

# Hierarchical Linear Regression Analysis of Achievements in Advanced Mathematics

Jian Qiao, Junjie Du, Zaixing Li

China University of Mining and Technology (Beijing), Beijing  
Email: 18801360907@163.com

Received: Oct. 1<sup>st</sup>, 2019; accepted: Oct. 14<sup>th</sup>, 2019; published: Oct. 21<sup>st</sup>, 2019

---

## Abstract

In the analysis of the higher mathematics scores of a 211 college students, the innovative ideas start from the relationship of students nesting in the teacher, instead of adopting the usual practice of students nesting in the class, using a two-level hierarchical linear model. Analysis has broken the limitations of ordinary regression models in analyzing data with nested structures. At the individual level, students' higher mathematics grades are mainly influenced by their math scores and gender in college entrance examinations. Among them, the math scores of college entrance examinations have positive effects on higher mathematics scores, and the average scores of female students are higher than those of male students. At the teacher level, students' higher mathematics scores are mainly influenced by the gender of the teacher. The average number of students taught by male teachers is higher than that of the female teachers.

## Keywords

Hierarchical Linear Model, Higher Mathematics Scores, Data with Nested Structure, Individual Level, Group Level

---

# 高等数学成绩的分层线性回归分析

乔 舰, 杜俊洁, 李再兴

中国矿业大学(北京), 北京  
Email: 18801360907@163.com

收稿日期: 2019年10月1日; 录用日期: 2019年10月14日; 发布日期: 2019年10月21日

---

## 摘 要

在对某211高校学生的高等数学成绩进行分析时, 创新思路从学生嵌套于老师这一关系入手, 而没有采用学生嵌套于班级这一通常做法, 采用两水平的分层线性模型对其进行分析, 打破了普通回归模型在分

析具有嵌套结构数据的局限性。在个体水平上, 学生的高等成绩主要受其高考数学成绩和性别的影响, 其中高考数学成绩对高数成绩具有正效应, 女同学的高数成绩平均高于男同学; 在老师水平上, 学生的高等数学成绩主要受老师性别的影响, 男老师所教授学生的高数成绩平均高于女老师所教授的学生。

## 关键词

分层线性模型, 高等数学成绩, 具有嵌套结构的数据, 个体水平, 组群水平

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

高等数学作为大学几乎所有专业的一门核心基础课, 它的作用可谓是举足轻重。它不仅可以为其他课程的学习打下基础, 还可以拓宽学生的思维, 培养学生的逻辑能力, 因此分析学生的高等数学成绩对大学制定学生培养计划至关重要。但是学生的数学学习能力不是天生就具备的, 学生的高等数学成绩不仅受学生个体自身变量的影响, 比如学生的智力水平、努力程度和其高考数学成绩等, 也可能受任课老师的影响, 一位同学被不同老师教授可能其的高数成绩会有所不同。

此时学生的高数成绩为具有嵌套结构的数据。学生与学生之间本来并无关系, 但是当被同一位任课老师所教授后, 他们之间便存在了某种联系, 这种关系被叫做组内同质, 而不同任课老师所教授的学生的成绩具有异质性。具有嵌套结构的数据不能满足普通回归分析所要求的独立性与方差齐性[1], 此时用普通回归处理具有嵌套结构的数据会产生错误, 为了处理这一问题 Lindley 和 Smith 于 1972 年首先提出了分层线性模型[2]。本文基于分层线性模型的建立引入了可能对学生高等数学成绩有影响的变量, 通过逐次引入并检验这些变量来定量分析各变量对学生高等数学成绩的影响。

## 2. 模型介绍

在此以两水平分层线性模型为例,

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \varepsilon_{ij} \quad (2.1)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}z_{1j} + e_{0j} \quad (2.2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}z_{1j} + e_{1j} \quad (2.3)$$

