Operations Research and Fuzziology 运筹与模糊学, 2016, 6(1), 27-35
Published Online February 2016 in Hans. http://www.hanspub.org/journal/orf
http://dx.doi.org/10.12677/orf.2016.61004

Hans汉斯

Parameter Estimation of the Fuzzy Logistic Regressive Model with LR Typed Fuzzy Coefficients

Yi Chen, Lili Wei

School of Mathematics and Computer, Ningxia University, Yinchuan Ningxia
Email: 308325191@qq.com, weill886@163.com

Received: Jan. 29th, 2016; accepted: Feb. 20th, 2016; published: Feb. 25th, 2016

Abstract

The classical logistic regression model is appropriate for the problems of binary variable. Since the observations were not crisp value, response variable was often between 0 and 1 and no probability distribution can be considered for response variable; hence error can not be completely regarded as random aspect. Therefore, by combining the classical logistic regression model with fuzzy sets theory, a fuzzy logistic regression model of crisp input and fuzzy output data is constructed; the coefficients and outputs are LR type fuzzy numbers. Considering the possibilities of success instead of the probabilities, the possibilities of success are described by some linguistic terms. Then the distance between two fuzzy numbers is constructed by cut sets. The least squares estimation of fuzzy parameters is obtained in proposed model based on the distance. Finally, the capability index MSI = 0.54 showed that the proposed model is effective of an ordinary one in the modeling Lupus.

Keywords

Fuzzy Nonlinear Regression, Possibility Odds, Fuzzy Least Squares Method, Capability Index Format

系数为LR-型模糊数的模糊Logistic回归模型的参数估计

陈　怡，魏立力

文章引用: 陈怡, 魏立力. 系数为LR-型模糊数的模糊Logistic回归模型的参数估计[J]. 运筹与模糊学, 2016, 6(1): 27-35.
http://dx.doi.org/10.12677/orf.2016.61004

宁夏大学数学计算机学院，宁夏 银川
Email: 308325191@qq.com, weill886@163.com

收稿日期：2016年1月29日；录用日期：2016年2月20日；发布日期：2016年2月25日

摘　要

针对二分类变量问题，经典Logistic回归是合适的。由于观测结果的不精确，响应变量往往介于0，1之间，且没有概率分布，误差也不能完全看作随机性现象。为此，将经典Logistic回归模型与模糊集理论相结合，构建了具有清晰输入－模糊输出的一类模糊Logistic回归模型，其中系数与输出均用LR-型模糊数表示。用成功的可能性替代概率，这些可能性可以由一些语义词描述。然后基于截集构造了模糊数之间的距离，利用此距离得到了上述模型中模糊参数的最小二乘估计。最后将模型应用在狼疮中并通过相容性指数MSI = 0.54说明该模型的有效性。

关键词

模糊非线性回归，可能性优势，模糊最小二乘法，相容性指数

1. 引言

基于模糊集理论的模糊回归分析模型对于处理模糊或不精确数据的分析提供了强有力的工具。近几十年来主要发展了两类主要的模糊回归方法。第一类是 1982 年 Tanaka [1]提出的可能性回归，该模型具有模糊系数和离散输入变量，从而使得模糊性最小化；另一类是 1987 年 Celmins 和 Diamond [2] [3]同时提出的模糊最小二乘法。目的是使得模糊数之间的距离最小化。这些模糊回归模型都是线性模型，近年来非线性模糊回归模型成为模糊回归分析的一个研究热点。

作为最流行的非线性模型之一的经典 Logistic 回归的响应变量服从伯努利分布[4]。但实际中由于不同原因导致观测结果不精确，变化的模型误差不能完全归功于随机性现象，故将 Logistic 回归模型和模糊集理论相结合作为一种新的模型，即模糊 Logistic 回归模型。

文献[5]提出模糊类 Logistic 模型，并利用模糊分类最大似然算法估计模型中的参数；文献[6] [7]研究了具有清晰解释变量、模糊响应变量的模糊 Logistic 回归模型，提出并介绍了可能性优势；文献[8]在最小绝对偏差方法的基础上对具有清晰输入-模糊输出数据的模糊 Logistic 回归模型的参数做出估计；文献[9]利用 Dk 距离对 LR-型模糊数的多元模糊线性回归模型进行了研究；文献[10]通过将模糊观测数据用区间来表示，然后利用区间的左、右端点和中点的数据集求出传统线性回归模型相应的回归系数；文献[11]应用模糊结构元理论，研究了系数为有界闭模糊数的多元线性回归模型。

