

城镇高中生家庭教育投入对学习成绩的影响研究

摘要

少年兴旺，则国家兴旺，因此学生的学习成绩，一直是学校、老师和广大家长，甚至是整个社会共同的重点关注。关于影响学生学习成绩的相关分析研究，也是层出不穷。包括先天个体智力水平的差异，后天养成不同的学习习惯、个体差异化的自主学习欲望和自律能力等等。除此之外，我们认为，家庭教育投入，也是影响学生学习成绩的重要外部因素，包括家庭对于择校费的投入，会导致学生接受不同水平的高中教育，家庭对于课外学习费的支出，会影响学生知识的累积程度，最终导致成绩排名的差距，等等。因此，本文通过相关矩阵等信息，讨论了各种不同的家庭教育投入，对学生学习成绩排名的影响；并通过因子分析的方法对影响城镇高中生学习成绩的因素进行降维，分析影响学生学习成绩的各种因子；并基于上述结果，通过回归分析建模，给出重要教育支出作为因变量，学习成绩的具象表现——成绩排名作为自变量的回归方程；最后对模型的优缺点进行分析，并以此给出调整教育投入比例的合理化建议，从而提高学生的学习成绩。

问题一：主要研究各种不同的家庭教育投入，对学生学习成绩排名的影响程度。本文选取特定城市为长春市，收集该城市典型城镇高中生，关于家庭教育投入与其在校成绩排名比例的实际数据，通过相关矩阵等信息，对9项影响城镇高中生学习成绩的因素进行影响程度分级。

问题二：主要通过因子分析的方法对包括学校教育支出和学校外教育支出在内的，两大教育支出中的9项影响城镇高中生学习成绩的因素进行降维，分析影响学生学习成绩的各种因子。

问题三：主要研究在问题一、问题二的基础上，分别通过最小二乘方法和Logit模型两种方式，建立回归方程，以求更加贴近实际的，建立以重要家庭投入为自变量，以学生学习成绩为因变量的回归方程，来更加合理的刻画，城镇高中生家庭教育投入对学习成绩的影响，并对两种回归模型的优缺点进行合理分析。

问题四：主要根据前三个问题的具体研究结果和模型的构建，结合实际学习情况，从调整各类家庭教育投入比例的角度，对提高学生的学习成绩，提出有建设性的合理化建议。

关键词：家庭教育投入 因子分析 最小二乘法 Logit 回归模型

一. 问题重述

教育是成功的基石，学习成绩自然是家长和学生的心头之重，因此，如何从多方面合理高效的提高学生的学习成绩，一直引起广泛的关注和讨论。由于不同学生的智力水平存在差异，也拥有不同优良程度的的学习习惯、自主学习欲望和自我约束能力，考虑从家庭教育投入方面，探讨对学生的学习成绩的影响，也是一个值得分析研究的问题。

家庭教育投入的哪些方面，与学生的学习成绩密切相关，又是否可以通过合理的数学模型，去刻画对学生学习成绩影响最重要的几项家庭教育投入与学习成绩排名的关系，为科学提高学生成绩提供定量依据，为此，请大家尝试解决以下问题：

1. 请自行设计调查问卷，收集不同类型的家庭教育投入和学生学习成绩排名的实际数据，并用合理的分析方法，对影响城镇高中生学习成绩的不同种类家庭教育投入进行影响程度分级。

2. 请利用适当的分析方法，对不同种类家庭教育投入进行合理降维。

3. 请利用适当的建模方法，合理定量刻画城镇高中生家庭教育投入与学习成绩排名的关系。

4. 根据研究结果，从调整各类家庭教育投入比例的角度，对提高学生的学习成绩，提出有建设性的合理化建议。

二. 问题分析

如何提高学生的学习成绩，一直是一个棘手的问题，虽然不能以成绩论英雄，但是学生成绩的好坏，的确在很大程度上，影响着学生的未来，高中时期，尤为重要。由于先天基因决定了大部分学生的智力水平都处在平均水平之中，而且每个人的自我评价存在差异，所以关于良好的学习习惯、自主学习的欲望和自我约束能力等等，外部影响学生学习成绩的因素，在学生个体之间，不能明确量化描述。

