**一、调查方面：**

数据来源及覆盖范围（困难程度，多样性）：我组对国际关系学院、黑龙江大学、北京大学、东北财经大学、合肥工业大学等来自29个省市自治区XXXX个全国普通高等院校的大学生进行随机性匿名问卷调查和访谈。接受调查共352人。调查对象 中男女生比例为1:1.77, 文科类、理工类比例为1.41：1。选取的数据比较合理。 主要采用网络问卷调查法。通过微信朋友圈，QQ、EMAIL、留言板等方式让各大高校的部分学生网上答卷。二是问卷调查法。问卷题型主要采用选择题的形式。本次调查共收回问卷发103份，有效问卷103份；三是一对一谈话或网络视频等形式获取信息，三种形式形成互补，相互促进，使我们的研究更具客观性。

二、EXCEL方面：

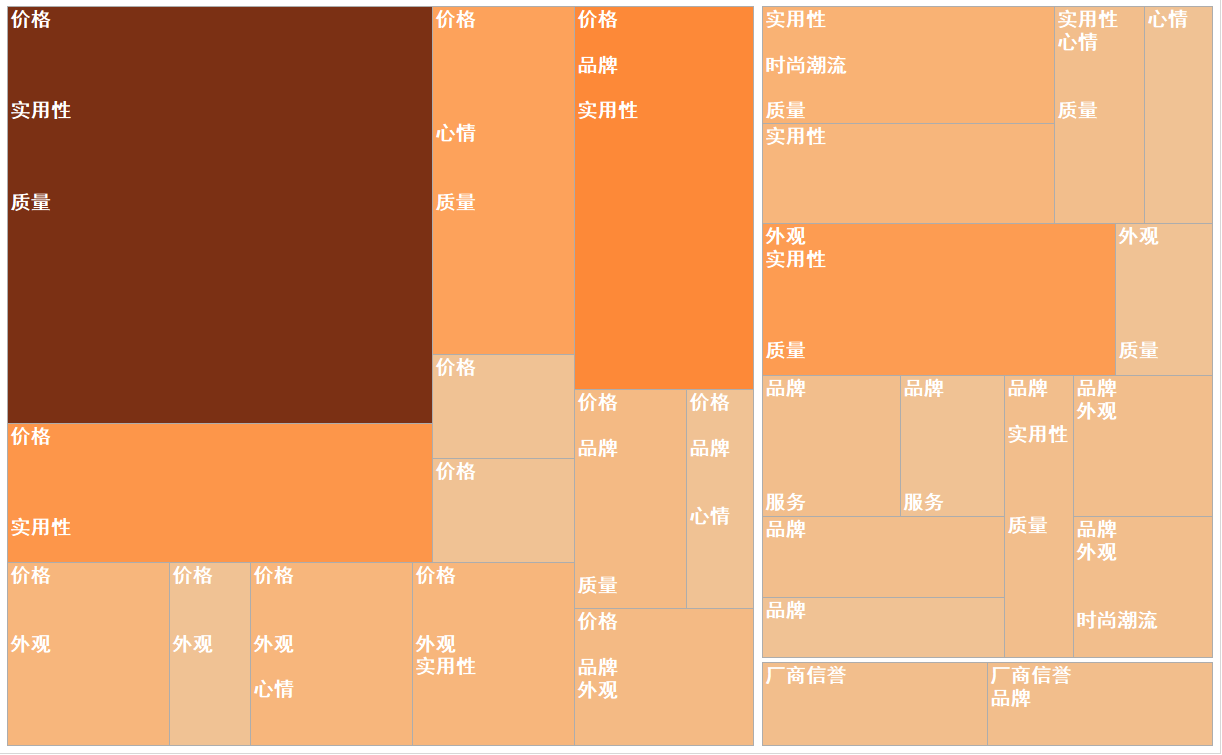
使用EXCEl录入并统计数据。分析数据特点； 1.数据选择： 样本：均为在校大学生。大一大二学生居多，大三大四学生偏少。分类：把大学生消费分为了基本生活消费、休闲娱乐消费、形象消费、学习消费、恋爱消费五个主要方面。又调查分析了影响消费的因素，记账情况、冲动消费和后悔情况，理财能力的自我评分等数据。2. 类型：包括离散型数据111个，连续型数据11个，文字类数据3.数据处理总量为：249kb。

