

---

## 基于打折力度概念的“薄利多销”模型

### 摘要

本文主要对某商场 2016 年 11 月 30 日至 2019 年 1 月 2 日的销售流水相关数据进行分析,通过汇总、拼接、删失(删除缺失值)、排序、补全缺失值等数据处理得到商场每日销售额和利润率,然后选取商品折扣率、商品利润率以及折扣响应度指标,计算出商场每日最终的打折力度。依据每日商品营业额、利润率、打折力度,建立了打折力度与商品营业额、利润率的鲁棒线性回归模型,并进一步分析了商品类别对该模型的影响。

针对问题一,为了计算出商场每天的营业额和利润率,本文主要根据附件 1 附件 2 的销售流水记录来进行分析。对于大部分非打折商品成本价缺失(共 783524 条记录缺失成本价),本文采用两种方式补全成本价缺失值。第一种方式利用商品 ID 遍历成本价未缺失的历史流水数据,从而填补了 71.46%的缺失部分;对于在历史数据查询不到商品 ID 的缺失记录,依据先验信息零售商利润率一般在 20%到 40%之间,第二种方法则直接生成随机的商品利润率(0.2 到 0.4),然后再根据销售价反算出缺失的商品成本价。最终依据完整的销售流水记录表汇总得到所有商品每日的营业额和对应的利润率。

针对问题二,根据附件 3 的折扣信息表,利用促销时间、价格、数量等变量,计算出商品折扣率、商品利润率以及折扣响应度指标,利用较为简单直观的公式输出一个粗粒度的打折力度,并与问题一数据融合计算出商场每日最终的打折力度。

针对问题三,根据问题一和问题二对商场每天的销售额和利润率与打折力度,利用迭代加权最小二乘法,建立了打折力度为自变量、销售额利润率为因变量的线性回归模型,结果表明:打折力度与销售额存在正相关关系,打折力度与利润率存在负相关关系。

针对问题四,通过附件 4 的商品信息表,对不同类别商品的价格弹性进行分析,按照问题一与问题二的计算方式来作图分析,考察不同类别商品下对该问题结果的影响。发现价格弹性高的商品,对打折力度的强度反应很灵敏,而弹性低的商品类别,销售额与打折力度之间没有很强的相关性,且利润率变化相对平缓。

**关键词:** 打折力度 营业额 利润率 迭代加权最小二乘法 需求价格弹性

---

## 一、问题重述

打折促销，是“薄利多销”策略的一种应用。“薄利多销”，顾名思义就是通过降低商品价格，吸引更多的消费者达到更多的销售量，虽然单品利润降低但总体利润增加。日常生活频现打折活动，打折不仅可以促进本身打折品的销售，更妙地是它可以吸引更多潜在的消费者，形成对同商场其他未打折商品的购买力。

然而，打折必须精准定位，不同打折策略对商场营业额和利润率会有不同的影响。由经济学需求价格弹性概念知道<sup>[1]</sup>，薄利多销策略只对价格弹性高的商品有效，打折过低或过高都可能适得其反<sup>[2]</sup>，而任何商品都打折等盲目打折行为甚至可能造成亏损<sup>[3]</sup>。因此，确立打折力度与营业额、利润率的关系十分重要。

如前所述，依据零售商超市粗糙的数据流水记录数据，确立打折力度与营业额、利润率的关系，挖掘更有效精准的打折策略，必须先清洗、整合干净的交易数据，选取合适的指标量化“打折力度”这一概念。建立三者的相关关系，进而分析不同类别的商品对模型关系的影响。

所用数据集介绍：附件 1 和附件 2 是某商场自 2016 年 11 月 30 日起至 2019 年 1 月 2 日的销售流水记录，附件 3 是折扣信息表，附件 4 是商品信息表，附件 5 是数据说明表。因此，所要解决问题如下：

- (1) 计算该商场从 2016 年 11 月 30 日到 2019 年 1 月 2 日每天的营业额和利润率
- (2) 选取合适指标衡量商场每天打折力度，并计算商场从 2016 年 11 月 30 日到 2019 年 1 月 2 日每天的打折力度。
- (3) 分析打折力度与商品销售额以及利润率的关系。
- (4) 考虑商品的大类区分，打折力度与商品销售额以及利润率的关系有何变化？

