

# “拍照赚钱”任务的合理定价探究

## 摘要

针对问题一，我们用定位软件绘制出已结束任务的地理位置分布图。对比了完成的和未完成的任务的地理位置分布，和不同价格区间的任务的地理位置分布，发现任务的价格分布与其距市中心的距离成正相关。同时发现任务完成情况呈现地域性，另外考虑该市区的情况。剔除掉一些极端情况的特殊点后，用 EXCEL 拟合所有的样本点，拟合度达 99%，得到定价模型。最终我们分析出附件一中项目失败的原因主要有二，一是没有考虑地区的经济水平，在每个地区的基础定价都是一样的；二是价格随任务偏远程度的变化不够层次化。

针对问题二，我们以任务的完成情况为 0-1 变量，采用 logit 模型，以完成情况为被解释变量，以任务的价格，任务距市中心的距离和该市的人均 GDP 为解释变量，用 SPSS 软件拟合出各解释变量的系数，得到计算任务完成率的模型，为评价体系。带入附件一中数据，实际任务完成率与 logit 模型的理论完成率相差不到 1%，评价体系有效。我们将任务距市中心的距离据 logistic 模型计算，代替定价模型中的原距离，用 EXCEL 拟合优化后的模型成功的样本点，拟合度达到 94%，定价优化有效。将优化后的定价模型带入 logit 模型，发现提高了 62% 的任务的完成率。

针对问题三，通过 matlab 对分布的任务聚类，所有的任务以任务包的形式出现。任务包完成分为被接受和被执行两步。考虑任务包在以其为中心一定距离范围内的被接受率（与该范围内会员的分布有关，与会员接受任务的限额有关），得到接受的概率。以任务包中的任务的平均距离代替任务包的距离，用问题二中优化后的定价模型对任务包定价。再用问题二中的 logit 模型，计算出一个任务包的完成率。考虑违约风险（与该范围内限额大于任务包中任务数量的会员的信誉值有关，与总会员的信誉水平有关），完成率乘上守约概率得到任务包执行概率。接受概率乘上执行概率得到该任务包的考虑违约的完成率。以任务包的完成率代替其中每个任务的完成率，求平均值得到所有任务的完成率。

针对问题四，根据附件三中的数据，将新任务的分布聚合，得到任务包的分布，用问题三中任务包形式下的完成概率模型，借助 matlab，计算出最终的所有任务完成的平均概率。

**关键词：**任务打包    聚类分析    logit 模型    线性定价    matlab

# 正文

## § 1 问题重述

### 1.1 引言

“拍照赚钱”是移动互联网下的一种自助式服务模式。用户下载 APP，注册成为 APP 的会员，然后从 APP 上领取需要拍照的任务（比如上超市去检查某种商品的上架情况），赚取 APP 对任务所标定的酬金。这种基于移动互联网的自助式劳务众包平台，为企业提供各种商业检查和信息搜集，相比传统的市场调查方式可以大大节省调查成本，而且有效地保证了调查数据真实性，缩短了调查的周期。因此 APP 成为该平台运行的核心，而 APP 中的任务定价又是其核心要素。如果定价不合理，有的任务就会无人问津，而导致商品检查的失败。

### 1.2 问题的提出

- 1.研究附件一中项目的任务定价规律，分析任务未完成的原因。
- 2.为附件一中的项目设计新的任务定价方案，并和原方案进行比较。
- 3.实际情况下，多个任务可能因为位置比较集中，导致用户会争相选择，一种考虑是将这些任务联合在一起打包发布。在这种考虑下，如何修改前面的定价模型，对最终的任务完成情况又有什么影响？
- 4.对附件三中的新项目给出你的任务定价方案，并评价该方案的实施效果。

## § 2 问题分析

### 2.1 问题一

根据附件 1 中已完成项目的数据，由于经纬度数值不能直观地反映信息，需要进一步挖掘，我们利用软件定位出每个样本点的实际位置，观察分析失败的和成功的项目的分布图，和不同价位区间的项目的分布图，对比两图，分析出项目至地区中心的距离与其价格呈正相关，且项目完成情况呈区域性分布。引入虚拟变量区分四个区域，人均 GDP 和项目至地区中心距离为解释变量，任务价格为被解释变量，拟合出最佳曲线，解释附件一中的定价规律，分析出失败的原因。

