

6. Regressão multinomial (e logística)

6.1. Num estudo sobre a sinistralidade rodoviária num dado distrito e ano obteve-se a seguinte informação sobre a gravidade dos acidentes e o local do acidente:

	Dentro de uma localidade	Fora de uma localidade
Sem vítimas	4229	1381
Com vítimas	1046	599

- A gravidade dos acidentes está significativamente relacionada com o local do acidente?
- Calcule duas chances (*odds*) e duas razões de chances (*OR*) à sua escolha, e interprete

6.2. Nos últimos anos, a atividade física regular tem vindo a ser bastante incentivada pelos profissionais da educação e da saúde e pelos responsáveis políticos. Também tem vindo a ser alvo de interesse do mercado, quer na oferta de diferentes serviços, quer nas tentativas de resposta aos tipos da procura que são cada vez mais diversificados. Na Carta Europeia do Desporto, a atividade física é vista num sentido mais amplo, englobando competição e prazer.

Com o objetivo de caracterizar a prática desportiva no concelho de Évora dos residentes com 15 ou mais anos de idade, em 2011 o CIMA-UE/DMAT e a DD-CME realizaram uma sondagem para a Câmara Municipal de Évora apresentando-se no ficheiro AFsub.csv (com *encoding* UTF-8) alguns dados obtidos.

Nome da variável	Descrição	Codificação
id	Identificação do número do respondente	
sexo	Variável sexo do respondente	0 = Feminino 1 = Masculino
idade	Idade do respondente	
instrucao	Grau de instrução do respondente	1 = 1.º ciclo ou inferior 2 = 2.º ciclo 3 = 3.º ciclo 4 = Secundário 5 = Superior
modalidade	Principal modalidade desportiva que o respondente pratica	1 = BTT/Ciclismo 2 = Caminhada 3 = Fitness 4 = Outra

- Avalie a existência de valores omissos no conjunto de dados.
- Crie uma variável que identifique se um residente é praticante de caminhadas.
- Realize uma análise preliminar que permita analisar a existência de relação entre a variável ser praticante de caminhadas com o sexo, idade e nível de instrução.
- Ajuste um modelo de regressão logística para um dado indivíduo residente no concelho ser praticante de caminhadas em função do sexo, da idade e do nível de instrução (nesse caso categorizando entre os que possuem ensino superior e os restantes).
 - Ajuste o modelo logístico múltiplo preliminar, mantendo apenas as significativas.
 - Avalie a possibilidade de juntar categorias, caso se aplique.
 - Avalie o pressuposto de linearidade.
 - Avalie a possível existência de multicolineariedade.

- v) Avalie a possível existência de interações entre as variáveis presentes no modelo e que possam ter algum sentido.
 - vi) Avalie a adequabilidade do modelo e a bondade do ajustamento (recorrendo a testes e a medidas).
 - vii) Avalie a capacidade discriminativa do modelo.
- e) Admita que para o modelo ajustado na alínea anterior já foi realizada uma análise de resíduos que não revelou observações influentes nem *outliers*.
- i) Escreva a equação que descrevem o modelo ajustado e interprete os coeficientes.
 - ii) Obtenha as estimativas intervalares, a 95%, para cada um dos parâmetros do modelo.
 - iii) Estime as probabilidades associadas às preferências dos residentes, do sexo feminino, com idade 30, 40, e 50 anos.
 - iv) Compare, em função da idade (a variar entre 15 e 80 anos), as probabilidades de dois residentes serem praticantes de caminhadas, diferenciando por sexo. Faça a representação gráfica incluindo bandas de confiança a 90%.

6.3. Retome o exercício 6.2 anterior.

- a) Realize uma análise preliminar que permita analisar a existência de relação entre a variável modalidade com o sexo, idade e nível de instrução.
- b) Ajuste os modelos multinomiais (de categoria de referência) univariados.
- c) Ajuste um modelo de regressão multinomial de efeitos simples do tipo de modalidade em função do sexo, da idade e do nível de instrução.
- d) Avalie a possível existência de interações entre as variáveis presentes no modelo e que possam ter algum sentido.
- e) Averigue a possibilidade de juntar categorias e valide o pressuposto de linearidade.
- f) Admita que para esse modelo já foi realizada uma análise de resíduos que não revelou observações influentes nem *outliers*. Avalie a adequabilidade do modelo e a bondade do ajustamento.
- g) Escreva o conjunto de equações que descrevem o modelo ajustado e interprete os coeficientes.
- h) Estime as probabilidades associadas às preferências dos residentes, do sexo masculino, em função da sua idade.
- i) Avalie a capacidade discriminativa do modelo

6.4. Na Universidade de Évora, foi realizada a dissertação de Mestrado intitulada “Análise Estatística da Aptidão Física em Ambiente Escolar” que procurou identificar fatores que podem influenciar a aptidão física em ambiente escolar e o rendimento escolar dos alunos.

Foram recolhidos dados em três escolas (0 – E.B. Dr. Hernâni Cidade de Redondo, 1 – E. B. Conde de Vilalva de Évora, 2 - Escola Secundária de Castelo de Paiva) junto de uma amostra de alunos destas escolhas. No ficheiro *imc.xls*, estão disponíveis as variáveis: idade, o sexo, IMC, IMC2cat (0 - IMC fora da zona saudável (ZS), 1 – IMC dentro da ZS) e IMC3cat (0 - IMC dentro da ZS, 1 – IMC acima da ZS, 2 – IMC abaixo da ZS).

Ajuste um modelo de regressão multinomial que permita identificar os fatores associados às diferentes categorias de IMC.