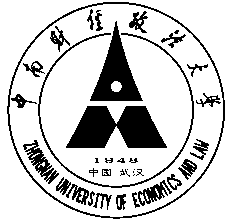
****

**本科生学年论文**



|  |  |  |
| --- | --- | --- |
| 论文题目 | **：** | **外卖平台啤酒销量数据对战略制定的启发** |
| 姓名 | **：** | **郑中天** |
| 学号 | **：** | **1603100142** |
| 班级 | **：** | **统计学中英班** |
| 年级 | **：** | **2017级** |
| 专业 | **：** | **统计学** |
| 学院 | **：** | **统计与数学 学院** |
| 指导教师 | **：** | **宁瀚文 教 授** |
| 完成时间 | **：** | **2020年 7 月13日** |

**作者声明**

本论文报告是在老师的指导下由本人独立撰写完成的，没有剽窃、抄袭、造假等违反道德、学术规范和其他侵权行为。对本论文报告的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。因本论文报告引起的法律结果完全由本人承担。

特此声明。

|  |  |  |
| --- | --- | --- |
| 作者专业 | ： | **统计学** |
| 作者学号 | ： | **1603100142** |
| 作者签名 | ： |  |
|  | **年 月 日** | |

**由啤酒销量数据看百威英博外卖平台战略制定**

**郑中天**

***Inspiration for strategy advisement drawn from online beer sales data set***

*Zheng, Zhong Tian*

**2020年7月13日**

**摘要**

随着互联网经济的快速发展, 外卖已经成为食品消费的重要渠道, 其中涉及物流，营销，战略布局等多个问题。扩大在外卖渠道的销量及布局，探索与之相关的营销策略显然成为亟待探索的重要问题。

本文围绕外卖数据做了相应的可视化，剖析了总订单和啤酒订单的转化关系，不同产品的受欢迎程度，变量对上述关系影响的重要程度，并就策略层面进行讨论。

本文系运用广义线性模型中的泊松，二项分布混合模型[[1]](#footnote-1)，及回归决策树[[2]](#footnote-2)进行讨论。全文分3个部分：首先，是通过可视化， 探索性分析可能存在的数量关系，其次，是在假定各家订单总量已知的前提下，讨论其中如何转化为各品牌啤酒订单的，并建立决策树讨论哪个属性影响最大。最后，重点关注销量特别高的商家，并就特殊营销方式进行讨论。以及辩证性讨论模型的不足。

**关键词：**啤酒销量；外卖；泊松分布；决策树；

**Abstract**

With the rapid development of Internet economy, take-out has become an important channel for food consumption, which involves logistics, marketing, strategic layout and other issues. To expand the sales volume and layout in the take-out channel and explore the relevant marketing strategy has obviously become an important issue to be explored.

This paper makes corresponding visualization around the takeout data, analyzes the conversion relationship between total orders and beer orders, the popularity of different products, and the importance of variables on the above relationship, and discusses the strategic level.

In this paper, poisson, binomial mixed model and regression decision tree in generalized linear model are discussed.This paper is divided into three parts: firstly, it analyzes the possible quantitative relationship through visualization and exploration; secondly, it discusses how to convert the total amount of each order into the order of each brand of beer under the assumption that the total amount of each order is known, and establishes a decision tree to discuss which attribute has the greatest influence.Finally, focus on merchants with particularly high sales volume and discuss special marketing methods.And the deficiency of dialectical discussion model.

**Key words:** beer sale; food delivery; poisson generalized linear model; decision tree;

Contents

[引论 - 1 -](#_Toc66493962)

[**（一）研究目的与意义** - 1 -](#_Toc66493963)

[**（二）主要研究内容** - 1 -](#_Toc66493964)

[**（三）数据集概览** - 1 -](#_Toc66493965)

[一、市场占有及销量的可视化分析 - 2 -](#_Toc66493966)

[**（一）公司旗下子公司品牌及市场份额** - 2 -](#_Toc66493967)

[**（二）**啤酒订单和各变量数量关系 - 4 -](#_Toc66493968)

[二、正式建模及分析 - 7 -](#_Toc66493969)

[结论与建议 - 13 -](#_Toc66493970)

[主要参考文献 - 14 -](#_Toc66493971)

**引论**

**（一）研究目的与意义**

外卖作为O2O[[3]](#footnote-3)的一种特殊形式，以其高流量，相对低单价，时效性的特点，成为收餐饮，快消食品企业，包装，物流企业的重点关注对象。通过研究啤酒公司订单这一指标与各个变量的数量关系，以及各品牌的分布与销量，可以指导企业优化定价，对产品的开发和铺货也有启发意义，最终实现扩大市场占有率，提高利润。

