

西南财经大学

Southwestern University of Finance and Economics

课程论文

学年学期： 2020-2021 学年第一学期

课程名称： 抽样调查与应用

论文题目： 财大学生外卖偏好调查报告

学 院： 统计学院

年级专业： 2018 级经济统计学

小组成员： 谢瀚洋 41809061 郑岳 41809062

尹肇国 41809049 李铮言 41809059

马雪岩 41809050

目录

一、问卷题目调整.....	3
二、描述性统计与图形分析.....	4
2.1 问卷组成.....	4
2.2 单一问题分析.....	4
2.2.1 最常单人外卖消费区间.....	5
2.2.2 平均外卖消费次数.....	5
2.2.3 各个价格区间最受欢迎的外卖.....	5
2.2.4 各种外卖出现的价格区间.....	6
2.3 交叉问题分析.....	8
2.3.1 最常外卖消费区间与生活费呈正相关.....	8
2.3.2 各个园区平均点外卖次数有差异.....	9
2.3.3 男女外卖偏好差异.....	9
2.3.4 男女购买外卖金额差异.....	10
2.3.5 各个区间外卖购买量差异.....	10
三、聚类分析.....	11
3.1 数据预处理.....	11
3.2 聚类方法选择.....	12
3.3 k-modes 算法介绍.....	12
3.4 利用 k-modes 算法进行数据分析.....	13
3.4.1 各类别基本特征分析.....	13
3.4.2 各类别各价格区间内外卖品类偏好分析.....	14
四、结论.....	21
4.1 偏好描述.....	21
4.2 给商家建议.....	21

一、问卷题目调整

在 2020 年 12 月 15 日我们从大一大二三大四以及研究生这五个群体中各抽取 10 人进行预调查问卷的填写，所发放的 50 份问卷得到的结果显示，第四个问题中回答全部集中在 B、C、D、E 三个选项中，其余选项并未涉及（如下图），因此调整问卷将原问题中 A、F、G、H、I 删去，以减少调查者的思考时间，让调查者有更好的填写问卷体验。

第4题： 您单人外卖消费金额最经常在哪个区段？（实际支付金额） [单选题]

选项 ↓	小计 ↓	比例
5~10	0	0%
11~15	8	24.24%
16~20	10	30.3%
21~25	8	24.24%
26~30	7	21.21%
31~35	0	0%
36~40	0	0%
41~45	0	0%
46及以上	0	0%
本题有效填写人次	33	

[表格](#) [饼状图](#) [圆环图](#) [柱状图](#) [条形图](#) [隐藏零数据](#)

而第五个问题是“您平均每周点多少次外卖？”，这是一个开放性的主观问题，由于没有进行事先的说明，调查者回答格式不统一（如下图），所以考虑在此题上方增加备注“请填写阿拉伯数字”。

以便于最后对调查结果进行数据分析。

序号	提交答卷时间	答案文本	查看答卷
1	1月2日 00:07	1~3	查看答卷
2	1月2日 00:18	3-5	查看答卷
3	1月2日 00:23	1	查看答卷
4	1月2日 00:44	2	查看答卷
5	1月2日 00:47	1	查看答卷
6	1月2日 00:50	1	查看答卷
7	1月2日 00:53	2-8	查看答卷
8	1月2日 00:54	0-2	查看答卷
9	1月2日 00:56	四五次	查看答卷
10	1月2日 01:00	1	查看答卷

其余的问题的调查结果中并没有出现与预期或者现实有明显区别的情况，而且对预调查的调查者进行采访发现除了第四题和第五题，其他问题都比较明确易懂，所以将其余问题保留，仅对第四题和第五题进行修改。

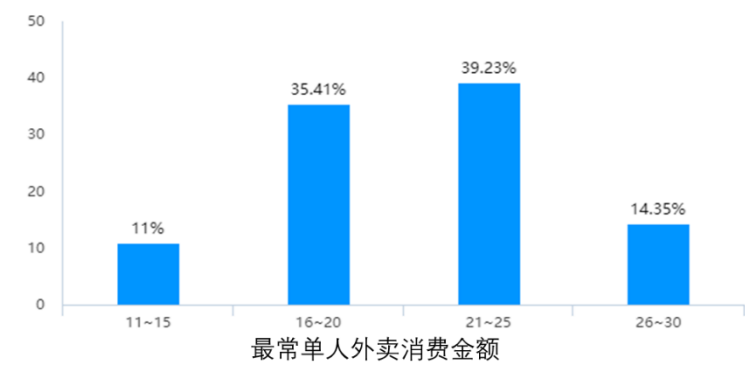
二、描述性统计与图形分析

2.1 问卷组成

正式调查通过问卷星发送网上问卷，目的是调查（西财）大学生外卖消费倾向，总共收到来自调查总体西财学生的 209 份有效问卷，男女参填人数分别为 88 与 121，其中有大一 37 人、大二 54 人、大三 42 人、大四 42 人、研究生 34 人。

2.2 单一问题分析

2.2.1 最常单人外卖消费区间



通过此次问卷，我们了解到近四份之三参与该问卷的西财学生主要单人外卖消费区间为 16-25 元，反之 15 元以下仅占 11%，26 以上的仅占 14.35%，可见西财大学生主要点的外卖是在 20 元左右的，略微高于在食堂的消费。

2.2.2 平均外卖消费次数

通过此次问卷对于各个受调查者的平均消费次数统计，我们计算出参与调查的 209 人平均每周外卖消费次数为 3.8038 接近 4 次，也就是说一周有大概一半的天数(如果不在一天点多次外卖)会点外卖，可见研究外卖相关问题的重要性。