三、TABLEAU方面：  
1. 大图：地图（全国各个城市）

调查对象 中男：女= 1：1.77， 文科：理工科= 1：1.41

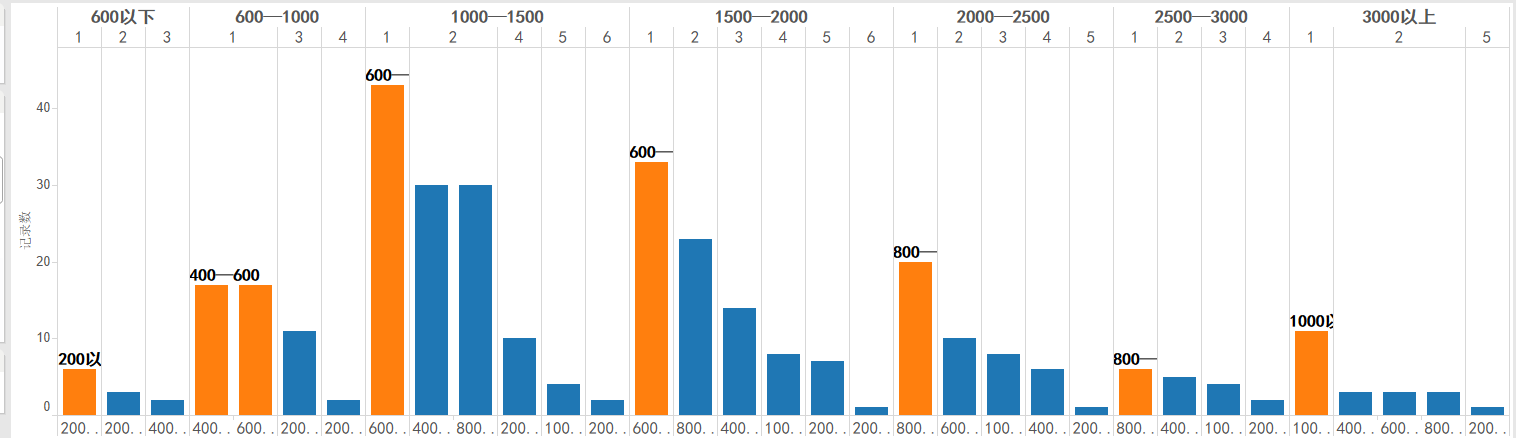


2. 突出显示表（消费因素）

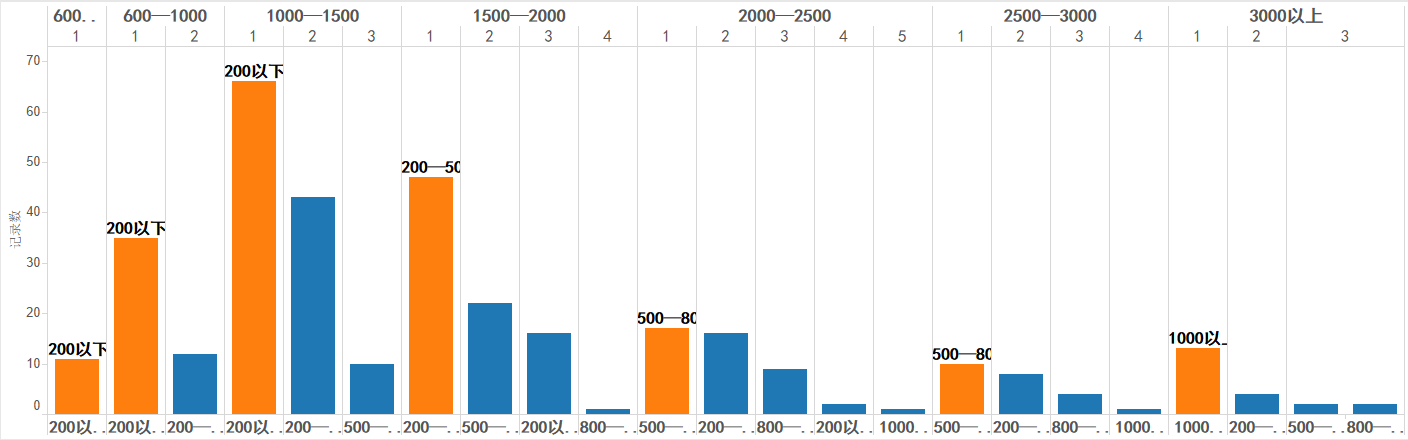