**（二）主要研究内容**

本文广义线性模型，针对各家店啤酒总订单关于各变量建模，针对

分析和探讨，具体研究内容为：

第一部分为引论，介绍了为什么要分析销量数据，提及了外卖平台对快消的战略意义。

第二部分利用可视化分析进行探索性分析，直观地体现了以百威英博为首地啤酒大厂拥有的子品牌数量及定位；销量和外卖平台，平均售价等变量的数量关系。

第三部分对啤酒订单建模，假定该店的订单数（exposure）由该店的吸引力，受欢迎程度，风味等核心因素决定，讨论变量对啤酒订单影响；

第四部分对套餐这一特殊性营销方式进行讨论，同时刻画在高端高定价市场的战略。

**（三）数据集概览**

|  |  |  |
| --- | --- | --- |
| **表格名** | **字段名** | **字段意义** |
| 门店数据 | check\_id | 店铺ID，用于关联两张表格 |
|  | poc\_name\_dummy | 门店名 |
|  | platform\_enum | 平台 |
|  | sub\_channel\_enum | 菜系 |
|  | open\_time | 营业时间 |
|  | rating\_score\_num | 店铺评分 |
|  | order\_num | 该店产生月售单数 |
|  | beer\_sku\_num | 该店在售啤酒商品数量 |
|  | abi\_beer\_sku\_num | 该店在售ABI啤酒商品数量 |
|  | beer\_sales\_by\_volume\_num(hl) | * 啤酒总销量（HL） |
|  | abi\_beer\_sales\_by\_volume\_num(hl) | ABI啤酒总销量（HL） |
|  | poc\_chain\_name\_dummy | 连锁名 |
|  | poc\_chain\_num | 连锁店的数量 |
|  | night\_flg | 该店是否夜间运营 |
|  | sku\_num | 该店在售商品数量 |
|  | sales\_volume\_num | 该店全部产品订单数量加总 |
|  | sales\_value\_cny\_amt | 该店全部产品订单金额加总 |
|  | beer\_sales\_volume\_num | 该店产生啤酒订单数量加总 |
|  | beer\_sales\_value\_cny\_amt | 该店产生啤酒订单金额加总 |
|  | abi\_beer\_sales\_volume\_num | 该店产生ABI啤酒订单数量加总 |
|  | abi\_beer\_sales\_value\_cny\_amt | 该店产生ABI啤酒金额加总 |
| 菜品数据 | check\_id | 店铺ID，用于关联两张表格 |
|  | poc\_name | 店铺名 |
|  | platform\_enum | 平台 |
|  | sku\_id | sku的ID |
|  | sku\_name | sku名称 |
|  | sku\_cls\_desc | 商品分类 |
|  | price\_num | 商品价格，可能为空值 |
|  | discount\_price\_num | 促销价，没有空值 |
|  | sales\_volume\_num | 销量 |
|  | sales\_value\_cny\_amt | 销售额，使用促销价进行计算 |
|  | beer\_brand\_enum | 啤酒品牌 |
|  | beer\_sub\_brand\_enum | 啤酒子品牌 |
|  | beer\_manufacturer\_enum | 啤酒生产商 |
|  | single\_pack\_num | 单品数量，每个商品包含几支啤酒 |
|  | single\_pack\_beer\_volume\_num(l) | 单品体积（L为单位） |
|  | whole\_pack\_beer\_volume\_num(hl) | 整包体积（HL为单位） |
|  | single\_pack\_unit | 单品规格（瓶/罐/桶等） |

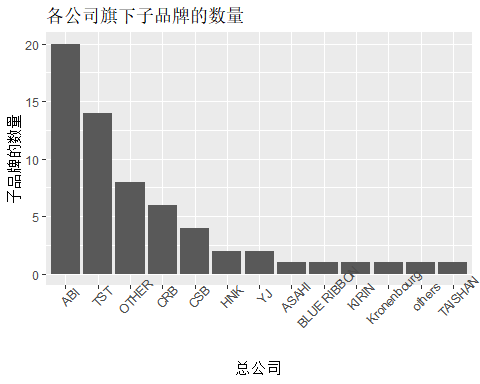
**一、****市场占有及销量的可视化分析**

**（一）公司旗下子公司品牌及市场份额**

**1．市场占有及销量的可视化分析**

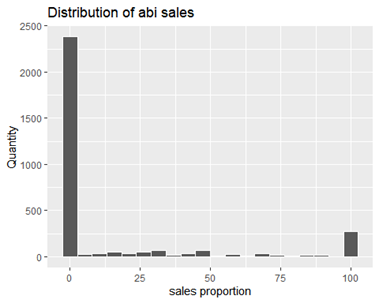
**（1）**在各个公司中，ABI百威英博拥有最多的子公司数目20家，TST青岛啤酒其次，

OTHER 中包含千岛湖，白熊，蓝星花，艾丁格，美乐，瓦伦丁6款品牌。子品牌最少的是TAISHAN，法国Kronenbourg，日本Kirin等。

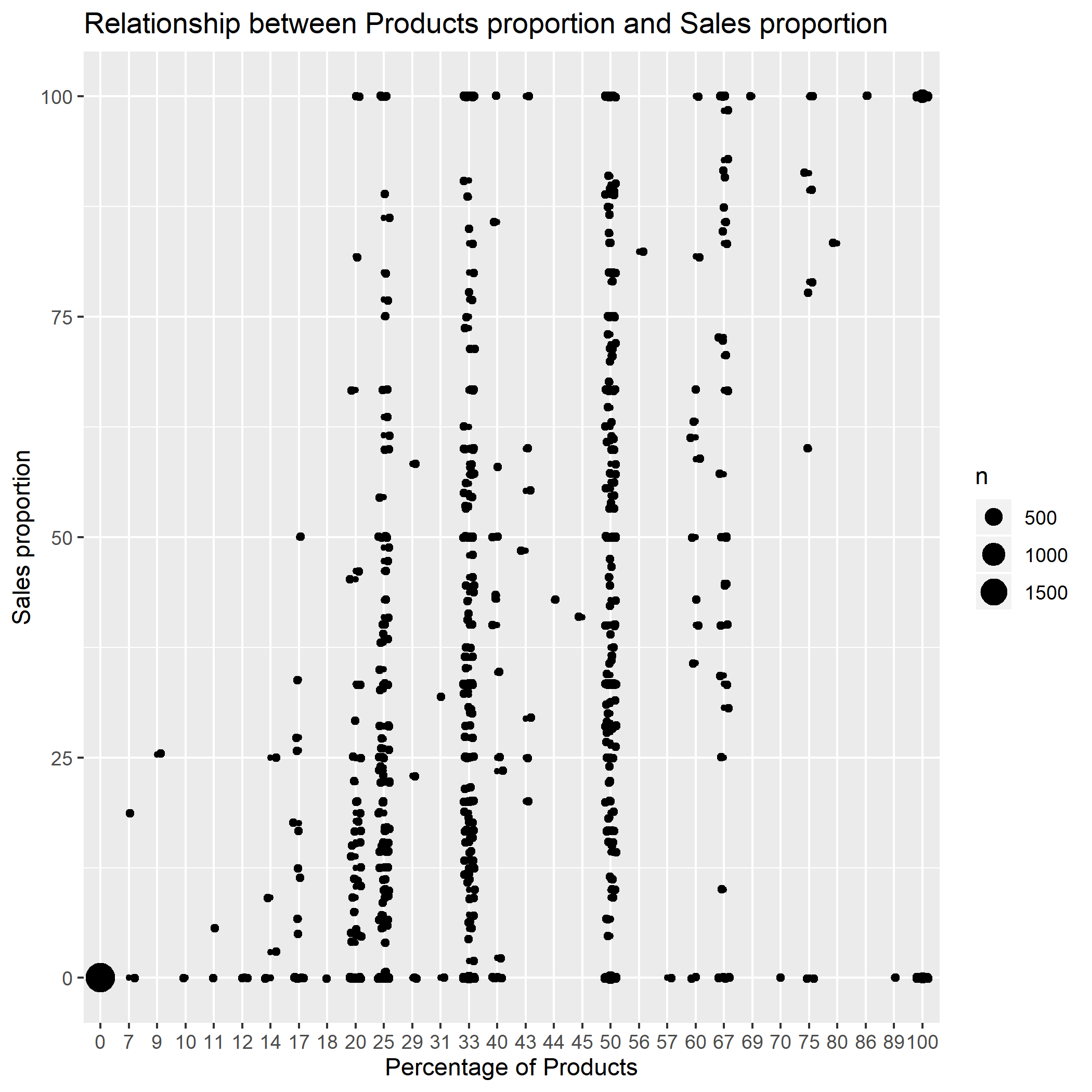


（2）本文从百威英博立场出发， 故重点分析其市场占有。

在下图中，可以看到有2380个商铺没有任何ABI旗下产品的销售, 其中分为有ABI产品但是没有卖出去, 可能是其所进货产品在该店铺不受欢迎，更换产品将有利于销售；1930店铺没有进任何ABI产品，那么对其进行有针对性的推销将有利于增加销量。