2.2.3 各个价格区间最受欢迎的外卖

	无	奶茶	面条/粉	粥	甜点/蛋糕	便当	炸鸡	烧烤	汉堡披萨	干锅鸡公煲	火锅冒菜	包子饺子抄手	特色小吃	中餐外卖	其他
10以下	83	27	18	42	5	1	1	1	0	0	0	21	9	0	1
11到15	40	62	30	20	5	9	2	0	0	1	1	19	17	2	1
16到20	11	71	30	5	10	35	11	2	6	2	6	10	6	3	1
21到25	5	35	4	3	12	35	55	5	16	9	11	0	5	12	2
26到35	20	1	1	1	6	6	55	14	21	26	14	1	5	31	7
36到45	49	0	0	0	5	2	3	32	29	20	18	1	3	41	6
46以上	74	0	0	0	1	1	1	35	22	17	14	4	0	23	17

通过此次问卷调查，我们得出了各个价格区间中前三经常被购买的外卖，他们分别是：

10 元以下：粥 奶茶 面条/粉

11-15 元：奶茶 面条/粉 粥

16-20 元：奶茶 便当 炸鸡

21-25 元：炸鸡 奶茶 便当（并列第二）

26-35 元：炸鸡 中餐外卖 干锅鸡公煲

36-45 元：中餐外卖 烧烤 披萨汉堡

46 以上：烧烤 中餐外卖 披萨汉堡

结合 2.2.2 最常单人外卖消费区间的分析，我们可以大致推测在 16-20 与 21-25 区间具有统治力的奶茶、炸鸡与便当应当是西财大学生最经常购买的外卖。

2.2.4 各种外卖出现的价格区间

本问卷对于各个区间最常点的外卖统计不仅可以得出各个区间外卖受欢迎程度，还可以得出，各种外卖在同学们心中的可接受价格范围。

	无	奶茶	面条/粉	粥	甜点/蛋糕	便当	炸鸡	烧烤	汉堡披萨	干锅鸡公煲	火锅冒菜	包子饺子抄手	特色小吃	中餐外卖	其他
10以下	83	27	18	42	5	1	1	1	0	0	0	21	9	0	1
11到15	40	62	30	20	5	9	2	0	0	1	1	19	17	2	1
16到20	11	71	30	5	10	35	11	2	6	2	6	10	6	3	1
21到25	5	35	4	3	12	35	55	5	16	9	11	0	5	12	2
26到35	20	1	1	1	6	6	55	14	21	26	14	1	5	31	7
36到45	49	0	0	0	5	2	3	32	29	20	18	1	3	41	6
46以上	74	0	0	0	1	1	1	35	22	17	14	4	0	23	17