因此，本文在假设上述因素均忽略不考虑的情况下，为了探索学习成绩与各类家庭教育投入的关系，利用调查问卷，收集了 556 份(385 份可用)长春城镇高中生家庭教育投入与在校成绩排名比例的实际数据，将问题具体化。通过合理的分析方法，研究 9 项影响长春城镇高中生学习成绩的家庭教育投入因素，并建立起相应的数学模型进行研究。

问题一：主要研究不同类型的家庭教育投入，对学生学习成绩排名的影响程度。利用长春市城镇高中生的家庭教育投入与其在校成绩排名比例的实际数据，

通过相关矩阵等信息，对 9 项影响城镇高中生学习成绩的因素进行影响程度分级。

问题二：主要通过因子分析的方法对 9 项影响城镇高中生学习成绩的因素进行降维，为数学模型的建立，做定性的前期研究。

问题三：主要研究在问题一、问题二的基础上，充分考虑现实中学生的学习情况，以数据散点图为依据，分别通过最小二乘方法和 Logit 回归模型，合理拟合数据，建立以重要的家庭投入为自变量，学生学习成绩为因变量的回归方程，来更加合理的刻画，城镇高中生家庭教育投入对学习成绩的影响。

问题四：主要根据前三个问题的具体研究结果和模型的构建，结合实际学习情况，从调整各类家庭教育投入比例的角度，对提高学生的学习成绩，提出有建设性的合理化建议。

三. 模型假设

- [1]假设不考虑学生先天的智力水平差异
- [2]假设不考虑学生自我约束能力和自主学习欲望差异
- [3]假设不考虑不同学校，成绩排名比例的学习实力差异

四. 问题一：不同种类家庭教育投入的影响程度分级

本文前期进行了长春城镇高中生家庭教育投入与在校成绩排名比例的问卷调查，具体收集的影响学生学习成绩的 9 项家庭教育投入如下。

学校教育支出是指为了介绍学校教育，家庭所必须的支出。

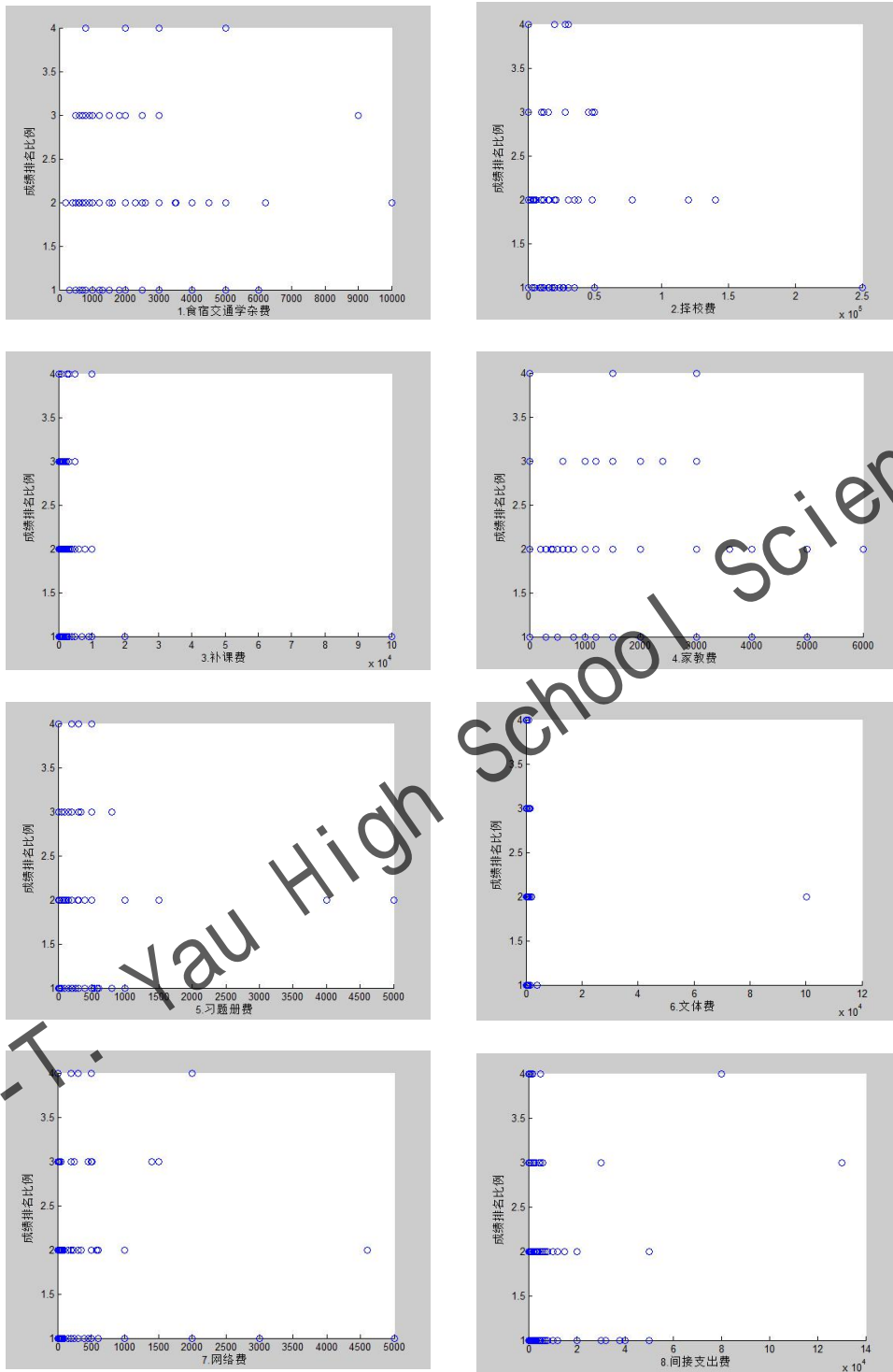
1. 每月一般支出(学杂费、课本文具资料费、在校食宿费、为上学的交通费)。
2. 择校费。