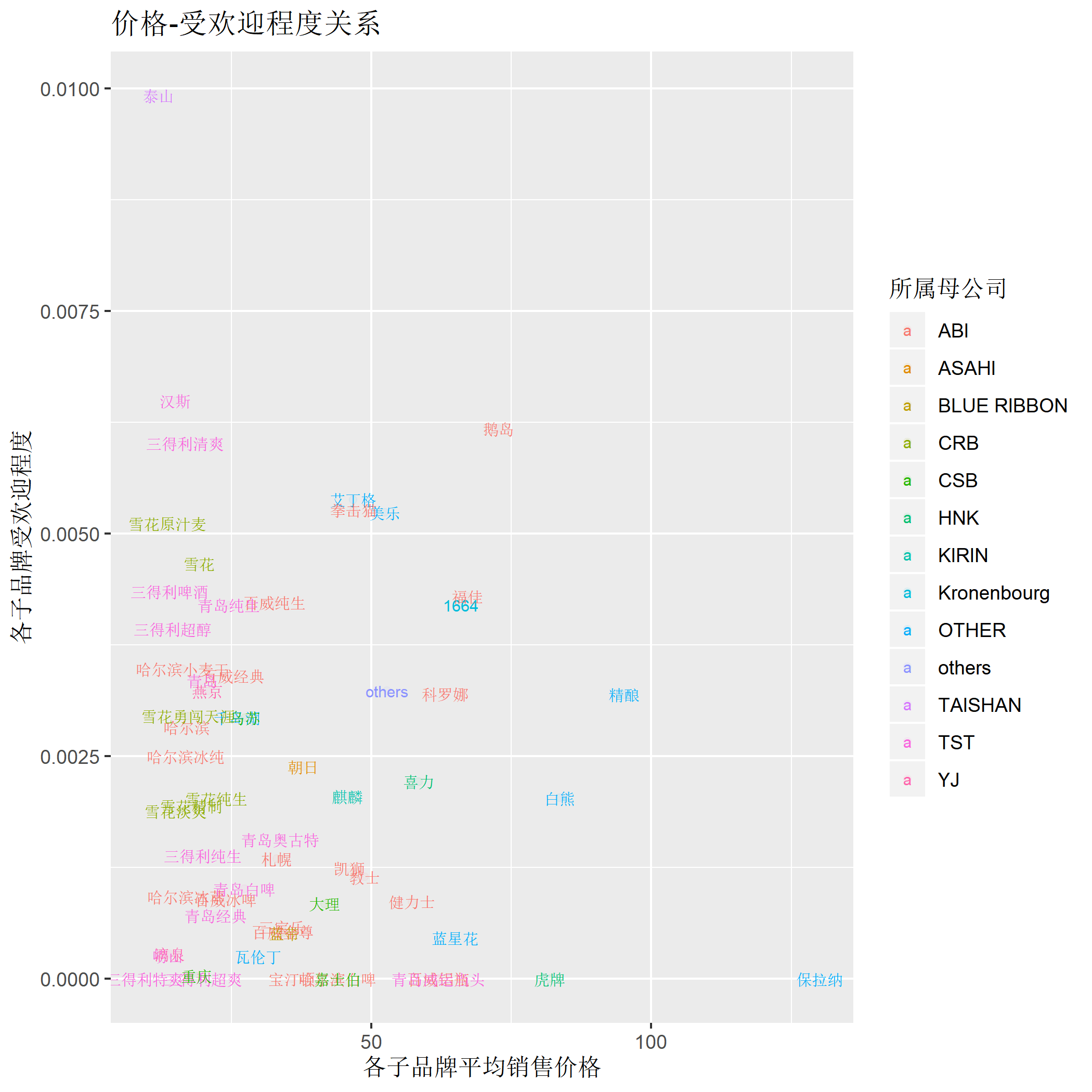
在有百威啤酒的店铺中，我们考虑百威啤酒占总啤酒数量的比例对该店的市场占有会不会有影响。在下图中可以看到，随着ABI在各店的占比增加，卖出的份额平均来说随之增加，虽然线性关系较弱。



**（二）**啤酒订单和各变量数量关系

根据常识，影响各个店铺的啤酒销量的影响因素有该店的菜系总类，营业时间，是否夜间开店，店面评分，合作的外卖平台，啤酒的种类及数量这些店面属性，以及啤酒的平均定价（利用总售价除以总销量得到）。通过一个店中某一品牌的**订单数量除以总订单数量**，以衡量该品牌啤酒在该店的受欢迎程度。

在下图中可以看到，大多数啤酒品牌的平均受欢迎程度形成了一条曲线，符合经济学的供需原理，同时有以鹅岛为典型的一系列高端产品，虽然价格较高，但是游离于曲线之外，很容易作出品牌效应带来了额外收益的推理，体现了消费者愿意为高质量产品额外付费。

