

基于交叉列联表分析和回归分析的大学生月消费水平相关因素影响实证分析

李昊哲, 李秀森*, 于 淼

山东理工大学, 数学与统计学院, 山东 淄博
Email: *leexiusen@163.com

收稿日期: 2021年7月22日; 录用日期: 2021年8月16日; 发布日期: 2021年8月23日

摘 要

对山东理工大学的202名同学进行问卷调查, 总共回收有效调查问卷202份, 结合问卷调查的数据对问卷的每个选项赋予数值, 导入spss, 运用描述统计、交叉列联表分析、相关分析、回归分析的方法, 对大学生的月消费水平影响因素相关性进行研究分析。

关键词

大学生月消费水平, 交叉列联表分析, 回归分析, 方差分析

Empirical Analysis Based on Correlation Analysis and Cross-Correlation Table Analysis of College Students' Monthly Consumption Related Factors

Haozhe Li, Xiusen Li*, Miao Yu

School of Mathematics and Statistics, Shandong University of Technology, Zibo Shandong
Email: *leexiusen@163.com

Received: Jul. 22nd, 2021; accepted: Aug. 16th, 2021; published: Aug. 23rd, 2021

Abstract

A questionnaire survey was conducted with 202 students from Shandong University of Technology.

*通讯作者。

文章引用: 李昊哲, 李秀森, 于淼. 基于交叉列联表分析和回归分析的大学生月消费水平相关因素影响实证分析[J]. 统计学与应用, 2021, 10(4): 714-720. DOI: 10.12677/sa.2021.104073

A total of 202 valid questionnaires were collected. The data from the questionnaire was combined with a numerical value for each option of the questionnaire; spss were imported; descriptive statistics were used; cross-contingency table analysis and correlation analysis were used. The regression analysis method is used to research and analyze the correlation of the influencing factors of the monthly consumption level of college students.

Keywords

College Students' Monthly Consumption Level, Cross-Contingency Table Analysis, Regression Analysis, Variance Analysis

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

目前,国内外对于大学生消费的影响因素主要集中于消费心理、消费行为、消费结构进行研究分析。胡娟[1]就对当代大学生的消费水平、消费结构、消费方式、消费心理特点及其影响因素及对策做了分析;刘莉[2]通过研究大学生消费行为的调查分析,道出大学生消费行为受到家庭、社会、同学的影响;张亮亮[3]通过假设检验,拟合优度检验方法分析了男生与女生月消费均值以及非理性消费比例。Soycon Shim [4],从大学生的消费行为角度分析了大学生月消费影响因素。

论文主要从大学生的日常消费内容,以及大学生的平时学习成绩与学习习惯的角度来探寻影响大学生月消费的因素。以山东理工大学数学与统计学院的学生为调查对象,设计问卷调查,回收有效问卷 202 份。

2. 各影响因素与月消费水平的交叉列联表分析

根据收集到的样本数据编制交叉列联表,在交叉列联表的基础上,进行多个变量在不同取值情况下的数据分布情况,从而进一步的深入分析变量之间的相互影响和关系。

Table 1. Cross-contingency table analysis

表 1. 交叉列联表分析

	在校期间的平均月消费				总计
	1000 内	1000~1500	1500~2000	2000 以上	
大一	8 36.40%	8 36.40%	2 9.10%	4 18.20%	22 100.00%
大二	6 15.80%	22 57.90%	10 26.30%	0 0.00%	38 100.00%
大三	4 7.10%	26 46.40%	18 32.10%	8 14.30%	56 100.00%
大四	10 11.60%	40 46.50%	20 23.30%	16 18.60%	86 100.00%
男	14 14.90%	52 55.30%	16 17.00%	12 12.08%	94 100.00%
女	14 13.00%	44 40.70%	34 31.50%	16 14.80%	108 100.00%

年级和性别与在校期间的平均月消费的交叉列联表(见表 1)可分析出皮尔逊相关系数(见表 2)与肯德尔系数的 P 值(见表 3)分别为 0.268 和 0.119。无法拒绝这两个变量间不存在相关性的原假设。

Table 2. Chi-square test

表 2. 卡方检验

	值	自由度	渐进显著性(双侧)
皮尔逊卡方	11.119a	9	0.268
似然比	12.92	9	0.166
线性关联	3.109	1	0.078
有效个案数	202		

