MODELO FUZZY E MODELO DE REGRESSÃO LOGÍSTICA PARA TOMADA DE DECISÃO NA CLASSIFICAÇÃO DE OBESIDADE.

Diego Augusto Queijo[1];
Liciana Vaz de Arruda Silveira[2];
Lia Thieme Oikawa Zangirolani[3].

1) INTRODUÇÃO

A obesidade tem sido considerada uma epidemia global pela Organização Mundial da Saúde (OMS) devido ao aumento alarmante de sua prevalência em diversas partes do mundo[(1)]. Este aumento está distribuído em quase todas as raças e sexo, e atinge tanto crianças como adultos[(2,3)].

Obesidade é comumente definida como um excesso de gordura corporal, porém diante da dificuldade em mensurar tal gordura diretamente, esta tem sido definida como um excesso de peso mais do que um excesso de gordura corporal, que tem como desdobramento a ocorrência de doenças associadas e/ou prejuízos à saúde do indivíduo[(4)]. Atualmente, o excesso de peso é verificado por meio de um Índice de Massa Corpórea (IMC) de 25 até 29,9 Kg/m², e a obesidade de um IMC de 30 Kg/m² ou mais, de acordo com a OMS[(1)].

A identificação das causas da obesidade não é trivial e objetiva. Especialistas reconhecem que a obesidade é uma doença crônica, de difícil tratamento, denominada multifatorial, envolvendo em sua gênese diversos aspectos, entre eles: o consumo alimentar, aspectos ambientais, genéticos, psicossociais, entre outros.

O objetivo deste trabalho foi desenvolver dois modelos, um baseado em teoria dos conjuntos *Fuzzy* e outro no modelo de regressão logística, para tomada de decisão na classificação de obesidade.

1-Graduando do Curso de Física Médica – UNESP – Botucatu, SP.
2- Departamento de Bioestatística – IB – UNESP – Botucatu, SP.
3 -EPIGEO - FCM - UNICAMP – Campinas-SP.

2) METODOLOGIA

2.1) TEORIA DOS CONJUNTOS *FUZZY*

Nas ciências em geral, nos confrontamos freqüentemente com diversos problemas que envolvem incertezas e imprecisão, dos quais o processo de tomada de decisão acaba apoiando-se em conceitos vagos estranhos à lógica clássica e em parâmetros de natureza subjetiva(5).

A lógica *fuzzy* pode ser utilizada para representar a extensão da lógica clássica para uma mais flexível com objetivo de formalizar conceitos imprecisos e também onde se aplicam conjuntos *fuzzy* a diversas teorias e tecnologias para processar informações imprecisas, por exemplo, em processos de tomada de decisão.

Podemos dizer que a lógica *fuzzy* é uma ferramenta capaz de capturar informações vagas, em geral descritas em uma linguagem natural e convertê-las para um formato numérico.

Definição: seja U um conjunto (clássico); um subconjunto *fuzzy* F de U é caracterizado por uma função $\varphi_F : U \rightarrow [0,1]$, chamada função de pertinência do subconjunto *fuzzy* F. O índice F na função de pertinência é usado em analogia à função característica de subconjunto clássico(6).

Outros conceitos importantes de conjuntos *fuzzy* dizem respeito a cardinalidade e altura desses conjuntos. A cardinalidade de um conjunto é o número total de elementos no conjunto. Uma vez que os elementos podem pertencer parcialmente a um conjunto *fuzzy*, uma generalização natural da noção clássica de cardinalidade consiste em pesar cada elemento pelo seu grau de pertinência. Sendo assim, a cardinalidade de um conjunto *fuzzy* é definida por:

$$Card\ (A) = \sum_{x_i} \mu_A(x_i)$$

onde A é um conjunto *fuzzy* e os x_is são os elementos do conjunto Universo.

A altura de um conjunto *fuzzy* é o maior valor de pertinência da sua função de pertinência:

$$hgt(A) = \max_{x_i} \mu_A(x_i)$$

Os conjuntos *fuzzy* com altura igual a 1 são chamados normais e aqueles cuja altura é inferior a 1 são chamados subnormais. Na figura 1 tem-se um exemplo de função de pertinência triangular subnormal(7).