学校外教育支出是指为高中生在学校之外的教育目的而支出的费用。

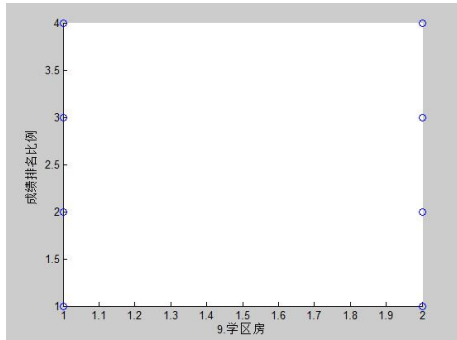
3. 每月参加课外学习班(补习班、各类兴趣班、特长班)投入的费用。
4. 每月聘请家教的费用。
5. 每月购买课外书及习题册投入的费用。
6. 每月在体育、文化活动中投入的费用。
7. 每月在网络、虚拟平台投入的费用。
8. 每月为孩子购买电脑、大件乐器、体育用品、通讯工具投入的费用。
9. 是否因孩子上学购买学区房或租房(1 为是，2 为否)。

对于学生的学习成绩，分为前 20%，20%-50%，50%-80%和后 20%四级，为了更加直观的看出各种家庭教育支出和学生学习成绩排名比例的关系，下面分别给出，以 9 种家庭教育支出为自变量，学生学习成绩排名比例为因变量的实际数据散点图如下：

图一. 家庭教育支出与学生学习成绩排名比例散点图



2018S. -T. Yau High School Science Award



从图一散点图中不难看出，9 种家庭教育支出与学生学习成绩排名比例，均没有一种固定的变化趋势，但是为了深入分析出一些问题，本文给出如下 9 种家庭教育支出与学生学习成绩排名比例的相关系数如下表：

表一. 9 种家庭教育支出与学生学习成绩排名比例相关系数表

	食宿交通 学杂费	择校费	补课费	家教费	习题册费	文体费	网络费	间接支出 费	学区房	成绩排名
相关 食宿交通	1.000	.191	.219	.095	.182	-.019	.222	.132	-.029	.088
学杂费		1.000	.696	.229	.064	-.010	.001	.081	-.022	.083
择校费			1.000	.059	.087	-.005	.021	.042	.002	-.011
补课费				1.000	.215	-.011	.155	.191	-.174	.108
家教费					1.000	.011	.108	.003	-.102	.048
习题册费						1.000	.016	.024	.040	.031
文体费							1.000	.017	-.041	.073
网络费								1.000	.004	.119
间接支出 费									1.000	.052
学区房										1.000
成绩排名										

