“拍照赚钱”的任务定价

摘　要

随着移动互联网的快速发展，市场上产生了一种众包服务模式，如“拍照赚钱”App 平台。任务定价是平台运行的核心要素。本文研究“拍照赚钱”的任务定价规律。全文分为四部分内容。

对于问题 1，本文研究任务的定价规律且分析任务未完成的原因。首先通过单位分格法和任务中心圈法找到密度指标，即会员密度和任务密度。其次利用 MATLAB 软件编程求出每个任务所在单元格对应这两个指标的数据。再次建立多元线性回归模型，利用 SPSS 软件建模 $p_i = 73.416 - 0.05 * x_{1i} - 0.17 * x_{2i}$。最后定性分析得出：定价不合理导致了任务完成率不高。

对于问题 2，本文给附件 1 中的任务制定新的定价方案。首先基于问题 1 中的定价模型，引入了新变量——任务与其最近会员的距离，作多元回归分析，得到新的定价模型。其次利用任务平均价格的变化和任务完成率的变化构建评价指标体系，比较新旧模型，得出新模型更优。

对于问题 3，研究打包定价模型。根据广州、深圳地区的人均日工资水平及会员的平均日工作能力、任务的平均定价（60 元~70 元），每包内平均任务数的取值小于 4 为宜。利用改进的单位分格法，得到打包定价模型，在任务平均价格基本保持不变的情况下，任务完成率得到大幅度提升，达到 84.23% 。

对于问题 4，本文采用附件三中新项目的任务打包定价方案，利用单位分格法，调整了相关参数，得到新的模型，实施效果令人满意。

本文针对“拍照赚钱”App 平台定价问题研究得到相应价格取值规律，研究结果对自助服务的科学指导有一定的理论意义。

关键词　众包服务 定价 单位分格法 任务中心圈法 多元线性回归

一、问题重述

近几年，移动互联网的迅速发展，促使市场上产生了自助式服务众包平[1]台。这种平台可以为企业提供大量的、有效的、真实的商业信息及数据，并且节省大量时间与成本。而对于该平台来说，APP 是其运行的核心，任务定价是其关键问题。所以给任务进行合理定价，是很重要的；否则，部分任务会无人问津，致使商品检查失败。

接下来，本文就根据题目所给的附件数据，分析影响任务定价的因素，以及影响任务完成情况的因素。

问题 1，要求研究附件一中的任务定价规律，并分析其中任务未完成的原因。考虑到影响任务定价的因素有很多，其中起决定性作用的因素是供需关系。本文把平台看成需求方，会员看成是供给方，然后仅从供给与需求方面，来寻找任务定价规律；再根据定价规律，来寻找影响任务未完成的因素。

问题 2，要求给附件中的任务制定一个新的定价方案，得到各个任务的新标价，以及总体的新任务完成率，再并与题目所给的任务标价及总体的任务完成率进行比较，分析哪个方案的效果更好。

问题 3，在实际生活中，平台所发布的任务，大多数位置可能比较集中，这容易引起会员争相选择。因此，出现了一种将位置比较集中的任务进行打包的发布任务方式。在这种情况下，任务定价方案应该如何改变；改变之后，又会如何影响任务的完成情况。

问题 4，对于给出的新项目，应该如何确定任务定价方案，并对该方案进行实施效果分析。

二、问题分析

劳务众包是一个新兴产业，与传统的市场调查方式相比，它的调查成本较低，并且数据真实性较高。但是定价不合理，就会造成任务难以完成。所以定价对劳务众包平台来说是一个关键因素。下面将对任务定价进行分析。

对于问题 1，在实际生活中，影响的商品定价的因素有很多，其中市场上的供需关系对定价起着重要性作用。经分析，本文“拍照赚钱”众包公司平台定价规律和快递业定价类似，都是以成本为导向的定价原则。因此，本文从供需方面对任务定价进行分析。其中，将众包平台作为需求方，会员作为服务供给方。因此，为了寻找定价规律，需要将任务的定价与平台发出的任务、做任务的会员这供求两方联系起来。只有寻找到了衡量供求双方的合适数据，才能较好的利用数据以及多元线性回归的模型寻求任务定价的规律。

若要分析未完成任务的原因，应分别从考虑地域和供求关系角度出发，进行定性、定量两方面进行全面分析。

对于问题 2，题目要求设计新的定价方案，换言之，就是将平台原有的定价方案进行改进。考虑到问题 1 得到的平台原有定价规律为多元线性回归模型，为了改进此模型，本文考虑引入新的自变量，对其进行更准确的描述与拟合，达到改进原定价模型，创建新定价模型的目的。

对于问题 3，以第一问的单位格法为基础，将方格内的任务进行打包，接着建立包的价格模型，由此求出包的价格，再代入任务完成率模型中，求出完成率，最后与未打包的完成率进行比较和分析。

对于问题 4，首先作出对附件 3 任务点的散点图，根据问题 3 任务密度来选择打包的策略，即定价模型。接着沿用问题 2 中的任务中心圈法，来求每个新任务所对应的单位圆内的会员密度、任务密度及最近会员距离。采用不打包和打包两种策略的定价模型，分别求出不同新任务的价格和完成率，进行比较和分析。

三、模型假设

1、假设任务及会员的经纬度可投影为平面坐标。
2、假设平台需在当天内将所有任务发放完毕。
3、假设会员接任务后必须当天完成该任务。
4、假设会员接包后必须当天完成该任务包内所有任务。
5、假设未完成的任务均是未被会员选择的任务。
6、为保障服务质量，假设信誉值低于 5 的会员不分配打包任务。

四、符号说明

符号	含义	单位
$P_i^{'}$	第 i 个任务包的价格	元
p_i	第 i 个任务的价格	元
F_m	所给任务的总完成期望	%
u_j	任务包总量中的第 j 个任务包的完成率	%
N	任务包总量	个
f_i	第 i 个任务的完成概率	—
x_{1i}	第 i 个任务的单位圆内的会员密度	—
x_{2i}	第 i 个任务的单位圆内的任务密度	—
x_{3i}	第 i 个任务的单位圆内的距离	公里

五、模型建立与求解

5.1 问题 1

定价是企业运营的重要问题之一。据经济学知识，一般企业定价法有成本导向定价法和顾客导向定价法。前者是基于单位产品的成本，加上预期利润，来给产品进行定价，是从供给方进行考虑；后者一般是根据某产品以往的价格和市场上需求的变化情况，在一定幅度内，对价格进行调整，这是从需求方进行考虑。

