

《统计建模》课程报告

**题目：牙膏模型再分析
旅客出行交通方式选择建模**

成员：3227042017 陈奕昕（组长）

3227042038 杨礼佳

3227022004 杨洁婧

3227042018 吴育雄

3227042015 黄荣鹏

应用统计学专业 2022 级

完成时间：2025 年 5 月 15 日

授课教师：陈征

目录

组员分工.....	4
1 摘要	5
2 牙膏销售问题再分析-多项式模型	8
2.1 增添影响牙膏销售的变量因素	8
2.1.1 增加“顾客商品存量”变量.....	8
2.1.2 增加优惠百分比变量.....	8
2.2 模型拟合	8
2.2.1 模型整体评估.....	11
2.2.2 顾客商品存量变量的分析.....	11
2.2.3 广告投入费用变量的分析.....	12
2.2.4 商品优惠百分比变量的分析.....	13
2.3 小结	13
3 牙膏销售问题再分析—微分方程模型	14
3.1 模型拟合	14
3.1.1 微分方程模型理论与模型构建.....	14
3.1.2 改进微分模型一.....	14
3.1.3 改进微分模型二.....	15
3.2 模型整体评估	15
3.2.1 广告投入费用变量的分析.....	17
3.2.2 销售周期变量的分析.....	18
3.2.3 灵敏度分析.....	18
3.3 小结	20
4 牙膏销售问题再分析总结	20
5 《高铁时代下旅客出行交通方式选择及现代运输业均衡发展研究》报告 ^[2] 的讨论分析.....	21
5.1 原报告不足之处	21
5.1.1 宏观与微观分析失衡.....	21
5.1.2 运输方式选择集覆盖不完整.....	21
5.1.3 数据的量化标准及挖掘深度和广度不足.....	22
5.1.4 运输走廊的细化研究和详细对比的缺失.....	23
5.1.5 接驳交通模型考量的缺失.....	24
5.1.6 Logit 模型与效用理论的衔接与结果展现缺失.....	24
5.2 对于报告内容的改进补足.....	24
5.2.1 效用函数的影响因素.....	25
5.2.2 效用函数再计算.....	26
5.2.3 层次分析法.....	28
5.2.4 选择概率函数.....	30
5.3 个体数据再分析	31
5.3.1 调查问卷收集与 probit 回归模型建立.....	31

5.3.2	模型结果解读.....	33
5.4	实例建模小结	34
5.4.1	企业或交通优化建议	35
6	参考文献.....	36
7	附录	36

组员分工

陈奕昕（组长）：牙膏微分方程模型建模;高铁模型城市对数据获取;效用函数改进计算;以及高铁民航模型优化 PPT 制作和发表。

杨礼佳：编制牙膏微分模型结果图的 R 程序，牙膏微分方程模型报告撰写以及部分 PPT 制作和发表。

杨洁婧：查找文献，撰写高铁报告不足之处报告分析，整体报告撰写与修正，牙膏模型 PPT 制作和发表。

吴育雄：牙膏多项式模型的建模与分析，设计调查问卷，构建并撰写高铁民航模型的报告以及编制相应的 R 程序，以及个体数据建模分析 PPT 制作和发表。

黄荣鹏：构建牙膏多项式模型，撰写高铁报告总结，分析实例建模以及高铁民航原模型分析 PPT 制作和发表。

1 摘要

牙膏模型再分析

摘要

目的 本研究针对牙膏销量预测问题，优化原有多项式模型并构建微分方程模型。**方法** 多项式模型引入“顾客商品存量”等变量，微分方程模型基于阻滞增长理论，引入“优惠百分比”变量并变形。**结果** 优化后的多项式模型 R^2 为 0.9007，各微分方程模型的 R^2 平均为 0.879，较原模型拟合效果略差，但捕捉到销量随广告投入先增后趋饱和的规律。**结论** 优化后的多项式模型发现存量与销量呈倒 U 型关系，广告投入、优惠百分比对销量有正向驱动；微分方程模型更符合销售量趋于市场饱和值的现象。

关键词 牙膏销售量预测；顾客商品存量；阻滞增长模型；微分方程；灵敏度分析

Reanalysis of the toothpaste model

Abstract

Objective Aiming at the problem of toothpaste sales volume prediction, in this study, the original polynomial model was optimized, and the differential equation model was also used for reanalysis. **Methods** In the polynomial model, variables such as "customer commodity stock" were introduced. The differential equation model was based on the theory of logistic growth, and was transformed after introducing the variable of "percentage of discount". **Results** The goodness of fit of the optimized polynomial model was 0.9007, which was better than that of the original model. The goodness of fit of each differential equation model was 0.872, 0.881 and 0.885 respectively. Although the fitting effect was slightly worse than that of the original model, it was more in line with the phenomenon that the sales volume tends to reach the market saturation value. **Conclusion** The optimized polynomial model reveals that there is an inverted U-shaped relationship between the stock and the sales volume. Advertising investment and the percentage of discount have a positive driving effect on the sales volume. After optimization, the differential equation model captures the law that the sales volume first increases with advertising investment and then tends to saturate. Both types of models provide certain theoretical support for the prediction of toothpaste sales volume and the formulation of strategies.

Keywords Toothpaste sales volume prediction; Customer commodity stock; Logistic growth model; Differential equation; Sensitivity analysis

旅客出行交通方式选择建模

摘要

目的 本研究针对《高铁时代下旅客出行交通方式选择及现代运输业均衡发展研究》报告的局限性，聚焦其理论模型衔接不足、数据利用广度欠缺及运输方式选择集覆盖不完整等问题，旨在优化旅客出行选择模型，为现代运输业均衡发展提供更精准的方法论支持。**方法** 通过扩展运输方式选择集，引入层次分析法量化效用函数权重系数，结合混合 Logit 模型与效用理论，解析旅客异质性偏好。同时，基于个体问卷调查数据建立 Probit 回归模型，从微观视角补充原报告的缺失。**结果** 研究发现短途运输方式及细分铁路类型对运输竞争格局影响显著；经济性、快速性等权重因旅客收入、出行目的而异；高铁在中短途具优势，长距离出行民航占主导；改进后的混合 Logit 模型显著提升预测精度。**结论** 研究揭示高铁与民航的竞合关系及客群细分策略，为运输走廊布局、票价调控及服务质量提升提供理论依据。建议高铁强化中短途竞争力，民航优化长航线服务，二者协同完善接驳系统。

关键词 高速铁路；民航；运量分担；层次分析法；BL 模型；效用函数

Passenger travel mode selection modeling

Abstract

Objective Aiming at the limitations of the report titled "Research on the Choice of Travel Modes by Passengers in the Era of High-Speed Rail and the Balanced Development of the Modern Transportation Industry", this study focuses on issues such as the insufficient connection of theoretical models, the lack of extensive data utilization, and the incomplete coverage of the choice set of transportation modes. It aims to optimize the passenger travel choice model and provide more accurate methodological support for the balanced development of the modern transportation industry. **Methods** By expanding the choice set of transportation modes, the Analytic Hierarchy Process is introduced to quantify the weight coefficients of the utility function. The Mixed Logit model is combined with utility theory to analyze the heterogeneous preferences of passengers. At the same time, based on the individual questionnaire survey data, a Probit regression model is established to supplement the lack of macro-analysis in the original report from a micro perspective. **Results** The study finds that short-distance transportation modes and the subdivided railway types have a significant impact on the transportation competition pattern; the weights of economic efficiency, rapidity, etc. vary according to passengers' income and travel purposes; high-speed rail has an advantage in medium and short distances, and civil aviation dominates for long-distance travel; the improved Mixed Logit model significantly improves the prediction accuracy. **Conclusion** The study reveals the competitive and cooperative relationship between high-speed rail and civil aviation and the passenger segmentation strategy, providing a theoretical basis for the layout of transportation

corridors, the regulation of ticket prices, and the improvement of service quality. It is recommended that high-speed rail strengthen its competitiveness in medium and short distances, civil aviation optimize the service of long-haul routes, and the two work together to improve the connection system.

Keywords High-speed railway; Civil aviation; Traffic volume sharing; Analytic Hierarchy Process; BL model; Utility function

2 牙膏销售问题再分析-多项式模型

2.1 增添影响牙膏销售的变量因素

2.1.1 增加“顾客商品存量”变量

首先，基于牙膏作为典型日常消耗品的消费特征——消费者购买后会经历固定使用周期，且短期内（2-3 个销售周期）一般不会重复购买，本研究引入“顾客商品存量”变量（定义为过去 3 个月累计销量），以量化市场饱和程度。

根据库存理论，消费者持有的产品存量直接影响其即时购买决策。当存量超过阈值时，整体市场需求因前期购买透支而趋于饱和，导致当期销量下降；反之，随着时间推移，当首期购买的牙膏接近消耗完毕（即存量降至较低水平）时，消费者的替换需求集中释放，推动销量回升。这种高存量抑制即时需求而低存量则刺激购买行为效应在消耗品行业中尤为显著，因此我们将“顾客商品存量”正式定义为反映市场供需动态的核心解释变量。

对于第 i 个销售周期，顾客存量=第 $i-1$ 个销售周期的销量+第 $i-2$ 个销售周期的销量+第 $i-3$ 个销售周期的销量。

2.1.2 增加优惠百分比变量

增加价格差与公司销售价格的交互项作为变量之一，即将本公司的销售的价格与其他公司的销售价格的差值除以本公司的价格，可以理解为优惠或者贵了多少百分比（价格差/公司销售价格）。在实际意义中反映了消费者心理对“绝对价格差”（优惠/贵了多少元）与“相对价格比”（优惠/贵了多少倍）存在感知差异。

2.2 模型拟合

基于原文的方法，将价格差作为变量带入模型拟合后，本研究通过划分固定价格差的不同区间进行可视化分析（图 1），探究固定价格差广告费用与销售量的关系。结果显示，不同的价格差所对应的广告费用的代表性并不是很高，数据中价格差与广告费用呈现一定正相关联，即当价格差较小时广告费的投入水平较低，价格差较大时广告费的投入相应增加，但价格差变量对模型的解释力较为有限，提示其在建模过程中代表性较弱。

进一步对固定价格比对广告费和销售量展开可视化分析（图 2），发现此时不同的价格比对应的广告费用分布区间较宽泛，变量代表性显著提升。同时通过对比价格差模型（X1）和价格比模型（X3）可以看出，将价格比替换价格差作为变量带入模型，*model3* 和 *model4* 的拟合效果无显著差异，进一步验证了价格比变量在模型中具备良好的解释效力。

$$\text{Model 1: } Y \sim x_1 + x_2$$

$$\text{Model 2: } Y \sim x_1 + x_2 + x_2^2$$

$$\text{Model 3: } Y \sim x_1 + x_2 + x_2^2 + x_1x_2$$

$$\text{Model 4: } Y \sim x_3 + x_2 + x_2^2 + x_3x_2$$

表 1 多项式模型结果表

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	27	1.5337				
2	26	1.2733	1	0.26041	6.1164	0.02054 *
3	25	1.0644	1	0.20887	4.9059	0.03610 *
4	25	1.0593	0	0.00506		