其中  $y_{ij}$  表示在第  $j$  组中的第  $i$  个个体的水平 1 结局测量; 其中  $i=1,2,\dots,N$  ( $N$  是总样本量),  $j=1,2,\dots,J$  ( $J$  是水平 2 的单位数)。式(2.1)是分析  $y_{ij}$  变异的水平 1 方程。水平 1 截距  $\beta_{0j}$  中的下标  $j$  表示水平 1 截距跨水平 2 单位变化, 水平 1 回归系数  $\beta_{1j}$  中的下标  $j$  表示水平 1 回归系数跨水平 2 单位变化。水平 1 变量  $x_{ij}$  对  $y_{ij}$  的效应随水平 2 单位而变化。与水平 1 随机回归系数  $\beta_{0j}$  和  $\beta_{1j}$  相对应的是两个水平 2 方程, 在这些方程中水平 1 随机回归系数变成了因变量。此时分层线性模型能够评估个体水平的解释变量与结局测量是如何受组群水平变量调节的, 这就是分层线性模型的核心思想。在这里模型还假设水平 1 和水平 2 残差符合正态分布, 而且水平 1 残差与水平 2 残差相互独立, 这些假设可表述为:

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$$\begin{bmatrix} e_{0j} \\ e_{1j} \end{bmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{e0}^2 & \sigma_{e01}^2 \\ \sigma_{e01}^2 & \sigma_{e1}^2 \end{pmatrix} \right]$$

$$Cov(\varepsilon_{ij}, e_{0j}) = 0, Cov(\varepsilon_{ij}, e_{1j}) = 0$$

### 3. 学生高等数学成绩建模分析

本文研究取自某 211 学校 2014 级不同学院两个学期的平均高数成绩和学生与任课老师的一些基本资料。根据一些常识与变量之间的相关性分析，本文初始选定自变量：学生的性别(标注为 *sex*)、学生依老师中心化后的高考数学成绩(标注为 *cgaokao*)、任课老师的性别(标注为 *xb*)、任课老师的教龄(标注为 *jiaoling*)、任课老师的职称(标注为 *zc*)，因变量为学生的高等数学成绩(标注为 *gaoshu*)。现得到学生依老师中心化后的高等数学成绩与高考数学成绩的关系如图 1：

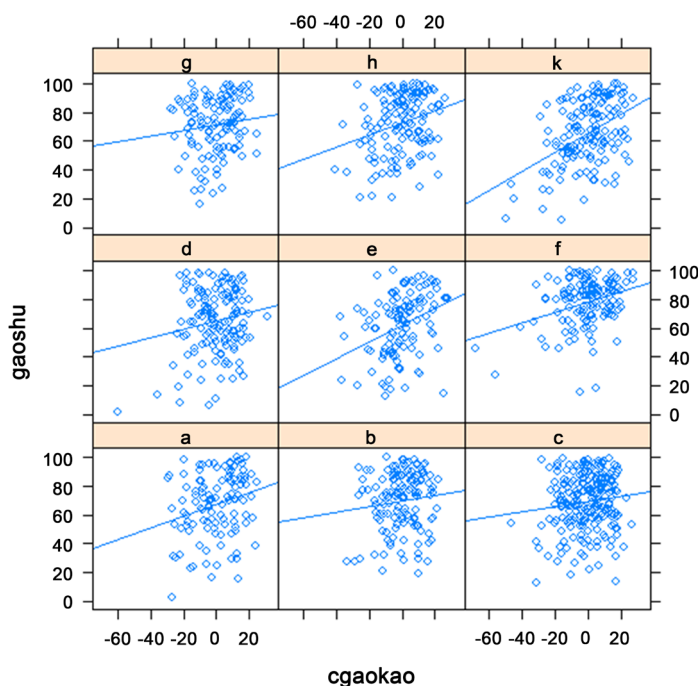


Figure 1. Regression analysis between the high scores of students in a 211 university and the math scores (centered) of the college entrance examination

图 1. 某 211 大学学生的高数成绩与学生高考数学成绩(已中心化)之间的回归分析图

由图可知各位老师所教授学生的高数成绩与高考数学成绩之间的回归关系有显著不同，因此可初步得出运用分层线性模型是合理的。

1) 现在建立空模型，它是最简单的随机效应模型：

$$gaoshu_{ij} = \beta_{0j} + \varepsilon_{ij} \tag{3.1}$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{3.2}$$

其中： $gaoshu_{ij}$ ：第  $j$  位老师所教授的第  $i$  位学生的高数成绩。 $u_{0j}$ ：第  $j$  位老师所教授学生的高数成绩均值与总均值  $\gamma_{00}$  之间的差异。 $\varepsilon_{ij}$ ：每位学生的高数成绩与老师所教授学生的高数成绩均值之间的差异。