本文对具有清晰输入-模糊输出的模糊 Logistic 回归模型的参数进行估计，其中输出与系数均是 LR-型模糊数。其次由于二分观测结果的模糊性，响应变量没有概率分布。这种模糊性可以通过可能性来评估和度量，用一些语义词描述可能性，并将语义词看作 LR-型模糊数。然后基于截集构造了模糊数之间的距离，利用此距离得到上述模型中模糊参数的最小二乘估计。最后将模型应用在临床案例中并通过相容性指数获得模型拟合的具体情况。

2. Logistic 回归

2.1. 经典 Logistic 回归

经典 Logistic 回归模型主要研究二分类的响应变量与影响结果的一些解释变量之间的关系。在这里

解释变量可以是离散型、连续型或混合型，且没有假设分布。响应变量 Y={0，1}(失败/成功)通常服从伯努利分布，即 $E(Y)=P(Y=1)=\pi$，$0<\pi<1$。

β_0 与 β_i，i≥分别是回归截距和回归系数，Y_{ij} 是第 i 个个体的第 j 个观测值，$\pi(\boldsymbol{x})=P(Y_{ij}=1|x_1,x_2,\cdots,x_n)$ 表示 $\boldsymbol{x}$ 的每一个预测变量 x_i 处“成功”的概率，则拥有 n 个预测变量的 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$ 的 Logistic 回归模型为

$$\pi(\boldsymbol{x})=\frac{\exp\left(\beta_0+\sum_{i=1}^{n}\beta_i x_i\right)}{1+\exp\left(\beta_0+\sum_{i=1}^{n}\beta_i x_i\right)},$$

经过 Logit 转化得到

$$\operatorname{logit}\left[\pi(\boldsymbol{x})\right]=\ln\left(\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}\right)=\beta_0+\beta_1x_1+\cdots+\beta_nx_n,$$

其中，表达式 $\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}$ 称为概率优势，$\beta_i\ (i=0,1,\cdots,n)$ 是模型中的参数。

当 $Y_{ij}=1$，并且控制 $\boldsymbol{x}$ 中的一部分预测变量 $x_j\ (i\neq j)$ 不变时，预测变量 x_i 每增加 1 个单位对多元 Logistic 回归模型的优势影响为 e^{β_i} 倍。

2.2. 模糊 Logistic 回归

在临床研究中，由于缺少合适的仪器或明确的标准，个体样本无法分类[8]。在实际中，更可行的措施是用语言变量来表示响应变量，这时响应类别是相对模糊的且不能认为其服从伯努利分布。这种模糊性可以通过成功的可能性来评估和度量。Pourahmad 等人在文献[7]提出了“可能性优势”的概念：

定义 2.1：设 $\mu_i,i=1,2,\cdots,m$ 是第 i 个个体成功可能性，$\mu_i=\operatorname{poss}(Y_i\approx1)$。成功可能性有两种情况：1) 精确值，$\mu_i\in\mathbb{R}$，$0\le\mu_i\le1$；2) 语义词，$\mu_i\in\{\cdots,$低,中,高,$\cdots\}$，用合适的模糊数来定义 μ_i，且 μ_i 支撑的并覆盖了整个(0, 1)区间。比值 $\frac{\mu_i}{1-\mu_i},i=1,2,\cdots,m$ 是第 i 个个体的可能性优势。

文献[7]讨论了第一种情况。本文主要讨论第二种情况，即具有模糊二分预测结果的 Logistic 回归模型，成功可能性由语义词 $\mu_i\in\{$非常低,低,中,高,非常高$\}$ 来代替。μ_i 通常定义为 LR-型模糊数，如下

$$\begin{aligned}&\text{非常低}=(0.02,0.01,0.16)_{LR},\quad \text{低}=(0.25,0.15,0.15)_{LR},\\&\text{中}=(0.50,0.15,0.15)_{LR},\quad \text{高}=(0.75,0.15,0.15)_{LR},\\&\text{非常高}=(0.98,0.18,0.01)_{LR}.\end{aligned}\tag{1}$$