广告费用与销售量关系（颜色：价格差区间，大小：原始价格）

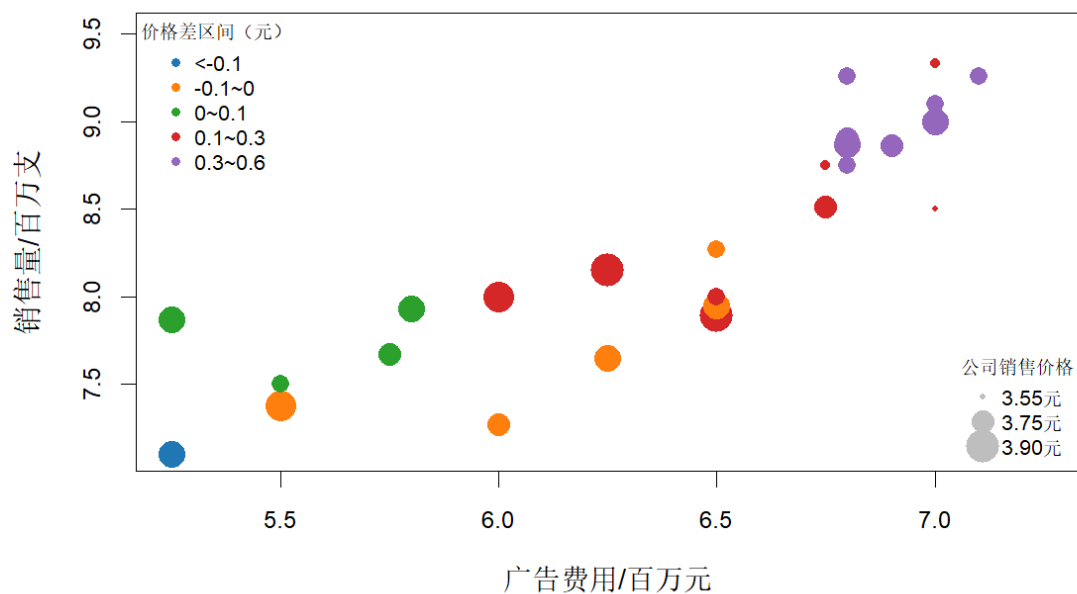


图 1 广告费用与销售量关系图（以价格差区分）

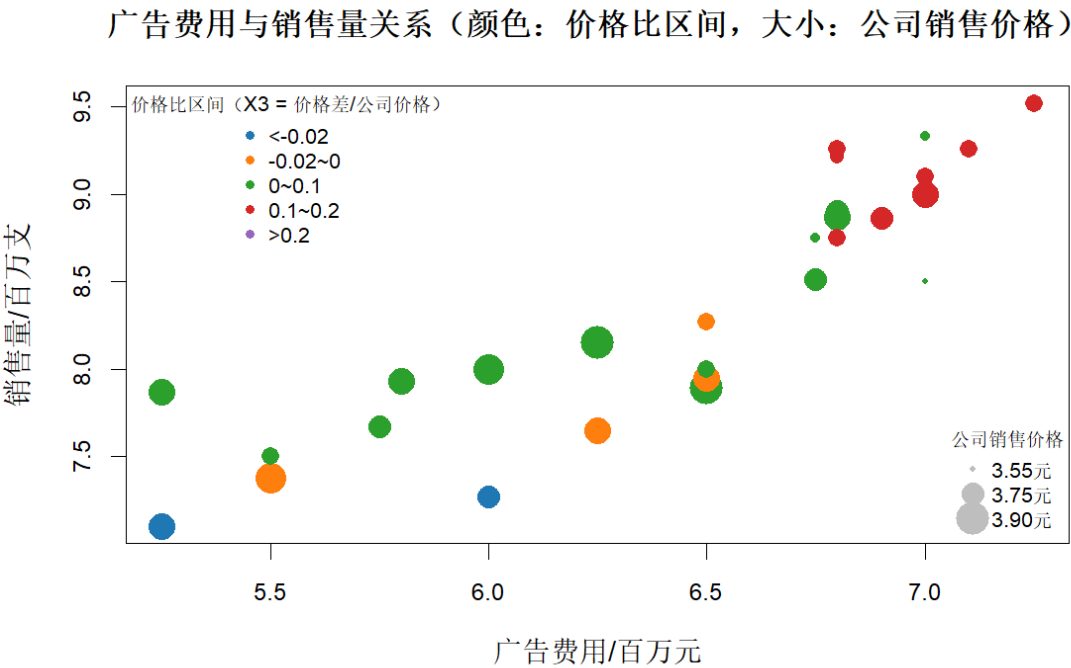


图 2 广告费用与销售量关系图（以价格比区分）

在变量体系中，价格差与公司销售价格、其他销售价格等变量存在较显著的信息重叠，导致变量间相关性较高。若盲目将高度相关的变量全部纳入模型进行分析，很可能引发模型的多重共线性问题。这会对模型性能产生多方面负面影响，如模型参数估计的稳定性被破坏，参数估计值的方差增大，标准误随之上升，导致参数显著性检验失效，在这种情况下，即使模型整体拟合效果较好，各解释变量对因变量的真实影响也难以准确识别，甚至可能还会出现符号异常或系数绝对值失真的情况，导致模型解释力下降。此外，高度相关的变量会放大模型对训练数据的拟合偏差，引发过拟合的现象，即模型在训练数据上呈现虚假的高拟合精度（如 R^2 较高），但在新数据预测预测时泛化能力显著下降，而模型对某些变量的微小变化呈现过度敏感性，会导致预测误差显著增大。这种缺陷不仅会降低结论的可靠性，还可能误导实际决策，增加商业判断失误的风险。

综合以上考量，最终选择纳入线性模型的自变量为：顾客存量，优惠百分比，广告投入，因变量为销售量，建立线性模型：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_2^2 + \beta_5 x_3 + \beta_6 x_3^2 + \varepsilon \quad (1)$$

其中， x_1 为顾客存量， x_2 为优惠百分比， x_3 为广告投入。

使用 R 语言进行分析结果如下：

表 2 线性模型分析结果

	Estimate	Std.Error	tvalue	Pr(> t)
(Intercept)	-20.31748	18.61797	-1.091	0.28812
poly(市场存量, 2, raw = TRUE)1	2.92204	1.36168	2.146	0.04433
poly(市场存量, 2, raw = TRUE)2	-0.0574	0.02722	-2.109	0.04777
poly(广告费用, 2, raw = TRUE)1	-3.52493	1.96349	-1.795	0.08774
poly(广告费用, 2, raw = TRUE)2	0.33564	0.15998	2.098	0.04881
poly(优惠率 2, raw = TRUE)1	8.41936	2.19181	3.841	0.00102
poly(优惠率 2, raw = TRUE)2	-27.57636	14.23319	-1.937	0.06694

2.2.1 模型整体评估

首先，在拟合优度方面，多元决定系数 R^2 为 0.9236，调整后决定系数 R^2 为 0.9007，表明模型解释了约 90.07% 的销售量方差，拟合效果比较不错。F 统计量为 40.32 ($p < 0.0001$)，说明模型整体高度显著，自变量联合对销售量具有强的解释能力。

残差分析结果显示，残差标准差（回归标准误）为 0.2062，基于 20 个残差自由度，反映预测误差整体处于较低水平。残差分布范围为 $[-0.276, 0.350]$ ，经检验未发现极端异常值。总体上看，模型在样本内拟合精度与残差规范性方面表现稳定，具有较强的预测可靠性。

2.2.2 顾客商品存量变量的分析

在模型参数估计结果中，一次项系数 ($\beta_1=2.922, p=0.044$) 呈现显著正相关，表明市场存量初期增加时，消费者因囤货需求可能会推高短期销量。二次项系数 ($\beta_2=-0.057, p=0.048$) 显著负相关，说明存量与销量间存在倒 U 型曲线关系：当存量处于较低水平时，销量随存量增加而上升，反映消费者需要囤积一定量的产品以满足日常需求；而当存量超过临界值后，消费者因库存积压和市场趋于饱和减少购买，导致销量下降。

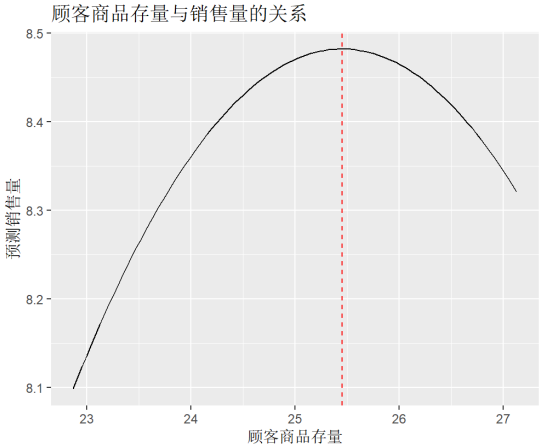


图 3 顾客商品存量与销售量的关系图

结合可视化分析（图 3），在低存量阶段，销量随存量增加而上升，与消费者储备日常用品的行为模式一致；在高存量阶段，因市场饱和及需求透支，消费者囤货后需求减少，销量随存量增加而下降，数据趋势与实际商业逻辑吻合。具体来看，在存量 23~25.5 区间内，销量从 8.1 线性增至 8.5 左右，销量持续上升，未出现下降趋势，表明该阶段仍处于倒 U 型曲线的

上升期，消费者购买意愿尚未触及拐点。从模型分析预测来看，顾客商品存量阈值约为 25.5，当顾客牙膏存量超过这一水平时，出于避免浪费的理性决策，购买意愿显著降低，进而导致销量下降。这符合需求理论中“库存调节效应”，也与市场供需动态变化规律一致。

2.2.3 广告投入费用变量的分析

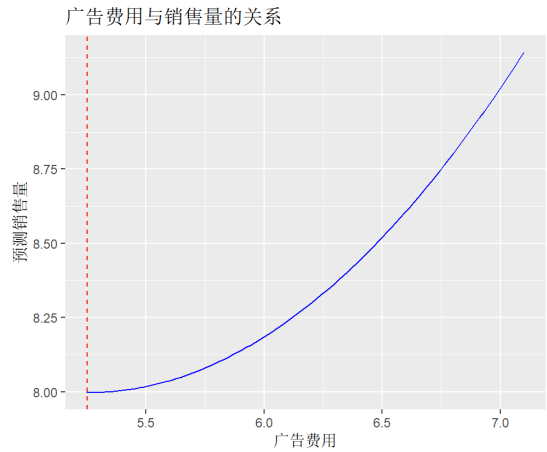


图 4 广告费用与销售量的关系图

结合广告费用与销售量关系图及模型估计结果，具体分析如下：广告费用的线性项系数显著为正（ $p=0.048$ ），与图中广告投入增加伴随预测销量上升的趋势一致，表明广告支出对销量具有明确的正向驱动作用，这既符合基本商业逻辑，也与现实中广告拉动消费的直观认知一致。然而，模型的二次项系数不显著或呈边缘显著（ $p=0.088$ ），这提示在当前模型设定下，尚未充分捕捉到广告费用与销售量的非线性关系，如可能存在边际效益递减等复杂关系。这可能意味着在广告费用的特定区间内，增加广告投入带来的销售增长可能趋于平缓甚至停滞，但现有模型结构未能有效反映这一特征，构成模型构建的潜在局限。因此模型需要进一步调整，以更好地反映广告费用与销售量之间的潜在非线性关系。

尽管存在上述不足，模型结果仍为理解广告投入的销售效应提供了关键依据，即显著的线性关系证明基础广告预算对销量的拉动作用稳定存在，有助于公司在制定广告策略时做出更加科学的决策。

此外，针对模型局限性，可以通过简化模型（如移除不显著的二次项）或交叉验证方法进一步检验泛化能力，避免因非必要变量引入导致的估计偏差。另外，结合业务场景特征，探索更贴合实际的非线性形式（如取对数变换等），优化广告费用区间，以提高模型变量的解释能力。

2.2.4 商品优惠百分比变量的分析

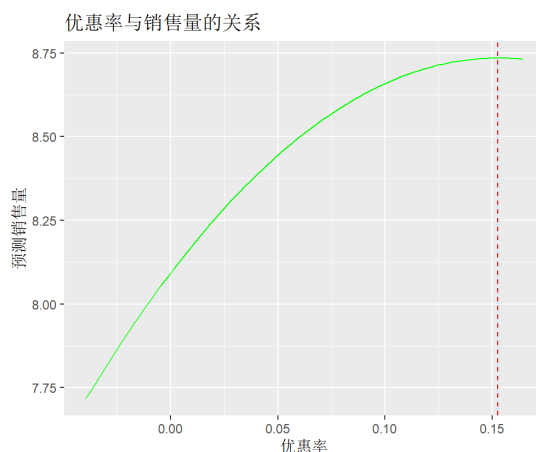


图 5 优惠率与销售量的关系

结合优惠率与销售量的模型估计及可视化分析，优惠率一次项系数高度显著 ($p=0.001$)，且当优惠率从 0% 增至 15% 时，预测销售量从 7.75 稳步增至 8.00，符合“提高优惠率直接刺激销量”的商业逻辑，表明在现有数据范围内，优惠率对销量的正向驱动效应明确且稳定。二次项边际递减趋势，二次项系数为负值 ($p=0.067$)，处于显著性边界，提示优惠率可能存在边际效应递减，即随着优惠力度超过某一临界值，销量提升幅度可能趋于平缓甚至反转。但当前可视化结果未呈现显著的拐点特征，这可能与消费者牙膏存量的调节作用相关，也可能受限于数据覆盖范围，但此结果不一定可靠。

