The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS Spring Semester 2018–2019

MAS474 Extended linear models - SOLUTIONS **2 hours**

Restricted Open Book Examination.
Candidates may bring to the examination lecture notes and associated lecture material (but no textbooks) plus a calculator which conforms to University regulations.
Answer all questions. Total marks 60.

Please leave this exam paper on your desk
Do not remove it from the hall

Registration number from U-Card (9 digits)
to be completed by student

Blank

1 (i) (a) (Routine)

$$V_{ijk} = \mu + b_i + b_{ij} + \epsilon_{ijk} \checkmark\checkmark$$

where $b_i \sim N(0, \sigma_1^2)$ $\quad b_{ij} \sim N(0, \sigma_2^2)$ $\quad \epsilon_{ijk} \sim N(0, \sigma^2) \checkmark\checkmark$

(b) (Routine)

$$\hat{\mu} = 18.8 \checkmark \quad \hat{\sigma_1^2} = 142.7 \quad \hat{\sigma_2^2} = 1.1 \quad \hat{\sigma^2} = 0.0079 \checkmark$$

(ii) (Routine) Use the GLRT

$$\begin{aligned} L &= -2(\log l(\hat{\psi}_r) - \log l(\hat{\psi})) \checkmark \\ &= -2(-151.3 - -53.7) \\ &= 123.2 \checkmark \end{aligned}$$

Compare this with a χ_1^2 distribution $\checkmark$.

Find very strong evidence to reject H_0 $\checkmark$.

(iii) (Unseen)

$$\hat{\mu} = \frac{1}{60} \sum_{ijk} V_{ijk} \checkmark$$

Thus

$$\begin{aligned} \mathbb{V}\mathrm{ar}(\hat{\mu}) &= \mathbb{V}\mathrm{ar}\left(\frac{1}{60} \sum_{i=1}^{10} \sum_{j=1}^{3} \sum_{k=1}^{2} (\mu + b_i + b_{ij} + \epsilon_{ijk}) \right) \checkmark \\ &= \frac{1}{60^2} \mathbb{V}\mathrm{ar}\left(60\mu + \sum_{i=1}^{10} 6b_i + \sum_{i=1}^{10} \sum_{j=1}^{3} 2b_{ij} + \sum_{i=1}^{10} \sum_{j=1}^{3} \sum_{k=1}^{2} \epsilon_{ijk} \right) \checkmark \\ &= \frac{1}{60^2} \left(\sum_{i=1}^{10} 36\sigma_1^2 + \sum_{i=1}^{10} \sum_{j=1}^{3} 4\sigma_2^2 + \sum_{i=1}^{10} \sum_{j=1}^{3} \sum_{k=1}^{2} \sigma^2 \right) \checkmark \\ &= \frac{\sigma_1^2}{10} + \frac{\sigma_2^2}{30} + \frac{\sigma^2}{60} \checkmark \\ &= \frac{142.7}{10} + \frac{1.1}{30} + \frac{0.0079}{60} = 14.3 \checkmark \end{aligned}$$

And so the standard error of $\hat{\mu}$ is $\sqrt{14.3} = 3.77 \checkmark$.

(iv) (Unseen) We can use QQ plots and residual plots $\checkmark$. There are 3 levels of residual $\checkmark$ we can look at:

- Level 0 $V_{ijk} - \mu$
- Level 1 $V_{ijk} - \mu - \hat{b}_i$
- Level 2 $V_{ijk} - \mu - \hat{b}_i - \hat{b}_{ij}$

$\checkmark$

Be generous - anything sensible gets the marks.

2 (i) (a) (Unseen)

$$L(\theta) = \prod w^{Y_i}(1-w)^{1-Y_i} e^{-\lambda(1-Y_i)-\mu Y_i} \lambda^{x_i(1-Y_i)} \mu^{x_i Y_i}/x_i! \checkmark\checkmark\checkmark$$

Thus

$$l(\theta; \mathbf{x}, \mathbf{Y}) = -\mu \sum Y_i - \lambda \sum (1-Y_i) + \log\lambda \sum (1-Y_i)x_i + \log\mu \sum Y_i x_i - \sum \log x_i! + \log w \sum Y_i + \log(1-w) \sum (1-Y_i) \checkmark\checkmark$$

(b) (Unseen)

$$\begin{aligned} \mathbb{E}(Y_i|x_i,\theta) &= \mathbb{P}(Y_i = 1|x_i,\theta) \checkmark \\ &= \frac{\mathbb{P}(x_i|Y_i = 1,\theta)\mathbb{P}(Y_i = 1,\theta)}{\mathbb{P}(x_i|\theta)} \checkmark \\ &= \frac{e^{-\mu}\mu^{x_i} w/x_i!}{e^{-\mu}\mu^{x_i} w/x_i! + e^{-\lambda}\lambda^{x_i}(1-w)/x_i!} \checkmark \\ &= \frac{w\mu^{x_i}e^{-\mu}}{w\mu^{x_i}e^{-\mu} + (1-w)\lambda^{x_i}e^{-\lambda}} \checkmark \end{aligned}$$

