Ear Infections in Swimmers

Keywords: Poisson regression, overdispersion

Description

The data come from the 1990 Pilot Surf/Health Study of NSW Water Board. The first column takes values 1 or 2 according to the recruit's perception of whether (s)he is a Frequent OCean Swimmer, the second column has values 1 or 4 according to recruit's usually chosen swimming location (1 for non-beach, 4 for beach), the third column has values 2 (aged 15-19), 3 (aged 20-25), or 4 (aged 25-29), the fourth column has values 1 (male) or 2 (female) and finally, the fifth column has the number of self-diagnosed ear infections that were reported by the recruit.

Download

Data file (tab-delimited text)

Source

Val Gebski, from a private communication from Cameron Kirton of the New South Wales Water Board, Sydney, Australia.

Hand D.J., Daly F., Lunn A.D., McConway K.J., Ostrowski E. (1994). *A Handbook of Small Data Sets*. London: Chapman & Hall. Data set 328

Analysis

```
> glm.inf <- glm(Infections~Swimmer*Location*Age*Sex,family=poisson)
> round(anova(glm.inf,test="F"),2)
Analysis of Deviance Table

Poisson model

Response: Infections

Terms added sequentially (first to last)
                         Df Deviance Resid. Df Resid. Dev F Value Pr(F)
                       NULL                286     824.51
                    Swimmer  1    34.70       285     789.81   10.98  0.00
                   Location  1    25.16       284     764.65    7.96  0.01
                        Age  2     8.58       282     756.07    1.36  0.26
                        Sex  1     0.63       281     755.43    0.20  0.65
           Swimmer:Location  1     1.69       280     753.74    0.54  0.46
                Swimmer:Age  2     6.38       278     747.36    1.01  0.37
               Location:Age  2     3.92       276     743.44    0.62  0.54
                Swimmer:Sex  1     0.23       275     743.21    0.07  0.79
               Location:Sex  1    11.12       274     732.09    3.52  0.06
                    Age:Sex  2     1.78       272     730.31    0.28  0.75
       Swimmer:Location:Age  2     3.67       270     726.63    0.58  0.56
       Swimmer:Location:Sex  1     0.24       269     726.39    0.08  0.78
            Swimmer:Age:Sex  2     0.19       267     726.20    0.03  0.97
           Location:Age:Sex  2    13.94       265     712.26    2.21  0.11
Swimmer:Location:Age:Sex     2     8.54       263     703.72    1.35  0.26
```

The data is too dispersed to be Poisson (residual deviance 703.7 on 263 df). If the variance can be taken to be phi*mu, then it appears the only effects are main effects for frequence ocean Swimmer and Location.

```
> glm.inf <- glm(Infections~Swimmer+Location,family=poisson)
> tapply(fitted(glm.inf),list(Swimmer,Location),mean)
      NonBeach     Beach
Occas 2.261286 1.3596173
 Freq 1.224948 0.7365101
```

Obviously swimmers report fewer ear infections if they are frequent ocean swimmers, and if they usually swim at the beach.

```
> plot(fitted(glm.inf),residuals(glm.inf))
```

The residual plot shows no reason to doubt the assumed mean-variance relationship.