（图同 2.2.3）

奶茶：该外卖在 10 元以下到 25 元均有许多购买量，但超过 25 元一般少有人最常选择购买。

面条/粉：10 元以下具有一定的购买量，11-20 元间为最常购买区间，20 元往上购买量急剧减少。

粥：外卖中最为廉价的一种，仅在 15 元以下具有较多的购买，似乎同学们不愿意为喝粥付出过多的钱。

甜点/蛋糕：仅在 16-25 元间有一定购买，且购买量较少。

便当：16-25 元间均有不少的购买，符合 20 元上下吃份快餐的预期。

炸鸡：作为在学生群体中比较火热的外卖品，从 16 元开始就有购买，直到 20-35 元为主要购买区间

烧烤：属于外卖中较为昂贵的种类，从 26 元开始有较多的购买，且直到 46 以上价格越高购买量越多。

披萨汉堡：从 20 元往上都有一定程度的购买量，且随价格分布较平均。

干锅鸡公煲：与披萨汉堡类似，但发力价格比披萨汉堡略高。

火锅冒菜：在 21 元以上有一定量的购买，但不管在哪个区间购买量都有限。

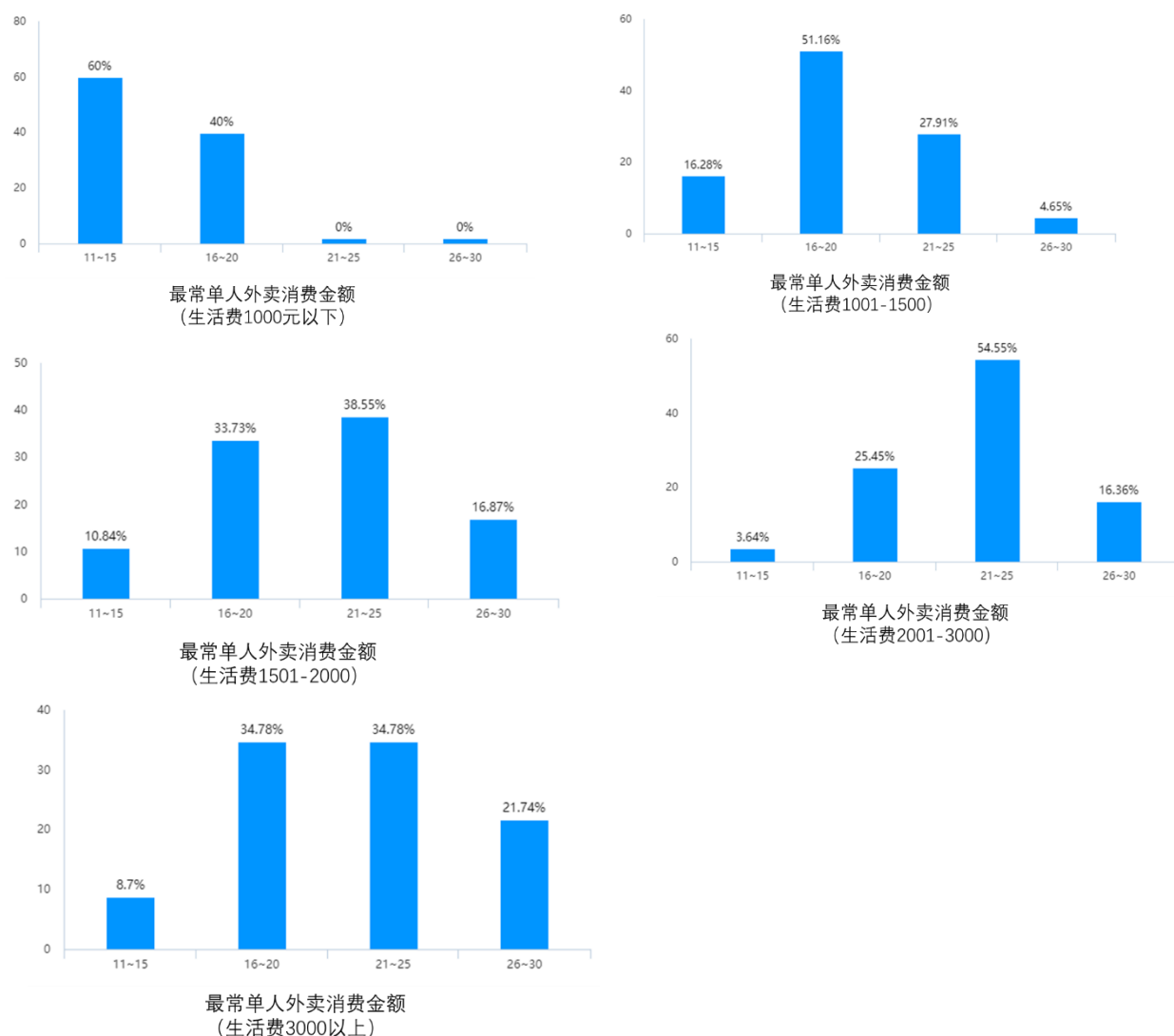
包子饺子抄手：在 10 元下到 20 元均有一定购买，之后断崖式下跌，可见学生基本能接受的价格为 20 以下。

特色小吃：也许是因为特色小吃包含范围较广，其在较多区间均有购买，但也许是其“小吃”的特点，使得该商品在 11-15 元最能让人接受。

中餐外卖：在 21 元往上有较大量的购买，最常购买区间在 26-45 元。

2.3 交叉问题分析

2.3.1 最常外卖消费区间与生活费呈正相关



通过此次对于外卖消费区间调查的问卷，同时通过观察不同生活费区间的最常外卖消费金额我们可以直观的感受最常外卖消费金额这一数值与生活费存在着一定程度的正相关，最常外卖消费区间会随着生活费的上升有一定程度的上升。

生活费 3000 以上的学生反而比之 2001-3000 的 11-20 元外卖消费更多可能是因为一部分 3000 以上生活费的是通过自己赚取的，拥有更为勤俭的价值观，而 21-25 区间较少则可能是因为一部分该区间

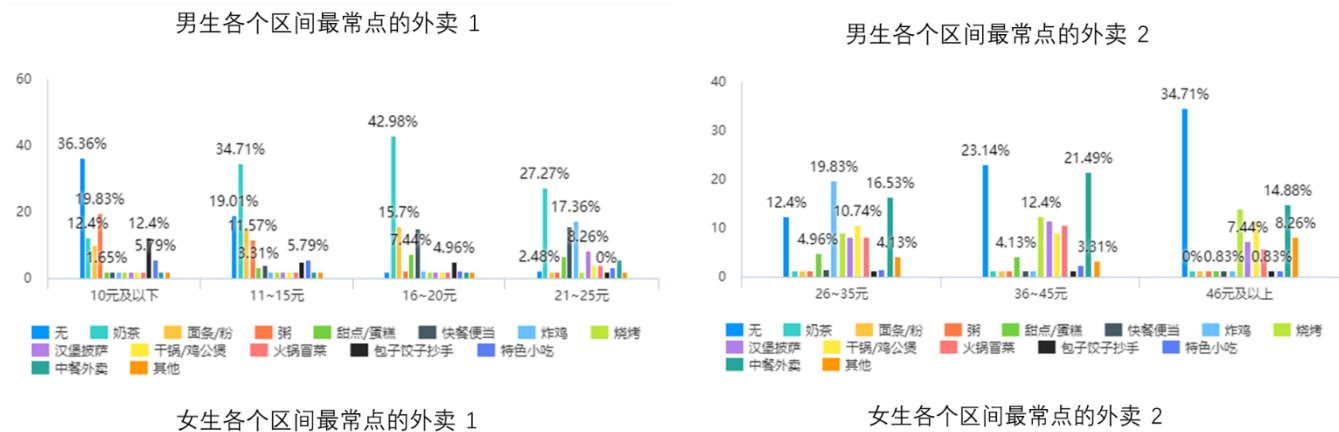
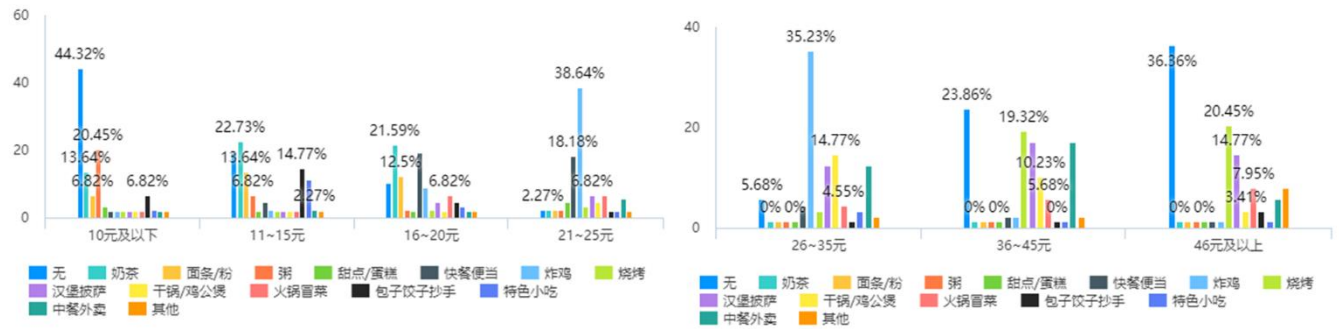
学生进入 26-30 区间。

