Queen Mary University of London

Main Examination period 2022 – May/June – Semester B

MTH5131: Actuarial Statistics – SOLUTIONS

$[C]$ = Similar to Coursework, $[U]$ = Unseen

Question 1 [10 marks]. [U]

(a) It is not simple random sampling because each subset does not have the same probability of being chosen and it is not stratified sampling because the students are not divided inbto groups. **[4]**

(b) This is stratified sampling because the students are divided into groups from which a number of students are chosen randomly. **[3]**

(c) This is neither because students are not chosen randomly. The students who go to the library may not be typical. **[3]**

Total marks so far: 10

Please remove the showmarks command before passing the exam to the checker

---

Question 2 [12 marks]. [C]

(a) The average of the first column is $(19 + 22 + 6 + 3 + 2 + 20)/6 = 12$. The average of the second column is $(12 + 6 + 9 + 15 + 13 + 5)/6 = 10$. The centred matrix is therefore

$$X = \begin{pmatrix} 7 & 2 \\ 10 & -4 \\ -6 & -1 \\ -9 & 5 \\ -10 & 3 \\ 8 & -5 \end{pmatrix}.$$

The sample covariance matrix is

$$\frac{1}{6-1}X^T X = \begin{pmatrix} 86 & -27 \\ -27 & 16 \end{pmatrix}.$$

[5]

 Continue to next page

(b) The characteristic polynomial of $X^T X$ is

$$(86-\lambda)(16-\lambda)-27^2=\lambda^2-102\lambda+647$$

The eigenvalues are

$$\lambda=\frac{102\pm\sqrt{102^2-4\times 647}}{2}=95.20407, 6.795928$$

The component corresponding to 95.20407 satisfies

$$\begin{pmatrix} 86-95.20407 & -27 \\ -27 & 16-95.20407 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

so

$$y=(86-95.20407)x/27=-0.3408916x$$

Taking $x=1$, $\begin{pmatrix} 1 \\ -0.3408916 \end{pmatrix}$ is an eigenvector. Normalising gives the component

$$\frac{1}{\sqrt{1^2+0.3408916^2}}\begin{pmatrix} 1 \\ -0.3408916 \end{pmatrix}=\begin{pmatrix} 0.9465153 \\ -0.3226591 \end{pmatrix}$$

[7]

Total marks so far: 22

Please remove the showmarks command before passing the exam to the checker

---

Question 3 [16 marks]. [C]

(a) (i) With **x** production and **y** salaries, we have

$$\sum_{i=1}^{n} x_i = 11800,\ \sum_{i=1}^{n} x_i^2 = 32540000,\ \sum_{i=1}^{n} y_i = 1520,\ \sum_{i=1}^{n} y_i^2 = 661000,\ \sum_{i=1}^{n} x_i y_i = 4429000,$$

$$S_{xx} = 4692000,\ S_{yy} = 198920,\ S_{xy} = 841800 \Rightarrow r = \frac{841800}{\sqrt{4692000 \times 198920}} = 0.8713461$$

[5]

(ii) Under H_0, the t-value with $5-2=3$ degrees of freedom is is

$$\frac{0.8713461 \times \sqrt{5-2}}{\sqrt{1-0.8713461^2}} = 3.0758$$

The upper 2.5% point of the t_3 distribution is 3.182. We accept H_0 and conclude that $\rho = 0$.

OR

The p-value is 0.0543131. We accept H_0 and conclude that $\rho = 0$. **[5]**

(b) The ranks of **x** are $(1, 2, 4, 5, 3)$ and the ranks of **y** are $(2, 3, 4, 5, 1)$ We make a table of concordant and discordant pairs:

Rank1	Rank2	C	D
1	2	3	1
2	3	2	1
4	4	2	0
5	5	1	0
3	1		

Totalling the columns gives $n_c = 8$ and $n_d = 2$ Thus

$$\tau = \frac{8-2}{(5)(4)/2} = 0.6$$

[6]

Total marks so far: 38

Please remove the showmarks command before passing the exam to the checker

Question 4 [13 marks]. [C]

(a) The likelihood is

$$\begin{aligned} L(\theta; \underset{\sim}{y}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(y_i-5)^2}{2\theta}\right) \\ &= (2\pi\theta)^{-n/2} \exp\left(-\sum_{i=1}^{n} \frac{(y_i-5)^2}{2\theta}\right) \end{aligned}$$

and so

$$\ln L(\theta; \underset{\sim}{y}) = -\frac{n}{2}\ln(2\pi\theta) - \sum_{i=1}^{n} \frac{(y_i-5)^2}{2\theta}$$

$$\frac{d}{d\theta}\ln L(\theta; \underset{\sim}{y}) = -\frac{n}{2\theta} + \sum_{i=1}^{n} \frac{(y_i-5)^2}{2\theta^2}$$

and

$$\frac{d^2}{d\theta^2}\ln L(\theta; \underset{\sim}{y}) = \frac{n}{2\theta^2} - \sum_{i=1}^{n} \frac{(y_i-5)^2}{\theta^3}.$$

Therefore the Fisher Information is

$$\mathbb{E}\left(-\frac{n}{2\theta^2} + \sum_{i=1}^{n} \frac{(Y_i-5)^2}{\theta^3}\right) = -\frac{n}{2\theta^2} + \frac{n\mathrm{Var}(Y_i)}{\theta^3} = -\frac{n}{2\theta^2} + \frac{n\theta}{\theta^3} = \frac{n}{2\theta^2}$$