Pourahmad 将对数 $\ln\frac{\mu_i}{1-\mu_i}$ 转化得到的可能性优势 $w_i,i=1,2,\cdots,m$ 看作观测结果，这些观测结果的隶属函数通过扩张原理以及 μ_i 的隶属函数来确定，如下：

$$w_i(y)=\sup_{\forall x:f(x)=y}\mu_i(x),$$

$f(x)=\ln\frac{x}{1-x},0<x<1$ 是一对一的函数，因此

$$w_i\left(\ln\frac{\mu_i}{1-\mu_i}\right)=\mu_i\left(\frac{e^x}{1+e^x}\right).\tag{2}$$

设有一组精确的解释变量以及模糊的观测结果$\left(x_{i1},x_{i2},\cdots,x_{in},\frac{\mu_i}{1-\mu_i}\right)$，用解释变量回归对数转化后的可能性优势$w_i,i=1,2,\cdots,m$，即模糊 Logistic 回归模型为：

$$W_i=\ln\frac{\mu_i}{1-\mu_i}=\sum_{j=0}^{n}A_jX_{ij},i=1,2,\cdots,m, \tag{3}$$

其中A_j是 LR-型模糊数，$x_{i0}=1$，x_{ij}是正实数。

3. 模糊最小二乘法

模糊最小二乘法由 Celmins 和 Diamond 同时提出，是最小二乘法的模糊扩展。因此一个合适的模糊数的距离定义是必要的。

定义 3.1 [12]：设 E 是函数空间，则对，基于函数之间的距离为：

$$d(u,v)=\left[\int_0^1 f(\alpha)d^2\left((u)_\alpha,(v)_\alpha\right)\mathrm{d}\alpha\right]^{\frac{1}{2}},$$

其中$d^2\left((u)_\alpha,(v)_\alpha\right)=\left[a_1(\alpha)-b_1(\alpha)\right]^2+\left[a_2(\alpha)-b_2(\alpha)\right]^2$，且$(u)_\alpha=\left[a_1(\alpha),a_2(\alpha)\right]$，$(v)_\alpha=\left[b_1(\alpha),b_2(\alpha)\right]$分别是$u$，$v$的$\alpha$截集，函数$f(\alpha)$是$d^2\left((u)_\alpha,(v)_\alpha\right)$的权重因子，在区间[0,1]上单调递增，满足$f(0)=0$，$\int_0^1 f(\alpha)\mathrm{d}\alpha=0.5$。通常将$f(\alpha)=\alpha$看作是权重函数。

为了获得最佳模型，预测变量W_i以及观测结果w_i之间误差平方和(SSE)应该最小。由定义 3.1 得到

$$\mathrm{SSE}=\sum_{i=1}^{m}\left(d(w_i,W_i)\right)^2,$$

其中$d(w_i,W_i)=\left[\int_0^1 f(\alpha)d^2\left((w_i)_\alpha,(W_i)_\alpha\right)\mathrm{d}\alpha\right]^{\frac{1}{2}}$。

为了不失一般性，我们假设$A_j=\left(a_j,s_j,t_j\right)_{LR},j=1,2,\cdots,n$，估计出的结果$W_i=\left(f_i(a),f_i(s),f_i(t)\right)_{LR}$，$i=1,2,\cdots,m$也是 LR-型模糊数。其中$f_i(a)=\sum_{j=0}^{n}a_jx_{ij}$，$f_i(s)=\sum_{j=0}^{n}s_jx_{ij}$，$f_i(t)=\sum_{j=0}^{n}t_jx_{ij}$，$x_{i0}=1$。通过计算得到：

$$(W_i)_\alpha=\left[(\alpha-1)f_i(s)+f_i(a),(1-\alpha)f_i(t)+f_i(a)\right].$$

为了计算$(w_i)_\alpha$，设$(\mu_i)_\alpha=\left[b_{i1},b_{i2}\right]$，此时

$$(w_i)_\alpha=\left[\ln\frac{b_{i1}}{1-b_{i1}},\ln\frac{b_{i2}}{1-b_{i2}}\right],$$