### 2.2. 问题二

为了对优化后的价格的效果量化，需要一个评价体系，以任务的完成率为指标，构造

logit 模型，得到目标函数即任务完成的概率。比较计算出的理论完成率与实际完成率，判断评价体系的有效性。根据问题一中分析出的任务未完成的原因，将原价格模型中的距地区中心距离 logit 计算化，使得近处的价格适宜上升，远处的价格适宜下降，将改动后的价格模型与任务成功的样本点进行拟合。计算每一个任务的完成概率值，观测多少个任务的完成概率将上升。

### 2.3. 问题三

根据附件二中会员的位置和会员的信誉度，考虑了将任务打包后任务包的被接受情况和被接受后的违约情况。在一定范围内距离近的会员愿意接受，再根据会员的接受任务的限额，考虑会员接受任务包的能力，得到一个任务包被接受的概率。计算能接受任务包的会员的信誉度的数学期望，据其信誉在愿接受任务包的会员中的排名情况来计算违约风险。再有问题二中的价格影响的任务完成率模型，计算任务包在一定价格下能被完成的概率。最后得出所有的任务被完成的平均概率。

### 2.4. 问题四

根据附件三中的数据，将新任务的分布聚合，得到任务包的分布，用问题三中任务包形式下的完成概率模型，借助 matlab，计算出最终的所有任务完成的平均概率。

## § 3 名词解释与符号说明

表 1.名词解释与符号说明

序号	符号	说明
1	$x_1$	任务点距市中心的距离
2	$x_2$	该市的人均 GDP
3	$D_{\text{佛山}}$	$D_{\text{佛山}} = \begin{cases} 1, & \text{佛山} \\ 0, & \text{不是佛山} \end{cases}$
4	$D_{\text{广州}}$	$D_{\text{广州}} = \begin{cases} 1, & \text{广州} \\ 0, & \text{不是广州} \end{cases}$
5	$D_{\text{东莞}}$	$D_{\text{东莞}} = \begin{cases} 1, & \text{东莞} \\ 0, & \text{不是东莞} \end{cases}$
6	$\text{logit}(\pi_x)$	在给定 $x$ 的条件下，变量=1 的概率
7	$G$	评价体系的评分值，即在人均 GDP，任务价格，所在市区，距市中心距离和任务价格给定情况下任务完成的概率
8	$y$	任务价格
9	$y'$	优化后的价格

10	$N$	任务包的数目
11	$i$	任务包的编号
12	$n_i$	第 $i$ 个任务包里任务的个数
13	$J$	第 $i$ 个任务包里任务的编号
14	$A$	距离范围内预订该任务包的会员个数
15	$k$	距离范围内预订该任务包的会员编号
16	$q_k$	该会员的任务领取限额
17	$c$	会员的信誉值
18	count	计数函数
19	$\beta$	能接受到任务包的会员数
20	$x_{i1}$	第 $i$ 个任务包距市中心的优化后的距离
21	$G_{i\text{任务包}}$	考虑任务包所在市，任务包离所在市市中心的距离，该市人均 GDP，任务包价格的条件下，第 $i$ 个任务包的完成率
22	$P_{i1}$	一定范围内，有能力接受任务包的会员占范围内会员数的比例，即任务包能被接受的概率
23	$P_{i2}$	考虑违约风险下，会员不会违约的概率
24	$P_{i\text{任务包}}$	考虑违约风险，任务包接受力的条件下，任务包能完成的概率
25	$P$	$N$ 个任务包总体的完成概率

## § 4 模型假设

假设 1.假设附件一中未完成任务只受到任务价格和地理位置的影响

假设 2.假设市民的消费能力和该市的人均 GDP 成正比

假设 3.假设市政府所在位置为该市的市中心

假设 4.假设以任务到市中心的距离最小来划分任务所在区

## § 5 模型建立与求解

### 5.1 问题一的建立与求解

在问题一中，依据 834 个已结束任务的经纬度，利用地图定位软件分别定位出成功的、失败的任务地理位置分布图像，及对应不同价格的任务的地理位置分布图像，将活动范围划分为深圳、佛山、东莞和广州四个地区，发现任务的价格与任务的偏远程度（任务完成地点距离地区中心的距离）呈现正相关。最终以 4 个地区，任务偏远程度及地区的经济发展（以人均 GDP 为依据）作为三个解释变量，任务价格为被解释变量，用 EXCEL 进行回归，得到附件一中的定价模型，拟合度达到 99%。