2.3.2 各个园区平均点外卖次数有差异

通过对每个园区平均点外卖次数调查，我们可以得出人品园学生平均每周点 3.8 次外卖，植物园学生平均每周点 3.6 次外卖，研究生园则是 4.35 次。

对于这些数据我们进行分析，人品园与植物园均为本科生，且都靠近校门，点外卖的次数相差不大，所差的 0.2 可能来自于双边食堂口味好坏；而研究生园所住的为本科生，课业较为繁忙，且对于其他的园区来说距离食堂较远，因此研究生园的学生点外卖次数较多。

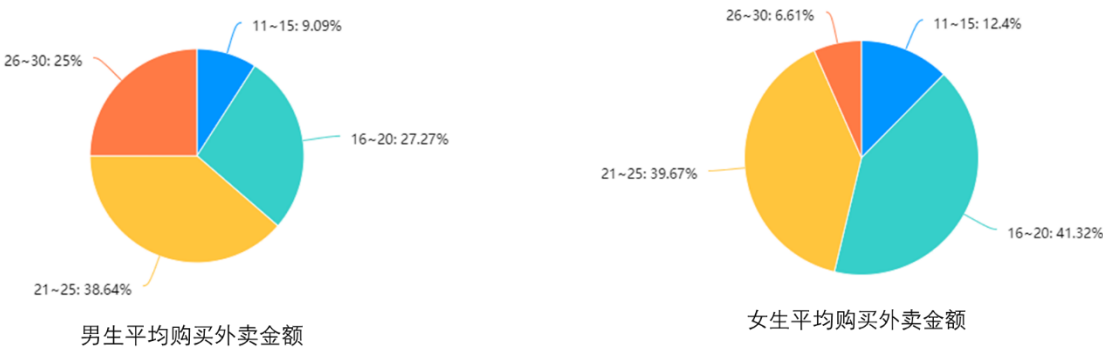
2.3.3 男女外卖偏好差异



通过对于男女各个区间最常点的外卖分性别分析，我们可以直观

的看出男女之间外卖偏好有差异。男生在炸鸡这一外卖的偏好几乎是女生的两倍，而女生在价格较高的奶茶（16-25 元）上的购买比例则远远超过男生。但同时也存在诸如快餐便当这类男女购买差异较小的外卖。

2.3.4 男女购买外卖金额差异



通过对男女生平均购买外卖进而进行分析，我们可以很明显的看出男女生间购买外卖金额有差异，男生的外卖消费明显比女生更多，这可能是因为男生的食量较女生而言更大。

2.3.5 各个区间外卖购买量差异

将收集到的问卷中价格区间外卖品类表变成 0，1 数据（打勾的为 1，没打勾的为 0），再将点外卖频率数据乘以此表中的 0 或 1 作为该调查单位的不同价格区间对外卖品类的打分值，再将所有问卷中该表的数据加和，形成以下表格反映外卖在各个区间上的购买量。

	无	奶茶	面条, 粉	稀饭	甜点蛋糕	快餐便当	炸鸡	烧烤	汉堡披萨	干锅鸡公	火锅冒菜	包子饺子	特色小吃	中餐外卖	其他
10元及以下	270	116	63	170	27	4	8	2	0	0	0	89	45	0	1
11-15元	129	224	123	65	27	34	12	0	0	8	4	82	77	8	2
16-20元	27	249	115	14	46	132	49	12	27	12	24	43	30	13	2
21-25元	8	139	15	6	50	125	196	19	72	42	43	0	21	48	11
26-35元	51	2	2	7	19	31	213	60	80	96	65	3	18	118	30
36-45元	153	0	0	0	15	12	11	134	111	91	84	4	6	160	15
46元及以上	242	0	0	0	2	2	4	140	82	80	59	13	0	100	71

通过观察表格，我们得出各个区间购买量前三的外卖：

10 元以下：粥 奶茶 包子饺子抄手

11-15 元：奶茶 面条/粉 包子饺子抄手

16-20 元：奶茶 便当 面条/粉

21-25 元：炸鸡 奶茶 便当

26-35 元：炸鸡 中餐外卖 干锅鸡公煲

36-45 元：中餐外卖 烧烤 披萨汉堡

46 以上：烧烤 中餐外卖 披萨汉堡

发现与上文得到的最受欢迎外卖表有差异，但差异不大。

三、聚类分析

为了更进一步对西财学生的外卖倾向进行探究并更好地帮助外卖商家进行消费者画像，考虑对问卷调查数据进行聚类分析。

3.1 数据预处理

将问卷星收集到的 209 份问卷的结果以 Excel 表格形式导出，并用 python 对数据进行清理，删除没有用的特征，如序号、提交时间、来源、用户 ip 等。再将分类型特征进行虚拟变量的构造，最后得到