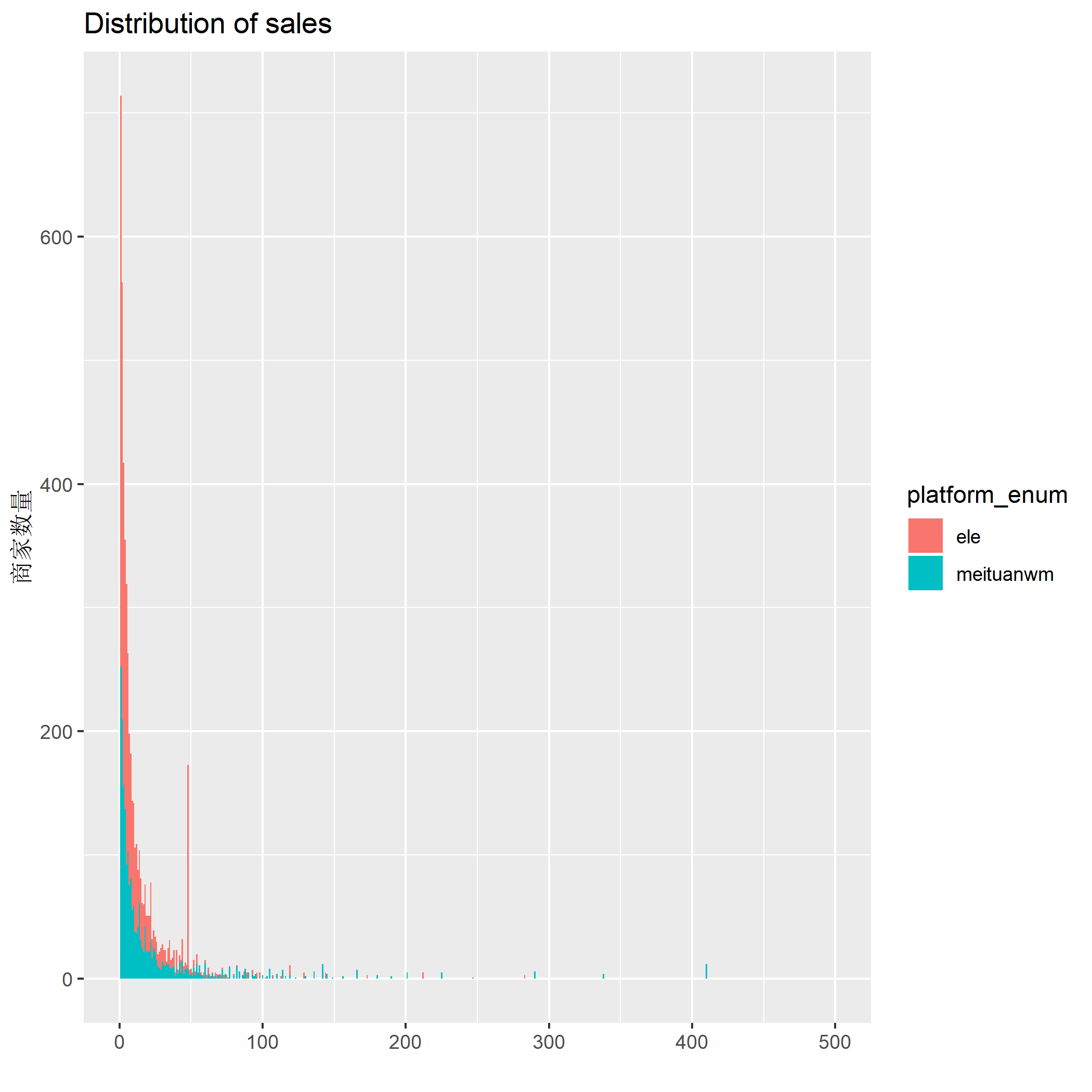


各品牌啤酒每升平均价格的百分位数

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 5% | 25% | 50% | 75% | 95% |
| 16.00000 | 20.00000 | 29.09091 | 70.65915 | 10.00000 |

观察到数据集中的营业时间open\_time含有过多因子，以形如['10:30:00-13:40:00', '16:50:00-23:30:00']的格式存储，具有结构性特征，考虑用正则表达式将其数量化，得到各店营业的总时间open\_sum。这一变量还有更多可能的信息可以挖掘，例如是否包含某一特定时间点/段，在该数据集中已经给出是否夜间运营，故不再深究。

对于总订单转化为啤酒订单，如图所示, 一大部分订单没有转化为啤酒订单，故单纯地用Poisson模型将无法解释这种现象，考虑使用二项分布-泊松混合模型，图中似乎饿了么更少配送含夹带啤酒销售的订单。



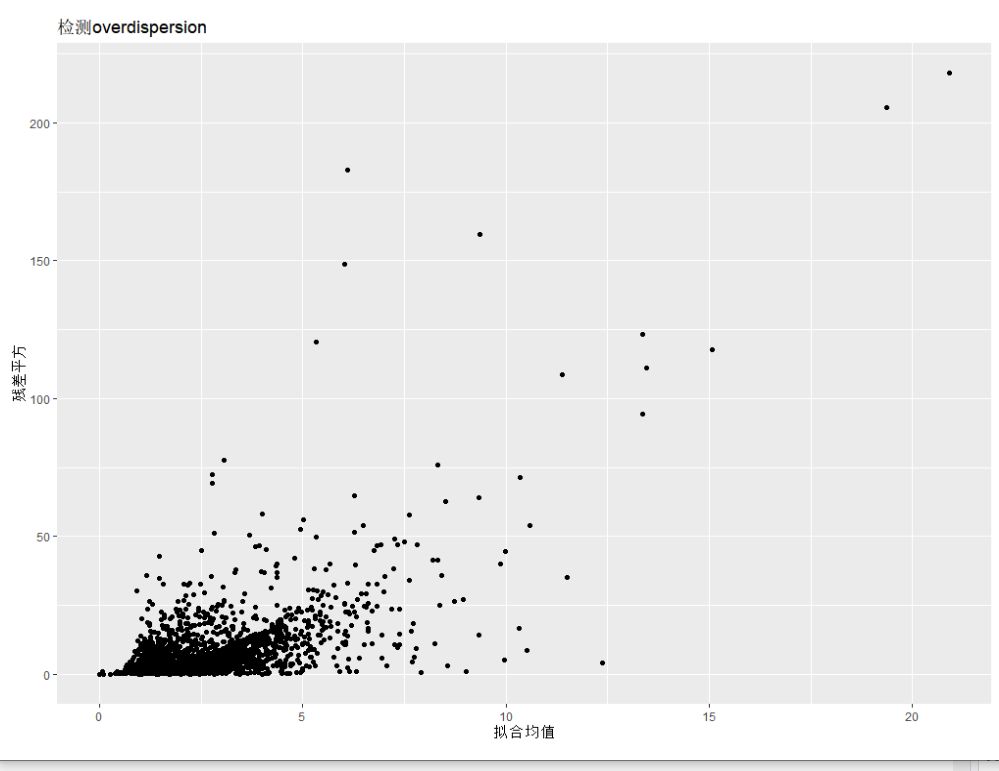
**二、正式建模及分析**

(一)Hurdle方法-广义线性模型

商家的订单数显然是由于其口味等核心竞争力决定, 每一单订单有可能会点啤酒，由数据集可见，这个转化率并不高，普遍低至千分之一。建模使用二项分布和泊松分布混合模型,来解释模型中过多的0订单,我们将每一家商铺总订单作为泊松分布中的暴露数exposure，同时也作为二项分布中的总次数N

各个牌子的啤酒订单关于offset（总订单数）(啤酒品牌beer\_sub\_brand\_enum +折后价格discount\_price\_num +菜品数量sku\_num+啤酒产品数量beer\_sku\_num +平台platform\_enum+菜系sub\_channel\_enum +评分rating\_score\_num+ 夜间营业night\_flg+营业总时长open\_sum+每个商品包含几支啤酒single\_pack\_num + 单位single\_pack\_unit建立hurdle模型

检测overdispersion[[4]](#footnote-4)



可见残差都在y=x, 与模型假设相违背, 处理这种情况, 我们考虑改用负二项分布

Call:

hurdle(formula = sales\_volume ~ beer\_sub\_brand\_enum + discount\_price\_num + platform\_enum +

sku\_num + sub\_channel\_enum + beer\_sku\_num + rating\_score\_num + night\_flg + open\_sum +

single\_pack\_unit + single\_pack\_num, data = ., offset = log(sales\_volume\_num), dist = "negbin",

link = "logit")