## 二、基本假设

- (1) 假设 ID 相同的商品三年内成本价变化不大
- (2) 假设商品成本的利润率一般在 20%-40%之间
- (3) 忽略单品买赠对评估打折力度的影响
- (4) 假设当天内等时间段的打折力度相等，即打折力度一天对时间段无差别
- (5) 假设每天打折力度是当天所有商品打折力度之和

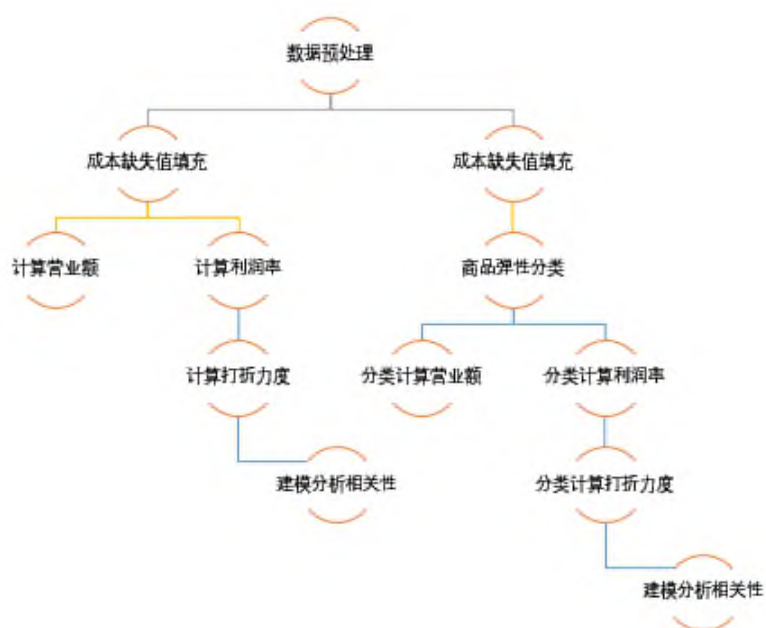
### 三、符号说明

为简化问题分析和数据处理，对符号做如下规定：

<b>Y :</b>	每日营业额
<b>R:</b>	每日利润率
<b>n:</b>	销售数量
<b>Sp:</b>	商品实际销售价格
<b>Cp:</b>	商品成本价
<b>Bp:</b>	商品标价
<b>Dp:</b>	打折力度
<b>Da:</b>	打折活跃度
<b>Dr:</b>	打折率
<b>Lc:</b>	限购数量
<b>Sc:</b>	购买数量

### 四、模型的建立与求解

本节主要求解四个小问，其求解思路过程流程图如下：



## 4.1 问题一的建模与解答

### 4.1.1 商品销售数据可视化探索

通过对附件 1 和附件 2 的 1221855 多条记录进行分析（见图 1），发现在这些订单中，有近 8% 的订单属于无效订单（96674 条），无效订单并没有形成实际购买，对于每日销售额  $Y$  和利润率  $R$  的计算是无用的，因此本小问计算对这些记录进行了删除处理。