3. 在月总消费与月基本消费等消费的关系 ，橙色突出显示最多的

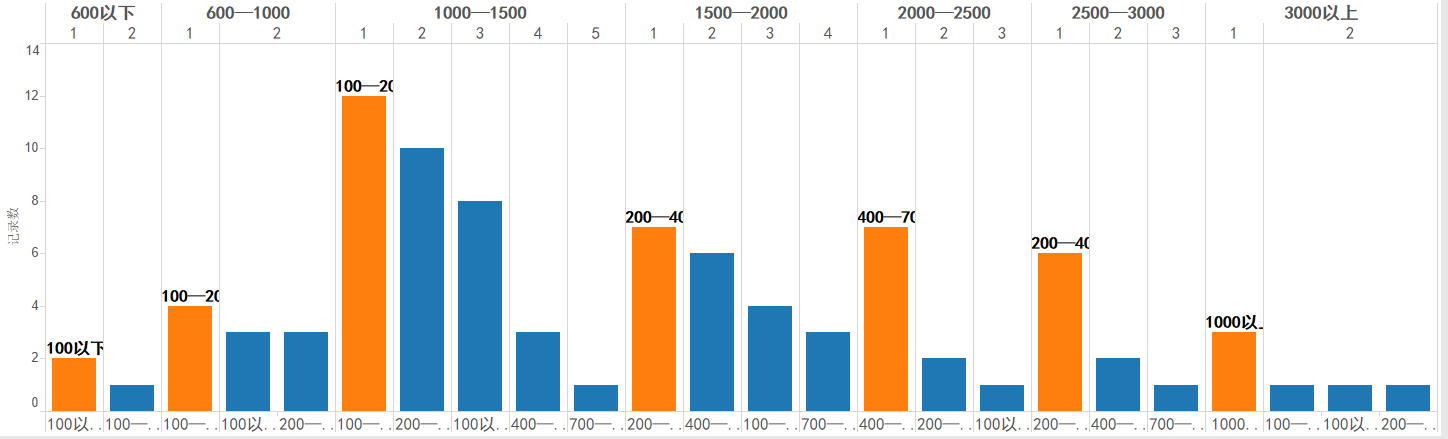
月均基本生活消费 随着月总消费增大而增大



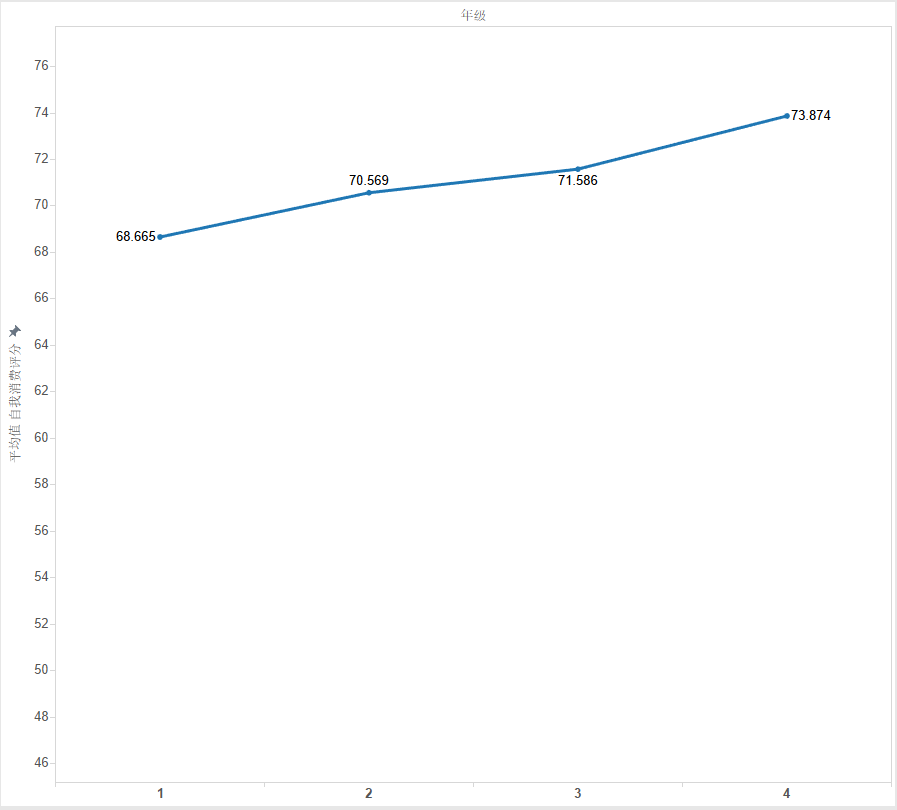
月均休闲娱乐消费



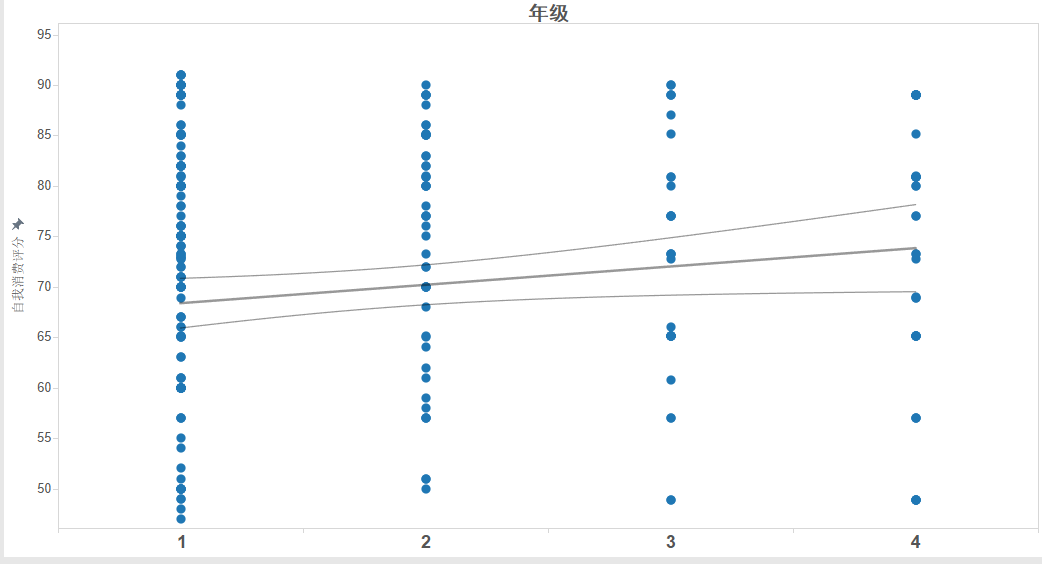
