

## MI613 - Análise de Dados Categóricos

### 2ª Lista de Exercícios - Regressão Logística e Regressão de Poisson

1. Utilize os dados abaixo para ajustar um modelo de regressão logística. Agrupe as categorias de resposta Razoável e Muito bom e utilize o logito de Razoável/Muito bom contra ruim. Especifique sua matriz  $X$ , interprete os parâmetros e explique brevemente os passos para obtenção do melhor modelo.

Centro	Tratamento	Resposta			Total
		Ruim	Razoável	Muito Bom	
1	A	6	6	3	15
1	B	2	7	6	15
2	A	4	7	6	17
2	B	2	4	11	17
3	A	11	19	6	36
3	B	6	12	17	35

2. Os indivíduos abaixo foram classificados de acordo com sua residência (rural, urbana), nível de "stress" no dia a dia (baixo, alto), e atitude com relação a uma específica política de saúde (favorável, não favorável). Ajuste um modelo adequado para explicar a atitude com relação a política de saúde de acordo com a residência, nível de stress e possíveis interações.

Residência	"Stress"	Atitude		Total
		Não Favorável	Favorável	
Rural	Baixo	14	58	72
Rural	Alto	112	116	228
	Total	126	174	300
Urbana	Baixo	162	66	228
Urbana	Alto	64	8	72
	Total	226	74	300

3. Utilize os três conjuntos de dados entregues em sala e ajuste modelos de regressão logística apropriados. Observe que as variáveis resposta de interesse são: LOW (peso abaixo do normal=0, peso normal =1); IMPROVEMENT (nenhuma melhora=0, alguma=1, muita=2); CHD (ausência de doença das coronárias=0, presença de doença das coronárias=1).

Para o conjunto de dados droga.dat

(disponível na página [www.ime.unicamp.br/hildete/mi613.html](http://www.ime.unicamp.br/hildete/mi613.html))

as variáveis são: sexo, tratamento, idade e resposta, onde Feminino=1, Masculino=0, droga=1, placebo=0, nenhum=0, algum=1, muito=2. Descreva brevemente como você chegou ao melhor modelo, interprete os parâmetros de acordo com a codificação usada para a matriz  $X$ . Imagine que o pesquisador lhe entregou os dados para análise e você precisa explicar para ele quais os resultados encontrados a partir desse modelo. Faça um pequeno relatório sobre sua análise e suas interpretações. Comente sobre as interpretações das razões de odds, intervalos de confiança etc.

4. O conjunto de dados Col20.sd2 é um arquivo SAS que contém variáveis relacionadas com nível de colesterol de 873 indivíduos. A variável resposta é o colesterol total (COLESTER), que depois foi categorizada (COL) de acordo com o nível de colesterol (alto=1, médio=2 ou baixo=3). Estes níveis são definidos de acordo com o seguinte critério:

se  $COLESTER \geq 240$ , COL=1;

se  $200 \leq COLESTER < 240$ , COL=2;

se  $COLESTER < 200$ , COL=3.

É importante salientar que o metabolismo dos indivíduos é alterado de acordo com a idade e que as mulheres depois da menopausa mudam bastante seu metabolismo. O objetivo é ajustar um modelo que explique o nível de colesterol total em função de covariáveis de interesse como, por exemplo, idade, sexo, obesidade, fumo e outras.

Abaixo está o código em SAS para a leitura inicial dos dados:

```
libname insas "local do arquivo";
data vera2;
    set insas.Col20;
    if coles='>= 240' then col=1;
    if coles='200 |-- 240' then col=2;
    if coles=' < 200' then col=3;
    if sexo='M' then sex=1; else sex=0;
run;
proc contents data=vera2;
run;
```

Ajuste um modelo de regressão logística para o logito de alto contra médio ou baixo. Interprete e comente sobre os resultados.

5. Analise os dados do arquivo leucem2.sas e leucemia.sas. Encontre o melhor modelo e interprete os resultados.

6. Analise os dados do arquivo melanoma.sas. Encontre o melhor modelo e interprete os resultados.