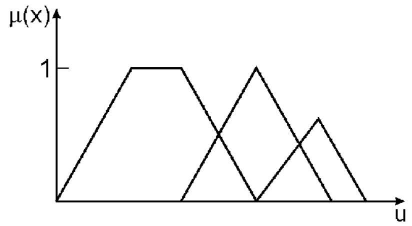

Figura 1: Exemplo de função de pertinência do tipo linear por partes

Uma variável lingüística *fuzzy* é uma variável cujo valor é expresso qualitativamente por um termo lingüístico (que fornece um conceito à variável) e quantitativamente por uma função de pertinência. De fato, uma variável lingüística é caracterizada por {n, T, X, m(n)} onde n é o nome da variável (por exemplo, Índice de Massa Corpórea, IMC), T é o conjunto de termos lingüísticos de n (baixo peso, estrófico, sobrepeso, obesidade Tipo I, obesidade tipo II, etc.), X é o domínio (Universo) de valores de n sobre o qual o significado do termo lingüístico é determinado (obesidade tipo I, por exemplo, entre 30 e 34,9 IMC) e m(t) é uma função semântica que assinala para cada termo lingüístico $t \in T$ o seu significado, que é um conjunto *fuzzy* em X (ou seja, $m : T \rightarrow (X)$ onde (X) é o espaço dos conjuntos *fuzzy*).

A inferência baseada em regras *fuzzy* pode ser compreendida como um funcional que mapeia um conjunto de entradas do sistema para um conjunto de saídas (como em um esquema de interpolação).

A regra *fuzzy* é uma unidade capaz de capturar algum conhecimento específico, e um conjunto de regras é capaz de descrever um sistema em suas várias possibilidades.

Cada regra *fuzzy*, da mesma forma que uma afirmação clássica, é composta por uma parte antecedente (a parte Se) e uma parte conseqüente (a parte Então), resultando em uma estrutura do tipo:

Se {antecedentes} Então {conseqüentes}.

Os antecedentes descrevem uma condição (premissas), enquanto a parte conseqüente descreve uma conclusão ou uma ação que pode ser esboçada quando as premissas se verificam.

2.2) MODELO DE REGRESSÃO LOGÍSTICA

Em Epidemiologia, a regressão logística tem como objetivo descrever a relação entre um resultado (variável dependente ou resposta) e um conjunto simultâneo de variáveis explicativas (preditoras ou independentes), mediante um modelo que tenha bom ajuste, que seja biologicamente plausível e obedeça ao princípio da parcimônia.

Em qualquer problema de regressão a quantidade sendo modelada é o valor médio da variável resposta dado os valores das variáveis independentes. Esta quantidade é chamada média condicional e será expressa por E(Y|x), em que Y denota a variável resposta e x denota os valores das variáveis independentes.

O modelo de regressão logística é descrito por:

$$P(x) = E(Y \mid x) = \mathrm{P(Y = 1 \mid x)} = \frac{\exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)}{1 + \exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_k\right)}$$

em que Y = 1 significa a presença da resposta, **x** = (x1; x2;...; xp) representa as covariáveis (fatores de risco), o parâmetro β_0 é o intercepto, e o β_k(k = 1;...; p) são os p parâmetros de regressão.

Estimação dos Parâmetros

A estimação dos parâmetros em regressão logística é feita, em geral, pelo método de máxima verossimilhança. Para aplicação deste método é necessário, inicialmente, construir a função de verossimilhança, a qual expressa a probabilidade dos dados observados como uma função dos parâmetros desconhecidos. Os estimadores de máxima verossimilhança dos parâmetros serão os valores que maximizam esta função L(β):

$$L(\beta) = \prod_{i=1}^{n} \left(P(x_i)\right)^{y_i} \left(1 - P(x_i)\right)^{1-y_i}$$

3) RESULTADOS

Os modelos acima descritos foram utilizados para estudar as causas da obesidade na população de moradores da região do Distrito Sul de Campinas.

3.1) MODELO FUZZY

Em um primeiro estudo, no modelo *fuzzy*, foram consideradas como variáveis de entrada, a porcentagem de gordura corporal (GC) e a relação cintura quadril (RCQ) dos indivíduos e, como saída o IMC (Índice de Massa Corpórea), classificando-os como nãoobseso (normal ou sobrepeso) e obeso, levando em consideração dados da OMS. Foi aplicado um modelo para cada sexo.