本文先作出了任务和会员经纬分布图 5.1.1，并根据价格高低，利用颜色浅深对任务数据点进行了标记，颜色越浅价格越高。

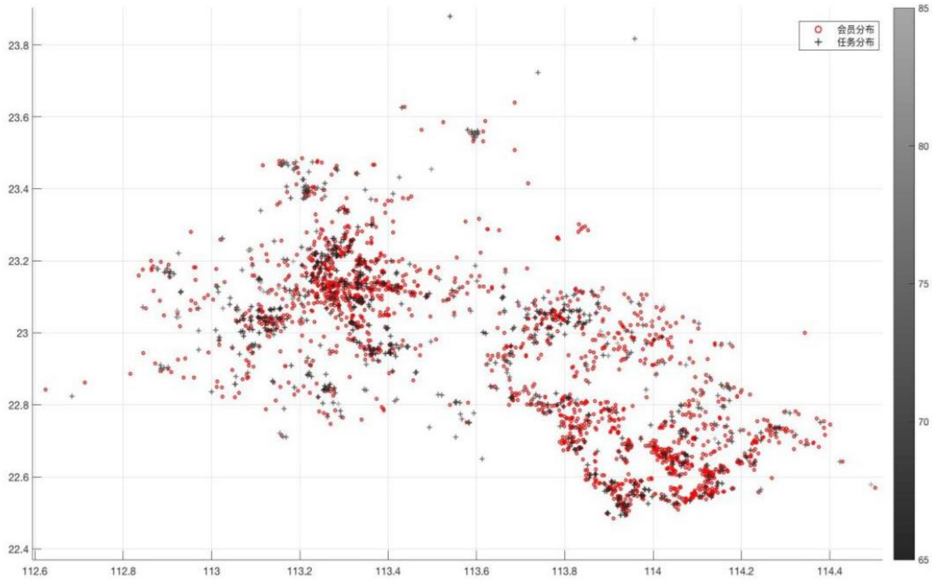

图 5.1.1 任务及会员分布图

由图，可得数据点密集程度越高的区域，任务的价格越低（颜色越深），所以会员或任务密集程度必为定价的相关因素之一。

由分析结合经济学知识，本题众包 “拍照赚钱”外包公司与快递公司业务性质相近，目前市场上快递公司都是用的成本导向定价，所以本题中定价模型也为成本导向定价。将发布任务的平台看作需求方，将提供服务的会员看作供给方，可初步推测影响任务价格的因素大致分为以下几个方面：

1、供给量，即提供众包服务的会员的人数及其配额情况。

2、需求量，即平台所发布的任务数量。

3、供需比，即供给量与需求量之比。

4、距离因素，即某任务与离它最近会员的距离。

5、任务所在地区的人均工资水平。

本文将以任务价格为因变量，利用单元分格法和任务中心圈法这两种不同的分析方式来探索以上几个因素对价格的影响，找出定价规律。

5.1.1 数据的预处理

（1）异常数据的修正

在整理数据时注意到，编号为 B1175 会员的经纬度数据颠倒，其经度为 113.131483 度、纬度为 23.031824 度，因此将其调换，进行了修正。

表 5.1.1　异常数据处理表

原始异常数据					
会员编号	会员 GPS 纬度	会员 GPS 经度	限额	预订任务开始时间	信誉值
B1175	113.131483	23.031824	1	6:36	19.9231
将数据更正为					
会员编号	会员 GPS 纬度	会员 GPS 经度	限额	预订任务开始时间	信誉值
B1175	23.031824	113.131483	1	6:36	19.9231

（2）偏远数据的剔除

从整体会员的分布来看，绝大多数会员位于纬度属于(22.1,23.8)区间，经度属于(112.5,114.6)的区间内，即 1877 个会员中仅有 10 个位于该范围之外。为了方便分析，所以本文将这十个会员数据剔除。被剔除会员的基本信息如下。

表 5.1.2　偏远剔除数据表

会员编号	会员 GPS 纬度	会员 GPS 经度	预订任务限额	预订任务开始时间	信誉值
B0005	33.65205	116.97047	66	6:30:00	20919.0667
B0007	29.560903	106.239083	15	6:30:00	15729.3601
B0022	27.124487	111.017906	1	6:30:00	6494.3768
B0039	21.679227	110.922443	88	6:30:00	1691.0543
B0048	20.335061	110.178827	30	6:30:00	1365.3053
B0082	21.202247	110.417157	29	6:39:00	527.6239
B0136	24.80413	113.605786	23	6:36:00	229.1596
B0472	21.498823	111.106315	1	6:30:00	39.8463
B1708	24.293405	116.122526	1	7:24:00	0.3707
B1727	21.53332	111.229119	5	6:33:00	0.3678

（3）剔除数据的比例

所剔除会员数据点数量为 10，而会员数据点总数即会员人数为 1877，因此剔除数据比为

$$p_t = \frac{n_t}{n} = \frac{10}{1877} = 0.532\% \qquad (5.1.1)$$

其中，n_t 为剔除数据个数，n 为总数据个数，p_t 为剔除数据所占比重。由(5.1.1)可得，剔除数据仅占总数的 0.532%，对总体无较大影响。

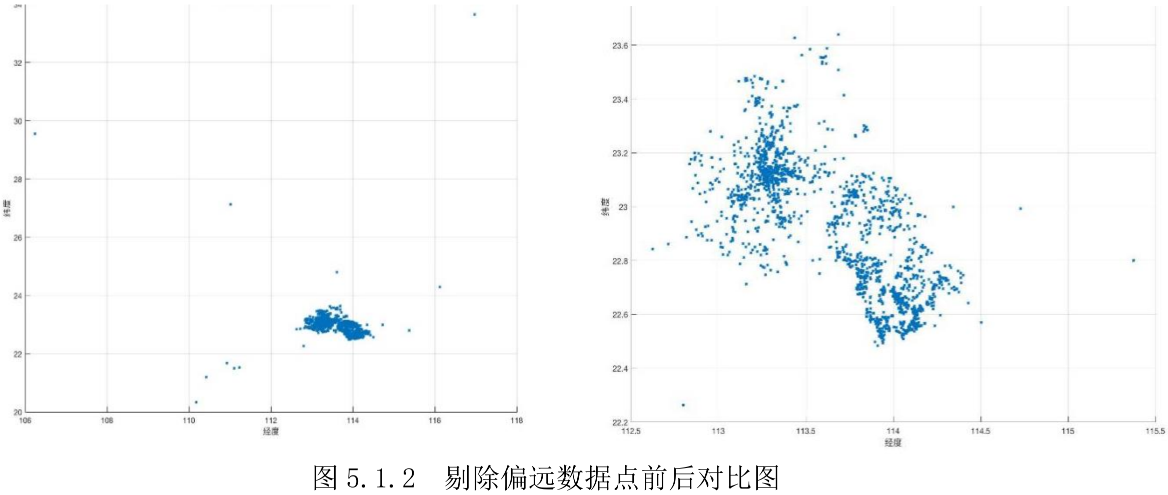

图 5.1.2　剔除偏远数据点前后对比图

如图 5.1.2 所示,剔除偏远位置会员的坐标数据后,会员分布情况较为清晰，且有利于进行下一步分析。

而对于附件 1 中已完成的 835 个任务坐标分布数据，因其比较集中且无特偏远情况，在选定范围内，且无异常数据，所以不需要做剔除、修正处理。

5.1.2　单元分格法

经过查询百度地图，题中所给出的经纬度范围大致为广东省广州、深圳及佛山一带地区，其位于北回归线附近且面积不大 如图所示。因此，本文将任务及会员的在地球表面经纬度的分布近似看成平面中点的分布，其经纬度则为该点对应的平面坐标。接下来，以 0.05 纬度、0.05 经度（距离差）为步长，约为 7km，结合经度范围区间(112.5,114.6)及纬度范围区间(22.1,23.8)，将所在区域地图划分为了成 34×42=1428 个单元方块格，如图所示。这样，就将会员和任务的离散点划分到了各个单元格里，进而可计算出每个单元格内的会员数量、任务数量以及它们的比值,模型如下。

$$\begin{cases} m=(114.6-112.5)/0.05=42; \\ n=(23.8-22.1)/0.05=34; \\ J_s=112.5+s*0.05, s=1,2,3,\ldots,42; \\ W_t=22.1+t*0.05, t=1,2,3,\ldots,34. \end{cases} \quad (5.$$

