

MOOC 视频观看情况与课程得分探究

大作业说明文档

2015012623 耿紫音

一、问题提出

本程序的思路来自于现实情况和数据资料：从理论上来说，对于 MOOC 课程的学习，视频是课程内容的主要来源，因此 MOOC 看得越多，学生对课程内容理解越充分，相应课程得分就应该越高。但从现实生活中看，很多人学习 MOOC 课程时，往往受到课业压力、学习动力、话题兴趣、视频不计成绩等因素的影响，难以将每个视频都完完整整看一遍，尽管如此，还是有很多人天资禀赋，或凭借小号、朋友的帮助等原因完成小测、期末考试，并在总评拿到不错的分数；从 2016 年计算机文化基础 MOOC 部分得分数据来看，无论最后学习成绩是 A、B、F，又抑或是清华本校学生、外校学生、社会学生，不同选课人群中都有“异常现象”存在——即没怎么看视频、但课程总评分数名列前茅，或看过不少视频，但课程成绩不尽理想。

因此，我们做出假设，视频观看情况和课程得分存在一定程度的正相关关系；但相关关系不强。

基于以上猜测，本程序主要从不同指标的描述性统计特征入手，分层分析观看视频和课程成绩之间的相关关系，验证假设，并进一步对不同学生的学习效率进行分析，从而在学生学习方法、MOOC 课程设置等方面提出自己的建议。

二、代码实现

0. 数据预处理&清洗

首先，在程序中对表格进行行读取、拼接、数据类型转换。

其次，排除已退课、半退课的账号。其中半退课指的是虽然在选课状态，但没有任何作业完成情况的账号，具体体现在无期末成绩、无平时作业。应注意的是，本筛选没有对视频完成情况进行任何要求，即肯定了“天才”的存在——“我不看视频，但我能答对题”，但一定程度上也纳入了选课用户的“小号”——“我不看视频，我答题只是来刷正确答案”。

最后，对过滤掉账号后的选课学生数据，画出不同分数区间课程得分分布图（清华学生：紫色；校外学生：绿色；社会学生：蓝色）。横向分层对比发现，过滤掉一部分账号

之后，聚集于 0 分的大量数据点消失了，三类学生的成绩分布形态基本一致，均为单峰分布。值得注意的是，社会学生的整体成绩比预期要好得多，与清华学生一样在高分区出现峰值，优于外校学生。

