

# Statistical Analysis of Student Score

Di Gong<sup>1</sup>, Rong Jiang<sup>2</sup>

<sup>1</sup>Shangwen Middle School, Shanghai

<sup>2</sup>Department of Mathematics, Tongji University, Shanghai

Email: [987521005@qq.com](mailto:987521005@qq.com), [jrtrying@126.com](mailto:jrtrying@126.com)

Received: Dec. 11<sup>th</sup>, 2013; revised: Jan. 15<sup>th</sup>, 2014; accepted: Jan. 25<sup>th</sup>, 2014

Copyright © 2014 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



---

## Abstract

Student test score is an important topic in teaching. Taking a midterm grade test scores as an example, by using the basic ideas and methods of statistics for each of the students and teachers, the individual student grades and overall achievement information table were presented. The new method based on grade information, "increased probability", is proposed to give advice to the school leaders on which subject will be highly valued. The principles based on individual score, "breakthrough probability", "baseline" and "goal", are proposed to help students understand their own image of concrete achievements and future specific goal.

## Keywords

Student Score; Probability and Statistics; Linear Regression

---

# 学生成绩的统计分析

龚笛<sup>1</sup>, 姜荣<sup>2</sup>

<sup>1</sup>上海市尚文中学, 上海

<sup>2</sup>同济大学数学系, 上海

Email: [987521005@qq.com](mailto:987521005@qq.com), [jrtrying@126.com](mailto:jrtrying@126.com)

收稿日期: 2013年12月11日; 修回日期: 2014年1月15日; 录用日期: 2014年1月25日

---

## 摘要

学生考试成绩的分析是教学研究的一个重要课题。以一个年级的期中考试成绩为案例, 用基本统计的思

想和方法提取出对学生和老师各自有用的信息,给出年级总体和学生个人成绩信息表。对年级成绩信息,提出“提高概率”一个新工具,来给学校建议接下来重视哪门学科;对学生个人成绩信息,提出“突破概率”、“基准线”和“奋斗目标”,有助于学生形象具体的了解自己的成绩,以及未来努力的具体目标。

## 关键词

学生成绩; 概率统计; 线性回归

## 1. 引言

考试是教学过程中的重要环节,是检验教学效果和教学质量的重要方法。通过对学生成绩的统计分析,可以了解学生在学习过程中是否达到了预期目标,帮助教师及时发现教学中存在的问题及薄弱环节,有助于教师及时调整教学内容,也有利于教师和教学管理人员评定教学质量和比较各种教学方法的效果,为教学改革提供反馈信息。而对于学生来说,有助于了解自己的全局地位,对于初中生而言是极其重要的一件事情,有利于学生了解自己的不足以及优势[1]。

如果能够在大型考试后,通过一些统计分析得出一些充分反映学生总体的成绩水平信息,以及针对每个学生的具体情况提出相应的建议对于现今的初中教育是很有意义的。

下面选取上海市尚文中学 2013 级预备班 221 位学生的期中成绩数据,结合基本统计的思想和方法,给出学生成绩的信息。

## 2. 年级成绩统计分析

对于教师和学校领导给出表 1: 年级成绩信息表。使得教师和学校领导对学生这次考试成绩有个总体的了解。具体通过均值来了解总体的平均分;中位数是中间位置人的分数;标准差反映出分数的波动程度,标准差大的说明,该学科更容易拉开差距;峰度反映中均值附近人的多少,峰度越大,均值附近人越多;偏度反映出卷子的难易程度;极差为最好与最差的差距大小;最后提出“提高概率”一个新工具,来给学校建议接下来重视哪门学科。

Table 1. Grade score information table

表 1. 年级成绩信息表

	语文	数学	英语	总分
均值	68.33	81.96	75.24	225.53
中位数	69	85	77.5	228
标准差	8.48	13.40	16.02	34.43
峰度	-0.01	7.70	0.98	2.68
偏度	-0.47	-1.95	-0.97	-1.18
极差	42	93	82	208.5
最小值	42	7	17	70
最大值	84	100	99	278.5
提高概率	21.53%	34.98%	43.49%	-
样本量	221	221	221	221

