

血氧饱和度水平的影响因素及实证分析

杨佳鹏 俎毓伟 潘旭阳

华北理工大学理学院 河北省唐山市 063210

【摘要】脉搏血氧计通常用于监测患者的血氧饱和度水平，在连续监测期间，最好使用模型来描述血氧饱和度的模式。本文以血氧饱和度的影响因素为研究内容，运用实证分析对年龄，BMI，性别，吸烟史或当前吸烟状况及血氧饱和度水平进行多元线性回归分析。我们选择对36名测试人员的相关信息进行分析。首先，建立多元线性回归模型，并通过异方差和多重共线性进行检验，并对预测结果进行改进。借助三维空间图像对各个相关变量进行进一步的分析讨论，最终得出结论：随着年龄的增长不会降低或增加血氧饱和度，即年龄与血氧饱和度几乎无关。继续分析其他变量与年龄的关系，发现只有BMI与年龄有关，即随着年龄的增长，人们的BMI会有很小程度的下降。最后，对各模型的可行性和准确性进行客观评价。

【关键词】多元线性回归；BP神经网络；逐步回归；血氧饱和度

1 引言

为了弄清血氧饱和度问题可变性的起源，下面的背景值得一提。

1.1 血氧饱和度

人体的代谢是一个生物氧化过程，代谢过程中所需要的氧气，通过呼吸系统进入人体血液，与血液中的红细胞血红蛋白(Hb)，结合成氧合血红蛋白(HbO₂)，然后输送到人体组织细胞的各个部位。通过血氧饱和度来测量血液携带氧气的的能力，这被认为与患者的血氧含量密切相关。

$$SaO_2 = \frac{c_{HbO_2}}{c_{HbO_2} + c_{Hb}} \times 100\%$$

c_{HbO_2} 、 c_{Hb} 是血液中氧化血红蛋白和还原血红蛋白的浓度。

1.2 血氧饱和度变异性的意义

血液的血氧饱和度与呼吸道畅通和肺部状况密切相关。一般血液的血氧饱和度至少为95%，甚至98%~100%。但当气道堵塞或影响肺部的疾病，如睡眠呼吸暂停、哮喘、肺炎或纤维化时，血氧饱和度可以降低。血氧饱和度过低，可能有气喘，气短等症状，需要吸氧治疗。因此，血氧饱和度的变异性数据可以作为相关疾病严重程度的指标。

1.3 研究问题

脉搏血氧计通常测量病人的血氧饱和度水平。在连续监测期间，我们希望能够使用该模型来描述血氧饱和度的模式。我们有来自36个人的数据，每个受试者在1Hz的频率下进行了大约一个小时的血氧饱和度测

试。我们还记录了以下有关参与者的信息，包括年龄、BMI、性别、吸烟史和当前吸烟状况，以及可能影响结果的任何重要医疗条件。

问题1：我们想用这些数据来找到氧饱和度变化的典型模式，这样我们就可以用几个参数来描述一个人。

问题2：我们想知道血氧饱和度序列的模式是否与年龄有关，以及老年人与年轻人相比有什么特点发生了变化。

2 问题的描述

2.1 如何建立模型的 SaO₂ 可变性?

这个问题是一个常见的多元回归分析问题。我们可以利用36个已知个体的数据进行分析，首先采用多元线性回归分析，然后结合异方差检验和多重共线性检验来判断多元线性回归方程的拟合效果。

在数据处理方面，我们需要对现有的数据进行预处理。部分字符采用热编码方式加工成数字，便于后期编码。

2.2 如何确定 SaO₂ 是否与年龄有关?

采用逐步回归方法确定血氧饱和度序列模型与年龄相关变量之间的关系，达到预期效果。在实际应用过程中，对无量纲数据进行处理。

也需要重点分析老年人随着时间的推移的生物学或医学特征，一些变量会影响最终的血氧饱和度序列模式。

3 假设与模型

3.1 基本模式

3.1.1 符号

符号定义	符号描述
c_{HbO_2}	血液中氧血红蛋白的浓度
c_{Hb}	血液中血红蛋白浓度降低
$X_0 \sim X_7$	受试者的特征
$\beta_0 \sim \beta_7$	回归参数
ε	随机误差项

3.1.2 假设

●假设研究使用了准确的数据。

原因：我们假设患者已经照顾了指尖皮肤，在连续脉搏血氧测定时，选择了一种适合患者皮肤状况的脉搏检测探头。在监测过程中避免了探头压迫引起的局部皮肤损伤。此外，受试者在测量血氧饱和度之前没有进行剧烈运动，避免了所有其他影响测量数据的客观因素。

