

ME 714 A - Análise de dados discretos

Primeiro semestre de 2017

Trabalho (Parte I). Data da entrega: até o dia 05/06/2017, às 16h00.

### 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 (com exceção da Questão 1) e uma análise preditiva, têm de ser apresentados. 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.

## Questões

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, que 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  $B\pi = D$ , 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  $A\pi = X\beta$ , para estimar, somente, as medidas de desempenho definidas no item b), para cada tipo de selante, apresentando estimativas pontuais e intervalares.
2. Os dados constantes na tabela abaixo são provenientes de um experimento desenvolvido para avaliar a germinação de um determinado tipo de semente segundo três condições experimentais: nível da temperatura (21°C, 42°C e 62°C); nível da umidade (baixo, médio e alto) e temperatura da germinação (11°C e 21°C). A tabela abaixo apresenta o número de sementes que germinaram após cinco dias para cada 100 sementes submetidas a cada condição experimental. Assuma que as contagens de cada casela ( $Y_{ijk}$ ) são tais que  $Y_{ijk} \overset{ind.}{\sim} \text{binomial}(100, p_{ijk})$ . Esses dados também encontram-se no arquivo “sementes.dat”. Note que também é importante verificar quais interações são significativas, bem como os efeitos principais e também fornecer a combinação entre os níveis dos fatores que gera a melhor taxa de germinação (quanto maior, melhor).

Responda os itens:

Temperatura da Germinação	Nível da Umidade	Nível da Umidade		
		21°C	42°C	62°C
11°C	baixo	98	96	62
11°C	médio	94	79	3
11°C	alto	92	41	1
21°C	baixo	94	93	65
21°C	médio	94	71	2
21°C	alto	91	30	1

- Realize uma análise descritiva apropriada (gráficos, medidas resumo etc). Apresente gráficos dos perfis das proporções observadas, apropriados.
- Ajuste um modelo de regressão logística apropriado (verificando se é possível reduzi-lo), sempre realizando-se as análises de diagnóstico necessárias (mesmo nos modelos intermediários).
- Com o modelo selecionado no item b), prediga (pontual e intervalarmente) as proporções de germinação observadas e previstas. Apresente os resultados num gráfico.