

队伍编号	201266
题号	A

## 无车承运人平台线路定价问题

### 摘 要

本文以无车承运人平台的角度，利用量化分析和相关性分析分析影响平台定价的相关指标，根据平台现有定价建立平台线路定价预测与评价模型，最后结合数据分析及建模结果提出平台定价改进方案。

针对问题一，根据附件 1 所提供数据，进行数据预处理。对处理后数据用进行 Pearson 相关系数筛选出总里程、运输时长、业务类型、计划发车时间、运输等级、需求紧急程度、需求类型 2、车辆吨位 8 个价格决定因素。建立因子分析模型，用因子分析模型对 8 个因素进行验证，并确认此 8 个因素存在联系，为平台指导价的定价指标。

针对问题二，根据问题一得出的 8 项指标，建立层次分析模型获取其权重，再将指标权重带入秩和比综合评价模型中，通过 Python 编程，结合各指标量化后的数据与线路价格，得出评价结果，将线路价格分为 3 个等级，1 级最好，2 级较好，3 级最差。最终，线路价格中评价较好及以上的结果占比 97.81%，超过 95%，故可知线路成交价定价较优。

针对问题三，建立三项调价策略。建立两个回归模型预测价格，第一个模型预测第一次报价，同时也为总成本定价；第二个模型预测线路成交价，即最后一次报价；取两次预测价格的平均值作为第二次报价。建立决策树模型对价格进行决策，判断司机是否会接单，若接单则拒绝下一次报价。通过将回归预测模型与决策树模型相结合的方式建立报价模型，并预测出附件 2 中的订单价格填入附件 3 中。针对附件 3 中的报价结果，用问题 2 的秩和比模型进行评价，发现模型报价结果中评价较好及以上的结果占比为 98.05%，超过上次评价，故预测结果较优。

针对问题四，针对以上分析结果，对无车承运人平台从制定合适的定价体系、建立合适的订单接受评测体系、与续签实际承运人签订具体合同、完善制度建设，提升技术水平四个方面提出意见。

关键词：无车承运人平台、价格评价、定价预测、Pearson 相关系数、因子分析、秩和比综合评价、回归预测、决策树、神经网络

# 目录

一、问题提出.....	1
1.1 背景.....	1
1.2 问题重述.....	1
二、基本假设.....	1
三、符号说明.....	1
四、问题分析.....	2
4.1 问题一的分析.....	2
4.2 问题二的分析.....	2
4.3 问题三的分析.....	2
4.4 问题四的分析.....	3
五、模型的建立与求解.....	3
5.1 问题一模型的建立与求解.....	3
5.1.1 数据预处理.....	3
5.1.2 问题一模型的建立与求解：Pearson 相关系数模型.....	4
5.1.2 问题一结果检验：因子分析模型.....	6
5.2 问题二模型的建立与求解.....	7
5.2.1 问题二模型的建立.....	7
5.2.2 问题二模型的求解.....	10
5.3 问题三模型的建立与求解.....	13
5.3.1 调价策略.....	13
5.3.2 报价模型的建立与求解.....	13
5.3.3 报价结果的评价.....	15
5.4 建议信.....	16
六、评价与改进.....	17
6.1 模型的评价.....	17
6.1.1 模型优点.....	17
6.1.2 模型缺点.....	18
6.2 模型的改进.....	18
七、参考文献.....	18
八、附录.....	19

## 一、问题提出

### 1.1 背景

随着社会经济的发展，货运需求呈现多元化、个性化的趋势，也使得货运市场对物流企业的要求越来越高，推动了“无车承运人”模式的发展。截至目前为止，“无车承运人”模式在我国仅仅处于初步发展阶段，它的运作模式还不成熟，并且在国内缺少可以借鉴的成功经验。所以，“无车承运人”模式的发展在很多方面还需要不断的完善<sup>[1]</sup>。

本文以无车承运人的视角，在平台定价的相关研究的基础上，通过数学建模的方法帮助某无车承运人平台解决定价问题。既有利于扩大无车承运平台的利润空间并提高平台的竞争优势，对物流平台的转型发展也具有重要的意义。

### 1.2 问题重述

(1) 通过定量分析的方法，研究影响无车承运人平台进行货运线路定价的主要因素有哪些，并说明理由。

(2) 通过建立数学模型，对已经成交货运线路历史交易数据中的定价进行评价。

(3) 建立关于线路定价的数学模型，给出附件 2 线路任务的三次报价以及总成本定价，并填充在表格中；给出调价策略；评价定价。

(4) 根据研究，给无车承运人平台写一封不超过一页的建议信。

## 二、基本假设

①假设承运人无放弃订单行为，每个承运人都能完成订单。

②假设每个订单都会被接单。

③假设定价仅受附件 1 所给因素影响。

④假设所有因缺失数据过多而筛选掉的数据对定价无影响。

⑤假设季节变化不影响运输。

⑥假设量化处理的数据无错误无丢失。

## 三、符号说明

符号	意义	符号	意义
$X_1$	总里程	$Y$	线路指导价格
$X_2$	运输时长	$Z$	线路成交价格

$X_3$	路编号	$x_{ij}, (i=1,2,3,\dots, j=1,2,3,\dots)$	指标 $X_i$ 的第 $j$ 行数据
$X_4$	业务类型	$y_j, (j=1,2,3,\dots)$	指标 $Y$ 的第 $j$ 行数据
$X_5$	标的展示策略	$w_i$	第 $i$ 层
$X_6$	运输等级	$R_{ij} (j=1,2,3,\dots)$	$X_i$ 的效益性指标
$X_7$	需求紧急程度	$RSR_i$	秩和比
$X_8$	需求类型 1	$f$	频数
$X_9$	需求类型 2	D	训练数据集
$X_{10}$	计划发车时间	B	特征
$X_{11}$	车辆长度	$g(D,B)$	信息增益
$X_{12}$	车辆吨位	$H(x)$	信息熵
$X_{13}$	是否续签	$H(X Y)$	条件熵