月均恋爱消费



4. 自我消费评分折线图（从大一到大四）



5. 散点图 大学生自我消费评分随入学时间的变化（大一到大四）



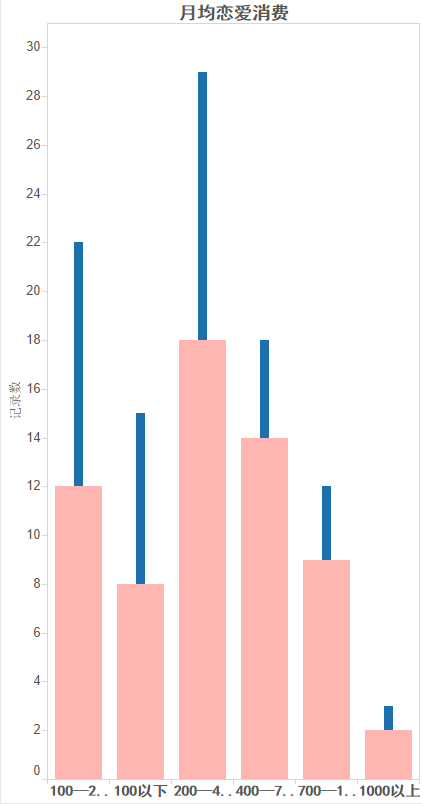
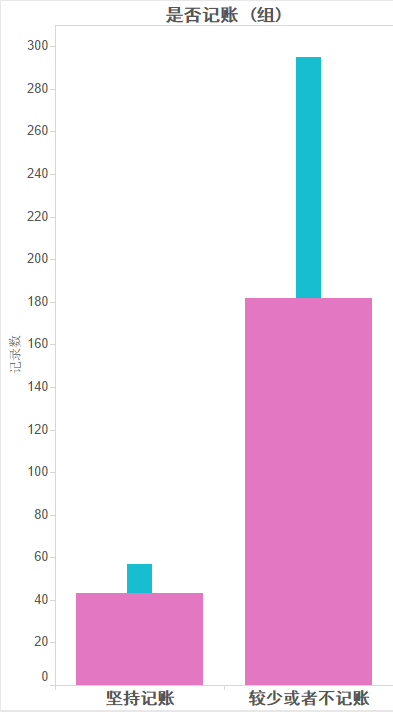
针对 自我消费评分 (给定 年级) 计算线性趋势模型。

