CITY UNIVERSITY OF HONG KONG STUDENTS' UNION
http://www.cityusu.net/

Name: CHAN King Yeung
SID: 1155119394

STAT4006

Problem Sheet 4

Question 1

a) $\log\left(\frac{\hat{\pi}_1}{\hat{\pi}_2}\right) = \log\left(\frac{\hat{\pi}_1}{\hat{\pi}_3}\right) - \log\left(\frac{\hat{\pi}_2}{\hat{\pi}_3}\right)$

$= (1.785 + 1.044x_1 + 0.733x_2) - (1.554 + 0.254x_1 - 0.106x_2)$

$= 0.231 + 0.79x_1 + 0.839x_2$

b) $e^{1.044 \pm 1.96(0.259)} \approx (1.7098, 4.7192)$

the estimated odds for females response 'Yes' rather than 'no' on life after death are at least 1.7 times and at most 4.7 times those for males.

c) $\hat{\pi}_1(x_1=1, x_2=0) = \frac{e^{1.785+1.044}}{1+e^{1.785+1.044}+e^{1.554+0.254}}$

≈ 0.7046

d) for white males $(x_1=0, x_2=0)$,

$\log\left(\frac{\hat{\pi}_1}{\hat{\pi}_2}\right) = \hat{\alpha}_1 = 0.231$ $\Rightarrow \hat{\pi}_1 > \hat{\pi}_2$

$\log\left(\frac{\hat{\pi}_2}{\hat{\pi}_3}\right) = \hat{\alpha}_1 = 1.554$ $\Rightarrow \hat{\pi}_2 > \hat{\pi}_3$

for black females $(x_1=1, x_2=1)$,

$\log\left(\frac{\hat{\pi}_1}{\hat{\pi}_2}\right) = \hat{\alpha}_1 + \beta_i^G + \beta_i^R = 1.86$ $\Rightarrow \hat{\pi}_1 > \hat{\pi}_2$

$\log\left(\frac{\hat{\pi}_2}{\hat{\pi}_3}\right) = \hat{\alpha}_1 + \beta_i^G + \beta_i^R = 1.692$ $\Rightarrow \hat{\pi}_2 > \hat{\pi}_3$

e) for all the $\hat{\beta}$ in either $\log\left(\frac{\hat{\pi}_1}{\hat{\pi}_2}\right)$ or $\log\left(\frac{\hat{\pi}_1}{\hat{\pi}_3}\right)$ are positive, that is, female $(x_1=1)$ always get a higher value than males for each race.

f) $H_0: \beta^G = 0$ vs $H_1: \beta^G \neq 0$

$G^2 = 46.74 - 0.69$
$= 46.05$

$df = 2$

p-value $\approx 1.0009 \times 10^{-10}$

Since p-value < 0.05, we reject H_0 at $\alpha = 0.05$.
We conclude that opion is not independent of gender, given race.

Question 2

a) for $i < j$, $\text{logit}[P(Y \le i)] - \text{logit}[P(Y \le j)] = (\alpha_i - \alpha_j) + (\beta_i - \beta_j)x$. For the nature of cumulative logit model, the above result cannot be positive. When $\beta_i > \beta_j$, large positive x leads to positive result. Similarly, when $\beta_i < \beta_j$, large negative x lead to positive result as well. Thus, for $x \in \mathbb{R}$ violates the nature of the model which is improper in cumulative probabilities

b) for x is a binary indicator, it avoid the problem in a) while the help from α_i to adjust the proper cumulative probabilities. From the nature of cumulative logit model, we need monotone increasing in $\alpha_i + \beta_i$. Thus, the requirement of ordering constraint on α_i help to achieve the nature.

Question 3

complete independence model

$(X,Y,Z) = \log(\mu_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$

joint independence model

$(X,YZ) = \log(\mu_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ}$

$(Y,XZ) = \log(\mu_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ}$

$(Z,XY) = \log(\mu_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}$

conditional independence model

$(XZ,YZ) = \log(\mu_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$

$(XY,ZY) = \log(\mu_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{kj}^{ZY}$

$(YX,ZX) = \log(\mu_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ji}^{YX} + \lambda_{ki}^{ZX}$

homogeneous association model

$(XY,YZ,XZ) = \log(\mu_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ} + \lambda_{ik}^{XZ}$

saturated model

$(XYZ) = \log(\mu_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ} + \lambda_{ik}^{XZ} + \lambda_{ijk}^{XYZ}$

Question 4

λ_{ij}^{XY} determines the $\log(\theta)$

$\Rightarrow \log(\theta) = \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY}$

$= 0.1368$

$\Rightarrow \theta = e^{0.1368}$

≈ 1.1466

Question 5

a)

model	AIC	BIC
(A, C, M)	1294	1316.92
(AC, M)	853.8	882.45
(AM, C)	949.6	978.25
(CM, A)	544.2	572.85
(AM, CM)	199.8	234.18
(AC, AM)	509.4	543.78
(AC, CM)	104	138.38
(AC, AM, CM)	54.4	94.51

the parsimonious model is (AC, AM, CM) under AIC and BIC.

b) loglinear model:

$\log(\mu_{ijk}) = \mu + \lambda_i^A + \lambda_j^C + \lambda_k^M + \lambda_{ij}^{AC} + \lambda_{ik}^{AM} + \lambda_{jk}^{CM}$

logit model:

$\text{logit}(\pi_i) = \log\left(\frac{\mu_{i1k}}{\mu_{i2k}}\right)$

$= \log(\mu_{i1k}) - \log(\mu_{i2k})$

$= (\lambda_1^C - \lambda_2^C) + (\lambda_{i1}^{AC} - \lambda_{i2}^{AC}) + (\lambda_{1k}^{CM} - \lambda_{2k}^{CM})$

$= \alpha + \beta_i^A + \beta_k^M$

Question 6

a) model (GH, GI, HI) is a homogeneous model. Thus, its degree of freedom is $(I-1)(J-1)(K-1) = 1$.

b) interaction terms in model (GH, HI), (GI, HI) or (GH, GI, HI) are significant since the p-values of these models are greater than 0.05 in the LRT. Although model (GH, GI, HI) fits the best among adequate models, i.e., highest p-value, it is more complex and difficult to interpret. Thus, we will choose either model (GH, HI) or (GI, HI) depending on the interested study.