结合商品优惠百分比（优惠率）与销售量关系图及模型结果来看，模型合理性存在两方面潜在局限。其一，由于数据量有限，优惠率数据仅覆盖至 15%，接近模型计算的理论最优值，但未包含超过拐点后的样本（如 20% 以上优惠），数据范围可能不足以完整验证二次效应；其二，二次项 p 值为 0.067，接近显著性边界，模型中包含了不显著变量，可能夸大了优惠率的独立解释力。另外，受限于数据规模，无法通过扩展数据范围（如测试更高优惠率对于销量的影响）来验证模型分析的拐点后的销量变化，致使模型对非线性关系的解析能力存在边界。

尽管存在上述约束，模型的核心结论依然具有实践价值。显著的一次项效应表明，在当前数据范围内提升优惠率对销量的促进作用真实可靠，可为促销策略制定提供有效的信息与决策。后续研究可通过扩大数据样本（覆盖更广优惠区间）或引入交互项（如结合存量变量）进一步优化模型，以更精准地捕捉优惠策略的复杂效应。

2.3 小结

在此回归模型中，当部分高阶项（如二次项）未通过显著性检验但模型整体显著时，需在解释效力与预测精度间权衡，保留不显著变量虽可能维持模型复杂度，但也可能会增加过拟合风险，降低模型在新数据中的泛化能力，同时可能干扰核心变量的真实解释；而移除冗余变量虽能简化模型结构、提升可解释性并降低多重共线性风险，但可能因遗漏潜在非线性关系而导致预测精度轻微损失。

对此，建议采取以下策略：首先，基于商业销售行业理论的逻辑与实际情况，优先剔除既

无统计显著性又缺乏商业解释价值的变量（如意义较弱的不显著项），避免因无关变量引入导致的模型失真；对于一些具备实际意义的边缘显著变量，可暂时保留并通过交叉验证评估其对模型稳定性的实际贡献，防止因严格显著性标准误删重要变量。

针对同时存在显著一次项与二次项的变量（如优惠率），需结合市场实际场景校准模型结论，例如在制定优惠策略与广告策略时，同步考虑理论拐点与消费者心理与行为，确保策略既符合统计规律又具备落地可行性。

3 牙膏销售问题再分析—微分方程模型

从多项式模型的结构特征来看，其预测结果显示销售量 Y 随广告费用 X_2 的增加呈单调递增趋势，当 X_2 趋向无穷大时， Y 的理论预测值亦趋于无穷大。这一结论与市场实际运行规律相悖。现实中商品销售受市场容量限制，最终会达到市场饱和阈值，而非无限增长。

基于牙膏销售量随广告投入的增长轨迹与人口动力学中的阻滞增长模型具有相似的饱和特征，研究引入微分方程建模方法，在经典阻滞增长理论框架基础上，进行适应性改进，通过构建三个差异化的微分方程模型并开展拟合效果对比分析。

3.1 模型拟合

3.1.1 微分方程模型理论与模型构建

以广告费用 X_2 为自变量，价格差 X_1 为已知的变量，销售量为因变量，结合阻滞增长模型，建立如下微分方程模型：

$$\begin{cases} \frac{dy}{dx_2} = (a + bx_1)y(1 - \frac{y}{y_m}) \\ y(0,0.5) = 7.5 \end{cases} \quad (2)$$

其中 y_m 代表市场饱和量， a ， b 是待定系数。

根据已知的数据，可以得到待定系数 a ， b 以及市场饱和量 y_m 的估计值，因此，销售量 Y 与价格差 x_1 、广告费用 x_2 之间的关系式如下：

$$= \frac{14.3184}{1 + 0.9091e^{-(0.1678+0.3118x_1)-(x_2-5.5)}} \quad (3)$$

令其为 *Model 1*，其中市场饱和量 $y_m = 14.3184$ ，表示理论最大销售量。增长速率由 $0.1678 + 0.3118x_1$ 决定，价格差 x_1 越大，广告费用对销售量的促进作用越强。当广告费用 x_2 增加时，指数项衰减，导致销售量 Y 逐渐逼近饱和值 y_m 。

利用牙膏销售量模型的解析表达式，结合已知的价格差 x_1 、广告费用 x_2 的数据，对销售量 Y 进行预测^[1]。

3.1.2 改进微分模型一

在原微分方程模型的基础上，将原式中的价格差 x_1 更改为优惠百分比变量 x_3 （即价格差/本公司销售价格），并增加其与广告费用 x_2 的交互项，建立相应的微分方程：

$$\begin{cases} \frac{dy}{dx_2} = (a + bx_3 + cx_3x_2)y(1 - \frac{y}{y_m}) \\ y(0,0.5) = 7.5 \end{cases} \quad (3)$$

带入数据拟合得到参数估计值如下：

$$a = 0.36448, b = -49.97142, c = 8.22518, y_m = 9.52961$$

令其为 *Model 2*。

3.1.3 改进微分模型二

在改进微分模型一的基础上，我们保留了交互项，将优惠百分比变量更改为广告费用的平方，然后建立相应的微分方程，最终得到销售量 Y 与广告费用 X_2 、优惠百分比变量 X_3 的关系式：

$$\begin{cases} \frac{dy}{dx_2} = (a + bx_2^2 + cx_3x_2)y(1 - \frac{y}{y_m}) \\ y(0,0.5) = 7.5 \end{cases} \quad (4)$$

带入数据拟合得到参数估计值如下：

$$a = -0.444, b = 0.0177, c = 0.3801, y_m = 10.2100$$

对于各个项的解释如下：

$a = -0.444$ 为基础广告效应，负值可能反应广告投入初期成本压力。

$bx_2^2 = 0.0177$ 为广告费用的平房项，体现边际效益递增。

$cx_3x_2 = 0.3801x_3x_2$ 指优惠百分比的线性效应。

令其为 *Model 3*。

3.2 模型整体评估

表 3 模型拟合优度

	R2	RMSE	MAE
<i>Model 1</i>	0.8718409	0.2397805	0.1898445
<i>Model 2</i>	0.8814728	0.2305941	0.1851646
<i>Model 3</i>	0.8853875	0.2267541	0.1895415

对三个模型的拟合优度指标进行比较分析发现，其值均高于 80%，显示出三个模型均具备较强的拟合效能与预测潜力。其中 *Model 3* 的综合性能最佳，在数据拟合吻合度与预测精度上都有较好的表现，相较于其他模型具有更优的应用价值。

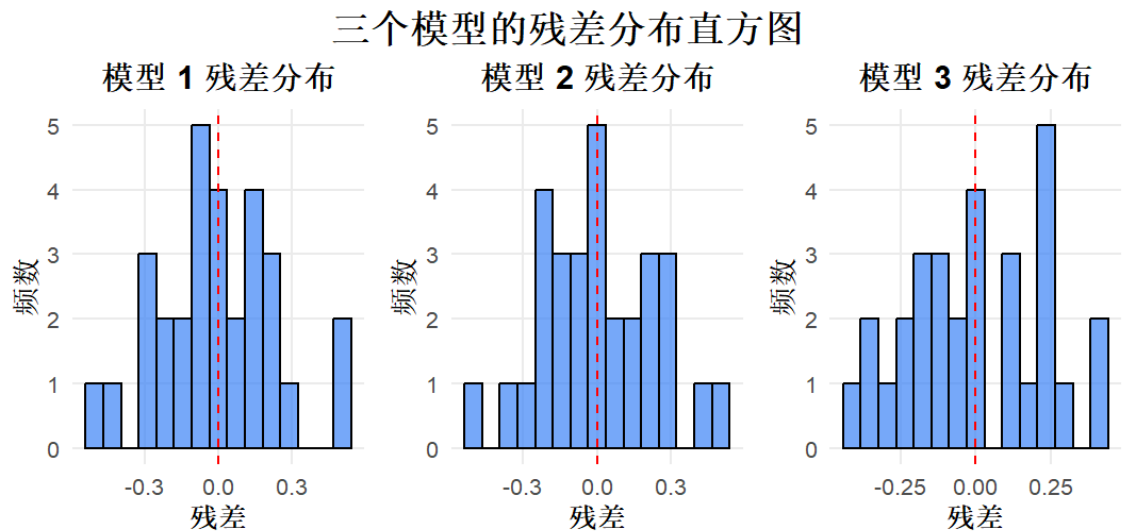


图 6 残差分布直方图

从残差分布来看（图 6），模型 1 的残差呈现单峰钟形趋势但对称性不足，中心区域频率密度较高而两侧渐次降低，这表明模型对大部分数据的拟合有一定效果，但仍存在一定程度的系统性偏倚，可能存在部分数据未被有效拟合，从而致使残差分布偏离理想的正态分布。

模型 2 的残差表现出显著的中心聚集特征，零值附近频率密度峰值突出且分布区间较窄，但仍存在轻微左偏或右偏的非对称形态，即存在偏离理想分布的状况。这说明该模型在处理大部分数据时的拟合精度尚可，但在处理边缘数据时存在局部拟合误差，残差序列的正态性检验可能存在一定偏离。

模型 3 的残差在零值附近分布较为集中，相对更趋近于正态分布，其频率密度曲线与正态分布的拟合度优于其他模型。这表明该模型的残差表现相对出色，对数据的拟合程度较高，对数据生成过程的随机扰动项具有更强的解释力，随机误差项的统计特性得到有效控制，能够较好地解释数据中的变异性。

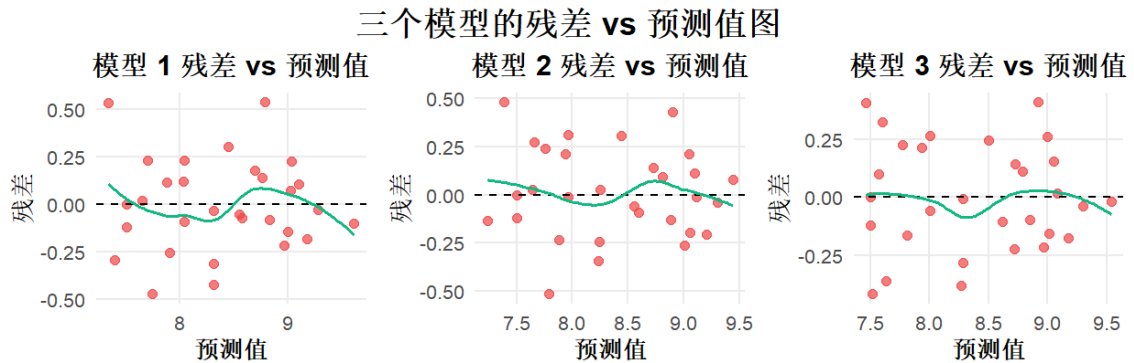


图 7 残差与预测值关系

从残差与预测值的关系来看（图 7），模型 1 的残差虽以零值为中心上下波动，但绿色拟合曲线显著弯曲，表明存在未被捕捉的非线性关系，这意味着模型可能遗漏了关键的非线性特征，致使预测出现系统性偏差。