(c) (Routine)

$$Q(\theta|\theta^{(m)}) = -\mu \sum p_i - \lambda \sum (1-p_i) + \log\lambda \sum (1-p_i)x_i + \log\mu \sum p_i x_i - \sum \log x_i! + \log w \sum p_i + \log(1-w) \sum (1-p_i) \checkmark\checkmark$$

$$\frac{dQ}{dw} = \frac{1}{w}\sum p_i - \frac{1}{1-w}\sum(1-p_i)$$

Setting $\frac{dQ}{dw} = 0$ and solving gives $\hat{w} = \frac{\sum p_i}{n}$ ✓.

$$\frac{dQ}{d\lambda} = -\sum(1-p_i) - \frac{1}{\lambda}\sum x_i(1-p_i)$$

Setting $\frac{dQ}{d\lambda} = 0$ and solving gives $\hat{\lambda} = \frac{\sum x_i(1-p_i)}{\sum(1-p_i)}$ ✓.

$$\frac{dQ}{d\mu} = -\sum p_i - \frac{1}{\mu}\sum x_i p_i$$

Setting $\frac{dQ}{d\mu} = 0$ and solving gives $\hat{\mu} = \frac{\sum x_i p_i}{\sum p_i}$ ✓.

(d) (Routine) The expression for w is the average value of $\mathbb{P}(Y_i = 1|x_i,\theta)$. The expression for λ is the average number of emails received by lecturers when weighted by the probability they are arts lecturers, and the expression for μ is the average number of emails received by lecturers when weighted by the probability they are science lecturers. ✓✓✓

(ii) (Unseen)

(a) NMAR ✓

(b) MCAR ✓

(c) MAR ✓

Question 2 continued on next page

2 (continued)

3 (i) (a) (Routine) Multiple imputation using chained equations ✓

It creates $m = 5$ ✓complete datasets by replacing the missing values using a variety of imputation methods.✓The approach begins by filling in missing values by sampling with replacement ✓. It then

- replaces missing values in `dist` and `climb` using linear regression given the other values
- replaces missing values in `time` using unconditional mean imputation ✓✓

It uses stochastic imputation ✓meaning that a random error is added and parameter uncertainty is accounted for ✓(Bayesian regression approach).

It then iterates through the 4 steps above until convergence ✓.

(maximum of 6 ✓)

(b) (unseen) We can estimate the expected value by combing the parameter estimates from each imputed datasets

$$\begin{aligned}\mathbb{E}(\beta_1|Y_{obs}) &= \frac{1}{m}\sum \hat{\beta}_1^{(i)} \checkmark \\ &= \frac{6.46 + 6.15 + 6.50 + 6.16 + 6.41}{5} = 6.336\end{aligned}$$

To calculate the variance, we need to use the corrected version of the posterior variance

$$Var(\theta|Y_{obs}) \approx \bar{V} + (1 + \frac{1}{m})B \checkmark$$

where

$$\bar{V} = \frac{1}{5}(0.49 + 0.56 + 0.47 + 0.52 + 0.37) = 0.482$$

is the average within imputation variability. ✓

and

$$B = (\frac{1}{4}\sum \beta_3^{(i)2} - \bar{\beta}_3^2) = 0.029 \checkmark$$

is the between imputation variability which can be read from the R output (or computed).

Thus

$$Var(\theta|Y_{obs}) \approx 0.482 + (1 + 1/5)0.029 = 0.5168$$

and so the estimated standard error is $\sqrt{0.5168} = 0.72$ ✓.

 Question 3 continued on next page

3 (continued)

(ii) (a) (Unseen)

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 7 \\ 1 & 14 \\ 1 & 1 \\ \vdots & \\ 1 & 14 \end{pmatrix} \checkmark \qquad \beta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \checkmark$$

$$Z_a = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \checkmark \qquad a = \begin{pmatrix} a_1 \\ \vdots \\ a_5 \end{pmatrix} \checkmark$$

$$Z_b = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 7 & 0 & 0 & 0 & 0 \\ 14 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & 0 & 14 \end{pmatrix} \checkmark \qquad b = \begin{pmatrix} b_1 \\ \vdots \\ b_5 \end{pmatrix} \checkmark$$

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{53} \end{pmatrix} . \checkmark$$

(b) (Routine) `lmer(Weight   Days + (1| RatID) + (Days-1|RatID)`

$\checkmark\checkmark$

The answer `lmer(Weight   Days + (Days|RatID)` gets 1 mark only.

End of Question Paper