公式: 自我消费评分 =1.81782\*年级 日期 +66.5604

R 平方值: 0.0133667

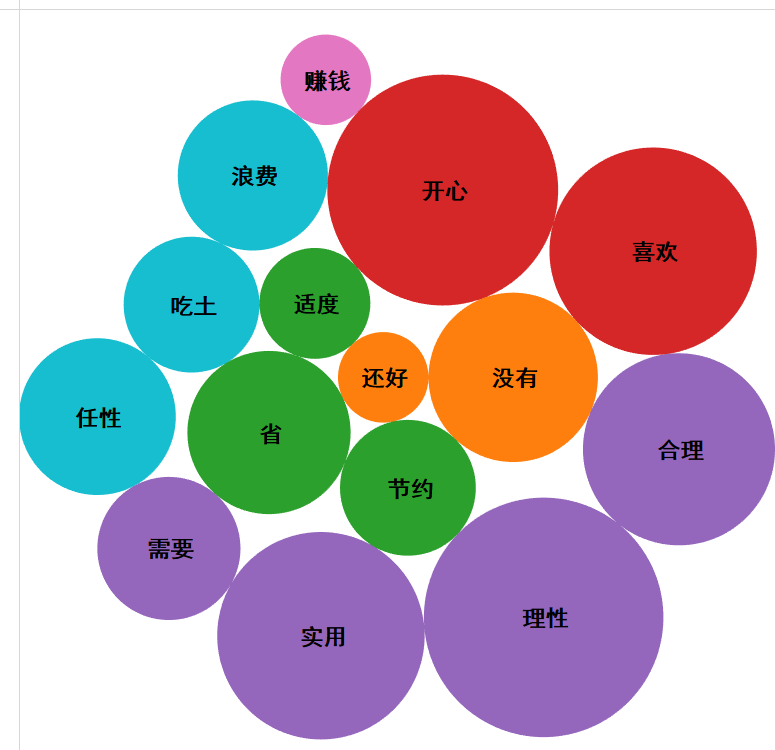
P 值: 0.0454079 < 0.05表示该模型有统计学意义

6. 对比柱状图

图表结合问卷的男女比专业比，可得出，在高恋爱消费的学生中，文史类学生比理工类多。坚持记账的女生比男生多。

7. 关键字气泡图



大学生对自己消费观念的一句话总结可以提炼出如下的高频关键字。主要包括5个方面，不同类的用颜色区分。

1. 无观念派(橙色)2.理性消费派（紫色）3.勤俭节约派（绿色）4.购物享乐派（红色）5. 月光派（蓝色）

新的消费名词：吃土

学生趋势词：赚钱

**四、 SPSS方面**

先量化数据

1. **问卷的信度效度分析**

借助SPSS进行信度分析，可以通过信度分析判断问卷问题测量的稳定性和可靠性

计算目前最常用的信度系数克朗巴哈系数，其结果如表1所示。

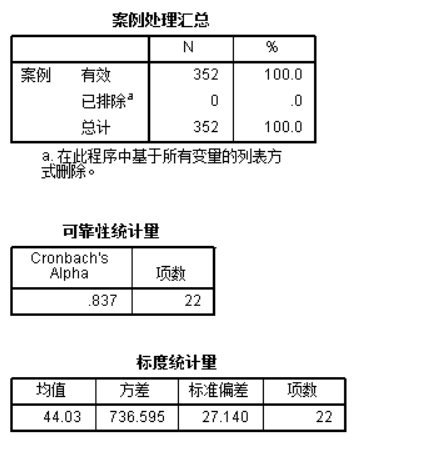
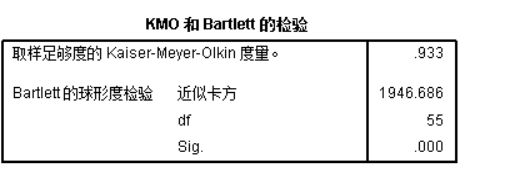