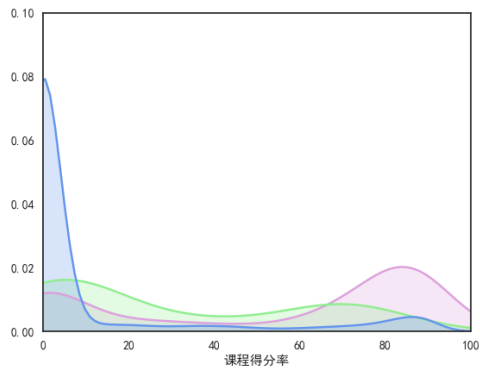


Figure 1 清洗前数据

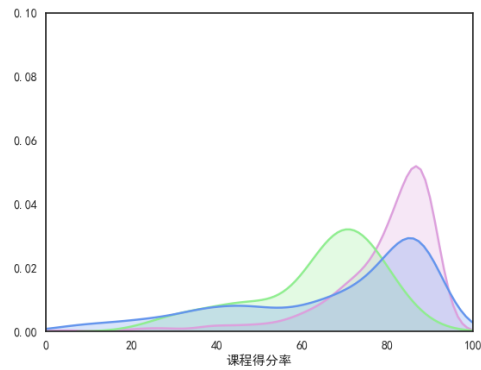


Figure 2 清洗后数据

1. 描述性统计

在数据清洗的基础上，进行不同得分项纵向对比，画出课程得分率、课程学习比例、习题作业的得分率、理论考试的得分率四组数据的箱线图。

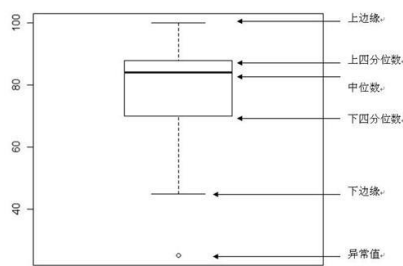


Figure 3 补充-箱线图介绍

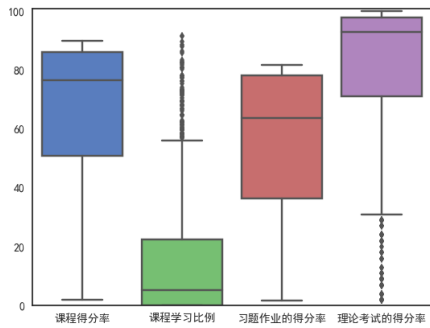


Figure 4 清华学生各部分得分箱线图

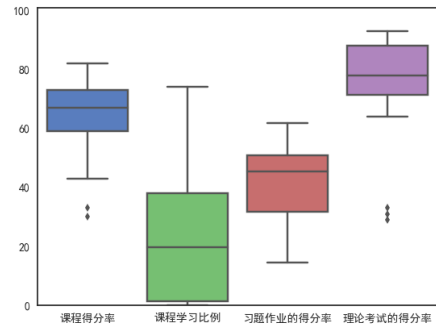


Figure 5 校外学生各部分得分箱线图

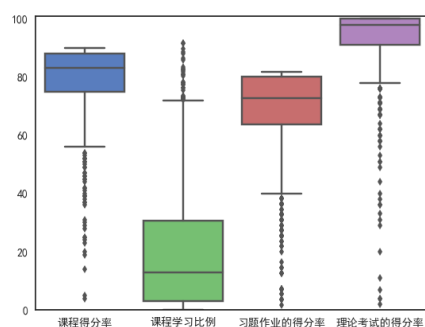


Figure 6 社会学生各部分得分箱线图

可以看出，相比而言，课程得分率（总评）、习题作业的得分率（平时作业，占总评 45%）、理论考试的得分率（期末考试，占总评 55%）均较高，而课程学习比例（视频观看情况，不计入总评）在不同类型学生中普遍较低。

以上结果反映出，从整体完成情况来看，学生在课程学习中不同部分的投入上，课程学习上的投入相对其他部分较少。特别通过在不同类型学生的横向比较，我们发现清华学生这一问题尤其严重。

