**学生考试成绩分析论文（SPSS方法）**喻金燕

**摘要**  
本文利用 Kaggle 提供的学生考试成绩数据集，结合 SPSS 20 软件，做了一些基本的数据分析，包括描述性统计、相关分析和回归分析。主要想看看学习时长、睡眠、出勤率和以前的成绩对最后考试成绩有什么影响。结果总体上符合常识，比如学习时间多和成绩高是正相关的，以前成绩好的同学通常在最后考试中表现也不错。虽然研究规模有限，但结论对学生学习习惯和教育管理提供了参考价值，同时也提示我们单纯依赖某一个因素来预测成绩是不够的，学习成绩往往是多方面因素共同作用的结果。

**1 背景与意义**  
学习成绩问题一直是教育研究中绕不开的话题。家长关心孩子成绩好不好，老师关心教学效果如何，学生也在意自己能不能考好。通常人们有一些直观的经验判断，比如“只要多学一定能考好”，或者“熬夜复习一定会影响发挥”。这些判断到底对不对？在数据层面能不能找到证据？这正是本文要探讨的。

在教育学研究中，学习投入被认为是影响成绩的重要因素，学生成绩和学习时间、学习机会密切相关。Bloom (1984) 的掌握学习理论也强调，花更多时间往往能帮助学生掌握知识点。另一方面，心理学领域的研究发现，睡眠与记忆力、专注力息息相关，长期缺乏睡眠可能导致注意力下降和学习效率降低。

国内方面，随着“双减政策”的实施，课业负担和课外补习的变化让学生的学习时间和休息情况变得更加复杂。很多老师和家长都在思考：究竟怎样的学习习惯才最有效？在国外，许多研究发现学生的学习时间和睡眠习惯与成绩显著相关，但在中国，由于教育制度、考试压力和课外补习的影响，这种关系可能会更复杂。特别是在“双减政策”之后，学生的作息和学习模式发生了明显变化，因此探讨这些因素在本土环境下的作用就更有价值。