样本量为 209，特征数为 104 的数据集。

3.2 聚类方法选择

由于 k-means 算法是一种简单且实用的聚类算法，所以一开始考虑使用传统的 k-means 算法进行聚类。但是传统的 k-means 算法只适用于连续属性的数据集，而对于离散属性的数据集，计算簇的均值以及点之间的欧式距离就变得不合适了。而 k-modes 算法作为 k-means 算法的一种扩展，适用于离散属性的数据集于是决定使用 k-modes 算法对数据进行聚类

3.3 k-modes 算法介绍

K-modes 是数据挖掘中针对分类属性型数据进行聚类采用的方法，其算法思想比较简单，时间复杂度也比 K-means、K-medoids 低，大致思想如下：

假设有 N 个样本，共有 M 个属性，均为离散的，对于聚类数目标 K ：

step1: 随机确定 k 个聚类中心 C_1, C_2, \dots, C_k , C_i 是长度为 M 的向量, $C_i = [C_{1i}, C_{2i}, \dots, C_{Mi}]$ 。

step2: 对于样本 $x_j (j=1, 2, \dots, N)$ ，分别比较其与 k 个中心之间的距离（这里的距离为不同属性值的个数，假如 $x_1 = [1, 2, 1, 3]$, $C_1 = [1, 2, 3, 4]$ $x_1 = [1, 2, 1, 3]$, $C_1 = [1, 2, 3, 4]$ ，那么 x_1

与 C1 之间的距离为 2)。

step3: 将 x_j 划分到距离最小的簇, 在全部的样本都被划分完毕之后, 重新确定簇中心, 向量 C_i 中的每一个分量都更新为簇 i 中的众数。

step4: 重复步骤二和三, 直到总距离 (各个簇中样本与各自簇中心距离之和) 不再降低, 返回最后的聚类结果。

3.4 利用 k-modes 算法进行数据分析

3.4.1 各类别基本特征分析

考虑到将用户类别不明确, 于是对数据进行多种类数的聚类, 对比发现聚为 4 类时聚类效果最好, 各类别基本特征聚类结果如下表:

表 1 类别基本特征表								
类别	数量	男	女	大一	大二	大三	大四	研究生
1	73	0.30137	0.69863	0.219178	0.684932	0.082192	0.013699	0
2	59	0.694915	0.305085	0.20339	0.050847	0.135593	0.610169	0
3	36	0.416667	0.583333	0.027778	0	0.027778	0	0.944444
4	41	0.243902	0.756098	0.195122	0.02439	0.658537	0.121951	0

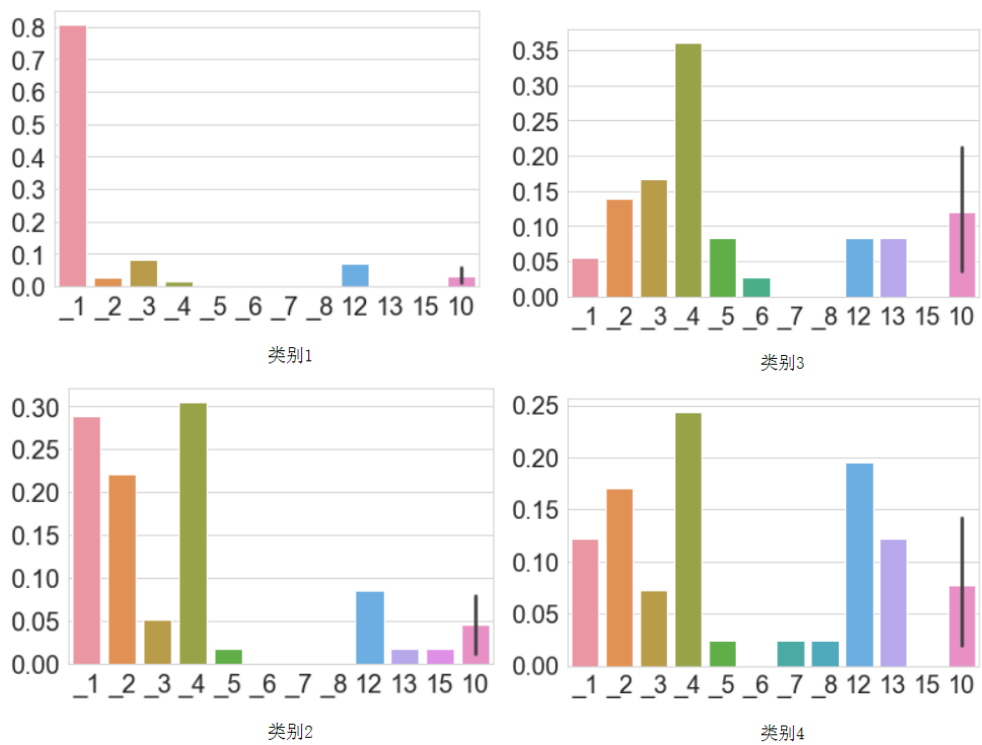
由表格数据可知, 类别 1 大二学生偏多, 女生偏多的特征; 类别 2 中大四学生偏多, 男生偏多; 类别 3 中研究生偏多, 男女比例与填问卷男女比例相近, 说明类别三基本可以说明研究生特征, 也说明研究生和本科生偏好有较大差别; 类别 4 大三学生偏多, 女生偏多。

