11º FEPEG FÓRUM ENSINO • PESQUISA EXTENSÃO • GESTÃO
UNIVERSIDADE, SOCIEDADE E POLÍTICAS PÚBLICAS
ISSN: 1806-549X
Realização: Unimontes Universidade Estadual de Montes Claros
Apoio: FAPEMIG FADENOR

UM BREVE ESTUDO SOBRE ANALISE DESCRIMINANTE

Autores: JANAINA NEVES SOUZA, JANAINA NEVES SOUZA

Introdução

A análise discriminante é uma técnica da estatística multivariada utilizada para discriminar e classificar objetos. Segundo KHATTREE & NAIK (2000) é uma técnica da estatística multivariada que estuda a separação de objetos de uma população em duas ou mais classes. A discriminação ou separação é a primeira etapa, sendo a parte exploratória da análise e consistem em se procurar características capazes de serem utilizadas para alocar objetos em diferentes grupos previamente definidos. A classificação ou alocação pode ser definida como um conjunto de regras que serão usadas para alocar novos objetos (JOHNSON & WICHERN,1999). Contudo, a função que separa objetos pode também servir para alocar, e, o inverso,regras que alocam objetos podem ser usadas para separar. Normalmente, discriminação e classificação se sobrepõem na análise, e a distinção entre separação e alocação é confusa. Segundo REGAZZI (2000) o problema da discriminação entre dois ou mais grupos, visando posterior classificação, foi inicialmente abordado por Fisher (1936). Consiste em obter funções matemáticas capazes de classificar um indivíduo X (uma observação X) em uma de várias populações ?i, (i=1, 2,..., g), com base em medidas de um número p de características, buscando minimizar a probabilidade de má classificação, isto é, minimizar a probabilidade de classificar erroneamente um indivíduo em uma população ?i, quando realmente pertence a população ? j, (i?j) i,j=1, 2, ..., g.

Desenvolvimento

Quando o pesquisador estiver interessado em estudar somente dois grupos de variáveis dependentes, a técnica é chamada de Análise Discriminante Simples. No entanto, em muitos casos, há o interesse na discriminação entre mais de dois grupos, sendo a técnica, assim, denominada de Análise Discriminante Múltipla (MDA).

Os objetivos principais desses dois tipos de análises são parecidos: (i) identificar as variáveis que melhor discriminam dois ou mais grupos; (ii) utilizar estas variáveis para desenvolver funções discriminantes que representem as diferenças entre os grupos; (iii) fazer uso das funções discriminantes para o desenvolvimento de regras de classificação de futuras observações nos grupos.

1. 1. Modelagem da Análise Discriminante
    1. Construção da regra de classificação em duas populações

Inicialmente temos duas populações e um conjunto de observações independentes de cada população. Se a distribuição de probabilidade das características medidas dos elementos amostrais de cada população for conhecida será possível utilizar o principio da máxima verossimilhança para construir uma regra que minimize a chance de se classificar um elemento amostral incorretamente.

Tendo os dados é possível calcular a razão entre duas distribuições de probabilidades, chamada de razão de verossimilhança entre duas populações, definidas por:

ISSN: 1806-549X
Realização: Unimontes Universidade Estadual de Montes Claros · Secretaria de Desenvolvimento Econômico, Ciência, Tecnologia e Ensino Superior · Minas Gerais Governo de Todos · PIBID Unimontes Programa Institucional de Bolsa de Iniciação à Docência
Apoio: FAPEMIG · FADENOR

Image not found or type unknown

Que no caso da distribuição normal, torna-se:

Image not found or type unknown

1. 1. 1. Construção da regra de classificação em varias populações

Para o caso de mais de duas populações envolvidas temos:n1

Seja Image not found or type unknown a função densidade da população i, i=1,2,...,n. e deseja-se elaborar a regra que tenha mínimas probabilidades de erro. Neste caso é dado:

Para um vetor de observações Image not found or type unknown fixo calcula-se o valor da densidade Image not found or type unknown para cada população, i=1,2,...,n, sendo o elemento amostral classificado na população que tiver o maior valor de Image not found or type unknown ou seja, classifica-se o elemento amostral naquela população k, tal que :

Image not found or type unknown

No caso particular em que o vetor aleatório X em cada população tem distribuição normal p variada, esta regra é equivalente a classificar o elemento com vetor observado x naquela população k.

A matriz de covariância amostral pode ser definida por:

Image not found or type unknown

Métodos

1.1.3 **Método forward:** Esse procedimento parte da suposição de que não há variável no modelo, apenas o intercepto. A ideia do método é adicionar uma variável de cada vez. A primeira variável selecionada é aquela com maior correlação com a resposta. Supondo que essa variável seja Image not found or type unknown, calculamos a estatística F para testar se ela realmente é significativa para o modelo. A variável entra no modelo se a estatística F for maior do que o ponto crítico, chamado de Fi ou F para entrada. Notemos que Fi é calculado para um dado ? crítico.

1.1.4 **Método backward:** Enquanto o método Forward começa sem nenhuma variável no modelo e adiciona variáveis a cada passo, o método Backward faz o caminho oposto; incorpora inicialmente todas as variáveis e depois, por etapas, cada uma pode ser ou não eliminada.