Pearson residuals:

Min 1Q Median 3Q Max

-1.644 -0.610 -0.364 0.181 34.004

Count model coefficients (truncated negbin with log link):

Estimate Std. Error z value Pr(>|z|)

(Intercept) -7.707e+00 7.391e-01 -10.428 < 2e-16 \*\*\*

beer\_sub\_brand\_enumothers -4.841e-01 4.323e-01 -1.120 0.262768

beer\_sub\_brand\_enum艾丁格 -7.221e-01 7.951e-01 -0.908 0.363771

beer\_sub\_brand\_enum白熊 -9.443e-01 5.885e-01 -1.605 0.108591

beer\_sub\_brand\_enum百威冰啤 -1.567e+01 1.874e+03 -0.008 0.993331

beer\_sub\_brand\_enum百威纯生 -1.356e+00 5.346e-01 -2.537 0.011189 \*

beer\_sub\_brand\_enum百威金尊 -1.464e+00 8.180e-01 -1.790 0.073461 .

beer\_sub\_brand\_enum百威经典 -7.415e-01 4.317e-01 -1.717 0.085901 .

beer\_sub\_brand\_enum朝日 -7.590e-01 5.010e-01 -1.515 0.129821

beer\_sub\_brand\_enum大理 -2.154e+00 1.433e+00 -1.503 0.132882

beer\_sub\_brand\_enum鹅岛 -9.450e-01 9.183e-01 -1.029 0.303411

beer\_sub\_brand\_enum福佳 -7.682e-01 4.649e-01 -1.652 0.098470 .

beer\_sub\_brand\_enum哈尔滨 -5.532e-01 4.865e-01 -1.137 0.255527

beer\_sub\_brand\_enum哈尔滨冰纯 -9.395e-01 6.891e-01 -1.363 0.172749

beer\_sub\_brand\_enum哈尔滨冰萃 -1.639e+00 8.396e-01 -1.952 0.050928 .

beer\_sub\_brand\_enum哈尔滨小麦王 -2.486e-01 5.342e-01 -0.465 0.641653

beer\_sub\_brand\_enum汉斯 6.879e-01 4.874e-01 1.411 0.158191

beer\_sub\_brand\_enum健力士 -1.557e+00 1.178e+00 -1.321 0.186478

beer\_sub\_brand\_enum教士 -2.427e-01 8.336e-01 -0.291 0.770920

beer\_sub\_brand\_enum精酿 -1.180e+00 4.408e-01 -2.678 0.007411 \*\*

beer\_sub\_brand\_enum凯狮 -1.932e+00 8.774e-01 -2.202 0.027649 \*

beer\_sub\_brand\_enum科罗娜 -5.569e-01 4.930e-01 -1.130 0.258567

beer\_sub\_brand\_enum蓝带 -1.666e+00 1.132e+00 -1.472 0.141030

beer\_sub\_brand\_enum蓝星花 -1.815e+01 1.900e+03 -0.010 0.992381

beer\_sub\_brand\_enum崂山 -1.793e+00 1.120e+00 -1.601 0.109347

beer\_sub\_brand\_enum漓泉 -2.221e+00 1.511e+00 -1.471 0.141416

beer\_sub\_brand\_enum麒麟 -3.658e-01 7.539e-01 -0.485 0.627471

beer\_sub\_brand\_enum千岛湖 -1.478e+00 8.102e-01 -1.824 0.068109 .

beer\_sub\_brand\_enum青岛 -3.787e-01 4.357e-01 -0.869 0.384793

beer\_sub\_brand\_enum青岛奥古特 4.903e-01 1.002e+00 0.490 0.624465

beer\_sub\_brand\_enum青岛白啤 -1.429e+00 7.244e-01 -1.973 0.048481 \*

beer\_sub\_brand\_enum青岛纯生 -1.034e+00 4.381e-01 -2.361 0.018223 \*

beer\_sub\_brand\_enum青岛经典 -1.220e+00 4.606e-01 -2.649 0.008061 \*\*

beer\_sub\_brand\_enum拳击猫 1.525e+00 1.127e+00 1.353 0.176013

beer\_sub\_brand\_enum三宝乐 6.715e-02 1.991e+00 0.034 0.973097

beer\_sub\_brand\_enum三得利超醇 -3.280e-01 4.746e-01 -0.691 0.489470

beer\_sub\_brand\_enum三得利纯生 -1.682e-01 8.044e-01 -0.209 0.834384

beer\_sub\_brand\_enum三得利啤酒 -5.598e-02 4.371e-01 -0.128 0.898094

beer\_sub\_brand\_enum三得利清爽 -4.451e-01 5.131e-01 -0.868 0.385621

beer\_sub\_brand\_enum泰山 1.351e+00 1.123e+00 1.203 0.229062

beer\_sub\_brand\_enum瓦伦丁 -1.027e+00 1.538e+00 -0.668 0.504398

beer\_sub\_brand\_enum乌苏 -5.366e-01 4.842e-01 -1.108 0.267748

beer\_sub\_brand\_enum喜力 -1.513e+00 6.299e-01 -2.402 0.016301 \*

beer\_sub\_brand\_enum雪花 -4.390e-02 4.360e-01 -0.101 0.919794

beer\_sub\_brand\_enum雪花纯生 -9.598e-01 4.476e-01 -2.144 0.031997 \*

beer\_sub\_brand\_enum雪花淡爽 -4.646e-01 1.128e+00 -0.412 0.680319

beer\_sub\_brand\_enum雪花勇闯天涯 -5.141e-01 4.416e-01 -1.164 0.244415

beer\_sub\_brand\_enum雪花原汁麦 -2.032e-01 4.473e-01 -0.454 0.649705

beer\_sub\_brand\_enum燕京 -9.595e-01 6.693e-01 -1.433 0.151737

beer\_sub\_brand\_enum札幌 -8.983e-02 1.566e+00 -0.057 0.954256

beer\_sub\_brand\_enum重庆 -1.583e+01 1.581e+03 -0.010 0.992010

discount\_price\_num -1.639e-02 2.255e-03 -7.269 3.63e-13 \*\*\*

platform\_enummeituanwm 5.438e-02 6.617e-02 0.822 0.411161

sku\_num -4.167e-03 4.523e-04 -9.212 < 2e-16 \*\*\*

sub\_channel\_enum川湘菜 2.546e+00 5.873e-01 4.336 1.45e-05 \*\*\*

sub\_channel\_enum地方菜系 2.292e+00 5.881e-01 3.898 9.70e-05 \*\*\*

sub\_channel\_enum韩式料理 2.135e+00 6.302e-01 3.387 0.000706 \*\*\*

sub\_channel\_enum汉堡/披萨 1.790e+00 6.124e-01 2.922 0.003475 \*\*

sub\_channel\_enum火锅/串串香 2.290e+00 5.988e-01 3.824 0.000131 \*\*\*

sub\_channel\_enum简餐便当 2.080e+00 5.863e-01 3.548 0.000388 \*\*\*

sub\_channel\_enum饺子馄饨 1.968e+00 6.070e-01 3.242 0.001186 \*\*

sub\_channel\_enum卤味鸭脖 1.188e-01 1.436e+00 0.083 0.934085

sub\_channel\_enum冒菜/麻辣烫 1.146e+00 6.140e-01 1.866 0.062085 .

