INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

7 October 2011 (pm)

Subject CT3 — Probability and Mathematical Statistics
Core Technical

Time allowed: Three hours

INSTRUCTIONS TO THE CANDIDATE

1. *Enter all the candidate and examination details as requested on the front of your answer booklet.*

2. *You must not start writing your answers in the booklet until instructed to do so by the supervisor.*

3. *Mark allocations are shown in brackets.*

4. *Attempt all 10 questions, beginning your answer to each question on a separate sheet.*

5. *Candidates should show calculations where this is appropriate.*

Graph paper is NOT required for this paper.

AT THE END OF THE EXAMINATION

Hand in BOTH your answer booklet, with any additional sheets firmly attached, and this question paper.

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator from the approved list.

1 The first 20 claims that were paid out under a group of policies were for the following amounts (in units of £1,000):

3.2	2.1	6.3	4.0	3.8	4.4	6.5	7.8	2.8	5.2
7.0	8.1	4.4	5.8	1.7	2.8	5.0	3.2	3.7	4.4

For these data $\sum x = 92.2$.

(i) Calculate the mean of these 20 claim amounts. [1]

The next 80 claims paid out had a mean amount of £5,025.

(ii) Calculate the mean amount for the first 100 claims. [2]
[Total 3]

2 The claims which arose in a sample of policies of a certain class gave the following frequency distribution for the number of claims per policy in the last year:

Number of claims x	0	1	2	3	4 or more
Number of policies f	15	20	10	5	0

Calculate the third order moment about the origin for these data. [3]

3 A random sample of 60 adult men who live in Leeds includes 21 who have visited Majorca. An independent random sample of 70 adult women who live in Leeds includes 28 who have visited Majorca.

Calculate a 98% confidence interval for the proportion of adults who live in Leeds who have visited Majorca. [4]

4 The random variables X and Y are related as follows:

X conditional on $Y = y$ has a $N(2y, y^2)$ distribution.
Y has a $N(200, 100)$ distribution.

Derive the unconditional variance of X, $V[X]$. [3]

5 Consider the random variable X taking the value $X = 1$ if a randomly selected person is a smoker, or $X = 0$ otherwise. The random variable Y describes the amount of physical exercise per week for this randomly selected person. It can take the values 0 (less than one hour of exercise per week), 1 (one to two hours) and 2 (more than two hours of exercise per week). The random variable $R = (3 - Y)^2(X + 1)$ is used as a risk index for a particular heart disease.

The joint distribution of X and Y is given by the joint probability function in the following table.

	Y		
X	0	1	2
0	0.2	0.3	0.25
1	0.1	0.1	0.05

(i) Calculate the probability that a randomly selected person does more than two hours of exercise per week. [1]

(ii) Decide whether X and Y are independent or not and justify your answer. [2]

(iii) Derive the probability function of R. [3]

(iv) Calculate the expectation of R. [2]

[Total 8]

6 The number of claims made by each policyholder in a certain class of business is modelled as having a Poisson distribution with mean λ.

(i) Derive an expression for the probability, p, that a policyholder in this class has made at least one claim. [2]

The claims records of 20 randomly chosen policyholders were examined and the number of policyholders that made at least one claim in a year, X, was recorded.

(ii) (a) State the distribution of the random variable X and its parameters.

(b) Derive an expression for the maximum likelihood estimator of the probability p given in (i) using your answer in (ii)(a).

[4]

(iii) Show that, in the case $X = 5$, the maximum likelihood estimate (MLE) of p is $\hat{p} = 0.25$ and hence calculate the MLE of λ. [3]

It is now found that of the five policyholders who had made at least one claim there were four who had made exactly one claim and one who had made two claims.

(iv) Calculate the MLE of λ given this additional information. [4]

[Total 13]

7 The total amounts y_{ij} (in £ millions) paid out under a certain type of policy issued by four different companies A, B, C, D in each of six consecutive years were as follows:

Company							*Total*
A	2.870	3.125	3.000	2.865	2.890	3.060	17.810
B	3.105	3.200	3.300	2.975	3.210	3.150	18.940
C	2.800	2.985	3.060	2.900	2.920	3.050	17.715
D	2.830	2.600	2.765	2.690	2.600	2.700	16.185

For these data, $\Sigma_i \Sigma_j y_{ij} = 70.650$ and $\Sigma_i \Sigma_j y_{ij}^2 = 208.828$.

