

Modelos lineares generalizados - Lista de Exercícios 3

Aviso exercícios para entregar: 23/10 - Data de entrega: 25/10

Exercício 1. Faça o Exercício 16 do Capítulo 2 livro de Gilberto Paula versão 2013.

OBS1: Não é necessário ajustar o modelo normal heterocedástico, pois não o discutimos em detalhes no curso.

OBS2: Interprete as estimativas dos parâmetros no modelo que você escolheu.

Exercício 2. Faça o Exercício 19 do Capítulo 2 livro de Gilberto Paula versão 2013.

Exercício 3. Considere o banco de dados auto-resumido.xls descrito em

'<http://archive.ics.uci.edu/ml/datasets/Auto+MPG>' (note que X_8 é uma variável qualitativa). Assuma que um modelo linear generalizado com resposta gama e função de ligação logarítmica é adequado para o ajuste dessa variável resposta.

- (a) Construa diagramas de dispersão envolvendo o logaritmo da variável resposta e cada uma das variáveis preditoras. O que esses gráficos parecem sugerir?
- (b) Selecione as variáveis utilizando algum método de seleção de variáveis e descreva resumidamente o método utilizado.
- (c) Apresente os resultados do modelo final e interprete as estimativas dos parâmetros em termos do problema.
- (d) Realize uma análise de diagnóstico completa do modelo final. O modelo gama com função de ligação logarítmica parece ser adequado?
- (e) Parece haver algum ponto alavanca, discrepante ou influente? Caso haja um ou mais, ajuste o modelo com e sem esses pontos e compare os resultados. Algum desses pontos parecem alterar consideravelmente alguma das conclusões obtidas?
- (f) Obtenha o pseudo R^2 do modelo e comente.

Exercício 4. Considere o banco de dados e o modelo final que você encontrou no exercício 3.

- (a) Teste se a média da variável resposta varia em função da variável qualitativa X_8 .
- (b) Teste se média da variável resposta varia entre os níveis 1 e 2, 1 e 3 e 2 e 3 da variável X_8 .
- (c) Obtenha um intervalo de confiança para a razão entre a média da variável resposta entre os níveis 2 e 3 da variável X_8 e interprete-o.
- (d) Obtenha um intervalo de confiança para a média da variável resposta se $x_2 = 4$, $x_3 = 150$, $x_4 = 100$, $x_5 = 2300$, $x_6 = 17$, $x_7 = 80$, $x_8 = 1$.
- (e) Seguindo algum critério razoável, categorize a variável x_7 e troque em seu modelo final x_7 como quantitativa pela x_7 categorizada. De que forma você acharia que seria melhor incluir x_7 no modelo neste caso?
- (f) Inclua no modelo final interação entre a variável qualitativa e as demais variáveis preditoras e teste se as interações podem ser retiradas.

Exercício 5. Considere o banco de dados e o modelo final que você encontrou no exercício 3.

- (a) Ajuste o modelo com duas outras funções de ligação. Você acha que seria melhor considerar o modelo com função de ligação logarítmica ou com uma das outras funções de ligação?
- (b) Troque a distribuição da variável resposta para a normal e depois para a normal inversa. Você acha que seria melhor usar uma dessas duas distribuições ao invés da distribuição gama?