sub\_channel\_enum米粉面馆 1.740e+00 5.906e-01 2.946 0.003216 \*\*

sub\_channel\_enum奶茶果汁 3.934e+00 1.564e+00 2.516 0.011883 \*

sub\_channel\_enum轻食沙拉 2.583e+00 6.921e-01 3.733 0.000190 \*\*\*

sub\_channel\_enum日料寿司 2.509e+00 6.734e-01 3.726 0.000195 \*\*\*

sub\_channel\_enum烧烤/小龙虾 2.608e+00 5.837e-01 4.469 7.86e-06 \*\*\*

sub\_channel\_enum生鲜 5.476e+00 7.363e-01 7.437 1.03e-13 \*\*\*

sub\_channel\_enum特色小吃 3.374e+00 6.035e-01 5.590 2.27e-08 \*\*\*

sub\_channel\_enum甜品饮品 1.522e+00 9.682e-01 1.572 0.115869

sub\_channel\_enum西餐 2.090e+00 8.764e-01 2.385 0.017068 \*

sub\_channel\_enum香锅干锅 2.023e+00 6.185e-01 3.272 0.001070 \*\*

sub\_channel\_enum炸鸡炸串 2.278e+00 5.971e-01 3.816 0.000136 \*\*\*

beer\_sku\_num 1.723e-02 1.863e-03 9.245 < 2e-16 \*\*\*

rating\_score\_num -1.052e-01 2.639e-02 -3.988 6.67e-05 \*\*\*

night\_flg 5.176e-01 6.928e-02 7.471 7.94e-14 \*\*\*

open\_sum -6.454e-03 6.910e-03 -0.934 0.350317

single\_pack\_unit瓶 -9.350e-02 1.194e-01 -0.783 0.433507

single\_pack\_unit桶 -8.461e-01 8.819e-01 -0.959 0.337402

single\_pack\_unit扎 -4.126e+00 1.062e+00 -3.884 0.000103 \*\*\*

single\_pack\_num 7.556e-02 2.785e-02 2.713 0.006670 \*\*

Log(theta) -8.356e-01 7.122e-02 -11.734 < 2e-16 \*\*\*

Zero hurdle model coefficients (binomial with logit link):

Estimate Std. Error z value Pr(>|z|)

(Intercept) -1.184e-01 9.198e-01 -0.129 0.897558

beer\_sub\_brand\_enumothers -2.055e-01 7.566e-01 -0.272 0.785932

beer\_sub\_brand\_enum艾丁格 -1.412e+00 1.384e+00 -1.020 0.307783

beer\_sub\_brand\_enum白熊 4.344e-02 1.317e+00 0.033 0.973677

beer\_sub\_brand\_enum百威冰啤 1.324e+01 2.400e+03 0.006 0.995597

beer\_sub\_brand\_enum百威纯生 -6.687e-01 8.662e-01 -0.772 0.440124

beer\_sub\_brand\_enum百威金尊 -5.554e-01 1.061e+00 -0.523 0.600789

beer\_sub\_brand\_enum百威经典 -1.797e-01 7.454e-01 -0.241 0.809497

beer\_sub\_brand\_enum朝日 1.023e+00 9.947e-01 1.028 0.303766

beer\_sub\_brand\_enum大理 1.383e+01 1.636e+03 0.008 0.993253

beer\_sub\_brand\_enum鹅岛 -1.119e+00 1.594e+00 -0.702 0.482756

beer\_sub\_brand\_enum福佳 -1.429e+00 7.808e-01 -1.830 0.067220 .