---

## 四、问题分析

### 4.1 问题一的分析

为进行量化分析，需先对表一的数据进行筛选，从运输、货主、承运人角度筛选可用于影响平台无人车定价的因素，再对筛选出来的因素进行量化处理和剔除无效数据处理。之后可选用 Pearson 相关系数分析，分析出与平台定价有关的指标。筛选出来的指标为确认其确实为有效指标，拟采用因子分析，确认各指标之间有联系，若有联系则筛选出的指标确为平台定价指标，若无联系，则需再次进行分析。

### 4.2 问题二的分析

本题可采用秩和比综合评价模型对成交定价进行评价，然而秩和比综合评价模型的评价需要各指标的权重，权重不可随意决定，故可采用层次分析法，得出各指标的权重。将权重带入计算 Profit 和 RSR，进行分档，根据数据的分档结果即可评价定价。

### 4.3 问题三的分析

本题可用线性回归模型对价格预测，预测结果可用决策树模型进行评价，判断预测结果是否可能被司机选中。

建立两个线性回归预测模型，第一个模型是根据运输对平台定价进行预测，预测结果既是第一次报价，也是总成本定价。第二个模型是综合运输、货主、承运人三方以及线路成交价进行预测。考虑现实因素，估价因一次比一次高，才会增加司机接单的可能性，故用成交价相比指导价调高的数据建立模型，因此，第二个模型的预测价格应当会高于第一个模型的预测价格。

为平台效益考虑，报价不应一次上调过多。第一次报价用第一个预测价格，若无司机选中，则未选中的订单第二次报价取第一个预测价格和第二个预测价格的平均值。第二次报价仍未选中的订单，第三次报价再选用第二个预测价格。

争对报价可用 MATLAB 进行决策树分析，判断是否有司机会选中该定单。

#### 4.4 问题四的分析

本题基于前三小题的结果，建立的数学模型和大量文献，对无车承运人平台提出一些建议。

### 五、模型的建立与求解

#### 5.1 问题一模型的建立与求解

##### 5.1.1 数据预处理

(1) 数据筛选：针对本题给出的附件 1 数据，从货主需求、运输本身、实际承运人三个角度考虑，筛选出以下下可用条件（表 5.1.1）。

对象	条件
运输本身	总里程、运输时长、路编号、业务类型、标的展示策略、运输等级
货主需求	需求紧急程度、需求类型 1、需求类型 2
实际承运人	计划发车时间、车辆长度、车辆吨位、是否续签
平台	线路指导价

表 5.1.1 数据筛选结 (ps: 运输时长=计划到达时间-计划靠车时间)

（2）量化处理：已知表中大部分数据为非数值型数据，需对此进行量化处理，对非数值型数据处理规则如下。

变量名称	属性名	量化表示
业务类型	速运	1
	重货	2
需求类型 1	普通	1
	区域发运	2
需求类型 2	计划	1
	临时	2
是否续签	已分拨续签	1
	续签 ECP 审批驳回或撤销	2
	非续签	3
	未知	4
需求紧急程度	常规订单	1
	紧急订单	2
	特急订单	3
标的展示策略	BDC	1
	DIR	2
运输等级	一级运输	1
	二级运输	2
	三级运输	3

表 5.1.2 非数值型变量处理规则

此外，对于时间的处理为：

$$\begin{aligned} \text{发车时间} &= \text{发车点小时数值} \\ \text{运输时长} &= \frac{\text{计划到达时间} - \text{计划停车时间}}{24h} \end{aligned}$$

（3）异常数据剔除：表格数据存在数据缺失现象，需将带有缺失数据即原表格显示值为“N”的数据所在行进行删除处理。

5.1.2 问题一模型的建立与求解：Pearson 相关系数模型

Pearson 相关系数适用于测量两变量的相关性。设线路指导价为 Y, 其他变量为  $X_i$ , 则各变量与线路指导价的总体相关系数为

$$\rho = \frac{\text{cov}(X_i,Y)}{\sqrt{\text{var}(X_i)}\sqrt{\text{var}(Y)}}$$

其中， $\text{cov}(X_i,Y)$  是两变量的协方差； $\text{var}(X_i)$ 、 $\text{var}(Y)$  是变量  $X_i$  和 Y 的方差。

然而，总体相关系数是未知的，故需要用样本相关系数来计算相关系数。设  $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$ ， $Y = (y_1, y_2, \dots, y_n)$  分别为来自  $X_i$  和  $Y$  的两个样本，则样本相关系数为

$$r = \frac{\sum_{j=1}^n (x_{ij} - \bar{x}_i)(y_j - \bar{y})}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_i)^2 \sum_{i=1}^n (y_j - \bar{y})^2}}$$

因使用的数据量足够大，故可用样本相关系数代替总体相关系数<sup>[7]</sup>。

用 SPSS 软件对进行相关系数分析得，各变量与线路指导价的相关系数如下（表 5.1.3）：

线路指导价格			
	Pearson 相关性	Sig. (双尾)	个案数
线路指导价格	1		15962
总里程	0.991**	0	15962
运输时长	0.974	0	15962
路编号	0.001	0.944	15962
业务类型	-0.215**	0	15962
发车时间	0.168**	0	15962
表的展示策略	0.014	0.074	15962
运输等级	-0.647**	0	15962
需求紧急程度	0.081**	0	15962
需求类型 1	0	0.986	15962
需求类型 2	0.128**	0	15962
车辆长度	0.765**	0	15962
车辆吨位	0.813**	0	15962
是否续签	-0.003	0	15962