则

$$d^2\left((w_i)_\alpha,(W_i)_\alpha\right)=\left[\ln\frac{b_{i1}}{1-b_{i1}}-(\alpha-1)f_i(s)-f_i(a)\right]^2+\left[\ln\frac{b_{i2}}{1-b_{i2}}-(1-\alpha)f_i(t)-f_i(a)\right]^2,$$

$$\mathrm{SSE}=\sum_{i=1}^{m}\int_0^1\left[\left[\ln\frac{b_{i1}}{1-b_{i1}}+(1-\alpha)f_i(s)-f_i(a)\right]^2+\left[\ln\frac{b_{i2}}{1-b_{i2}}-(1-\alpha)f_i(t)-f_i(a)\right]^2\right]\mathrm{d}\alpha,$$

SSE 仅依赖模型系数$f_i(a),f_i(s),f_i(t)$。为求解 SSE 的最小值，令偏导$\frac{\partial}{\partial a_j}\mathrm{SSE}$，$\frac{\partial}{\partial s_j}\mathrm{SSE}$，以及$\frac{\partial}{\partial t_j}\mathrm{SSE}$等于 0，得到

$$\begin{aligned}
&6\sum_{j=0}^{n}a_j\sum_{i=1}^{m}x_{ij}x_{i\tau}-\sum_{j=0}^{n}\left(s_j-t_j\right)\sum_{i=1}^{m}x_{ij}x_{i\tau}=6\sum_{i=1}^{m}z_ix_{i\tau},\\
&\sum_{j=0}^{n}s_j\sum_{i=1}^{m}x_{ij}x_{i\tau}-2\sum_{j=0}^{n}a_j\sum_{i=1}^{m}x_{ij}x_{i\tau}=-12\sum_{i=1}^{m}k_ix_{i\tau}\\
&\sum_{j=0}^{n}t_j\sum_{i=1}^{m}x_{ij}x_{i\tau}+2\sum_{j=0}^{n}a_j\sum_{i=1}^{m}x_{ij}x_{i\tau}=12\sum_{i=1}^{m}l_ix_{i\tau},
\end{aligned}\tag{4}$$

其中 $\tau=0,1,\cdots,n$

公式(4)的矩阵表示为

$$\begin{aligned}
&6\boldsymbol{Aa}-\boldsymbol{A}\left(\boldsymbol{s}-\boldsymbol{t}\right)=6\boldsymbol{Z},\\
&\boldsymbol{As}-2\boldsymbol{Aa}=-12\boldsymbol{K},\\
&\boldsymbol{At}+2\boldsymbol{Aa}=12\boldsymbol{L},
\end{aligned}$$

$$\boldsymbol{A}=\boldsymbol{X}'\boldsymbol{X},\quad \boldsymbol{X}=\begin{pmatrix}1 & x_{11} & x_{12} & \cdots & x_{1n}\\ 1 & x_{21} & x_{22} & \cdots & x_{2n}\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & x_{m1} & x_{m2} & \cdots & x_{mn}\end{pmatrix}_{m\times(n+1)},$$

$$\boldsymbol{Z}=\left(\sum_{i=1}^{m}z_ix_{i0},\sum_{i=1}^{m}z_ix_{i1},\cdots,\sum_{i=1}^{m}z_ix_{in}\right)',$$

$$\boldsymbol{K}=\left(\sum_{i=1}^{m}k_ix_{i0},\sum_{i=1}^{m}k_ix_{i1},\cdots,\sum_{i=1}^{m}k_ix_{in}\right)',$$

$$\boldsymbol{L}=\left(\sum_{i=1}^{m}l_ix_{i0},\sum_{i=1}^{m}l_ix_{i1},\cdots,\sum_{i=1}^{m}l_ix_{in}\right)'.$$

当 $\operatorname{rank}\left(\boldsymbol{X}\right)=n+1$， $\boldsymbol{A}^{-1}$存在。则 SSE 的最小值具有唯一解[12]：

$$\begin{aligned}
&\boldsymbol{a}=\boldsymbol{A}^{-1}\left(3\boldsymbol{Z}-6\boldsymbol{K}-6\boldsymbol{L}\right),\\
&\boldsymbol{s}=\boldsymbol{A}^{-1}\left(6\boldsymbol{Z}-24\boldsymbol{K}-12\boldsymbol{L}\right),\\
&\boldsymbol{t}=\boldsymbol{A}^{-1}\left(-6\boldsymbol{Z}+12\boldsymbol{K}+24\boldsymbol{L}\right).
\end{aligned}\tag{5}$$