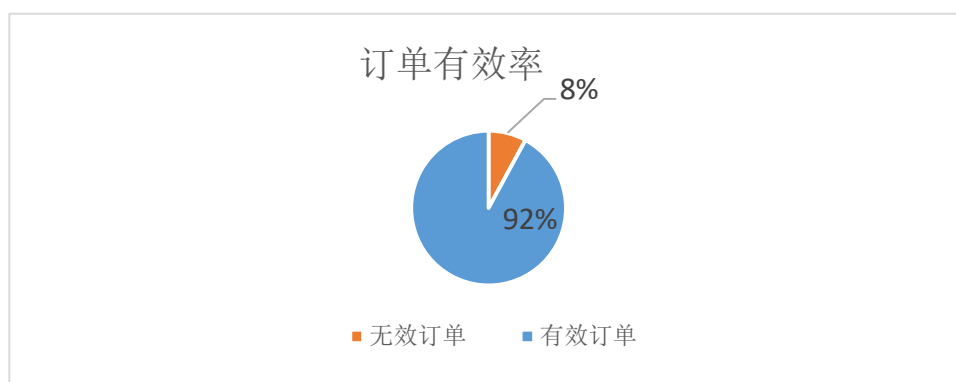


图 1

计算有效订单中每天商品的营业额，再对每天所有商品的营业额进行求和得到每日营业额  $Y$ 。要计算每日的总利润率，需知道每日各个销售商品的成本价，再依据成本价  $C_p$ 、实际销售价  $S_p$  计算。经统计附件一、二中销售流水记录中有有效订单的成本价缺失严重，有效订单中成本价缺失条数 720850，占比约 64%（见图 2），需想办法对这些成本价缺失值记录填充再行计算日利润率  $R$ 。

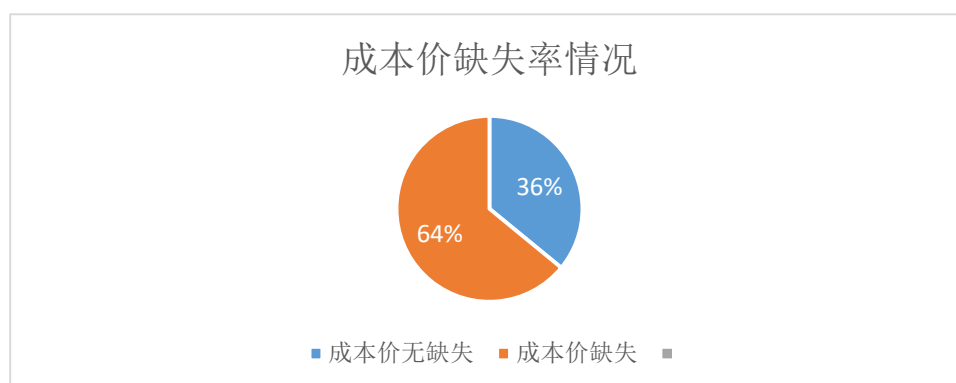


图 2

填补成本价缺失的方法如前文所述，首先根据商品 ID 一一对应的方法，在有成本价的历史订单中遍历（由假设 1 三年内相同商品 ID 的成本价相差不大），从而补全部分缺失商品与历史订单相同商品 ID 的成本价，该方法共填充缺失条目数为 519901，值得说明的是该方法历史遍历往往发现多条记录，本文选择与之售价或标价相同的记录对应的成本价。对于剩余的 2 万左右的成本价缺失记录，本文依据题目中所给假设 2，商品零售利润率在 20%~40% 之间，随机生成商品利润率  $r_i \in [0.2, 0.4]$ ，再依据商品利润率和商品标价反算出缺失条目的成本价。

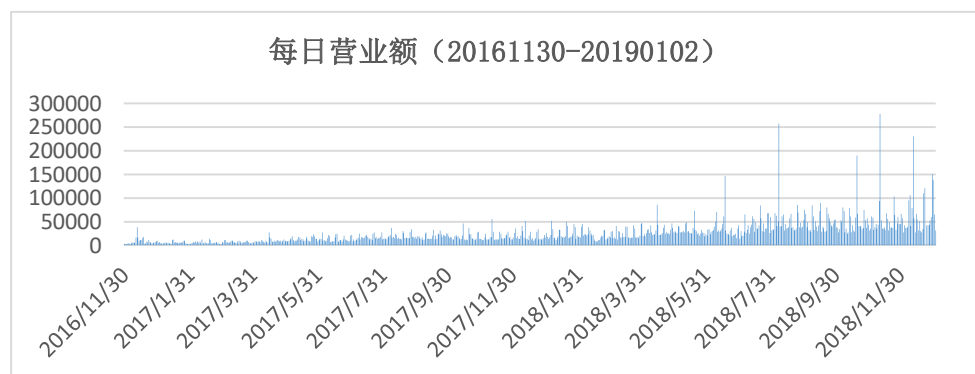
#### 4.1.2 问题一的求解

问题要求结合商品销售流水记录，利用 SAS 将附件 1 与附件 2 进行拼接并按照日期排序，筛选出有效订单(is\_finished=0 表示该订单记录但并未成功付款，视为未成功订单)，根据有效订单的单品销售价及销售量，计算出指定时间段的每天营业额 Y。每天营业额 Y 计算公式如下：

