

UNIVERSIDADE FEDERAL FLUMINENSE

Programa de Mestrado e Doutorado em Engenharia de Produção

Disciplina: Estatística Multivariada – 2016/2

Lista: Logistic Binary Regression

Professor: Valdecy Pereira, D. Sc.

email: valdecy.pereira@gmail.com

1) O Data Set LR-01 é um questionário feito por uma fabricante de carros. O questionário levantou a percepção de atributos que homens (GENDER = 1) e mulheres (GENDER = 2) consideram mais importantes em um carro. No total 93 pessoas responderam o questionário. Os atributos foram medidos em uma escala Likert de 5 pontos (1 = Nenhuma Importância; 2 = Pouca Importância; 3 = Alguma Importância; 4 = Muita Importância; 5 = Extremamente Importante) e são listados a seguir:

VARIÁVEL	DESCRIÇÃO
P01	Qualidade dos Acessórios
P02	Facilidade de Dirigir
P03	Conforto
P04	Visibilidade
P05	Segurança na Estrada
P06	Alta Velocidade
P07	Poder de Aceleração
P08	Absorção de Impacto
P09	Transmissão
P10	Design
P11	Pintura
P12	Ausência de Barulho
P13	Ar Condicionado
P14	Economia de Combustível
P15	Economia nos Reparos
P16	Valor Estimado do Carro Usado
P17	Espaço
P18	Atendimento da Garantia
P19	Disponibilidade das Peças
P20	Qualidade da Assistência Técnica
P21	Preço das Peças
P22	Preço dos Reparos

a) Determine um modelo, em que todos os coeficientes de regressão sejam significativos.

- b) Verifique se o modelo final é adequado pelo teste de razão de verossimilhança.
- c) Analise os pseudo r-quadrado (Cox & Snell e Nagelkerke) do modelo final.
- d) Interprete as alterações médias (aumento ou diminuição da chance) do modelo final.
- e) Ache a tabela de classificação para um *cutoff* de 0,5.
- f) Analise a curva ROC.
- h) O modelo com as variáveis SAVING ON FUELS, CONFORT, POWER Of ACCELERATION e SAVING ON REPAIRS é superior do que o modelo final? Explique.

2) O Data Set LR-02 são dados do olho direito de 618 crianças que podem não ter miopia (MYOPIC = 0) ou ter miopia (MYOPIC = 1). A miopia é um distúrbio visual que acarreta uma focalização da imagem antes desta chegar à retina. Uma pessoa míope consegue ver objetos próximos com nitidez, mas os distantes são visualizados como se estivessem embaçados (desfocados). Para uma visão mais apurada, o ponto focal dos raios luminosos deve convergir para uma área próxima aos receptores de luminosidade (localizados na retina). No olho míope, a imagem concentra-se na frente da retina, de modo que a imagem resultante sobre a própria retina é borrada. A imagem borrada cria problemas para uma variedade de tarefas de visualização à distância e requer o uso de óculos ou lentes de contato para corrigir o problema. Os fatores de risco para o desenvolvimento da miopia têm sido debatidos por um longo tempo e incluem fatores genéticos (por exemplo, história familiar de miopia) e da quantidade e tipo de atividade visual realizada (por exemplo, estudar, ler, assistir TV, usar o computador/jogar vídeo game, e esportes / atividades ao ar livre). Há fortes evidências de que ter pais míopes aumenta a chance de uma criança se tornar míope, e as provas mais fracas são que certos tipos de atividades visuais também podem aumentar a chance de uma criança se tornar míope. Os dados são:

VARIÁVEL	DESCRIÇÃO
AGE	Idade (em anos)
GENDER	Gênero. 0 = Masculino ; 1 = Feminino.
SPHEQ	Refração Esférica Equivalente (m^{-1})
AL	Comprimento Axial (mm)
ACD	Profundidade da Câmara Anterior (mm)
VCD	Profundidade da Câmara Vítrea (mm)
SPORTHR	Quantidade de Tempo Dedicado para Praticar Atividades ao Ar Livre (horas/semana).
READHR	Quantidade de Tempo Dedicado a Leitura (horas/semana).
COMPHR	Quantidade de Tempo Dedicado a Computadores e/ou Vídeo Games (horas/semana).
STUDYHR	Quantidade de Tempo Dedicado para Estudos (horas/semana).
TVHR	Quantidade de Tempo Dedicado para Assistir TV (horas/semana).
MOTHERMY	Mãe com Miopia. 0 = Não ; 1 = Sim.

FATHERMY	Pai com Miopia. 0 = Não ; 1 = Sim.
----------	---------------------------------------

UTILIZE O MÉTODO FOWARD (CONDICIONAL)

- Determine um modelo, em que todos os coeficientes de regressão sejam significativos.
- Verifique se o modelo final é adequado pelo teste de razão de verossimilhança.
- Analise os pseudo r-quadrado (Cox & Snell e Nagelkerke) do modelo final.
- Interprete as alterações médias (aumento ou diminuição da chance) do modelo final.
- Ache a tabela de classificação para um *cutoff* de 0,5.
- Ache e analise a curva ROC.
- A amostra permite que a evidência científica pode ser fortalecida? Explique.

