INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

18 September 2019 (am)

Subject CS1B – Actuarial Statistics
Core Principles

Time allowed: One hour and forty-five minutes

INSTRUCTIONS TO THE CANDIDATE

1. *You are given this question paper and data files.*

2. *Mark allocations are shown in brackets.*

3. *Attempt all questions. Begin your answer to each question on a new page.*

*If you encounter any issues during the examination, please contact the Examinations Team at
T. +44 (0) 1865 268 255*

© Institute and Faculty of Actuaries

1   In a very small empirical study a sample from a random variable X is observed. The data can be entered into R using the following code:

```
x =
c(0.22,0.38,1.28,0.54,0.56,1.36,0.55,0.37,0.43,0.46,0.62,
0.54,0.54,0.51,0.44,0.68,0.55,0.30)
```

   (i)   Estimate the expected value of X.   [2]

   (ii)   Calculate a 95% confidence interval for the expected value of X assuming that X is Normally distributed.   [3]

   (iii)   Construct a confidence interval for the expected value of X using the bootstrap method with 10,000 bootstrap replications.   [5]

   (iv)   Comment on the differences between the confidence intervals in parts (ii) and (iii).   [3]
   [Total 13]

2   The dataset "Interest_rates.csv" contains a series of returns on bonds of maturities 1 year, 5 years, 10 years, 15 years, 20 years and 30 years (i.e. bonds that provide a return of the principal investment after 1, 5, 10, 15, 20 and 30 years respectively).

   (i)   (a)   Calculate the Pearson correlation coefficient matrix for this data.
           (b)   Comment on the correlations between the data using the matrix from part (i)(a).

   [8]

   (ii)   (a)   Perform a reduction of the dimension of the data using principal component analysis with the method of singular value decomposition. Your answer should include a summary of the principal component analysis.
           (b)   Suggest with reasons, using the output of the R analysis, how many components of the transformed data should be retained.
   [10]
   [Total 18]

3   Use the command `set.seed(2019)` to initialise the random number generator. When you execute any R code, make sure you run the entire R script including the line `set.seed(2019)`.

Consider a random sample $X_1, \ldots, X_n$ from an Exponential distribution with parameter λ and define $Y = \sum_{i=1}^{n} X_i$.

(i)   State the distribution of Y, giving all the parameters of the distribution.   [3]

(ii)   Perform simulation of a sample $x_1, \ldots, x_n$ with sample size $n = 15$ from an exponential distribution with parameter $\lambda = 2$.   [2]

(iii) Calculate the value of Y for the sample in part (ii).   [1]

(iv)   Perform 1,000 repetitions of parts (ii) and (iii) to obtain a Bootstrap sample $y_1, \ldots, y_B$ from the random variable Y with $B = 1,000$.   [8]

(v)   Plot a histogram showing the relative frequencies of the sample $y_1, \ldots, y_B$ from part (iv).   [2]

(vi)   (a)   Compare graphically the histogram in part (v) to the density of a suitable Normal distribution. You might find the following R command useful:

```
curve(dnorm(x,mean= , sd= , add=TRUE, lwd=2,col="red"))
```

       (b)   Comment on your findings in part (vi)(a) in the context of the Central Limit Theorem.

                                                                                        [5]
                                                                              [Total 21]

4   A recent study suggests that the maximum heart rate of a person, related to age in years, is given by the equation:

$$\text{Max Rate} = 220 - \text{Age}$$

Suppose this is to be empirically proven and 15 people of varying ages are tested for their maximum heart rate. The following data are collected:

Age (years)	18	23	25	35	65	54	34	56	72	19	23	42	18	39	37
Max Rate	202	186	187	180	156	169	174	172	153	199	193	174	198	183	178

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

```
x = c(18,23,25,35,65,54,34,56,72,19,23,42,18,39,37)

y = c(202,186,187,180,156,169,174,172,153,199,193,174,
    198,183,178)
```

(i)   Plot the fitted line for the regression of Max Rate on Age.   [5]
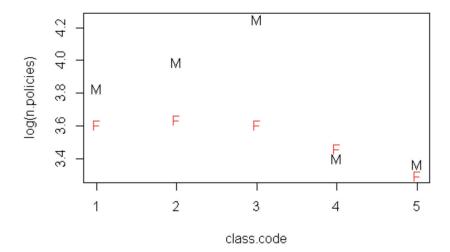
(ii)   Comment on the results.   [2]

A researcher reviews the plot in part (i) and suggests the slope should be equal to –1.

(iii)   Calculate the *p*-value of a hypothesis test for this suggestion, by creating a suitable test statistic.   [7]

(iv)   Comment on the researcher's suggestion, using your answer to part (iii).   [2]

[Total 16]

5 The data given in the file `policies_data.RData` show the numbers of policies (`n.policies`) by sex of policyholder (`sex.code`; 1 for male, 2 for female) and class of business (`class.code`; 5 different classes) from a certain insurance portfolio.

  (i)   (a)   Construct a plot of the logarithm of the number of policies (on the *y* axis) against the class of business.

        (b)   Comment on the relationship in the data based on your plot in part (i)(a).

[5]

In the plot produced in part (i) we can distinguish between male and female policyholders. The plot is shown below, with "M" and "F" showing male and female policyholders respectively:

  (ii)   Comment on the relationship in the data based on this plot.   [2]

For the remainder of the question you will need to ensure that the sex and class variables are treated as categorical variables (factors). You can use the following R code:

```
class.code = as.factor(class.code)
sex.code = as.factor(sex.code)
```

  (iii)   Fit a generalised linear model analysis to the data, using a Poisson distribution, with the numbers of policies as the response variable and the class of business as the only factor. Your answer should include estimates of the parameters, corresponding *p*-values and a brief interpretation of their effect.   [8]

  (iv)   Fit a second Poisson generalised linear model analysis to the data, using the numbers of policies as the response variable and both the class of business and the sex of the policyholders as factors. Your answer should include estimates of the parameters, corresponding *p*-values and a brief interpretation of their effect.   [8]

(v)   Determine, using the deviance, which of the two models used in parts (iii) and (iv) provides a better fit to the data. Your answer should include the null hypothesis, the *p*-value of the relevant test and a clear conclusion.   [6]

(vi)   Calculate the predicted number of policies for male policyholders when the class of business is 2, based on the model chosen in part (v).   [3]

[Total 32]

END OF PAPER