模型 2 的残差分布有围绕零值波动的趋势，但在预测值较高的区间，残差波动幅度显著

扩大，疑似存在异方差性；同时，绿色拟合曲线的起伏揭示了潜在的非线性关系未被模型捕捉。这反映了该模型对不同范围预测值的拟合能力存在差异。

相比之下，模型 3 的残差呈现以零值为中心形成紧致的分布态势，且绿色拟合曲线趋近水平状态，说明残差序列与预测值之间的非线性关联度较小，模型对数据的拟合相对出色。这表明该模型有效控制了变量间的非线性影响和随机扰动项的异质性，残差的条件均值保持相对稳定，从而在数据拟合和预测精度上展现出更优的稳健性。

3.2.1 广告投入费用变量的分析

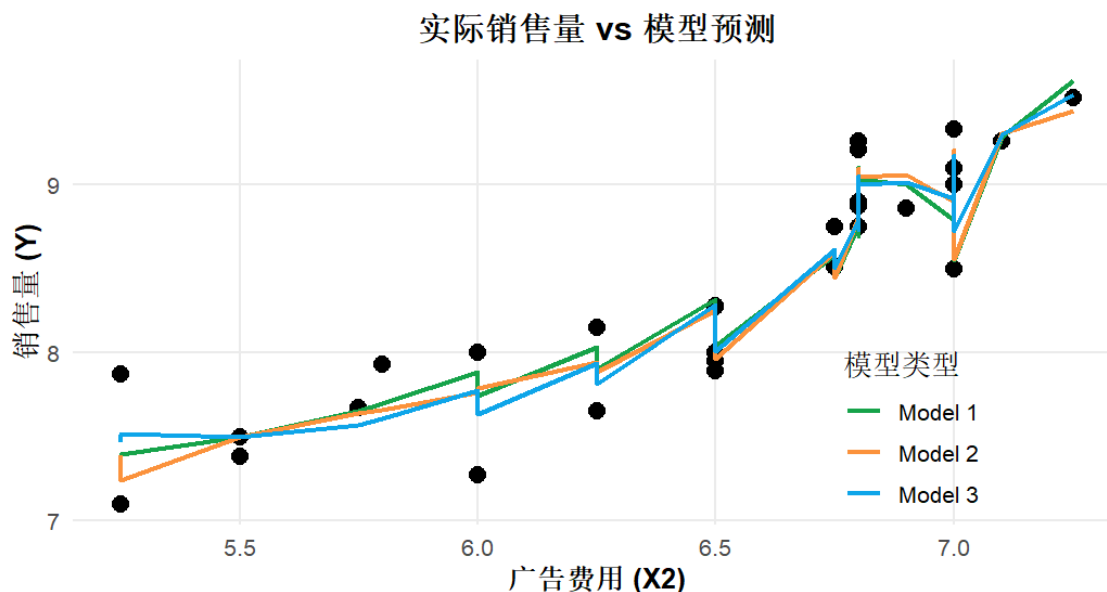


图 8 实际销售量 vs 广告费用

图 8 中黑色散点序列是实际销售量随广告费用变化的数据体现，不同颜色曲线代表各模型的预测值。整体来看，实际销售量与模型预测值均随广告费用增加呈现上升趋势，但增长轨迹表现出显著的非线性特征，与线性模型的单调增长假设形成对比。

在广告费用较低区间，模型预测值与实际值有一定偏差，各模型预测曲线存在明显差异，说明各模型在低广告费用时对销售量的预测思路 and 结果差异明显。

在广告费用较高区域，部分模型的预测值与实际销售量的吻合度显著提升，但仍存在波动和偏差。这表明模型在处理广告费用与销售量关系时，虽能捕捉到大致上升趋势，但在细节和准确性上还有提升空间，可能需要进一步考量广告投入与销售转化间的复杂非线性关系。

综合来看，模型预测的偏差分布与广告费用的边际效应相关，提示未来研究可通过引入交互项、滞后变量或非参数估计方法，进一步优化模型。

3. 2. 2 销售周期变量的分析

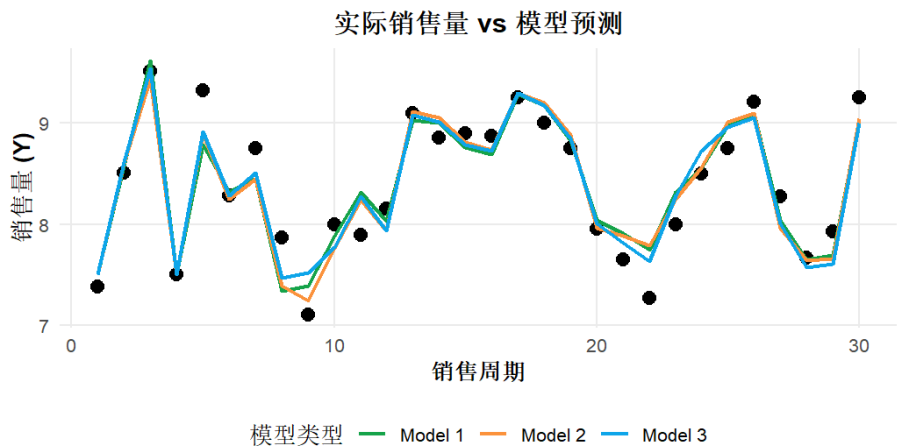


图 9 销售周期预测与实际对比

图 9 中黑色散点代表实际销售量数据，各颜色折线代表各微分方程模型的预测结果。研究发现，模型预测动态与实际销售轨迹呈现阶段性拟合特征，其偏差分布与销售周期的阶段性特征密切相关。

如图 9 所示，各模型预测值与实际销售量的走势既存在重合部分，也有明显差异：在部分销售周期节点（如起始阶段和中间某些时段），模型预测值与实际值偏差较大，折线与散点间距显著，体现出模型在这些阶段对销售量的预测准确性不足；而在另一些周期时段内，折线与散点贴合较为紧密，表明模型在这些时段对实际销售情况的拟合效果较好。

总体来看，不同模型的折线波动情况存在差异，这反映出各模型对销售周期与销售量关系的捕捉能力有所不同，部分模型可能未充分考量销售周期中影响销售量的复杂因素。

3. 2. 3 灵敏度分析

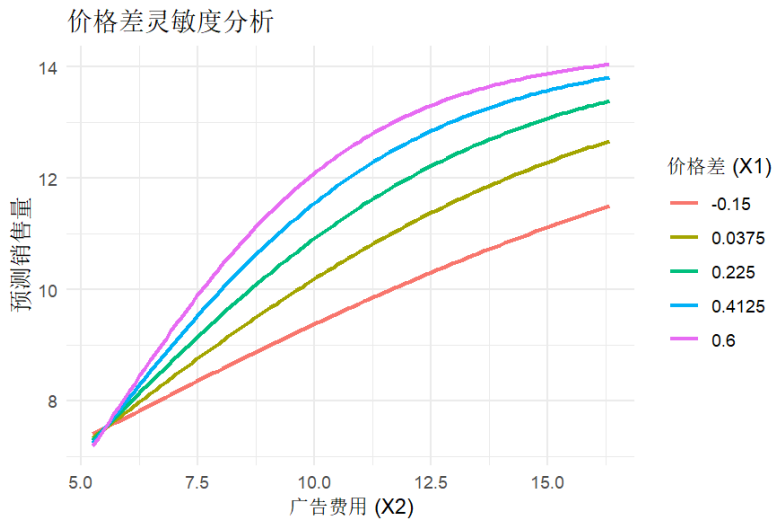


图 10 model 1 灵敏度曲线

图 10 呈现了不同价格差（X1）与广告费用（X2）对预测销售量的影响。总体来看，随着广告费用的增加，预测销售量持续上升。不同颜色线条代表的价格差显示，价格差越大，在相

同广告费用下预测销售量越高，且随着广告费用增加，价格差对销售量的影响被进一步放大。

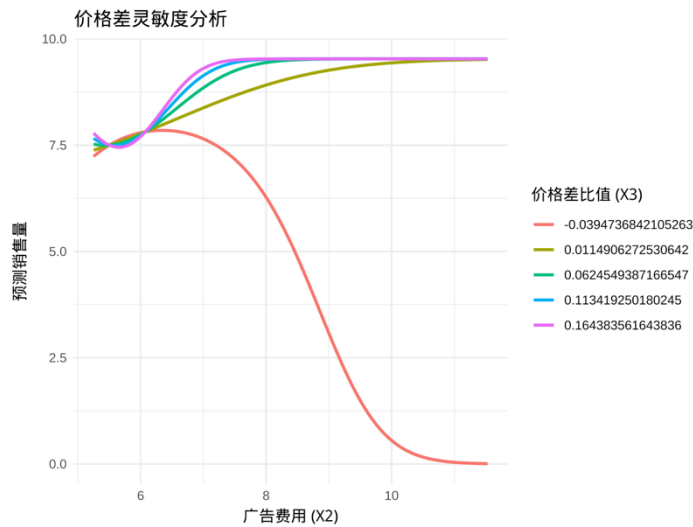


图 11 model 2 灵敏度曲线

图 11 呈现了不同价格差比值 (X3) 与广告费用 (X2) 对预测销售量的影响。总体看来，预测销售量并非呈现单一的持续上升趋势。不同价格差比值 (X3) 对应的曲线走势差异明显，说明价格差比值与广告费用对预测销售量存在复杂的交互影响。

当价格差比值为正时，在广告费用增加初期，预测销售量有不同程度的上升。这表明当本公司产品相对有价格优势（价格差比值为正）时，适当增加广告费用能促进销售量提升。当价格差比值为负时，在广告费用增加初期，预测销售量有小幅上升，但随着广告费用进一步增加，销售量急剧下降。这说明当本公司产品价格处于劣势（价格差比值为负）时，广告费用增加到一定程度会起到反作用。

从模型参数层面看，当价格差比值 X3 为负时， bX_3 会产生较大的正值，极大削弱 a 所代表的基础增长趋势；同时，交互项 cX_2X_3 也为负，随着广告费用 X2 增加，该项数值更负，进一步抑制销售量增长，甚至导致销售量下降，反映在灵敏度曲线上就呈现出反常走势。

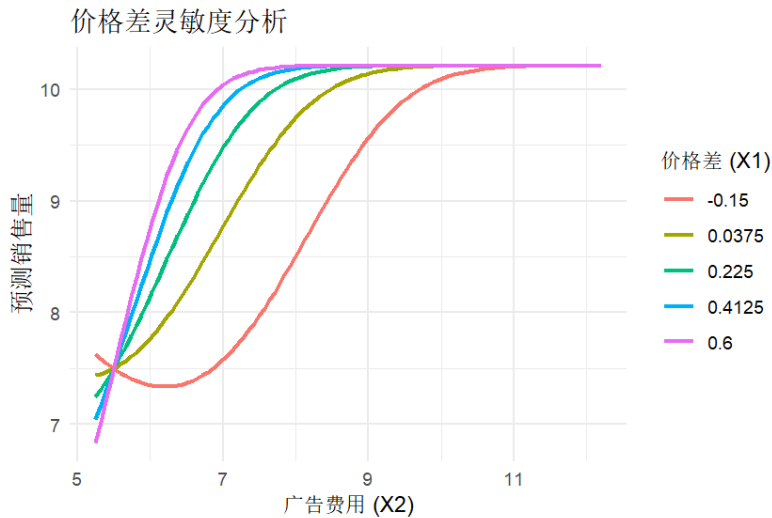


图 12 model 3 灵敏度曲线

图 12 展示了广告费用 (X_2) 与预测销售量的关系。前期随着广告费用增加, 预测销售量快速上升, 达到一定值后增长变缓并趋于平稳, 呈现饱和状态。不同价格差下, 销售量达到平稳的广告费用值和最终销售量不同。正价格差能借助广告投入更有效提升销售量并更快饱和, 负价格差可能抑制销售, 需更多广告投入扭转。