A decisão de retirada da variável é tomada baseando-se em testes F parciais, que são calculados para cada variável como se ela fosse à última a entrar no modelo.

O menor valor das estatísticas F parciais calculadas é então comparado com o F crítico, calculado para um dado valor ? crítico. Se o menor valor encontrado for menor do que F critico, elimina-se do modelo a covariável responsável pelo menor valor da estatística F parcial.

1.1.5 **Método stepwise:** Stepwise é uma modificação da seleção Forward em que cada passo todas as variáveis do modelo são previamente verificadas pelas suas estatísticas F parciais. Uma variável adicionada no modelo no passo anterior pode ser redundante para o modelo por causa do seu relacionamento com as outras variáveis e se sua estatística F parcial for menor que F critico , ela é removida do modelo.

Assim, a regressão Stepwise requer dois valores de corte: Fi e F critico . Alguns autores preferem escolher Fi=Fc mas isso não é necessário. Se Fi< Fc: mais difícil remover que adicionar; se Fi>Fc : mais difícil adicionar que remover.

ISSN: 1806-549X

EXEMPLOS DE APLICAÇÃO

1- **Ecologia:** algumas espécies de insetos são muito similares. Neste caso a analise discriminante pode auxiliar na classificação de insetos bastando que, para cada espécie candidata, se tenha informações sobre seu perfil geral em relação a algumas características morfológicas. A tradução destas medidas em função matemática fará com que a classificação do inseto possa ser mais precisa do que a mera inspeção visual. O mesmo se aplica a classificação de outros organismos ou plantas.

2- **Medicina:** A criação de mecanismos que possam identificar os fatores de riscos ou distinguir doenças que tenham similaridade relativa aos sintomas apresentados pelo paciente é de suma importância. Se, por exemplo, tivermos medidas laboratoriais, comportamentais e sociais de pacientes portadores e não portadores de uma doença ou vírus, será possível conhecer o perfil dos dois grupos de pacientes e construir uma regra que permitira a classificação de novos pacientes como prováveis ou não prováveis de terem a respectiva doença ou vírus, A analise discriminante permite também identificar variáveis que podem estar associadas com a doença e que poderão auxiliar o medico na sua decisão de como tratar o paciente.

3- **Finanças:** É importante para um banco identificar se uma pessoa que está pleitando um empréstimo bancário será inadimplente ou não.Assim o banco pode buscar informações em seus arquivos sobre algumas variáveis dos clientes que obtiveram empréstimo e honraram o pagamento de seus compromissos de acordo com o contrato firmado e daqueles com os quais isto não ocorreu.Com as informações uma regra matemática poderá ser elaborada para classificar e identificar os possíveis inadimplentes antes de conceder-lhe o empréstimo. Exemplo de variáveis: renda media mensal, profissão, numero de cartões de credito, estado civil, idade, numero de filhos, etc.

4- **Ensino:** Escolas usam processos de seleção para a escolha de candidatos a seus cursos de pós-graduação. Algumas variáveis são importantes, como histórico escolar, curriculum vitae, cartas de referencias, experiência profissional e outras. Com estas variáveis é possível construir uma regra de classificação que permita discriminar os estudantes com maior potencial para concluir o curso daqueles que tem menos potencial.

É certo que todo o processo de tomada de decisões traz consigo um possível erro de decisão, logo é preciso construir uma regra de classificação que minimize o numero de classificações incorretas, o erro de dizer que um elemento amostral pertence a uma população quando na verdade ele pertence à outra.

Conclusão/Conclusões/Considerações finais

ISSN: 1806-549X

A análise discriminante pode ser utilizada em conjunto com outras técnicas estatísticas multivariadas, como a analise de componentes principais e analise de agrupamento. De agrupamento (cluster). Se um pesquisador tiver p-variaveis-resposta, ele poderá reduzir a informação amostral para k componentes principais e utilizá-las como variáveis discriminantes para obter a regra de classificação. Para a aplicação da AD, é necessário quem a divisão dos elementos amostrais em n grupos tenha sido feita previamente, qualquer divisão exige o uso de critérios.

Agradecimentos

Professor Dr. Rômulo Barbosa Veloso

Referências bibliográficas

[1] Introdução à Estatística de Mário F. Triola, Editor: Livros Téc. e Cient. Editora

[2] JOHNSON, R. A.; WICHERN, D. W. Applied multivariate statistical analysis. 4th ed. Upper Saddle River, New Jersey: Prentice-Hall, 1999, 815 p.

[3] KHATTREE, R. & NAIK, D.N. Multivariate data reduction and discrimination with SAS software. Cary, NC, USA: SAS Institute Inc., 2000. 558 p

[4] FÁVERO, L. P.; BELFIORE, P.; SILVA, F. L.; CHAN, B. L. Análise de dados: modelagem multivariada para tomada de decisões. Rio de Janeiro: Elsevier, 2009

http://igce.rc.unesp.br/Home/Departamentos47/geologiaaplicada/9.discriminante.pdf acesso em 09 de setembro de 2017