表 5.1.3 相关系数分析结果（本表仅为与线路指导价的相关性）

经 Pearson 相关系数分析得，总里程、运输时长、业务类型、计划发车时间、运输等级、需求紧急程度、需求类型 2、车辆长度、车辆吨位的数据右边都有两颗星，表示在 0.01 的显著性水平之下，与指导价格是显著相关的。而路编号、表的展示策略、需求类型 1、是否续签的数据的右边没有星号，即表示与指导价格不相关。又因为车辆长度的数据与车辆吨位数据相对应，即对车辆长度的数据进行分析，就是对车辆吨位的数据进行分析，故舍弃车辆长度。

综上可得，与指导价格相关的指标为总里程、运输时长、业务类型、计划发车时间、运输等级、需求紧急程度、需求类型 2、车辆吨位。

5.1.2 问题一结果检验：因子分析模型

下面通过因子分析对筛选出的各指标之间是否有联系进行验证，若各指标之间有联系，则可以做因子分析，即 Person 相关系数得出的 8 个指标即为与指导价格相关的主要判断指标，若不能做因子分析，即表示筛选出的指标有误。

① KMO 和 Bartlett 球形度检验

因子分析的前提是变量具有较强的相关性，并不是所有的变量指标都适合进行因子分析，因此需要进行 KMO 和 Bartlett 检验。当 KMO 检验是否适合进行因子分析的标准如下：KMO>0.9 时，非常适合；0.8<KMO<0.9，适合；0.7<KMO<0.8，一般适合；0.5<KMO<0.7，勉强适合；KMO<0.5，不适合。Bartlett 的显著性水平>0.05 时,选取的变量不适合做因子分析,Bartlett 的显著型水平<0.05 时,说明原始变量之间存在相关性,比较适合作因子分析。

用 SPSS 分析得出检验结果如表 5.1.4 所示,由表 5.1.4 可知 KMO 系数为 0.626,且 Bartlett 的显著性为 0.000<0.5，适合做因子分析。

KMO 取样 适切性量数	巴特利球形度检验		
	近似卡方	自由度	显著性
0.626	103794.4	28	0

表 5.1.4 KMO 和 Bartlett 球形度检验

② 合适的公因子数量

利用统计软件 SPSS 中的主成分模块选项,计算得出数据如表 5.1.5 所示，选取了 3 个主成分,其值分别为 3.408、1.511 和 1.126，总方差解释为 75.555%，第 1 种成分的特征值解释总方差为 42.596%，第 2 种成分的特征值解释总方差为 18.890%，第 3 种成分的特征值解释总方差为 14.070%。

进一步对数据进行基于观测特征值的碎石检验,如图 5.1.1 所示。由特征值大于 1 的成分数为 3,同时在图形变化最大处之上的主成分也是 3,和上述分析结果一致<sup>[2]</sup>。

成分	初始特征值			提取载荷平方和		
	总计	方差百分比	累积%	总计	方差百分比	累积%
1	3.408	42.696	42.596	3.408	42.696	42.596
2	1.511	18.890	61.485	1.511	18.890	61.485
3	1.126	14.070	75.555	1.126	14.070	75.555



4	0.918	11.470	87.025
5	0.601	7.514	94.538
6	0.347	4.342	98.880
7	0.071	0.881	99.761
8	0.019	0.239	100.00

表 5.1.5 总方差解释

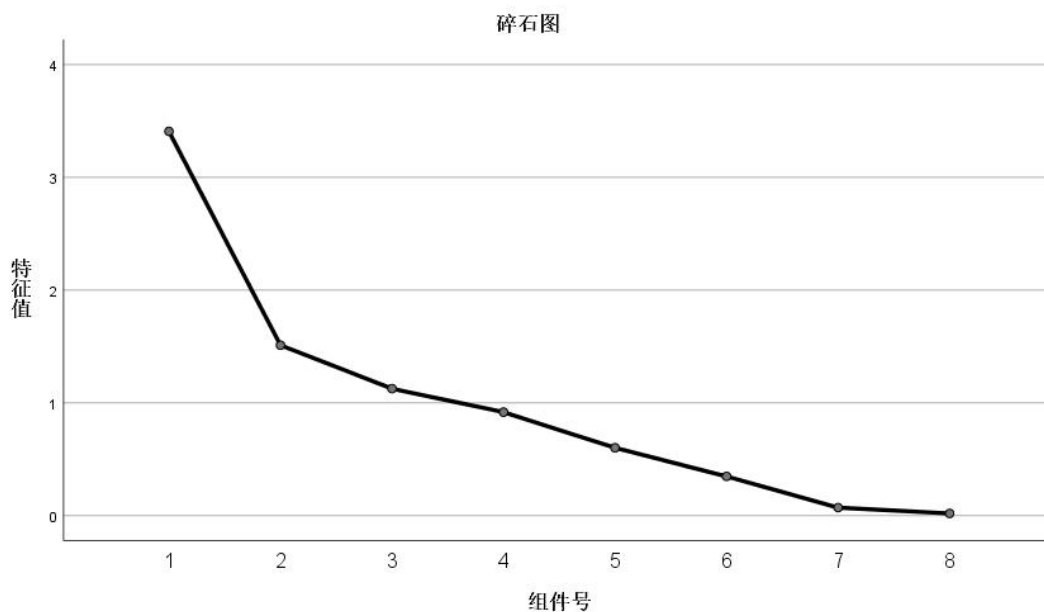


图 5.1.1 碎石图

综上，此八项指标不仅内部有相关性，并且可以提取出 3 个公因子。

由此可证，与指导价格相关的指标为总里程、运输时长、业务类型、计划发车时间、运输等级、需求紧急程度、需求类型 2、车辆吨位。

## 5.2 问题二模型的建立与求解

### 5.2.1 问题二模型的建立

#### (1) 获取权重：层次分析模型

##### ①指标体系选取

本模型以运输本身、货主需求、粮食企业为一级指标，将上题选取出的 8 个指标作为二级指标，对线路价格进行对比。其中，运输本身指标包括：总里程、运输时长、业务类型、运输等级；实际承运人指标包括计划发车时间、车辆吨位；货主需求指标包括需求紧急程度、需求类型 2<sup>[3]</sup>，具体如图 5.2.1 所示。