### 5.1.1 多元回归预备知识

设影响因变量  $y$  的自变量有  $m$  个, 记  $x_1, x_2, \dots, x_m$ , 随机因素总和用  $\varepsilon$  表示, 多元线性回归模型记作

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \varepsilon, \quad \varepsilon \sim N(0, \delta^2)$$

已知一组独立数据  $(y_i, x_{i1}, x_{i2}, \dots, x_{im}), i=1, \dots, n, n>m$ , 记

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \beta = (\beta_0, \beta_1, \dots, \beta_m)^T$$

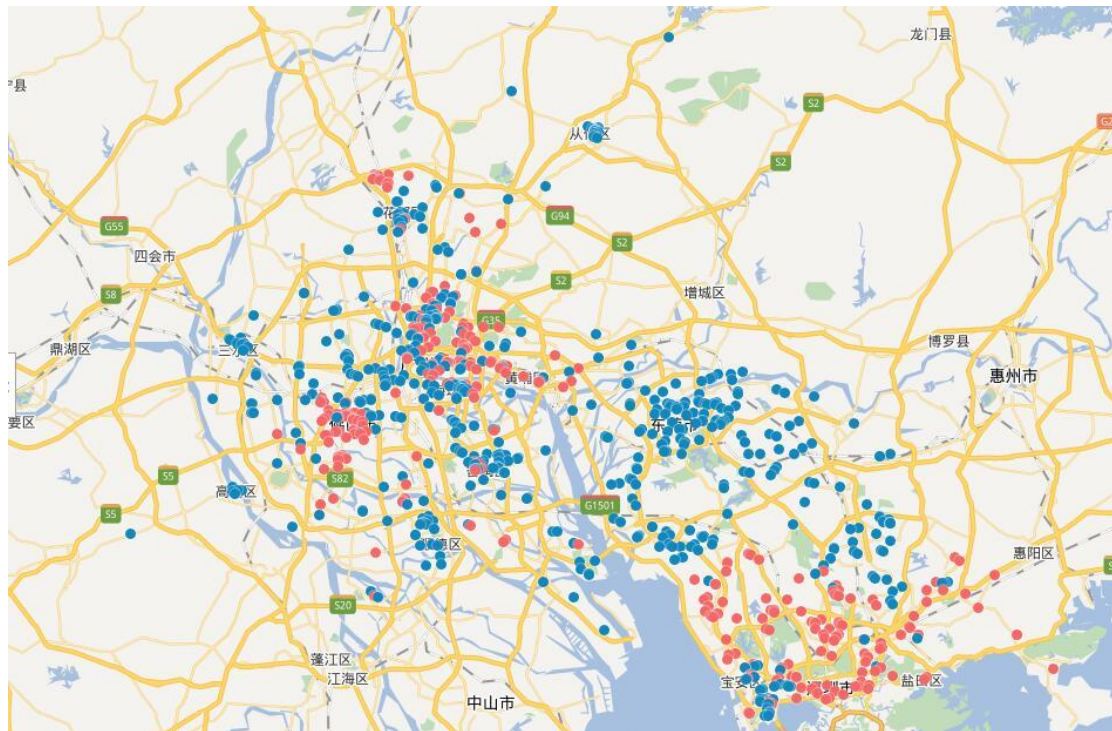
代入模型得

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \delta^2 I)$$

按照最小二乘准则估计模型系数  $\beta$ , 估计值记作  $\hat{\beta}$ , 可得

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

### 5.1.2 Step1.数据信息挖掘



(注: 蓝色点代表完成任务, 红色点代表未完成任务)