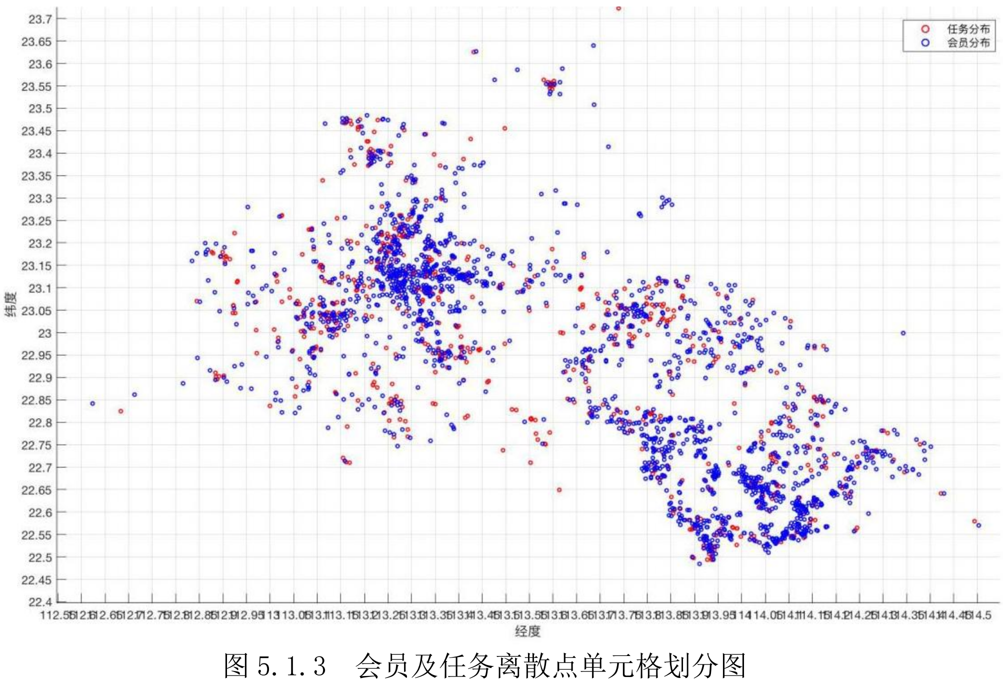

图 5.1.3　会员及任务离散点单元格划分图

由于每一个任务 m(1<m<835)都必定会落在某个特定的单元格,因此有且只有一个单元方块格坐标(f,e)与之对应。且对于每一个单元格，都可以记录有多少个任务、会员落入了其中，即每个单元格内的任务、会员数。利用 MATLAB 软件循环语句进行编程，可实现上述步骤。代码 density1.m 见附录。

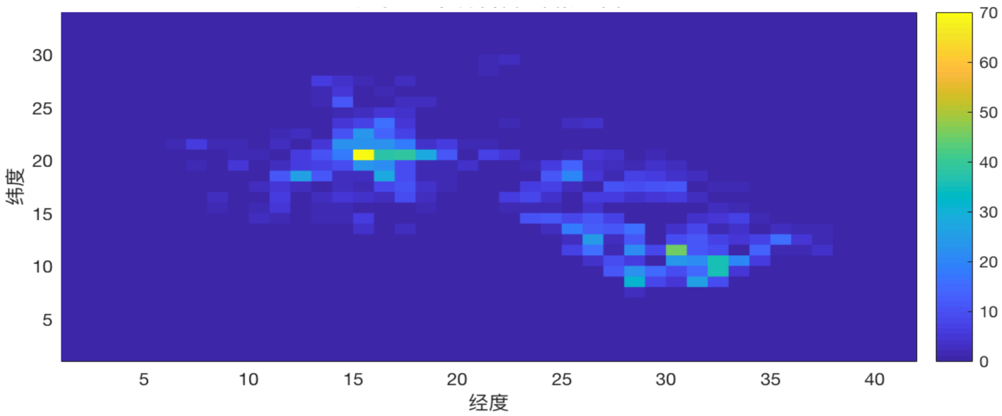

图 5.1.4　任务密度色度标记图

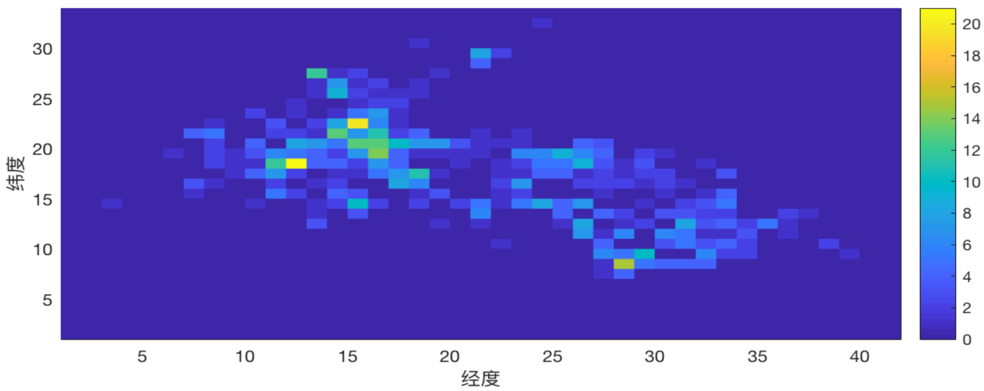

图 5.1.5　会员密度色度标记图

图 5.1.4 和图 5.1.4 中，右侧色度条颜色深浅表示每个单元格内的任务或会员密度大小，颜色越浅密度越高。上图为广东地区任务密度色度标记图，下图为广东地区会员密度色度标记图。从中可以看出，任务及会员的均主要分在两大密集区域。其中左上角密集区域为广州市，其密度程度呈圆形放射状向四周递减。右下角密集区域为深圳市，其密集程度呈扇形放射状由南至北递减。和经济学上“商圈”的概念类似。结合实际的地理和经济发展情况，发现上述单元分格法比较切合现实。

为了深入研究“拍照赚钱”众包劳务公司的定价规律，本文考虑采用任务中心圈法进一步分析。

5.1.3　任务中心圈法

任务中心圈法的思想是，以每个任务为圆心，以 5km 位半径作单位圆（5km 为市内借助公交系统的经济出行半径），计算落在此单位圈内的会员数和任务数，即每个任务圈内的会员密集程度和任务密集程度。此外，还可计算出完成每个任务的最小距离，即此任务与距离它最近会员的距离。

如图 5.1.6 所示，图中红点表示会员，黑点表示任务，任务中心圈是以任务为圆心，以 5km 为半径作的单位圆，故共有 835 个任务圈。

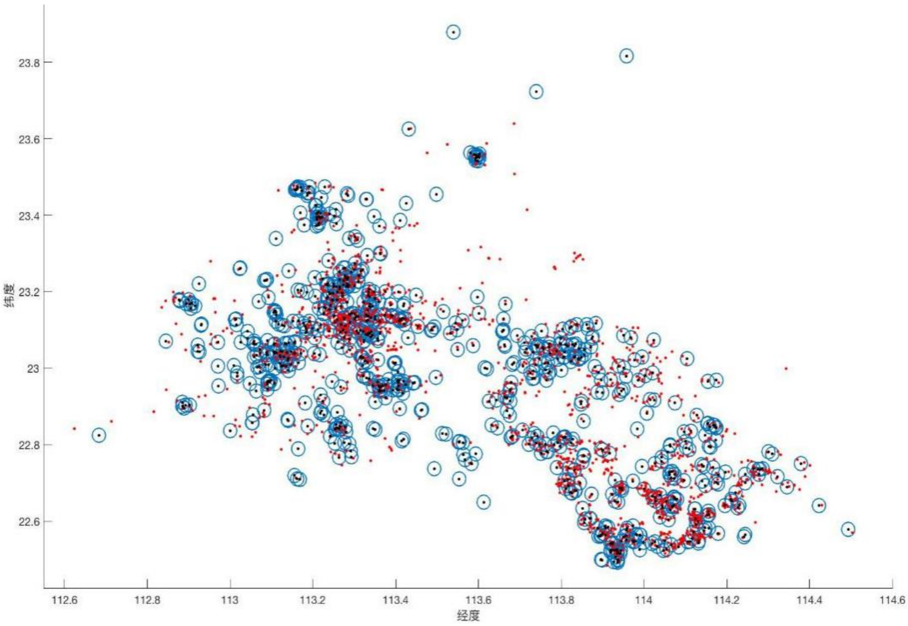

图 5.1.6 任务中心圈示意图

使用 MATLAB 软件进行编程，对每一个任务，用 for 循环语句，运用两点 A(x1,y1) B(x2,y2)距离公式：

$$|AB| = \sqrt{(x1 - x2)^2 + (y1 - y2)^2} \tag{5.1.3}$$

可求得每一个会员与该任务之间的距离，即对每个任务得到一个 1876x1 的向量。又因有 835 个这样的任务，所以最终可得 1867x835 的矩阵。对矩阵每列求最小值，即可得代表每个任务对应最小距离的向量。

