MAS473 Extended linear models
2013-14 Exam Solutions

(1)(i) Model is

$$Y_{ijk} = \beta_i + b_j + b_{ij} + \epsilon_{ijk},$$

[1 mark]

where

Y_{ijk} is the blood clotting time for drug i, volunteer j and replicate k,

$i = 1, \ldots, 3$, $j = 1, \ldots, 10$ and $k = 1, 2$

[1 mark]

$b_j \sim N(0, \sigma_1^2)$

$b_{ij} \sim N(0, \sigma_2^2)$

$\epsilon_{ijk} \sim N(0, \sigma^2)$.

[1 mark]

(1)(ii)

$$\begin{aligned} \hat{\sigma}_1^2 &= 0.001150, \\ \hat{\sigma}_2^2 &= 0.030827, \\ \hat{\sigma}^2 &= 0.008216. \end{aligned}$$

[1 mark]

(1)(iii)

$$\widehat{Var}(Y_{ijk}) = \widehat{Var}(b_i) + \widehat{Var}(b_{ij}) + \widehat{Var}(\epsilon_{ijk}) = 0.040193.$$

[1 mark]

(1)(iv)

$$\begin{aligned} \widehat{Cov}(Y_{ijk}, Y_{ijk'}) &= \widehat{Cov}(b_i + b_{ij} + \epsilon_{ijk}, b_i + b_{ij} + \epsilon_{ijk'}) \\ &= \widehat{Var}(b_i) + \widehat{Var}(b_{ij}), \\ \Rightarrow \widehat{Cor}(Y_{ijk}, Y_{ijk'}) &= (0.001150 + 0.030827)/0.040193 = 0.796. \end{aligned}$$

[M1 A1]

(1)(v)(a)

$$\begin{aligned} Var(\bar{Y}_{i..}) &= Var\left(\frac{1}{20}\sum_{j=1}^{10}\sum_{k=1}^{2}(\beta_i + b_j + b_{ij} + \epsilon_{ijk})\right) \\ &= \frac{Var(b_i)}{10} + \frac{Var(b_{ij})}{10} + \frac{Var(\epsilon_{ijk})}{20} \end{aligned}$$

[2 marks]

Estimated standard error is

$$\sqrt{\frac{0.001150}{10} + \frac{0.030827}{10} + \frac{0.008216}{20}} = 0.06007$$

[1 mark]

(1)(v)(b)

$$\begin{aligned} Cov(\bar{Y}_{i..}, \bar{Y}_{i'..}) &= \frac{1}{400}Cov\left(\sum_{j=1}^{10}\sum_{k=1}^{2}(\beta_i + b_j + b_{ij} + \epsilon_{ijk}), \sum_{j=1}^{10}\sum_{k=1}^{2}(\beta_{i'} + b_j + b_{i'j} + \epsilon_{i'jk})\right) \\ &= \frac{1}{400}Cov\left(\sum_{j=1}^{10}\sum_{k=1}^{2} b_j, \sum_{j=1}^{10}\sum_{k=1}^{2} b_j\right) \\ &= \frac{Var(b_j)}{10} \end{aligned}$$

[3 marks]

Estimated correlation is

$$\frac{0.001150/10}{0.06007^2} = 0.032.$$

[1 mark]

(1)(v)(c) The standard error is smaller, because under a fixed effect model, there is no variability in the mean of the drug 1 observations due to volunteer or volunteer-drug interaction random effects. **[1 mark]**

In the fixed effects model, the `drug1` term is the expected clotting time, averaged over the 10 volunteers in the study.

[1 mark]

In the mixed effects model, the `drug1` term is the expected clotting time, averaged over the population of volunteers from which the 10 were drawn.

[1 mark]

(1)(vi) Test hypothesis $H_0 : \sigma^2_{b_2} = 0$.

[1 mark]

Using GLRT, we compute

$$-2(9.289661 - 22.15275) = 24.189,$$

greater than $\chi^2_{1;0.99}$, so strong evidence against H_0. Conclude that there is evidence of significant random interaction effects between drug and volunteer: the difference in blood clotting times between drugs will vary between volunteers. **[1 mark for test, 1 for conclusion]**

(2)(i) $E(y_i) = exp(\eta_i)/(1 + exp(\eta_i))$**[1 mark]**
$\eta_i = \beta_0 + \beta_1 s_i + \beta_2 t_i + \beta_3 s_i t_i$**[1 mark]**
where s_i is 1 if the ith surface is B (0 otherwise) and t_i is the thickness for the ith tyre. **[1 mark]**.