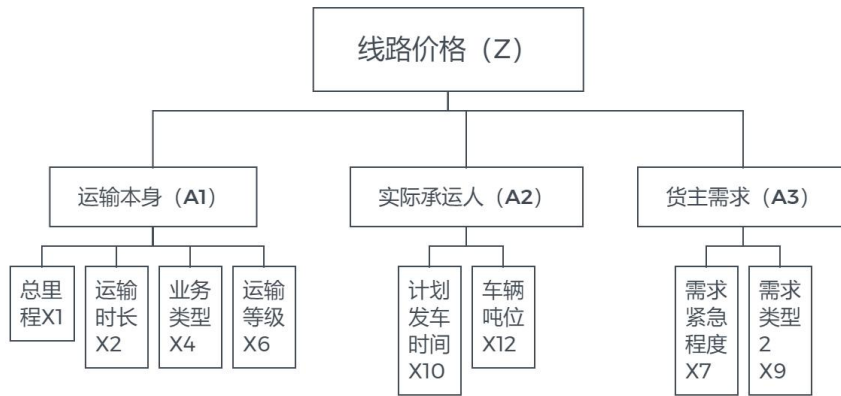


图 5.2.1 层次分析模型架构图

## ②建立判断矩阵与权重求解

Z	$A_1$	$A_2$	$A_3$	$\omega_0$
$A_1$	1	3	5	0.64
$A_2$	$\frac{1}{3}$	1	3	0.26
$A_3$	$\frac{1}{5}$	$\frac{1}{3}$	1	0.10

表 5.2.1 准则层  $A_i$  对目标层 Z 的判断矩阵

$A_1$	$X_1$	$X_2$	$X_6$	$X_4$	$\omega_1$
$X_1$	1	3	5	7	0.59
$X_2$	$\frac{1}{3}$	1	2	5	0.25
$X_6$	$\frac{1}{5}$	$\frac{1}{3}$	1	3	0.10
$X_4$	$\frac{1}{7}$	$\frac{1}{5}$	$\frac{1}{3}$	1	0.06

表 5.2.2 指标层  $X_i$  对准则层  $A_1$  的判断矩阵

$A_2$	$X_{10}$	$X_{12}$	$\omega_2$
$X_{10}$	1	2	0.67
$X_{12}$	$\frac{1}{2}$	1	0.33

表 5.2.3 指标层  $X_i$  对准则层  $A_2$  的判断矩阵

$A_3$	$X_7$	$X_9$	$\omega_3$
$X_7$	1	2	0.71
$X_9$	$\frac{1}{3}$	1	0.29

表 5.2.4 指标层  $X_i$  对准则层  $A_3$  的判断矩阵

③一致性检验

第一步：计算一致性指标 C.I.。

$$C.I.=\frac{\lambda_{\max}-n}{n-1}$$

第二步，确定相应的平均随机一致性指标 R.I.=0.537。

第三步，计算一致性比率 C.R.并进行判断。

$$C.R.=\frac{C.I.}{R.I.}=0.030$$

检验结果 C.R.<0.1，一致性检验通过，逻辑合理，可继续进行分析。

④组合权重求解

目标层	准则层	准则层权重	指标层	指标层权重	组合权重
线路价格 Z	运输本身 $A_1$	0.64	总里程 $X_1$	0.59	0.359
			运输时长 $X_2$	0.25	0.168
			业务类型 $X_6$	0.10	0.075
			运输等级 $X_4$	0.06	0.035
	实际承运人 $A_2$	0.26	计划发车时间 $X_{10}$	0.67	0.065
			车辆吨位 $X_{12}$	0.33	0.194
	货主需求 $A_3$	0.10	需求紧急程度 $X_7$	0.71	0.087
			需求类型 2 $X_9$	0.29	0.017

表 5.2.5 组合权重求解

(2) 评价：秩和比综合评价模型

通过层次分析模型计算出了各指标的权重，下面将指标权重带入秩和比模型，对线路成交价格进行评价，下面开始建立秩和比模型。

#### ① 指标说明

8 项指标全部为效益型指标，效益型指标计算公式如下：

$$R_{ij} = 1 + (n-1) \frac{x_{ij} - \min(x_{i1}, x_{i2}, \dots, x_{in})}{\max(x_{i1}, x_{i2}, \dots, x_{in}) - \min(x_{i1}, x_{i2}, \dots, x_{in})}$$

#### ② 计算秩和比

在整次秩和比法中，只考虑元素的相对大小，不考虑具体值，计算秩和：

$$RSR_i = \frac{1}{n} \sum_{j=1}^m w_j R_{ij}$$

其中  $w_i$  为第  $i$  个评价指标的权重， $\sum_{i=1}^m w_i = 1$

#### ③ 确定 RSR 的分布(转化为概率单位)

RSR 的分布是指用概率单位 Probit 表达的值特定的累计频率。Probit 模型是一种广义的线性模型，服从正态分布。其转换方法为：

Step 1 编制 RSR 频数分布表，列出各组频数  $f$ ，计算各组累计频数  $\sum f$ ；

Step 2 确定各组 RSR 的秩次范围及平均秩次；

Step 3 计算累计频率  $\frac{\bar{R}}{n} \times 100\%$ ，最后一项记为  $1 - \frac{1}{4n}$  进行修正；