$$Y = \sum_{i=1}^I Sp_i \times n_i \quad (1),$$

其中  $Sp_i$  商品 i 的销售价格， $n_i$  为商品 i 的销售量，I 为当天不同商品销售总记录条数。

由公式(1)计算，2016 年 11 月 30 至 2019 年 1 月 2 日每日的营业额图形如下：



在利润率的计算上，本文先利用商品 ID 在设定随机数来补全缺失成本价格，设置 0.2~0.4 之间的随机数利润率,每日利润率 R 的计算公式如下：

$$R = \frac{Y}{\sum_{i=1}^I Cp_i \times n_i} - 1 \quad (2),$$

其中 Y 为已计算好的日营业额， $Cp_i$  为商品 i 的成本价， $n_i$  为商品 i 的销售量，I 为当天不同商品销售记录条数。商场每日营业额 Y 和利润率 R 计算如下表：(所有计算数据参见文件“问题 1 营业额及利润率.xlsx”)

时间	日营业额(Y)	日利润率(R)
30Nov2016	2833.70	11.74%
1Dec2016	2346.20	19.47%
2Dec2016	2338.30	17.09%
3Dec2016	3270.40	16.21%
4Dec2016	3881.68	15.77%

5Dec2016	2362.60	22.77%
6Dec2016	2253.10	19.8%
7Dec2016	5312.88	16.38%
8Dec2016	5890.72	16.58%
9Dec2016	5750.58	17.59%
10Dec2016	5755.26	16.02%
11Dec2016	16415.22	17.09%
12Dec2016	37236.08	15.45%
13Dec2016	16547.00	15.82%
14Dec2016	9172.90	17.91%
15Dec2016	11167.40	17.21%
16Dec2016	11062.28	16.95%
17Dec2016	14744.44	15.61%
18Dec2016	17424.66	16.17%
19Dec2016	3151.70	17.52%

表 1

## 4.2 问题二的建模与解答

问题二要求建立合适的指标量化打折力度,并计算商场 2016 年到 2019 年期间每天的打折力度。由日常经验知,五折比九折的打折力度要高,因此打折率指标可以反应单品的打折力度,单品打折率  $Dr=Sp/Bp$ (售价/标价)。

另外,同样的折扣率的两个商品,如 A 商品成本 1 元,门店标价 10 元,打五折实际售价 5 元;B 商品成本 2 元,门店标价 10 元,打五折售价 5 元,消费者应当购买 B 商品才更实惠,因为商场实际让利给消费者更多,打折力度更大,即同样的折扣率、原始利润率更低的商品折扣力度更大(商家利润率低,让利给消费者的实惠就更多,二者是竞争关系)。因此为简化问题,相同折扣率的商品,利润率  $r$  更低的商品打折力度更大,选择利润率  $r$  作为第二个量化打折力度的指标。

最后,由附件 3 限购数量和实际销售数量知,消费者用“脚”投票,实际销售量越接近限购数量,该打折活动的消费者活跃度越高,表示该商品打折活动很奏效,打折力度很大,折扣活跃度也应纳入衡量打折力度的指标中。由上述,这三个指标综合考量才形成打折力度的合理量化。

综上分析,在日期不存在附件 3 折扣信息表情形下,本文采用如下公式尝试计算单品打折力度  $Dp$ :

$$Dp = \frac{1 - Dr}{r} \quad (3),$$

公式(3)恰能反应折扣力度与折扣率、利润率的关系,且分子分母均在(0,1),计算数值变化不大,可以作为折扣力度的衡量。由于题中要求出 2016 年 11 月 30 号到 2019 年 1 月 2 号每日打折力度,而折扣信息表中只有 2017 年 3 月份以后的折扣信息数据,本文结合由附件 1 和附件 2 数据补全 2016 年 11 月 30 号至 2017 年 3 月份的打折力度(该时间段的打折力度按照公式(3)计算,并汇总成每日打折

