

INSTITUTE AND FACULTY OF ACTUARIES

Curriculum 2019

SPECIMEN EXAMINATION

Subject CS1A – Actuarial Statistics

Time allowed: Three hours and fifteen minutes

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 11 questions, beginning your answer to each question on a new page.*
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 (CT3 September 2010)

In a survey, a sample of 10 policies is selected from the records of an insurance company. The following data give, in ascending order, the time (in days) from the start date of the policy until a claim has arisen from each of the policies in the sample.

297 301 312 317 355 379 404 419 432+ 463+

Some of the policies have not yet resulted in any claims at the time of the survey, so the times until they each give rise to a claim are said to be censored. These values are represented with a plus sign in the above data.

- (i) Calculate the median of this sample. [2]
- (ii) State what you can conclude about the mean time until claims arise from the policies in this sample. [2]

[Total 4]

2 (CT3 April 2014)

Let X be a random variable with probability density function:

$$f(x) = \begin{cases} \frac{1}{2}e^x & ; x \leq 0 \\ \frac{1}{2}e^{-x} & ; x > 0 \end{cases}$$

- (i) Show that the moment generating function of X is given by:

$$M_X(t) = (1 - t^2)^{-1},$$

for $|t| < 1$. [3]

- (ii) Hence find the mean and the variance of X using the moment generating function in part (i). [3]

[Total 6]

3 (CT3 September 2011)

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.

X	Y		
	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]

[Total 6]

4 (CT3 April 2013)

Consider a random sample, X_1, \dots, X_n , from a normal $N(\mu, \sigma^2)$ distribution, with sample mean \bar{X} and sample variance S^2 .

- (i) Define carefully what it means to say that X_1, \dots, X_n is a random sample from a normal distribution. [2]
- (ii) State what is known about the distributions of \bar{X} and S^2 in this case, including the dependencies between the two statistics. [3]
- (iii) Define the t -distribution and explain its relationship with \bar{X} and S^2 . [2]

[Total 7]

5 (CT6 April 2014)

The heights of adult males in a certain population are Normally distributed with unknown mean μ and standard deviation $\sigma = 15$.

Prior beliefs about μ are described by a Normal distribution with mean 187 and standard deviation 10.

- (i) Calculate the prior probability that μ is greater than 180. [2]

A sample of 80 men is taken and the mean height is found to be 182.

- (ii) Calculate the posterior probability that μ is greater than 180. [4]
- (iii) Comment on your results from parts (i) and (ii). [2]

[Total 8]

6 (CT6 September 2010)

An office worker receives a random number of e-mails each day. The number of emails per day follows a Poisson distribution with unknown mean μ . Prior beliefs about μ are specified by a gamma distribution with mean 50 and standard deviation 15. The worker receives a total of 630 e-mails over a period of ten days.

Calculate the Bayesian estimate of μ under all or nothing loss. [7]

7 (CT6 April 2010)

An insurance company is modelling claim numbers on its portfolio of motor insurance policies using a Poisson distribution, whose mean depends on the age and gender of the policyholder.

(i) Suggest a link function for fitting a generalised linear model for the mean of the Poisson distribution. [1]

(ii) Specify the corresponding linear predictor used for modelling the age and gender dependence as:

- (a) age + gender
- (b) age + gender + age \times gender

[4]

[Total 5]

8 (CT6 April 2013)

An insurance company believes that individual claim amounts from house insurance policies follow a gamma distribution with distribution function given by:

$$f(y) = \frac{\alpha^\alpha}{\mu^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-\frac{\alpha}{\mu}y} \text{ for } y > 0$$

where α and μ are positive parameters.

(i) Show that the gamma distribution can be written in exponential family form, giving the natural parameter and the canonical link function. [5]

The insurance company has data for claim amounts from previous claims. It believes that the claim amount is primarily influenced by two variables:

x_i the type of geographical area in which the house is situated. This can take one of 4 values.

y_i the category of the age of the house where the three categories are 0–29 years, 30–59 years and 60 years +.

It wishes to model claim amounts using this data and the generalised linear model from part (i) with canonical link function. The insurance company is investigating models which take into account these variables and has the following table of values:

Model	Choice of predictor	Scaled Deviance
A	1	900
B	Age	789
C	Age +location	544
D	Age * location	541

- (ii) Explain, by analysing the scaled deviances, which model the insurance company should use.

[6]

[Total 11]

9 (Based on CT3 September 2015)

An insurance company has calculated premiums assuming that the average claim size per claim for a certain class of insurance policies does not exceed £20,000 per annum. An actuary analyses 25 such claims that have been randomly selected. She finds that the average claim size in the sample is £21,000 and the sample standard deviation is £2,500. Assume that the size of a single claim is normally distributed with unknown expectation α and variance σ^2 .