从消费者心理和市场行为角度看, 消费者对价格较为敏感。当本公司产品价格高于竞争对手 (价格差比值为负) 时, 消费者购买意愿会降低。初期广告投入较小时, 其投放质量和数量无法得到保证, 广告宣传反而会使消费者更加关注价格劣势, 从而使消费者在比较价格后, 仍更倾向于选择价格更低的竞品。而到后期, 广告投入达到一定规模时, 本公司产品已经具备一定品牌效应, 因此价格差比的影响被削弱, 销售量随广告投入的提升而增加, 直至其达到市场饱和值。

3.3 小结

基于牙膏销售量、价格差及广告费的实证数据, 本研究分析了回归分析模型在拟合市场动态时的局限性。基于牙膏销售终将趋近市场饱和和量的客观规律, 我们构建了基于阻滞增长理论的微分方程预测模型, 通过系统建模与参数校准, 最终得到了牙膏销售量的预测模型。研究进一步通过灵敏度分析, 验证了模型对市场关键变量波动的响应机制, 确认其适配于现实销售场景的动态特征。

在微分方程构建的三个模型中, Model 3 的综合性能最优。该模型在拟合精度与趋势捕捉能力上显著优于其他模型, 为销售预测与市场策略制定提供了更可靠的量化支持。

4 牙膏销售问题再分析总结

在牙膏销售模型的研究中, 多项式模型与微分方程模型从不同维度对牙膏销售量展开了分析与预测。

多项式模型纳入顾客商品存量、优惠百分比和广告投入等解释变量, 其拟合优度达约 90.07%, 整体稳定性表现良好, 但部分变量的二次项显著性不足。具体而言, 顾客商品存量对销量的影响呈现倒 U 型曲线特征; 广告投入对销量具有正向促进作用, 然而模型对其非线性效应的捕捉不够充分; 优惠百分比的一次项对销量有显著的促进作用, 二次项则显示出边际效应递减趋势, 但受限于数据规模, 相关假设的验证仍不够充分。

微分方程模型基于阻滞增长模型构建, 三个子模型的拟合优度均超过 80%, 其中 Model 3 的表现最佳, 该模型的残差分布更为集中且趋近正态分布, 残差与预测值间的非线性关联度较低, 在数据拟合与预测精度方面展现出显著优势。不过, 各模型在处理广告费用与销售量的关系时, 虽能捕捉到二者的总体变化趋势, 但在细节刻画和精准度上仍有提升空间。

综合来看, 两类模型为解析牙膏销售规律提供了多维度的理论视角, 对销售策略的制定具有一定的实践指导意义。未来研究可进一步优化模型设定, 以提升对复杂销售动态的解释能力。

5 《高铁时代下旅客出行交通方式选择及现代运输业均衡发展研究》报告^[2]的讨论分析

5.1 原报告不足之处

5.1.1 宏观与微观分析失衡

在研究出行交通方式的选择中，旅客的个体特征包含有年龄、收入、学历、出行目的及其所在城市等都可能影响到出行交通方式选择的特征。

原研究^[2]仅围绕经济效益、速度效益和舒适性三方面，探究其对出行交通方式选择的影响，而忽略了个体特征这一关键变量。事实上，个体特征在出行决策中具有较为重要的影响，从年龄维度看，年长旅客出于健康因素考量，可能倾向于选择平稳舒适的出行方式；收入水平不同，旅客对交通方式各影响因素的权重分配亦有差异，高收入群体或许更注重舒适性，而中低收入群体则更关注经济性；学历层次影响旅客认知水平，进而影响其对交通方式安全性、便捷性等方面的综合判断；出行目的直接关系到行程时间规划，商务出行往往更强调时效性，而休闲旅行则可能更看重性价比；此外，旅客所在城市的交通基础设施建设水平及各运输方式的发展水平，也显著影响其出行选择偏好。

原研究^[2]仅在宏观层面分析总体数据，而并未对旅客的个体特征进行微观层面分析，无法对精准刻画不同客群的选择偏好，这导致模型对真实出行决策的解释力受限。

5.1.2 运输方式选择集覆盖不完整

原文^[2]在运输方式选择集的构建上存在明显局限性，仅聚焦于高铁与飞机这两种中长途运输方式，而忽略了其他重要的运输方式，这违反了离散选择模型的“全面性假设”。在短途出行（距离小于 500 公里）中，公路客运和自驾是主流的运输方式。根据 2014 年的数据，公路客运周转量占比超过 60%，然而，这些短途运输方式未被纳入模型，导致短途运输市场的竞争格局分析出现偏差。同时，报告还忽略了动车和城际铁路等细分铁路类型，它们在不同的运输距离和市场需求中发挥着独特的作用。此外，对于水运（如长江流域的轮渡）以及国际运输（如跨境高铁与航空的协同）等特殊场景缺乏相应的分析，这使得研究结论的适用范围受到极大限制。

报告仅关注高铁与飞机之间的两两竞争关系，忽略了综合运输体系中多种运输方式协同发展的内在逻辑。在实际的运输网络中，短途公路运输可以作为高铁运输的喂给和疏散方式，实现旅客在高铁站与最终目的地之间的便捷衔接；长途航空与高铁可以通过航线与高铁线路的合理布局，形成优势互补的运输网络。由于报告未能充分考虑这些协同关系，导致关于“运输业均衡发展”的研究局限于对独立业态的策略建议，未能构建起基于全方位竞争 - 协同网络的系统性解决方案。因此，为了提高运输方式选择模型的准确性和研究结论的普适性，报告需要进一步完善运输方式选择集，纳入短途运输方式、细分铁路类型以及其他特殊运输场景，并深入研究多种运输方式之间的竞争与协同关系，构建更加全面、系统的分析框架。

5. 1. 3 数据的量化标准及挖掘深度和广度不足

1. 经济方面

在票价分析与时间成本考量方面，原研究存在明显局限性。

一方面原研究^[2]在处理数据时仅采用民航 0.81 元/公里、高铁 0.47 元/公里的基准票价，未能体现票价的动态变化特征与分类差异。且在实际运营中，机票在淡季折扣可低至 0.3 元/公里，另外，民航设有头等舱、经济舱，高铁划分一等座、二等座等不同票价等级。这些票价的动态变化和分类设置对旅客出行决策的影响显著，但却在研究中被忽略。

另一方面，原研究^[2]将时间成本统一设定为 22.69 元/小时，以人均 GDP 为计算依据，未分层考虑不同收入群体的时间价值差异。实际上，高收入群体时间成本可达 50-100 元/小时，不同收入分层人群对时间成本的敏感度存在显著差异，其在出行方式选择中的权重分配不相同。这种时间成本与出行决策间的复杂关联性可能是影响模型拟合效果的潜在因素。

2. 时间方面

出行时间的周期影响：以年为周期来观察不同交通方式周转量当期值（如下图 13，部分数据缺失未处理），可以发现在不同的时期不同的出行交通方式的周转量当前值以及每个交通方式的占比是不同的并且存在一定的周期性，如旅客的出行数量的高峰期会集中在暑假、国庆以及春节假期，这三个时期的交通方式周转量的占比也存在一定的差异。原文中并未将时间上的周期影响考虑到模型拟合中，会存在一定的局限性。

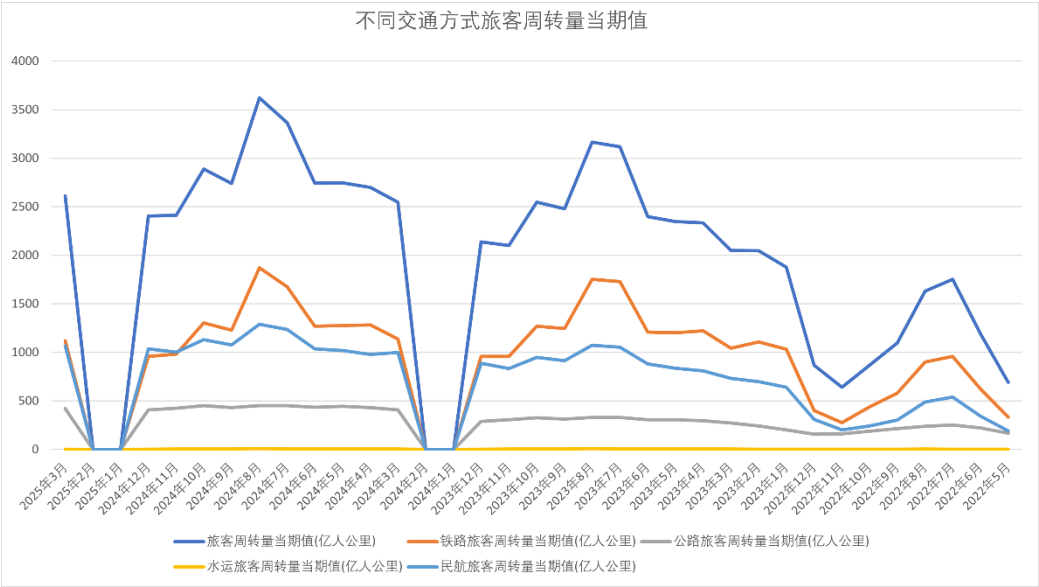


图 13 不同交通方式旅客周转量变化图

总出行时间：在实际旅客出行中出行时间的计算并非仅考虑在某种交通工具在行驶过程中的时间即并非单纯的某种交通工具的行驶速度与乘坐时间，在出行过程中还需要考虑到等候时间（如航空出行的候机时间，铁路出行中出站入站换乘时间等枢纽内的出行时间），而且大部分旅客出行存在预留充足的缓冲时间来避免市区交通的不稳定性，这个缓冲时间预留的

大小又同时与城市交通规划、市区交通保障性、旅客的时间价值观念等因素有关。高铁换乘出行相对于民航运输便捷了很多，一般候车检票上车大概需要十五分钟左右时间，而民航出行需要提前半个小时办理登机手续，过安检等，大约需要四十分钟左右，由此也可以看出在出行时间的衡量中不能只简单考虑对于交通工具运行的时间还需要考虑换乘等候的时间。

准时性：随着工作生活节奏的加快，旅客的时间观念逐步增强，出行的计划性要求也会越来越高，从而对运输方式的运输效率提出更多的要求，那么需要运输的可靠和保障性来体现。在旅客的出行决策时，会有倾向性的选择准时可靠的运出方式来减少运输晚点带来的行为计划、工作计划无法实现的情况，避免因为计划改变而带来的损失。通过相关数据，两种运输方式的准时率不同，铁路运输的准时率远远高于航空运输，高铁的准点率能接近于 100%，而航空的准点率只有 80%，也就是说民航的晚点率达到了 20%左右，由此看出对于准时性是影响选择决策的一个影响因素。

3. 舒适性方面

原文中对于舒适性的效用作用的因素有旅客疲劳恢复时间、振动加速度、以及人均面积。那么我们对于舒适度的参考因素可以为人均可旅行时间、人均面积、乘坐环境、交通工具的安全性以及服务质量等。原文对于舒适度因素和量化标准并未给出详细的解释以及数据支撑，存在一定的主观性，但是原文中所引入模型的因素在实际意义中具有一定的效用，可以作为参考来进行建模。

5. 1. 4 运输走廊的细化研究和详细对比的缺失

运输走廊为两地间产生一定规模交通流，为承担交通流所建设的运输线路的集合，其中的构成要素包括城市为走廊所链接的节点，运输线路，运输工具，要素流以及外部环境。对于原文的研究背景，我们可以将衔接两个城市，由高速铁路和民航两种运输方式组成，服务于中远距离旅客运输的条形地带

然而，本文在分析运输走廊对交通方式选择的影响时，主要集中在高铁和航空两种交通方式的竞争关系上，而对公路和水运等其他运输方式的综合影响考虑较少。此外，虽然提到了一些具体的案例，如京沪高铁对航空业的冲击，但对于运输走廊的整体布局和规划如何影响旅客选择缺乏深入探讨。

例如，文中提到高铁对短途航空客运市场产生了不同程度的分流作用，如郑州至西安高铁投入运营后，幸福航空停运了郑州至西安的航班。这表明高铁在短途运输中的优势明显，但对于中长途运输，尤其是超过 1000 公里的运输距离，航空仍然占主导地位。然而，文中并未详细分析不同距离区间内旅客选择高铁或航空的具体影响因素，如票价、旅行时间、舒适度等。此外，文中对运输走廊的基础设施建设、运输能力、运行速度和准时性等基础设施特性因素的分析较为简略，未能充分结合实际数据进行深入探讨。