表 5.1.3 符号说明表

符号	意义	大小	符号	意义	大小
PPLNUM	每个任务圈内会员数量	835×1	TASNUM	每个任务圈内任务数量	835×1

对于求出第 m 个任务中心圈内的会员数和任务数，则可将 for 循环和 if 判断语句相结合。会员数为 PPLNUM 向量，任务数为 TASNUM 向量，皆为 835x1 大小。若满足两点距离小于 5km 则对应向量第 m 行的元素加一。最后输出的两个 PPLNUM 与 TASNUM 向量，即得到会员密度和任务密度的数据。

根据上述所得到的数据，再结合附件 1 中的共 522 个已完成任务的标价数据，建立多元线性回归模型[2]，模型如下：

$$\begin{cases} PPLNUM(i) = x_{1i}; \\ TASNUM(i) = x_{2i}; \\ p_i = a_0 + a_1 * x_{1i} + a_2 * x_{2i} \end{cases} \tag{5.1.4}$$

其中，a_0、a_1、a_2 为参量，p_i 表示第 i 个任务的价格，x_{1i} 表示第 i 个任务对应的

任务中心圈的会员密度，x_{2i}表示以第i个任务对应的任务中心圈的任务密度。由于任务 i 所在区域为单位圆，所以可用其圈内任务或会员数量代表其密度。

然后利用 SPSS 软件[3]进行求解，结果如表 5.1.4

表 5.1.4　相关系数矩阵

指标		任务的价格	会员密度	任务密度
Pearson 相关系数	任务的价格	1	-0.493	-0.521
	会员密度	-0.493	1	0.732
	任务密度	-0.521	0.732	1
Sig. （单尾）	任务的价格	.	0	0
	会员密度	0	.	0
	任务密度	0	0	.

根据 Pearson 相关性分析结果，如表 5.1.4 所示，可以看出任务的价格与会员密度、任务密度的显著性水平均为 0，小于显著性水平 0.05，说明任务的价格与这两个变量有显著性相关性。会员密度的相关系数为-0.493，表明任务的价格与其具有显著地负相关性，即任务所对应单位圆内的会员密度越高，则任务的价格越低。

任务密度的相关系数为-0.521，说明任务的价格与任务密度具有显著性地负相关关系，即任务所对应单位圆内的任务密度越高，则任务的价格越低。

接下来，以任务的价格为因变量，以会员密度和任务密度为自变量，建立多元线性回归模型,利用 SPSS 软件进行求解,输出结果如表 5.1.5 和表 5.1.6 所示。

表 5.1.5　ANOVA[a]

	平方和	自由度	均方	F	Sig.
回归(R)	3606.583	2	1803.291	110.269	0
残差	8487.49	519	16.354		
合计	12094.073	521			

对定价模型进行检验，检验结果如表所示，F 值为 110.269，对应的显著性水平是 0，小于置性水平 0.05，则拒绝原假设，说明该定价模型通过了检验。

表 5.1.6　定价模型系数表

定价模型	非标准化系数		标准系数	Sig.
	B	标准错误	贝塔	
常量	73.416	0.309	—	0
会员密度	-0.05	0.011	-0.242	0
任务密度	-0.17	0.027	-0.343	0

根据表 5.1.6，可写出定价模型如下：

$$p_i = 73.416 - 0.05 * x_{1i} - 0.17 * x_{2i} \tag{5.1.5}$$

从表中可以看出，会员密度和任务密度的系数为-0.05 和-0.17，对应的 Sig. 均为 0，小于显著性水平 0.05，则说明这两个变量的系数都通过了检验。任务的价格与会员密度和任务密度具有负相关关系，这说明在任务对应的单位圆内，会员或任务分布越密集，该任务的价格越低。

5.1.4　模型检验

俗话说："有钱能使鬼推磨"，价格的高低是影响任务完成与否的决定性因素，即在某种程度上说，价格可决定任务的完成情况。所以，原定价方案下，没有被完成的 313 个任务，大多数是定价偏低造成的。根据新定价方案，原方案中未完成的任务的标价应该是要高于原标价，才能提高任务完成率。但与此同时也应该考虑控制成本，即提价也应该有个上限。将上述 313 个未完成任务的数据，代入模型(5.2.5)求出每个任务的新标价，与原标价作差，得到大于 0 的任务数共计 176，占比达到 56.23%，因此说明该模型可以对任务进行较为合理的定价。

5.1.5　任务未完成的原因

本文假设未完成的任务均是未被会员选择的任务。本文将利用控制单一变量法，从任务地点和任务价格两个方面对任务完成情况进行分析。

从任务地点方面进行分析，利用 MATLAB 软件对附件 1 中任务坐标数据及附件 2 中会员坐标数据绘制联合分布散点图，并用黑、橙两色区分未完成和已完成任务。未完成任务主要集中于图中所标记的三个区域：A、B 和 C。不难发现，会员密度较高的区域中，任务完成率往往较低；而在会员密度较小的地区，任务的完成率反而更高。其原因有二：一，会员密集区域任务定价较低，不足以吸引会员接单。二，对比地图，A 为佛山，B 为广州市，C 为深圳市，市民平均工资水平高，会员可能对定价为六、七十块的拍照任务兴趣不大。

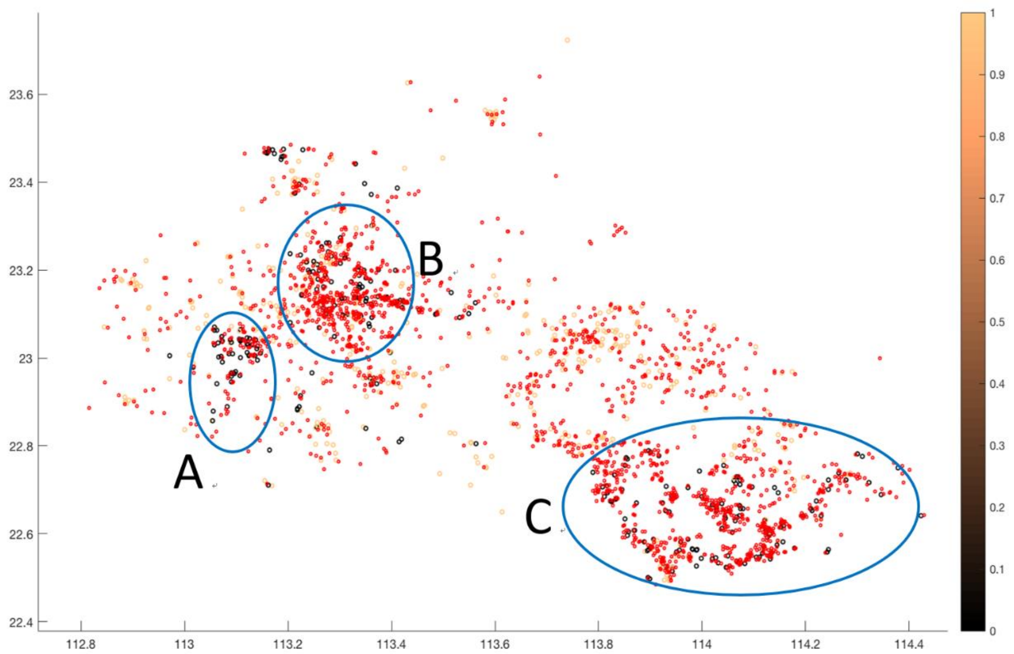

图 5.1.7　任务未完成情况分布图

另一方面，从任务价格方面进行分析。本文任务价格 p 划分为以下三个组别：60~69.5、70~79.5、80 以上（包括 80），从而得到各价格段的任务完成情况，利用 Excel 对数据汇总，结果如表 5.1.7 所示。不难发现，价格越高，任务完成率越高，任务未完成率越低；从而可以初步推断任务完成率与任务价格成正相关关系，即任务未完成率与任务价格成负相关关系。

