

Statistical Analysis of Major Risk Factors Affecting Glycosylated Hemoglobin Levels in Diabetic Patients

Junyu Jiang, Tianfang Zhang*

College of Mathematics and Information Science, Jiangxi Normal University, Nanchang Jiangxi
Email: 1301108065@qq.com, *tifa8804@163.com

Received: Oct. 3rd, 2017; accepted: Oct. 20th, 2017; published: Oct. 30th, 2017

Abstract

The aim of this study is to identify the major risk factors for HbA1c in diabetic patients. Through the collection of a hospital from January 2015 to December 2016 in diabetic patients with clinical and laboratory data, 10 cases were randomly sampled from representative and perfect data. Then using the method of Latin hypercube sampling processing of these data, we derived a new data sampling group. After that, the Minitab software was used to analyze the sampling data by multiple linear regression analysis, and the reliable conclusions of both medical significance and statistical significance were obtained.

Keywords

Glycosylated Hemoglobin, Super Latin Squares, Correlation Coefficient, Regression Analysis

影响糖尿病患者糖化血红蛋白含量的主要危险因素统计分析

江峻瑜, 张天芳*

江西师范大学数学与信息科学学院, 江西 南昌
Email: 1301108065@qq.com, *tifa8804@163.com

收稿日期: 2017年10月3日; 录用日期: 2017年10月20日; 发布日期: 2017年10月30日

*通讯作者。

摘要

本文的目的在于找出影响糖尿病患者糖化血红蛋白含量的主要危险因素。通过收集某医院在2015年1月至2016年12月间的糖尿病患者的临床及实验室数据资料, 从中随机抽取10例具代表性、完善性的数据资料, 然后使用超拉丁方抽样的方法对这些数据资料进行加工处理, 得到一组全新的抽样数据。之后使用minitab软件工具对抽样数据作多元线性回归分析, 得出同时具有医学意义和统计学意义的可靠结论。

关键词

糖化血红蛋白, 超拉丁方, 相关系数, 回归分析

Copyright © 2017 by authors 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. 背景

糖尿病是以持续高血糖为其基本生化特征的一种慢性全身性代谢性疾病, 主要是由于体内胰岛素分泌绝对或相对不足, 从而导致以糖代谢紊乱为主的糖、蛋白质、脂肪代谢紊乱的一种综合病症[1]。目前糖尿病诊断标准更新的主要变化是将糖化血红蛋白(HbA1c)应用于糖尿病的诊断, 糖化血红蛋白是血液中红细胞内的血红蛋白与血糖结合的产物, 其结合生成糖化血红蛋白是不可逆反应, 并与血糖浓度成正比, 且保持120天左右, 所以可以观测到120天之前的血糖浓度。糖化血红蛋白测试通常可以反映患者近8~12周的血糖控制情况[2]。2010年1月国际专家委员会发表声明将糖化血红蛋白(HbA1c)作为糖尿病诊断指标, 确定 $HbA1c \geq 6.5\%$ 为诊断切点[3], 这也进一步确定了HbA1c在糖尿病诊断史上的地位。

考虑到 $HbA1c \geq 6.5\%$ 被作为糖尿病的诊断切点。本文对影响糖尿病患者糖化血红蛋白(HbA1c)含量的主要危险因素统计分析, 为临床合理治疗与护理以及应用药物提供科学依据。

2. 资料与方法

数据来源于某医院计算机系统中2015~2016年糖尿病患者的临床及实验室数据资料, 包括糖化血红蛋白(Y, %)、年龄(X_1 , 岁)、体重指数(X_2 , kg/m^2)、总胆固醇(X_3 , mmol/L)、收缩压(X_4 , mm/Hg)和舒张压(X_5 , mm/Hg)等数据资料[4]。现从中随机抽取代表性的10例, 数据见下表1。

2.1. 超拉丁方抽样

当试验点之间有负相关时, 方差 $Var(y(x)) = n$ 可以减少, 即可以提高估计总体均值的精度。基于这种思想, McKay *et al.* (1979)提出了超拉丁方抽样方法(LHS), 该方法是采用两步随机化, 可以给出总体均值的无偏估计, 且其渐近方差比SRD的小, 其本质是分层抽样方法。假如试验次数为 n , LHS方法首先对区域进行分层, 即区域的每一维都等分为 n 个小区间, 这样试验域就等分为 n^s 个小方格, 然后在 n^s 个方格中选取 n 个方格, 使得任一行和任一列都仅有一个方格被选中, 最后在选中的 n 个方格中各自随机选取一个点组成最后的 n 个试验点, 这种方法使试验域 C^s 内任一点都可能被抽到[5]。给定试验点数 n 和因素个数 s , LHS的构造过程分为两步, 具体的构造方法如下。