以南京到上海的运输走廊为例^[3]，高铁的运行时间和频率明显优于公路和水运，这使得高铁在短途运输中具有显著优势。然而，对于一些没有高铁覆盖的地区，公路运输仍然是主要的

出行方式。文中对这些运输走廊的基础设施特性和运输能力对旅客选择的影响分析不足，未能充分说明不同运输方式在运输走廊中的互补性和竞争性。

总的来说，本文在分析运输走廊对交通方式选择的影响时，主要集中在高铁和航空的竞争关系上，而对其他运输方式和运输走廊的整体布局考虑较少。未来的研究可以进一步结合实际数据，深入分析不同运输走廊的基础设施特性和运输能力对旅客选择的影响，为我国运输业的均衡发展提供更全面的建议。

5.1.5 接驳交通模型考量的缺失

高速铁路与民航同属高端运输方式，在实际出行场景中，均需依赖其他运输方式完成城市交通的接驳。具体来看，高速铁路站点一般选址于市中心或近郊区域，与城市接驳交通的衔接较为便捷高效；而民航机场对净空有一定的要求，往往布局在远离城市中心的郊区，城市交通接驳不方便且距离较远，运输费用较高。因此，在分析高速铁路与民航的客运分担率时，除考虑区域交通的运营情况外，还需进一步分析城市接驳交通在运输时间、成本等方面的差异，及其对旅客出行方式选择的影响机制。但是原研究在运用模型分析两种高端运输方式的选择行为时，缺乏多种交通方式的考虑和研究以及对于接驳的运输方式的便捷性，时间以及成本的考虑，存在明显的分析局限性。

5.1.6 Logit 模型与效用理论的衔接与结果展现缺失

报告虽分别阐述了效用理论与多元 Logit 模型，但在效用理论应用与模型构建中存在显著缺陷，致使理论与实践脱节，研究结论缺乏深度与实证支撑。

其一，效用理论中的经济、速度、舒适度等指标，如票价、行程时间、疲劳恢复时间等，未作为 Logit 模型的解释变量直接输入，导致“偏好分析”与“选择行为建模”无法有效衔接，难以准确反映旅客出行决策逻辑。同时，研究仅停留在理论公式推导，未通过 Logit 模型估计各效用因素的权重系数与边际效应，无法量化票价上涨、行程时间变化等因素对选择概率等的具体影响，缺乏实证支撑。

其二，结果展现形式较单一，未通过可视化手段（如距离 - 价格弹性热力图、客群细分选择概率对比表）展现不同运输方式的选择概率随变量变化的规律，难以直观呈现模型结论对运输策略的指导意义。

5.2 对于报告内容的改进补足

首先简单梳理原报告的建模主要目的及方法。研究旨在从宏观和微观两个角度来分析旅客出行交通方式选择行为，通过建立数学模型对旅客运输量数据进行分析，寻找我国运输网络的均衡发展路径。基于效用理论，研究试图挖掘不同业务场景下旅客出行需求，并探讨消费群体需求特征差异及其成因。为此，报告采用多元 Logit 模型建立旅客出行方式的选择模型，着重分析选择弹性和概率，并期望据此推断不同消费群体的选择偏好倾向，进而优化运输工具供给策略，从而为我国运输业发展提供决策参考，同时为市场发展阶段和策略制定提供新的方法论，促进交通资源实现最优化配置。

从整体框架来看，研究构思完整、逻辑清晰，但是在研究各个方向的步骤与过程的落实层面存在显著不足。首先是宏观与微观维度，报告仅停留在大量列举历年各个出行方式的旅客周转量、高铁里程等宏观数据，缺乏对数据的深入分析，这与前文提及的研究局限相呼应这里不再赘述。

此外，由于数据支撑不足，“效用函数”公式改动的合理性验证、旅客属性的完整性等考量，都较难做出有效的改进补足。值得关注的是，研究存在一项重大缺陷：报告虽提出要基于效用函数与 logit 模型分析不同群体的选择弹性和概率，从而为运输工具的供给提供建议，但是研究报告中关键的概率分析结果是缺失的，这削弱了研究结论的说服力与建议的可行性，而且这些缺失的结果应该是能够依据其报告及参考文献当中的既有信息进行推导计算。

因此我们将在假定原报告公式与数据基本可靠的前提下，对相关数据和参数展开验证计算，尝试填补报告中缺失的概率结果。

5.2.1 效用函数的影响因素

对于高速铁路与民航运输方式的效用函数，我们使用了快速性、经济性、舒适性和安全性四个指标。根据其报告研究以及其引用的参考文献可得到其计算的准则：

对于经济效用：

$$U_{i,e} = I_i * d_i \quad (5)$$

I_i 表示该交通方式的票价， d_i 表示该交通方式 OD 对（出发地与目的地）的距离。同一个 OD 对采用民航运输和高速铁路运输方式的运输距离是不同的，这主要受到两种运输方式运输线路的影响，由于高铁轨道基础建设时要考虑自然地形并衔接主要城市节点，因此其线路不能像航线一样顺直，其非直线系数较高。经济效用模型表示旅客在出行中的价格与距离的乘积，可分析旅程费用对旅客对的性价比情况。国务院批准的《民航国内运价改革方案》确定国内机票价格基准价（2014 年新修订）为 0.81 元/人·km；高铁的出行票价参照京津、武广、郑西、沪宁等几条相同设计时速且已通车运营的高铁票价定价标准计算为 0.47 元/km。

对于速度时间效用：

$$U_{i,v} = \left(\frac{d_i}{V_i} + t_{i,wait} \right) \times \bar{T} \quad (6)$$

其中 V_i 表示速度指标， $t_{i,wait}$ 表示该交通方式枢纽内的出行时间， \bar{T} 表示选择出行方式的时间成本。每个人的时间都具有一定的时间成本，在牺牲时间耗费在交通的过程中同时也在丧失时间成本。当然，时间成本也具有一定的差异性。 \bar{T} 为时间成本，单位时间成本即人均时间价值，是指每个人平均每小时能创造的时间价值，其可根据统计年鉴的年度人均 GDP 与年度人均工作时间的比值计算求得。2014 年我国人均 GDP 为 45380 元，全年工作时间按 2000h 计（扣除法定节假日全年工作 250d，每天按 8h 工作制），则 2014 年的人均时间成本为 22.69 元/h。

对于舒适度效用：

$$U_{i,c} = \frac{J}{[1 + a \times e^{(-b_i \times t_i)}]} \times \bar{T} \quad (7)$$

其中， J 表示旅客极限疲劳恢复时间， b 表示旅行时间的疲劳恢复时间系数。这和目前的各种交通方式的各自行业的服务水平也存在一定的关系。 a 为常量； b_i 为该交通方式的单位旅行时间的疲劳恢复时间系数； t_i 为旅行时间； \bar{T} 为时间成本。各参数的取值分别为： $J=16$ ， $a=69$ ， $b_{\text{民航}}=0.28$ ， $b_{\text{高铁}}=0.59$ 。由此可以求出不同旅行时间 t 下的旅行疲劳恢复时间。不同的交通方式具有不同的舒适度。

在原报告的基础上，我们增加了安全性指标。对于安全性效用，国际上普遍认为，相比航空运输，高速铁路更安全也更被公众认可。我国民航运输已十年以上无重大事故，并执行常态化的安全检查。从我国实际情况来看，可认为旅客在安全偏好程度上对民航与高速铁路没有多大差距，根据李冬梅的研究，将其安全性参数设置为高铁 $U_{i,s} = 0.99$ ；民航 $U_{i,s} = 0.97$ 。

其中快速性、经济性和舒适性指标之间相互独立且能够线性互补，适用于效用函数的加法准则。而安全性是旅客运输方式选择时的首选属性，即只有在安全性较好时，旅客才会考虑该种运输方式。因此，安全性与经济性、快速性和舒适度这几项指标相互独立，但不能够线性补偿，适用于效用函数的乘法准则。

确定性效用分析模型公式如下：

$$U_i = W_{4i} U_{i,s} (W_{1i} U_{i,e} + W_{2i} U_{i,v} + W_{3i} U_{i,c}) \quad (8)$$

其中， $U_{i,e}$ 、 $U_{i,v}$ 、 $U_{i,c}$ 、 $U_{i,s}$ 分别非关于经济效用、速度时间效用，舒适度效用及安全性效用方面的考量得分。 W_{1i} 、 W_{2i} 、 W_{3i} 、 W_{4i} 分别为各属性因素的权重系数。

5.2.2 效用函数再计算

原报告中计算了 20 个特定 OD 对的经济效应，速度时间效用及安全性效用，由于其没有提供原始数据，在此我们通过 python 爬虫与高德地图 API 获取城市对的相关数据进而计算其效用函数。

表 4 各城市对交通出行数据表

航线	直线距离	驾车距离	驾车时间	飞机时间	高铁时间
南京-北京	898.30	1005.42	10.97	2.10	3.50
南京-青岛	468.94	570.15	6.52	1.60	2.10
南京-上海	269.64	296.71	3.96	1.30	1.40
武汉-上海	687.47	797.22	8.82	1.90	2.80
武汉-北京	1052.68	1141.21	11.92	2.30	4.00
武汉-青岛	829.86	1046.61	11.08	2.00	3.30
无锡-重庆	1334.47	1541.38	16.78	2.70	4.90
无锡-宁波	220.55	273.91	3.37	1.30	1.20
无锡-成都	1549.84	1816.48	19.42	2.90	5.70
郑州-北京	623.97	692.29	7.63	1.80	2.60
郑州-上海	829.47	946.45	9.64	2.00	3.30
郑州-广州	1292.14	1454.99	16.23	2.60	4.80
深圳-厦门	463.94	572.02	6.23	1.60	2.00
深圳-北京	1943.09	2164.71	22.73	3.40	7.00
深圳-上海	1213.24	1417.74	15.75	2.50	4.50

常州-北京	955.33	1050.30	11.22	2.20	3.70
常州-西安	1065.33	1211.50	12.53	2.30	4.10
常州-哈尔滨	1654.94	2109.72	22.02	3.10	6.00
重庆-北京	1458.54	1753.28	19.64	2.80	5.40
重庆-合肥	1050.79	1285.83	14.05	2.30	4.00

效用函数计算结果如表 5 出行选择概率结果表所示。

表 5 出行选择概率结果表

航线	飞机			高铁		
	U_{ie}	U_{iv}	U_{ic}	U_{ie}	U_{iv}	U_{ic}
南京-北京	727.623	47.649	9.23175	422.201	79.415	37.23311
南京-青岛	379.8414	36.304	8.052449	220.4018	47.649	17.29792
南京-上海	218.4084	29.497	7.416927	126.7308	31.766	11.63292
武汉-上海	556.8507	43.111	8.741086	323.1109	63.532	25.52132
武汉-北京	852.6708	52.187	9.749198	494.7596	90.76	48.30875
武汉-青岛	672.1866	45.38	8.983153	390.0342	74.877	33.471
无锡-重庆	1080.921	61.263	10.87001	627.2009	111.181	75.14916
无锡-宁波	178.6455	29.497	7.416927	103.6585	27.228	10.37513
无锡-成都	1255.37	65.801	11.4763	728.4248	129.333	107.1056
郑州-北京	505.4157	40.842	8.505385	293.2659	58.994	22.85957
郑州-上海	671.8707	45.38	8.983153	389.8509	74.877	33.471
郑州-广州	1046.633	58.994	10.57862	607.3058	108.912	71.69388
深圳-厦门	375.7914	36.304	8.052449	218.0518	45.38	16.35151
深圳-北京	1573.903	77.146	13.13848	913.2523	158.83	172.0798
深圳-上海	982.7244	56.725	10.29481	570.2228	102.105	62.05154
常州-北京	773.8173	49.918	9.487041	449.0051	83.953	41.365
常州-西安	862.9173	52.187	9.749198	500.7051	93.029	50.83362
常州-哈尔滨	1340.501	70.339	12.11525	777.8218	136.14	120.936
重庆-北京	1181.417	63.532	11.16917	685.5138	122.526	94.24114
重庆-合肥	851.1399	52.187	9.749198	493.8713	90.76	48.30875