表 5.1.7　任务按价格分组的完成情况表

价格分组	任务数	完成数	完成率%
60~69.5	510	275	53.92
70~79.5	285	213	74.74
80 以上	40	33	82.50

为了进一步确定任务完成率与任务的价格之间的关系，本文根据附件中的各任务的价格，将数据按表5.1.8方式进行汇总。从该表中可以看出，任务完成率与任务的价格的关系并非是一种简单的线性相关关系。因此，可以通过曲线拟合来得到两者之间的关系。

表 5.1.8　按任务的价格分组的任务完成情况表

任务的价格	任务数	完成数	完成率	ln 完成率	任务的价格	任务数	完成数	完成率	ln 完成率
65.0	65	35	53.85	3.99	71.0	4	3	75.00	4.32
65.5	150	75	50	3.91	71.5	5	4	80.00	4.38
66.0	103	46	44.66	3.80	72.0	60	42	70.00	4.25
66.5	63	35	55.56	4.02	72.5	9	7	77.78	4.35
67.0	38	18	47.37	3.86	73.0	10	7	70.00	4.25
67.5	23	17	73.91	4.30	73.5	5	3	60.00	4.09
68.0	30	25	83.33	4.42	74.0	5	1	20.00	3.00
68.5	11	6	54.55	4.00	74.5	2	1	50.00	3.91
69.0	19	12	63.16	4.15	75.0	78	59	75.64	4.33
69.5	8	6	75.00	4.32	80.0	13	9	69.23	4.24
70.0	96	77	80.21	4.38	85.0	27	24	88.89	4.49
70.5	11	9	81.82	4.40	–	–	–	–	–

接下来，本文利用SPSS中的曲线估计进行多次拟合，发现S曲线的拟合效果是当中最好的，所以可以建立价格关于任务完成度的模型。

$$f_i = \exp(a - b / p_i) \tag{5.1.6}$$

其中，f_i表示第i个任务对应的完成率，p_i表示第i个任务的价格，a、b为参数。

根据模型检验发现，F值为7.035，且对应的显著性为0.015，小于显著性水平0.05.，则拒绝原假设，说明该模型具有统计意义。

表 5.1.9 ANOVA

	平方和	自由度	均方	F	显著性
回归(R)	0.232	1	0.232	7.035	0.015
残差	0.66	20	0.033		

表 5.1.10 模型系数表

	非标准化系数		标准系数	t	显著性
	B	标准错误	贝塔		
1 / 任务的价格	-118.112	44.53	-0.51	-2.652	0.015
（常量）	5.861	0.631		9.284	0

根据模型系数表，可以写出模型为：

$$f_i = \exp(5.861 - 118.112 / p_i) \tag{5.1.7}$$

在实际生活中，不考虑其他情况下，任务的价格越高，则会员更倾向于选择该任务，即两者是一种正相关关系。从表中可以看出任务完成率与任务的价格是呈正相关关系的，即任务的价格越高，任务完成率越大。这说明模型符合实际情况。

5.2 问题 2

题目要求给附件一中的任务，设计一种新的定价方案，并与问题 1 中的任务定价方案进行比较。

本文在问题 1 任务定价规律的建模分析中，得到了任务的价格与会员密度、任务密度均具有负相关关系的结论；由于问题 1 中模型求解只用了完成任务的数据，得到的效果不是很好。这也就说明该任务定价方案(5.1.5)不是很合理，可能是某些影响任务定价的因素考虑不够，或是其中参量需要进行调整。因此，在制定新的任务定价方案时，本文引入了任务 i 对应的最小距离，即该任务与距离它最近会员的距离。

因此，本文沿用问题 1 中的任务中心圈法，结合 MATLAB 软件进行编程（代码 distance1.m 见附件 2），得到每个任务的距离数据；再建立多元线性回归模型，利用 SPSS 软件进行求解；之后，根据所得到的新的定价模型，计算出各个任务的新标价，并根据问题 1 中所得到的任务完成率与任务的价格的函数模型(5.1.7)，计算出每个任务对应的完成概率，将新、旧模型进行总体比较。

5.2.1 模型的建立与求解

本文分析了会员密度、任务密度以及距离这三个自变量与因变量任务价格的关系。根据数据特点，建立了多元线性回归模型如下。

$$p_i = a_0 + a_1 * x_{1i} + a_2 * x_{2i} + a_3 * x_{3i} \tag{5.2.1}$$

其中，a_0、a_1、a_2、a_3为参量，p_i表示第i个任务的价格，x_{1i}表示第i个任务所对应单位圆内的会员密度，x_{2i}表示以第i个任务对应单位圆内的任务密度，

x_{3i} 表示第 i 个任务离最近会员的距离。

接下来，本文将任务中心圈法与 MATLAB 软件相结合，求出 835 个任务所在地的会员密度、任务密度以及每个任务点到最近会员的距离数据（见附表）。将所得数据与附件中对应的 835 个任务的价格数据，一起导入到 SPSS 软件中进行求解。

在求解过程中，发现模型的拟合优度即 R^2 均偏低，但模型和各变量的系数均能通过检验。通过观察距离数据，发现该数据大多小于自然数 e，且波动比较小。为了提高模型的拟合优度，本文对该数据进行取自然对数处理。之后再与未取自然对数的情况进行比较，将输出结果进行整理，得到表 5.2.1 至表 5.2.3。

根据 Pearson 相关性分析结果，如表 5.2.1 所示，可以看出任务的价格与距离、任务对应单位圆内的会员数及任务数的显著性水平均为 0，小于显著性水平 0.05，说明任务的价格任务与这三个变量有显著性相关。这三个变量相关系数分别为 0.335、-0.508、-0.491，说明任务的价格与最近会员距离具有显著性的正相关关系，与任务对应单位圆内的任务密度和会员密度具有显著性的负相关关系。

表 5.2.1　相关系数矩阵

		任务的价格	会员密度	任务密度	ln 距离
Pearson 相关系数	任务的价格	1	-0.508	-0.491	0.335
	会员密度	-0.508	1	0.664	-0.408
	任务密度	-0.491	0.664	1	-0.308
	ln 距离	0.335	-0.408	-0.308	1
Sig.(单尾)	任务的价格	.	0	0	0
	会员密度	0	.	0	0
	任务密度	0	0	.	0
	ln 距离	0	0	0	.