●假设 36 名受试者处于相同的身体状况，没有其他疾病影响结果。

原因：由于模型研究影响氧饱和度的因素，因此有必要确保其他外部条件是一致的，以便进行有意义的比较。

●假设一个人的血氧饱和度水平是可以预测的。

原因：虽然每个人身体的血氧饱和度不一定相似，但一般人的身体水平在一定范围内略有波动。

3.1.3 模型的基础

这个问题是一个常见的多元回归分析问题。在回归分析中，如果有两个或两个以上的自变量，称为多元回归。事实上，一个现象往往与多个因素有关，通过多个自变量的最优组合联合预测或估计因变量比只预测或估计一个自变量更有效和实用。

$$Y_i = \beta_0 + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \dots + \beta_k * X_{ki} + \varepsilon_i$$

Y 被称为因变量或解释变量，是解释变量，是随机扰动，是前者的参数，表达式由一个单位的变化引起的变化程度。

$$X_k \in \beta_k X_k X_k Y_i$$

为了研究血氧饱和度的变异性，我们使用年龄、吸烟史和 BMI 作为解释变量。建立多元回归模型，建立的函数模型为：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + u$$

利用 Python 进行线性拟合，可以得到如下回归方程：

$$Y = 100.11773 - 0.11352X_1 + 0.01218X_2 - 0.20303X_3 + 0.20303X_4 - 0.78399X_5 - 0.20640X_6 + 0.99039079X_7 + 0.581167$$

3.1.4 解决方案和结果

根据建立的模型，我们使用 Python 对模型进行拟合。通过测试者的 BMI、年龄、性别、吸烟史和当前吸烟状

况等信息被用来预测血氧饱和度。预测结果如图 1 所示：

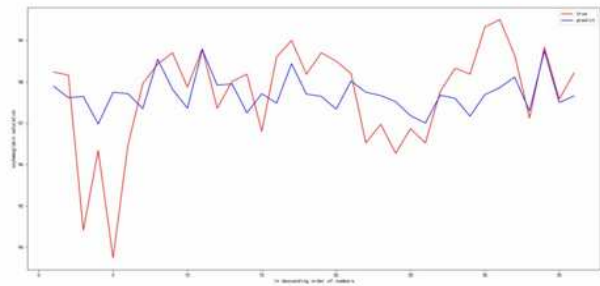


图 1 多元线性回归 - 血氧饱和度预测

图中，横轴表示测试的序号，纵轴表示测试仪一小时内连续测量血氧饱和度的平均值，红色折线表示测试仪实际测量的血氧饱和度值，蓝色折线表示测试仪预测的血氧饱和度值。通过观察图 1，我们发现上述建立的多元线性回归模型可以很好地预测测试人员血氧饱和度的波动趋势，但波动程度无法准确预测。

我们用 Python 绘制了自变量为年龄和 BMI 的三维坐标图，因变量为血氧饱和度。三角形表示已经吸烟的人，圆形表示从未吸烟的人，正方形表示目前正在吸烟的人，M 表示该人是男性，F 表示该人是女性。绘制的三维空间坐标图如图 2 所示：

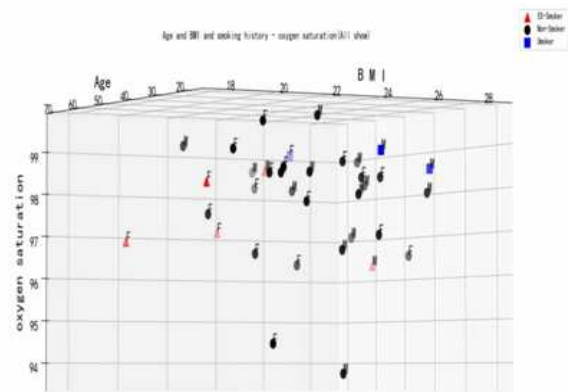


图 2 年龄和 BMI 的三维坐标

通过观察图像，我们发现大多数受试者的血氧饱和度和：

图 2 受试者血氧饱和度的变化正常，只有两名受试者的血氧饱和度较低。我们还发现，吸烟或吸烟史似乎对血氧饱和度没有太大影响。

通过旋转三维空间图像，截取了表示 BMI 的横坐标，表示血液血氧饱和度横截面的纵坐标，并以 BMI 为自变量，测试仪的血氧饱和度为因变量。我们观察到，氧饱和度正常的女性的 BMI 普遍低于男性的 BMI。如下图 3 所示：

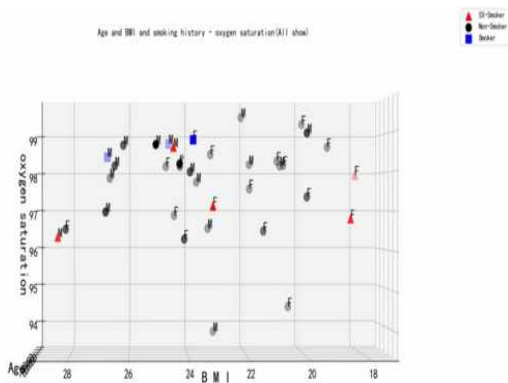


图3 测试对象 BMI 的变化

利用 Python 对 36 个已知测试人员的数据和相关信息作为测试数据进行预测,并基于原始坐标点绘制预测面。通过观察,我们发现大部分数据点都附着在预测面上,因此预测面可以更好地预测血氧饱和度的变异性。如下图 4 所示:

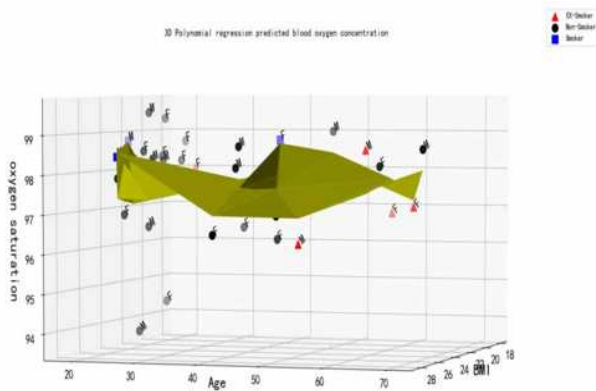


图4 多元线性回归 - 血氧饱和度的三维预测

3.1.4 型号检查

● 异方差性检验

如果存在异方差,则会导致 T 检验失败,破坏模型的预测函数。为了检验随机误差线方差是否随解释变量观测值的变化而变化,这里选择 White 检验,以确定是否存在异方差。根据检验结果, P 值大于 0.05,接受零假设。该模型不存在异方差。

● 多重共线性检查

多重共线性是指线性回归模型中解释变量之间由于精确的相关性或高度的相关性而产生的失真或估计困难。即当变量之间存在多重共线性时,参数估计的经济意义将是不合理的,因此这里进行多重共线性检验。我们通过 Python 绘制了系数之间的相关热图,从图 5 可以看出,变量之间没有严重的多重共线性。

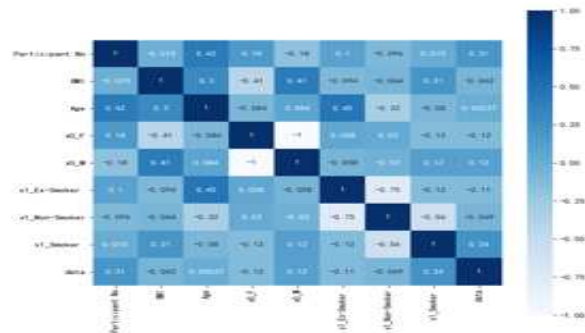


图5 系数相关热图

3.2 结果分析

通过观察表 1,我们发现血氧饱和度与年龄关系的权重为 0.000373,影响非常,非常小。同时,也可以看出 BMI 与年龄关系的权重为 0.3043,表明 BMI 特征与自变量年龄之间存在明显的关系。

通过对图 6 的观察,我们发现随着年龄的增长,血氧饱和度没有呈现下降的趋势。因此,老年人的血氧饱和度不一定低于年轻人,甚至可能高于年轻人。通过看图 7,我们发现 BMI 随着年龄的增长而降低,但不显著,因此老年人的 BMI 可能低于年轻人。

4 结论

通过建立多元回归模型以及观察三要素图像,并根据每个变量的权重,我们可以得出结论,随着年龄的增长,人们的血氧饱和度不会受到很大的影响。也可以说,老年人的血氧饱和度可能不低于年轻人的血氧饱和度。

不过我们也利用过梯度回归树模型进行建模,预测精度会大幅提高。最后得出结论:随着年龄的增长,血氧饱和度不呈现下降或上升的趋势, BMI 会随着年龄的增长出现小程度的降低。

【参考文献】

- [1] 血氧饱和度 (2007.11):<https://baike.baidu.com/item/%E8%A1%80%E6%B0%A7%E9%A5%B1%E5%92%8C%E5%BA%A6/6567714?fromtitle=%E6%B0%A7%E9%A5%B1%E5%92%8C%E5%BA%A6&fromid=8596210&fr=aladdin>
- [2] 苍兰阁 .(2020.4.21)<https://www.bilibili.com/video/BV1qt4y127BN?from=search&seid=1530300084272032557>.
- [3] 多元线性回归,多重共线性检验 (2020.4.5)<https://blog.csdn.net/u011517132/article/details/105323135>
- [4] 多元线性回归 (2017.5)<https://baike.baidu.com/item/%E5%A4%9A%E5%85%83%E5%9B%9E%E5%BD%92%E6%A8%A1%E5%9E%8B/4425841?fr=aladdin>
- [5] 刘洪山,王卫星,孙道宗,高鹏,陈文斌,林楚婷,谢家兴.基于 GA-BP 神经网络的果园需水量预测 [J/OL].排灌机械工程学报,2020.12.06:1-6. <http://kns.cnki.net/kcms/detail/32.1814.TH.20201202.1727.016.html>. BP 神经网络
- [6] 尚华娟.逐步回归分析.(2009.1)<https://baike.baidu.com/item/%E9%80%90%E6%AD%A5%E5%9B%9E%E5%BD%92/585832?fr=aladdin>