

# 基于线性回归模型对大学生心理数据的研究

安 彤

北京建筑大学 北京 102600

**摘 要:** 近年来大学生心理健康问题约来越发受到重视,有调查显示大学生的日常行为与心理状况有关系。根据某校学生SCL-90的心理自评测试表给出的打分指标,以及学生的指标评分,通过聚类分析法设定等级区分心理健康危险程度,结合学生的消费和图书馆刷卡次数,建立多元线性回归模型进行分析,判断学生的日常行为是否跟心理状况有关。然后,基于分析结果给图书馆刷卡总次数、学生消费总次数、学生消费总金额进行权重赋值,与原有的十三个指标合并计算,得出新等级评定方式,同时与学生真实的心理评级数据作对比分析,检验模型。最后,考虑到学生进出图书馆次数、消费次数和消费金额对学生心理状况影响的权重很小,需要考虑更多因素。基于所得结果给出更科学合理有效的心理测评方案,为大学生心理测评提供更科学、更多元化参考方案。

**关键词:** 心理健康; 聚类分析; 线性回归模型; 心理测评

## Study on the Psychological Data of College Students Based on the Linear Regression Model

Tong An

Beijing University of Civil Engineering and Architecture, Beijing 102600

**Abstract:** In recent years, the mental health problems of college students have been paid more and more attention. Some surveys show that the daily behavior of college students is related to their psychological status. According to one school students' SCL-90 psychological self-evaluation test given scoring indicators, and the students' score, through clustering analysis set level to distinguish mental health risk, combined with the students' consumption data and the number of visits to the library, establish multiple linear regression model analysis, determine whether the students' daily behavior is related to psychological condition. Then, based on the analysis results, the total number of visits to the library, the total number of students consumed and the total amount of students consumed are weighted and calculated with the original 13 indicators to get the new rating method. At the same time, we compare with the real psychological rating data of students to test the model. Finally, considering that the number of visits to the library, consumption times and consumption amount affect students' psychological status very slightly, more factors need to be considered. Based on the obtained results, a more scientific, reasonable and effective psychological evaluation program is given to provide a more scientific and diversified reference scheme for the psychological evaluation of college students.

**Keywords:** Mental health; Cluster analysis; Linear regression model; Psychological assessment

### 1. 概述

近年来,由心理问题所引发的大学生自伤、自杀事件日益增加,大学生作为国家重要的储备人才,心理危机事件给危机学生本人及其家庭、社会带来重大的损失。为预防大学生心理危机事件的发生,各高校

陆续开展了对新生或全体学生的心理健康状况普查工作,在心理普查实际工作中会产生大量的、多样化的数据,通过挖掘和利用这些数据,进而改进心理健康教育工作、提高学生心理健康水平、提升高校人才培养质量,具有极强的现实意义。本文选取了某校学生心理数据、在校消费的部分数据以及进入图书馆的部分刷卡次数数据。通过建立线性回归模型进行分析并得出结论。

**作者简介:** 安彤(2000.9-),女,汉族,北京人,北京建筑大学,本科在读,研究方向:数学建模。

## 2. 模型假设

2.1 对于所有心理状况的分级全部假设分为四级（正常、轻度、中度、重度），并赋值为0、-1、1、2。

2.2 对于数据中的冗余数据进行处理，如进出图书馆两次时间间隔小于1分钟应归并一起，算为1次。

2.3 对于指标14设置为整数。

## 3. 模型建立与求解

### 3.1 根据学生的心理数据进行等级评分

根据学生SCI-90得分以及评分标准进行数据分析，采用SPSS软件进行迭代求解的聚类分析算法——K-均值聚类进行分级。

将数据导入SPSS，进行分类--K-均值聚类操作，设置最大迭代次数为20，根据模型假设将所有的数据分成4类，即将K设定为4。得到初始聚类中心，经过11次迭代，此时聚类中心发生变化均为0，得到最终的聚类中心，并将每个成员变量所属的分级导出。

由最终聚类中心可得，将数据分为四类，聚类3对应正常，根据模型假设的设定值对应为0；聚类1对应为轻度，根据模型假设的设定值对应为-1；聚类4对应中度，根据模型假设的设定值对应为1；聚类2为重度，根据模型假设的设定值对应为2。

### 3.2 判断学生的日常行为与心理状况是否有关

为研究学生的日常行为是否跟心理状况有关，根据提供的数据图书馆刷卡次数、消费次数、消费金额，建立线性回归模型，分析处理因变量学生心理状况与自变量学生进入图书馆的刷卡数据、消费数据、消费次数的关系。线性回归模型记为：

$$\begin{cases} y = \beta_0 + \beta_1 x_i + \varepsilon \\ E(\varepsilon) = 0, D(\varepsilon) = \sigma^2 \end{cases}$$

