INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

25 September 2020 (am)

Subject CS1B – Actuarial Statistics
Core Principles

Time allowed: One hour and forty-five minutes

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.

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

1 An employee at a café has been trained to set the coffee machine so that an espresso coffee portion results in 2.0 grams of coffee being placed into a cup. Knowing that variations are expected, the employee pours eight portions and measures the amounts to be 1.95, 1.80, 2.10, 1.82, 1.75, 2.01, 1.83 and 1.90.

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

```
amounts=c(1.95,1.80,2.10,1.82,1.75,2.01,1.83,1.90)
```

(i) Calculate an 80% confidence interval for the mean size of espresso coffee portions. [4]

(ii) Comment on whether the mean portion of coffee is equal to 2.0 grams. [2]
[Total 6]

2 A researcher has collected the following data on a group of students, regarding whether they passed or failed an exam and whether or not they attended tutorials:

Number of students	*Exam passed*	*Exam failed*
Attended tutorials	132	27
Did not attend tutorials	120	51

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

```
exam.success = matrix(c(132,120,27,51),ncol=2,nrow=2)
```

The reasercher wants to establish whether tutorial attendance is independent of exam success, using a chi-square test.

(i) State the hypotheses of this test. [2]

(ii) Calculate the expected frequencies for the data under the null hypotheses in part (i). [3]

(iii) Perform the test stating clearly your conclusion. [6]
[Total 11]

3 A machine in a sweet factory fills bags of sweets to weigh 500 grams. The actual weight of the sweet bags is known to follow a Normal distribution. The sweet manufacturer believes that the machine is under-filling the sweet bags. A sample of 10 sweet bags is taken and weighed, as summarised below.

Bag	1	2	3	4	5	6	7	8	9	10
Weight (grams)	474.11	512.01	493.64	495.03	518.13	486.03	494.48	501.76	498.83	503.02

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

```
weight=c(474.11,512.01,493.64,495.03,518.13,486.03,
494.48,501.76,498.83,503.02)
```

(i) Perform a suitable *t*-test to determine whether the sweet bags are being consistently under-filled, stating the hypotheses and the level of significance used in the test. [8]

(ii) Propose an interpretation of your conclusion in part (i). [2]
[Total 10]

4 Data were collected on average alcohol and cigarette consumption per adult per year for a number of countries. The data are given in the file `smoking_data.RData` and contain the following information:

country: the country concerned;
alcohol: alcohol consumption per adult per year (litres/year);
cigarettes: number of cigarettes consumed per adult per year.

(i) (a) Construct a scatterplot of the data with alcohol consumption on the x axis.
(b) Comment on the relationship between the data on alcohol and cigarette consumption based on your plot in part (i)(a).
[5]

(ii) Calculate Pearson's correlation coefficient between the data on alcohol and cigarette consumption. [2]

An analyst suggests using the following R code to modify the data:

```
alcohol.2 = alcohol[-c(6,16)]
cigarettes.2 = cigarettes[-c(6,16)]
```

(iii) Explain what the above code does and give a justification for its use. [3]

For the remainder of the question, use the modified data (`alcohol.2, cigarettes.2`), as produced by applying the R code above.

(iv) (a) Construct a scatterplot with alcohol consumption on the x axis.
(b) Calculate Pearson's correlation coefficient between the new data on alcohol and cigarette consumption.
(c) Comment on your answers in parts (ii) and (iv)(b).
[6]

(v) Perform a hypothesis test for the null hypothesis that Pearson's population correlation coefficient is equal to zero, against the alternative that it is positive. You should report the p-value of the test and a clear conclusion. [5]

A media bulletin has reported that 'higher alcohol consumption causes higher cigarette consumption'.

(vi) Comment on whether this report is justified based on your analysis in parts (iv) and (v). [2]

(vii) Perform a simple linear regression analysis on the new data using a model of the form $y = \alpha + \beta x + \varepsilon$ (cigarette consumption, y, on alcohol consumption, x), where the error terms ε independently follow a $N(0, \sigma^2)$ distribution.

Your answer should show the fitted line plotted on the data scatterplot and report the estimate of parameter σ. [5]

(viii) State the proportion of the total variability of the responses explained by the model, based on your output in part (vii). [1]

(ix) Plot a graph of the residuals of the model fitted in part (vii) against the explanatory variable. [2]

(x) Comment on the validity of the model, based on your output in part (ix) . [3]

[Total 34]

5 A waiting time random variable X follows an Exponential distribution with rate λ parameterised as $f(x) = \lambda e^{-\lambda x}$ ($x > 0, \lambda > 0$).

The rate λ has a Gamma prior distribution with parameters α and s. A Bayesian credibility model provides that the posterior mean of $1/\lambda$ can be expressed as

$$Z \times \bar{x} + (1 - Z) \times \frac{s}{\alpha - 1}, \text{ where } Z = \frac{n}{\alpha + n - 1}$$

with n being the number of past waiting times observed.

Assume that the parameters of the prior Gamma distribution are $\alpha = 5$ and $s = 1$.

(i) Determine an estimate of the posterior mean of $1/\lambda$ assuming $n = 10$ by implementing $M = 3{,}000$ Monte Carlo repetitions. [14]

(ii) Determine an estimate of the posterior mean of $1/\lambda$ and the value of $\bar{x}$ when $n = 1{,}000$, by implementing $M = 3{,}000$ Monte Carlo repetitions. [15]

(iii) Plot the histograms of the samples of the posterior mean of $1/\lambda$ and of $\bar{x}$ obtained in part (ii). [4]

(iv) Compare, by visual inspection of the graphs in part (iii), the distributions of the posterior mean of $1/\lambda$ and the distribution of $\bar{x}$ when $n = 1{,}000$. [2]

(v) Comment on the behaviour of the credibility model as n increases, relating your answer to your findings in part (iv). [4]

[Total 39]

END OF PAPER