对于效用函数模型当中的系数，在原报告当中只是笼统概括为“由于每个人的偏好具有不同性，是一个浮动的值，但特定群体具有一定的相似性，通过对数据进行散点图的绘制，可以线性回归得到近似值”，这其实是不恰当的。

影响旅客出行选择的各项因素的权重系数确定非常重要，因为权重系数的精确度最终影响到运量分担模型的可信度。为了得到各个属性因素的权重系数，使用层次分析法进行分析。从而与 logistic 回归模型结合进而得到不同旅客交通方式的选择概率。

5.2.3 层次分析法

层次分析法的特点是定性与定量相结合，是将部分定性问题定量化分析的决策方法。首先，它以设定的标度定量化旅客的主观判断，即在测算影响旅客决策的各个因素的权重值时，先通过交通调查数据，分析旅客侧重的结果，确定各项因素的权重；为细化权重值，提高模型预测精确度，对旅客进行分类，按照收入情况、运输距离和出行目的等分旅客类型测算权重系数；效用函数的各效用项的权重值能体现不同出行属性、旅客属性等对各个运输属性的影响。

分析旅客出行数据信息，得到旅客对各影响因素选择偏好的比例，以此作为对各影响因素的评价标准，能准确评价旅客对出行方式选择时考虑的对各个影响因素的偏好程度，在一定范围内避免了专家法的主观性。将各因素两两对比分析，得到判断矩阵：

$$A = (a_{ij})_{m \times n} (i, j = 1, 2, \dots, n) \quad (9)$$

规范化判断矩阵的各列，得到：

$$\bar{A} = (\bar{a}_{ij})_{m \times n} (i, j = 1, 2, \dots, n) \quad (10)$$

其中：

$$\bar{a}_{ij} = a_{ij} / \sum_{i=1}^n a_{ij} (i, j = 1, 2, \dots, n) \quad (11)$$

对规范化的矩阵的规范列平均处理：

$$W_i = \frac{1}{n} / \sum_{i=1}^n a_{ij} (i, j = 1, 2, \dots, n) \quad (12)$$

得到特征向量：

$$W = [W_1 \dots W_n]^T (i, j = 1, 2, \dots, n) \quad (13)$$

W_i 即为各项影响因素的权重。

因素权重值确定

层次分析法中的层次包括目标层、准则层和措施层（如图 14），以各类旅客集选择能使自身效益最大化的交通方式为目标层来构建层次分析模型。目标层为最高层，为选择效用最大的运输方式；准则层是第二层，考虑准时性可与快速性合并分析，选择效用最大的运输方式的指标为经济性、快速性、舒适性、安全性；方案层是第三层，即选择的效用最大的运输方式。本次研究的是高速铁路和民航两种运输方式，因此，把各类旅客集对高速铁路和民航运输方式的选择问题转化为方案层两种运输方式效用高低的排序问题。

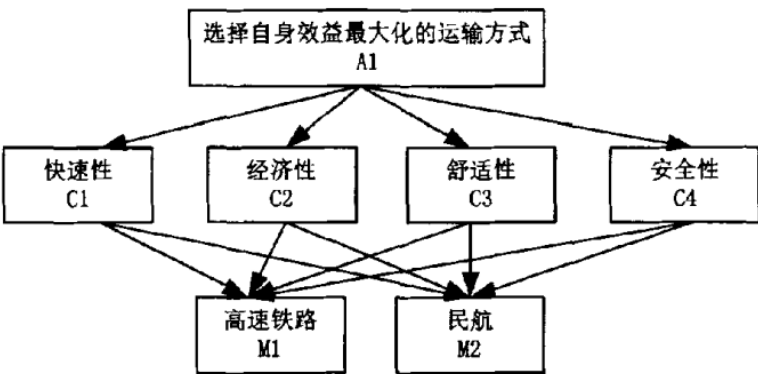


图 15 层次结构图

表 6 层次分析法重要性标度表

重要性标度	含义
1	两个因素同样重要
3	两者相比，前者略重要
5	两者相比，明显前者重要
7	两者相比，前者特别重要
9	两者相比，前者极其重要
2, 4, 6, 8	上述比较的折中情况

在某些特殊情况中，层次分析法的判断矩阵具有一致性和传递性。在普通情况下，不要求层次分析法的判断矩阵严格满足一致性和传递性，但要符合一定的逻辑规律，例如：A 比 B 重要，B 比 C 重要，则 A 与 C 对比时 A 更重要；若对比时出现 A 比 C 的重要度较 A 比 B 的要小，则不符合逻辑规律。

因此，再矩阵判断中要求满足基本的一致性，并进行一致性检验，以确定判断矩阵具有合理的逻辑，并继续分析判断结果。在我们的研究中，参考了文献^[3]层次分析法的参数

根据层次分析法，可得每一种不同属性的旅客对于各个指标偏好的不同权重，并得到以下的数据结果：

表 7 旅客类型权重数据表

旅客类型	快速性权重	经济性权重	舒适性权重	安全性权重
客流 1（高收入公商务）	0.398	0.011	0.153	0.437
客流 2（中等收入公商务）	0.257	0.123	0.174	0.445
客流 3（低收入公商务）	0.264	0.232	0.111	0.398
客流 4（高收入休闲度假）	0.341	0.022	0.154	0.483
客流 5（中等收入休闲度假）	0.201	0.154	0.16	0.485
客流 6（低收入休闲度假）	0.126	0.41	0.077	0.387

5.2.4 选择概率函数

结合上面的权重系数，以及效用函数模型，计算出各种人群对于各个选择偏好的得分，进一步建模求解不同旅客出行的交通方式选择概率。二项 Logit 模型公式如下：

$$\begin{aligned} P_1 &= b \int_{-\infty}^{+\infty} w * \exp(-by) dy \\ &= \int_0^1 \frac{w * \exp(-by)}{w * \exp(-by)[1 + \exp(bv_2 - bv_1)]} dw \\ &= \frac{1}{1 + \exp(bv_2 - bv_1)} \\ &= \frac{\exp(bv_1)}{\exp(bv_1) + \exp(bv_2)} \end{aligned} \tag{14}$$

解得：

$$P = \frac{\exp(bv_1)}{\exp(bv_1) + \exp(bv_2)} \tag{15}$$

其中 b 是效用函数中的尺度参数， $U_i = b \cdot v_i + \epsilon_i$

U_i 是第 i 个选项的总效用 v_i 是第 i 项的系统效用（可观测部分，例如价格、时间、距离等线性组合） ϵ_i 是误差项，服从 Gumbel 分布；

根据以上步骤，计算对于不同类型的人群对于不同的出行目的以及不同的 OD 对的情况下的出行选择高铁出行的概率，结果如下表 8 所示：

表 8 各城市对分客流交通方式选择概率表

航线	高收入公商务	中收入公商务	低收入公商务	高收入休闲度假	中收入休闲度假	低收入休闲度假
南京-北京	0.374663	0.401279	0.439927	0.376825	0.410296	0.473706
南京-青岛	0.324579	0.364945	0.416925	0.328532	0.377922	0.462407
南京-上海	0.442096	0.460241	0.48	0.44451	0.465751	0.50021
武汉-上海	0.341538	0.375724	0.422682	0.344606	0.387047	0.464114
武汉-北京	0.340567	0.372959	0.420299	0.343072	0.383962	0.4608
武汉-青岛	0.218769	0.281241	0.36695	0.223582	0.301762	0.434765
无锡-重庆	0.169893	0.241745	0.340858	0.175099	0.265363	0.417021
无锡-宁波	0.204347	0.289492	0.384837	0.212829	0.31436	0.453267
无锡-成都	0.166062	0.240214	0.341755	0.171433	0.264442	0.418708
郑州-北京	0.436414	0.452113	0.473024	0.43806	0.457287	0.493541
郑州-上海	0.231761	0.287493	0.365144	0.235969	0.305951	0.427081
郑州-广州	0.282937	0.325889	0.38921	0.285929	0.340505	0.440793
深圳-厦门	0.380374	0.409753	0.445755	0.383643	0.419044	0.479391
深圳-北京	0.190375	0.256846	0.351971	0.194698	0.279059	0.424139
深圳-上海	0.188955	0.262042	0.358603	0.194083	0.285279	0.426543

常州-北京	0.344518	0.377303	0.424469	0.347195	0.388379	0.465573
常州-西安	0.200932	0.267083	0.358475	0.205813	0.28885	0.429469
常州-哈尔滨	0.179937	0.248498	0.347104	0.184469	0.271531	0.423385
重庆-北京	0.173304	0.244682	0.34373	0.178262	0.268143	0.418295
重庆-合肥	0.214582	0.275741	0.362608	0.218946	0.296185	0.431412

总的来说，由于舒适度与时间的差异，高铁相对于飞机来说速度较慢，相同的路程时，出行时间就会更长，对于出行的疲劳舒适方面的影响也就更大，所以说高铁相对于飞机出行来说有着天然的劣势，所以在飞机与高铁二选一之间时，高铁总是稍低；

从计算出来的具体结果来看，不同群体选择高铁的概率存在比较大的差异。低收入休闲度假旅客选择高铁的平均概率最高，达到了 44.72%，而高收入公商务旅客选择高铁的平均概率最低，仅为 27.03%。这种差异不仅体现在不同收入等级之间，也体现在不同出行目的之间。具体来说，低收入群体，无论是公商务旅客还是休闲度假旅客，选择高铁的平均概率都显著高于高收入群体，这背后反映的是经济水平对出行成本敏感度的直接影响，低收入群体由于预算限制，更倾向于选择价格相对亲民的高铁作为出行方式，以在满足基本出行需求的同时，控制交通支出，而高收入群体由于经济实力较强，可能更注重出行的时间效率和舒适度，因此在长途或时间敏感的出行中，可能更倾向于选择飞机等其他交通方式，此外一个就是，经济水平因素差异还与出行目的相互作用，共同影响着旅客的高铁选择行为，休闲度假旅客相对于公商务旅客来说，更愿意为高铁旅行支付一定的溢价，因为他们更看重高铁旅行本身可能带来的舒适体验和沿途风景，而公商务旅客则可能更注重出行的时间成本，因此在时间效率上更具优势的其他交通方式可能成为他们的首选。

此外，出行目的也对高铁选择概率产生比较大的影响。对于高收入、中收入和低收入旅客来说，休闲度假旅客选择高铁的平均概率都高于公商务旅客。这可能是因为休闲度假旅客更注重旅行体验，而高铁旅行本身可能被视为一种旅行体验，因此他们更倾向于选择高铁。

航线特性也是影响高铁选择概率的重要因素，不同航线间的高铁选择概率差异也非常大，这可能与航线距离、高铁服务便利性、竞争交通方式等因素有关。南京-上海航线的平均概率最高，这可能是因为该航线距离较短，高铁在时间上的差别没有那么大了，且高铁服务便利，使得了更多的旅客选择。而常州-哈尔滨航线的平均概率相对较低，这可能是因为该航线距离较长，对比飞机，高铁在时间上的优势不明显，导致旅客选择高铁的概率降低。

5.3 个体数据再分析

5.3.1 调查问卷收集与 probit 回归模型建立

在本研究视角下，高铁、飞机等大型运输交通工具的普及与发展，无疑是影响居民出行意向选择的重要因素，但同样重要的是，调查居民出行交通工具选择意向需进行多重维度综合考量：一方面，需关注受访者的统计学特征，涵盖年龄、收入、家庭结构等个人属性，同时结合出行距离、目的地类型等出行场景，以此识别不同群体的差异化需求；另一方面，需深入