图 1.任务地理位置分布

可以看出, 任务完成情况与地域分布明显相关, 其中, 深圳的任务完成情况最差, 广州和佛山的市中心的任务完成率相对偏低, 东莞的完成情况最佳。



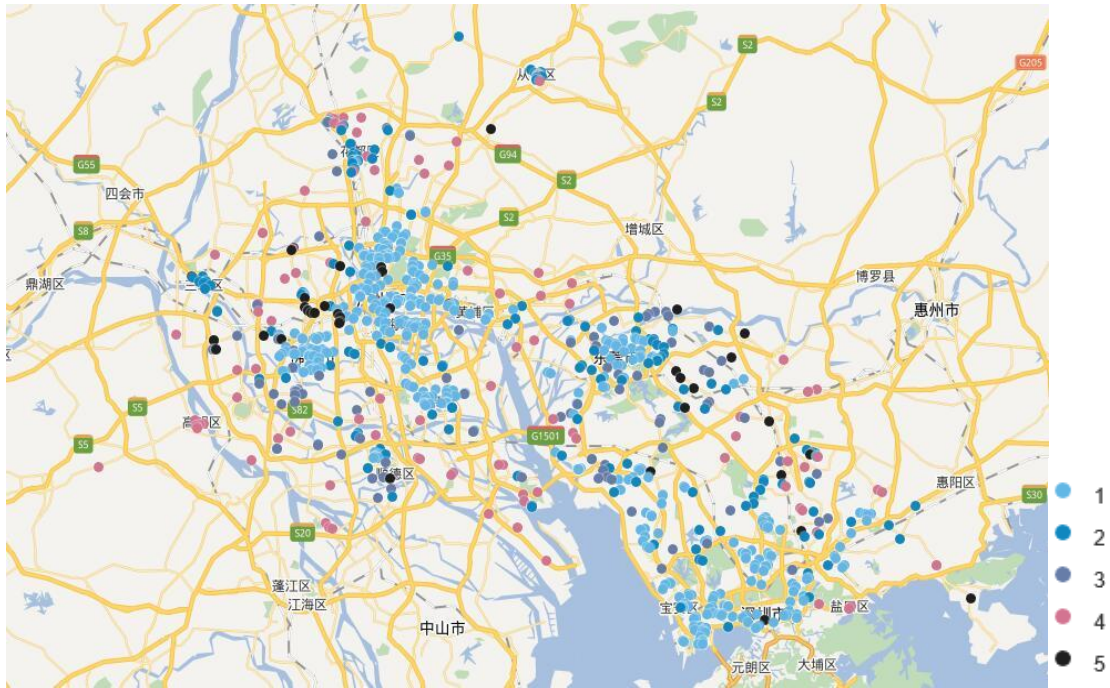


图 2.不同价格的任务的地理位置分布

表 2.价格区间

1	价格为 65、65.5、66、66.5、67.5
2	价格为 68、68.5、69、69.5、70
3	价格为 70.5、71、71.5、72、72.5
4	价格为 73、73.5、74、74.5、75
5	价格为 80、85

可以看出，与地区中心距离越远，定价越高。广州、佛山市中心价格可能偏低，不能满足市中心人民的机会成本，失败率偏高。且可以看出东莞的距离价格分布合理，任务基本完成，深圳的距离价格分布没有层次差别，任务未完成率偏高。

表 3.各地区任务成功率

共 834 个样本。		
深圳，184 个样本	成功 63 个	成功率 34.24%
佛山，160 个样本	成功 100 个	成功率 62.5%
东莞，183 个样本	成功 172 个	成功率 93.99%
广州，307 个样本	成功 187 个	成功率 60.91%

## Step2. 构建模型

(1)、变量设定：选取解释变量距市中心距离，佛山，广州，各市人均 GDP，东莞和深圳。记为

$x_1, D_{\text{佛山}}, D_{\text{广州}}, x_2, D_{\text{东莞}}$

其中

$$D_{\text{佛山}} = \begin{cases} 1, & \text{佛山} \\ 0, & \text{不是佛山} \end{cases}; D_{\text{广州}} = \begin{cases} 1, & \text{广州} \\ 0, & \text{不是广州} \end{cases}; D_{\text{东莞}} = \begin{cases} 1, & \text{东莞} \\ 0, & \text{不是东莞} \end{cases}$$

任务标价为被解释变量，记为  $y$

(2)、相关系数检验：

表 4. 任务价格与各解释变量的相关性

	市中心 距离	广州	东莞	佛山	人均 GDP	任务标 价
市中心距 离	1					
广州	0.099077	1				
东莞	0.040999	-0.40397	1			
佛山	-0.11072	-0.37124	-0.25793	1		
人均 GDP	-0.04197	-0.20155	-0.47944	-0.2739	1	
任务标价	0.349273	-0.17431	0.124996	0.203605	-0.15667	1