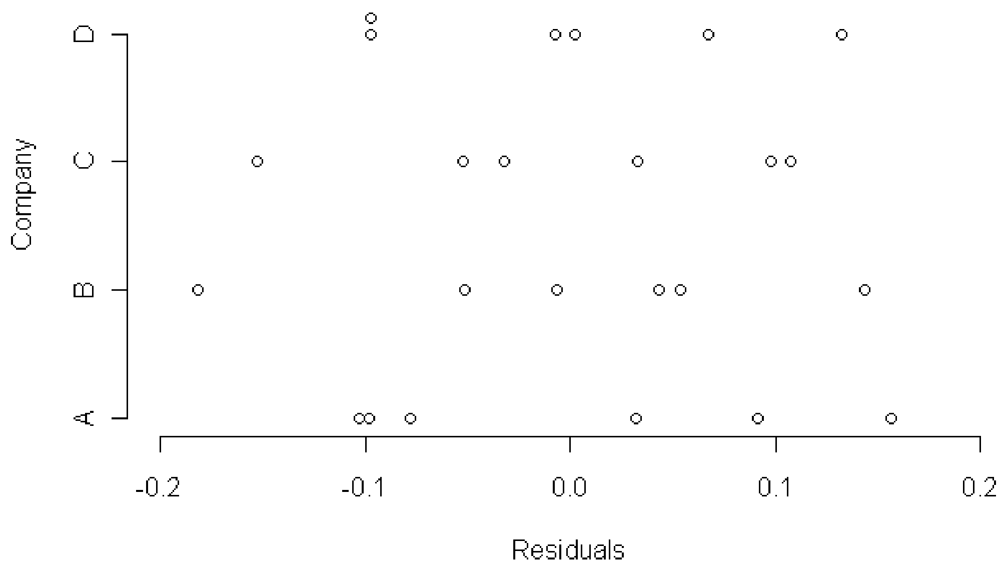
Consider the ANOVA model $Y_{ij} = \mu + \tau_i + e_{ij}$, $i = 1, \ldots, 4$, $j = 1, \ldots, 6$, where Y_{ij} is the jth amount paid out by company i, and $e_{ij} \sim N(0, \sigma^2)$ are independent errors.

The ANOVA table for these data is given below.

Source	*DF*	*SS*	*MS*
Company (between treatments)	3	0.640	0.213
Residual	20	0.212	0.0106
Total	23	0.852	

(i) Test the hypothesis that there are no differences in the means of the amounts paid out under such policies by the four companies (the company means), stating your conclusions clearly. [2]

(ii) Comment briefly on the validity of the test performed in (i), using the plot of the residuals given below. [2]



(iii) (a) Calculate the least significant difference between pairs of company means using a 5% significance level.

(b) List the company means in order, illustrate the non-significant pairs using suitable underlining, and comment briefly.

[6]
[Total 10]

8 Consider a random sample $X_1, \ldots, X_n$ from a Poisson distribution with expectation $E[X_i] = \lambda$. An estimator $\hat{\lambda}$ for the parameter λ is given by the observed mean of the sample, that is:

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i \ .$$

(i) Derive formulae for the expected value and the variance of $\hat{\lambda}$ in terms of λ and n. [3]

Assume in parts (ii) to (v) that the true parameter value is $\lambda = 0.25$.

(ii) Calculate the exact probability that $0.2 \le \hat{\lambda} \le 0.3$ if the sample size is $n = 10$. [3]

(iii) Calculate the approximate probability that $0.2 \le \hat{\lambda} \le 0.3$ if the sample size is $n = 10$ using the following:

(a) the normal approximation to $\sum_{i=1}^{n} X_i$ with continuity correction

(b) the normal approximation to $\sum_{i=1}^{n} X_i$ without continuity correction.

[6]

(iv) Comment on the differences in your answers in parts (ii) and (iii). [2]

(v) Calculate the minimal required sample size n for which the probability that $0.2 \le \hat{\lambda} \le 0.3$ is at least 0.95, using the normal approximation without continuity correction [4]

Suppose a random sample of size $n = 400$ gives the estimate $\hat{\lambda} = 0.27$.