下面介绍几个统计量的含义：

## 2.1. 标准差

标准差定义是总体各单位标准值与其平均数离差平方的算术平均数的平方根。它反映数据间的离散程度[2]。

对本文数据反映的是拉开成绩容易程度。从表 1 中该项可以看出本次考试中，按拉开成绩容易程度排序：英语 > 数学 > 语文。

## 2.2. 峰度

正态分布的峰度为 3。小于 3 称分布具有不足的峰度，大于 3 称分布具有过度的峰度[2]。

从表 1 中该项可以看出，数学具有过度的峰度，说明均值附件的人数很多。而语文和英语均值附件的人比较少，语文相对英语这种情况更严重。

## 2.3. 偏度

偏度是统计数据分布偏斜方向和程度的度量，是统计数据分布非对称程度的数字特征。正态分布的偏度为 0，两侧尾部长度对称。偏度小于 0 时数据位于均值左边的比位于右边的少，直观表现为左边的尾部相对于与右边的尾部要长，因为有少数变量值很小，使曲线左侧尾部拖得很长；偏度大于 0 时数据位于均值右边的比位于左边的少，直观表现为右边的尾部相对于与左边的尾部要长，因为有少数变量值很大，使曲线右侧尾部拖得很长；而偏度接近 0 则可认为分布是对称的[2]。

从表 1 中该项可以看出，语文、数学和英语的偏度都小于 0，而从可以看出这次考试三门课都是高分的人数较多，卷子偏简单。按卷子容易程度排序：数学 > 英语 > 语文。