Table 1. Data of 10 diabetic patients
表 1. 10 例糖尿病患者的数据资料

编号	X_1	X_2	X_3	X_4	X_5	Y
1	49	32.19	6.0	148	90	7.8
2	64	25.24	7.0	151	80	7.4
3	66	24.26	4.8	156	87	7.2
4	48	26.18	7.6	139	86	7.5
5	47	32.07	5.7	157	89	7.7
6	59	25.19	6.6	156	81	7.1
7	76	27.26	5.4	140	85	6.9
8	74	24.94	7.3	150	82	7.9
9	65	24.34	2.5	142	93	6.8
10	53	24.43	7.1	145	84	7.3

步骤 1 取 s 个独立的 $\{1, \dots, n\}$ 的随机置换 $\pi_1(1), \dots, \pi_j(n), \dots, \pi_s(n)$, $j=1, \dots, s$, 将它们作为列向量组成一个 $n \times s$ 设计矩阵, 称为超拉丁方设计(Latin hypercube design, LHD), 记为 $LHD(n, s)$, 它的第 k 行第 j 列的元素记为 $\pi_j(k)$ 。

步骤 2 取 $[0, 1]$ 上 ns 个均匀分布的独立抽样, $U_{ij} \sim U(0, 1)$, $i=1, \dots, n$, $j=1, \dots, s$ 。记 $\mathbf{x}_k = (x_{k1}, \dots, x_{ks})'$, 其中

$$x_{kj} = \frac{\pi_j(k) - U_{kj}}{n}, k=1, \dots, n; j=1, \dots, s. \quad (2.1)$$

则设计 $D = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ 即为一个 LHS 设计, 并记为 $LHS(n, s)$ 。

2.2. 由 10 例糖尿病患者数据资料生成超拉丁方设计

对抽样数据进行处理, 令 $x_{kj} = \frac{\pi_j k - \bar{X}_j}{n}$, 其中 $n=20$, $k=1, \dots, 20$, $j=1, \dots, 6$, 并且使 $Y = X_6$, \bar{X}_j 是组 X_j 数据的均值。

为了减小误差, 对六组数据的处理有所不同, X_1 组数据是每组的每个数据减去该组均值然后除以 2 得到一个新的 $[0, 10]$ 上的取值, X_2 、 X_3 、 X_4 、 X_5 四组数据是每组的每个数据减去该组均值然后得到一个新的 $[0, 10]$ 上的取值, X_6 (即 Y) 组数据是每组的每个数据减去该组均值然后乘以 10 得到一个新的 $[0, 10]$ 上的取值。其中 $\bar{X}_1 = 60.1$ 、 $\bar{X}_2 = 26.61$ 、 $\bar{X}_3 = 6$ 、 $\bar{X}_4 = 148.4$ 、 $\bar{X}_5 = 86.7$ 、 $\bar{Y} = \bar{X}_6 = 7.36$ 。由超拉丁方抽样的构造过程可知: 1) 它很容易产生; 2) 它可以处理试验次数 n 与因素个数 s 较大的问题; 3) 与完全随机抽样相比, 它估计 y 的样本均值的样本方差更小。而且它的表现比完全随机抽样稳定, 故在实际中大量使用, 能够很好的配合计算机试验, 帮助我们高速、有效地处理越来越复杂的问题。

表 2 给出了根据表 1 数据所得到的超拉丁方(LHS)设计。

3. 统计数据分析

对于表 2 的数据资料用 minitab 统计软件计算的主要结果见表 3 和表 4。

由上表 3 可见, $F = 2.72$, $P \leq 0.01$, 此回归方程有统计学意义。由上表 4 可见, 自变量 X_2 、 X_3 、 X_4 按 $\alpha = 0.05$ 水平有统计学意义, 但 X_1 和 X_5 无统计学意义。

Table 2. A Latin hypercube design based on data from 10 patients with diabetes mellitus**表 2.** 由 10 例糖尿病患者的数据资料生成的超拉丁方设计

编号	X_1	X_2	X_3	X_4	X_5	Y
1	5.55	5.58	0	0.4	3.3	4.4
2	1.95	1.37	1	2.6	6.7	0.4
3	2.95	2.35	1.2	7.6	0.3	1.6
4	6.05	0.43	1.6	9.4	0.7	1.4
5	6.55	5.46	0.3	8.6	2.3	3.4
6	0.55	1.42	0.6	6.6	5.7	2.6
7	7.95	0.65	0.6	8.4	1.7	4.6
8	6.95	1.67	1.3	1.6	4.7	5.4
9	2.45	2.27	3.5	6.4	6.3	5.6
10	3.55	2.18	1.1	3.4	2.7	0.6