其中y表示心理状况等级， $x_i$ 影响心理状况因素（图书馆刷卡总次数、消费总次数、消费总金额）， $\beta_0$ 为常数项， $\varepsilon$ 为剩余误差。

3.2.1 分析学生进入图书馆刷卡总次数与心理状况是否有关

根据学生的真实心理状况对正常、轻度、中度、重度进行整合，分别为0、-1、1、2。结合学生进入图书馆的部分刷卡数据，对数据重新整理，利用MATLAB中线性回归语句进行模型建立，得到结果如下。

$\widehat{\beta}_0=0$ ,  $\widehat{\beta}_1=2.7496e-0.5$ ,  $\widehat{\beta}_0$ 的置信区间为[1.0e-0.3,  $\infty$ ],  $\widehat{\beta}_1$ 的置信区间为[-0.4661, 0.5211], 可知回归模型 $y=(2.7496e-0.5)x_1$ 成立。

即心理状况与图书馆刷卡次数为线性关系，刷卡次数越多相对心理状况评级越高。

### 3.2.2 分析学生消费总次数与心理状况是否有关

根据学生消费数据，计算得出消费总次数，对数据

重新编辑，用MATLAB中线性回归语句进行模型建立，得到结果如下。

$\widehat{\beta}_0=0$ ,  $\widehat{\beta}_1=0.0026$ ,  $\widehat{\beta}_0$ 的置信区间为[0.0021, 0.0032],  $\widehat{\beta}_1$ 的置信区间为[ $-\infty, +\infty$ ], 可知回归模型 $y=0.0026x_2$ 成立。

从结果可以得到学生消费总次数与心理状况呈微弱线性关系。

### 3.2.3 分析学生消费总金额与心理状况是否有关

根据学生消费的数据，计算得出消费总金额，对数据重新编辑得到结果。利用MATLAB进行编写回归分析程序得到的结果如下。

$\widehat{\beta}_0=0$ ,  $\widehat{\beta}_1=3.1882e-0.4$ ,  $\widehat{\beta}_0$ 的置信区间为[1.0e-0.3,  $\infty$ ],  $\widehat{\beta}_1$ 的置信区间为[0.2507, 0.3869], 可知回归模型 $y=(3.1882e-0.4)x_3$ 成立。

即心理状况与学生消费总金额呈线性关系，消费总金额越大相对心理状况评级越高。

### 3.3 给出新的等级评定方案

根据前两小节的分级设定，参考原始的计算方法，设定一个新的指标14，

$$\text{指标 } 14 = \frac{\text{加权消费次数} + \text{加权消费金额} + \text{加权去图书馆次数}}{3}$$

取上节MATLAB线性回归结果中的R方作为权重占比。根据数据计算得出指标14的值，结合原始的13个指标再次进行聚类分析，设置最大迭代次数为20，根据模型假设将所有的数据分成4类，即将K设定为4。得到初始聚类中心经过11次迭代后，此时聚类中心发生变化均为0，得到最终的聚类中心，并将每个成员变量所属的分级导出。

由最终聚类中心可得，将数据分为四类，聚类1对应正常，根据模型假设的设定值对应为0；聚类4对应为轻度，根据模型假设的设定值对应为-1；聚类2对应中度，根据模型假设的设定值对应为1；聚类3为重度，根据模型假设的设定值对应为2。

### 3.4 设计更科学合理有效的心理测评方案

根据建立的回归模型发现，学生在图书馆的刷卡次数，消费次数，消费金额总数与分级后学生们的心理测评状况的（从正常到严重）成线性关系，即：学生去图书馆次数越多的，心理状况评级越高，心理状况越不正常，消费次数，消费金额总数也是如此。虽然回归分析后我们发现相关系数普遍偏低，但也可能引起小幅度心理变化，因此给三种影响因素赋予占比很低的权重，结合原有的SCI-90十三个指标的数据添加一个指标——“加权消费次数+加权消费金额+加权去图书馆次数分数和除以3”，一共14个指标数作为新的评级方案。

同时由于消费金额、消费次数以及进入图书馆刷卡

数据对学生心理状况影响较小,因此建议增加更多影响因素,比如睡眠时间、体育锻炼时间,家庭结构、教育方式、家庭环境等对学生心理状况的影响。在这互联网飞速发展的时代,网络也是影响学生的一大重要因素,由此可以对平时网上浏览的信息,短视频观看的类型以及所接触游玩的游戏类型的数据进行收集重新进行心理等级分级。让这个评测方案更科学以及安全。