(2)(ii) $n_1, ..., n_5 = 100, 50, 40, 50, 50$ **[1 mark]**
$y_1, ..., y_5 = 75/100, 25/50, 11/40, 10/50, 3/50$ **[1 mark]**

(2)(iiia) With $\eta_i = \beta_0 + \beta_1 s_i$:
Test 1, $H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$. $\Delta D = 5.82 > 3.84$ so H_0 rejected. **[1 mark]**
With $\eta_i = \beta_0 + \beta_1 s_i + \beta_2 t_i$:
Test 2, $H_0 : \beta_2 = 0$ $H_1 : \beta_2 \neq 0$. $\Delta D = 176.61 > 3.84$ so H_0 rejected. **[1 mark]**
With $\eta_i = \beta_0 + \beta_1 s_i + \beta_2 t_i + \beta_3 s_i t_i$:
Test 3, $H_0 : \beta_3 = 0$ $H_1 : \beta_3 \neq 0$. $\Delta D = 0.33 < 3.84$ so H_0 not rejected. **[1 mark]**
So both main effects needed but not the interaction. H_1 must be specified in all tests. **[1 mark]**

(2)(iiib) $\eta_3 = 5.091 - 1.744 \times 3.4 = -0.839$ **[1 mark]**
$\hat{\mu_3} = \frac{exp(-0.839)}{1+exp(-0.839)} = 0.302$ **[1 mark]**
Pearson residual is $\frac{11/40-0.302}{\sqrt{0.302*(1-0.302)/40}} = -0.372$ **[1 mark for 11/40, 1 mark for answer]**

(2)(iiic) odds of Y=1 for thickness 5mm is $exp(5.091 - 5 \times 1.744) = 0.027$
odds of Y=1 for thickness 3.4mm is $exp(5.091 - 3.4 \times 1.744) = 0.432$
[1 mark] for both right
So odds ratio is $0.027/0.432 = 0.0625$ **[1 mark]**
log OR is difference in the log odds so equals $1.6\hat{\beta}_2$ so variance is $1.6^2 \times 0.0446$ **[1 mark]**
95% CI for logOR is $log(0.0625) \pm 1.96 \times 1.6 \times \sqrt{0.047} = (-3.44, -2.11)$**[1 mark]**

So 95% CI for OR is (0.03,0.12) and so significant evidence that probability of splitting is not the same for the two tyre thicknesses. **[1 mark]**

(2)(iiid) Any suitable that include the same thickness in both linear predictors but different surfaces **[2 marks]**

(3)(i) Controlled variables are considered to have fixed marginal totals, response variables are not. **[1 mark]**
Minimal model contains `helmet*collision` only in linear predictor. **[1 mark]**

(3)(ii) Percentages are **[1 mark]**

helmet			no helmet		
collision	serious	minor	collision	serious	minor
lorry	52%		lorry	60%	
car	19%		car	46%	
pedestrian	11%		pedestrian	14%	

Helmets don't seem to affect the probability of a serious injury in collisions with pedestrians or lorries but they seem to reduce it in collisions with cars. **[1 mark]**
The probability of a serious injury is lower in collisions with pedestrians compared with cars and in collisions with cars compared to lorries.**[1 mark]**

(3)(iii) Terms are

- intercept
- indicator variable for helmet
- 2 indicator variables for collision
- indicator variable for outcome
- 2 interaction indicator variables for collision*helmet

[1 mark]
$n = 12, p = 7$ so $df = 5$. **[1 mark]**

(3)(iv) $\eta_{ijk} = \beta_0 + \alpha_i + \gamma_j + \tau_{ij} + \delta_k + \theta_{ik}$
with $\alpha_1 = \gamma_1 = \delta_1 = 0$
and $\tau_{1.} = \tau_{.1} = \theta_{1.} = \theta_{.1} = 0$ with the notation correctly defined **[2 marks]**.

(3)(v)

Model	Residual Deviance	Df
A helmet*collision+outcome	25.643	5
B helmet*collision+outcome*helmet	20.751	4
C helmet*collision+outcome*collision	8.371	3
D helmet*collision+outcome*helmet+outcome*collision	2.318	2

Comparison	change in df	change in res dev	chi-sq threshold
A,B	1	4.89	3.84
A,C	2	17.27	5.99
C,D	1	6.06	3.84
B,D	2	18.43	5.99

2 marks for each of the three nested comparisons, all have Δ res.dev greater than threshold so final model is chosen.

Yes seems consistent. The need for the interaction terms not in the minimal model says that the probability of a serious outcome depends on both helmet and collision type and the particular combination of helmet and collision type.**[1 mark]**

(3)(vi) $\mu = 54 \times \frac{10+17}{(10+17)+(44+20)}$ **[1 mark for 54, M1 for collapsing across levels of helmet, 1 mark for correct values in the fraction]** $= 54 \times \frac{27}{91} = 16.02198$ **[1 mark]**