表 5.2.2 是对模型进行检验，检验结果表明，F 值为 121.237，对应的显著性水平是 0，小于置性水平 0.05，则拒绝原假设，说明该模型通过了检验，即该模型具有意义。

表 5.2.2　模型检验表

模型	R	R^2	F	Sig.
未取自然对数	0.552	0.304	121.237	0
取自然对数	0.563	0.317	128.51	0

表 5.2.3　系数表

模型	常量	会员密度	任务密度	距离
未取自然对数	72.375	-0.053	0.119	0.168
取自然对数	72.712	-0.046	0.119	0.662

注：每项变量的系数所对应的显著性均小于 0.05

从表 5.2.3 中可以看出，任务对应单位圆内的会员密度、任务密度和距离的系数分为-0.046、-0.119 和 0.662，所以可以得到模型为：

$$p_i = 72.712 - 0.046 * x_{1i} - 0.119 * x_{2i} + 0.662 * x_{3i} \qquad (5.2.2)$$

根据输出结果，可知会员密度、任务密度的系数分别为-0.046 和-0.119，且所对应的 Sig.均为 0，小于显著性水平 0.05，则拒绝原假设，说明会员密度和任务密度的系数通过了检验，对任务的价格具有显著性的负相关性，即表明任务的价格随着会员密度和任务密度的增加而降低。

距离的系数是 0.662，对应的 Sig.同样为 0，小于 0.05，说明最近会员距离的系数也通过了检验，且该变量对任务的价格具有显著性的正相关关系，表明任务的价格会随着距离的增加而增加。由此，可说明会员希望任务的价格应随着自身所花费的成本成正相关关系，即任务的价格随着成本的上升而增加。

从表 5.2.2 和表 5.2.3 中可以看出，对距离这一变量数据取自然对数后，模型的拟合优度虽然只有 31.7%，但是与未取自然对数的模型的拟合优度相比，有所提高，表明取自然对数的模型的效果更好。

最后，本文根据模型（5.2.2），计算出每个任务的新标价 p_i'，再计算 835 个任务的新标价的均值 $\overline{p}$，与任务的原标价均值作差，再除以原标价均值 $\overline{p}$，即得到任务的价格变化的任务率 Δp 。

接着根据模型（5.2.3），按同样步骤计算出任务完成的变化率 Δf 。

$$\begin{cases} \overline{p}' = \dfrac{1}{n}\sum_{i=1}^{n} p_i' \\ \overline{p} = \dfrac{1}{n}\sum_{i=1}^{n} p_i \end{cases} \Rightarrow \Delta p = \frac{\overline{p}' - \overline{p}}{\overline{p}} = \frac{\dfrac{1}{n}\left(\sum_{i=1}^{n} p_i' - \sum_{i=1}^{n} p_i\right)}{\dfrac{1}{n}\sum_{i=1}^{n} p_i}$$
$$\begin{cases} \overline{f}' = \dfrac{1}{n}\sum_{i=1}^{n} f_i' \\ \overline{f} = \dfrac{1}{n}\sum_{i=1}^{n} f_i \end{cases} \Rightarrow \Delta f = \frac{\overline{f}' - \overline{f}}{\overline{f}} = \frac{\dfrac{1}{n}\left(\sum_{i=1}^{n} f_i' - \sum_{i=1}^{n} f_i\right)}{\dfrac{1}{n}\sum_{i=1}^{n} f_i} \qquad (5.2.3)$$

再将任务的价格的变化率与任务完成的变化率相减，得到的差值记作 k，可作为评价任务定价方案的指标。

$$k = \Delta f - \Delta p \qquad (5.2.4)$$

评价标准为：在其他条件保持不变的情况下，任务价格的变化率越小越好，任务完成率的变化率越大越好；当 k>0 时，每单位价格带来的完成率更大，即价

格利用率更高，并且差越大，定价模型越优，所以该判别式可以较准确的表示定价模型的优劣。

表 5.2.4　新旧定价方案的比较

	平均标价	任务完成率%	标价的变化率%	任务完成率的变化率%	k
原定价方案	69.1108	62.51	—	—	—
定价方案(5.1.5)	69.1198	63.55	0.013	1.664	1.651
定价方案(5.2.5)	74.1198	71.32	7.234	12.227	4.993

表 5.2.4 中的新定价方案是根据上诉步骤计算得到，可以看出任务价格的变化率、任务完成率的变化率、k 分别为 0.013 %、1.664% 、1.651，均大于 0，各指标值均很小，则说明该方案的定价效果比初始定价方案的效果略好。所以有必要对该定价方案进行调整，即对模型（5.2.2）中的各参量进行微调。本文选择了一个最为简单的调整方法，即对模型中的常量进行调整。本文是将常量由原来的 72.712 调整为 77.712，再按照上述步骤计算得到的结果见表中定价方案(5.2.5)。

表 5.2.4 中的定价方案(5.2.5)的各指标明显比定价方案(5.1.5)的大，且均大于 0，特别是 k=4.993>0，说明定价方案(5.2.5)将任务的平均标价调高的同时，任务完成率也随之提高，但前者的变化率小于后者的变化率，表明该定价方案的效果较好。

$$p_i = 77.712 - 0.046 * x_{1i} - 0.119 * x_{2i} + 0.662 * x_{3i} \qquad (5.2.5)$$

5.3　问题 3

本题要求建立任务打包发布的情况下的新模型，并将其与问题 2 中的模型进行比较，分析新模型对任务完成情况的影响。所以，本文以沿用问题 1 的单位分格法，建立了任务包价格模型(5.3.2)。

首先，将调整单位分格法中步长即格子间距看作一个变量 b，如在问题 1 中 b=0.05。其次，将位于方格内的任务打包，即看作一个包。即可用单位分格法对整个地区的任务进行打包。接着，建立包的价格模型，由此求出包的价格，再代入完成概率模型(5.1.7)中，求出包的完成概率，最后与未打包的单个任务完成概率进行比较和分析。

考虑到第一问的范围中区间的经纬度变化值较大，于是固定范围，缩小区间的经纬度变化值，改良单位分格法的模型如下

$$\begin{aligned} & m = 2.1/b; n = 1.7/b; \\ & J_s = 112.5 + s*b, s = 1,2,3,\dots,m; \\ & W_t = 22.1 + t*b, t = 1,2,3,\dots,n. \end{aligned} \qquad (5.3.1)$$

换而言之，改进问题 1 中的单位分格法，使其适应于打包问题的重点就是找到合适的网格间距 b，使根据改进后单位分格法划分出来的打包方案是合理和有效的。

经计算可得，当纬度变化 0.02 度，实际距离变化值为 2.2264km，经度变化 b 度，实际距离变化值为 2.0643km。该距离在题目所给范围所占比例非常

小，满足将包内任务点的距离可忽略不计的条件。于是本题取 b=0.02。

5.3.1　模型的建立

假设一个任务包内有 n 个任务，会员接包后会完成该任务包内所有任务。

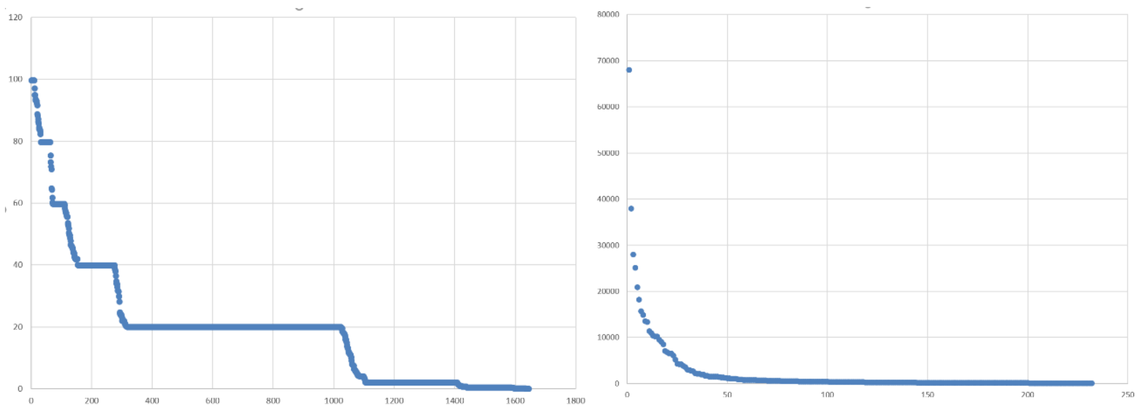

图 5.3.1　会员信誉值散点图

由图 5.3.1 可知，会员信誉值高于 100 是呈幂函数分布；会员信誉值低于 100 是呈分段函数分布，其中，信誉值低于 5 的会员有 750 个，为保障服务质量，假设信誉值低于 5 的会员不分配打包任务。

