

# Evaluation Model of Diabetes Therapeutic Effect Based on Multiple Linear Regression

Xuan Liu, Zhiming Wang\*

School of Information Science and Technology, Hunan Agricultural University, Changsha Hunan  
Email: \*1009916073@qq.com

Received: May 24<sup>th</sup>, 2019; accepted: Jun. 7<sup>th</sup>, 2019; published: Jun. 14<sup>th</sup>, 2019

---

## Abstract

This paper studied the relationship between the re-admission rate of diabetic patients and its related eigenvectors. The diagnostic model was made by ridge regression and linear fitting. The relationship between the re-admission rate and its corresponding characteristic variables was obtained by stepwise regression. The analysis of model results shows that: the success rate of diabetes treatment in hospitals is 74.32%; insulin and metformin can be used to control diabetes in low and middle income groups; tolazamide can be used to control diabetes in high income groups. The treatment model of this problem is helpful to the rational treatment of diabetes mellitus patients.

## Keywords

Ridge Regression, Multiple Regression, Stepwise Regression, Diabetes Mellitus

---

# 基于多元线性回归的糖尿病治疗效果评价模型

刘璇, 王志明\*

湖南农业大学信息科学技术学院, 湖南 长沙  
Email: \*1009916073@qq.com

收稿日期: 2019年5月24日; 录用日期: 2019年6月7日; 发布日期: 2019年6月14日

---

## 摘要

本文针对糖尿病患者的再次入院率与其相关特征向量的关系进行研究, 采用岭回归以及线性拟合得到评价指标体系, 通过逐步回归得到了再次入院率与其相应特征变量之间的关系模型。对模型结果分析表明: 医院对

\*通讯作者。

糖尿病患者治疗成功率为74.32%; 中低水平收入者可考虑选用insulin、metformin对糖尿病进行控制; 高水平收入者可考虑选用tolazamide对糖尿病进行控制。本问题出的治疗模型有助于糖尿病患者的合理治疗。

## 关键词

岭回归, 多元回归, 逐步回归, 糖尿病

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着生活方式的改变, 导致肥胖和超重人口数量增加, 我国糖尿病的发病率呈现出逐年增高的趋势。根据国际糖尿病联盟(IDF)预测到 2040 年, 中国糖尿病患者数量将达到 1.51 亿人, 相比于 2015 年糖尿病患者数量增加近 50%, 因此控制并降低糖尿病的患病率十分必要[1], 中南大学谢玉秀[2]通过分析住院 2 型糖尿病患者各项数据, 采用 Logistic 逐步回归分析筛选相关危险因素, 得到对于 2 型糖尿病, 不应只是单纯的控制血糖或追求其他单个指标的控制, 而应该注重在降糖、降压、降脂方面的综合治疗; 王珍[3]等人通过分析糖尿病患者 3 种不同的血糖状态在 TG、胰岛素抵抗指数(HOMA-IR)、胰岛素分泌指数(HOMA- $\beta$ )、血清总免疫球蛋白 E(IgE)和 C-RP 的分布, 得到结论 HOMA-IR 与 HOMA- $\beta$  是公认的糖尿病的危险因子, 为糖尿病的治疗提供了理论方向; 徐秀菊[4]通过对糖尿病患者进行药物注射治疗, 比较其腹血糖、体重指数、糖化血红蛋白等各项指标, 得到格列美脲与胰岛素联合使用医治 2 型糖尿病的效果明显, 能够有效减少餐后 2 h 血糖含量、空腹血糖含量、糖化血红蛋白含量, 降低胰岛素的使用量, 对体重指数影响很小, 值得在临床上推广应用; John Doupis [5]等人提出二肽基肽酶-IV (DPP4)的抑制导致胰高血糖素样肽-1 (GLP-1)和胃抑制性多肽(GIP)的血液浓度增加, 这导致胰岛素分泌的葡萄糖依赖性刺激增加, 从而导致血糖水平降低, 为糖尿病的治疗提供了高效性药物以及具有临床疗效的最新药物; Knowler WC [6]等人通过比较注射 metformin 和改善生活方式下血液中葡萄糖浓度, 得到生活方式的改变与 metformin 均能降低糖尿病发病率高危人群, 但生活方式干预是比 metformin 更有效, 为糖尿病患者提供除药物治疗外的其它不产生副作用的治疗方式; 朱彩蓉[7]等人通过 Markov 状态转移决策树模型对新药罗格列酮钠治疗糖尿病的长期效果进行评价, 发现 Markov 状态转移决策树模型是评价药物治疗长期效果的有效模型, 为糖尿病的药物治疗效果提供模型支持。

本文利用岭回归消除数据之间的共线性提取权重较高的指标, 并进行多元线性拟合得到评价指标体系; 采用逐步回归的方法, 建立再次入院率与各特征变量之间的多元线性回归模型, 分析各特征变量对再次入院率的影响; 根据模型结果对医院的治疗方案进行评价并对糖尿病患者给出合理的治疗建议。

