

### Observações

- Sempre apresente as estimativas pontuais e intervalares, bem como os testes de nulidade individuais dos parâmetros de cada modelo ajustado.
  - Sempre realize análises de diagnóstico apropriadas (desvio, análise residual e análise preditiva). Pelo menos o gráfico de envelopes para os RDC's e uma análise preditiva, tem de ser apresentadas (com exceção da Questão 1). Caso não se utilize outras técnicas, deve-se apresentar as devidas justificativas.
  - Mesmo que o modelo não se ajuste bem ao conjunto de dados, continue a análise com ele, fazendo as devidas ressalvas (de que o modelo não está bem ajustado e de que alguma alternativa teria de ser considerada, sugerindo ao menos uma alternativa).
  - Sempre deixe o intercepto no modelo, ainda que ele não seja significativo.
  - Apresentar (numérica e/ou graficamente) as estimativas (pontuais e intervalares) das quantidades de interesse (médias e proporções).
  - Para se testar hipóteses de interesse, utilize quantos testes (análise do desvio, teste da RV, teste  $C\beta = M$  etc) julgar necessário, apresentando sempre as devidas justificativas, a menos que se solicite testes específicos.
  - Utilize as técnicas de seleção de modelo que julgar conveniente, apresentando sempre as devidas justificativas, a menos que se solicite técnicas específicas.
  - Sempre apresente as interpretações dos resultados.
  - As metodologias utilizadas (estimação, resíduos etc) não precisam ser explicadas, somente comentadas com as respectivas referências. Entretanto, todos os modelos utilizados devem constar no relatório, com as respectivas interpretações dos parâmetros.
1. Os dados constantes no arquivo WitzelGrandeSinger2000.xls são oriundos de um estudo realizado na Faculdade de Odontologia da USP, em que cada um de três avaliadores (A,B,C) classificou 72 molares selados com diferentes materiais, segundo o nível de microinfiltração, numa escala que varia de 0 (=sem microinfiltração) até 4 (=microinfiltração intensa). Ou seja, temos um total de  $72 \times 3 = 216$  dentes, sendo, também, 72 dentes para cada tipo de material. Considere então, as quantidades totais de dentes selados por cada

tipo de material (Allbond, Optibond e Scotchbond) foram fixadas. Mais detalhes sobre o estudo podem ser encontrados em Witzel, Grande & Singer (2000). Assim, temos 72 observações por tipo de material. O interesse principal é classificar os tipos de selante quanto ao nível de microinfiltração (quanto menor o valor, melhor o desempenho do selante). Responda os itens:

- a) Gere uma tabela de contingência (nível de microinfiltração  $\times$  material) apropriada para responder à pergunta de interesse, lembrando da suposição dos totais fixados, escrevendo o modelo probabilístico gerador (e seu nome). Faça uma análise descritiva, apresentando as estimativas pontuais e intervalares de cada probabilidade associada à tabela de contingência definida.
  - b) Defina, como medida de desempenho de cada tipo de material, uma média dos escores, ponderada pela probabilidade de classificação em cada categoria. Teste hipóteses do tipo  $\mathbf{B}\boldsymbol{\pi} = \mathbf{D}$ , para comparar os selantes através da medida de desempenho definida, classificando-os com relação aos seus desempenhos.
  - c) De acordo com o resultados dos testes feitos no item b), ajuste um modelo do tipo  $\mathbf{A}\boldsymbol{\pi} = \mathbf{X}\boldsymbol{\beta}$ , para estimar, somente, as medidas de desempenho definidas no item b), para cada tipo de selante, apresentando estimativas pontuais e intervalares.
2. Considere o conjunto de dados constante no arquivo (meninas.txt) que trata sobre a idade de ocorrência de menarca de garotas de Varsóvia. Tem-se o interesse em saber como a idade impacta na ocorrência de menarca. Em cada coluna apresentam-se, respectivamente, o número de garotas que apresentaram menstruação, o número de garotas entrevistadas e a idade média do grupo. Para analisar os dados, considere o seguinte modelo de regressão logística:  $Y_i \stackrel{ind.}{\sim} \text{Binomial}(m_i, p_i)$ ,  $m_i$  : é o número de garotas entrevistadas com idade  $i$ ,  $Y_i$  : é o número de meninas que já apresentaram menarca com idade  $i$  e  $\text{logit}(p_i) = \beta_0 + \beta_1(x_i - \bar{x})$ ,  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ,  $n = 25$  e  $x_i$  : é a idade média do grupo  $i$ . Para maiores detalhes veja Paula (2013) página 241. Responda os itens:
- a) Realize uma análise descritiva apropriada (gráficos, medidas resumo etc).
  - b) Ajuste o modelo de regressão proposto (verificando se é possível reduzi-lo), sempre realizando-se as análises de diagnóstico necessárias.
  - c) Com o modelo selecionado no item b), prediga (pontual e intervalarmente) as proporções de meninas a apresentar menarca, em cada uma das idades observadas. Apresente os resultados num gráfico.
  - d) Apresente as estimativas pontuais e intervalares, segundo o modelo selecionado no item b), das idades para que 25%, 50% e 99% das meninas já tenham apresentado menarca.

3. O arquivo `quine.dat` apresenta um banco de dados relativo à uma pesquisa realizada na Austrália que teve, entre outros objetivos, comparar a assiduidade de estudantes (representada pelo número de faltas durante um certo período), considerando diversos fatores. Aqui, temos interesse em apenas dois deles: etnia (A: aborígine e N: não aborígine) e desempenho escolar (0: insuficiente, 1: suficiente). Na primeira coluna consta a etnia (A:aborígine, N:não aborígine), na quarta o desempenho (SL:insuficiente, AL:suficiente), enquanto que na última consta o número de faltas. Para maiores detalhes veja Paula (2013) página 312. Considere como grupo de referência os estudantes aborígenes com desempenho insuficiente.
- Realize uma análise descritiva apropriada (histogramas, boxplots, gráficos de perfis, medidas resumo etc).
  - Ajuste um modelo de regressão de Poisson log-linear para analisar o efeitos dos dois fatores e da interação no número de faltas, sob a parametrização casela de referência, (verificando, adequadamente, se é possível reduzi-lo), sempre realizando-se as análises de diagnóstico necessárias.
  - Com o modelo selecionado no item b), prediga (pontual e intervalarmente) os números de falta de cada um dos quatro grupos. Apresente os resultados numa tabela e num gráfico (perfis médios ajustados).