根据检验得老师这一第 2 水平解释变量对学生的高考数学成绩与高数成绩之间的关系有显著影响，此时使用分层线性模型是合理的。

此时进行协方差参数估计得：水平 1 随机截距方差估计为  $\hat{\sigma}_{u_0}^2 = 18.34$ 、水平 1 残差方差估计为  $\hat{\sigma}^2 = 416.86$ 。组内相关系数  $ICC = 0.04$ ，即学生的高数成绩中大约有 4% 的变异是由老师的不同引起的，每位老师所教授学生的高数成绩均值有显著不同。下面建立带宏观解释变量的随机截距模型。

2) 带宏观解释变量的随机截距模型：

$$gaoshu_{ij} = \beta_{0j} + \varepsilon_{ij} \quad (3.3)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}jiaoling_j + \gamma_{02}xb_j + \gamma_{03}zc_j + u_{0j} \quad (3.4)$$

其中： $jiaoling_j$ ：第  $j$  位老师的教龄。 $xb_j$ ：第  $j$  位老师的性别，为虚拟变量， $xb = 1$  表示为男性， $xb = 0$  表示为女性。 $zc_j$ ：第  $j$  位老师的职称，为虚拟变量， $zc = 1$  表示为副教授， $zc = 0$  表示为讲师。

经过回归系数的显著性检验，老师的性别是显著的，而教龄与职称均不显著，因此最终将变量  $xb$  引入模型。得到最终的分层线性模型：

$$gaoshu_{ij} = \beta_{0j} + \varepsilon_{ij} \quad (3.5)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}xb_j + u_{0j} \quad (3.6)$$

此时进行协方差估计得，模型中组内方差的估计值为  $\hat{\sigma}^2 = 416.86$ ，与空模型中组内方差的估计值  $\hat{\sigma}^2 = 416.86$  基本相同。组间方差的估计值为却从 18.34 下降到 6.95。这说明了水平 2 解释变量  $xb$  能解释组间变异，但不能解释组内变异。此时组内相关系数  $ICC = 0.02$ 。

模型解释的学生高数成绩组间变异可通过 Raudenbush & Bryk 方法或 Snijders & Bosker 方法进行估计 [3]。此时采取 Raudenbush & Bryk 方法计算可得约有 62.1% 的学生高数成绩跨老师水平的变异可由组水平解释变量  $xb$  解释；采用 Snijders & Bosker 方法计算可得约有 53.6% 的学生高数成绩跨老师水平的变异可由组水平解释变量  $xb$  解释。

3) 现在考虑是否有个体因素会影响个人的高数成绩，因此将依老师中心化后的个体变量高考成绩和学生性别纳入模型 2 得到模型 3：

$$gaoshu_{ij} = \beta_{0j} + \beta_{1j}cgaokao_{ij} + \beta_{2j}sex_{ij} + \varepsilon_{ij} \quad (3.7)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}xb_j + u_{0j} \quad (3.8)$$

经过回归系数显著性检验得中心化后的高考数学成绩与学生性别均显著，因此将两个变量都纳入分层线性模型。模型拟合统计量的结果表明将水平 1 的解释变量引入模型可以显著改善模型拟合度。

使用 Raudenbush & Bryk 方法，仅有 7.5% 的水平 1 变异可被模型解释，有 66.9% 的水平 2 变异可被模型解释；而用 Snijders & Bosker 方法则约有 10% 的水平 1 变异可被模型解释，有 58.8% 的水平 2 变异可被模型解释。在模型 3 中水平 2 解释方差为 66.9%，而在模型 2 中为 62.1%。这一结果说明了水平 1 解释变量既可在个体水平，也可在组群水平上影响结局变异。

4) 水平 1 解释变量斜率的随机性检验，建立模型 4：

$$gaoshu_{ij} = \beta_{0j} + \beta_{1j}cgaokao_{ij} + \beta_{2j}sex_{ij} + \varepsilon_{ij} \quad (3.9)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}xb_j + u_{0j} \quad (3.10)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (3.11)$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \quad (3.12)$$

对回归系数的随机性进行显著性检验，得到结果依老师中心化后的学生高考成绩在老师与老师之间有显著性差别，即依老师中心化后的高考数学成绩的回归系数是随机回归系数。学生性别回归系数的随机性不显著，说明学生性别对学生依老师中心化后的高数成绩的影响不随老师的变化而变化。得到

