NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

B. COMM ACTUARIAL SCIENCE

STATISTICS FOR INSURANCE 1 – CIN 2111

JULY SUPPLEMENTARY EXAMINATION PAPER

Instructions

- Answer all questions
- In addition to this paper you should have available Actuarial Tables , Graph paper and electronic calculator.

---

1 The number of claims per policy during 1995 was noted for a sample of 50 household insurance policies. It turned out that no claims arose under 40 of the policies, one claim arose under each of 5 of the policies, two claims arose under each of 3 of the policies, three claims arose under 1 policy, and four claims arose under 1 policy.

Calculate the sample standard deviation of the number of claims per policy in 1995.

[2marks]

2 Suppose that the joint distribution of two random variables X and Y is as follows:

	$y = 0$	$y = 1$	$y = 2$
$x = 0$	0	0.25	0
$x = 1$	0.25	0	0.25
$x = 2$	0	0.25	0

(a) Determine whether X and Y are correlated.

(b) Determine whether X and Y are independent. **[2 marks]**

3 A random sample of size 25 is taken from a normal distribution with variance $\sigma^2 = 30$. The sample variance is an observation of a random variable S^2 which has its own sampling distribution with its own mean and variance.

Calculate the value of $\text{Var}(S^2)$. **[3 marks]**

4 A discrete random variable can take three possible values, 1, 2 and 4, with respective probabilities $\frac{1}{2} + \theta$, $\frac{1}{2} - 2\theta$ 0and θ.

A random sample of 50 observations of this variable results in the respective observed frequencies 30, 16 and 4.

Calculate the method of moments estimate of the parameter θ. **[3 marks]**

5 A random sample of 1000 people from the voters' roll in the UK was taken and 128 members of the sample knew what an actuary does.

Calculate a 99% confidence interval for the percentage of the whole electorate who know what an actuary does. **[2 marks]**

6 A random sample of size 25 from a normal distribution has sample mean 15 and sample standard deviation 2.

Calculate a 95% confidence interval for the population mean. **[3 marks]**

7 Let X have a Poission distribution with mean 10, and let $Y = 2X + 3$. Derive the cumulant generating function of Y. **[3 marks]**

8 Two different and independent response variables, X and Y, are observed, both of which have the same unknown mean θ. The variance of X is known to be 4 times that of Y. The estimator

$$\theta^* = a\overline{X} + (1 - a)\,\overline{Y}, 0 < a < 1$$

is based on random samples of X and Y of sizes m and n respectively, and with means $\overline{X}$ and $\overline{Y}$ respectively.

Show that θ^* is unbiased for θ and that the mean square error of θ^* is minimised when $a = m/(m + 4n)$.0 **[4 marks]**

9 The following table gives the ages of 100 men (in years) in the form of a grouped frequency distribution, where the ages are in groups of width five years, with the exception of the final group.

Age last birthday:	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–64
Number of men:	1	2	10	16	22	20	15	14

(i) Draw a histogram of the data.

(ii) When originally collected the data had been in 5-year age groups throughout the whole range (nine groups), and, using central values of the age groups (e.g. 22.5 for the first group), the mean age was calculated to be 45.1 years.

Reconstruct the original table with 5-year age groups. **[5 marks]**

10 Insurance policies providing "holiday insurance" cover are such that 60% of policies give rise to no claims, 30% give rise to single claims, and 10% give rise to two claims each.

Most claims are genuine, but some are fraudulent, and you may assume for simplicity that claims are independent of one another in this respect. Further you may assume that the probability that any claim is a genuine claim is 0.8.

The sizes of claims which arise as the only claim on a policy are distributed normally with mean £2000 and standard deviation £500. The sizes of claims which arise as one of two claims on a policy are distributed normally with mean £1500 and standard deviation £300.

Calculate the expected number of genuine claims for more than £1500 which will arise under a group of 1000 such policies.
[6 marks]

11 The amounts of claims on a particular type of policy are distributed about a mean of £10,000 and with standard deviation £1,000. In a given week an insurance company receives 50 such claims.

Assuming that all such claims are independent of one another, calculate the approximate probability that the total amount of these claims exceeds £515,000.
[4 marks]

12 Suppose that a total of x claims are recorded in the first t_0 months of operation of a motor insurance portfolio comprising n new policies. You may assume that each policy gives rise to a maximum of one claim and that claims arise independently of one another.

Suppose also that, for each policy, the time (months) to a claim occurring is distributed as a variable T with:

$$P(T \leq t) = 1 - \exp(-\beta t)$$

0

where β is an unknown parameter with $\beta > 0$.

(i) Write down the likelihood of observing x claims and hence determine the maximum likelihood estimate of β in terms of x, n, and t_0. **[4marks]**

(ii) Hence estimate the underlying mean time to a claim occurring on a policy.

[2 marks]

[Total 6 marks]

13 An experiment was carried out to compare the ease of operation and efficiency of six electronic calculators. A clerk, familiar with the calculators, was given 18 similar sets of
calculation to perform, with a random assignment of three tasks to each calculator. The times (sec.) taken to complete the tasks were as follows:

Calculators

I	*II*	*III*	*IV*	*V*	*VI*
150	108	98	130	90	90
140	112	118	160	120	112
130	110	108	100	90	104

Perform an analysis of variance and state your conclusion. **[6marks]**

14 (i) By using the normal approximation to the chi-square distribution with v degrees of freedom, χ^2_v, show that the upper and lower 5% points of the distribution of a χ^2_{120} random variable are approximately 145.5 and 94.52 respectively, that is, show:

$$P(\chi^2_{120} < 94.52) \simeq 0.05 \text{ and } P(\chi^2_{120} > 145.5) \simeq 0.05.$$ **[3 marks]**

(ii) Let X be a gamma (α, λ) random variable (with mean α/λ).

