NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

FACULTY OF COMMERCE

DEPARTMENT OF INSURANCE AND ACTUARIAL SCIENCE

B COMM ACTUARIAL SCIENCE

ACTUARIAL STATISTICS II [CIN 2211]

JULY 2006 SUPPLEMENTARY 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. Suppose that the joint probability distribution of two random variables X and Y is given by the following table:

		Y		
		2	*4*	*6*
	1	0.2	0.0	0.2
X	2	0.0	0.2	0.0
	3	0.2	0.0	0.2

(i) Show that X and Y are uncorrelated, but are not independent. **[3]**

(ii) Leaving the probabilities in the first and third rows of the table the same, change the entries in the second row so that X and Y are independent. **[2]**

[Total 5]

Q2. The number of claims N which arise from a portfolio of business is modelled as a Poisson variable with mean μ. The claim amounts $X_i : i = 1, 2, ..., N$ are modelled as independent gamma variables each with parameters α and β and are independent of N. Let S be the total claim amount arising from this portfolio.

(i) Obtain expressions for the mean and standard deviation of S in terms of μ, α and β, using general results for the mean and variance of S. **[2]**

(ii) Consider the case where $\mu = 100$ and the individual claim amounts have mean 100 and standard deviation 50.

(a) Calculate the mean and standard deviation of the total claim amount S.

(b) Calculate an approximate value for the probability that the total claim amount S exceeds \$12,500, giving a brief justification of your approach. **[4]**

[Total 6]

Q3. The continuous random variables X and Y have a bivariate probability density function

$f(x, y) = 2$ for $0 < x + y < 1, x > 0, y > 0$.

The conditional distribution of X given $Y = y$ is a uniform distribution with probability density function

$$f(x|y) = \frac{1}{1-y} \quad 0 < x < 1-y$$

and the marginal distribution of Y is a beta distribution with probability density function

$f(y) = 2(1-y) \quad 0 < y < 1$.

(i) Show that the conditional expectation of X given $Y = y$ is

$$E(X|Y = y) = \frac{1-y}{2}$$

and obtain the conditional variance of X given $Y = y$. **[3]**

(ii) Verify in this case that $\text{var}(X) = \text{var}(E(X|Y)) + E(\text{var}(X|Y))$. **[3]**

[Total 6]

Q4. A simple model for the movement of a stock price is such that, independently in each time period, the stock either:

goes up with probability $(\frac{1}{4} - \theta)$;

stays the same with probability $(\frac{5}{8} + 2\theta)$;

goes down with probability $(\frac{1}{8} - \theta)$.

(i) Determine the range of admissible values of the parameter θ. **[2]**

(ii) (a) Calculate the probability that the stock goes down in one time period, in the case $\theta = 0.1$.

(b) Calculate the probability that the stock stays the same for two consecutive time periods, in the case $\theta = 0$.

(c) Calculate the probability that, in four time periods, the stock goes up twice and stays the same twice, in the case $\theta = -0.2$. **[4]**

(iii) Data are collected for 80 consecutive time periods and yield the following observed frequencies:

change in stock	up	same	down
no. of time periods	24	35	21

(a) (1) Write down an expression for $L(\theta)$, the likelihood of these data, and show that

$$\frac{\partial}{\partial\theta}\log L(\theta) = 0$$ reduces to the quadratic equation

$$5120\theta^2 - 468\theta - 95 = 0$$

(2) Explain why one of the roots of this quadratic yields the maximum likelihood estimate of θ and hence determine. this estimate. **[5]**

(b) (1) Calculate the expected frequencies using the model with the maximum likelihood estimate of θ.

(2) Hence perform a χ^2 goodness of fit test of the model and state your conclusion clearly. **[6]**

(c) Comment briefly on what additional information would be needed for this data in order to investigate the validity of the assumption of independence used in this model, and comment briefly on how the validity might be checked. **[2]**

[Total 19]

Q5. Consider a linear regression model in which responses Y_i are uncorrelated and have expectations βx_i and common variance δ^2 (i=1,…,n), i.e Y_i is modelled as a linear regression through the origin:

$$E(Y_i / x_i) = \beta x_i \text{ and } V(Y_i / x_i) = \delta^2 \text{ (i=1,…,n)}$$

(i) (a) Show that the least squares estimator of β is $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}$

(b) Derive the expectation and variance of $\hat{\beta}_1$ under the model. **[5]**

(ii) An alternative to the least squares estimator in this case is:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i} = \frac{\overline{Y}}{\overline{x}}$$

(a) Derive the expectation and variance of $\hat{\beta}_2$ under the model.

(b) Show that the variance of $\hat{\beta}_2$ the estimator is at least as large as that of the least squares estimator $\hat{\beta}_1$. **[5]**

(iii) Now consider an estimator $\hat{\beta}_3$ of β which is a linear function of the responses, i.e. an estimator which has the form $\hat{\beta}_3 = \sum_{i=1}^{n} a_i Y_i$, where $a_1,...a_n$ are constants.

(a) Show that $\hat{\beta}_3$ is unbiased for β if $\sum_{i=1}^{n} a_i x_i = 1$ and that the variance of $\hat{\beta}_3$ is $\sum_{i=1}^{n} a_i \delta^2$.

(b) Show that the estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ above may be expressed in the form $\hat{\beta}_3 = \sum_{i=1}^{n} a_i Y_i$ and hence verify that $\hat{\beta}_1$ and $\hat{\beta}_2$ satisfy the condition for unbiasedness in (iii)(a).

