Queen Mary
University of London

Main Examination period 2023 – May/June – Semester B

MTH5131: Actuarial Statistics - SOLUTIONS

Duration: 2 hours

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

Question 1 [10 marks]. $[U]$

(a) The study took a random sample of people but does not randomly assign people to different groups. The study simply observed whether the people were light, moderate, or heavy smokers and their stress level. Thus, this was an observational study. Other sensible answers accepted. **[4]**

(b) This study is descriptive because it summarises the data and is not intended to enable the user to draw any specific conclusions. Other sensible answers accepted. **[3]**

(c) This is called truncated data. **[3]**

Total marks so far: 10

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

---

Question 2 [12 marks].

(a) $[U]$ For each $1 \leq i \leq n$,

$$A\mathbf{u}_i = \sum_{j=1}^{n} \lambda_j \mathbf{u}_j \mathbf{u}_j^T \mathbf{u}_i = \sum_{j=1}^{n} (\mathbf{u}_i \cdot \mathbf{u}_j) \lambda_j \mathbf{u}_j = \lambda_i \mathbf{u}_i$$

because the $\mathbf{u}_i$ are orthonormal. Therefore the λ_i are all eigenvalues. **[7]**

(b) $[C]$ The average of the first column is $(2+4)/2 = 3$. The average of the second column is $(3+5)/2 = 4$. The centred matrix is therefore

$$X = \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}.$$

The sample covariance matrix is

$$\frac{1}{2-1} X^T X = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}.$$

Clearly, or by calculation, components are

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

[5]

Total marks so far: 22

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

---

Question 3 [17 marks]. [C]

(a) (i) The differences in rank are $d_1 = 0$, $d_2 = 0$. $d_3 = 1$, $d_4 = 1$, $d_5 = 3$, $d_6 = -5$. Thus, $\sum_{i=1}^{5} d_i^2 = 36$ and

$$r_s = 1 - \frac{6 \times 36}{6 \times 35} = -0.02857143$$

[5]

(ii) Under H_0, the t-value with $6 - 2 = 4$ degrees of freedom is

$$\frac{r_s\sqrt{6-2}}{\sqrt{1-r_s^2}} = -0.0571662$$

The lower 2.5% point of the t_4 distribution is -2.776. We don't reject H_0; there isn't enough evidence to say that they are correlated.

[5]

(b) We make a table of concordant and discordant pairs:

Rank1	Rank2	C	D
1	6	0	5
2	1	4	0
3	3	2	1
4	4	1	1
5	2	1	0
6	5		

Totalling the columns gives $n_c = 8$ and $n_d = 7$ Thus

$$\tau = \frac{8-7}{(6)(5)/2} = 0.0667$$

[7]

Total marks so far: 39

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

---

Question 4 [8 marks]. [C]

 Continue to next page

(a) The bias of the estimator is

$$\text{bias}(T_n) = \mathbb{E}(T_n) - \pi = \frac{nm\pi}{(n+1)m} - \pi = \frac{n\pi}{n+1} - \pi = -\frac{\pi}{n+1}.$$

The variance of the estimator is

$$\text{Var}(T_n) = \frac{nm\pi(1-\pi)}{(n+1)^2m^2} = \frac{n\pi(1-\pi)}{(n+1)^2m}.$$

The mean square error is

$$\text{MSE}(T_n) = \text{bias}(T_n)^2 + \text{Var}(T_n) = \frac{\pi^2}{(n+1)^2} + \frac{n\pi(1-\pi)}{(n+1)^2m}.$$

[6]

(b) Since $\lim_{n\to\infty} \text{MSE}(T_n) = 0$, the sequence of estimators T_n is consistent. **[2]**

Total marks so far: 47

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

---

Question 5 [19 marks]. $[C, U]$

(a) The likelihood is

$$L(\theta, \underline{y}) = \prod_{i=1}^{n} \frac{\theta 2^{\theta}}{y_i^{\theta+1}} = \frac{\theta^n 2^{n\theta}}{(\prod_{i=1}^{n} y_i)^{\theta+1}}$$

and so

$$\ln L(\theta, \underline{y}) = n \ln \theta + n\theta \ln 2 - (\theta + 1) \sum_{i=1}^{n} \ln y_i$$