根据表一，有些意外的了解到，第八项教育支出——间接支出费，和学生学习成绩排名比例的相关性最大，为 0.119，即家庭为购买电脑、大件乐器、体育用品、通讯工具等等投入的费用，而不是传统的补课费和习题册费等等；而且通过图一可以看出，忽略一些离群点，成绩排名后 20% 的同学，在间接支出费方面的教育投入，真的远不如成绩排名前三档的同学。

不过这也符合现在的教育模式和节奏，在普遍独生子女的社会中，学生的教育是头等大事，几乎每个学生都会进行不同程度的补习，而且重在引导培养学生的自主学习能力和多方面发展，共同促进学生学习成绩的提升。因此购置电脑等设备，可以方便学生线上查阅世界各地的学习资料，购置乐器和体育用品，可以丰富培养学生的课外兴趣，不做“书呆子”，正如一种现象，学的好的孩子，玩的也好。

对比来看，第三项教育支出——学生的补课费，和学生学习成绩排名比例的相关性，不仅最小，而且还有负相关的影响，为-0.011。从图一可以看出，除去排名比例前 20%处存在离群点，处于各个成绩排名比例的学生，都在补课费上，有旗鼓相当的教育支出；但是也能看出，成绩在后 20%的同学，在补课费上的支出，稍逊于前三档。

而这也符合实际情况，在竞争激烈的高中学习中，普遍学生都需要一定的补课费支出，去巩固乃至扩展所学知识，稳定提升自己的学习成绩，这一传统的教育支出项目，已经成为一种必然，没有使成绩进一步加分的能力，而且如果一味的只补课，也许会产生一些适得其反的结果。

从地域角度看来，第一项教育支出——食宿交通学杂费，和学生学习成绩排名比例的相关性为 0.088，相关性排第三，第九项教育支出——是否租买学区房，和学生学习成绩排名比例的相关性为 0.052，相关性排第六，不难看出，如果缩短学生上学补课的距离，节省学生的学习休息时间，也可以在一定程度上，提升学生的学习成绩。

根据上述结果，可见 9 种不同家庭教育投入对城镇高中生学习成绩的影响程度分级如下表二：

表二. 家庭教育投入对城镇高中生学习成绩的影响程度分级表

	一级	二级	三级	四级
等级	间接支出费 家教费	食宿交通学杂费 择校费 网络费	学区房 习题册费 文体费	补课费

传统的教育投入模式已经不能更好的提高学生的成绩，一味的投入补课费，并不能提高学生的学习兴趣，不如从提高间接支出费和网络费等方面入手，利用新媒体的力量，扩展学生的眼界和思路，并且提高家教费的支出，给学生更加个性化的课后辅导等等。

五. 问题二：因子分析

在问题一中，我们对 9 项家庭教育投入，进行了影响程度分级，为了进一步了解分析 9 项投入的内部联系，下面，本文将通过因子分析的方法提取公共因子，来更加合理的概括影响学生学习成绩的投入因素，具体的 SPSS 分析输出结果如下：

表三. KMO 和 Bartlett 的检验

取样足够度的 Kaiser-Meyer-Olkin 度量。	.530
Bartlett 的球形度检验	近似卡方
	399.350
	df
	36
	Sig.
	.000

表四. 相关系数表

	食宿交通 学杂费	择校费	补课费	家教费	习题册费	文体费	网络费	间接支出 费	学区房
相关									
食宿交通	1.000	.191	.219	.095	.182	-.019	.222	.132	-.029
学杂费		1.000	.696	.229	.064	-.010	.001	.081	-.022
择校费			1.000	.059	.087	-.005	.021	.042	.002
补课费				1.000	.215	-.011	.155	.191	-.174
家教费					1.000	.011	.108	.003	-.102
习题册费						1.000	.016	.024	.040
文体费							1.000	.017	-.041
网络费								1.000	.004
间接支出									1.000
费									
学区房									

从表三和表四的巴特利特球度检验、KMO 检验和相关矩阵可见，在 KMO 中，p 值为 0.000，小于显著性水平 0.05，拒绝原假设，表明相关矩阵与单位矩阵有显著差异，KMO 为 0.53，说明可以进行因子分析；虽然大部分相关系数都偏低，线性关系不强，但是也可以提取公共因子，对 9 项家庭教育投入进行公因子提取，在一定程度上解释大部分的原始信息，具体的旋转成份矩阵如下表五：