2. 数据分析及可视化

2.1 相关关系

在数据预处理和描述性统计的基础上，绘制出每一类型学生课程观看情况与最后课程得分的相关关系。

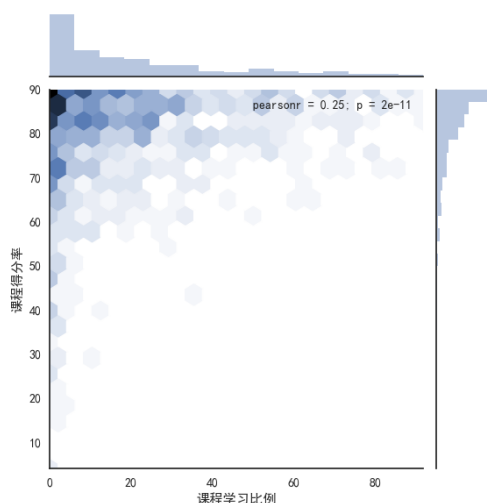


Figure 7 清华学生视频观看与课程得分相关关系

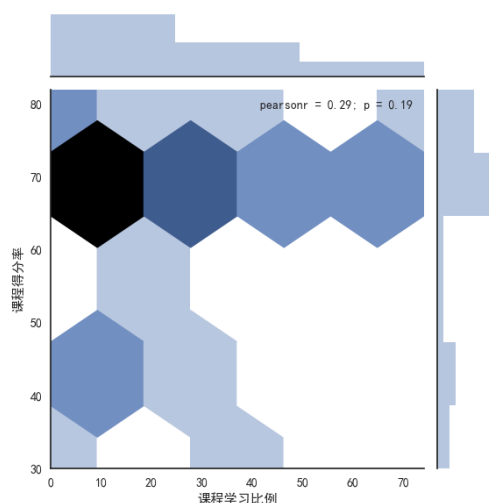


Figure 8 外校学生视频观看与课程得分相关关系

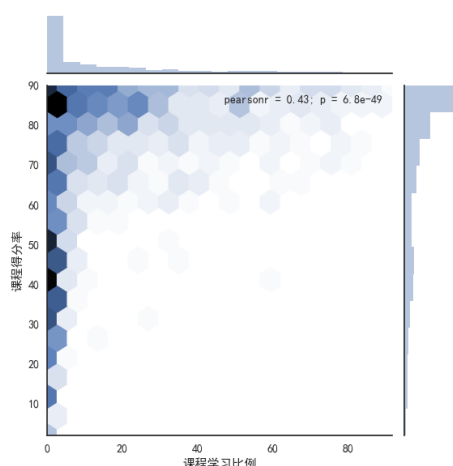


Figure 9 社会学生视频观看与课程得分相关关系

可以看出，对于每一类学生，视频观看情况和课程得分率的关系基本符合问题假设：

课程学习比例和课程得分率在清华学生和外校学生中，存在弱相关关系(pearsonr: 0.20-0.40)，而对于社会学生，两个指标有中等相关关系(pearsonr: 0.40-0.60)。

总体来说，课程学习比例和课程得分率有较弱的正相关关系。

2.2 学习视频效率分析

在相关关系基础上，我们进一步对选课学生视频学习效率进行了探究。

定义“视频得分比=课程学习比例/课程得分率-1”，该指标反映不同选课学生看视频对最后得分的效率程度。

若视频得分比 <0 ，则说明课程学习比例小于课程得分率，即近似认为该学生通过更少的看视频付出，收获了更高的分数；而若视频得分比 >0 ，则说明课程学习比例大于课程得分率，即近似认为该学生学习视频的效率不高；若视频得分比恰好 $=0$ ，课程学习比例=课程得分率，则近似认为“一分耕耘，一分收获”。视频得分比越高，说明学习视频效率越低；视频得分比越低，说明学习视频效率越高。特别的，视频得分比最小值为-1，此时课程学习比例为0，但课程仍有得分（在平时作业或期末考试中取得成绩），可以说是“天才中的战斗天才”了——但也不排除单纯只是其他学习用户的“小号”；同时，视频得分比无最大值。

绘制出学生课程得分率与视频得分比分布的散点图，并用红黄蓝线画出分数引导线（ $x=60$ 分，90分，100分），用绿线画出效率引导线（ $y=0$ ），如下图所示：

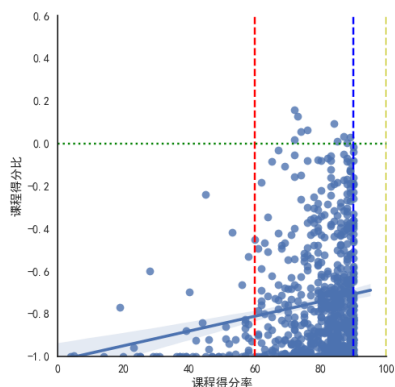


Figure 10 清华学生视频学习效率分布

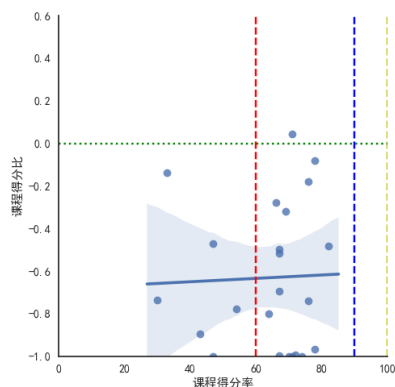


Figure 11 校外学生视频学习效率分布

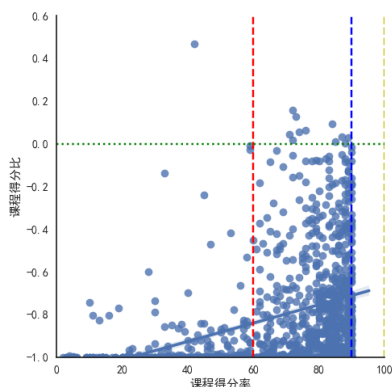


Figure 12 社会学生视频学习效率分布

可以看出：

①大部分学生视频得分比还是分布在 <0 的区间，说明学生视频学习效率普遍较高，均能通过观看较少的视频获得更高比例的分数。

②有大量视频得分比数据为-1，“天才”和“小号”——都不少。

③横比来看，清华学生普遍学习效率、课程得分最高，体现在视频得分比基本上分布在 <0 的区间，而课程得分基本分布在 >80 的区间；外校学生学习效率较高（ <0 ），但课程得分不高（大多分布在60-80的区间，且没有90+）；相比来说，社会学生学习效率不如清华、外校学生高，但课程得分较高，和清华学生一样基本上集中在80-100的区间。