3.4.2 各类别各价格区间内外卖品类偏好分析

各外卖品类对应编号如下表：

外卖品类编号对应表							
无	奶茶	面条，粉	粥类	甜点蛋糕	快餐便当	炸鸡	烧烤
1	2	3	4	5	6	7	8
汉堡披萨	干锅鸡公煲	火锅冒菜	包子饺子抄手	特色小吃	中餐外卖	其他	
9	10	11	12	13	14	15	

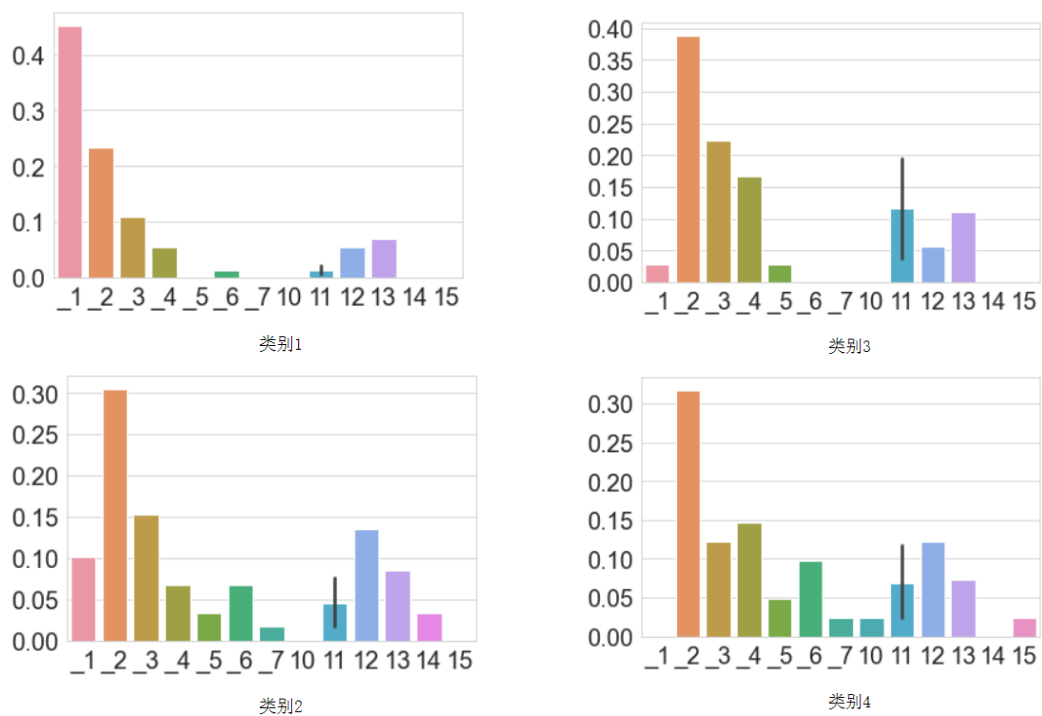
各类别在 10 元以下价格区间的外卖选择偏好图如下：



由图可知，类别 1 基本不购买 10 元以下的外卖，而类别 2, 3, 4 对 10 元以下的外卖接受度不低，其中粥类和奶茶接受度相对较高，其他外卖品类在这个价格区间内不占优，原因很可能是其他外卖品类

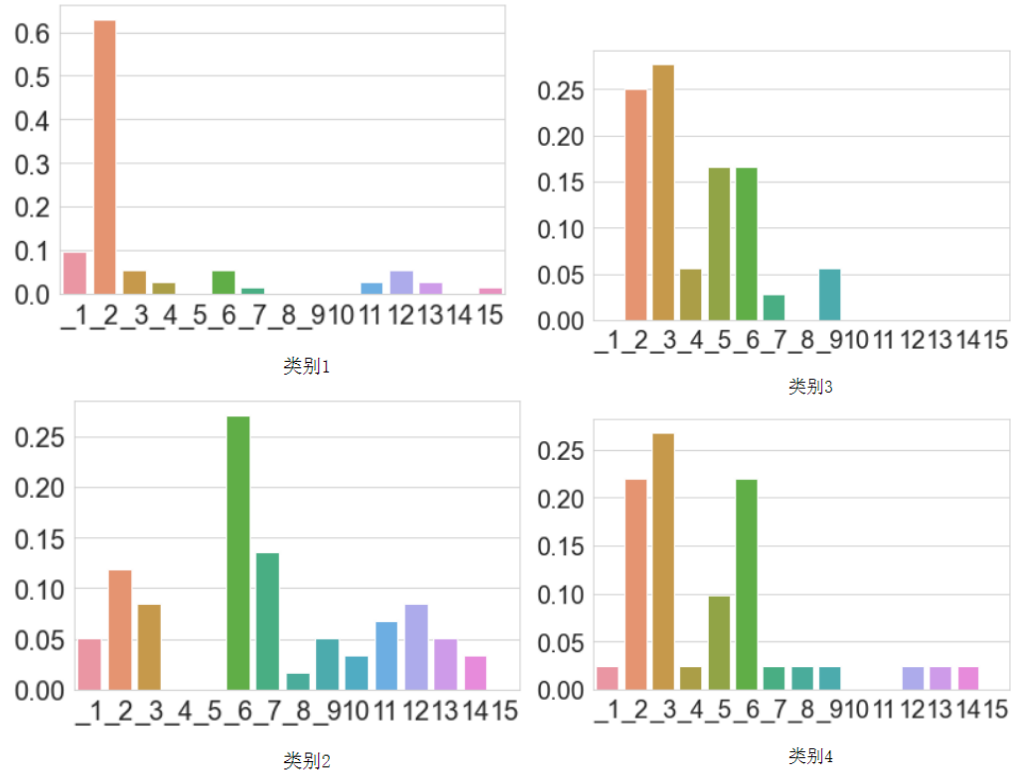
在保证质量时的成本超出了 10 元，若是降到这个区间，只能是便宜没好货，没好货就不能长期吸引消费者。所以除了奶茶和粥类，其他外卖品类的价格区间不适于定于 10 元以下区间。