Abaixo podemos observar o gráfico de superfície obtido através do modelo de regras *fuzzy*, utilizando a função de pertinência triangular e o método de defuzificação Mamdani, para o sexo feminino.

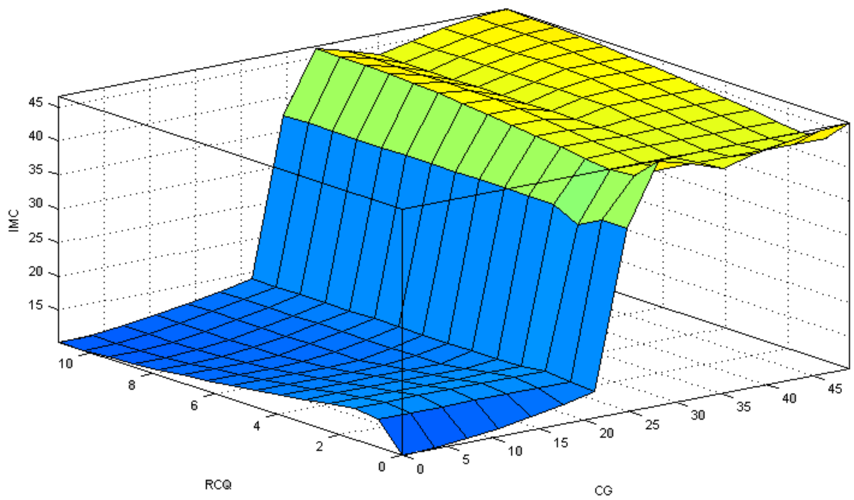

Gráfico: IMC em relação RCQ e GC para o sexo feminino.

3.1) MODELO LOGÍSTICO

Para ajustar este modelo aos dados usou-se o PROC LOGISTIC do pacote estatístico SAS.

Dentre as variáveis avaliadas no estudo sobre obesidade na região de moradores do Distrito Sul de Campinas, tanto a % de gordura corporal categorizada quanto a razão cintura quadrial categorizada, levando em consideração os diferentes sexos, mostraram-se significativas para classificação da obesidade.

Tabela1: Estatística de Wald e valor p para os parâmetros segundo o ajuste do modelo logístico.

Parâmetros	gl	Estatística Wald	Valor p
Intercepto	1	0,0026	0,9590
% Gordura Corporal	2	18,0675	<0,0001
Razão Cintura Quadril	1	5,7059	0,0169
Interação	2	2,1479	0,1428

Observando a Tabela 1, pode-se concluir que indivíduos com 40 anos ou mais têm 2 vezes mais chance de obesidade. Comparações entre o modelo de regressão *fuzzy* e o modelo de chances proporcionais serão feitas através de curva ROC.

REFERÊNCIAS

(1) World Health Organization. **Physical status: the use and interpretation of anthropometry**. Report of a WHO Expert Committee. WHO Technical Report Series. Geneva; 1995.
(2) Blumenkrantz M. **Obesity: the world's metabolic disorder** [online]. Beverly Hills, 1997. Available from www: URL:http://www.quantumhcp.com.obesity.htm
(3) Popkin BM & DOAK, CM. **The obesity epidemic is a worldwide phenomenon**. Nutrition Reviews, Washington DC, v.56, n.4 (Pt1), p.106-114, 1998.
(4) Ogden C.L., Yanovski S.Z., Carrol M.D. and Flegal K.M. **The Epidemiology of Obesity**. Gastroenterology, 132: 2087-2102, 2007.
(5) Souza, CA. **Teoria de conjuntos *fuzzy* e regressão logística na tomada de decisão para realização de cintilografia das paratiróides.** Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Saúde Pública da Universidade de São Paulo, 2007, 84p.
(6) Barros LC, Bassanezi RC. **Tópicos de Lógica *Fuzzy* e Biomatemática.** Coleção IMECC, 2006, 344p.
(7) Ortega NRS. **Aplicação da Teoria de Conjuntos *Fuzzy* a Problemas da Biomedicina.** Tese de Doutoramento submetida ao Instituto de Física da Universidade de São Paulo para obtenção do Título de Doutor em Ciências, 2001, 166p.
(8) Tanaka H, Uejima S and Asai K. **Linear Regression Analysis with Fuzzy Models,** IEEE Transactions on Systems, Man and Cybernetics, v. 12, n. 6, p.903-907, 1982.