## 2.4. 提高概率

提高概率定义为：当总分增加 1 分的情况下，这 1 分来自某个学科的概率。

采用简单的总分对学科的线性回归[3]来观察他们之间的依赖关系(见图 1~3)。其中线性回归方程

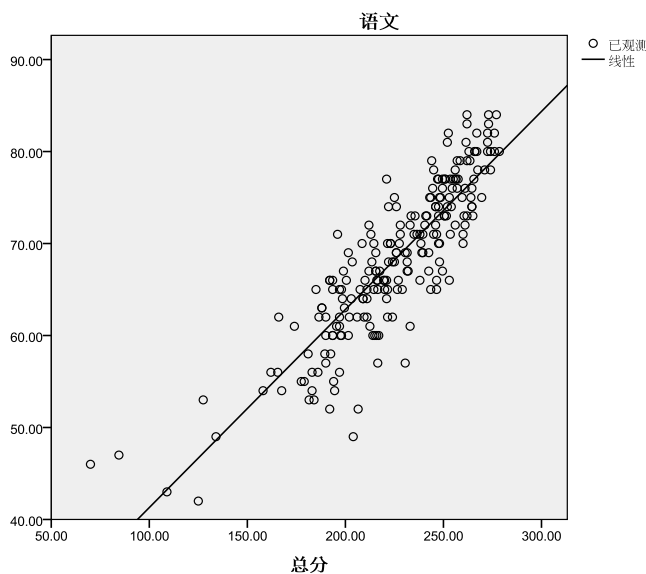


Figure 1. The total score and Chinese in linear regression

图 1. 总分和语文成绩的线性回归

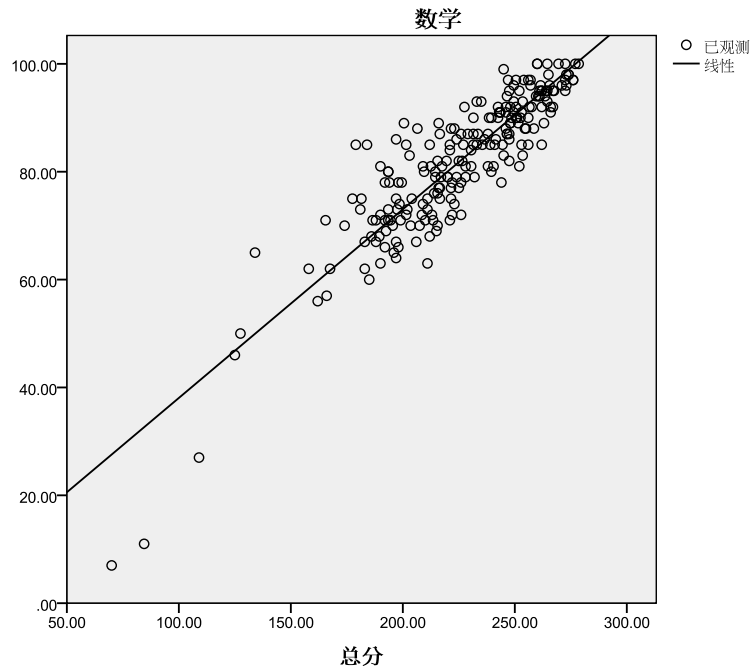


Figure 2. The total score and Math in linear regression  
图 2. 总分和数学成绩的线性回归

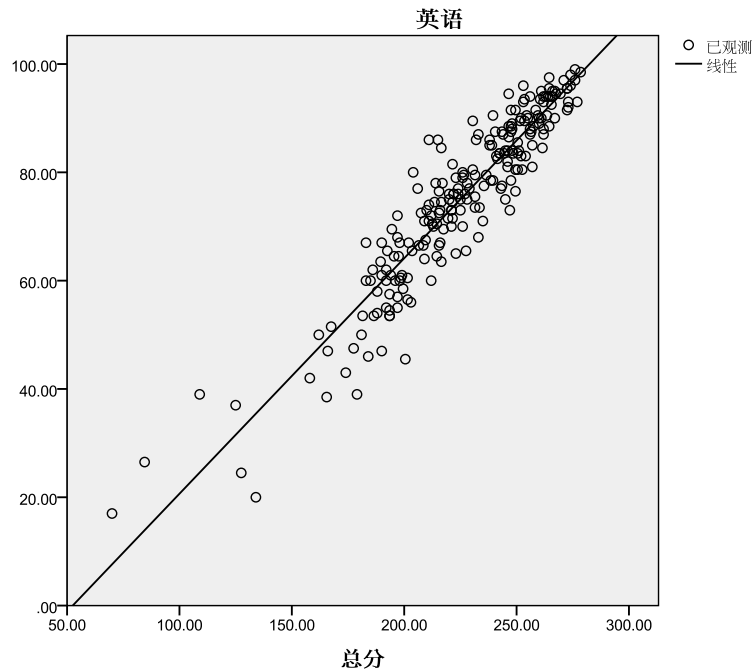


Figure 3. The total score and English in linear regression  
图 3. 总分和英语成绩的线性回归

的系数见表 2。表 2 中的斜率项代表总分增加 1 分，学科增加的分数。因此，我们定义：

$$\text{提高概率} = \text{斜率} \times 100\%$$

因为这样三门学科相加可能不为 1，所以需要正则化。

从表 1 中该项可以看出, 按提高概率从大到小排序: 英语 > 数学 > 语文。由此结果学校接下来应该重视英语, 其次数学, 再次为语文。

### 3. 学生个人成绩统计分析

在每次大型考试后, 如何让学生了解到自己的成绩和水平程度, 这也是一个学生和教师共同关心的问题。建议采用如下的方法给每位学生制作他们个人成绩信息表。(以下以学生沈 XX 的成绩为例)

#### 3.1. 累积分布

累积分布:  $F(x) = p(X \leq x)$ , 其中  $X$  代表学科所有人的成绩,  $x$  为学生个人学科成绩。累积概率反映了比自己差的百分比。

从表 3 该项可以看出, 该学生数学较好, 英语中等, 语文较差, 总体处于中游水平。

#### 3.2. 突破概率

为了下次考试成绩更好, 自然要问要考虑在哪买学科上更努力对自己成绩提高最有帮助? 一个直观的认识: 如果某位同学一个学科当前成绩的累积分布  $F_i(x_i)$  较大时, 则在这个学科上提高成绩的难度应该比其他学科中累积分布  $F_j(x_j)$  较小的要大; 也就是说如果总分提高 1 分, 那么着 1 分出现在学科  $i$  中的概率小于出现在其他学科  $j$  中的概率。

基于上面的分析, 用  $p_i(x_i) = 1 - F_i(x_i)$  表示 1 分出现在学科  $i$  中的概率, 那么  $q_i(x_i) = F_i(x_i)$  就是 1 分不出现在学科  $i$  的概率(当前学科  $i$  的分数为  $x_i$ )。2.4 中讨论过提高概率, 对全年级学生学科提高概率高的, 对个人学科提高应该有促进作用, 所以下面把学科提高概率作为权重。