Table 3. Symmetry measurement table

表 3. 对称测量表

		对称测量			
		值	渐近标准误差 ^a	近似 T ^b	渐进显著性
有序到有序	肯德尔 tau-b	0.142	0.09	1.56	0.119
	斯皮尔曼相关性	0.16	0.103	1.61	0.111c
区间到区间	皮尔逊 R	0.176	0.105	1.782	0.078c
有效个案数		202			

由学习成绩与在校期间的平均月消费的交叉列联表(见表 4)可分析出皮尔逊相关系数(见表 5)和肯德尔相关系数的 P 值(见表 6)分别为 0.22 和 0.936, 无法拒绝这两个变量之间不存在相关性的原假设。因此可以说明大学生月消费水平与学习成绩之间相关性较低。

Table 4. Cross table of average monthly consumption and academic performance

表 4. 平均月消费与学习成绩交叉表

在校期间的平均月消费与学习成绩交叉表					
		学习成绩			总计
		上游	中游	末尾	
在校期间的平均月消费	1000 以内	12	14	2	28
	1000~1500	16	70	10	96
	1500~2000	18	28	4	50
	2000 以上	10	12	6	28
总计		56	124	22	202

Table 5. Chi-square test

表 5. 卡方检验

卡方检验			
	值	自由度	渐进显著性(双侧)
皮尔逊卡方	8.260a	6	0.22
似然比	8.153	6	0.227
线性关联	0.011	1	0.916
有效个案数	202		

Table 6. Symmetry measurement table
表 6. 对称测量表

		对称测量			
		值	渐近标准误差 ^a	近似T ^b	渐进显著性
名义到名义	列联系数	0.275			0.220
有序到有序	肯德尔tau-b	-0.008	0.101	-0.081	0.936
	斯皮尔曼相关性	-0.009	0.112	-0.089	0.930 ^c
区间到区间	皮尔逊R	0.010	0.112	0.104	0.917 ^c
协议测量	Kappa	0.107	0.058	1.834	0.067
有效个案数		202			

a) 未假定原假设。

b) 在假定原假设的情况下使用渐近标准误差。

c) 基于正态近似值。

3. 各影响因素与月水平消费的回归分析

为了更精确的研究各因素与月消费水平之间的相关性，月消费水平与大学生学习成绩之间的相关性，以及影响大学生的月消费水平影响因素，计算了相关系数并进行了相关系数的显著性检验，最后为了更进一步了解各因素是怎样影响月消费水平的，进行了线性回归分析。

如表 7 所示，设因变量成绩为 Y ，在校平均月消费为 X_1 ，每月在恋爱方面的支出为 X_2 ，作业完成情况为 X_3 ，是否挂过科为 X_4 ，上课期间玩手机程度为 X_5 ，听课认真程度为 X_6 。

Table 7. Variables removed
表 7. 除去的变量

输入/除去的变量 ^a		
模型	输入的变量	除去的变量
1	听课认真程度	
2	每月在恋爱方面的支出	
3	是否挂过科	
4	上课期间会玩手机吗	
5	在校期间的平均月消费	
6	作业完成情况	

因变量：学习成绩

如表 8 所示，六个回归模型的回归系数的显著性检验。其中包括了标准化的 B 估计值和标准错误，还有 Beta 分布，还有对每一个回归系数的 t 检验的结果，以及共线性问题的相关统计量：容差和方差膨胀系数。

在模型中，在校期间的平均月消费的回归系数为 0.001， t 值为 0.556 ($p < 0.01$)，容忍度为 0.832，方差膨胀系数为 1.202 (容忍度越接近 0 表示变量之间存在线性关系，同时如果方差膨胀系数大于 10，则表示自变量之间存在线性问题)。

每月在恋爱方面的支出的回归系数-0.003， t 值为-0.784 ($p < 0.01$)，容忍度为 0.805，方差膨胀系数为 1.242。

Table 8. Coefficient
表 8. 系数

模型	系数 ^a					共线性统计	
	未标准化系数		标准化系数	<i>t</i>	显著性	容差	VIF
	B	标准错误	Beta				
(常量)	49.38	15.48		3.19	0.002		
在校期间的平均月消费	0.001	0.003	0.058	0.556	0.579	0.832	1.202
每月在恋爱方面的支出	-0.003	0.003	-0.083	-0.784	0.435	0.805	1.242
您的作业完成情况	0.406	0.156	0.311	2.605	0.011	0.633	1.579
是否挂过科	1.993	2.547	0.088	0.783	0.436	0.723	1.383
上课期间会玩手机吗?	0.121	0.092	0.131	1.309	0.194	0.902	1.109
听课认真程度	-0.271	0.147	-0.183	-1.846	0.068	0.921	1.085

