

Universidade Federal do Paraná - Departamento de Estatística
CE071 - Análise de Regressão Linear
Prof. Cesar Augusto Taconeli

Lista de exercícios - seleção de modelos e regressão com covariáveis categóricas

Nota: Os exercícios desta lista foram extraídos do livro **Linear models with R**, de Julian J. Faraway. Todas as bases de dados estão disponíveis no pacote `faraway` do R. Antes de iniciar qualquer análise, consulte a documentação da base, verifique o contexto dos dados e a descrição das variáveis. Adicionalmente, todas as questões devem ser precedidas por uma análise descritiva/exploratória, composta por gráficos e medidas descritivas pertinentes.

Exercício 1- Use o conjunto de dados `prostate` com `lpsa` como resposta e as demais variáveis como explicativas. Implemente os seguintes métodos de seleção de covariáveis para determinar o “melhor” ajuste:

- a) Eliminação *backward* baseada no teste F;
- b) Seleção *stepwise* baseada no critério de informação de Akaike;
- c) Seleção *stepwise* baseada no critério de informação Bayesiano;
- d) Todas as regressões possíveis, baseado no R^2 ajustado;
- e) Todas as regressões possíveis, baseado no C_p de Mallows.

Compare os modelos resultantes da aplicação dos cinco métodos.

Exercício 2- Semelhante ao exercício 1, mas usando a base de dados `divusa`. Considere a variável `divorce` como resposta e as demais variáveis como explicativas.

Exercício 3- Use a base de dados `seatpos` e considere `hipcenter` como a variável resposta.

- a) Ajuste o modelo de regressão com as oito variáveis explicativas. Comente a respeito do efeito do comprimento da perna na resposta;
- b) Obtenha o intervalo de predição 95% considerando o vetor de médias das covariáveis;
- c) Conduza a seleção de covariáveis baseada no AIC. Repita os itens a e b usando o modelo selecionado. Compare os resultados produzidos pelos dois modelos.

Exercício 4- Os patrimônios (em bilhões de dólares) para 232 bilionários são apresentados na base de dados `fortune`.

- a) Apresente o gráfico do patrimônio *vs* idade separado de acordo com as diferentes regiões do mundo;
- b) Encontre uma transformação para a resposta que permita linearizar a relação entre as variáveis;
- c) Ajuste os quatro modelos sugeridos em sala de aula (nulo, retas coincidentes, retas paralelas e retas concorrentes). Selecione um deles com base na análise dos testes F para modelos encaixados;
- d) Para o modelo selecionado realize o diagnóstico do ajuste.

Exercício 5- *Ankylosing spondylitis* é uma forma crônica de artrite. Um estudo foi conduzido para determinar se alongamento diário do quadril melhora a mobilidade dos portadores. Os dados estão disponíveis na base `hips`. O ângulo de flexão do quadril antes do estudo é um preditor para o ângulo de flexão após o estudo (resposta).

- a) Apresente o gráfico de dispersão separando os grupos tratado e controle;
- b) Ajuste o modelo de regressão e determine se há efeito de tratamento;
- c) Calcule as diferenças entre as flexões antes e depois e teste se essa diferença varia entre os grupos tratamento e controle. Compare os resultados desta análise aos gerados pelo modelo ajustado anteriormente;

- d) Verifique se há outliers. Explique porque podemos remover os três casos com `fbef` menor que 90. Ajuste um modelo de regressão apropriado e verifique se há efeito de tratamento;
- e) Qual é a estimativa do efeito de tratamento? Apresente um IC 95%;
- f) Cada indivíduo teve as medidas de flexão tomadas para cada uma das pernas. Desta forma, temos duas observações por indivíduo na base. Explique que tipo de dificuldade isso pode gerar na análise quanto às suposições do modelo;
- g) Calcule o ângulo de flexão médio para cada sujeito e repita a análise com a base de dados reduzida. Indique eventuais diferenças nos resultados.

Exercício 6- Ajuste um bom modelo de regressão para a base de dados `uswages` com `wages` como resposta e todas as demais variáveis como explicativas.

Exercício 7- Use a base de dados `cars` com `distance` como resposta e `speed` como variável explicativa.

- a) Faça o gráfico de `distance` vs `speed`;
- b) Ajuste o modelo de regressão linear simples para o par de variáveis e acrescente a reta ajustada ao gráfico;
- c) Ajuste o modelo de regressão com termo quadrático. Reresente o modelo ajustado no gráfico;
- d) Agora, use `sqrt(dist)` como resposta e ajuste novamente o modelo. Represente também o ajuste desse modelo no gráfico;
- e) Usando a função `loess`, obtenha também a regressão não paramétrica, ajustada por polinômios locais, para `distance` em função de `speed`. Compare o ajuste deste modelo aos obtidos previamente.

Exercício 8- A base de dados `aatemp` apresenta as temperaturas médias anuais (em °F) num período de 150 anos na cidade de Ann Arbor, Michigan.

- a) Faça o gráfico da temperatura média vs ano. Na sua opinião, parece haver uma variação linear na temperatura?
- b) Ajuste um polinômio de ordem 10 e, usando o método backward, reduza o máximo possível o grau do modelo. Represente o modelo ajustado sobre o gráfico obtido no item anterior. Use o modelo para prever a temperatura média em 2020;

Exercício 9- O nível de carbonatação de certa bebida é afetado pela temperatura do produto e pela pressão no processo de enchimento. Doze observações foram obtidas e os dados resultantes são apresentados na sequência:

Carbonatacao	Temperatura	Pressao
2.60	31.0	21.0
2.40	31.0	21.0
17.32	31.5	24.0
15.60	31.5	24.0
16.12	31.5	24.0
5.36	30.5	22.0
6.19	31.5	22.0
10.17	30.5	23.0
2.62	31.0	21.5
2.98	30.5	21.5
6.92	31.0	22.5
7.06	30.5	22.5

- a) Ajuste o modelo polinomial de segunda ordem aos dados;

- b) Teste a significância da regressão;
- c) O termo de interação contribui significativamente para o ajuste?
- d) Os termos de segunda ordem contribuem significativamente para o ajuste?
- e) Selecione o modelo mais simples que se ajusta bem aos dados. Apresente o gráfico da superfície de resposta.