在这种背景下，本文的分析有三个意义：

（1）理论意义：用数据验证常见的教育假设，比如学习时间、睡眠和成绩的关系；

（2）实践意义：为学校和老师提供一些管理思路，比如是否需要严格要求出勤、如何合理安排作息；

（3）个人启示：学生自己可以通过数据结果认识到，合理学习和良好作息可能比单纯延长学习时间更有用。

**2 数据来源与数据介绍**  
本文使用的数据集来自 Kaggle 平台公开的 **Student Exam Scores**。数据包含若干学生的学习和生活情况，主要变量包括：

student\_id：学生唯一 ID；

hours\_studied：每天学习时长；

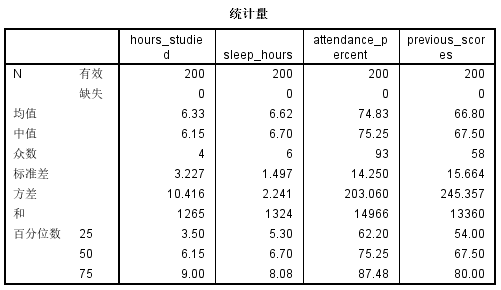
sleep\_hours：平均每天睡眠时间；

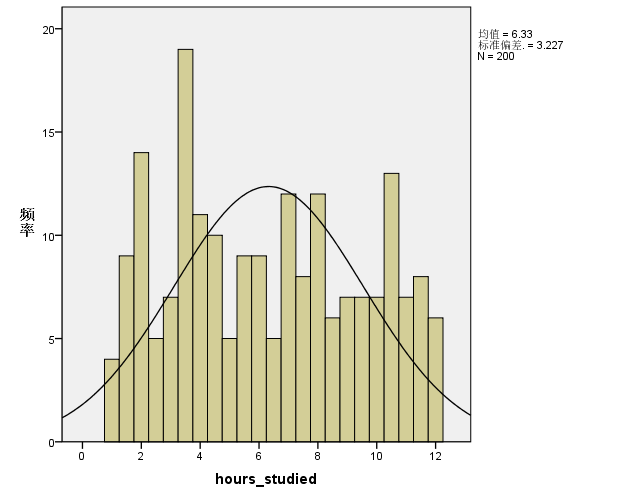
attendance\_percent：课堂出勤率；

previous\_scores：以往平均成绩；

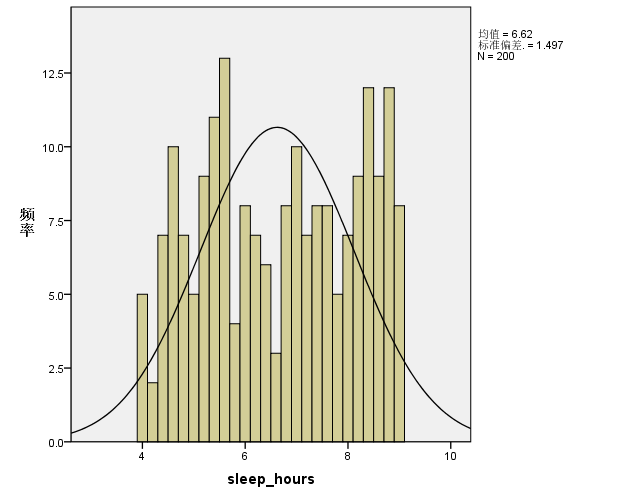
exam\_score：最终考试成绩(因变量)。

**3 简单的描述性统计分析**

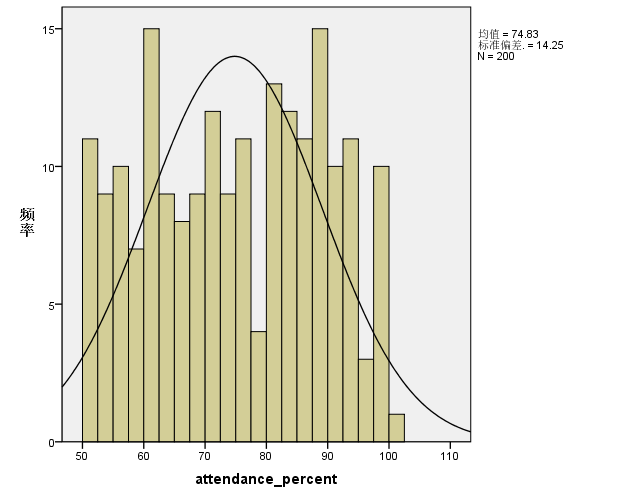




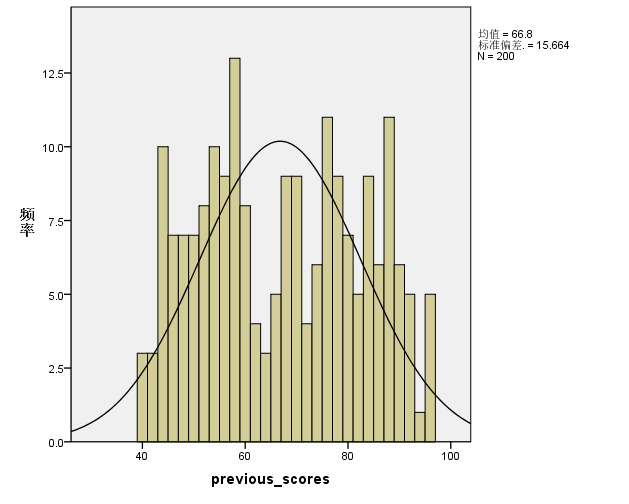
hours\_studied：平均 6.3 小时，标准方差3.22，说明大多数学生处在一个中等水平，但有部分同学少于5小时的学习。