(c) It can be shown that, subject to the condition $\sum_{i=1}^{n} a_i x_i = 1$, the variance of $\hat{\beta}_3$ is minimised by setting $a_i = \dfrac{x_i}{\sum_{i=1}^{n} x_i^2}$

Comment on this result. **[8]**

[Total 18]

Q6. A psychologist conducted an investigation into the effect of alcohol on reaction times using 10 male and 10 female subjects. Each subject was given two tests on different days, during which his/her reaction times were recorded.

Before each of the tests, the subject drank a glass of liquid. Some glasses contained a fixed quantity of alcohol and others contained a liquid which had a similar colour and taste but no alcohol. Each subject drank one glass of each type. The order of presentation was randomized, independently for each subject.

The data below give the reaction times, in units of 0.01 seconds. Also given is the difference between the reaction time with alcohol and the reaction time without alcohol for each subject (reaction time with alcohol minus reaction time without).

Males

Subject No.	1	2	3	4	5	6	7	8	9	10
With alcohol	45	51	35	43	51	54	51	49	44	52
Without alcohol	40	54	21	31	44	47	39	33	32	56
Difference	5	-3	14	12	7	7	12	16	12	-4

Females

Subject No.	1	2	3	4	5	6	7	8	9	10
With alcohol	47	54	58	48	60	46	55	74	56	49
Without alcohol	39	40	42	30	51	41	55	68	47	40
Difference	8	14	16	18	9	5	0	6	9	9

(i) (a) Construct a 95% confidence interval for the mean difference between the reaction times with and without alcohol for the males, using the 10 difference values.

(b) Construct a similar 95% confidence interval based on the female difference values.

(c) Comment briefly on the two confidence intervals. **[8]**

(ii) (a) Perform a two-sample t-test to investigate whether the alcohol effect differs between males and females.

(b) Show that the variances in the male and female samples are not significantly different at the 5% level, and comment briefly with reference to the test conducted in (ii)(a). **[8]**

[Total 16]

Q7. The following data give the invoiced amounts for work carried out on 12 jobs performed by a plumber in private customers' houses. The durations of the jobs are also given.

duration x (hrs)	1	1	2	3	4	4	5	6	7	8	9	10
amount y ($)	45	65	80	95	100	125	145	180	180	210	330	240

$$\sum x_i = 60, \sum x_i^2 = 402, \sum y_i = 1972, \sum y_i^2 = 343725, \sum x_i y_i = 11570$$

The plumber claims to calculate his total charge for each job on the basis of a single call-out charge plus an hourly rate for the time spent working on the job.

(i) (a) Draw a scatter plot of the data on graph paper and comment briefly on your plot.

(b) The equation of the fitted regression line of y on x is $y = 22.4 + 25.4x$ and the coefficient of determination is R^2 = 87.8% *(you are not asked to verify these results).*

Draw the fitted line on your scatter plot. **[3]**

(ii) (a) Calculate the fitted regression line of invoiced amount on duration of job using only the 11 pairs of values remaining after *excluding the invoice* for which $x = 9$ and $y = 330$.

(b) Calculate the coefficient of determination of the fit in (ii)(a) above.

(c) Add the second fitted line to your scatterplot, distinguishing it clearly from the first line you added (in part (i)(b) above).

(d) Comment on the effect of omitting the invoice for which $x = 9$ and $y = 330$.

(e) Carry out a test to establish whether or not the slope in the model fitted in (ii)(a) above is consistent with a rate of $25 per hour for work carried out. **[12]**

[Total 18]

Q8. A researcher studying the claims experience of a company (in a particular class of business) records the payments on 100 recent claims. The payments (in units of $1000, and sorted) are given below.

Payments
0.30 0.89 0.96 1.16 1.67 1.77 1.93 1.98 2.07 2.09 2.30 2.48
2.58 2.78 3.00 3.19 3.21 3.21 3.25 3.31 3.34 3.37 3.66 3.95
4.16 4.18 4.60 4.72 4.73 4.76 5.01 5.17 5.21 5.63 5.72 6.00
6.13 6.17 6.24 6.37 6.47 6.48 6.87 7.05 7.16 7.21 7.51 7.72
7.74 8.00 8.00 8.03 8.04 8.54 9.11 9.18 9.49 9.59 10.00 10.36
10.85 11.08 11.22 11.27 11.38 11.45 11.69 11.78 12.27 12.30 12.50 13.04
13.28 13.43 13.48 13.85 14.27 14.31 14.49 14.55 14.62 14.68 14.70 14.83
15.67 15.70 15.77 16.28 16.44 17.17 17.89 18.03 18.12 20.72 22.00 24.33
25.41 28.30 31.00 32.80

For these 100 observations: $\sum x = 952.75$, $\sum x^2 = 13584.5217$.

The researcher wants to examine whether or not the exponential distribution provides a good description of the distribution of the payments.

(i) (a) Calculate the sample mean and standard deviation for the 100 payments, and comment briefly on whether or not you think the exponential distribution will provide an acceptable fit to the data.

(b) Specify the fitted exponential distribution, giving a brief justification of your approach.

Note: you are not required to give any mathematical derivation in your justification. **[5]**

(ii) The researcher decides to conduct a formal chi-squared goodness-of-fit test of the exponential distribution to the data, using five equi-probable intervals (i.e. intervals each with associated probability 0.2).

(a) Show that the value x which is exceeded with probability p by an exponential variable with mean μ satisfies $x = -\mu \log p$.

(b) Calculate the values which divide the positive real numbers into five equi-probable intervals for your fitted exponential distribution.

[4]

(iii) Conduct the formal goodness-of-fit test outlined in part (ii) above and comment on the result. **[5]**

[Total 14]