- (i) Calculate a 95% confidence interval for α based on the sample of 25 claims. [3]
- (ii) Perform a test for the null hypothesis that the expected claim size is not greater than £20,000 at a 5% significance level. [3]
- (iii) Calculate the largest expected claim size, α_0 , for which the hypothesis $\alpha \leq \alpha_0$ can be rejected at a 5% significance level based on the sample of 25 claims. [2]

The insurer is also concerned about the number of claims made each year. It is found that the average number of claims per policy was 0.5 during the year 2011. When the analysis was repeated in 2012 it was found that the average number of claims per policy had increased to 0.6. These averages were calculated on the basis of random samples of 100 policies in each of the two years. Assume that the number of claims per policy per year has a Poisson distribution with unknown expectation λ and is independent from the number of claims in any other year or for any other policy.

- (iv) Perform a test at 5% significance level for the null hypothesis that $\lambda = 0.6$ during the year 2011. [3]

- (v) Perform a test to decide whether the average number of claims has increased from 2011 to 2012. [3]
[Total 14]

10 (Based on CT3 April 2010)

The size of claims (in units of £1,000) arising from a portfolio of house contents insurance policies can be modelled using a random variable X with probability density function (pdf) given by:

$$f_X(x) = \frac{ac^a}{x^{a+1}}, \quad x \geq c$$

where $a > 0$ and $c > 0$ are the parameters of the distribution.

- (i) Show that the expected value of X is $E[X] = \frac{ac}{a-1}$, for $a > 1$. [2]
- (ii) Verify that the cumulative distribution function of X is given by

$$F_X(x) = 1 - \left(\frac{c}{x}\right)^a, \quad x \geq c \quad (\text{and } = 0 \text{ for } x < c). \quad [2]$$

Suppose that for the distribution of claim sizes X it is known that $c = 2.5$, but a is unknown and needs to be estimated given a random sample x_1, x_2, \dots, x_n .

- (iii) Show that the maximum likelihood estimate (MLE) of a is given by:

$$\hat{a} = \frac{n}{\sum_{i=1}^n \log\left(\frac{x_i}{2.5}\right)}. \quad [3]$$

- (iv) Derive the asymptotic variance of the MLE \hat{a} , and hence determine its approximate asymptotic distribution. [4]

In the current year, claim sizes are assumed to follow the distribution of X with $a = 6$, $c = 2.5$. Inflation for the following year is expected to be 5%.

- (v) Calculate the probability that the size of a claim arising from this portfolio in the following year will exceed £4,000. [3]
[Total 14]

11 (CT3 September 2010)

An investigation concerning the improvement in the average performance of female track athletes relative to male track athletes was conducted using data from various international athletics meetings over a period of 16 years in the 1950s and 1960s. For each year and each selected track distance the observation y was recorded as the average of the ratios of the twenty best male times to the corresponding twenty best female times.

The data for the 100 metres event are given below together with some summaries.

<i>year t:</i>	1	2	3	4	5	6	7	8
<i>ratio y:</i>	0.882	0.879	0.876	0.888	0.890	0.882	0.885	0.886
<i>year t:</i>	9	10	11	12	13	14	15	16
<i>ratio y:</i>	0.885	0.887	0.882	0.893	0.878	0.889	0.888	0.890

$$\Sigma t = 136, \quad \Sigma t^2 = 1496, \quad \Sigma y = 14.160, \quad \Sigma y^2 = 12.531946, \quad \Sigma ty = 120.518$$

(i) Draw a scatterplot of these data and comment briefly on any relationship between ratio and year. [3]

(ii) Verify that the equation of the least squares fitted regression line of ratio on year is given by:

$$y = 0.88105 + 0.000465t. \quad [4]$$

(iii) (a) Calculate the standard error of the estimated slope coefficient in part (ii).

(b) Determine whether the null hypothesis of “no linear relationship” would be accepted or rejected at the 5% level.

(c) Calculate a 95% confidence interval for the underlying slope coefficient for the linear model.

[5]

Corresponding data for the 200 metres event resulted in an estimated slope coefficient of:

$$\hat{\beta} = 0.000487 \text{ with standard error } 0.000220.$$

(iv) (a) Determine whether the “no linear relationship” hypothesis would be accepted or rejected at the 5% level.

(b) Calculate a 95% confidence interval for the underlying slope coefficient for the linear model and comment on whether or not the underlying slope coefficients for the two events, 100m and 200m, can be regarded as being equal.

(c) Discuss why the results of the tests in parts (iii)(b) and (iv)(a) seem to contradict the conclusion in part (iv)(b).

[6]

[Total 18]

END OF PAPER