Thus,

$$\mathrm{CRLB}(\theta) = \frac{\left(\frac{d\theta}{d\theta}\right)^2}{n/2\theta^2} = \frac{2\theta^2}{n}.$$

[7]

(b)

$$\mathbb{E}\left(\frac{(Y_1-5)^2 + (Y_2-5)^2 + \cdots + (Y_n-5)^2}{n}\right) = \frac{n\mathrm{Var}(Y_i)}{n} = \frac{n\theta}{n} = \theta$$

and so the estimator is unbiased.

$$\mathrm{Var}\left(\frac{(Y_1-5)^2 + (Y_2-5)^2 + \cdots + (Y_n-5)^2}{n}\right)$$

$$= \frac{n\mathrm{Var}(Y_i-5)^2}{n^2} = \frac{n(2\theta^2)}{n^2} = \frac{2\theta^2}{n} = \mathrm{CRLB}(\theta)$$

and so the estimator is MVUE. **[6]**

Total marks so far: 51

Please remove the showmarks command before passing the exam to the checker

---

Question 5 [13 marks]. [C]

(a) The expectation of the Y_i is

$$\mathbb{E}(Y_i) = \int_0^1 y\theta y^{\theta-1} dy = \frac{\theta}{\theta+1}.$$

Therefore $\mathbb{E}(\overline{Y}) = \frac{n\theta/(\theta+1)}{n} = \frac{\theta}{\theta+1}$ as well. The bias is

$$\mathbb{E}(\overline{Y}) - \frac{\theta}{\theta+1} = 0.$$

The variance is

$$\mathrm{Var}(\overline{Y}) = \frac{n\mathrm{Var}(Y_i)}{n^2} = \frac{\mathrm{Var(Y_i)}}{n}$$

Thus,

$$\mathrm{MSE} = \frac{\mathrm{Var(Y_i)}}{n} + 0^2 \to 0 \text{ as } n \to \infty$$

and $\overline{Y}$ is consistent. **[6]**

(b) The bias is now

$$\frac{n}{n+1}\mathbb{E}(\overline{Y}) - \frac{\theta}{\theta+1} = \frac{\theta}{\theta+1}\left(\frac{n}{n+1} - 1\right) = -\frac{\theta}{(n+1)(\theta+1)}$$

The variance is

$$\left(\frac{n}{n+1}\right)^2 \mathrm{Var}(\overline{Y}) = \left(\frac{n}{n+1}\right)^2 \frac{\mathrm{Var(Y_i)}}{n}$$

Thus

$$\mathrm{MSE} = \left(\frac{n}{n+1}\right)^2 \frac{\mathrm{Var(Y_i)}}{n} + \left(\frac{\theta}{(n+1)(\theta+1)}\right)^2 \to 0 \text{ as } n \to \infty$$

and $\frac{n}{n+1}\overline{Y}$ is consistent. **[7]**

Total marks so far: 64

Please remove the showmarks command before passing the exam to the checker

---

Question 6 [11 marks]. [U]

(a) The likelihood is $\text{Binomial}(24, \theta)$, so

$$L(\theta; 3) = \binom{24}{3} \theta^3 (1-\theta)^{21} \propto \theta^3 (1-\theta)^{21}.$$

and the prior is

$$f(\theta) \propto \theta^{0.5}(1-\theta)^{0.5}$$

. Therefore, the posterior density is

$$f(\theta|3) \propto \theta^3(1-\theta)^{21} \times \theta^{0.5}(1-\theta)^{0.5} = \theta^{3.5}(1-\theta)^{21.5}$$

and the posterior is $\text{Beta}(4.5, 22.5)$ distributed. **[7]**

(b) The mode of the posterior is found by

$$\frac{d}{d\theta}\theta^{3.5}(1-\theta)^{21.5} = 0$$

$$3.5\theta^{2.5}(1-\theta)^{21.5} - 21.5\theta^{3.5}(1-\theta)^{20.5} = 0$$

$$3.5(1-\theta) - 21.5\theta = 0$$

so

$$\theta = \frac{3.5}{3.5 + 21.5} = 0.14$$

[4]

Total marks so far: 75

Please remove the showmarks command before passing the exam to the checker

---

Question 7 [13 marks]. [C,U]

(a) For the Normal/Normal model, in the formula for the credibility factor

$$Z = \frac{n}{n + \frac{\sigma_1^2}{\sigma_2^2}} = \frac{n}{n + \frac{400}{500}} = \frac{n}{n + \frac{4}{5}}$$

where n should be taken as the number of years of past data.

Year	Pure Premium	Credibility factor at the start of year	Average pure premium based on number of years of past data available at the start of year	At the start of the year, the credibility estimate of the pure premium in the coming year
1	760	0.000	0.000	700.000
2	735	0.556	760.000	733.333
3	790	0.714	745.500	733.939

[13] 4 for each column, 1 for Z formula

Total marks so far: 88
Please remove the showmarks command before passing the exam to the checker

Question 8 [12 marks]. [U]

(a) The distribution is Gamma, the link is identity, the linear predictor is $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3$, where x_1 is size, x_2 is the factor new, and x_3 is the number of beds. **[3]**

(b) size:beds is the least significant coefficient, so the next model to be be checked is the model in (a), except the linear predictor is now $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2$ **[4]**

(c) The price would be

$$44.3759 + 0.0740 \times 2064 + 22.7131 \times 3 + 0.0100 \times 2064 * 3 = 327.1712$$

so the price is £$327, 171.$. **[5]**

Total marks so far: 100
Please remove the showmarks command before passing the exam to the checker

End of Paper.