beer\_sub\_brand\_enum哈尔滨 -8.860e-03 8.548e-01 -0.010 0.991730

beer\_sub\_brand\_enum哈尔滨冰纯 5.093e-01 1.299e+00 0.392 0.694919

beer\_sub\_brand\_enum哈尔滨冰萃 -2.153e-03 1.340e+00 -0.002 0.998718

beer\_sub\_brand\_enum哈尔滨小麦王 2.974e-01 1.054e+00 0.282 0.777798

beer\_sub\_brand\_enum汉斯 1.957e+00 1.256e+00 1.559 0.119052

beer\_sub\_brand\_enum健力士 -1.950e+00 1.704e+00 -1.144 0.252427

beer\_sub\_brand\_enum教士 -2.979e+00 1.037e+00 -2.873 0.004069 \*\*

beer\_sub\_brand\_enum精酿 -8.109e-01 8.521e-01 -0.952 0.341267

beer\_sub\_brand\_enum凯狮 5.964e-01 1.348e+00 0.442 0.658277

beer\_sub\_brand\_enum科罗娜 -6.467e-01 8.282e-01 -0.781 0.434930

beer\_sub\_brand\_enum蓝带 -1.097e+00 1.510e+00 -0.726 0.467861

beer\_sub\_brand\_enum蓝星花 1.245e+01 1.625e+03 0.008 0.993886

beer\_sub\_brand\_enum崂山 1.414e+01 1.697e+03 0.008 0.993349

beer\_sub\_brand\_enum漓泉 1.421e+01 2.400e+03 0.006 0.995276

beer\_sub\_brand\_enum麒麟 8.123e-01 1.378e+00 0.589 0.555593

beer\_sub\_brand\_enum千岛湖 -1.318e+00 1.051e+00 -1.254 0.209944

beer\_sub\_brand\_enum青岛 -3.788e-01 7.521e-01 -0.504 0.614548

beer\_sub\_brand\_enum青岛奥古特 -4.967e-01 1.412e+00 -0.352 0.725009

beer\_sub\_brand\_enum青岛白啤 -1.123e+00 9.657e-01 -1.163 0.245003

beer\_sub\_brand\_enum青岛纯生 -4.610e-01 7.539e-01 -0.611 0.540873

beer\_sub\_brand\_enum青岛经典 -1.704e+00 7.592e-01 -2.244 0.024829 \*

beer\_sub\_brand\_enum拳击猫 1.254e+01 1.697e+03 0.007 0.994105

beer\_sub\_brand\_enum三宝乐 1.274e+01 1.662e+03 0.008 0.993881

beer\_sub\_brand\_enum三得利超醇 5.603e-01 8.649e-01 0.648 0.517133

beer\_sub\_brand\_enum三得利纯生 -7.746e-01 1.048e+00 -0.739 0.459691

beer\_sub\_brand\_enum三得利啤酒 8.921e-01 7.827e-01 1.140 0.254397

beer\_sub\_brand\_enum三得利清爽 2.718e-01 8.989e-01 0.302 0.762412

beer\_sub\_brand\_enum泰山 1.358e+01 1.643e+03 0.008 0.993407

beer\_sub\_brand\_enum瓦伦丁 -1.879e+00 1.711e+00 -1.098 0.271990

beer\_sub\_brand\_enum乌苏 -5.696e-01 8.351e-01 -0.682 0.495206

beer\_sub\_brand\_enum喜力 -1.530e+00 9.311e-01 -1.643 0.100280

beer\_sub\_brand\_enum雪花 8.565e-01 7.748e-01 1.105 0.268976

beer\_sub\_brand\_enum雪花纯生 -1.107e+00 7.589e-01 -1.459 0.144669

beer\_sub\_brand\_enum雪花淡爽 1.352e+01 1.648e+03 0.008 0.993456

beer\_sub\_brand\_enum雪花勇闯天涯 1.484e-01 7.691e-01 0.193 0.846972

beer\_sub\_brand\_enum雪花原汁麦 3.327e-01 7.952e-01 0.418 0.675673

beer\_sub\_brand\_enum燕京 -3.838e-01 1.099e+00 -0.349 0.726993

beer\_sub\_brand\_enum札幌 1.445e+01 2.400e+03 0.006 0.995196

beer\_sub\_brand\_enum重庆 -1.673e+00 1.623e+00 -1.031 0.302704

discount\_price\_num -1.462e-02 3.177e-03 -4.602 4.19e-06 \*\*\*

platform\_enummeituanwm -1.035e+00 1.081e-01 -9.580 < 2e-16 \*\*\*

sku\_num -4.551e-04 6.711e-04 -0.678 0.497662

sub\_channel\_enum川湘菜 2.000e+00 5.136e-01 3.894 9.86e-05 \*\*\*

sub\_channel\_enum地方菜系 2.031e+00 5.139e-01 3.952 7.74e-05 \*\*\*

sub\_channel\_enum韩式料理 1.464e+00 6.184e-01 2.367 0.017956 \*

sub\_channel\_enum汉堡/披萨 2.796e+00 6.104e-01 4.581 4.62e-06 \*\*\*

sub\_channel\_enum火锅/串串香 1.477e+00 5.388e-01 2.741 0.006134 \*\*

sub\_channel\_enum简餐便当 1.874e+00 5.096e-01 3.677 0.000236 \*\*\*

sub\_channel\_enum饺子馄饨 2.366e+00 6.171e-01 3.834 0.000126 \*\*\*

sub\_channel\_enum卤味鸭脖 1.539e+01 1.155e+03 0.013 0.989372

sub\_channel\_enum冒菜/麻辣烫 8.307e-01 5.426e-01 1.531 0.125762

sub\_channel\_enum米粉面馆 1.754e+00 5.230e-01 3.353 0.000800 \*\*\*

sub\_channel\_enum奶茶果汁 1.529e+01 2.400e+03 0.006 0.994917

sub\_channel\_enum轻食沙拉 2.793e+00 9.071e-01 3.079 0.002078 \*\*

sub\_channel\_enum日料寿司 1.609e+00 8.183e-01 1.966 0.049261 \*

sub\_channel\_enum烧烤/小龙虾 2.174e+00 5.040e-01 4.314 1.60e-05 \*\*\*

sub\_channel\_enum生鲜 5.408e-01 8.070e-01 0.670 0.502805

sub\_channel\_enum特色小吃 1.808e+00 5.613e-01 3.220 0.001280 \*\*

sub\_channel\_enum甜品饮品 1.581e+01 1.164e+03 0.014 0.989166

sub\_channel\_enum西餐 5.506e-01 7.505e-01 0.734 0.463178

sub\_channel\_enum香锅干锅 1.168e+00 5.474e-01 2.133 0.032885 \*

sub\_channel\_enum炸鸡炸串 2.091e+00 5.609e-01 3.729 0.000193 \*\*\*

beer\_sku\_num 1.393e-02 3.949e-03 3.527 0.000420 \*\*\*

rating\_score\_num 1.727e-01 3.295e-02 5.242 1.59e-07 \*\*\*

night\_flg 1.396e-01 1.172e-01 1.191 0.233649

open\_sum -2.961e-03 1.329e-02 -0.223 0.823671

single\_pack\_unit瓶 -1.874e-01 2.004e-01 -0.935 0.349603

single\_pack\_unit桶 1.190e+01 1.369e+03 0.009 0.993066

single\_pack\_unit扎 -1.274e+00 7.702e-01 -1.654 0.098113 .

single\_pack\_num 3.148e-02 4.804e-02 0.655 0.512281

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Theta: count = 0.4336

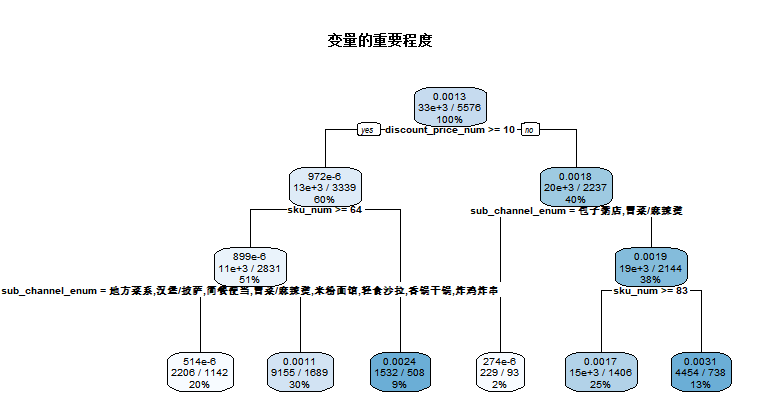
Number of iterations in BFGS optimization: 87

Log-likelihood: -1.08e+04 on 165 Df

模型解释：数据中0来源于二项分布,其他菜系相对于包子粥店,啤酒产品总数,评分的Odds multiplier显著且大于一,平均而言配送更少不含啤酒的订单；美团相对于饿了么的Odds multiplier则小于一，平均而言更少配送不含啤酒的订单；同时揭示了当单价增加1RMB，Odds（）减小到之前的exp(-1.462e-02)=0.9855倍。