Step 4 将累计频率换算为概率单位 Probit，Probit 为累计频率对应的标准正态离差  $u$  加 5。

④ 以累积频率所对应的概率单位  $Probit_i$  为自变量，以 RSR 值为因变量，计算直线回归方程，即：

$$RSR = a + b \times Probit$$

回归方程检验:对该回归方程,需要进行回归系数  $b$  的  $t$  检验法及拟合优度的 Pearson 相关系数检验

⑤ 通过检验回归方程输出 RSR 矫正值，并进行分档排序：

按照回归方程推算所对应的 RSR 估计值对评价对象进行分档排序<sup>[4]</sup>。

### 5.2.2 问题二模型的求解

用 Python 对 8 个指标进行分析，得出结果如下。

# 1. 确定 RSR 的分布

总里程	R1: 里程	R2: 运输时长	R3: 业务类型	R4: 运输等级	R5: 发车时间	R6: 吨位	R7: 紧急程度	R8: 需 2	RSR
238	7	12	1	2	1	2	1	1	0.000337485
1010	27	74	2	2	9	4	2	2	0.001548283
1100	27	76	2	2	7	4	3	2	0.001565635
1100	27	76	2	2	7	4	3	2	0.001565635
1100	27	76	2	2	7	4	3	2	0.001565635
1000	27	75	2	2	2	4	3	2	0.001534245
620	27	75	2	2	2	4	3	2	0.001534245
1000	27	75	2	2	2	4	3	2	0.001534245
860	27	76	2	2	7	4	3	2	0.001565635
238	7	12	1	2	1	2	1	1	0.000337485
860	27	76	2	2	7	4	3	2	0.001565635
770	27	76	2	2	7	4	3	2	0.001565635
238	7	12	1	2	1	2	1	1	0.000337485
1000	27	162	2	2	7	4	1	2	0.002497178
1713	30	166	2	2	2	4	3	2	0.00259778
238	7	12	1	2	1	2	1	1	0.000337485
259	7	21	1	2	15	2	3	1	0.000500475
259	7	15	1	2	16	2	3	1	0.000439061
259	7	25	1	2	16	2	3	1	0.000548242

表 5.2.6 部分综合评价结果的各指标的秩和 RSR 的值

RSR	f	Σ f	$\bar{\{R\}} f$	$\bar{\{R\}}/n*100\%$	Probit
0.0002	7	7	4	0.000251	1.519881
0.00021	83	90	49	0.00307	2.259768
0.00021	82	172	131.5	0.008238	2.601817
0.00022	1	173	173	0.010838	2.70401
0.00022	1	174	174	0.010901	2.706196
0.00022	1	175	175	0.010964	2.708371
0.00024	6	181	178.5	0.011183	2.715899
0.00025	85	266	224	0.014033	2.803647
0.00025	83	349	308	0.019296	2.931485
0.00026	84	433	391.5	0.024527	3.031878
0.00026	2	435	434.5	0.027221	3.076695
0.00026	7	442	439	0.027503	3.081168
0.00027	84	526	484.5	0.030353	3.124374
0.00028	77	603	565	0.035397	3.193198
0.00028	1	604	604	0.03784	3.223677
0.00028	2	606	605.5	0.037934	3.224817
0.00028	83	689	648	0.040596	3.256193

表 5.2.7 部分 RSR 值分布表

（注：表 5.2.6 与表 5.2.7 仅为结果展示用，因原数据过大无法完整截取，故此二表可用信息较少，具体结果见附件“RSR 分析结果报告.xlsx”）

2. 计算回归方程

```

                                OLS Regression Results
=====
Dep. Variable:                  y      R-squared:                  0.939
Model:                        OLS      Adj. R-squared:             0.939
Method:                    Least Squares  F-statistic:                1.194e+04
Date:                Sat, 23 May 2020  Prob (F-statistic):            0.00
Time:                19:02:11      Log-Likelihood:              5134.9
No. Observations:          778      AIC:                      -1.027e+04
Df Residuals:              776      BIC:                      -1.026e+04
Df Model:                  1
Covariance Type:            nonrobust
=====
                                coef      std err          t      P>|t|      [0.025      0.975]
-----
const          -0.0043      6.33e-05     -67.992      0.000      -0.004      -0.004
Probit          0.0012      1.09e-05     109.255      0.000      0.001      0.001
=====
Omnibus:                318.342      Durbin-Watson:            0.019
Prob(Omnibus):            0.000      Jarque-Bera (JB):         2580.631
Skew:                    1.636      Prob(JB):                 0.00
Kurtosis:                11.301      Cond. No.                 32.1
=====

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
回归直线方程为 : y = 0.00118951775187599 Probit - 0.004304765231914681

```

图 5.2.2 回归方程

t 检验统计量较大，说明模型的回归系数具有统计学意义；拟合优度达到了 93.9%，拟合效果较好，通过回归检验。

3. 归档与排序

等级	Probit	RSR
1	≥ 6.000377	≥ 0.002778
2	2.931485 ~ 6.000377	0.000253 ~ 0.002778
3	≥ 2.931485	≤ 0.000253

表 5.2.8 分档排序

其中等级 1 代表评价最好，等级 2 代表评价较好，等级 3 代表评价最差。研究分析结果报告的 15962 条可用数据可知：

等级	数据条数	占比
1	2535	15.88%
2	13078	81.93%
3	349	2.19%

表 5.2.9 分档结果

综上，成交定价中评价较好及以上的结果占比 97.81%，超过 95%，定价较优。

## 5.3 问题三模型的建立与求解

### 5.3.1 调价策略

①第二、第三次的报价只会比前面高不会比前面低。司机接单的目的是为了赚钱因为如果第一次报价给的价格不愿意接受，那么第二第三次的价格再降低，司机更不会接受。

②报价大于等于总成本定价。若低于总成本定价，会减少司机的接单率

③报价金额一次性变化不要太大，需考虑平台的受益。

### 5.3.2 报价模型的建立与求解

#### (1) 回归预测模型的建立

##### 1. 第一次价格预测

针对运输的 4 个指标以及线路指导价对所有可用数据用 SPSS 进行回归分析可得回归预测模型的系数（表 5.3.1）：

模型	未标准化系数		标准化系数		t	显著性	B 的 95%置信区间	
	B	标准错误	Beta				上限	下限
(常量)	-20.995	19.233			-1.902	0.275	-58.694	16.704
总里程	4.192	0.25	0.908		167.512	0	4.143	4.242
因变量：指导价格 运输时长	533.779	32.186	0.087		16.584	0	470.689	596.868
业务类型	56.077	7.135	0.009		7.859	0	42.091	70.064
运输等级	-3.542	8.109	-0.0011		-0.437	0.666	-19.436	12.352