$x_1, D_{\text{佛山}}, D_{\text{广州}}, x_2, D_{\text{东莞}}$  与  $y$  的相关系数的大小，代表了它们对  $y$  的解释能力的大小，相关系数越大，解释变量在模型中的位置越靠前。

根据相关系数。对解释变量排序，初步构建模型为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 D_{\text{佛山}} + \beta_3 D_{\text{广州}} + \beta_4 x_2 + \beta_5 D_{\text{东莞}}$$

### Step3. 模型求解

表 5. 解释变量解释效果

回归统计	
Multiple R	0.998337
R Square	0.996676
Adjusted R Square	0.995425
标准误差	4.000992
观测值	815

表 6. 回归系数

	Coefficients	标准误差	t Stat	P-value
Intercept	0	#N/A	#N/A	#N/A
市中心距离	0.162659	0.013746	11.83313	6.52E-30
佛山	46.85034	0.35322	132.6378	0
广州	37.92186	0.2949	128.5925	0
人均 GDP	12.48856	0.069623	179.3742	0

东莞	51.71668	0.360576	143.4281	0
----	----------	----------	----------	---

可以看出，贡献率高达 99.5%，说明自变量向量（ $x_1$ ， $D_{\text{佛山}}$ ， $D_{\text{广州}}$ ， $x_2$ ， $D_{\text{东莞}}$ ）对因变量  $y$  解释效果好。各自变量参数均通过显著性检验（ $t$  检验），模型解释有效。

## Step4. 结果分析

得到附件一中已结束项目的定价模型为

$$y = 0.162659x_1 + 46.85034D_{\text{佛山}} + 37.92186D_{\text{广州}} + 12.48856x_2 + 51.71668D_{\text{东莞}}$$

当距市中心距离每增加一个单位，定价提高 0.1627 元；

当所在城市的人均 GDP 每提高一万点，定价提高 12.4886 元；

在四个城市人均 GDP 在同一水平相等的情况下，即考虑除 GDP 外的地区因素（如地理，人口密度，城市文化等因素）

当地区在佛山，即  $D_{\text{佛山}}=1$ ， $D_{\text{广州}}=0$ ， $D_{\text{东莞}}=0$  时，价格比深圳提高了 46.8503 元；

当地区在广州，即  $D_{\text{广州}}=1$ ， $D_{\text{东莞}}=0$ ， $D_{\text{佛山}}=0$  时，价格比深圳提高了 37.92186 元；

当地区在东莞，即  $D_{\text{东莞}}=1$ ， $D_{\text{广州}}=0$ ， $D_{\text{佛山}}=0$  时，价格比深圳提高了 51.71668 元

而佛山、广州和东莞的地区带来的价格的相对差距为两两地区  $D_i$  之前的系数之差。即：

考虑除 GDP 外的地区因素，在东莞的定价比在广州的定价提高了 13.79482 元，比在佛山的定价提高了 4.86638 元；在佛山比在广州的定价提高了 8.92844 元。

经过分析，任务的完成情况与所在地区密切相关，价格分布与距离市中心距离相关。

可以看出，价格范围分划的梯度很小：0.5 元，虽然离市中心距离较远时价格有所提高，但定价时没有考虑该地区的消费水平，价格的小幅度提高可能不足以弥补距离增加所带来的耗费。另外，还应考虑当地的人均收入水平，若人均收入水平高，则人们去拍照赚钱的机会成本提高，若价格不能提高一定幅度，人们不会选择完成任务。比如深圳的经济发达，任务完成率也最低。

## 5.2 问题二的建立与求解

以任务完成率为指标，根据 logic 模型构建目标函数，用 SPSS 软件拟合出目标函数中的系数。对未优化的定价模型里的  $x_1$ （任务点距地区中心的距离）进行 logit 计算，使得近郊的价格上升一点，远郊的价格下降一点，并与任务成功完成的样本点拟合，用 EXCEL 回归得到新的定价模型  $y'$ ，带入新的定价模型  $y'$  到目标函数，对比与之前的  $y$  所得函数值的



差异，分析任务完成情况。

### 5.2.1 logit 模型预备知识

设Y为 0-1 变量,x 为影响 y 分布的条件,用 $\pi(x)$  表示在 x 发生的条件下 Y=1 的概率,  
即  $\pi(x) = P(Y = 1|x)$