最终模型:

$$gaoshu_{ij} = \beta_{0j} + \beta_{1j}cgaokao_{ij} + \beta_2sex_{ij} + \varepsilon_{ij} \quad (3.13)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}xb_j + u_{0j} \quad (3.14)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (3.15)$$

5) 考虑跨层交互作用, 建立模型 5:

$$gaoshu_{ij} = \beta_{0j} + \beta_{1j}cgaokao_{ij} + \beta_2sex_{ij} + \varepsilon_{ij} \quad (3.16)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}xb_j + u_{0j} \quad (3.17)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}xb_j + u_{1j} \quad (3.18)$$

经检验得学生的依老师中心化的高考数学成绩这一个体解释变量的效应并不受老师性别这一组群水平变量的影响。因此得到最终模型:

$$gaoshu_{ij} = \beta_{0j} + \beta_{1j}cgaokao_{ij} + \beta_2sex_{ij} + \varepsilon_{ij} \quad (3.19)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}xb_j + u_{0j} \quad (3.20)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (3.21)$$

模型检验得检验结果, 如表 1 所示:

**Table 1.** Test results of the final model

**表 1.** 最终所建模型的检验结果

固定效应			
变量名	估计值	标准差	P 值
截距项	72.08	1.57	<0.0001
老师性别	7.60	2.00	0.0068
学生依老师中心化后的高考 数学成绩	0.38	0.06	<0.0001
学生性别	-6.81	1.36	<0.0001
方差成分			
$Var(\varepsilon_{ij}): \hat{\sigma}^2$	382.84	-	-
$Var(u_{0j}): \hat{\sigma}_{u_0}^2$	5.88	-	<0.0001
$Var(u_{1j}): \hat{\sigma}_{u_1}^2$	0.02	-	0.0930
$Cov(u_{0j}, u_{1j}): \hat{\sigma}_{u_0u_1}^2$	-0.15	-	0.0544
-2LL = 11429; AIC = 11445; BIC = 11487			

由检验结果可知, 当学生依老师中心化后的高考数学成绩增加 1 分, 则该学生的高数成绩平均增加 0.38 分; 由于老师性别是虚拟变量(0-女性, 1-男性), 所以在保持其他条件不变的情况下, 男老师所教授的学生的的高数平均成绩比女老师所教授的学生的的高数平均成绩高 7.6; 在其他条件不变的情况下, 男同学的高数成绩在平均意义下比女同学低 6.81。

## 4. 总结

在本次分析的学生的的高数成绩与学生的高考数学成绩之间关系中, 结果表明不同老师所教授学生的高数成绩差异显著。这一差异一方面是由学生依老师中心化后的高考数学成绩决定的, 这是个体水平的解释变量, 它体现的是学生个人的学习能力。另一方面是由不同大学老师存在差异而决定的, 这是组群水平的解释变量。具体分析结果表明: 1) 在其它条件不变的情况下, 男老师对学生高数成绩影响程度较女老师更大。因此在进行教师招聘的过程中应该协调好男女教师的比例, 不应让比例失衡, 以此来降低由老师性别差异对学生高等数学成绩的影响。2) 在其它条件不变的情况下, 女同学的高数成绩普遍高于男同学。因此老师应该更加关注男同学的学习, 同时也可以搭配男女互助学习小组, 让女同学带动男同学的学习, 共同进步。老师也可以加强男同学的出勤率考察, 以此来督促男同学学习。3) 在其它条件不变的情况下, 拥有更好高考数学成绩的同学普遍拥有更高的高数成绩。因此在各高校招生时我们不仅需对高考总分进行划线, 也应对高考数学成绩有基本要求。同时老师也可以对基础薄弱的同学适当增加帮扶, 例如上课时多抽点基础薄弱的同学回答问题, 以此激励他们进步。

## 基金项目

本文受国家自然科学基金(项目编号: 11671398)支持。

## 参考文献

- [1] 何晓群, 刘文卿. 应用回归分析[M]. 第4版. 北京: 中国人民大学出版社, 2011: 18-20.
- [2] Stephen W. Raudenbush, Anthony S. Bryk. 分层线性模型: 应用与数据分析方法[M]. 第2版. 北京: 社会科学文献出版社, 2007: 15-21.
- [3] 王济川, 谢海义, 姜宝法. 多层统计分析模型: 方法与应用[M]. 第2版. 北京: 高等教育出版社, 2008: 1-73.