表1显示，Cronbachˊs Alpha系数为0.83大于0.8，可以推断此问卷的可信度较好，该评价问卷具有较高的内在一致性。

接着用SPSS对其进行效度分析。效度分析使用的是因子分析模型，分析结果如表2所示。



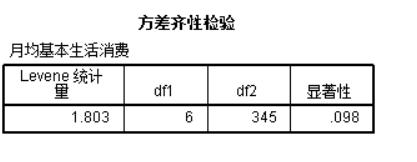
由表2的数据可知，问卷数据的KMO值为0.933，显著性水平为0.000小于0.005，通过了巴特利球型检验，说明大学生消费水平调查问卷是具有效度的。

1. **单因素方差分析**

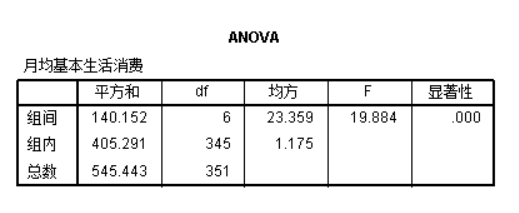
**方差分析是检验多个总体均值是否相等的统计方法.它是通过检验各总体的均值是否相等来判断分类型自变量对因变量是否有显著影响.**

月均总消费与月均基本生活消费的相关性(月均基本生活消费 随着月总消费增大而增大)

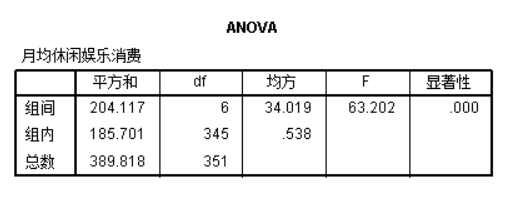
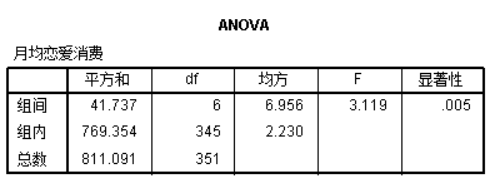
1. 先经过方差齐性检验：如果显著性都没有达到最小值0.05，就代表不显著，这证明方差是齐性的，可以继续进行单因素方差分析。

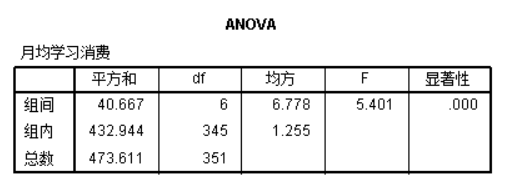
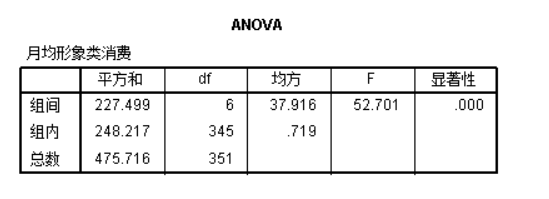