各类别在 11-15 元价格区间的外卖选择偏好图如下：



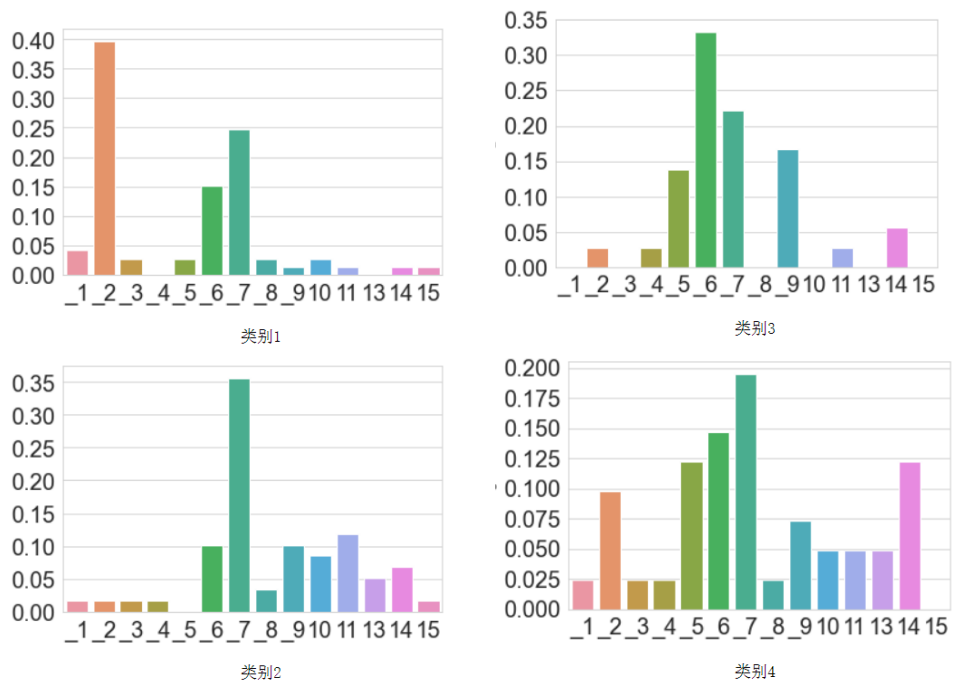
由图可知 11-15 元价格区间的外卖同样对类别 1 没有较大的吸引力，综合四个类别发现，这个价格区间内奶茶的需求都是最高的，其次是粉面和粥类。说明在这个价格区间内最能打的还是奶茶，其他外卖品类除了粉面与粥类，都不建议定在这个价格区间。

各类别在 16-20 元价格区间的外卖选择偏好图如下：



由图可知在 16-20 元价格区间内类别 1 仍然对奶茶有着强烈的偏好，而其他三类中，奶茶，粉面，快餐便当所受欢迎的程度也很高。

各类别在 21-25 元价格区间的外卖选择偏好图如下：



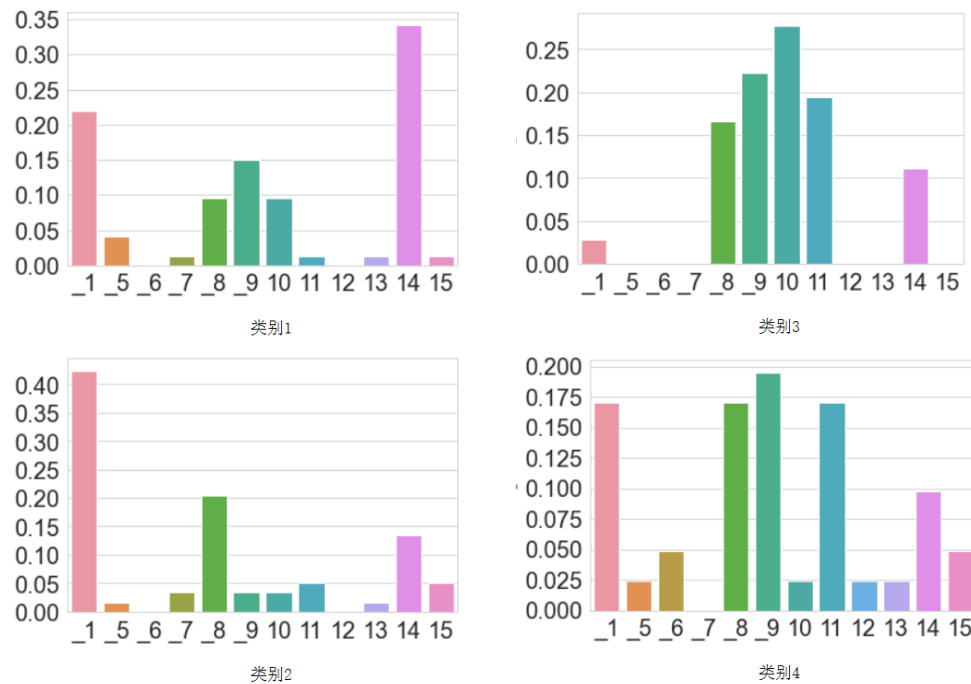
由图可知在 21-25 元价格区间内快餐便当和炸鸡受欢迎程度相对较高，而类别 1 仍然是最偏爱奶茶，而类别 1 女生偏多且大二学生偏多，可以进行推断：大二学生，以及女生对奶茶有着很强烈的偏好。

