

基于二分类 Logistic 回归模型的低出生体重儿影响因素研究

杨 威

邵阳职业技术学院公共课部 湖南 邵阳 422000

【摘要】本文针对是否娩出低出生体重儿作为二分类变量，选取对低出生体重可能有影响的 8 个指标，运用二分类 Logistic 回归模型进行分析，得到是否患有高血压是对低出生体重影响最强的因素，其次是本次妊娠前早产次数，而产妇产体重在本模型中是一个保护因素，最后给出了相关解释及干预对策。

【关键词】Logistic 回归模型；二分类；SPSS；低出生体重儿

低出生体重儿，即出生体重小于 2500 克。研究表明，低出生体重儿的死亡风险是正常出生体重新生儿的 20 倍以上，是五岁以下儿童死亡的主要原因之一，且低出生体重与成人的许多疾病密切相关。因此低出生体重的发生率将作为监测儿童健康的主要指标之一^[1]。本文采用二分类 Logistic 回归模型研究低出生体重儿的影响因素，并制定相应的干预对策。

1 二分类 Logistic 回归模型原理

假设某效应结果指标 Y false 是一个二分类变量，取值为 $Y=1$ false 时表示此效应发生， $Y=0$ false 时表示不发生。影响效应结果 Y false 的因素称之为协变量（即自变量，又称为解释变量，可以是连续性变量，也可以是离散性变量），有 $x_1, x_2, \dots, x_m$ false 共 m 个，因变量取值为 1 的概率 $p(y=1|x)$ false 就是要研究的对象，称为二分类 Logistic 回归模型，这是广义线性模型，可以系统地应用线性模型方法^[2]。二分类 Logistic 回归模型为：

$$\ln \frac{p}{1-p} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m \quad (1)$$

其中 $1-p$ false 表示优势比（即事件发生概率与不发生概率的比值），待估参数 $\alpha, \beta_1, \beta_2, \dots, \beta_m$ false 反映了优势比的变化情况。如果 β_i false 是正的，它的指数值一定大于 1，则优势比率会增加；反之，如果 β_i false 是负的，则优势比率会减小^[2]。

概率 p false 的计算公式为：

$$p = \frac{\exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m)}{1 + \exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m)} \quad (2)$$

2 数据来源及变量筛选

本文数据来源于某医院的新生儿出生记录数据，结果变量为是否娩出低出生体重儿（low），希望筛选出低出生体重儿的影响因素，考虑的影响因素有 lwt、age、smoke、ptl、ht、ui、ftv、race 8 个变量，变量的说明见表 1 所示。将其中 189 个样本作为训练样本，以

是否娩出低出生体重儿为二分类变量，利用 SPSS 软件进行二分类的 Logistic 回归分析。

表 1 变量及其说明

变量	变量说明
low	是否娩出低出生体重儿，低出生体重儿记为 1，否则记为 0
lwt	产妇体重
smoke	产妇在妊娠期间是否吸烟，吸烟记为 1，否则记为 0
ptl	本次妊娠前早产次数
ht	是否患有高血压，患高血压记为 1，否则记为 0
ui	应激性，有应激性记为 1，否则记为 0
ftv	随访次数
race	种族，白人记为 1，黑人记为 2，其他种族记为 3

3 模型的建立与求解

起始块 0 是拟合一个只有常数项的模型，通常把它称为基线模型。表 2 输出的是在块 0 中对所有候选变量的比分检验结果，其中 race 产生两个哑变量，因此其总自由度为 2。由表 2 可以发现，本次妊娠前早产次数（ptl）的比分检验统计量最大，其 p 值为 0.007，小于 SPSS 软件中默认入选变量的标准 $p=0.05$ ，因此下一步将它先选入模型^[1]。

块 1 是真正建立的模型，标题“块 1：方法 = 向前步进（似然比）”说明将会给出向前法的结果。

表 3 模型汇总

-2 对数似然	考克斯 - 斯奈尔 R 方	内戈尔科 R 方
227.893	.035	.050
223.583	.057	.080
217.220	.088	.124

表 3 输出了每一步时的 $-2\ln L$ ，可用于进行似然比检验，还输出了 Cox&Snell R 方和 Nagelkerke R 方两种伪决定系数。由表 3 可知，在拟合过的三个模型中， $-2\ln L$ 是在逐步减小的，且伪决定系数也在逐步增大。

表 2 不在方程中的标量

	变量	得分	自由度	显著性
步骤 0	产妇在妊娠期间是否吸烟	4.924	1	.026
	产妇体重	4.616	1	.032
	产妇年龄	2.407	1	.121
	本次妊娠前早产次数	7.267	1	.007
	是否患有高血压	4.388	1	.036
	应激性	4.205	1	.040
	随访次数	.934	1	.334
	种族	5.005	2	.082
	种族 (1)	4.787	1	.029
	种族 (2)	1.727	1	.189
总体统计		29.140	9	.001

