INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

24 April 2023 (am)

Subject CS1 – Actuarial Practice 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 on T. 0044 (0) 1865 268 873.

1 (i) Simulate 10,000 values from a N(0,1) distribution using an appropriate R command. You should save the generated values in R for later use. [1]

(ii) Simulate 10,000 values from a χ_4^2 distribution using an appropriate R command. You should save the generated values in R for later use. [2]

(iii) Discuss the main disadvantage of using the inverse transform method for simulating values from a chi-square distribution. [3]

(iv) Simulate 10,000 values from a t_4 distribution using your answers to parts (i) and (ii), explaining your answer.

You should save the generated values in R for later use. [5]

(v) Simulate 10,000 values from a t_{20} distribution by modifying your R code from parts (i), (ii) and (iv) where necessary.

You should save the generated values in R for later use. [5]

(vi) Demonstrate an important property of the t distribution by presenting and comparing two appropriate plots of the values simulated in parts (iv) and (v). [6]

[Total 22]

2 Consider a sample of 1,000 insurance policies. For each policy, there is a record of whether the policyholder has submitted at least one claim during the last 6-month period, the age of the policyholder and the duration of the policy. The number of policies, X, with one or more submitted claims is modelled as a random variable with a Binomial distribution $X \sim \text{Binomial}(1000, p)$.

(i) State the underlying assumptions that lead to the choice of a Binomial distribution for X. [2]

Use the command `load("CS1policies.Rdata")` to load the observed data into R. This will create a data frame called `CS1policies` with three entries in your R workspace. You can check the names of the entries with `names(CS1policies)`:

- `age`: age of the policyholder in years
- `duration`: duration of the policy in months
- `claimed`: takes the values 0 (no claim in 6-month period) or 1 (at least one claim in 6-month period).

An Actuary wishes to fit different Generalised Linear Models (GLMs) to the data, assuming that the number of policies with one or more submitted claims has a Binomial distribution and the link function of the GLM is the logit function.

(ii) Fit a GLM to the data such that p depends on the age of the policyholder and report the estimated parameter values. [5]

(iii) Fit a GLM to the data such that p depends on the duration of the policy and report the estimated parameter values. [5]

(iv) Compare the fit of the models in parts (ii) and (iii) using the deviance and the Akaike's Information Criterion (AIC). [3]

(v) Fit a GLM to the data, such that p depends on both the age of the policyholder and the duration of the policy, and report the estimated parameter values. [5]

(vi) Comment on which of the three models in parts (ii), (iii) and (v) is most appropriate, justifying your answer. [2]

[Total 22]

3 The file `heights.RData` contains an R data frame (`heights`) that consists of a random sample of heights of mothers (`mother.height`), fathers (`father.height`) and their daughters (`daughter.height`) in inches.

(i) Draw a scatterplot for each pair of data. [3]

(ii) Comment on the relationships between the pairs of data by referring to the plots from part (i). [4]

(iii) Fit a multiple linear regression model with the daughters' height as the response variable and mothers' and fathers' height as the explanatory variables. Your answer should show a summary of the output and the fitted equation. [5]

(iv) Assess the effect of each explanatory variable on the dependent variable by referring to the R output from part (iii). [4]

(v) Plot the residuals of the model fitted in part (iii) using a graph that is appropriate for checking if the estimated errors are independent of the explanatory variables. [5]

(vi) Comment on the fit of the model by referring to the plot from part (v). [2]

A suggestion is made that when the mother's height is 61 inches and the father's height is 63 inches, their daughter's height should be 67.5 inches.

(vii) Comment on whether this claim is likely by using an appropriate 95% predictive interval of the daughter's height based on the fitted model. [4]

[Total 27]

4 The table below shows the total claim amounts (£1,000s) per year, X_{ij}, over a 5-year period for five insurance companies' critical illness book of business.

		Year, j				
		2016	*2017*	*2018*	*2019*	*2020*
Insurance company, *i*	A	5,000	2,720	3,170	2,950	6,300
	B	3,680	3,360	3,900	2,910	8,020
	C	880	800	550	620	1,890
	D	6,150	3,880	5,780	5,220	7,100
	E	1,100	970	1,900	2,300	3,430

The data can be entered into R in matrix form using the following code:

```
claims <-
matrix(c(5000,3680,880,6150,1100,2720,3360,800,3880,970,
3170,3900,550,5780,1900,2950,2910,620,5220,2300,6300,
8020,1890,7100,3430),nrow=5,ncol=5)
```

(i) Calculate, using Empirical Bayes Creditability Theory (EBCT) Model 1, the following:

(a) $E[m(\vartheta)]$ [2]

(b) $\text{Var}[m(\vartheta)]$ [3]

(c) $E[s^2(\vartheta)]$ [2]

(ii) Calculate the expected claim amount for Company E, using your answers from part (i). [3]

The table below shows additional information that has been provided showing the number of claims, P_{ij}, over a 5-year period for the five insurance companies.

		Year, j				
		2016	*2017*	*2018*	*2019*	*2020*
Insurance company, *i*	A	580	540	490	500	650
	B	440	460	360	390	510
	C	180	160	170	120	220
	D	990	1,000	880	930	1,100
	E	210	200	175	170	130

The data can be entered into R in matrix form using the following code:

```
volumes <-
matrix(c(580,440,180,990,210,540,460,160,1000,200,490,
360,170,880,175,500,390,120,930,170,650,510,220,
1100,130),nrow=5,ncol=5)
```

(iii) Calculate, using EBCT Model 2, the following:

(a) $E[m(\vartheta)]$ [7]

(b) $\text{Var}[m(\vartheta)]$ [4]

(c) $E[s^2(\vartheta)]$ [3]

(iv) Calculate the expected claim amount for Company E, using your answers from part (iii), assuming that the number of claims for Company E in the following year remains at 130. [3]

(v) Comment on the difference between models EBCT Model 1 and EBCT Model 2 applied in earlier parts, by comparing your answers in parts (ii) and (iv). [2]

[Total 29]

END OF PAPER