数据中非0部分是建立在二项分布不为0的条件上的负二项分布，1664牌作为baseline，只有哈尔滨，汉斯，拳击猫，三得利啤酒，泰山和礼幌的有更大的平均转化率；值得注意的是相对于罐装，多数量的瓶装套装有更高的订单转化率，而扎啤则相反；美团平台的订单转化率更高；菜品数量上升，订单的平均转化率降低，而啤酒种类增多，转化率上升；就经营而言，夜间营业的店铺，啤酒订单所对应的转化率平均更高，但是经营时间的一味增加并不会使转化率提高。

(二)决策树描绘决定性特征



利用0.017的cp值(通过交叉验证法选取的等效树)剪枝得到如上决策树,可见在单价超过10的啤酒,在非包子粥/冒菜/麻辣烫店中, 菜品数低于83的平均而言拥有最高的订单转化率0.31%, 而在包子粥店, 冒菜店, 转化率惨淡, 几乎卖不出。而在啤酒单价低于10的店中, 菜品数量低于64的平均来说有第二高的转换率0.24%.

**结论与建议**

1. 对于特殊营销策略的讨论

我们无疑很感兴趣套餐这一特殊营销对订单转化率的影响, 然而在3000多家商铺中, 仅有48家使用了啤酒与食品组合成套餐这一营销策略 (算法---用正则表达式匹配其所在类别, 及产品在菜单上的名称) .

故将这些店铺挑出, 假定没有使用套餐, 用拟合的回归树及线性模型进行预测, 猜测不用套餐这一营销策略的预计啤酒订单. 如下图所示, 似乎套餐这一策略有一定的拉动作用, 然而数据集数量不够, 可能不具有严谨的统计意义。在实际运营时候应谨慎操作。

图片包含 照片, 游戏机, 桌子, 不同

描述已自动生成图片包含 游戏机, 照片, 桌子, 不同

描述已自动生成

1. 对模型不足的讨论

文中使用的二项分布模型，一定程度上解决了overdispersion的问题，但是模型的解释性仍有很多提升的空间，可以通过收集更多有关店面口味，或者营业街区等特征，得以更加精确地讨论订单的转化。

# 主要参考文献

[1] 吴学福刘振球 零膨胀计数资料几种模型方法的比较研究 <https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFQ&dbname=CJFDAUTO&filename=ZGWT202003003&uid=WEEvREcwSlJHSldRa1FhcTdnTnhYWU9Ba04yOWdveGtXeGF5bGY5ODRYUT0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4IQMovwHtwkF4VYPoHbKxJw!!&v=MjAzNzExVDNxVHJXTTFGckNVUjdxZllPZG5GaXZoVmI3TlB5cmNlckc0SE5ITXJJOUZaNFI4ZVgxTHV4WVM3RGg=>

[2] 陈辉林 夏道勋 基于CART决策树数据挖掘算法的应用研究 <https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFQ&dbname=CJFD2011&filename=MTJS201110071&v=MTYyNDFyQ1VSN3FmWU9kbkZpdmxVN3JNS0RuQmZiRzRIOUROcjQ5Q1pZUjhlWDFMdXhZUzdEaDFUM3FUcldNMUY=>

**[3]**颜贤斌 浅谈中国餐饮外卖O2O模式的现状与发展<https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFQ&dbname=CJFDLAST2015&filename=SCXH201522027&uid=WEEvREcwSlJHSldRa1FhcTdnTnhYWU9Ba053MU9SQlVOWFRmQTZVMmZlMD0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4IQMovwHtwkF4VYPoHbKxJw!!&v=MDQ0MzFuRmlybVU3N0xOaTdUWnJHNEg5VE9yWTlIWTRSOGVYMUx1eFlTN0RoMVQzcVRyV00xRnJDVVI3cWZZT2Q=>

**[4]** 韦 博 成 林 金 官 吕 庆 哲 回 归 模 型 中 异 方 差 或 变 离 差 检 验 问 题 综 述

<https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFQ&dbname=CJFD2003&filename=YYGN200302014&uid=WEEvREcwSlJHSldRa1FhcTdnTnhYWU9Ba05wLzlTL3h5UG43cEFJUmc0UT0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4IQMovwHtwkF4VYPoHbKxJw!!&v=MDEwMjl0TE1yWTlFWUlSOGVYMUx1eFlTN0RoMVQzcVRyV00xRnJDVVI3cWZZT2RuRml6bVdyM0JQRFRNWUxHNEg=>

1. 吴学福刘振球 零膨胀计数资料几种模型方法的比较研究 <https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFQ&dbname=CJFDAUTO&filename=ZGWT202003003&uid=WEEvREcwSlJHSldRa1FhcTdnTnhYWU9Ba04yOWdveGtXeGF5bGY5ODRYUT0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4IQMovwHtwkF4VYPoHbKxJw!!&v=MjAzNzExVDNxVHJXTTFGckNVUjdxZllPZG5GaXZoVmI3TlB5cmNlckc0SE5ITXJJOUZaNFI4ZVgxTHV4WVM3RGg=> [↑](#footnote-ref-1)
2. 陈辉林 夏道勋 基于CART决策树数据挖掘算法的应用研究 <https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFQ&dbname=CJFD2011&filename=MTJS201110071&v=MTYyNDFyQ1VSN3FmWU9kbkZpdmxVN3JNS0RuQmZiRzRIOUROcjQ5Q1pZUjhlWDFMdXhZUzdEaDFUM3FUcldNMUY=> [↑](#footnote-ref-2)
3. 颜贤斌 浅谈中国餐饮外卖O2O模式的现状与发展 <https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFQ&dbname=CJFDLAST2015&filename=SCXH201522027&uid=WEEvREcwSlJHSldRa1FhcTdnTnhYWU9Ba053MU9SQlVOWFRmQTZVMmZlMD0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4IQMovwHtwkF4VYPoHbKxJw!!&v=MDQ0MzFuRmlybVU3N0xOaTdUWnJHNEg5VE9yWTlIWTRSOGVYMUx1eFlTN0RoMVQzcVRyV00xRnJDVVI3cWZZT2Q=> [↑](#footnote-ref-3)
4. 韦博成林金官吕庆哲 回归模型中异方差或变离差检验问题综述

   <https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFQ&dbname=CJFD2003&filename=YYGN200302014&uid=WEEvREcwSlJHSldRa1FhcTdnTnhYWU9Ba05wLzlTL3h5UG43cEFJUmc0UT0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4IQMovwHtwkF4VYPoHbKxJw!!&v=MDEwMjl0TE1yWTlFWUlSOGVYMUx1eFlTN0RoMVQzcVRyV00xRnJDVVI3cWZZT2RuRml6bVdyM0JQRFRNWUxHNEg=> [↑](#footnote-ref-4)