3) O Data Set LR-03 representa o desempenho de 200 alunos pertencentes a diferentes tipos de escola. Um pesquisador deseja saber se com os dados sócio-econômicos e com os dados do desempenho escolar é possível prever o gênero do aluno [GENDER = 0 (Masculino) e GENDER = 1 (Feminino)] de modo que seja melhor do que 50%.

VARIÁVEL	DESCRIÇÃO
RACE	Etnia. 1 = Hispânico ; 2 = Asiático ; 3 = Afro-Americano e 4 = Branco.
SES	Classe Social. 1 = Baixa ; 2 = Média ; 3 = Alta.
SCHOOL	Escola. 1 = Pública ; 2 = Privada.
PROGRAM	Programa. 1 = Geral ; 2 = Acadêmica ; 3 = Técnica
READING	Nota Leitura.
WRITE	Nota Escrita.
MATH	Nota Matemática.
SCIENCE	Nota Ciências da Natureza.
SOCIAL	Nota Ciências Sociais

- Determine um modelo, em que todos os coeficientes de regressão sejam significativos.
- Verifique se o modelo final é adequado pelo teste de razão de verossimilhança.
- Analise os pseudo r-quadrado (Cox & Snell e Nagelkerke) do modelo final.
- Interprete as alterações médias (aumento ou diminuição da chance) do modelo final.
- Ache a tabela de classificação para um *cutoff* de 0,5.
- Ache e analise a curva ROC.

g) O modelo é adequado? Explique.

4) O Data Set LR-04 representa uma amostra de clientes com contratos fechados (com intervalo de confiança de 95%) composta por 100 clientes adimplentes (SITUAÇÃO = 0) e 100 clientes inadimplentes (SITUAÇÃO = 1) da Instituição X, uma instituição de microcrédito. O microcrédito é a concessão de crédito de baixo valor a pequenos empreendedores informais e microempresas sem acesso ao sistema bancário tradicional. Assim, o microcrédito é uma alternativa de acesso ao crédito para os microempreendedores que não possuem garantias reais a serem oferecidas quanto ao pagamento do empréstimo, pois o seu negócio possui características aquém daquelas exigidas pelos programas oficiais de financiamento. Devido a tais condições, estes microempreendedores buscam crédito, principalmente, em fontes informais, como por exemplo, agiotas. A Instituição X direciona a oferta do microcrédito de acordo com o perfil e a necessidade dos tomadores de crédito, e tem como característica o acompanhamento ao crédito concedido, através de profissionais especializados que vão até o local onde o cliente exerce sua atividade produtiva, para avaliar as necessidades e as condições de seu empreendimento, bem como as possibilidades de pagamento. Após a liberação do crédito, esses profissionais passam a acompanhar a evolução do negócio. A Instituição X deseja migrar de uma análise de crédito subjetiva (que necessita das percepções de seus profissionais especializados), para uma análise de crédito quantitativa. Para realizar essa migração, você foi contratado para desenvolver um modelo de concessão de crédito que os ajude a separar bons pagadores de maus pagadores. As seguintes variáveis estão disponíveis na base de dados da instituição:

VARIÁVEL	DESCRIÇÃO	MEDIDA
GÊNERO	Gênero do Cliente	(0 = Feminino; 1 = Masculino)
ESTCIVIL	Estado Civil do Cliente	(0 = Casado; 1 = Solteiro)
IDADE	Idade do Cliente	Em Anos
RLF	Receita Líquida Familiar	Em Reais
RLN	Receita Líquida do Negócio	Em Reais
NDEPEN	Número de Dependentes Financeiros do Cliente	Quantidade Discreta
NPARCEL	Número de Parcelas do Último Empréstimo	Quantidade Discreta
TEMPFUN	Tempo de Funcionamento	Em Anos
PERFILAC	Agente de crédito responsável pelo empréstimo que é incumbido de analisar e acompanhar o empréstimo em questão. O Perfil A diz respeito a agentes mais jovens, com nível superior e com novas técnicas de acompanhamento. O Perfil B diz respeito a agentes mais velhos, sem nível superior e com mais experiência.	(0 = Perfil A; 1 = Perfil B)
PERCENDIV	Percentual de Endividamento: $\frac{\text{Valor da Parcela do Empréstimo}}{\text{Receita Líquida do Negócio}}$	Em %

ATRASOSCA	Atrasos Maiores do que 60 dias em Créditos Anteriores	(0 = Não Possui Atraso; 1 = Possui Atraso)
-----------	--	---

- a) Determine um modelo, em que todos os coeficientes de regressão sejam significativos.
- b) Verifique se o modelo final é adequado pelo teste de razão de verossimilhança.
- c) Analise os pseudo r-quadrado (Cox & Snell e Nagelkerke) do modelo final.
- d) Interprete as alterações médias (aumento ou diminuição da chance) do modelo final.
- e) Ache a tabela de classificação para um *cutoff* de 0,5.
- f) Ache e analise a curva ROC.
- g) O modelo pode ser apresentado a Instituição X?