根据 logistic 模型:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

其反函数为:  $\ln \frac{\pi(x)}{1-\pi(x)} = \beta_0 + \beta_1 x$

记  $\ln \frac{\pi(x)}{1-\pi(x)}$  作  $\text{logit}(\pi_x)$  ,  $\beta_0$ 、 $\beta_1$  是回归系数

### 5.2.2 Step1.评价体系构建

记  $Z = \begin{cases} 1, & \text{任务完成} \\ 0, & \text{任务未完成} \end{cases}; \quad X = (y, x_1, x_2)$

得到目标函数  $G = \text{logit}(\pi(x)) = \beta_0 + \beta_1 y + \beta_2 x_1 + \beta_3 x_2$

价格优化:  $x'_1 = \text{logit}(x_1) = \ln \frac{x_1}{1-x_1}$

新定价模型:  $y' = \beta'_0 + \beta'_2 x'_1 + \beta'_2 D_{\text{佛山}} + \beta'_3 D_{\text{广州}} + \beta'_4 x_2 + \beta'_5 D_{\text{东莞}}$

### Step2. 模型求解

优化后的定价如下图:

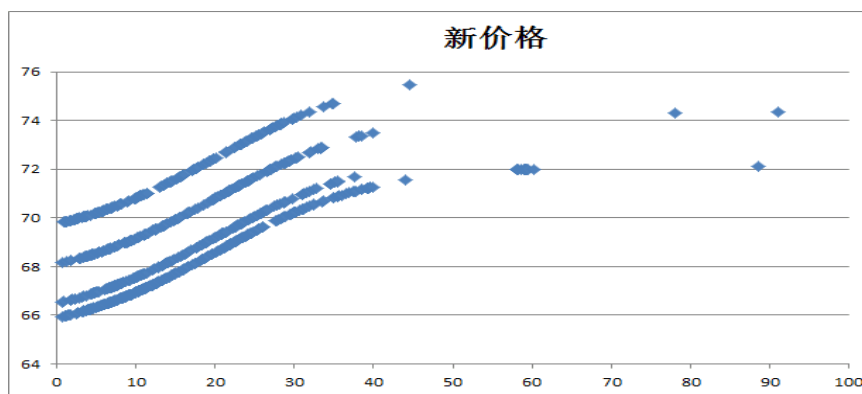


图 3.优化价格

从下至上依次为广州、深圳、东莞和佛山。

横轴为任务距市中心距离。

优化后，近郊距离的价格适宜上升，远郊距离的价格适宜降低。且得到了定价随距离的

层次分布。

目标函数 G 系数求解结果：

方程中的变量		B	S.E.	Wals	df	Sig.	Exp (B)
步骤 1 <sup>a</sup>	任务标价	.057	.021	7.490	1	.006	1.058
	市中心距离	.033	.008	15.879	1	.000	1.034
	人均GDP	-.507	.054	88.500	1	.000	.602
	常里	-2.577	1.398	3.397	1	.065	.076

图 4.评价体系系数回归结果（SPSS）

即 $\beta_0=-2.577$      $\beta_1=0.057$      $\beta_2=0.033$      $\beta_3=-0.507$

$G = \text{logit}(\pi(x)) = -2.577 + 0.057y + 0.033x_1 - 0.507x_2$

EXCEL 回归新定价模型结果：

表 7.优化模型的合理性

回归统计	
Multiple R	0.974071
R Square	0.948814
Adjusted R Square	0.946283
标准误差	15.5755
观测值	481

表 8.优化模型系数回归结果

	Coefficients	标准误差	t Stat	P-value
Intercept	0	#N/A	#N/A	#N/A
人均 GDP	4.857436	0.544569	8.919774	1.01E-17
佛山	10.14391	2.103421	4.822579	1.91E-06
广州	13.32511	1.964252	6.783813	3.49E-11
东莞	16.55476	1.901708	8.705208	5.23E-17
Logit ( $x_1$ )	16.63388	0.663908	25.05447	5.68E-89

可见，决定系数为 0.94，即成功任务的定价 94%可由这些解释变量来解释。解释变量通过显著性 t 检验，新定价模型为：

$$y' = 16.63388x_1' + 10.14391D_{\text{佛山}} + 13.32511D_{\text{广州}} + 4.857436x_2 + 16.55476D_{\text{东莞}}$$