Table 3. Variance analysis table of regression equation**表 3.** 回归方程的方差分析表

变异来源	自由度	SS	MS	F	p
回归	5	20.524	4.105	1.18	0.009448
残差误差	4	13.876	3.469		
合计	9	34.400			

Table 4. Estimation results of partial regression coefficients**表 4.** 偏回归系数估计结果

自变量	偏回归系数	系数标准误差	T	P
常数项	-0.01300	0.01712	-0.76	0.180
X_1	0.01119	0.01822	0.61	0.549
X_2	0.02565	0.03720	0.69	0.046
X_3	0.05274	0.03694	1.43	0.006
X_4	0.04701	0.01685	2.79	0.014
X_5	-0.00836	0.01723	-0.49	0.635

结果：体重指数、总胆固醇和收缩压对于糖尿病患者糖化血红蛋白的含量具有统计学意义上的显著影响，是影响糖尿病患者糖化血红蛋白含量的主要危险因素。

与前作参考文献[4]中的抽取 20 例糖尿病患者的数据资料用 SPSS 统计软件作多元线性回归分析进行比较，本文仅抽取 10 例糖尿病患者的数据资料，而且使用了超拉丁方抽样的方法对数据进行处理，得到更加科学、客观的一组全新数据，之后再作多元线性回归分析，仍然得出了与前作类似的符合医学生物学原理的结论。由此可见超拉丁方抽样方法的强大功能，仅仅根据少量数据也能够得到依靠大量数据才能够获得的统计结果。

结论：多个相关的研究显示，糖化血红蛋白可反映患者近 2 个月的平均血糖水平，研究发现其与糖尿病并发症的发生有关[6]；体重指数是目前评价肥胖的指标，肥胖患者普遍存在胰岛素抵抗、高胰岛素

血症、脂代谢异常、高血压等,是糖尿病慢性并发症的高危因素[7]。同时,国外有研究发现,收缩压为糖尿病肾病病变的独立危险因素;糖化血红蛋白还是糖尿病视网膜病变的主要危险因素;高胆固醇是糖尿病周围神经病变的主要危险因素[8]。

总之,体重指数、总胆固醇、收缩压和糖化血红蛋白是糖尿病慢性并发症发生的危险因素,特别是糖化血红蛋白含量更是糖尿病临床诊断的代表性指标,而且体重指数、总胆固醇、收缩压又是影响糖尿病患者糖化血红蛋白含量的主要危险因素,所以各危险因素相辅相成,时刻威胁着糖尿病患者的身体健康。除此之外,虽然影响糖尿病患者糖化血红蛋白(HbA1c)的主要危险因素我们已经找到,是体重指数、总胆固醇和收缩压,但是年龄、病程、高血压、高血糖、血脂紊乱也是影响糖化血红蛋白含量特别是糖尿病慢性并发症发生的危险因素。所以说严格控制所有的这些因素是提高糖尿病患者健康生活质量、预防和延缓糖尿病患者慢性并发症发生发展的重要措施,能最大限度地降低糖尿病对机体健康的危害。

基金项目

本项目由国家自然科学基金(基金号:11601201,11661076)支持。

参考文献 (References)

- [1] 张娟. 2008~2010年北京协和医院抗糖尿病药物使用统计分析[J]. 中国校医, 2012, 26(1): 598-600.
- [2] 杨婉思, 谢铨荣, 蔡巧玲. 1857例糖尿病住院病人分析[J]. 中国病案, 2006, 7(11): 40-41.
- [3] 王玲, 张卫华, 刘恒. 2型糖尿病住院患者慢性并发症患病率及相关因素分析[J]. 中国现代医生, 2014, 52(8): 31-33.
- [4] 马斌荣, 李康. 医学统计学[M]. 第5版. 北京: 人民卫生出版社, 2008: 105-106.
- [5] 方开泰, 刘民千, 周永道. 试验设计与建模[M]. 北京: 高等教育出版社, 2011: 207-209.
- [6] 刘蔚, 周翔海. 用糖化血红蛋白诊断糖尿病-WHO咨询报告[J]. 中国糖尿病杂志, 2011, 19(1): 2-5.
- [7] 王笠, 李琳, 王达, 等. 糖化血红蛋白的检测和临床应用[J]. 上海医学检验杂志, 2003, 18(2): 119-120.
- [8] 苏丽. 糖化血红蛋白与糖尿病及其并发症的相关性探讨[J]. 临床和实验医学杂志, 2011, 10(10): 771-772.

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