分析成本敏感度、舒适性偏好、环境意识及安全感知等交通工具属性偏好，并纳入公共交通覆盖率、限行政策、数字化工具使用习惯等外部环境因素，从而揭示政策导向、技术革新与基础设施建设对选择行为的作用机制。由此可见，在分析不同群体的出行意向选择时，基于真实调查的个人数据能够最为客观、准确地反映实际出行选择需求。然而，经广泛检索学术文献与公开数据库后发现，当前尚无可用的关于个人层面出行选择偏好数据，且原报告并没有使用个体数据进行研究，难以提供有效参考。基于此，我们设计了一套针对性调查问卷，围绕年龄、收入、出行频率等基础信息，以及出行成本、出行时间、出行舒适度的评分等指标展开调研，最终回收了有效问卷 89 份。经信度分析，问卷的 Cronbach. a 系数为 0.810，表明该问卷具有良好的内部一致性与可靠性。

表 9 问卷调查信度效度表

项目	因子 1	共同度
当你选择出行方式的时候，出行成本对于你的选择有多重要	0.8	0.647
当你选择出行方式的时候，出行时间对于你的选择有多重要	0.8	0.64
当你选择出行方式的时候，出行的舒适度对于你的选择有多重要	0.8	0.647
当你选择出行方式的时候，出行安全对于你的选择有多重要	0.78	0.616
特征根值（旋转前）	2.55	-
方差解释率%（旋转前）	63.76	-
累积方差解释率%（旋转前）	63.76	-
特征根值（旋转后）	2.55	-
方差解释率%（旋转后）	63.76	-
累积方差解释率%（旋转后）	63.76	-
KMO 值	0.784	-
巴特球形值	218.829	-

对数据建立 probit 回归模型，结果如下：

选择出行目的为工作出差，长出行距离（大于 600 千米）交通工具选择的数据进行建模。
首先是模型总体上的似然比卡方检验的结果显示，显著性 P 值 0.000***，水平上呈现显著性，拒绝原假设，因而模型是有效的。

表 10 probit 回归模型结果

似然比卡方	P	AIC	BIC
48.042	0.000***	68.042	94.094

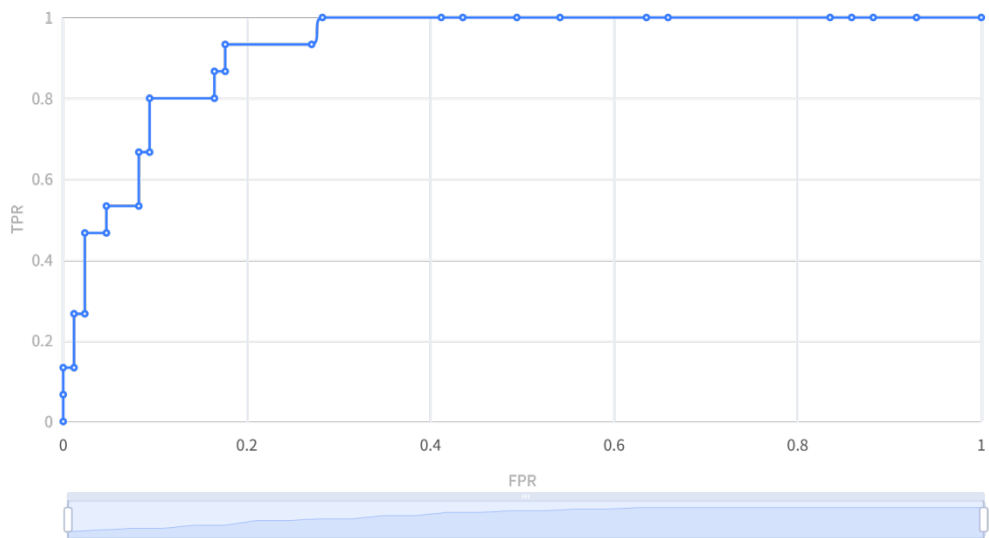


图 16 模型拟合 ROC 曲线

表 11 分类评价指标

准确率	召回率	精确率	F1	AUC
0.89	0.89	0.882	0.885	0.926

根据上图和上表，可以看出准确率达到 0.89，表明模型在整体上有较高的预测正确率；召回率为 0.89，表明模型能够有效识别出大部分正类样本；精确率为 0.882，表明模型在预测正类时有较高的准确性；F1 值为 0.885，表明模型在精确率和召回率之间取得了较好的平衡；AUC 值即 ROC 曲线下面积，是衡量模型整体分类性能的重要指标，AUC 值为 0.926，表明模型具有很好的区分能力，接近理想模型。

下面是各项的估计值以及显著性：

表 12 模型系数结果表

	回归系数	标准误差	z	P	边际效应	95%置信区间	
						上限	下限
常数	2.394	0.877	2.73	0.006***		0.675	4.112
出行时间重要性	-0.053	0.175	-0.302	0.762	-0.007	-0.395	0.289
出行舒适度重要性	-0.202	0.205	-0.985	0.325	-0.026	-0.605	0.2
出行成本重要性	-0.31	0.158	-1.964	0.049**	-0.04	-0.619	-0.001
出行安全重要性	-0.117	0.188	-0.625	0.532	-0.015	-0.485	0.251
月均收入_2.0	0.197	0.714	0.277	0.782	0.027	-1.202	1.597
月均收入_3.0	-1.36	0.648	-2.099	0.036**	-0.186	-2.631	-0.09
月均收入_4.0	-1.234	0.705	-1.749	0.080*	-0.178	-2.616	0.149
出行频率_2.0	-0.449	0.508	-0.884	0.377	-0.061	-1.444	0.546
出行频率_3.0	-0.256	0.65	-0.394	0.693	-0.031	-1.53	1.018

***、**、*分别代表 1%、5%、10%的显著性水平

5.3.2 模型结果解读

基于回归模型的分析结果，各变量系数及显著性水平揭示了不同因素对乘客选择高铁或

民航的差异化影响，具体解读如下：

1. **模型整体显著性**：模型整体有效（ $\beta=2.394$, $P=0.006$ ），具有统计学意义；
2. **出行时间因素**：无显著影响（ $\beta=-0.053$, $P=0.760$ ），可能与高铁/民航时间差异小或决策多因素综合有关；
3. **出行舒适度因素**：未达显著（ $\beta=-0.202$, $P=0.325$ ），两种交通方式舒适性优势不同且非核心决策依据；
4. **出行成本因素**：显著负向影响（ $\beta=-0.31$, $P=0.049$ ），这可能是因为在长距离运输场景下，民航单位里程成本可能低于高铁，尤其是当乘客能提前预订到特价机票时，而高铁票价相对固定且长距离总价较高。因此，成本敏感型乘客更倾向于选择民航出行；
5. **出行安全因素**：无显著差异（ $\beta=-0.117$, $P=0.532$ ），两者均属高安全性交通方式；
6. **收入水平因素**：中高收入群体（8000-11000 元： $\beta=-1.36$, $P=0.036$ ；11000+元： $\beta=-1.234$, $P=0.080$ ）更倾向民航，可能追求性价比（中收入）或高效出行（高收入）；
7. **出行频率因素**：无显著影响（ $\beta=-0.449/-0.256$, $P=0.377/0.693$ ），决策受行程紧急度、距离等多因素调控。

5.4 实例建模小结

本次实例分析最重要是从微观角度引入个人特征的数据带入模型进行分析，补充了原报告中只考虑宏观角度不考虑微观角度的不足。刻画不同客群的选择偏好，以满足不同层面上的出行选择的需求分析。

出行成本是影响乘客选择高铁或民航的重要因素之一。在长距离出行中，民航的单位里程成本可能低于高铁，尤其是当乘客能提前预订到特价机票时。而高铁的票价相对固定，且在长距离时总价较高，这使得成本敏感型乘客更倾向于选择民航。

月均收入在 8000-11000 元的乘客选择高铁的概率较低，可能是因为这一收入群体在出行时更注重性价比。他们可能愿意花时间寻找便宜的机票，而不是直接选择高铁。而高收入群体（月均收入 11000 元以上）选择高铁的概率也较低，但原因不同。他们可能更注重时间成本，愿意支付更高的价格以节省时间。

虽然出行时间重要性在模型中不显著，但在实际中，时间因素仍然是许多乘客选择交通方式的重要考虑。高铁在中短距离的出行中具有明显的时间优势，因为其无需提前到达车站、安检时间短、准点率高等特点。

本次实例收集通过线上公开问卷调查方式进行，关于本次调查问卷，虽然通过了信效度检验，但是在设计过程中存在设计时间不充分，问题设计缺少较为严谨的考量，考虑的指标不够完善，量化标准设计不够标准的问题，导致我们在本次研究由存在由部分指标结果显示不理想，数据粗糙等情况。同时本次问卷采集时间较短，导致采集的样本量数目较少，人群分布不够均匀等问题，也同样会导致结果与实际存在偏倚。

本次建模中对数据进行了基于效用模型的 logit 概率选择模型，建模可能存在一定的简

略，并未使用前面提及的运输走廊分析以及接驳模型的使用，需要不断的改进和完善对数据的挖掘和分析

在本次实例分析中分析结果的数据参数与原报告所得结果由较大的差别，这个可能是因为原报告所分析的数据来自十年前，高铁与民航的发展已经改变了许多，人们的交通出行观念也已经产生了改变，还有可能是因为样本量的不充足导致结果差异。

以上的小结仅在已有的数据上进行建模分析，由于可能存在上述提到的问题或不足之处导致结果并未完全反应实际情况导致模拟结果与实际中存在偏差，所以本次建模研究可能存在结果与显示存在偏差，还需要改进不足之处对建模进行完善，但是依然在部分显著性的指标中存在可以参考的地方。

5.4.1 企业或交通优化建议

1. 对高铁的建议

(1) 优化票价策略

针对成本敏感型乘客（如月均收入 8000-11000 元的群体），可以推出更多折扣票或套餐服务，提高其选择高铁的意愿。

在长距离路线上，可以考虑与航空公司竞争，推出更具竞争力的票价，尤其是在非高峰期。

(2) 提升服务质量

进一步提升高铁的舒适度和服务质量，以吸引高收入群体。例如，提供更高端的商务座或 VIP 服务，增加车上娱乐设施等。

加强高铁站的配套设施建设，如便捷的停车服务、快速安检通道等，减少乘客的出行时间成本。

(3) 加强中短距离路线的竞争力：

在中短距离（如 500-1000 公里）的路线上，突出高铁的时间优势，通过广告宣传等方式提高乘客的认知度。

增加班次密度，提供更灵活的出行选择，满足不同乘客的需求。

2. 对民航的建议

(1) 优化航线布局

针对长距离路线（如 1000 公里以上），进一步优化航线布局，增加航班频率，提高乘客的出行便利性。

加强与高铁的互补合作，例如在中短距离路线上与高铁进行联运，提供无缝衔接的出行服务。

(2) 提升价格竞争力

针对成本敏感型乘客，可以推出更多特价机票，尤其是在非高峰期。通过灵活的价格策略吸引价格敏感型乘客。

加强与旅游机构、酒店等的合作，推出套餐服务，提高整体性价比。

（3）提升服务质量

进一步提升机上服务质量，如提供更丰富的餐饮选择、更舒适的座椅、更优质的娱乐设施等，以吸引高收入群体。

加强机场的服务设施建设，如增加快速安检通道、提升候机体验等，减少乘客的出行时间成本。

6 参考文献

- [1] 王智峰,陈传军,孙丰云.牙膏销售量问题的微分方程模型[J].数学学习与研究,2021,(01):139-140.
- [2] 刘意程,李丝琦,高雅琦,等.高铁时代下旅客出行交通方式选择及现代运输业均衡发展研究[C]//中国统计教育学会.2015 年（第四届）全国大学生统计建模大赛论文.南京邮电大学,2015:587-610.
- [3] 李冬梅.运输走廊内高速铁路与民航的运量分担模型研究[D].东南大学,2015.

7 附录