表 4 方程中的变量

		B	标准误差	瓦尔德	自由度	显著性	Exp (B)
步骤 1	本次妊娠前早产次数	.802	.317	6.391	1	.011	2.230
	常量	-.964	.175	30.370	1	.000	.381
步骤 2	本次妊娠前早产次数	.823	.318	6.683	1	.010	2.277
	是否患有高血压	1.272	.616	4.270	1	.039	3.569
	常量	-1.062	.184	33.224	1	.000	.346
步骤 3	产妇体重	-.015	.007	5.584	1	.018	.985
	本次妊娠前早产次数	.728	.327	4.961	1	.026	2.071
	是否患有高血压	1.789	.694	6.639	1	.010	5.986
	常量	.893	.829	1.158	1	.282	2.441

表 4 中给出的是每一步逐步回归得到的模型中参数估计及其标准误差、Wald 卡方等。由表 4 可知,第 3 个模型即是最终模型。步骤 3 得到的模型含有的自变量是本次妊娠前早产次数 (ptl)、是否患有高血压 (ht)、产妇体重 (lwt), 并且它们的 p 值都小于 0.05。表 4 中 B 这一列是常数项和自变量系数的估计值, Exp(B) 是优势比, 记作 OR, 它可以理解为当某个事件发生的概率较小时, 自变量每增加一个单位所导致的这个事件发生概率的变化倍数。在这个最后得到的模型中, 可以看到是否患有高血压 (ht) 的系数最大, 故是否患有高血压 (ht) 是影响低出生体重儿最强的因素, 在其他变量保持不变的情况下, 患高血压的产妇产出低出生体重儿的概率是不患高血压的产妇的 5.986 倍; 类似地, 本次妊娠前早产次数 (ptl) 每增加 1 次, 娩出低出生体重儿的概率是原来的 2.071 倍; 产妇体重 (lwt) 的 OR 值是 0.985<1, 说明产妇体重在此模型中是一个保护因素。

最终预测的模型如下:

$$p = \frac{\exp(0.893 - 0.015lwt + 0.728ptl + 1.789h)}{1 + \exp(0.893 - 0.015lwt + 0.728ptl + 1.789h)} \quad (3)$$

4 结论解释与干预对策

本文对低出生体重儿的影响因素进行了调查研究。从分析结果可以得出以下结论: 是否患有高血压是对低出生体重儿影响最强的因素, 这是因为妊娠期有高血压疾病的孕妇由于胎盘着床浅, 影响了胎盘供血和供氧的能力, 导致长时间处于宫内缺氧状态的胎儿出现宫内窘迫的情况, 且随妊娠时间延长, 胎盘酶活性逐渐降低, 影响胎儿的生长发育, 出现胎儿宫内发育迟缓, 最终导

致了低出生体重儿的发生^[4]。其次是本次妊娠前早产次数, 而产妇体重是一个保护因素。为此我们提出以下建议。

首先要提高早期建卡率, 重视高危筛查, 加强追踪随访, 医疗保健机构要对孕妇产前进行系统的全身体格检查, 定期测量孕妇的血压。其次, 要重视和加强健康教育, 普及孕期营养干预知识, 特别对于文化水平低的孕妇, 对妊娠期保健知识缺乏, 应做为重点普及对象, 进而提高其孕期保健意识。最终, 一旦分娩低出生体重儿, 必须重视和加强低出生体重儿的管理。加强低出生体重儿体温、呼吸、营养、预防感染等方面的护理, 针对母亲孕期的高危因素, 积极预防并发症的发生。如果出现并发症, 需采取积极的治疗手段进行治疗, 减轻并发症对低出生体重儿生命的威胁和减少致残后遗症的发生^[3]。

【参考文献】

- [1] 张文彤, 董伟. SPSS 统计分析高级教程 [M]. 北京: 高等教育出版社, 2019.
 - [2] 于海妹, 李宪波. Logistic 回归模型在北京市空气质量等级预测中的应用 [J]. 中国环境管理干部学院学报, 2018 (1).
 - [3] 李浩, 朱茂灵, 宁旦, 刘永新. 国内低出生体重儿发生影响因素及防治研究 [J]. 中国保健营养, 2020 (20).
 - [4] 李六兰, 王蕾. 新生儿低出生体重的影响因素分析及护理干预对策 [J]. 护理实践与研究, 2020 (3).
- 作者简介: 杨威, 理学硕士, 邵阳职业技术学院公共课部讲师, 研究方向: 机器学习、大数据分析 with 统计建模。