表五. 旋转成份矩阵

	成份			
	1	2	3	4
补课费	.909	.085	.029	-.035
择校费	.905	.030	-.084	.102
网络费	-.109	.762	.030	.012
食宿交通学杂费	.272	.668	.088	.090
习题册费	.044	.513	-.370	-.008
学区房	.044	-.007	.761	-.010
家教费	.120	.210	-.540	.516
文体费	-.072	.097	.412	.344
间接支出费	.050	-.029	.041	.862

提取方法 :主成分分析法。

旋转法 :具有 Kaiser 标准化的正交旋转法。

a. 旋转在 4 次迭代后收敛。

根据表五的旋转成份矩阵, 可以将补课费、择校费归为第一类, 网络费、食宿交通学杂费和习题册费归为第二类, 学区房、家教费和文体费归为第三类, 间接支出费归为第四类。

该结果与问题一中的影响程度分级表结果相一致, 对学生学习成绩影响最大的四类家庭教育支出, 分别为间接支出费、家教费食宿交通、学杂费和择校费, 分别归在了不同的四类之中, 为公因子的提取, 提供信息。

因此, 第一公因子可以归纳为传统教育支出因子, 学生选择优质学校, 平时对课业的巩固, 都是最基本的教育支出。

第二公因子可以归纳为基本教育支出因子, 学生的学校生活, 离不开食宿交通, 与家长同学通过网络的联系, 对学习的知识进行巩固。

第三公因子可以归纳为加强教育支出因子, 无论是学区房的租买, 还是一对一昂贵的家教, 亦或是文体费, 都不是每一个家庭必须的选择, 但是可以对学生的学生成绩, 起到锦上添花的作用。

第四公因子可以归纳为新型教育支出因子, 购置电脑, 文体用具等, 方便学生开阔眼界, 丰富学生的课余生活, 玩的好, 学的好, 并对学生的学习成绩, 起最重要的影响。

六. 问题三：最小二乘方法和 Logit 回归模型的回归方程

由问题一和问题二的分析结果可见，我们虽然能够对影响学生学习成绩的教育投入进行等级分类，也可以提取四个公因子，来解释大部分的原始信息，以便我们对提高学生的学习成绩，提够合理化的建议，但是这只是定性的分析研究，本节，我们将进行定量的讨论分析。

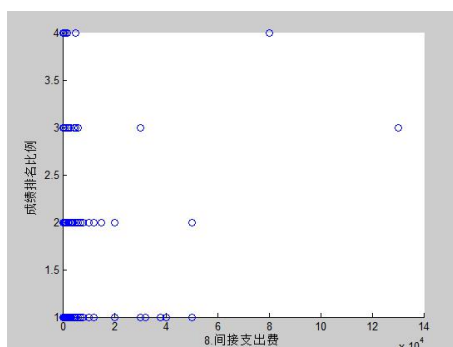
由表一可见，9 种影响学生学习成绩的教育投入中，交通食宿学杂费和补课费之间的相关性最大，为 21.9%，可见教育投入之间的交互作用影响不是很大，建模的时候，可以先忽略交互作用，建立合理的回归模型。

而且由图一的散点图可见，无法直接用排名前 20% 记为“1”，20%-50% 记为“2”，50%-80% 记为“3”，后 20% 记为“4”作为因变量，进行建模，由于我们想要研究的是如何提高学生的学习成绩，那么我们不妨可以考虑，分别以传统教育支出因子、基本教育支出因子、加强教育支出因子和新型教育支出因子的分段平均教育投入金额，作为自变量，该段对应排名中，所有排前 20% 的学生比例人数，作为因变量，进行建模，下面就是以新型教育支出因子——间接支出费为例，建立可以拟合数据的回归模型。