1. 显著性=0.000<0.05代表月均总消费对月均基本生活消费有显著影响。



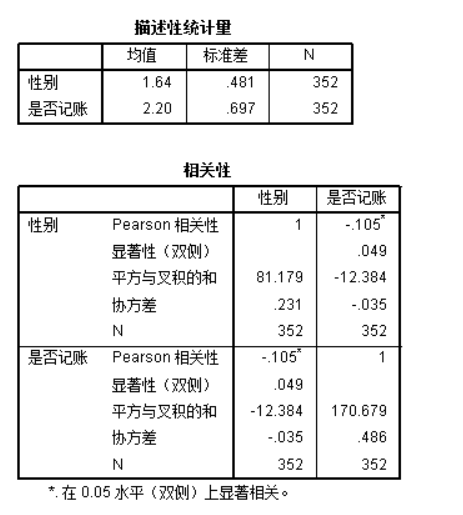
用同样方法分析其它消费：显著性均小于0.05

 ，



1. **双变量相关性分析——pearson相关系数检验**

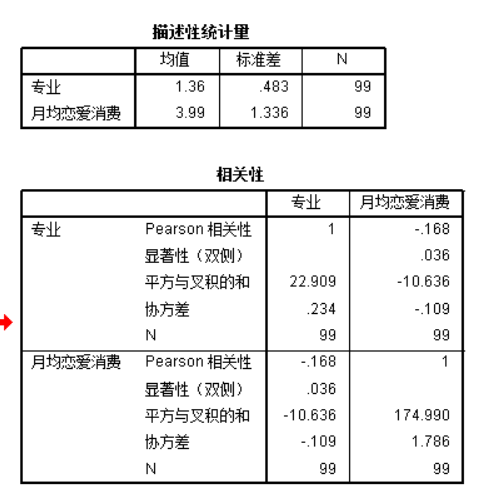
坚持记账的女生比男生多。



显著性（双侧）为0.049<0.05可证得记账情况与性别相关。

高恋爱消费的学生中，文史类学生比理工类多。

先使用SPSS选择个案，选出高恋爱消费的学生，再对文理科专业分组。最后使用双变量相关性分析。



显著性（双侧）为0.036<0.05可证得高恋爱消费情况与专业相关。

**4. 二元logistic回归模型**本文根据问卷统计结果，建立了二元logistic回归模型。 选取20个 是否记账 各个方面月平均消费 后悔程度 冲动消费 各项消费因素排序 的列数据，借助SPSS软件，运用二元logistic 回归分析的方法建立模型。下面以工科学生为例，对学生所在专业建立回归模型并求解（如表3）：



从拟合结果来看，月均总消费，休闲娱乐消费，学习消费和形象消费的sig显著值远小于其它自变量，故筛选出他们和常量建立模型的拟合方程式：

设四个自变量依次为X1，X2， X3，X4，

设得到模型的拟合方程式为：

 （1）

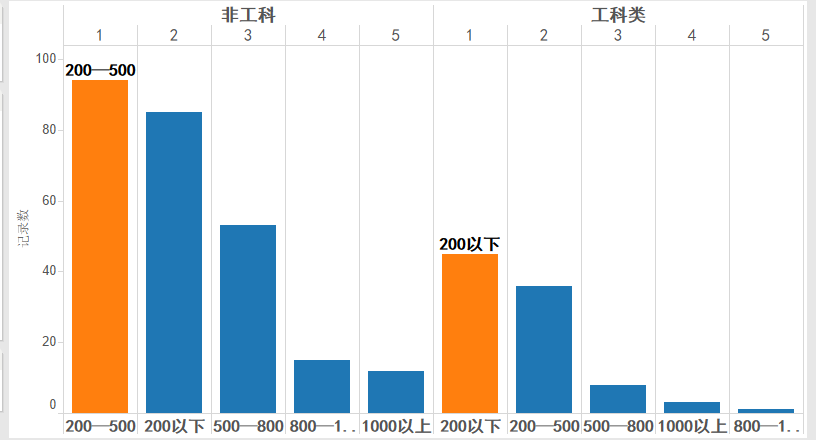
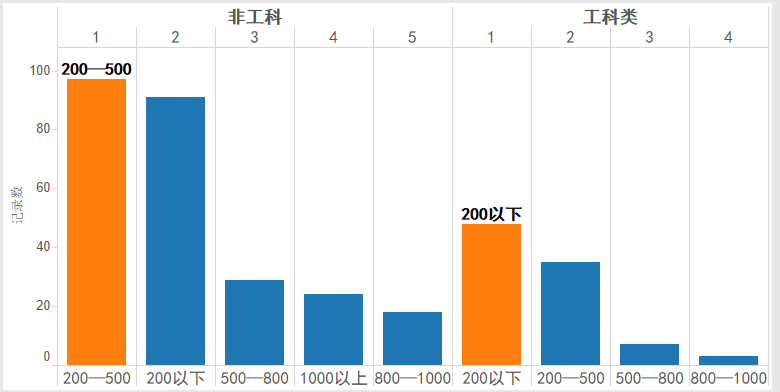
 (2)