(vi) Calculate a 95% confidence interval for λ. [3]

[Total 21]

9 In a recent study of attitudes to a proposed new piece of consumer legislation ("proposal X") independent random samples of 200 men and 200 women were asked to state simply whether they were "for" (in favour of) , or "against", the proposal. The resulting frequencies, as reported by the consultants who carried out the survey, are given in the following table:

	Men	*Women*
For	138	130
Against	62	70

(i) Carry out a formal chi-squared test to investigate whether or not an association exists between gender and attitude to proposal X.

Note: in this and any later such tests in this question you should state the *P-value* of the data and your conclusion clearly. [6]

At a subsequent meeting to discuss these and other results, the consultants revealed that they had in fact stratified the survey, sampling 100 men and 100 women in England and 100 men and 100 women in Wales. The resulting frequencies were as follows:

	England		*Wales*	
	Men	*Women*	*Men*	*Women*
For	82	66	56	64
Against	18	34	44	36

A chi-squared test to investigate whether or not an association exists between gender and attitude to proposal X in England gives $\chi^2 = 6.653$, while an equivalent test for Wales gives $\chi^2 = 1.333$.

(ii) (a) Find the *P-value* for each of the chi-squared tests mentioned above and state your conclusions regarding possible association between gender and attitude to proposal X in England and in Wales.

(b) Discuss the results of the survey for England and Wales separately and together, quoting relevant percentages to support your comments. [9]

(iii) A different survey of 200 people conducted in each of England, Wales, and Scotland gave the following percentages in favour of another proposal:

	England	*Wales*	*Scotland*
% in favour of proposal	62%	53%	58%

A chi-squared test of association between country and attitude to the proposal gives $\chi^2 = 3.332$ on 2 degrees of freedom, with *P-value* 0.189.

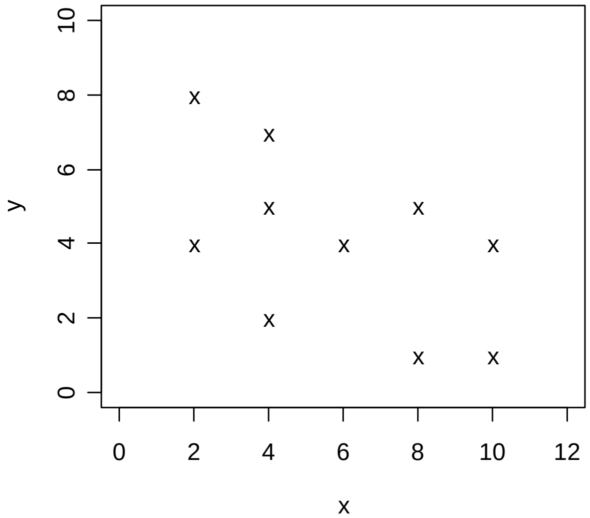
Suppose a second survey of the same size is conducted in the three countries and results in the same percentages in favour of the proposal as in the first survey. The results of the two surveys are now combined, giving a survey based on the attitudes of 1,200 people.

(a) State (or find) the results of a second chi-squared test for an association between country and attitude to the proposal, based on the overall survey of 1,200 people. [3]

(b) Comment briefly on the results. [1]

[Total 19]

10 Consider a situation in which integer-valued responses (y) are recorded at ten values of an integer-valued explanatory variable (x). The data are presented in the following scatter plot:



For these data: $\Sigma x = 58$, $\Sigma x^2 = 420$, $\Sigma y = 41$, $\Sigma y^2 = 217$, $\Sigma xy = 202$

(i) (a) Calculate the value of the coefficient of determination (R^2) for the data.

(b) Determine the equation of the fitted least-squares line of regression of y on x.

[7]

(ii) Calculate a 95% confidence interval for the slope of the underlying line of regression of y on x. [4]

(iii) (a) Calculate an estimate of the expected response in the case $x = 9$.

(b) Calculate the standard error of this estimate.

[3]

Suppose the observation ($x = 10$, $y = 8$) is added to the existing data. The coefficient of determination is now $R^2 = 0.07$.

(iv) Comment briefly on the effect of the new observation on the fit of the linear model. [2]

[Total 16]

END OF PAPER