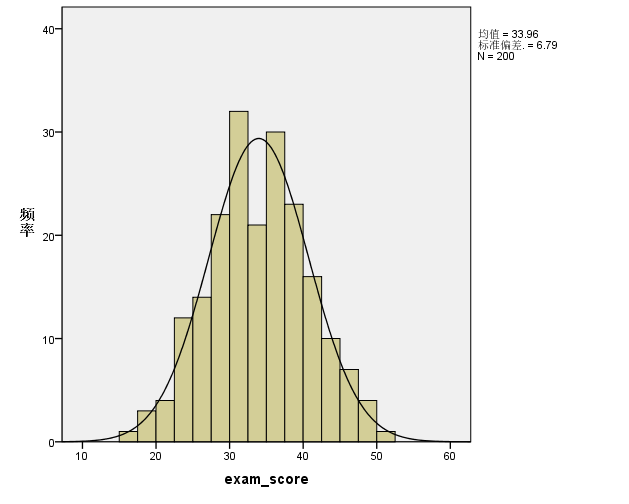
sleep\_hours：平均 6.3 小时，比较规律，但存在少部分同学睡眠不足。



attendance\_percent：平均 74.83%，标准偏差14.25，说明大部分人出勤率高，但有少部分同学出勤率低，整体偏差较大。



previous\_scores：平均 66.8 分，标准偏差15.66，差异比较明显，说明基础有高有低。



exam\_score：平均 33.95 分，标准偏差6.75，平均分较低，成绩分布较分散，较明显的符合正态分布。

**4 实证分析**

### **4.1 研究的方法**

本研究主要用两种方法：

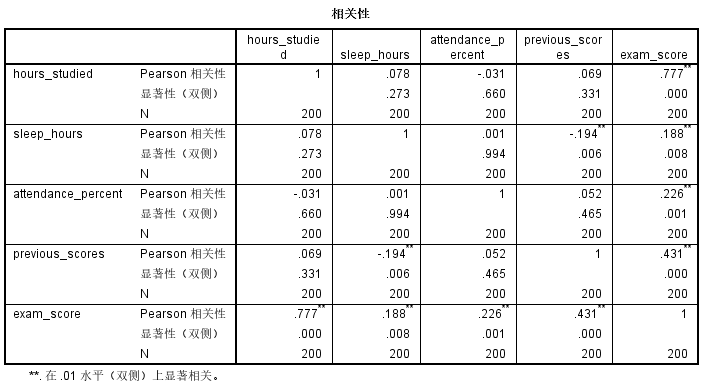
（1）相关分析：通过 Pearson 相关系数，判断各变量与考试成绩的关系；

（2）多元线性回归：在控制多个变量的情况下，估计它们对成绩的影响，并检验显著性。

### **4.2 数据与模型结合**

**相关分析结果：**

学习时长和成绩高度正相关（约 0.78，显著），说明学习投入确实重要；

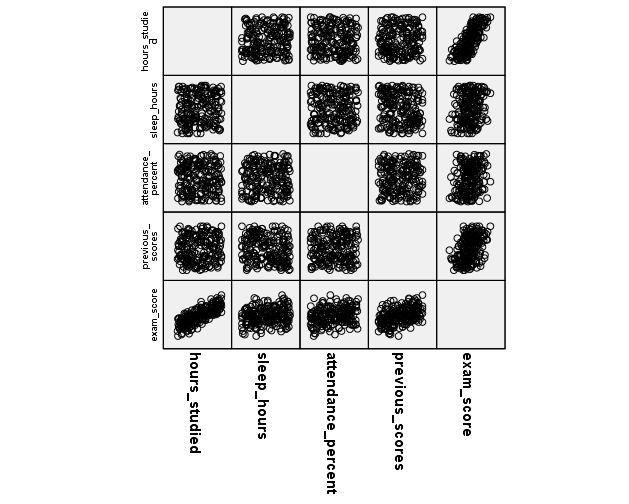


以前成绩和最终成绩相关系数为 0.43，说明学业基础对成绩有一定影响；

出勤率/睡眠和成绩正相关，相关性不算强。

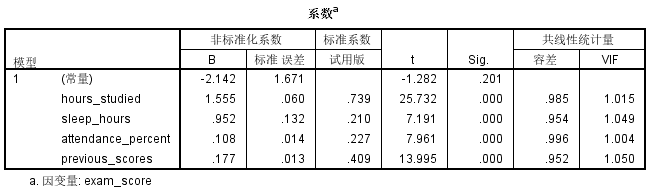
**回归分析结果：**

在回归前，通过散点图查看考试成绩因变量和其他变量之间的关系：

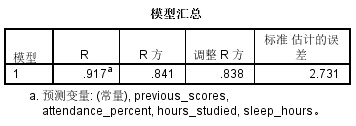


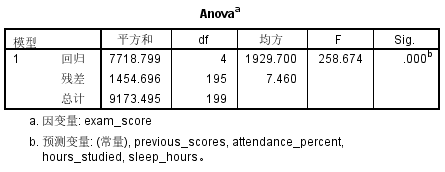
考试成绩和其他四个变量可以看出，有线性关系，特别是和学习时长有明显的线性关系。

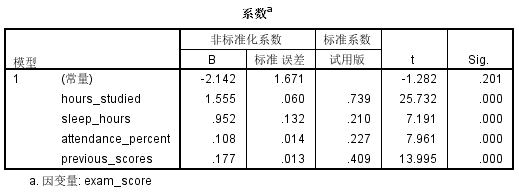
另外，检查了多重共线性，结果显示各变量的 VIF 值均小于 5，说明没有严明显的线性问题。



回归分析结果看到，模型的 R² 和调整后的 R² 都接近 1，说明整体拟合效果较好。总体检验显著性小于 0.01，说明模型通过统计检验。