根据广州、深圳地区的人均日工资水平及会员的平均日工作能力、任务的平均定价（60 元~70 元），每包内平均任务数的取值小于 4 为宜。

显然，任务包的价格由包内任务的数量和价格确定，于是本文建立如下任务包定价模型和任务包总完成期望模型。

$$\begin{cases} P_i^{'} = a \times \sum\limits_{k=1}^{n} P_k \\ F_{\mathrm{m}} = \dfrac{1}{N} \sum\limits_{j=1}^{n} f_j \end{cases} \tag{5.3.2}$$

其中 $P_i^{'}$ 为第 i 个任务包的价格；P_k 为该任务包中的第 k 个任务的价格；a 为修正值；F_m 为所给任务的总完成期望；f_j 为该任务包总量中的第 j 个任务包的完成概率；N 为任务包总量。由于数据量较大，本文使用 matlab 进行求解 (求解文件为 jiage.m)。在附件的支撑材料中，接着使用 EXCEL 统计软件对所求任务包进行完成概率求解，最后求出所给任务的总完成期望。

5.3.2　模型的求解

根据上述步骤可求出，有 478 个任务包，将其命名为包 1 至包 478。附录挑选了包 1 至包 37 的详细信息作为示例，全部数据在支撑材料“问题三打包及价格”中。

则可得，打包和未打包的总完成率期望、打包和未打包的总价格比较结果如表 5.3.1 所示。

表 5.3.1　打包与未打包情况对比表

		总完成期望	单个任务平均价格(元)
a=1	打包	0.842287	74.1198
	未打包	0.713147	74.1198
a=0.9	打包	0.773268	66.7078
	未打包	0.713147	74.1198

由表中信息可知，当 a=1 时，打包情况下的总完成期望达到 0.842287，即有 84.2287%的任务被完成，未打包情况下的总完成期望 0.713147，即有 71.3147%的任务被完成，打包情况下的总完成期望比未打包情况下的总完成期望高出 12.914%，同时单个任务平均价格没发生改变。

当 a=0.9 时，打包情况下的总完成期望不仅比未完成情况下的总完成期望高出 6.0121%，而且总价格下降了 10%。

根据表中信息，本文得出以下结论，打包发布的任务定价策略优于不打包发布的定价策略，为了使总完成期望较高且价格较合适，取 a=1 时较为合适。

图 5.3.2 是对任务进行打包后的整体价格分布图。从图中可以看出，打包分布后的情况，可以基本分为两个区域。其中一个区域是以某点为中心，呈圆形放射状向四周递减；另一个区域则是一个以边缘为中心，呈扇形放射状向四周递减。这与会员密度分布情况比较相似。

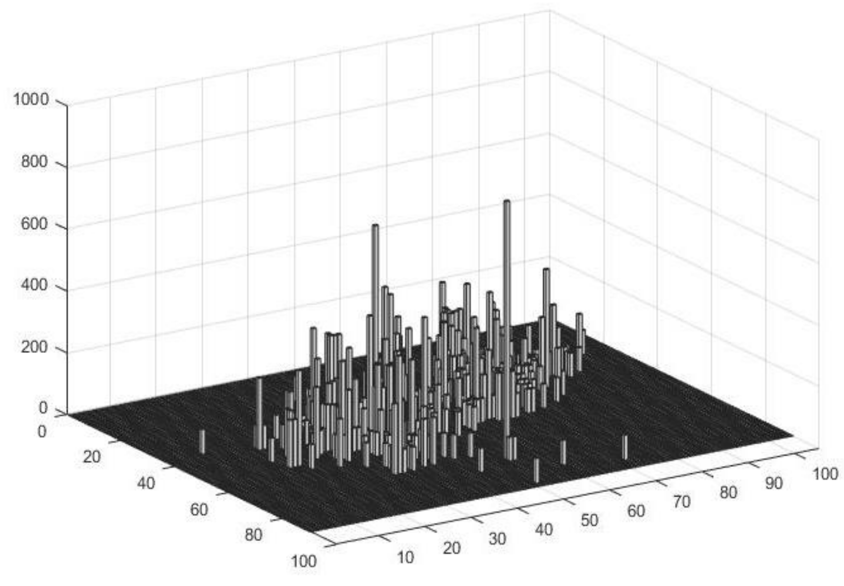

图 5.3.2　价格分布图

5.4　问题 4

该题采用问题 3 的单位格法打包策略，使用任务包定价模型和任务包总完成期望模型，求出平均任务标价和总完成期望。

首先对数据进行分析，作出附件 1 与附件 3 的散点图如下

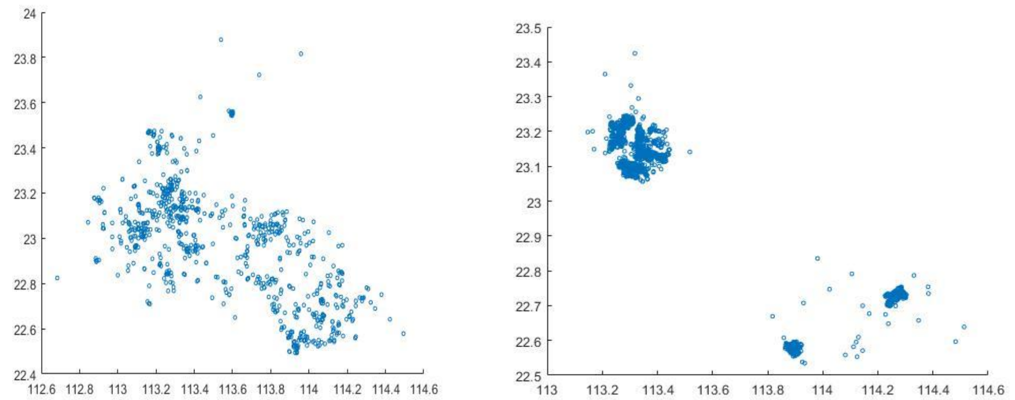

图 5.4.1　附件 1 与附件 3 散点图

从图中可知，新项目的任务点密集程度更高，且无任务点在所取范围之外。所以可以使用问题 3 的定价模型。

接着沿用问题 2 中的任务中心圈法，用求某任务周围会员密度、任务密度、和成本的 MATLAB 代码，来求每个新任务会员密度、任务密度、和成本。本文采用不打包和打包两种策略的定价模型，求出不同的新任务的定价和完成度，进行比较和分析。

经过多次尝试，最终取 b=0.005，利用问题 3 的 MATLAB 代码，修改经纬度变化值与生成矩阵的行列数，最终求得新任务的定价和完成度见附表。本文用打包和未打包的总完成率，打包和未打包的总价格的结果如下表

