INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

26 April 2022 (am)

Subject CS1 – Actuarial Statistics
Core Principles

Paper B

Time allowed: One hour and fifty minutes

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

If you encounter any issues during the examination please contact the Assessment Team at T. 0044 (0) 1865 268 873.

1 An engineer is considering the maximum number of people who can occupy a lift. In particular, the engineer wants to assess the probability of exceeding a maximum weight when eight people are allowed to use the lift at the same time, compared to nine people.

Assume that:

- the total weight of eight people chosen at random follows a normal distribution with a mean of 560 kg and standard deviation of 57 kg
- the total weight of nine people chosen at random follows a normal distribution with a mean of 630 kg and standard deviation of 61 kg.

(i) Calculate the probability that the total weight of eight people exceeds 650 kg. [2]

(ii) Calculate the probability that the total weight of nine people exceeds 650 kg. [2]

(iii) Comment on your answers to parts (i) and (ii). [2]

(iv) Calculate the interval for the central region containing 80% of the distribution of the total weight of eight people. [3]

Assume now that the total weight of eight people chosen at random follows a gamma distribution with parameters $\alpha = 96.5220$ and $\lambda = 0.1724$.

(v) (a) Calculate the interval for the central region containing 80% of the distribution of the total weight of eight people in this case.

(b) Comment on your answers to parts (iv) and (v)(a).

[3]

[Total 12]

2 (i) (a) Simulate a sample of 100 values from a Beta distribution with parameters $a = 3$, $b = 1$. Use the command `set.seed(12345)` to initialise the random number generator, before you start the simulation.

(b) Plot a histogram of the sample simulated in part (i)(a).

(c) Comment on the shape of the histogram produced in part (i)(b). [6]

(ii) (a) Perform 1,000 repetitions of the simulation in part (i)(a). You should compute and store the value of the mean of the sample for each repetition. Use the command `set.seed(12345)` to initialise the random number generator, before you start the simulation. [5]

(b) Plot a histogram of the 1,000 sample means computed in part (ii)(a). [2]

(c) Comment on the shape of the histogram produced in part (ii)(b) by referring to a fundamental statistical property. [2]

For the remaining parts of the question consider the following small sample of data (given here in R code):

```
y = c(4.9, 3.3, 2.2, 2.3, 1.6, 2.4, 4.7, 1.4, 1.7, 5.1)
```

(iii) Determine a 90% confidence interval for the population mean of these data, stating any assumptions that you make. [3]

(iv) Determine an estimate of the standard error of the sample mean using these data. [2]

(v) Determine an estimate of the standard error of the sample mean using bootstrap with a bootstrap sample size of 10,000. Use the command `set.seed(12345)` to initialise the random number generator, before you start the bootstrap sampling. [7]

(vi) (a) Determine a bootstrap 90% confidence interval for the population mean of the y data using the same bootstrap sample as in part (v).

(b) Comment on the two confidence intervals produced in parts (iii) and (vi)(a). [6]

[Total 33]

3 Consider a sample of 10,000 car insurance policies sold to drivers who have had a driving licence for no more than 5 years. The data are given in the file `CarInsurance.Rdata`, which contains the following four variables:

- `age`: a factor taking the values 0 for young and 1 for old; this refers to the age of the policyholder.
- `LY`: the number of years the policyholder has had a driving licence before the start of the current year.
- `NCD`: the number of years the policyholder has had a no claims discount before the start of the current year.
- `claims`: the number of claims the policyholder submitted during the current year.

It has been suggested that drivers belonging to the old age category have more years of no claims discount than younger drivers.

(i) Plot two histograms for the number of years with no claims discount, one for young policyholders and another for old policyholders. [5]

(ii) Comment on the suggestion that drivers belonging to the old age category have more years of no claims discount than young drivers, using the histograms from part (i). [2]

(iii) Calculate the proportion of:

(a) policyholders with more than 2 years of no claims discount among young policyholders.

(b) policyholders with more than 2 years of no claims discount among old policyholders.

[4]

(iv) Test the hypothesis that the proportion of policyholders with more than 2 years of no claims discount is equal in both groups (young and old policyholders). [6]

An analyst assumes that the number of claims per year has a Poisson distribution. The analyst wishes to fit a generalised linear model (GLM) to the observed claim counts using `age`, `LY` and `NCD` as explanatory variables.

(v) Fit a GLM to the claim counts with `age`, `LY` and `NCD` as the main effect explanatory variables. [3]

To simplify the integration of the model into business processes, an insurer wishes to use only models with at most two explanatory variables.

(vi) Determine which two explanatory variables (out of `age`, `LY` and `NCD`) should be used in a GLM for the claim counts if only two explanatory variables are allowed. [8]

(vii) Comment on the comparison of the fit of the models in parts (v) and (vi). [2]

[Total 30]

4 An actuary produces the following summary of claim amounts and number of policies sold for four of her insurance clients in the past 5 years.

		Year 1	*Year 2*	*Year 3*	*Year 4*	*Year 5*
Insurer A	Claims (£m)	56.2	60.7	62.0	69.1	59.6
	Policies	1,079	1,221	1,728	1,769	1,812
Insurer B	Claims (£m)	18.3	18.7	18.6	18.9	19.2
	Policies	697	624	588	615	520
Insurer C	Claims (£m)	130.4	110.3	123.2	129.6	124.4
	Policies	2,896	2,601	2,757	2,923	2,842
Insurer D	Claims (£m)	63.2	67.3	43.2	47.0	42.1
	Policies	1,321	765	721	743	730

The policies expected to be sold in the next year are as follows:

	Year 6
Insurer A	1,920
Insurer B	575
Insurer C	2,820
Insurer D	798

The data for years 1 to 5 are given in the file `claims.Rdata`, which contains the claims data (`claims_matrix`) and the policies data (`policies_matrix`).

(i) Calculate estimates of $E[m(\theta)]$, $E[s^2(\theta)]$ and $\text{Var}[m(\theta)]$, under the assumptions of the Empirical Bayes Credibility Theory (EBCT) Model 2. [10]

(ii) Calculate the credibility factors Z_i. [5]

(iii) Calculate the expected claim amounts for each of the four insurers in year 6. [10]

[Total 25]

END OF PAPER