We have

$$-\frac{d^2}{d\theta^2} \ln L(\theta, \underline{y}) = \frac{n}{\theta^2}$$

We are taking $\phi(\theta) = \frac{\theta}{\theta-1} = 1 + \frac{1}{\theta-1}$ and therefore $\phi'(\theta) = -\frac{1}{(\theta-1)^2}$. The Cramér-Rao lower bound is

$$\text{CRLB}(\phi) = \frac{[-1/(\theta-1)^2]^2}{n/\theta^2} = \frac{\theta^2}{(\theta-1)^4 n}$$

[7]

(b)

$$E(Y) = \int_2^\infty y f_Y(y) dy = \int_2^\infty \frac{\theta 2^\theta}{y^\theta} = \theta 2^\theta \frac{1}{-\theta+1} y^{-\theta+1} \Big|_2^\infty = \frac{2\theta}{\theta - 1}.$$

We have $\overline{y} = \frac{3+4+3+2+10}{5} = \frac{22}{5}$. Therefore

$$\frac{2\hat{\theta}}{\hat{\theta} - 1} = \frac{22}{5} \Rightarrow \tilde{\theta} = \frac{11}{6} = 1.8333$$

We see that $\tilde{\theta} > 1$ as required. **[5]**

(c) The likelihood is

$$L(\theta; \underline{y}) = \prod_{i=1}^5 \frac{\theta 2^\theta}{y_i^{\theta+1}} = \frac{\theta^5 2^{5\theta}}{(\prod_{i=1}^5 y_i)^{\theta+1}}$$

The log-likelihood is

$$\ell(\theta; \underline{y}) = 5 \ln \theta + 5\theta \ln 2 - (\theta + 1) \ln \left(\prod_{i=1}^5 y_i \right)$$

and its derivative is

$$\frac{d\ell}{d\theta} = \frac{5}{\theta} + 5 \ln 2 - \ln \left(\prod_{i=1}^5 y_i \right) \Rightarrow \frac{5}{\theta} + 5 \ln 2 - \ln 720 = 0$$

$$\Rightarrow \hat{\theta} = \frac{5}{\ln 720 - 5 \ln 2} = 1.6059$$

We see that $\hat{\theta} > 1$ as required. **[7]**

Total marks so far: 66
Please remove the showmarks command before passing the exam to the checker

---

Question 6 [8 marks]. [C,U]

(a) The likelihood is Geometric(θ) or Binomial($30, \theta$), so

$$L(\theta; 30) = (1 - \theta)^{29} \theta.$$

and the prior is

$$f(\theta) \propto \theta^0 (1 - \theta)^{18}.$$

Therefore, the posterior density is

$$f(\theta|3) \propto (1 - \theta)^{29} \theta \times (1 - \theta)^{18} = \theta (1 - \theta)^{47}$$

and the posterior is Beta($2, 48$) distributed. **[6]**

(b) The expectation of the posterior is $\frac{2}{2+48} = 0.04$ **[2]**

Total marks so far: 74
Please remove the showmarks command before passing the exam to the checker

---

Question 7 [13 marks]. [C]

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

$$Z = \frac{n}{n+\beta} = \frac{n}{n+2}$$

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	160	0.000	0.000	170.000
2	185	0.333	160.000	166.667
3	150	0.500	172.500	171.25

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

Total marks so far: 87

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

Question 8 [13 marks].

(a) $[U]$ Assets is a variable, Nation is a factor with 4 categories, and Sector is a factor with 10 categories. The degrees of freedom for the first model is $248 - (10 + 4 - 1) - 1 = 234$. The degrees of freedom for the second model is $248 - (10 + 4 - 1) = 235$. The degrees of freedom for the third model is $248 - 10 - 1 = 237$. The degrees of freedom for the fourth model is $248 - 4 - 1 = 243$. **[6]**

(b) $[C]$ The residual deviances for the three submodels and change in degrees of freedom are

Covariates	Δ Deviance	Δ DF
Nation, Sector	390.90	1
Assets, Sector	328.94	3
Assets, Nation	361.46	9

The p-values are $P(\chi_1 > 390.90) = 5.27 \times 10^{-87}$, $P(\chi_3 > 328.94) = 5.41 \times 10^{-71}$, $P(\chi_9 > 361.46) = 2.25 \times 10^{-72}$, all of which are very small. Therefore, all of the covariates in the largest model are significant. **[7]**

Total marks so far: 100

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

End of Paper.