考虑到$\left(\mu_i\right)_\alpha=\left[b_{i1},b_{i2}\right]=\left(b_i,m_i,n_i\right)_{LR}=\left[b_i-\left(1-\alpha\right)m_i,b_i+\left(1-\alpha\right)n_i\right]$，则有

$$\left(w_i\right)_\alpha=\left[\ln\frac{b_i-\left(1-\alpha\right)m_i}{1-b_i+\left(1-\alpha\right)m_i},\ln\frac{b_i+\left(1-\alpha\right)n_i}{1-b_i-\left(1-\alpha\right)n_i}\right],$$

由公式(2)解得 w_i 的隶属函数为：

$$w_i=\begin{cases}1-\dfrac{b_i-\dfrac{\mathrm{e}^x}{1+\mathrm{e}^x}}{m_i}, & b_i-m_i\le\dfrac{\mathrm{e}^x}{1+\mathrm{e}^x}\le b_i,\\[2ex] 1-\dfrac{\dfrac{\mathrm{e}^x}{1+\mathrm{e}^x}-b_i}{n_i}, & b_i<\dfrac{\mathrm{e}^x}{1+\mathrm{e}^x}\le b_i+n_i,\end{cases}$$

4. 拟合优度

论文利用文献[13]提出的相容性指数评估模型估计值与观测值的具体拟合情况。

定义 4.1 设 A，B 是两个模糊数，则 A 和 B 之间的相容性指数定义如下：

$$S_{UI}=\frac{\operatorname{Card}\left(A\cap B\right)}{\operatorname{Card}\left(A\cup B\right)},\ \operatorname{Card}A=\begin{cases}\int_x A\left(x\right)\mathrm{d}x, & x\text{为连续情形},\\ \sum_x A\left(x\right), & x\text{为连续情形},\end{cases}\tag{6}$$

其中∩与∪分别是两个模糊集的“最小”与“最大”运算。

定理 4.1 [13]：设 A，B 是两个模糊数，则

1) $0 \le S_{UI}(A,B) \le 1$;

2) $A = B \Leftrightarrow S_{UI}(A,B) = 1$;

3) $S_{UI}(A,B) = S_{UI}(B,A)$;

4) $A \cap B = \varnothing \Leftrightarrow S_{UI}(A,B) = 0$;

5) $A \subset B \subset C \Rightarrow S_{UI}(A,C) \le \min\left(S_{UI}(A,B), S_{UI}(B,C)\right)$.

定义 4.2 [14]：对于模糊 Logistic 回归模型，相容性指数是评估模型拟合优度的一种度量：

$$\mathrm{MSI} = \frac{1}{m}\sum_{i=1}^{m} S_{UI}(w_i, \mathrm{W}_i),\ 0 \le \mathrm{MSI} \le 1. \tag{7}$$

显然大的 MSI 对应更好的拟合优度。

5. 一个临床医学的实例分析

这个数值案例来自参考文献[7]。系统性红斑狼疮(Systematic Lupus Erythematosus，简称 SLE)是一种慢性自体免疫疾病，产生的抗体会攻击身体中的多个系统。由于这种疾病的潜伏期较长，因此对于狼疮没有良好的诊断试验。Physicians 尝试从之前的病史、试验以及现有的症状中收集信息。他们制定了 11 个标准来诊断对象是否患有 SLE [7]。通常，在做出诊断之前一个人至少要满足上述标准中的 4 个。那么如果一个人满足其中 3 个标准，这个人是否健康？或者对满足多于 3 个标准的患者，他们的患病严重程度是否相同？

SLE 在区分患者与正常人之间的界限并不清晰. 因此, 有研究者利用语义词对每个样本患病的可能性赋值：{非常低，低，中，高，非常高}。SLE 的患病可能性的定义见公式(1)。为了研究 SLE 患病的可能性优势与表 1 提到的影响因子之间的关系，提出如下模型：

$$\tilde{W}_i = \ln\frac{\tilde{\mu}_i}{1-\tilde{\mu}_i} = A_0 + A_1 x_{i1} + \cdots + A_5 x_{i5},$$