(a) Show that the moment generating function of X, $M(t)$, is given by:

$M(t) =$

(b) In the case when 2α is an integer, show that $2\lambda X$ is a $\chi^2_{2\alpha}$ random variable. **[5 marks]**

(iii) Let $(X_1, X_2, \ldots, X_n)$ be a random sample of a gamma (3, λ) random variable, with sample mean $\overline{X}$.

(a) Suggest a sensible estimator of λ.

(b) State the distribution of $2\lambda n\overline{X}$ **Error! Switch argument not specified.**.

(c) Construct an upper 95% confidence interval for λ, of the form $(0, U)$, based on $\overline{X}$.

(d) Construct a lower 95% confidence interval for λ, of the form (L, ∞), based on $\overline{X}$.

(e) Evaluate the intervals in (iii) (c) and (d) above in the case for which the total of a random sample of 20 observations had a value of $\Sigma x_i = 98$..

[9 marks]

[Total 17 marks]

15 (i) A manufacturing company produces screws of a particular size which are put into boxes of 150. On a particular day a random sample of such boxes is taken from each of the morning and afternoon production runs. The numbers of defective screws found in each sampled box are given in the following table:

Number of defectives per box

Morning:	28	17	18	16	20	12	11	10	18	17	20	25
Afternoon:	19	15	22	21	9	14	17	13	22	9		

(a) Test for a difference between the mean numbers of defectives produced in the morning and afternoon (you may assume that the underlying population variances are equal).

(b) Plot the data in an appropriate and simple way and comment briefly on the validity of the test in part (a).

[8 marks]

(ii) On another day screws are put into boxes of 100. The table below gives the numbers of defectives in twenty boxes sampled from this day's production run.

5	15	13	12	8	7	9	14	11	10
6	8	14	9	13	12	11	5	13	12

(a) Carry out a test to establish whether there is a difference between the proportions of defectives produced on the two days.

(b) Carry out a test to establish whether the proportion of defectives in boxes of 100 screws is more than 9%.

[9 marks]

[Total 17 marks]

16 The data given in the first three columns in the table below relate to the mortality experience of former male company employees who retired *through ill-health* in the period 1990–94:

x	d_x	e_x	y	$\hat{m}_x$	r_x
–7	29	1011	–3.537	–3.764	0.227

–6	17	853	–3.905	–3.684	–0.221
–5	32	991	–3.417	–3.605	0.188
–4	34	979	–3.343	–3.525	0.183
–3	31	1059	–3.516	–3.446	–0.071
–2	31	1142	–3.593	–3.366	–0.227
–1	25	1213	–3.872	–3.286	–0.585
0	46	1102	–3.155	–3.207	0.052
1	53	1088	–2.997	–3.127	0.130
2	68	1180	–2.824	–3.048	0.223
3	68	1256	–2.888	–2.968	0.080
4	79	1248	–2.727	–2.888	0.161
5	75	1365	–2.873	–2.809	–0.064
6	71	1205	–2.801	–2.729	–0.072
7	85	1250	–2.653	–2.650	–0.004

$$\Sigma y = -48.101 \quad , \quad \Sigma xy = 22.293$$

Column 1 — age at retirement (coded x = actual – 57 years)

Column 2 — number of deaths, d_x

Column 3 — exposure to risk of death, e_x

The values of y, tabulated in the fourth column, have been calculated using the formula

$$y = \log\{-\log(1 - \mathring{q}_x)\} \quad , \ \mathring{q}_x = d_x / e_x$$

0

where $\mathring{q}_x$ 0 denotes the crude (or estimated) probability that a life aged x dies before age $x + 1$. Therefore y is a measure of mortality.

(i) The regression model

$$y = \alpha + \beta x + \text{error}$$

has been fitted to these data using the least squares criterion and the corresponding fitted values $\hat{m}_x$ and residuals r_x are recorded in the last two columns above.

(a) Derive the least squares estimation equations for the parameters α and β and hence estimate the values of these parameters which give rise to the reported fitted values and residuals.

(b) Comment briefly on the adequacy of the fit. **[7 marks]**

(ii) (a) Express $\mathring{q}_x$ as a function of y and calculate the values of $\mathring{q}_x$ corresponding to the integer values 1, 0, –1, –2, –3, and –4 of y.

(b) Sketch the graph of y against $\mathring{q}_x$ and comment on the implications of its shape in the context of this question. [6]

(iii) A similar analysis has been conducted over the same age range on the reported mortality experience of former male company employees who retired *while healthy* in the period 1990–94, giving rise to the parameter estimates:

$$\hat{\beta} = 0.1330\ ,\ \hat{\alpha} = -4.357$$

Draw the fitted straight lines over the relevant age range, for the two cases (*ill-health* and *healthy*) on the same diagram, and comment briefly on the resulting graph. **[4 marks]**

[Total 17 marks]

END OF EXAMINATION