Step3. 结果分析

G=62.40%，即 834 个样本中任务成功完成率为 62.40%，与实际统计结果 62.58%相符合。则评价体系有效。

经过价格优化，使得距离地区中心较近的点价格有所上升，距离地区中心较远的点价格

有所下降,改善图 1,图 2 中所反映的失败任务位置集中在地区中心近点情况,同时控制了成本。结果显示,有 517 个样本的 G 值提高,即提高了 517 个样本的理论成功率。

## 5.3 问题三的建立与求解

对打包任务的定价与评估,我们分为以下四个阶段:

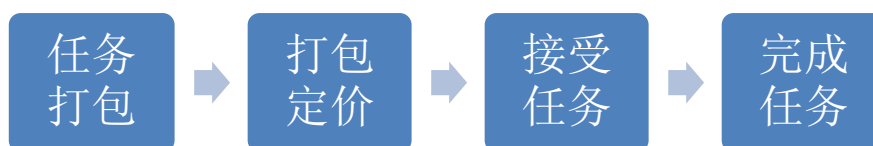


图 5.打包任务实施流程

(1)任务打包:通过分析新任务位置的分布,根据各个任务点分布的密集程度,利用 matlab 进行聚类,得到每个任务包中的任务个数以及该任务包中每个任务的位置。

(2)一个任务包的价格由该任务包中任务的个数以及各个任务的分布地点确定,为了更便捷地定价,对该任务包中每个任务的位置取平均(经纬度分别取数学期望),计算对应的价格来作为该任务包中每个任务的平均价格。

(3)每个会员根据自己的位置与限额,可以在网上对自己限额以内的任务包进行挑选以及预订。对于有多名会员预订单的任务包,系统将根据会员的信誉度排名以及该名会员剩余的限额对任务进行分配。

(4)影响任务完成情况的因素:一方面,信誉度越高的会员,接受任务后,完成任务的概率也越大;另一方面,会员离该任务包的距离越远,任务包被该会员预订的概率也就越小。

### 5.3.1 打包任务的定价模式

记任务包的个数为  $N$ 。记  $i=1,2,\dots,N$  为任务包的编号;记  $j=1,2,\dots,n_i$  为任务包中的任务的编号;以任务包为中心,  $R$  为半径,记此距离范围内预订该任务包的会员个数为  $\alpha$ , 其编号为  $k=1,2,\dots,\alpha$ ;  $q_k$  为该会员的限额

记信誉值为  $c$ ,  $\text{count}$  为计数函数,能接受到任务包的会员数为  $\beta$

#### Step1.聚类分析

录入新任务的经纬度,利用 matlab 进行聚类。

聚类分析指将物理或抽象对象的集合分组为由类似的对象组成的多个类的分析过程。在

本模型中，根据任务样本的分布情况，通过任务之间位置的相似性（距离的远近）对任务进行分类，一些彼此之间位置相似程度较大的任务聚合为一类，另外一些彼此之间位置相似程度较大的样本聚合为另一类，距离较近的聚合到一个小的分类单位，距离较远的又聚合到一个较大的分类单位，直到把所有样本都聚合完毕。

## Step2.任务包的定价

$$\bar{d}_{i\text{经度}} = \frac{1}{n_i} \sum_{j=1}^{n_i} d_{j\text{经度}} \quad \bar{d}_{i\text{纬度}} = \frac{1}{n_i} \sum_{j=1}^{n_i} d_{j\text{纬度}}$$

任务包定位 D ( $\bar{d}_{i\text{经度}}, \bar{d}_{i\text{纬度}}$ )

由 D 与所在市市中心的经纬度，换算出任务包至市中心的距离  $x_{i1}$

则第 i 个任务包的价格：

$$Y_{\text{任务包}i} =$$

$$n_i(16.63388x_{i1} + 10.14391D_{\text{佛山}} + 13.32511D_{\text{广州}} + 4.857436x_2 + 16.55476D_{\text{东莞}})$$

