

UNIVERSIDADE FEDERAL FLUMINENSE

Programa de Mestrado e Doutorado em Engenharia de Produção

Disciplina: Estatística Multivariada – 2016/2

Lista: Discriminant Analysis

Professor: Valdecy Pereira, D. Sc.

email: valdecy.pereira@gmail.com

1) O Data Set DA-01 é um conjunto de dados apresentado pelo pesquisador **Sir Ronald Aylmer Fisher** em 1936 no seu artigo seminal de apresentação e demonstração da sua recém criada técnica a **Análise Discriminante**. Os dados representam as medidas de 150 plantas divididas em três espécies de flores: *Iris Setosa* (categoria 1), *Iris Versicolor* (categoria 2) e *Iris Virginica* (categoria 3). Para cada planta foram feitas as seguintes medidas em cm: comprimento e largura das sépalas (SEPAL L e SEPAL W) e comprimento e largura das pétalas (PETAL L e PETAL W).

a) Qual é a acurácia do modelo?

2) O Data Set DA-02 é um questionário feito por uma fabricante de carros. O questionário levantou a percepção de atributos que homens (categoria 1) e mulheres (categoria 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

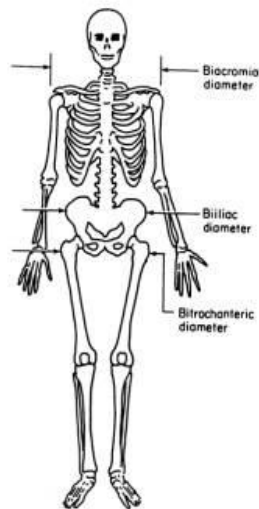
P17	Espaço
P18	Atendimento da Garantia
P19	Disponibilidade das Peças
P20	Qualidade da Assistência Técnica
P21	Preço das Peças
P22	Preço dos Reparos

a) Um vendedor de carros diz, pela sua experiência, que não é possível perceber diferença entre os grupos. Ele afirma que tanto homens quanto mulheres possuem percepções idênticas quanto a duas questões mais perguntadas pelos compradores - o atendimento da garantia (P18) e o Design (P10). O vendedor está correto, ou seja, os grupos não podem ser discriminados?

3) O Data Set DA-03 apresenta um levantamento de dados feito por um cientista forense - chamado Ambroise Paré - que devido a sua experiência de campo, alega que é possível inferir se um esqueleto pertence a um homem ou a uma mulher, baseado em certas medidas ósseas incomuns e que não são considerados pela ciência forense clássica (para medidas ósseas, é indicado a utilização dos ossos da pelvis pois permitem em média 90% de acurácia). Para provar sua teoria Ambroise selecionou uma amostra aleatória de 507 indivíduos vivos (tirados da população local e que permite um intervalo de confiança de 95%), sendo 260 mulheres (categoria 1) e 247 homens (categoria 2) entre 18 e 67 anos. Ambroise deseja confirmar se a sua teoria está correta e para isso coletou outras variáveis desses indivíduos, sendo assim possível comparar a efetividade das medidas ósseas em relação aos demais tipos de medidas. As variáveis coletadas foram:

VARIÁVEL	DESCRIÇÃO	MEDIDA
BIACROMIAL D	Diâmetro Bi-acrômio	Óssea (cm)
BIILIAC D	Diâmetro Bi-ilíaco	Óssea (cm) – Clássica
BITROCHANTERIC D	Diâmetro Bi-trocânter	Óssea (cm)
CHEST DEPTH	Profundidade Peitoral	Óssea (cm)
CHEST D	Diâmetro Peitoral	Óssea (cm)
ELBOW D	Diâmetro dos Cotovelos	Óssea (cm)
WRIST D	Diâmetro dos Pulsos	Óssea (cm)
KNEE D	Diâmetro dos Joelhos	Óssea (cm)
ANKLE D	Diâmetro dos Tornozelos	Óssea (cm)
SHOULDER G	Circunferência dos Ombros	Sistema Tegumentar (cm)
CHEST G	Circunferência Peitoral	Sistema Tegumentar (cm)
WAIST G	Circunferência da Cintura	Sistema Tegumentar (cm)
HIP G	Circunferência das Ancas	Sistema Tegumentar (cm)
THIGH G	Circunferência das Coxas	Sistema Tegumentar (cm)

BICEPS G	Circunferência dos Bíceps	Sistema Tegumentar (cm)
FOREARM G	Circunferência dos Antebraços	Sistema Tegumentar (cm)
KNEE G	Circunferência dos Joelhos	Sistema Tegumentar (cm)
CALF G	Circunferência das Panturrilhas	Sistema Tegumentar (cm)
ANKLE G	Circunferência dos Tornozelos	Sistema Tegumentar (cm)
WRIST G	Circunferência dos Pulsos	Sistema Tegumentar (cm)
AGE	Idade	Individual (anos)
WEIGHT	Peso	Individual (kg)
HEIGHT	Altura	Individual (cm)



Ambroise o contratou para que pudesse ajudá-lo a comprovar ou rejeitar a sua hipótese. Geralmente nos trabalhos de campo envolvendo crimes, nem todas as ossadas encontradas estão completas, surgindo então a necessidade de se obter alternativas significantes e que sejam tão boas quanto as medidas feitas em indivíduos vivos.

- Qual é a acurácia do modelo se todas as variáveis forem utilizadas?
- Qual é a acurácia do modelo se apenas as medidas ósseas forem utilizadas?
- Qual é a acurácia do modelo se apenas as medidas do sistema tegumentar forem utilizadas?
- Qual é a acurácia do modelo se apenas as medidas individuais forem utilizadas?
- A suposição de Ambroise pode ser confirmada estatisticamente?

4) O Data Set DA-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)

UTILIZE O MÉTODO STEPWISE (LAMBDA DE WILKS)

a) Qual deve ser a estratégia de concessão de microcrédito da Instituição X com base na sua modelagem?