表 5.3.1 第一次回归分析

故回归模型为：

$$y_j = -20.995 + 4.192x_{1j} + 533.779x_{2j} + 56.077x_{4j} - 3.542x_{6j},$$

##### 2. 第二次价格预测

针对所有的 8 个指标以及线路成交价对所有调价后价格调高的数据用 SPSS 进行回归分析可得回归预测模型的系数（表 5.3.2）：

模型	未标准化系数		标准化系数	t	显著性	B 的 95%置信区间	
	B	标准错误	Beta			上限	下限
（常量）	-356.328	285.752		-1.247	0.212	-916.478	203.821
总里程	6.13	0.059	1.006	104.724	0	6.015	6.245
运输时长	144.331	58.857	0.018	2.452	0.014	28.957	259.706
业务类型	55.763	20.864	0.005	2.672	0.008	14.854	96.653
运输等级	-8.327	1.022	-0.01	-8.149	0	-10.33	-6.324
发车时间	-0.524	10.902	-0.001	-0.048	0.962	-21.894	20.845
车辆吨位	194.445	73.229	0.026	2.655	0.008	50.897	337.993
需求紧急程度	25.648	9.91	0.003	2.588	0.01	6.221	45.074
需求类型 2	-39.932	18.629	-0.002	-2.144	0.032	-76.45	-3.415

表 5.3.2 第二次回归分析

回归模型为：

$$y_j = -356.328 + 6.13x_{1j} + 14.331x_{2j} + 55.753x_{4j} + 194.443x_{6j} - 8.327x_{10j} - 0.524x_{12j} + 25.648x_{7j} - 39.932x_{8j}$$

## （2）决策树模型的建立

①对当前样本集合，计算所有属性的信息增益

以线路指导价是否被接受做为特征 B，线路指导价格和 8 项指标做训练数据集 D，特征 B 对训练数据集 D 的信息增益  $g(D, B)$  为：

$$g(D, B) = H(D) - H(D|B)$$

信息熵： $H(x) = -\sum_{i=1}^n p(x_i) \log(p(x_i))$ ，其中  $p(x_i)$  代表随机事件  $x_i$  的概率。

条件熵： $H(X|Y) = H(X, Y) - H(x) = -\sum_{x,y} p(x, y) \log p(y|x)$

②选中信息增益最大的属性作为测试属性，把测试属性取值相同的样本划为同一个子样本集；若子样本集的分类属性只含有单个属性，则分支为叶子结点；否则对子样本集递归调用算法。