表 5.4.1　附件 1、附件 3 提升效果对比表

附件 3 任务		总完成期望	单个任务平均价格(元)
a=1	打包	0.706187	42.76465
	未打包	0.236952	42.76465

附件 1 任务		总完成期望	单个任务平均价格(元)
a=1	打包	0.842287	74.1198
	未打包	0.713147	74.1198

由此可知，附件 3 任务的打包与不打包的总完成期望相差 0.469235，附件 1 任务的打包与不打包的总完成期望相差 0.121940，两者相差较大。所以打包策略对任务点密集的地区的总完成期望的促进作用大于任务点不那么密集的地区。由此可以作出以下评价， 当任务点较密集时，使用打包策略能提高任务完成率。根据上述制定的定价方案，所得到的任务价格分布情况，如表 5.4.1。

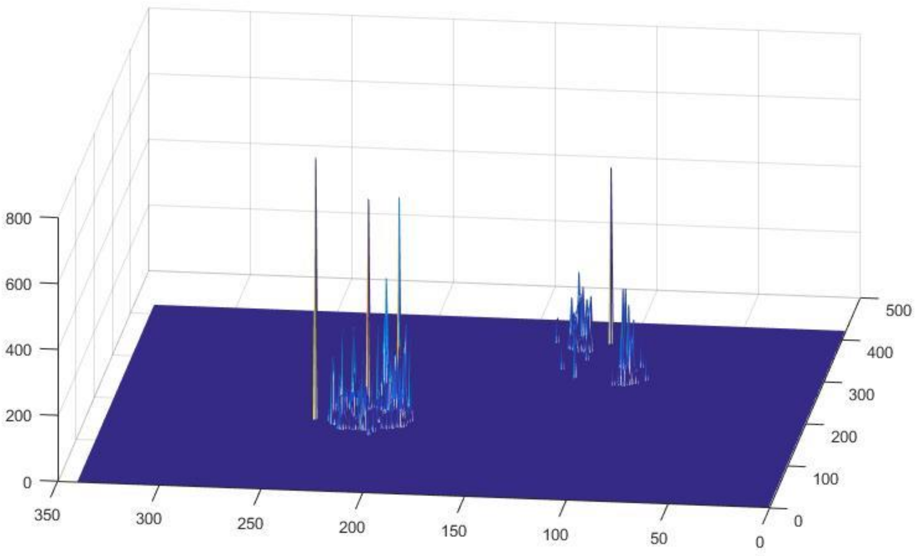

图 5.4.2　新任务的价格分布图

六、模型评价

6.1　优点

通过三次模型的优化，逐步将任务平均价格降低，而提高了任务完成率，最后采用了打包定价方案以后，任务的平均价格下降到了 64.12 元，而任务完成率达到了 93.23%，表明打包定价模型性能优良，在众包服务行业具有一定的推广应用价值。

6.2　不足

相对快递行业与出租车行业的定价模型[4]，本文中的模型仍较为复杂，任务定价与任务分布状况以及会员分布状况高度相关，定价波动较大，计算较为不便。

另外，本文未利用会员的预定配额、预定时间以及信誉值。

参考文献

[1]刘苏．专业服务众包模式应用研究[D].北京工业大学，2011.

[2]何晓群.多元统计分析(第四版)[M].北京：中国人民大学出版社，2015.

[3]宇传华.SPSS 与统计分析(第二版)[M].北京：电子工业出版社，2014.

[4]张启波．基于服务质量的快递企业服务定价[D].大连海事大学，2012.

[5]姜启源,叶其孝.数学建模[M].北京：机械工业出版社，2009.

[6]刘志新．基于顾客感知价值的快递服务定价研究[D].大连理工大学，2015.

[7]吴孟达.数学建模的理论与实践[M].长沙：国防科技出版社，1999.

附 录

附录 1：density1.m

问题 1 求距离、任务密度、会员密度函数

```
M=zeros(34,42);
N=zeros(34,42);
Z=zeros(1,835);
G=zeros(1,835);

for i=1:1867 %ppl density
    flag=0;
    X=xcood(i,:);
    for j=1:42%jingdu
        if X(2)>112.5+j*0.05 && X(2)<112.5+(j+1)*0.05
            for k=1:34 %weidu
                if X(1)>22.1+k*0.05 && X(1)<22.1+(k+1)*0.05
                    M(k,j)=M(k,j)+1;
                    flag=1;
                    break;
                    if flag == 1
                        break;
                    end
                end
            end
        end
    end
end

for m=1:835 %task density
    flag=0;
    Y=ycood(m,:);
    for e=1:42%jingdu
        if Y(2)>112.5+e*0.05 && Y(2)<112.5+(e+1)*0.05
            for f=1:34 %weidu
                if Y(1)>22.1+f*0.05 && Y(1)<22.1+(f+1)*0.05
                    N(f,e)=N(f,e)+1;
                    flag=1;
                    break;
                    if flag == 1
                        break;
```

```
                    end
                end
            end
        end
    end
end

for m=1:835 %task density
    flag=0;
    Y=ycood(m,:);
    for e=1:42%jingdu
        if Y(2)>112.5+e*0.05 && Y(2)<112.5+(e+1)*0.05
            for f=1:34 %weidu
                if Y(1)>22.1+f*0.05 && Y(1)<22.1+(f+1)*0.05
                    Z(m)=N(f,e); % how many task in every task place cell?
                    G(m)=M(f,e); % how many ppl in every task place cell?
                    flag=1;
                    break;
                    if flag == 1
                        break;
                    end
                end
            end
        end
    end
end

M
Z
G

for a=1:835
if G(a)==0
    DENS(a)=inf;
elseif G(a)~=0
    DENS(a)=Z(a)./G(a);
end
end

DENS
```

附录 2：distance1.m

问题 2 任务中心圈求距离、任务密度、会员密度函数

```
DIST = zeros(1867,835);

for tasks = 1:835
    TASKS = ycood(tasks,:);
    for members = 1:1867
    MEMBERS = xcood(members,:);
    DISTANCE1 = distance(TASKS(1),TASKS(2),MEMBERS(1),MEMBERS(2));
    DIST(members,tasks) = DISTANCE1*pi/180*6371;
    end
end

% compute the number of tasks and people and people within the 5km circle
TASNUM = zeros(835,1);
PPLNUM = zeros(835,1);

for tasks1 = 1:835
    TASKS1 = ycood(tasks1,:);

    % number of tasks in circle
    for tasks2 = 1:835
        TASKS2 = ycood(tasks2,:);
        if (distance(TASKS1(1),TASKS1(2),TASKS2(1),TASKS2(2)))*pi/180*6371 <= 5
            TASNUM(tasks1) = TASNUM(tasks1)+1;
        end
    end

    % number of ppl in circle
    for ppl2 = 1:1867
        PPL2 = xcood(ppl2,:);
        if (distance(TASKS1(1),TASKS1(2),PPL2(1),PPL2(2)))*pi/180*6371 <= 5
            PPLNUM(tasks1) = PPLNUM(tasks1)+1;
        end
    end

end


MINDIST = min(DIST)'
TASNUM
PPLNUM
DENSITY = TASNUM./PPLNUM
```

附录3：jiage.m

MATLAB软件编程求打包价格

```
M=zeros(90,105);
for m=1:835 %task density
    flag=0;
    Y=y(m,:);
    for e=1:105%jingdu
        if Y(2)>112.5+e*0.02 && Y(2)<112.5+(e+1)*0.02
            for f=1:90 %weidu
                if Y(1)>22.1+f*0.02 && Y(1)<22.1+(f+1)*0.02
                    M(f,e)=M(f,e)+Y(3);
                    flag=1;
                    break;
                    if flag == 1
                        break;
                    end
                end
            end
        end
    end
end
h=zeros(1,500);
k=zeros(1,500);
i=0;
for n=1:9450 %task density
        if M(n)>0
            i=i+1;
          h(i)=M(n);
          k(i)=N(n);
        end
end
```