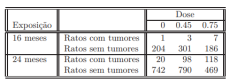
**Modelos Lineares Generalizados**

**LISTA DE EXERCÍCIOS nº 04**

1. Execute no SAS e R:

A fim de estudar os efeitos cancerígenos de um produto tóxico em ratos, foram administradas três diferentes doses da substância tóxica (0, 0.45 e 0.75 partes por 10 000) a algumas centenas de ratos, durante um de dois períodos de exposição (16 ou 24 meses). No final do período de exposição verificava-se a existência de tumores nos ratos. Os resultados da experiência foram os seguintes:



Ajustou-se um Modelo Linear Generalizado adequado para uma componente aleatória dicotômica, com função de ligação probit, considerando os preditores Dose e tempo de Exposição como variáveis numéricas. Considere: *glm(formula = cbind(com, sem) ~ Dose + Exposicao, family = binomial(probit), data = ratos)*

(a) Descreva em detalhes o tipo de modelo ajustado, indicando a relação considerada entre o surgimento de tumores e as variáveis preditoras.

(b) Comente a qualidade do ajustamento do modelo aos dados.

(c) Considera possível simplificar anteriormente o modelo sem prejuízo significativo na qualidade do ajustamento? Justifique formalmente.

Entretanto, é levantada a questão de que o baixíssimo número de diferentes valores dos preditores Dose e Exposição desaconselha a sua utilização como variáveis numéricas. Decidiu-se assim ajustar um novo modelo, com estes dois preditores considerados como fatores. Não se previram efeitos de interação entre os fatores. O ajustamento produziu os seguintes resultados:

*glm(formula = cbind(com, sem) ~ as.factor(Dose) + as.factor(Exposicao), family = binomial(probit), data = ratos)*

(d) Descreva em detalhes o modelo agora ajustado. Comente as semelhanças e diferenças com o modelo considerado inicialmente.

(e) Qual dos dois modelos é preferível?

2. Execute no SAS e R:

Considere um estudo dos efeitos sobre o sabor de vários aditivos de queijo. Os pesquisadores testaram quatro aditivos de queijo e obtiveram 52 classificações de resposta para cada aditivo. Cada resposta foi medida em uma escala de nove categorias, variando de forte (1) a excelente (9). Os dados, fornecidos em McCullagh e Nelder (1989, p. 175) na forma de uma tabela de frequência bidimensional de aditivo por classificação, são salvos no conjunto de dados Cheese usando o seguinte programa. A variável y contém a avaliação da resposta. A variável Additive especifica o aditivo de queijo (1, 2, 3 ou 4). A variável freq fornece a frequência com que cada aditivo recebeu cada classificação. A variável resposta é ordinal.

data Cheese;

do Additive = 1 to 4;

do y = 1 to 9;

input freq @@;

output;

end;

end;

label y='Taste Rating';

datalines;

0 0 1 7 8 8 19 8 1

6 9 12 11 7 6 1 0 0

1 1 6 8 23 7 5 1 0

0 0 0 1 3 7 14 16 11

;

run;

(a) Descreva em detalhes o tipo de modelo ajustado. (b) Comente a qualidade do ajustamento do modelo aos dados.