6.1 最小二乘方法

间接支出费和成绩排名比例的原始数据散点图如下图二所示。

图二. 间接支出费和成绩排名比例散点图



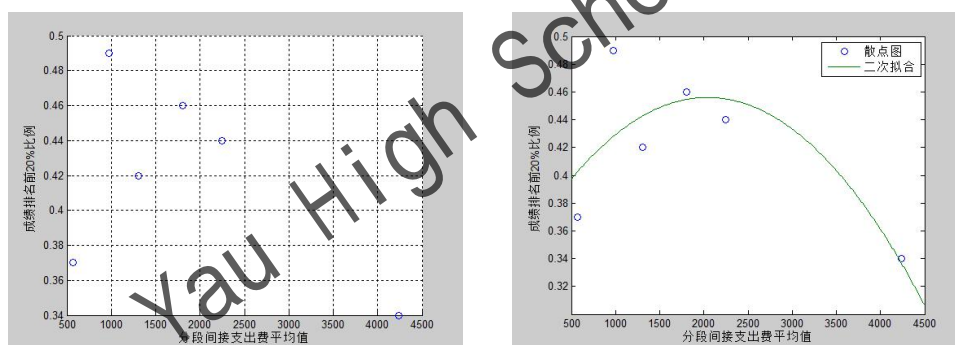
下面是为了进行建模，对数据的处理结果，见表六。

表六. 数据处理结果表

食宿交通学杂费 x	排名前 20%比例 y
565	0.37
978	0.49
1308	0.42
1800	0.46
2248	0.44
4234	0.34

根据表六，我们利用 MATLAB 得到图三的数据散点图，并且根据散点图的特点，决定对数据进行二次拟合，具体结果如下。

图三. 数据 (x, y) 的散点图和二次拟合图(最小二乘)



根据图三的两个图，我们可以看出，随着食宿交通学杂费的增大，因变量先增加后减小，可见，在一定程度的食宿交通学杂费教育投入下，学生成绩在前 20%的可能性，会随着金钱投入的增大而增加，但是，过犹不及，一味的过大增加教育投入，也会适得其反，学生的学习成绩，反而会下降，这与普遍的认知，也是相一致的。

一般的回归模型的形式为：

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$$

其中随机误差 ε 服从均值为 0, 的正态分布。

在我们的模型中，应为

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

利用最小二乘的方法，根据等式

$$\beta = (X^T X)^{-1} X^T Y$$

我们可以得到 β_0 , β_1 , β_2 的估计值，其中

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \dots & \dots & \dots \\ 1 & x_n & x_n^2 \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix}$$

具体的估计结果如下

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0.3532 \\ 0.0101 \\ -0.0002 \end{bmatrix}$$

即得到二次回归模型

$$\hat{y} = 0.3532 + 0.0101x - 0.0002x^2$$

分别将食宿交通学杂费带入上式，可以得到预测值和偏差比例如下表：

表七. 排名前 20% 比例的预测值 (最小二乘)

食宿交通学杂费 x	排名前 20% 比例	预测值	偏差比例
565	0.37	0.4039	9.16%
978	0.49	0.4328	11.67%
1308	0.42	0.4511	7.40%
1800	0.46	0.4702	2.22%
2248	0.44	0.4792	8.91%
4234	0.34	0.4223	24.21%

其中偏差比例为

$$\text{偏差比例} = \left| \frac{y_{\text{真值}} - y_{\text{预测值}}}{y_{\text{真值}}} \right|$$

偏差比例结果显示，预测结果在可接受范围之内，除去当食宿家庭学杂费过大时，偏差比例会高达 24.21%，其他情况，均不超过 12%。

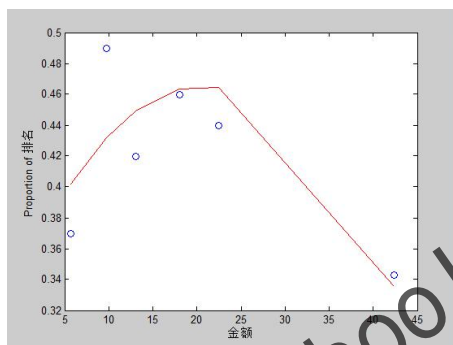
6.2. Logit 回归模型

该模型的优点在于，可以将[0,1]的范围，扩展到整个实数，具体模型表达式如下：

$$\text{logit}(\pi(x)) = \ln \frac{\pi(x)}{1-\pi(x)} = \beta_0 + \beta_1 x + \beta_2 x^2$$

利用 MATLAB 编程，得到 Logit 回归拟合以及参数估计如下。

图四. 数据的散点图和二次拟合图(Logit 回归)



$$\text{logit}(\pi(x)) = \ln \frac{\pi(x)}{1-\pi(x)} = -0.6294 + 0.0477x - 0.0012x^2$$