用该方程做预测，若y1>0.5，则符合工科大学生的消费水平特征；反 之，则不符合。

按照上述分析方法， 本文依次对其余8个专业建立logistics模型并进行求解，得出相应的拟合方程式。

其中工科大学生在月均休闲娱乐消费和形象消费上sig值最低，小于0.05，可以认为具有高显著性和统计学意义。这表明工科学生在这两项花费上明显不同于其他系。

通过Tableau图表可以观察到工科学生确实在休闲娱乐和形象上花费低于平均水平。

结论：

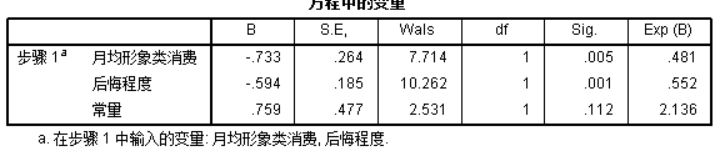
类似的，根据回归分析得出的EXP（B）指标，我们得出了一些有趣的结论：

1. 工科学生在休闲娱乐和形象方面花费低于平均水平。
2. 经济系大学生衣着打扮类开销大，且购物后很少后悔
3. 管理专业大学生更重视厂商信誉，且学习开支大
4. 读文史学的大学生购物时普遍看重服务，重视他人意见
5. 法律系学生看重价格，较少冲动消费
6. 理科生购物后很少后悔，且月总消费水平偏高

其余专业的大学生样本较少，不具有统计学意义

【附各专业分析图表】：

**经济学学生**

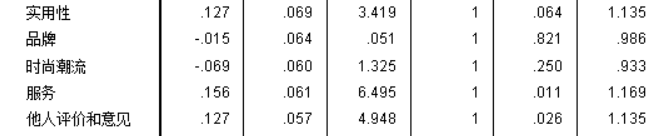


**管理类**



**文史学类：**

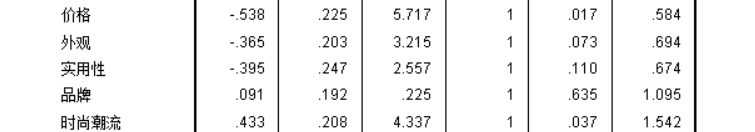
**（文化传播文学和政治学和语言）**



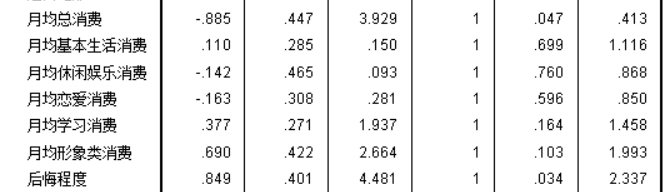
**工科**



**法学**



**自然科学**



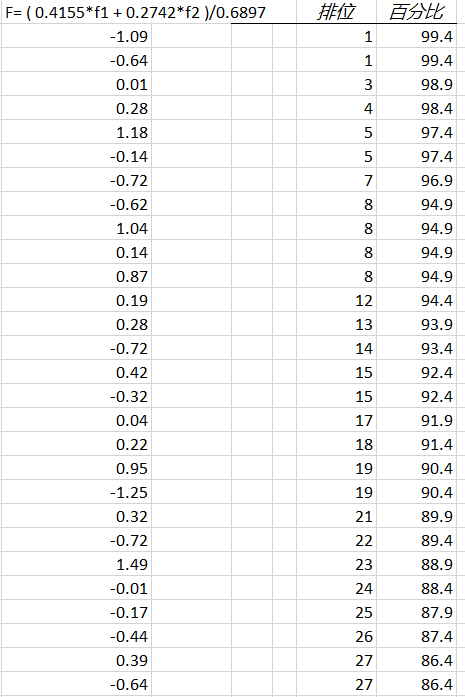
**五、模型建模——智慧消费指数**

精选了200份问卷（剔除自我消费评分偏离程度较大杂点），以自我消费评分，记账情况，冲动消费和后悔程度4个参数为自变量建模。最后得出了分布范围为0-100分的智慧消费指数。分数越高表示该大学生消费越理性。比如得了60分，则你在智慧消费方面打败了60%的小伙伴。

提取出两个主成分：成分1包括后悔程度和个人消费评分。成分2包括记账情况。



最后以各成分的方差贡献率占三个成分总方差贡献率的比重作为权重进行加权汇总，得出综合得分f : F = ( 0.4155\*f1 + 0.2742\*f2 )/0.6897



7.总结：数据的灵活性（处理方式多种多样扩展性：就是关于如何处理更大规模的数据。）；可开放性：扩大数据的开放程度，却又在保护个人隐私的基础上提出问题，不记录IP地址，手机号等针对个人身份信息（问卷星）；数据会影响到我们现在生活的哪些方面；