## 2. 数据预处理

通过对数据表<sup>1</sup>进行分析, 发现种族以及主治医师专业这两特征变量, 存在大量的缺失值, 因此本文认为种族以及主治医师专业缺失的患者数据为无用数据; 发现体重这项数据只有少数患者有信息, 在药物使用方面, acetohexamide、troglitazone、examide、citoglipton 这些单独使用药物以及 glimepiride-pioglitazone、metformin-rosiglitazone、metformin-pioglitazone 这些联合使用药物的患者使用人

<sup>1</sup>数据来源美国 130 家医院的糖尿病患者的治疗数据。

数均小于 10, 仅为所有就诊人数的万分之二, 在后续的研究中参考价值以及研究价值极小, 因此本文认为体重以及 acetohexamide、troglitazone、examide、citoglipton、glimepiride-pioglitazone、metformin-rosiglitazone、metformin-pioglitazone 这几类药物均为无效数据, 于是剔除两项无用数据以及六类无效数据, 然后将数据进行量纲化, 以便后续分析。

### 3. 模型建立

#### 3.1. 评价指标体系

本文选用再次入院作为治疗效果的评判, 未再次入院设为 0, 再次入院设为 1, 由于指标之间可能存在共线性, 因此本文选择岭回归[8]来对数据进行处理, 提取 7 项权重较高的指标分别为住院时间、诊断程序次数、门诊次数、住院次数、是否确诊为糖尿病患者、出院去处、诊断次数, 将其进行多元线性拟合得到系数表, 见表 1。

Table 1. Coefficient table

表 1. 系数表

模型	未标准化系数		标准化系数	t
	B	标准误差	Beta	
(常量)	0.201	0.007		30.073
time in hospital	0.005	0.001	0.027	8.518
num procedures	-0.012	0.001	-0.041	-13.239
number outpatient	0.020	0.001	0.051	16.704
number inpatient	0.078	0.001	0.197	64.034
diabetes Med	0.060	0.004	0.051	16.769
discharge disposition id	-0.002	0.000	-0.025	-8.051
number diagnoses	0.022	0.001	0.084	26.811

由此得到评价指标体系为:

$$Y = 0.201 + 0.005X_1 - 0.012X_2 + 0.02X_3 + 0.078X_4 + 0.06X_5 - 0.002X_6 + 0.022X_7$$

由评价指标体系可知, 住院天数、门诊次数、入院次数以及诊断次数越高, 表示该患者的健康状况越差, 因此该患者再次入院概率增大, Y 值越接近于 1, 表示医院的治疗效果越不理想, 该评价指标体系符合实际, 因此本文用其来评价医院对糖尿病患者治疗效果。

#### 3.2. 逐步回归

由于特征变量数量较多, 因此需进行变量筛选, 剔除与再次入院率不相关的因素。在初步进行变量筛选时, 本文发现出院去处、住院来源、挂号类别这三类数据中, 数据之间不存在关联, 彼此为相互独立的解释数据, 因此在考虑特征变量与再次入院率[9]的关系时, 将其剔除。对剩下的数据建立逐步回归模型为

$$y = 0.139 + 0.072x_1 + 0.021x_2 + 0.069x_3 + 0.023x_4 + 0.016x_5 - 0.008x_6 + 0.005x_7 - 0.029x_8 \\ + 0.005x_9 + 0.046x_{10} - 0.001x_{11} + 0.136x_{12} + 0.019x_{13} - 0.014x_{14} - 0.01x_{15} - 0.247x_{16}$$

其中模型中的变量  $x_1$  到  $x_{16}$  分别表示 number inpatient、number diagnoses、diabetes Med、number outpatient、number emergency、num procedures、time in hospital、metformin、age、repaglinide、insulin、acarbose、change、gender、num medications、tolazamide。

由回归拟合方程可知, 当其他特征变量不变的情况下, 再次入院的概率为 0.139, 由此可知, 当所有条件达到最优时, 该患者的健康状态较好, 再次入院率较小; 再次入院率与住院次数、诊断次数、确诊是否为糖尿病患者、门诊次数、紧急就诊次数、住院天数、年龄[10]、病情是否恶化以及 repaglinide、acarbose 两类药物的使用情况呈正相关关系, 即当这些特征变量数值越大时, 其再次入院率越高; 与诊断程序次数、药物使用次数、性别以及 metformin、insulin、tolazamide 三类药物使用呈负相关关系, 即当这些特征变量的数值越大时, 其再次入院的概率越低。

### 3.3. 治疗模型评价及建议

将所建立的评价指标体系运用于治疗效果的评价, 为减小随机误差, 本文随机选取 10,000 名就诊病人共选取 5 次, 避免偶然误差的产生, 由于再次入院的变量为 0, 1 两变量, 因此拟合方程的结果在[0,1]区间内, 即多元线性拟合得到的结果为再次入院的概率, 本文规定得到结果小于 0.5 的归为 0, 结果大于 0.5 的归为 1, 由此得到再次入院的概率如表 2。