记: 1: 语文; 2: 数学; 3: 英语。  $a_i$  = 第  $i$  个学科的提高概率,  $i = 1, 2, 3$ 。

$$\text{突破概率(语文)} = \frac{a_1 p_1 q_2 q_3}{a_1 p_1 q_2 q_3 + a_2 q_1 p_2 q_3 + a_3 q_1 q_2 p_3}$$

Table 2. Table on regression coefficient  
表 2. 回归系数表

	语文	数学	英语
斜率	0.2153	0.3498	0.4349
截距	19.7624	3.0749	-22.8374

Table 3. Table on individual student's score information

表 3. 学生个人成绩信息表

班级: 预备(06 班) 学号: 2201301XX 姓名: 沈 XX				
	语文	数学	英语	总分
成绩	66	94	86.5	246.5
排名	131	37	65	73
累积分布	40.72%	83.26%	70.59%	66.97%
突破概率	55.48%	12.45%	32.07%	-
基准线	73	89	84.5	-
奋斗目标	70	95	88.5	251

$$\text{突破概率(数学)} = \frac{a_2 q_1 p_2 q_3}{a_1 p_1 q_2 q_3 + a_2 q_1 p_2 q_3 + a_3 q_1 q_2 p_3},$$

$$\text{突破概率(英语)} = \frac{a_3 q_1 q_2 p_3}{a_1 p_1 q_2 q_3 + a_2 q_1 p_2 q_3 + a_3 q_1 q_2 p_3}。$$

从表 3 该项可以看出, 该学生学习重心应该放在语文上, 语文是使他最容易提高分数的学科, 而数学提高就比较难, 只要保持住就行。

### 3.3. 基准线

一般学生都有其各自的“优势学科”和“翘脚学科”, 那么相对自己水平(总分), 各学科应该为多少, 从中可以反映出自己在某学科有多少的优势以及劣势。对于“翘脚学科”更能说明自己应该到那个分数。

某学科的基准线可以理解为, 总分在某个水平上, 某学科对应的成绩。可以利用 2.4 中的线性回归方程, 把总分代入某学科的线性回归方程, 可以得到这个总分对应某学科的平均成绩。

从表 3 该项可以看出, 该学生语文是“翘脚学科”, 相差 7 分; 数学和英语是“优势学科”, 分别相差 5 分和 2 分。

### 3.4. 奋斗目标

以鼓励学生进步为目的, 为其设定合理的进步目标。累积分布  $F(x) = p(X \leq x)$ , 表示的是比自己成绩差的百分比; 那么相对的  $1 - F(x)$  即为比自己成绩好的百分比。

假设共有学生 100 名; 学生成绩为 80 名给其设定目标到 65 名应该是合理的, 不太难也不太简单; 而学生成绩为 50 名给其设定目标到 40 名; 而学生成绩为 20 名给其设定目标到 15 名。基于以上分析, 奋斗目标的累积分布应该为:

$$F(x) - (1 - F(x))/5 = 0.2 + 0.8F(x)$$

再根据奋斗目标的累积分布可以得到对于的成绩。

从表 3 该项可以看出, 该学生语文应提高 4 分, 因为现在考的比较差, 又加上他整体的水平, 所以 4 分比较合理; 数学应提高 1 分, 该生现在数学已经考的很好, 所以较难提高; 英语应提高 2 分, 总分应提高 4.5 分。

## 4. 结论

本文通过案例分析, 将统计方法与实际教学的需要相结合, 为教学工作者和学生提供了有用的信息, 从而及时发现教学双方在教学过程中存在的问题, 为改进教学工作提供定量话的科学依据, 真正发挥考试的导向、激励和反馈作用, 促进教学和学习质量的不断提高。并且提出的各种新的分析工具都是在很自然的前提下提出的, 具有一定的实际意义, 本文的方法可以很自然的应用到各种考试成绩的分析中。

## 致 谢

感谢编辑及审稿人对本文提出的宝贵意见。

## 参考文献 (References)

- [1] 苏斌, 谢友芹 (2006) 统计分析在学生成绩评估中的应用. *系统工程理论与实践*, 7, 134-140.
- [2] 杨振明 (2001) 概率论. 科学出版社, 北京.
- [3] 王松桂, 史建红, 尹素菊, 吴密霞 (2004) 线性模型引论. 科学出版社, 北京.