	课程得分（课程得分率）	学习效率（视频学习比）
清华学生	+	+
外校学生	-	+
社会学生	+	-

Table 1 不同学生得分、效率比较（相对值）

④随着分数的提高，视频得分比有增加的趋势，即成绩越高，学习效率趋向于更低。这是符合模型建立的，因为视频得分比随着课程得分率的增加有边际递减的趋势——具体

而言，学生可能只用看一两个视频，就可以取得不错的分数，但要想使课程得分从不错到优秀，只有加倍看更多的视频。

2.3 学生类别分析

在以上分析基础上，选择社会学生课程得分率和视频得分比，对不同学生类型进行划分：（仅供娱乐，切莫对号入座）

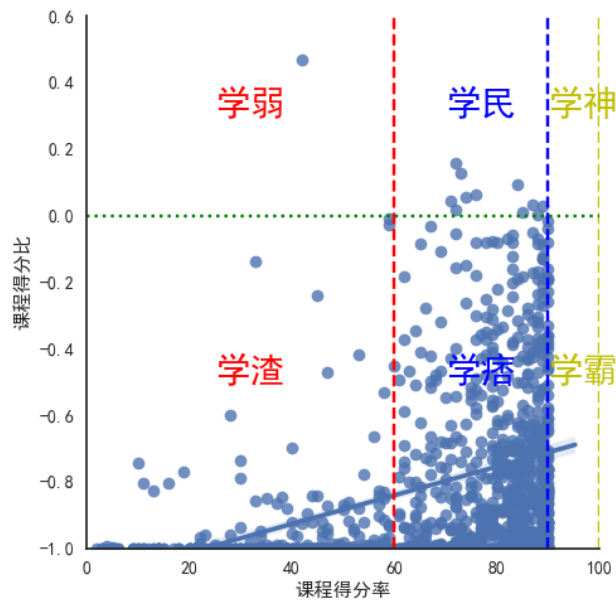


Figure 13 学生类型

课程得分率	F: 0-59	B: 60-89	A: 90-100
视频得分比>0 (低效)	学弱：熬油点灯，兢兢业业，学秃了也不能拯救我的 F	学民：好好学习，天天向上，跪求老师手下留情	学霸：热爱学习，日益消瘦
视频得分比≤0 (高效)	学渣：刻苦学习，效率也不低，但可能就是不及格的命	学痞：上课睡觉，下课玩闹，但成绩仍然不错	学神：爱学习，会学习
视频得分比=-1	学婊：我从不看视频，但我就是能得分，甚至满分（你来打我呀）		

Table 2 MOOC 学生图鉴

三、总结

从以上结果可以看出，学生学习 MOOC 课程时视频观看情况与最终课程得分有一定程度的正相关关系，但囿于种种因素，“事半功倍”、“事倍功半”的异常现象数见不鲜，这一结果将对 MOOC 选课学生、开课老师都有所启发。

对于选课学生来说，首先学生应该提升对 MOOC 视频重要性的认识，多看视频，提高成绩。第二，学生不应该仅仅追求课程成绩的提高，也应该时时反思自己的学习效率，思考自己是否花了更多的努力取得一般的学习效果，并改进自己的学习方法。

对于老师来说，老师可以通过这一结果对学生水平、视频安排有进一步的思考，相应在今后的课程安排中调整视频内容，在课程考核中加入视频完成程度的衡量，同时加强对小号刷题的监控。

同时，本程序实现过程中仍存在一些继续改进的空间：如没有排除小号因素的影响（可以近似认为没有视频播放记录，最后只有几个作业有成绩的账号为小号，但应明确小号也是本研究组成的重要部分），没有分析视频观看时间对课程得分的影响（如学生是每周坚持看视频、循序渐进，还是考试前填鸭看视频等，目前暂无数据），都有待后续进一步挖掘研究。