



1 Modelos Lineares Generalizados

Divulgação: 23/04/2018

1. Estudar o Capítulo 3 do livro texto¹.
2. Representar a distribuição Binomial(n, π) na forma da família exponencial. Obter a média e variância da distribuição, utilizando os resultados obtidos em sala de aula.
3. Representar a distribuição Poisson(μ) na forma da família exponencial. Obter a média e variância da distribuição, utilizando os resultados obtidos em sala de aula.
4. Representar a distribuição Gama(α, β) na forma da família exponencial. Obter a média e variância da distribuição, utilizando os resultados obtidos em sala de aula.
5. Representar a distribuição Exponencial(β) na forma da família exponencial. Obter a média e variância da distribuição, utilizando os resultados obtidos em sala de aula.
6. Obter média e variância de uma família exponencial na forma

$$f(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\}.$$

Utilizando a forma acima, confirmar os resultados obtidos para média e variância nas Questões 2 e 3.

Problemas do Livro Texto:

7. Problema 3.1.
8. Problema 3.3.
9. Problema 3.5.
10. Problema 3.9.
11. Problema 3.11.
12. Problema 3.22 (Justifique cada resposta).
13. Obter os EMVs e intervalos de confiança para um modelo logístico com duas variáveis explicativas, i.e.,
$$\text{logito}(\pi_i) = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2}$$
14. Para o problema das abelhas, obter os resultados apresentados utilizando o R.

¹A. Agresti, Introduction to Categorical Data Analysis, 2007.



2 Regressão Logística

Divulgação: 07/05/2018

15. Utilizar o R para refazer os os exemplos de Regressão Logística apresentados em aula.

Considere o livro texto *Introduction to Categorical Data Analysis*

16. Estudar o Capítulo 5 - livro texto.
17. Problema 4.7 - livro texto.
18. Problema 4.9 - livro texto.
19. Problema 4.15 - livro texto.
20. A tabela abaixo é baseada em estudo epidemiológico com 2484 sujeitos para avaliar se ronco (escores 0,2,4,5) é um possível fator de risco para doença do coração.

Ronco	Doença do Coração	
	Sim	Não
Nunca (0)	24	1355
Ocasional (2)	35	603
Quase toda noite (4)	21	192
Toda noite (5)	30	224

Utilize o SAS e/ou R para responder as questões seguintes:

- (a) Ajuste um modelo de regressão logística e obtenha as estimativas dos parâmetros.
(b) Calcule as probabilidades amostrais de casos de doenças de coração para cada nível de ronco. Obtenha as probabilidades estimadas e indique se o modelo parece ajustar bem os valores amostrados.
(c) Calcule e interprete o nível efetivo mediano $EL_{50} = -\hat{\alpha}/\hat{\beta}$.
(d) Calcule o acréscimo na chance de infarto para cada unidade acrescida no ronco.
(e) Faça um teste para qualidade do ajuste. Indique as hipóteses, estatística do teste, p-valor e conclusão.
(f) Se alterarmos a codificação da variável ronco no programa SAS, as estimativas dos parâmetros irão ser diferentes? Justifique?



21. Considere o problema *Horseshoe Crab* visto em aula. Sejam largura (L), cor (C) e espinha (E) as variáveis preditoras e suponha que foi feito um procedimento de eliminação *backward*. O procedimento foi iniciado com o modelo mais complexo (modelo saturado: $C * E * L$). Este modelo utilizou efeitos simples, interações de dois fatores e de três fatores. A tabela abaixo apresenta o ajuste para este modelo e outros modelos mais simples, na forma hierárquica.

Modelo	Preditor	Deviance	gl	AIC
1	$C * E * L$	170,44	152	212,4
2	$C * E + C * L + E * L$	173,68	155	209,7
3A	$C * E + E * L$	177,34	158	207,3
3B	$C * L + E * L$	181,56	161	205,6
3C	$C * E + C * L$	173,69	157	205,7
4A	$E + C * L$	181,64	163	201,6
4B	$L + C * E$	177,61	160	203,6
5	$C + E + L$	186,61	166	200,6

- Faça um teste de razão de verossimilhança comparando o modelo saturado com o modelo mais simples que remove apenas a interação de 3 fatores. O teste sugere que a interação de 3 fatores deve ser removida do modelo?
- Se queremos remover um termo no estágio seguinte, explique porque selecionaríamos o modelo $C * E + C * L$.
- Considerando o modelo 3C, se você tiver de eliminar mais termos e escolher entre os modelos 4A ou 4B, qual modelo seria o escolhido?
- É razoável simplificar o modelo ainda mais? Justifique.
- Dos modelos considerados na tabela, qual seria o escolhido pelo critério AIC?