Table 1. Fuzzy binary observations in SLE disease and the values of related risk factors
表 1. SLE 模糊二分观测数据以及相关的影响因子

	家族史	光暴露	ANA	Anti-DNA	ESR	优势
1	1	1	112	105	1	高
2	0	1	80	23	0	中
3	0	1	115	15	0	高
4	0	1	105	107	1	高
5	0	0	89	150	1	中
6	1	1	160	10	1	非常高
7	0	1	100	23	0	中
8	0	0	100	85	1	高
9	0	1	48	83	0	低
10	1	0	15	19	1	非常低
11	0	0	50	91	0	低
12	0	1	59	200	1	中
13	0	1	83	20	1	低
14	0	0	15	200	0	低
15	1	0	85	15	1	中

其中 $i=1,2,\cdots,15$， x_{i1} 是关于疾病的家族史； x_{i2} 是光暴露； x_{i3} 是 ANA 试验结果； x_{i4} 是 Anti-DNA 试验结果； x_{i5} 是 ESR 试验结果。

为了估计系数 $A_j=\left(a_j,s_j,t_j\right)_{LR}$, $j=0,1,\cdots,5$，

$$\boldsymbol{X}=\begin{pmatrix}1&1&1&\cdots&1\\1&0&1&\cdots&0\\\vdots&\vdots&\vdots&&\vdots\\1&1&0&\cdots&1\end{pmatrix},\ \boldsymbol{A}=\boldsymbol{X}'\boldsymbol{X}=\begin{pmatrix}15&4&9&1216&1146&9\\4&4&2&372&149&4\\9&2&9&862&586&5\\1216&372&862&119564&78864&808\\1146&149&586&78864&149738&711\\9&4&5&808&711&9\end{pmatrix}_{6\times6},$$

$$\boldsymbol{Z}=\left(-0.0416,1.0879,4.4673,749.8026,-123.4730,2.2147\right)',$$

$$\boldsymbol{K}=\left(-1.1466,-0.2661,0.0543,28.1208,-95.2428,-0.3837\right)',$$

$$\boldsymbol{L}=\left(1.1292,0.6326,1.3903,216.0201,53.8861,1.1341\right)',$$

由于 rank($\boldsymbol{A}$) = 6，公式(5)有唯一的解：

$$\boldsymbol{a}=\boldsymbol{A}^{-1}\left(3\boldsymbol{Z}-6\boldsymbol{K}-6\boldsymbol{L}\right)=\left(-4.1455,0.5106,-0.1338,0.0464,0.0099,-0.7191\right)',$$

$$\boldsymbol{s}=\boldsymbol{A}^{-1}\left(6\boldsymbol{Z}-24\boldsymbol{K}-12\boldsymbol{L}\right)=\left(0.3994,0.6258,0.0329,0.0051,0.0005,-0.2072\right)',$$

$$\boldsymbol{t}=\boldsymbol{A}^{-1}\left(-6\boldsymbol{Z}+12\boldsymbol{K}+24\boldsymbol{L}\right)=\left(1.5062,0.3930,0.0406,-0.0085,-0.0034,0.3622\right)'.$$

上述结果中 s_5，t_3，t_4 均小于 0，违反了模糊数的基本性质，但是这些数都非常接近 0，故令这些数等于 0，得到最终模型：

$$\begin{aligned}\tilde{W}=\ln\frac{\tilde{\mu}}{1-\tilde{\mu}}&=\left(-4.1455,0.3994,1.5062\right)_{LR}+\left(0.5106,0.6258,0.3930\right)_{LR}x_{i1}\\&+\left(-0.1338,0.0329,0.0406\right)_{LR}x_{i2}+\left(0.0464,0.0051,0\right)_{LR}\\&+\left(0.0099,0.0005,0\right)_{LR}x_{i4}+\left(-0.7191,0,0.3622\right)_{LR}x_{i5}.\end{aligned}\tag{8}$$