作业完成情况的回归系数为 0.406, *t* 值为 2.605 ($p < 0.01$), 容忍度为 0.633, 方差膨胀系数为 1.576。是否挂过科的回归系数为 1.993, *t* 值为 0.783 ($p < 0.01$), 容忍度为 0.723, 方差膨胀系数为 1.383。上课期间玩手机程度的回归系数为 0.121, *t* 值为 1.309 ($p < 0.01$), 容忍度为 0.902, 方差膨胀系数为 1.109。听课认真程度的回归系数为 -0.271, *t* 值为 -1.846 ($p < 0.01$), 容忍度为 0.921, 方差膨胀系数为 1.805。综合容忍度, 方差膨胀系数两方面考虑, 再结合共线性诊断, 共线性问题不是很严重。

从系数表可得出回归方程为:

$$Y = 0.001X_1 - 0.003X_2 + 0.406X_3 + 1.993X_4 + 0.121X_5 - 0.271X_6$$

回归方程显著性检验结果表明选择此模型具有合理性。变量均通过了回归系数显著性检验, 所以都不应该被剔除。共线性检验结果表明它们之间存在较弱或中等程度的共线性(如表 9 所示)。

Table 9. Collinearity diagnosis
表 9. 共线性诊断

模型	特征值	条件指标	共线性诊断 ^a						
			方差比例						
			(常量)	在校期间的平均月消费	每月在恋爱方面的支出	您的作业完成情况	是否挂过科	上课期间会玩手机吗?	听课认真程度
1	6.198	1.000	0.00	0.00	0.01	0.00	0.00	0.00	0.00
2	0.626	3.146	0.00	0.00	0.78	0.00	0.00	0.00	0.00
3	0.099	7.927	0.00	0.27	0.09	0.00	0.44	0.00	0.00
4	0.052	10.970	0.00	0.62	0.07	0.00	0.34	0.06	0.01
5	0.016	19.659	0.02	0.04	0.00	0.02	0.01	0.89	0.10
6	0.006	32.992	0.00	0.00	0.03	0.72	0.16	0.04	0.50
7	0.003	42.483	0.97	0.07	0.03	0.25	0.04	0.01	0.39

a. 因变量: 学习成绩

4. 方差分析检验

为了进一步证实模型整体性显著。合理性, 进行了方差分析检验, 以及标准化残差图。

如表 10 所示, 模型中的 *F* 值为 3.075 ($p < 0.01$), 达到了显著水平, 即六个变量的回归系数均不为 0, 表明了该回归方程具有显著性, 所以该回归模型具有显著性。

Table 10. Analysis of variance
表 10. 方差分析

ANOVA ^a					
模型	平方和	自由度	均方	<i>F</i>	显著性
回归	2123.658	6	353.943	3.075	0.009b
残差	10,589.453	92	115.103		
总计	12,713.111	98			

如图 1 所示为回归标准化残差值直方图，此直方图是用来检验样本观察值是否符合正态性基本假设，曲线基本呈钟形，样本观察值分布接近正态分布。

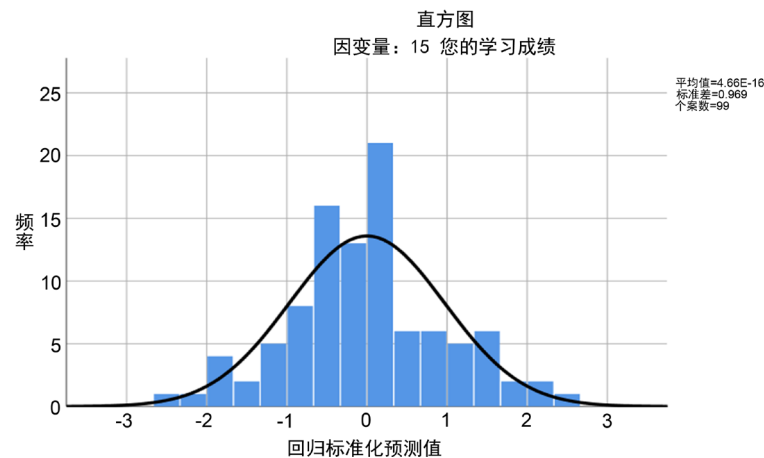


Figure 1. Histogram
图 1. 直方图

如图 2 所示为标准残差的正常概率分布图，也是用来检验样本的分布是不是符合正态分布的一种方法。分布图是一条从左下方到右上方的一条呈 45° 的直线，符合正态分布的标准。