分别将食宿交通学杂费代入上式，可以得到预测值和偏差比例如下表：

表八. 排名前 20%比例的预测值(Logit 回归)

食宿交通学杂费 x	排名前 20%比例	预测值 $\hat{\pi}(x)$	偏差比例
565	0.37	0.3625	2.03%
978	0.49	0.4089	16.55%
1308	0.42	0.4372	4.09%
1800	0.46	0.4602	0.04%
2248	0.44	0.4584	4.18%
4234	0.34	0.2522	25.82%

偏差比例结果显示，预测结果也在可接受范围之内，除去当食宿家庭学杂费过大时，偏差比例会高达 25.82%，其他情况，均不超过 17%。

对比上述两个模型，6.1 的优势在于模型简单，预测结果的偏差比例平均在 10%左右，没有特别精确的预测结果，但是偏差较平稳；6.2 的优势在于有些预

测结果十分精确，有五组偏差在 5%之内，但是模型较复杂一些，可以根据实际情况，自行选择。

七. 问题四:合理调整各类家庭教育投入比例的建议

根据对于前三个问题的分析研究,我们可以将家庭教育投入,分为四种类型,传统教育支出、基本教育支出、加强教育支出和新型教育支出。

针对新型教育支出,家庭应该引起最大程度的重视,应该重点引导培养学生的自主学习能力和多方面发展,共同促进学生学习成绩的提升。调整一些教育投入比例在购置电脑、乐器和体育用品等设备方面,方便学生线上查阅世界各地的学习资料,丰富培养学生的课外兴趣,但也要切记,过犹不及。

针对传统教育支出和基本教育支出,是每一个家庭的必备投入,不必攀比,但是也要适当投入。

针对加强教育支出无论是学区房的租买,还是一对一昂贵的家教,亦或是文体费,都不是每一个家庭必须的选择,但是可以对学生的学生成绩,起到锦上添花的作用。

八. 模型评价

8.1 模型优点

本文分别通过最小二乘方法和Logit 回归模型,对数据进行了合理拟合,对问题进行了定量描述,结果理想。前者模型简单,预测结果没有特别精确但是偏差较平稳;后者的优势在于有些预测结果十分精确,模型较复杂一些,可以根据实际情况,自行选择。

8.2 模型缺点

本文在收集数据上,存在一定的局限性,不能做到无差异抽样,而且数据结果表明,成绩偏好的学生较多,忽略了交互作用,模型可以进一步优化。

参考文献

- [1]姜启源, 谢金星, 叶俊. 数学模型(第四版)[M]. 北京: 高等教育出版社, 2010.
- [2]张尧庭. 多元统计分析引论[M]. 湖北: 武汉大学出版社, 2013.
- [3]杨坚争, 郑碧霞, 杨立钊. 基于因子分析的跨境电子商务评价指标体系研究[J]. 财贸经济, 2014(09):94-102.
- [4]施朝健, 张明铭. Logistic 回归模型分析[J]. 计算机辅程, 2005(03):74-78.

2018S. -T. Yau High School Science Award

附录

图三前两个图的 MATLAB 程序

```
X=[565 978 1308 1800 2248 4234]
Y=[0.37 0.49 0.42 0.46 0.44 0.34]
y1=scatter(X,Y)
grid on
figure(1)
p=polyfit(X,Y,2)
x2=500:10:4500
y2=polyval(p,x2)
figure(2)
plot(X,Y,'o',x2,y2)
xlabel('x')
ylabel('y')
dety=Y-polyval(p,X)
dety2=sum(dety.^2)
disp(sprintf('误差的平方和=%d',dety2));
legend('散点图','二次拟合')
```

图三第三个图的 MATLAB 程序

```
X=[565 978 1308 1800 2248 4234]';
Chd=[37 49 21 23 22 12]';
Total=[100 100 50 50 50 35]';
Proport=Chd./Total;
[b,dev]=glmfit([X X.^2],[Chd Total],'binomial','logit')
logitFit=glmval(b,[X X.^2],'logit');
plot(X,Proport,'o',X,logitFit,'r-');
xlabel('金额');
ylabel('Proportion of 排名')
b,bi=stats.se,dev
yhat=glmval(b,[X X.^2],'logit',stats,0.95)
```