to express the model as

$Y_i = \alpha + \beta(x_i - \bar{x}) + \varepsilon_i$: i=1,2,…,n Model II

to exploit the fact that $\sum(x_i - \bar{x}) = 0$

[Note that the α 's in models I and II differ.]

(i) Derive the least squares estimates $\hat{\alpha}$ of α and $\hat{\beta}$ of β using model II. **[4]**

(ii) Show that $\hat{\beta}$ is an unbiased estimator of β and that the variance of $\hat{\beta}$ is given by $Var(\hat{\beta}) = \delta^2 / S_{xx}$, where $S_{xx} = \sum(x_i - \bar{x})^2$ [4]

(iii) The following table gives data for 10 women from the maternity unit in a city hospital. The weight (*y* in kg.) of the baby at birth and the number(x) of cigarettes smoked per day is recorded for each woman.

x: 0	0	0	0	5	10	20	20	30	40
y: 3.9	3.3	4.1	3.6	3.5	3.1	3.0	2.7	2.4	2.6

$$\sum x = 125, \sum x^2 = 3425, \sum y = 106.54, \sum xy = 338.5$$

(a) Draw a scatter plot of this data and comment briefly.

(b) Calculate the estimated slope parameter and its standard error.

(c) Estimate the expected difference in baby weight for women who do not smoke and women who smoke 30 cigarettes per day. **[8]**

[Total 16]

Q7 A random sample of *n* claim sizes arising under a group of policies is examined. However, the data are censored at a known amount *c*. There are *m* claim sizes lower than *c* (the values of which are known) and *n* – *m* higher than *c* (the values of which are not known).

The known data are therefore *n*, *m* and the observed claim sizes $x_1, x_2, \ldots, x_n$

Suppose the claim sizes are exponentially distributed random variables with cumulative distribution function

$F(x) = 1 - e^{-\lambda x}$, $x > 0$ (= 0 otherwise)

(i) (a) Show, giving a clear explanation, that the likelihood function of these data is

$$L(\lambda; \underline{x}) = \lambda^m \exp[-\lambda\{\sum_{i=1}^{m} x_i + (n-m)c\}]$$

b) Show that the maximum likelihood estimate of λ is given by

$$\hat{\lambda} = \frac{m}{\sum_{i=1}^{m} x_i + (n-m)c}$$

.

(c) State the formula for the maximum likelihood estimate of the mean claim size μ = 1/λ and comment briefly on the result. **[8]**

(ii) In a particular case with c = \$10000, the observed data are n = 200, m = 173, and $\Sigma\, x_i$ = \$638,327.

(a) Calculate the maximum likelihood estimates of λ and of the mean claim size.

(b) Find the estimated asymptotic standard error of $\hat{\lambda}$, and hence calculate an approximate 95% confidence interval for the mean claim size. **[8]**

[Total 16]

Q8 Six insurance companies were being compared with regard to premiums being charged for house contents insurance for houses in a particular postcode region. Independent random samples of five policies from each company are examined and the premiums (in \$) were recorded.

Company	*A*	*B*	*C*	*D*	*E*	*F*
	151	152	175	149	123	145
	168	141	155	148	132	131
	128	129	162	137	142	155
	167	120	186	138	161	172
	134	115	148	169	152	141
Totals	748	657	826	741	710	744

$$\sum\sum y_{ij} = 4,426\ ; \ \sum\sum y_{ij}^2 = 661,796$$

(i) Compute an ANOVA table for these data, and show that there are no significant differences, at the 5% level, between mean premiums being charged by each company. **[5]**

(ii) State the assumptions required for the above analysis of variance, and, by drawing a suitable diagram of these data, comment briefly on the validity of these assumptions. **[4]**

(iii) Calculate a 95% confidence interval for the underlying common standard deviation δ using $\frac{SS_R}{\delta^2}$, as a pivotal quantity with a χ^2 distribution. **[4]**

(iv) A colleague points out that company C has the largest mean premium of \$165.20 and that Company B has the smallest mean premium of \$131.40 and suggests performing a *t*-test to compare these two companies.

(a) Perform this *t*-test, using the estimate of variance from the ANOVA table, and in particular show that there is a significant difference at the 1% level.

(b) Your colleague states that there is therefore a significant difference between the six companies.

Discuss the apparent contradiction with your conclusion in part (i) and explain the flaw in your colleague's argument. **[5]**

[Total 18]

***** END OF EXAMINATION*****