根据上述结果写出模型的公式为：

y = -2.142 + 1.555x\_1 + 0.952x\_2 + 0.108x\_3 + 0.177x\_4

其中，x1=学习时长，x2=睡眠时间，x3=出勤率，x4=以前成绩。

解释：

1. 学习时长系数为 1.555，显著，说明每天多学一小时，成绩平均提高 1.5 分左右；
2. 睡眠时间系数为 0.952，显著，说明睡眠对成绩影响很大，规律作息是保障学习效果的重要因素；
3. 出勤率和以前成绩系数虽然数值小，但仍显著，说明课堂参与和学业基础不能忽视。

**5 结论**

从数据分析可以看到：

A.学习时长：学习时间和成绩确实呈现明显正相关，但整体学习时长相差不大，而学习成绩相差大，说明学习效率差异大，仅仅增加学习时间并不总能保证成绩提升。

B.睡眠：睡眠对成绩的影响比预想的更大，说明熬夜赶工不是一个好习惯，规律作息更利于发挥。

C.以前成绩：说明成绩具有延续性，长期的学习积累会在最后考试中体现出来。

D.出勤率：虽然影响不大，但依然正相关，课堂参与至少能保证基本的学习状态。

值得一提的是，相关分析来看，睡眠时间与成绩的简单相关性较弱。但在回归分析中，睡眠的回归系数显著且接近1，说明在控制学习时长、出勤率和以往成绩后，睡眠的作用才得以显现。这可能意味着，睡眠对学习成绩的影响不是直接的，而是与学习时间、出勤等变量交织在一起，存在“抑制效应”。因此，不能单纯依赖相关系数来判断睡眠的重要性。

教育启示：

A.学校和老师应指导学生科学安排学习时间，不盲目追求“学习时长”，而是强调效率；

B.应重视学生的睡眠，避免过度熬夜，保证规律作息；

C.要对基础较差的学生进行针对性辅导，帮助他们逐步提升；

D.加强课堂管理，提高出勤率，让学生更多参与课堂互动。

不足与展望：

1. 数据样本有限，结果可能不适用于所有学生群体；
2. 研究中没有考虑心理状态、家庭背景等变量；
3. 未来可以尝试结合纵向数据，甚至用机器学习方法建立更复杂的预测模型。
4. 在教育启示里加“对家庭和学生的建议”。比如家长如何监督学生合理作息，学生如何自我管理时间。
5. 在展望里加“结合大数据和AI的可能性”，比如未来教育预测会更多用智能算法。