各类别在 26-35 元价格区间的外卖选择偏好图如下：



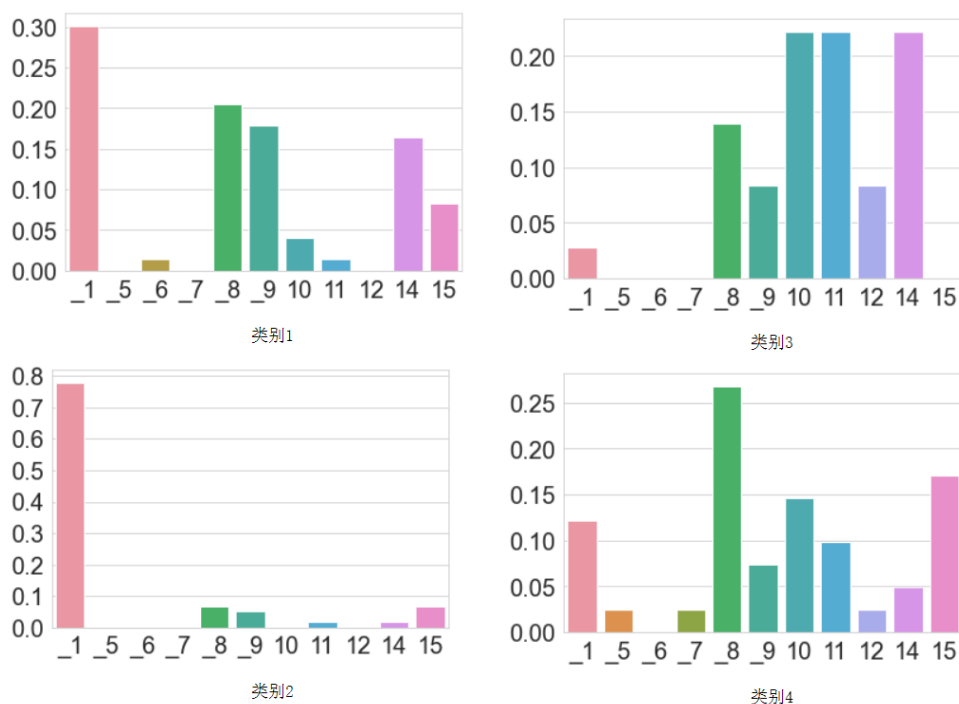
由图可知在 26-35 元价格区间内，炸鸡，干锅鸡公煲非常受欢迎，类别 1，2，34 都是最爱炸鸡，而类别 3 有很多其他的偏好。

各类别在 36-45 元价格区间的外卖选择偏好图如下：



由图可知在 36-45 元价格区间四个类别的图像都呈现中心高，两端高，中间低的性质，说明在这个区间内大家更偏向于不点外卖和点其他外卖，而在这个区间类别 3 和 4 点外卖的偏好相对较高，而且占比较多的种类为烧烤、汉堡披萨、干锅鸡公煲、火锅冒菜。

各类别在 46 元以上价格区间的外卖选择偏好图如下：



由图可得在 46 元以上价格区间内，类别 2 基本不点外卖，类别 1 也更偏向不点外卖，而类别 3 和 4 还是更偏向于点外卖，偏好的外卖品类包括烧烤、汉堡披萨、干锅鸡公煲、火锅冒菜，与在 36-45 元价格区间内的偏好一致。

综上所述，类别 1 对奶茶的偏好程度很高，在 11-25 元这个价格区间内都会优先选择奶茶作为外卖消费，而且不点的概率也相对较高，说明该群体特征明显，奶茶外卖商家可以重点关注这一群体，女生占比较多，大二的同学占比较多，可以进行对特定群体的宣传。

类别 2 对炸鸡，快餐便当的偏好程度较高，不点的概率仅次于类别 1，群体主要占比为大四学生和男生，炸鸡和快餐便当外卖商家可以注重对这些消费者的口味探寻。

类别 3 和 4 消费偏好较为类似，不介意高消费的外卖，对烧烤、

汉堡披萨、干锅鸡公煲、火锅冒菜这些种类较为偏好，而且也常点其他种类的外卖，新的商家可以重点关注这两个群体，主要包括研究生和大三学生。

四、结论

4.1 偏好描述

经过此次调查问卷，我们得出西财学生最能接受的外卖大多在 20 元上下，其中炸鸡、奶茶还有快餐便当最受同学们的欢迎，饺子包子和面条这类面粉类食物在低价位也十分具有统治力，同时男同学较女同学而言点的外卖价格相对较高。就外卖频率而言则是不同园区具有一定的差异，这取决于它们地理位置的差异。

4.2 给商家建议

由对类别 1 的分析可以给出建议：奶茶外卖商家可以多关注女生以及大二学生，价格区间可以在 11-25 元这个区间浮动，不要为了低价舍弃质量，也不要做的太过昂贵。

由对类别 2 的分析可以给出建议：炸鸡和快餐便当外卖商家可以注重对大四学生以及男生的口味探寻，可以进行定群体宣传。

对类别 3、4 的分析可以给出建议：烧烤、汉堡披萨、干锅鸡公煲、火锅冒菜的外卖商家可以多考虑大三学生和研究生群体，而且不需要太过在意价格，只要质量好就行。而其他的外卖品种商家也可以多考虑这两个群体，因为他们更愿意接受新形式的外卖。