

DEPARTMENT OF STATISTICS

STATS 330: STATISTICAL MODELLING

Assignment 5: Model answer for Question 5 (part 5).

Let μ_{ijk} be the Poisson mean for the cell corresponding to Leukoplakia = i , ($i = 1$: Yes, $i = 2$: No), Smoking = j ($j = 1$: Yes, $j = 2$: No), and Alcohol = k ($k = 1$: None, $k = 2$: 0-40 gm, $k = 3$: 41-80 gm, $k = 4$: > 80 gm). We can express the mean as

$$\mu_{ijk} = \exp((\text{Int}) + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\beta\gamma)_{jk} + (\alpha\gamma)_{ik} + (\alpha\beta\gamma)_{ijk}).$$

The corresponding multinomial probabilities are

$$Pr(L = i, S = j, A = k) = \frac{\mu_{ijk}}{\sum_{ijk} \mu_{ijk}}.$$

The logistic model is a model for the conditional probability $Pr(L = 1|S = j, A = k)$. Using the standard probability formulas, this can be written as

$$Pr(L = 1|S = j, A = k) = \frac{Pr(L = 1, S = j, A = k)}{Pr(L = 1, S = j, A = k) + Pr(L = 2, S = j, A = k)}.$$

We can express this in terms of the Poisson means as

$$\begin{aligned} & \frac{Pr(L = 1, S = j, A = k)}{Pr(L = 1, S = j, A = k) + Pr(L = 2, S = j, A = k)} \\ &= \frac{\mu_{1jk}}{\mu_{1jk} + \mu_{2jk}}. \end{aligned}$$

Now substitute the expressions for the means in terms of the main effects and interactions. After some algebra and cancelling, and using the fact that any main effect or interaction is zero when $i = 1$, we get

$$\begin{aligned} \frac{\mu_{1jk}}{\mu_{1jk} + \mu_{2jk}} &= \frac{1}{1 + \exp(\alpha_2 + (\alpha\beta)_{2j} + (\alpha\gamma)_{2k} + (\alpha\beta\gamma)_{2jk})} \\ &= \frac{\exp(-\alpha_2 - (\alpha\beta)_{2j} - (\alpha\gamma)_{2k} - (\alpha\beta\gamma)_{2jk})}{1 + \exp(-\alpha_2 - (\alpha\beta)_{2j} - (\alpha\gamma)_{2k} - (\alpha\beta\gamma)_{2jk})}. \end{aligned}$$

Thus, the Poisson model implies that logit of $Pr(L = 1|S = j, A = k)$ is $-\alpha_2 - (\alpha\beta)_{2j} - (\alpha\gamma)_{2k} - (\alpha\beta\gamma)_{2jk}$. The standard logistic model is

$$\text{logit } Pr(L = 1|S = j, A = k) = (\text{Int}) + \beta'_j + \gamma'_k + (\beta'\gamma')_{jk}.$$

(the 's just mean that β'_j is different from the β_j we used above.) This is the same as that derived from the Poisson model if we identify (Int) with $-\alpha_2$, β'_j with $-(\alpha\beta)_{2j}$, γ'_k with $-(\alpha\gamma)_{2k}$ and $(\beta'\gamma')_{jk}$ with $-(\alpha\beta\gamma)_{2jk}$. Looking at the regression summaries, we see that the estimates from fitting the two models match up as well.