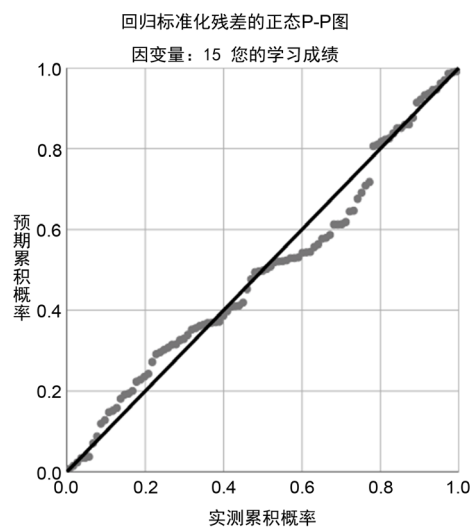


Figure 2. Normal P-P plot of regression standardized residuals
图 2. 回归标准化残差的正常 P-P 图

如图 3 所示的标准化残差值和标准化预测值交叉分布图。用来检验样本观察值是否符合正态分布以及残差值是否符合方差齐性的假定。散点图在 0 值上下水平随机分布，基本符合正态分布的同时也是齐性的。

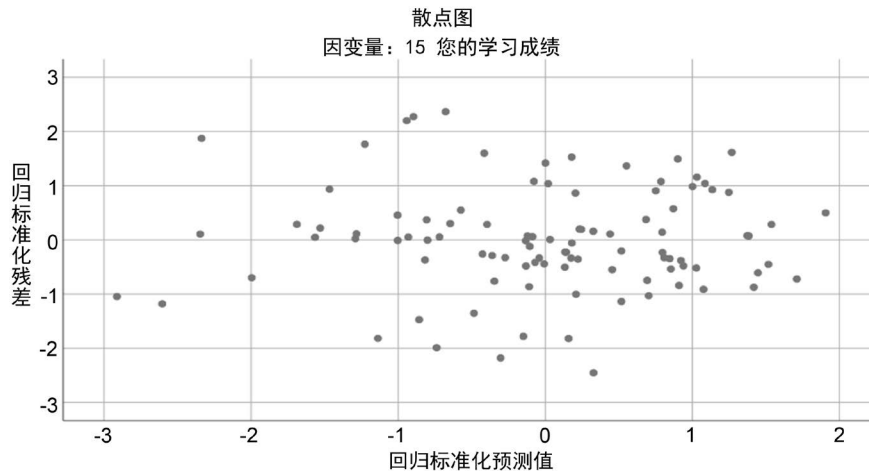


Figure 3. Scatter plot
图 3. 散点图

5. 结论

1) 由交叉列联表分析可以看出大学生月消费支出与大学生所在的年级之间不存在显著性的相关性，年级不同的差异，消费也会有所不同。并不是随着年级的增加，大学生的消费水平也会增加。

2) 性别与消费水平交叉列联表可知男生与女生的消费水平是不同的，在 1500 以内的男生占比要比女生高，而高于 1500 的消费称之为高消费，在这高消费的人群中女生要比男生占比多。

3) 线性回归研究结论中可以看出，月消费水平，恋爱方面的支出，作业完成情况，是否挂过科，听课认真程度与学生的成绩是有线性关系的，学习成绩好的学生往往在恋爱方面会有较少的支出，当然他们的作业完成情况也是较好的。

4) 由标准化预测值可以看出大学生的月消费水平普遍向更高的方向发展，这也许可能与社会的日新月异有关，大学生的需求也变得更高。

5) 在大学，学生应以学业为主，大学生应该合理的消费，着重注重学习，学习是我们普遍大学生的提升之路，大学生的恋爱不应以物质为基础，高消费的恋爱只会让双方陷入困境。大学毕竟是我们提升自我的空间，精神方面，学习方面都应是充实的。

参考文献

- [1] 张亮亮. 大学生月消费情况调查——以南京邮电大学为例[J]. 统计教育, 2009(2): 61-64.
- [2] 胡娟. 大学生消费心理和消费行为的研究[J]. 心理科学, 2003, 26(2): 297-300.
- [3] 刘莉. 大学生消费行为调查及影响因素研究——基于石家庄高校调查数据[J]. 统计与管理, 2017(1): 24-26.
- [4] Shim, S. and Maggs, J.L. (2009) A Psychographic Analysis of College Students' Alcohol Consumption: Implications for Prevention and Consumer Education. *Family and Consumer Sciences Research Journal*, **33**, 255-273. <https://doi.org/10.1177/1077727X04272365>