#### 4. 模型检验

4.1 针对心理状况与学生进入图书馆刷卡总次数的线性模型进行残差分析

根据残差分析结果可知,除第1、3、7、25、2、6点外,其余数据的参差离零点均较近,且参差的置信区间均包含零点,说明学生进入图书馆刷卡总次数与心理状况线性回归模型: $y = (2.7496e-0.5)x_1$ ,能较好地符合原始数据,模型成立。

4.2 针对心理状况与学生消费总次数的线性模型进行残差分析

根据残差分析结果可知,数据的参差离零点均较近,且参差的置信区间均包含零点,这说明学生消费总次数与心理状况线性回归模型 $y = 0.0026x_2$ 能较好地符合原始数据,无异常点,模型成立。

4.3 针对心理状况与学生消费总金额的线性模型进行残差分析

根据残差分析结果可知,数据的参差离零点均较近,且参差的置信区间均包含零点,这说明心理状况与学生消费总金额的线性模型回归模型 $y = (3.1882e-0.4)x_3$ 能较好地符合原始数据,无异常点,模型成立。

#### 4.4 针对新的等级评定方案结果与原始结果对比

将4.3中最后所得聚类结果与学生真实的心理评级数据进行对比,评级基本相同,其中有7个编号分级与学生真实的心理评级数据存在的误差较大。综上所述,根据对比结果可得,新的分级与原有分级差别不大,仅有少数数据存在误差较大。说明模型的合理性较强。

#### 5. 模型总结与评价

根据多元线性回归分析,得出学生心理状况与消费次数、消费数据、图书馆刷卡数据成线性关系,具有相互影响的作用,影响较小。根据4.3提出的新的等级评级方案与真实学生心理状况可得,新的分级与原有分级存在少数数据有误差。新的方案对于三种影响因素赋予占比很低的权重,在原有的十三个指标上添加一个指标——“加权消费次数+加权消费金额+加权去图书馆次数分数求和除以3”,一共14个指标。

从本例可以看出消费金额、消费次数以及进入图书馆刷卡数据对学生心理状况影响较小,为了更好的通过学生的行为习惯去衡量心理状况这个指标14还可以增加一些其他的数据,例如,学生在学校内上网的数据、平

时学生的学习数据、学生在校内的锻炼数据、学生在校内的心理咨询情况等等。通过增加这些数据去多维的衡量学生的行为习惯这一指标,使通过指标衡量出来的结果能够更加的准确,提高精准度。

#### 6. 模型的推广

本文所设计的心理评级模型主要针对的是大学生群体,同样,也可推广至高中生,中小學生,乃至成年人。对于高中生来说指标14可以涉及到平时校内的借书记录、参与社团情况等等。对于中小學生来说指标14涉及到的因素可以有学生的平时成绩,学生的兴趣爱好等等。而对于成年人来说,以公司为例,指标14可以涉及到平时工作效率,平时的上下班时间等等。

对于心理状况分级模型,可以推广应用于质量标准等级划分,如对器械、果蔬等质量检验,筛选出合格产品;在制定规则方面,通过该模型,对汽车车速、刹车灵敏检验、酒精检测仪灵敏度等因素对酒驾标准重新定义等。

对于本文中所使用的基础模型分析方法也可推广至其他方面。对于聚类分析方法来说,这是一种无监督学习,通过数据和一定的标准进行分类,可以用“物以类聚”来通俗定义。例如企业在进行客户画像的时候,可以根据客户的年龄、职业、收入、消费水平、兴趣爱好等等方面进行数据挖掘,通过所获的数据进行聚类分析,即把客户中具有相似特征的分为一组,不相似的再分为一组,从而得到精准的客户细分。

对于线性回归分析来说,这是一种应用极为广泛的数量分析法,用于分析事物之间的统计关系,侧重考察变量之间的数量变化规律。回归分析是通过规定因变量和自变量来确定变量之间的因果关系,帮助人们确定一个变量对于另一个变量的影响程度,对于合理的回归模型也可以用来预测数据,为预测数据提供精确的标准。例如,在分析房屋单价的影响因素时就可以采用多元线性回归的方法进行分析,进而得到各变量对于房屋单价的影响程度,进而建立多元回归方程,若方程的拟合程度好,基本不存在多重共线性情况,不存在异方差情况等,还可以拟合出来的多元线性回归方程进行预测。

#### 参考文献:

- [1]赵静,但琦.数学建模与数学实验[M].高等教育出版社,2008.
- [2]孙永明,杨进.基于BSTL与XGDT算法对多级别心理压力的评估[J].经济数学,2020,37(4):11.
- [3]狄玉峰.运用SCL-90测评大学生心理健康状况的调查研究[J].漯河职业技术学院学报,2010,9(02):141-142.
- [4]薛薇.基于SPSS的数据分析[M].中国人民大学出版社,2014.