用 MATLAB 编程，将数据带入，得到的决策树如下图所示（详细见附件 Tree.fig）。



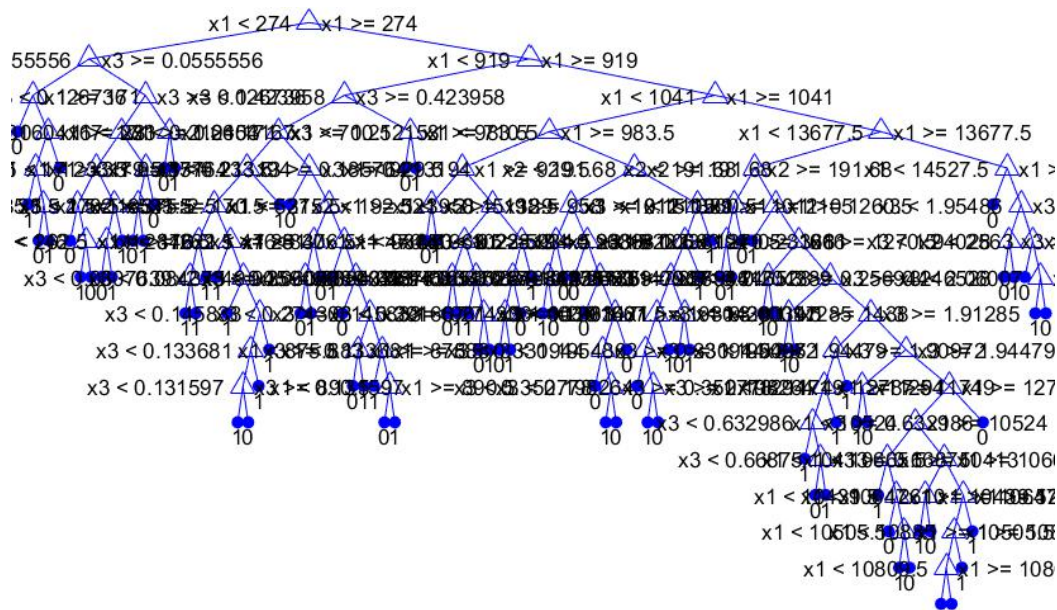


图 5.3.1 决策树模型

### (3) 报价模型的建立与求解

结合以上两个模型，建立报价模型。

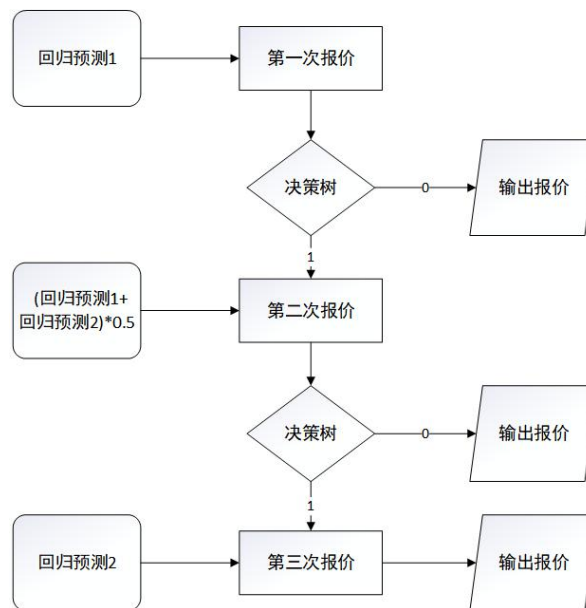


图 5.3.2 报价模型

将附件 2 数据带入模型，得到报价结果于附件“A201266 计算结果.xlsx”中。

### 5.3.3 报价结果的评价

用问题二的秩和比模型对报价结果进行评价，

```

OLS Regression Results

=====
Dep. Variable:          y      R-squared:          0.893
Model:                  OLS    Adj. R-squared:       0.892
Method:                 Least Squares    F-statistic:       1977.
Date:                   Sun, 24 May 2020    Prob (F-statistic): 5.57e-117
Time:                   17:32:21    Log-Likelihood:    1133.5
No. Observations:       239    AIC:               -2263.
Df Residuals:           237    BIC:               -2256.
Df Model:                1
Covariance Type:        nonrobust
=====

               coef      std err          t      P>|t|      [0.025      0.975]
-----
const         -0.0275      0.001    -30.443      0.000     -0.029     -0.026
Probit         0.0069      0.000     44.462      0.000      0.007      0.007
=====

Omnibus:                27.857    Durbin-Watson:          0.031
Prob(Omnibus):           0.000    Jarque-Bera (JB):        89.530
Skew:                    0.413    Prob(JB):                3.62e-20
Kurtosis:                5.882    Cond. No.                39.5
=====

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

回归直线方程为 : y = 0.006890545448915475 Probit - 0.02746952633939387

```

图 5.3.3 报价的回归方程

t 检验统计量较大，说明模型的回归系数具有统计学意义；拟合度为 89.3%，略低于上次评价，经分析，可能是因为我们加上了后一次报价要比前一次高的条件，故此拟合结果较合理。拟合效果较好，通过回归检验。

研究分析结果报告的 1489 条可用数据可知：

等级	数据条数	占比
1	233	15.65%
2	1227	82.40%
3	29	1.95%

表 5.3.3 分档结果

具体结果见附件“报价分析结果报告.xlsx”

综上，模型报价结果中评价较好及以上的结果占比 98.05%，超过上次评价，故报价结果较好。

## 5.4 建议信

(1) 制定合适的定价体系

首次报价考虑多重因素对价格的影响，如天气、路况等具体影响因素，保证首次报价能被大多数实际承运人接受。首次报价未被接受进行第二第三次报价时，考虑实际承运人方的因素，如预计发车时间对实际承运人接受价格的影响等。第二次报价还未被接受时，应适当调整价格，使得报价在允许范围尽可能被接受以确保每个订单能成功运输到达。

听取实际承运人的意见，根据他们在运输中的实际情况进行定价的更改使定价更好地让实际承运人接受。

## (2) 建立合适的订单接受评测体系

在首次定价发布之后，通过合适的体系来判断是否有实际承运人接受改报价的订单，并且通过以往订单的接受时间确定该订单该报价的接受时间，在一定时间无实际承运人接受订单的情况下，考虑进行第二次报价的发布来节约订单被接受的时间。

## (3) 与续签实际承运人签订具体合同

考虑与已经续签的实际承运人签订详细的合同，在遇到某订单无实际承运人接受的情况可将该单交付与本平台签约的实际承运人，减少报价过高以及订单无实际承运人接受导致的平台亏损的情况。

## (4) 完善制度建设，提升技术水平

目前，我国对于无车承运人的管理正处于初步阶段，在相关制度方面还不够健全和完善，同时也尚未出台专门的法律法规对其进行规范。因此，相关企业应提升对该问题的重视程度。相关部门之间应该加强协调与配合，包括单据、车型、数据等的标准化和统一化。在技术层面上，对相关数据进行实时化采集，同时还要进行全程跟踪，对数据进行全面收集和深入分析。通过各数据分析技术，能够推动着无车承运人逐渐开始向高质量、高效率以及低成本的方向发展。在技术的帮助之下，可有效降低在重复性上的精力和时间投入，最大限度的提升管理效率和工作效率，对无车承运人运营进行更大范围的管控和监督<sup>[5][6]</sup>。

# 六、评价与改进

## 6.1 模型的评价

### 6.1.1 模型优点

(1) Pearson 相关系数模型与因子模型将结合，可以很好的分析相关性并进行检验，减少相关性分析错误的可能性。

- (2) 层次模型很好的为秩和比模型提供了分析用的权重,使分析结果更具有合理性。
- (3) 回归预测模型表明自变量了和因变量之间的显著关系。
- (4) 决策树模型挖掘出来的分类规则准确性高, 便于理解。
- (5) 秩和比综合评价模型对指标的选择无特殊要求, 适用于各种评价对象, 由于计算时使用的数值是秩次, 可以消除异常值的干扰。

### 6.1.2 模型缺点

- (1) 使用层次分析法, 初始权重赋予较为主观。
- (2) 回归模型比较简单, 算法相对低级。
- (3) 决策树模型容易过拟合, 容易忽略属性之间的相关性。
- (4) Pearson 不适用于文本的相似性分析, 需对数据进行量化处理。
- (5) 指标转化为秩次时会失去一些原始数据的信息。当 RSR 值实际上不满足正态分布时, 分档归类的结果与实际情况会有偏差。

## 6.2 模型的改进

- (1) 层次分析模型的权重赋值较为主观, 可更换可得出更客观权重的模型。
- (2) 扩大指标选择范围, 如考虑季节变化, 天气影响, 政策因素的影响, 可得到更为全面的预测指标。

## 七、参考文献

- [1] 张秋月. “无车承运人”模式下承运平台的服务定价研究[D].西南交通大学:10-11,2019.
- [2] 杨怡,邓国和.我国农业大省的农业收入影响因素分析与建模[J].湖南文理学院学报(自然科学版),2019,31(04):1-6.
- [3] 李宸,樊重俊.基于层次分析法的粮食最低价格收购政策评价模型[J].改革与开放,2017(21):29-30+50.
- [4] Suranyi,RSR(秩和比综合评价法)介绍及 python3 实现,  
<https://zhuanlan.zhihu.com/p/38209882>, 2020/5/23
- [5] 丁华.无车承运人运营模式创新及发展路径[J].经济师,2018(05):245+247.
- [6] 贺登才. 无车承运人新业态全面推广的“行动指南”[N]. 现代物流报,2019-09-11(A01).