## 5.3.2 打包后任务完成情况的评价

根据问题二的任务完成率公式，有

考虑任务包所在市，任务包离所在市市中心的距离，该市人均 GDP，任务包价格的条件下，第 i 个任务包的完成率：

$$G_i = \text{logit}(\pi(Y_i))$$

在确定范围内的  $\alpha$  个会员愿意去预订该任务包，满足限额  $q_k \geq$  第 i 个任务包的任务个数  $n_i$  的会员 k 才能选到任务包，因此任务包接受概率为：

$$P_{i1} = \text{Count}(q_k \geq n_i) / \alpha$$

其中能接受到任务包的会员数为  $\beta$ ， $\beta = \text{Count}(q_k \geq n_i)$

已被接受的任务包，考虑违约风险之后的完成概率为  $P_{i2}$

取能接受到任务包的  $\beta$  个会员的信誉的平均值  $\bar{c}$

$$P_{i2} = \text{Count}(\bar{c} > c_k) / 1877 \quad (k=1,2,\dots,\alpha) \quad (1877 \text{ 为总会员数})$$

第 i 个任务包的完成概率  $P_i = P_{i1} * G_i * P_{i2}$

$$N \text{ 个任务打包后的所有任务的完成概率 } P = \frac{\sum_{i=1}^N P_i * n_i}{\sum_{i=1}^N n_i}$$

## 5.4 问题四的求解

我们以任务包中任务的平均经纬度为中心，考虑方圆 5 公里之内的会员接受任务的可能性。

根据附件 4 的数据，用 matlab 聚合，新任务的聚合情况如下表

表 5.任务聚合情况

每个任务包 中任务个数	任务包个数	每个任务包 中任务个数	任务包个数	每个任务包 中任务个数	任务包个数
1	799	6	10	11	1
2	232	7	6	21	1
3	82	8	2	32	1
4	40	9	1	33	1
5	20	10	3	43	1

共 1200 个任务包，经过 matlab 计算，用问题三中的任务包形式下的完成概率模型得出  $P=39.75\%$

## § 6 模型评价

### 6.1 优势与不足

1.优势 a.通过线性定价，定价公式简单易操作，同时保证了高拟合度。考虑了定价的地域性，引入人均 GDP 解释变量。

b.用 Logistic 模型优化了价格，得到较好的价格随地理位置的层次分布。

c.考虑打包定价时，创新地考虑了违约的风险因素和任务包被接受的概率。

2.不足 a.没有考虑网络预订中的突发情况

b.没有考虑会员预订任务包时的时间差距带来的完成率的影响。

## 附录 1：参考文献：

[1]姜启源. 谢金星. 叶俊. 数学模型[M]. 北京：高等教育出版社, 1998.

[2]基于 MATLAB 的高等数学问题求解. 清华大学出版社

[3][http://www.jnr.ac.cn/article/2016/1000-3037-31-9-1440.html#outline\\_anchor\\_22](http://www.jnr.ac.cn/article/2016/1000-3037-31-9-1440.html#outline_anchor_22)

line\_anchor\_22

## 附录 2：源程序：

### 源程序 1：

```
pack152.m  unpack152.m  +
1  for i=1:1:1200
2      for j=1:1:1877
3          dc(i,j)=sin(tc(i,1))*sin(mem(j,1))*cos(tc(i,2)-mem(j,2))+cos(tc(i,1))*cos(mem(j,1));
4          d(i,j)=6371.004*3.1415927/180.*acos(dc(i,j)); % 计算第i个任务中心点与第j个会员的距离
5          if d(i,j)<=5 % 如果距离小于5, 则纳入范围
6              k(i)=k(i)+1; % 计数范围内的所有会员
7              if mem(j,3)>=tc(i,3) % 如果限额足够, 则叠加
8                  s(i)=s(i)+1; % 计数范围内限额足够的会员数
9                  q(i)=q(i)+mem(j,4); % 求和限额足够的会员的信誉
10             end
11         end
12     end
13     c(i)=q(i)/s(i); % 求出平均信誉
14     p(i)=s(i)/k(i); % 求出任务包能被接收的概率
15     for r=1:1:1877
16         if c(i)>mem(r,4) % 计算平均信誉所在的排名位置
17             co(i)=co(i)+1;
18         end
19     end
20     pc(i)=co(i)/1877; % 不会违约的概率
21 end
22
23
```

### 源程序 2：

```
d1=pdist(d1);
```

```
z1=linkage(d1);
```

```
cluster(z1,1200);
```