**Table 2.** The hospital readmission rate of randomized selection of patients

**表 2.** 随机选取就诊病人再次入院率

随机选取次数	再次入院率
第一次	25.72%
第二次	25.91%
第三次	25.36%
第四次	25.96%
第五次	25.47%
平均值	25.68%

通过对再次入院率计算结果可知, 其治疗成功率为 74.32%, 因此医院对糖尿病患者的治疗效果较好。

分析数据表中 23 项药物以及联合药物的使用, 在初次变量筛选中剔除了七项药物无效数据, 以及在逐步回归分析中剔除了 11 类对再次入院率影响不显著的药物后, 最终得到 5 类影响显著的药物分别为 repaglinide、acarbose、metformin、insulin、tolazamide, 其中在逐步回归模型的分析中前两者药物的治疗效果较差, 因此在后续的治疗中只需使用 metformin、insulin、tolazamide 这三类药物。根据回归模型, tolazamide 在三种药物的使用中效果最好, 其次是 metformin, 在三者中效果较为不明显的为 insulin。查询三类药物的价格、使用量, 得到每天花费药物价格如表 3。

**Table 3.** Detailed prices list of drug

**表 3.** 药物价格详情表

药物	价格	使用量	每天花费药物价格
metformin	10 元/g	0.5 g/天	5 元
insulin	60 元/300 单位	6 单位/天	1.2 元
tolazamide	940 元/g	0.25 g/天	235 元

故综合考虑所有因素, 我们给出如下建议: 1) 在药物的选择上, 选择 metformin、insulin、tolazamide 这三类治疗效果好的药物; 2) 考虑患者收入水平, 中低水平收入者可考虑选用 insulin、metformin 对糖尿病进行控制, 高水平收入者可考虑 tolazamide 对疾病进行控制。

#### 4. 结论

通过评价指标体系评价可知, 医院的治疗模型对糖尿病患者的治疗效果较好, 其治疗成功率为 74.32%; 通过逐步回归模型以及结合药物价格得到, 中低水平收入者可考虑选用 insulin、metformin 对糖尿病进行控制, 其降低再次入院率的概率分别为 0.001、0.029, 高水平收入者可考虑 tolazamide 对疾病进行控制, 其降低再次入院率的概率为 0.247。

#### 基金项目

湖南省教育科学“十三五”规划项目(XJK016BXX009); 湖南农业大学 2018 年教改项目(67)。

#### 参考文献

- [1] 李红金, 赵敏, 綦雯雯, 王献伟, 李征. 2010 美国糖尿病协会糖尿病治疗指南[J]. 中国卒中杂志, 2011, 6(4): 316-324.
- [2] 谢玉秀. 住院 2 型糖尿病患者综合控制情况及慢性并发症相关危险因素分析[D]: [博士学位论文]. 长沙: 中南大学, 2013.
- [3] 王珍, 张红, 沈旭慧, 叶国芬, 金岷立, 钱莉, 张永红. 糖尿病前期与糖尿病相关因素的研究[J]. 中国糖尿病杂志, 2012, 20(6): 421-423.
- [4] 徐秀菊. 评价格列美脲联合胰岛素治疗 2 型糖尿病的临床疗效[J]. 世界最新医学信息文摘, 2015, 15(23): 115-125.
- [5] Doupis, J. (2008) AristidisVevesDPP4 Inhibitors: A New Approach in Diabetes Treatment. *Advances in Therapy*, **25**, 627 <https://doi.org/10.1007/s12325-008-0076-1>
- [6] Knowler, W.C., Barrett-Connor, E., Fowler, S.E., Hamman, R.F., Lachin, J.M., Walker, E.A. and Nathan, D.M. (2002) Reduction in the Incidence of Type 2 Diabetes with Lifestyle Intervention or Metformin. *The New England Journal of Medicine*, **346**, 393-403. <https://doi.org/10.1056/NEJMoa012512>
- [7] 朱彩蓉, 倪宗瓚, 孙克宏. 评价罗格列酮钠治疗 2 型糖尿病长期效果的 Markov 模型[J]. 卫生研究, 2005, 34(3): 355-358.
- [8] 杨楠. 岭回归分析在解决多重共线性问题中的独特作用[J]. 统计与决策, 2004(3): 14-15.
- [9] 马亮亮, 田富鹏. 基于糖尿病与民族因素的多元线性回归分析[J]. 山西大同大学学报(自然科学版), 2009, 25(4): 3-4, 19.
- [10] 黄琼刁, 邓万溪, 黄钦展, 林明欣. 糖尿病主要并发症年龄与体质之相关性研究[J]. 世界中医药, 2013, 8(3): 288-290.

**知网检索的两种方式：**

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择：[ISSN]，输入期刊 ISSN：2325-2251，即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：[sa@hanspub.org](mailto:sa@hanspub.org)