[7] 凌启东,张红伟,朱瑾,王建.基于 Pearson 相关系数的城市雾霾成因研究[J].海峡科技与产业,2019(12):20-22.

## 八、附录

### 1.秩和比模型 Python 代码

```
import pandas as pd
import statsmodels.api as sm
import xlrd
from scipy.stats import norm

def rsr(data, weight=None, threshold=None, full_rank=True):
    Result = pd.DataFrame()
    n, m = data.shape

    # 对原始数据编秩
    if full_rank:
        for i, X in enumerate(data.columns):
            Result[f'X{str(i + 1)}: {X}'] = data.iloc[:, i]
            Result[f'R{str(i + 1)}: {X}'] = data.iloc[:, i].rank(method="dense")
    else:
        for i, X in enumerate(data.columns):
            Result[f'X{str(i + 1)}: {X}'] = data.iloc[:, i]
            Result[f'R{str(i + 1)}: {X}'] = 1 + (n - 1) * (data.iloc[:, i].max() - data.iloc[:, i]) / (data.iloc[:, i].max() - data.iloc[:, i].min())

    # 计算秩和比
    weight = 1 / m if weight is None else pd.np.array(weight) / sum(weight)
    Result['RSR'] = (Result.iloc[:, 1::2] * weight).sum(axis=1) / n
    Result['RSR_Rank'] = Result['RSR'].rank(ascending=False)

    # 绘制 RSR 分布表
    RSR = Result['RSR']
    RSR_RANK_DICT = dict(zip(RSR.values, RSR.rank().values))
    Distribution = pd.DataFrame(index=sorted(RSR.unique()))
    Distribution['f'] = RSR.value_counts().sort_index()
    Distribution['Σ f'] = Distribution['f'].cumsum()
    Distribution[r'\bar{R} f'] = [RSR_RANK_DICT[i] for i in Distribution.index]
    Distribution[r'\bar{R}/n*100%'] = Distribution[r'\bar{R} f'] / n
    Distribution.iat[-1, -1] = 1 - 1 / (4 * n)
    Distribution['Probit'] = 5 - norm.isf(Distribution.iloc[:, -1])

    # 计算回归方差并进行回归分析
    r0 = pd.np.polyfit(Distribution['Probit'], Distribution.index, deg=1)
    print(sm.OLS(Distribution.index,
```

```

sm.add_constant(Distribution['Probit'])).fit().summary()
    if r0[1] > 0:
        print(f"\n 回归直线方程为: y = {r0[0]} Probit + {r0[1]}")
    else:
        print(f"\n 回归直线方程为: y = {r0[0]} Probit - {abs(r0[1])}")

    # 代入回归方程并分档排序
    Result['Probit'] = Result['RSR'].apply(lambda item: Distribution.at[item, 'Probit'])
    Result['RSR Regression'] = pd.np.polyval(r0, Result['Probit'])
    threshold = pd.np.polyval(r0, [2, 4, 6, 8]) if threshold is None else pd.np.polyval(r0,
threshold)
    Result['Level'] = pd.cut(Result['RSR Regression'], threshold, labels=range(len(threshold)
- 1, 0, -1))

    return Result, Distribution

```

```

def rsrAnalysis(data, file_name=None, **kwargs):
    Result, Distribution = rsr(data, **kwargs)
    file_name = 'RSR 分析结果报告.xlsx' if file_name is None else file_name + '.xlsx'
    Excel_Writer = pd.ExcelWriter(file_name)
    Result.to_excel(Excel_Writer, '综合评价结果')
    Result.sort_values(by='Level', ascending=False).to_excel(Excel_Writer, '分档排序结果
')
    Distribution.to_excel(Excel_Writer, 'RSR 分布表')
    Excel_Writer.save()

    return Result, Distribution

```

```

def excel_one_line_to_list(i):
    df = pd.read_excel("rsr.xlsx", usecols=[i],
                        names=None) # 读取项目名称列,不要列名
    df_li = df.values.tolist()
    result = []
    for s_li in df_li:
        result.append(s_li[0])
    return result

```

```

if __name__ == '__main__':
    name = ['里程', '运输时长', '业务类型', '运输等级', '发车时间', '吨位', '紧急程度', '需
2']

```

```

j = 1
excddata = {}
indexput = list
for i in name:
    indexput = excel_one_line_to_list(0)
    result = excel_one_line_to_list(j)
    excdata.update({i:result}) # 读所有行的 title 以及 data 列的值，这里需要嵌套
列表
    j = j+1
#print(excddata)
data = pd.DataFrame(excddata,index = indexput,
                    columns=['里程', '运输时长', '业务类型', '运输等级', '发车时
间', '吨位', '紧急程度', '需 2'])
rsrAnalysis(data,weight=[0.359,0.168,0.073,0.035,0.063,0.192,0.057,0.017],threshold =
[1,3,6,10])

```

## 2.决策树 Matlab 代码

```

clear all
clc
warning off
load data.mat
a = randperm(17451);
Train = data(a(1:15962),:);
Test = data(a(15963:end),:);
P_train = Train(:,3:end);
T_train = Train(:,2);
P_test = Test(:,3:end);
T_test = Test(:,2);
ctree = ClassificationTree.fit(P_train,T_train);
view(ctree);
view(ctree,'mode','graph');
T_sim = predict(ctree,P_test);

```