利用公式(8)的模型可以估计人们患有狼疮的可能性。例如，对于第 3 个样本(0, 1, 115, 15, 0)得到的估计结果为：

$$\tilde{W}_3=\ln\frac{\tilde{\mu}_3}{1-\tilde{\mu}_3}=\left(1.2052,1.0263,1.5468\right)_{LR}=\left(1.21,1.03,1.55\right)_{LR}.$$

根据扩张原理得到可能性优势的隶属函数为：

$$\frac{\tilde{\mu}_3}{1-\tilde{\mu}_3}(x)=\exp\left(\tilde{W}_3(x)\right)=\begin{cases}1-\dfrac{1.21-\ln x}{1.03}, & 0.18\le\ln x\le1.21\Leftrightarrow1.20\le x\le3.35,\\[2ex]1-\dfrac{\ln x-1.21}{1.55}, & 1.21<\ln x\le2.76\Leftrightarrow3.35\le x\le15.80,\end{cases}$$

以及具有狼疮可能性的隶属函数：

$$\tilde{\mu}_3(x)=\tilde{W}_3\left(\ln\frac{x}{1-x}\right)=\begin{cases}1-\dfrac{1.21-\ln\dfrac{x}{1-x}}{1.03}, & 0.55\le x\le0.77,\\[2ex]1-\dfrac{\ln\dfrac{x}{1-x}-1.21}{1.55}, & 0.77\le x\le0.94.\end{cases}$$

假设有一个新的样本，其信息为(0, 1, 110, 87, 0)，通过得到的模型估计患病的可能性优势为：

$$\tilde{W}_{new}=\ln\frac{\tilde{\mu}_{new}}{1-\tilde{\mu}_{new}}=(1.686,1.0368,1.5468)_{LR}=(1.69,1.04,1.55)_{LR}.$$

$$\frac{\tilde{\mu}_{new}}{1-\tilde{\mu}_{new}}(x)=\begin{cases}1-\dfrac{1.69-\ln x}{1.04}, & 1.92\le x\le 5.42,\\ 1-\dfrac{\ln x-1.69}{1.55}, & 5.42\le x\le 25.53,\end{cases}$$

以及具有狼疮可能性的隶属函数：

$$\tilde{\mu}_{new}(x)=\begin{cases}1-\dfrac{1.69-\ln\dfrac{x}{1-x}}{1.04}, & 0.66\le x\le 0.84,\\ 1-\dfrac{\ln\dfrac{x}{1-x}-1.69}{1.55}, & 0.84\le x\le 0.96,\end{cases}$$

并且根据定义 4.1、定义 4.2 得到模型的相容性指数为：

$$\mathrm{MSI}=\frac{1}{15}\sum_{i=1}^{15}S_{UI}(w_i,W_i)=\frac{1}{15}\times 8.1153=0.54.$$

MSI > 0.50，说明模糊 Logistic 回归具有很好的拟合结果。

在临床医学中，经典 Logistic 回归模型响应变量的取值为 0 (没患病)或 1 (患病)，而模糊 Logistic 回归响应变量的取值为五个语义词：{非常低，低，中，高，非常高}，这样更加符合实际情况。

6. 结论

在经典 Logistic 回归分析中，解释变量是没有假设分布的[15]，二分响应变量往往服从伯努利分布。但实际中二分响应变量的观测结果普遍模糊且没有概率分布。忽视这类观测结果是不合理的，这就需要一个新的模型。

本文利用模糊最小二乘法来估计具有清晰输入-模糊输出的模糊 Logistic 回归模型中的参数。成功可能性由语义词表示为：{非常低，低，中，高，非常高}，这些语义词支撑的并覆盖了(0, 1)区间。本文选择第二种定义，然后将每一个样本的成功可能性进行对数转化得到 w_i。基于两个模糊数的距离定义，利用模糊最小二乘法估计模型中的参数。同时给出评估模型的一个拟合优度准则，最后利用所提出的模型来研究一个关于狼疮的实例。

与之前的研究相比，本文的方法具有一些优点：

1) 将模糊线性回归扩展到模糊非线性回归；

2) 用可能性替代概率，语义词表示观测结果，使得模型更加完善；

3) 观测结果与系数均用 LR-型模糊数表示，计算简单且在实践中更常见；

4) 基于 α 截集构造模糊数之间的距离相比基于模糊数的三个点(左端点、中间、右端点)更加准确；

5) 利用相容性指数获得模型拟合的具体情况。

本文研究了清晰输入-模糊输出的模糊 Logistic 回归模型，此回归模型还可以扩展到输入输出及系数均为 LR-型模糊数或其它类型模糊数，利用模糊最小二乘法来估计相应模型中的参数。

致　谢

非常感谢我的导师魏立力老师对我的指导，从论文的定题，到参考文献的查阅，到写作、修改，到

最后的定稿，魏老师给了我耐心的指导和无私的帮助。魏老师的这种无私奉献的敬业精神令我钦佩，他不仅教会了我如何学习，也教会了我如何做人。在此我向魏老师表示我诚挚的谢意！

基金项目

国家自然科学基金资助项目(11261044)。创新项目：宁夏大学研究生创新项目(GIP2015034)。

参考文献 (References)

[1] Tanka, H., Uejima, S. and Asai, K. (1982) Linear Regression Analysis with Fuzzy Model. *IEEE Transactions on Systems, Man, and Cybernetics*, **12**, 903-907. http://dx.doi.org/10.1109/TSMC.1982.4308925

[2] Celmins, A. (1987) Least Squares Model Fitting to Fuzzy Vector Data. *Fuzzy Sets System*, **22**, 245-269. http://dx.doi.org/10.1016/0165-0114(87)90070-4

[3] Diamond, P. (1987) Least Squares Fitting of Several Fuzzy Variables. *Proceedings of the second IFSA Congress*, Tokyo, 20-25.

[4] Agresti, A. (2002) Categorical Data Analysis. John Wiley & Sons, New York. http://dx.doi.org/10.1002/0471249688

[5] Yang, M. and Chen, H. (2004) Fuzzy Class Logistic Regression Analysis. *Fuzziness and Knowledge Based Systems*, **12**, 761-780. http://dx.doi.org/10.1142/S0218488504003193

[6] Pourahmad, S., Ayatollahi, S.M.T. and Taheri, S.M. (2001) Fuzzy Logistic Regression: A New Possibilistic Model and Its Application in Clinical Vague Status. *Iranian Journal of Fuzzy Systems*, **1**, 1-17.

[7] Pourahmad, S., Ayatollahi, S.M.T., Taheri, S.M., *et al*. (2011) Fuzzy Logistic Regression Based on the Least Squares Approach with Application in Clinical Studies. *Computers and Mathematics with Application*, **62**, 3353-3365. http://dx.doi.org/10.1016/j.camwa.2011.08.050

[8] Namdari, M., Yoon, J.H., Abadi, A., *et al*. (2015) Fuzzy Logistic Regression with Least Absolute Deviations Estimators. *Soft Computing*, **19**, 909-917. http://dx.doi.org/10.1007/s00500-014-1418-2

[9] 梁艳, 魏立力. 系数为 LR-型模糊数的模糊线性最小二乘回归[J]. 模糊系统与数学, 2007, 21(3): 112-117.

[10] 张爱武. 系数为 LR-型模糊数的模糊回归模型的参数估计[J]. 模糊系统与数学, 2013, 27(6): 140-147.

[11] 岳立柱. 系数为一般模糊数的多元线性回归模型[J]. 统计与决策, 2015(3): 72-74.

[12] Xu, R. and Li, C. (2001) Multidimensional Least Squares Fitting with a Fuzzy Model. *Fuzzy Sets and Systems*, **119**, 215-223. http://dx.doi.org/10.1016/S0165-0114(98)00350-9

[13] Taheri, S.M. and Kelkinnama, M. (2012) Fuzzy Linear Regression Based on Least Absolutes Deviations. *Iranian Journal of Fuzzy Systems*, **9**, 121-140.

[14] Kelkinnama, M. and Taheri, S.M. (2012) Fuzzy Least Absolutes Regression Using Shape Preserving Operations. *Information Sciences*, **214**, 10 5-120.

[15] Domr, M., Zain, R., Kareem, S.A., *et al*. (2007) An Adaptive Fuzzy Regression Model for the Prediction of Dichotomous Response Variables. *Computational Science and Applications*, **15**, 14-19.