力度值), 对于附件 3 折扣信息表中存在限购数量  $L_c$  和购买数量  $S_c$ , 其打折力度公式计算如下:

$$Dp = \frac{1 - Dr}{r} + \frac{Sc}{Lc} \quad (4),$$

数据观察发现, 由于限购数通常还是很大的, 公式(4)中第二项  $\frac{Sc}{Lc}$  实际很小, 当接近 1 时表明打折活跃度也很高, 这里线性相加反应在折扣力度  $Dp$  中。最后, 本文将这(3)、(4)两种计算结果融合, 拼接成总表 2 如下: (所有计算数据参见文件“问题 2 商品折扣力度.xlsx”)

时间	营业额 Y	利润率 R	打折力度 $Dp$
30Nov2016	2833.70	0.115763026	2.02806
1Dec2016	2346.20	0.192372413	1.35932
2Dec2016	2338.30	0.175281984	1.36281
3Dec2016	3270.40	0.16535232	1.48605
4Dec2016	3881.68	0.156719561	1.38741
5Dec2016	2362.60	0.228914204	1.13271
6Dec2016	2253.10	0.190059093	1.47578
7Dec2016	5312.88	0.167720832	1.43409
8Dec2016	5890.72	0.170971981	1.51741
9Dec2016	5750.58	0.178246114	1.29463
10Dec2016	5755.26	0.16082193	1.40768
11Dec2016	16415.22	0.170418126	1.29086
12Dec2016	37236.08	0.154816697	1.48759
13Dec2016	16547.00	0.160560791	1.38423
14Dec2016	9172.90	0.180690837	1.24069
15Dec2016	11167.40	0.171880836	1.38703
16Dec2016	11062.28	0.171506814	1.48392
17Dec2016	14744.44	0.155224881	1.56064
18Dec2016	17424.66	0.162251897	2.02838
19Dec2016	3151.70	0.181803678	1.44158

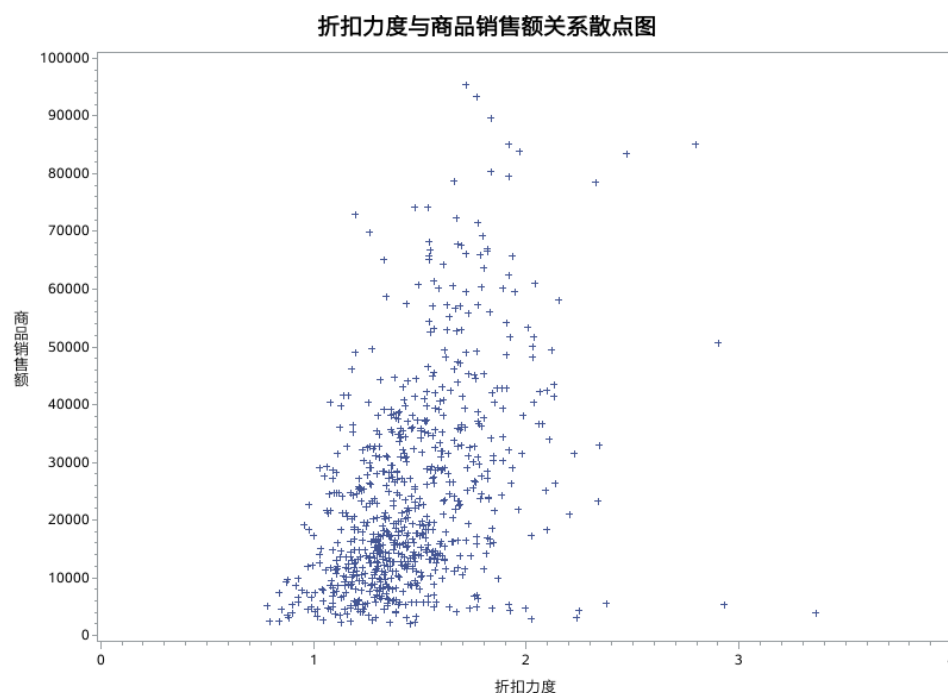
表 2

### 4.3 问题三的建模与解答

分析打折力度与商品销售额及利润率与之间的关系, 主要依赖于问题一和问题二得到的每日营业额、利润率、以及打折力度信息, 即 4.2 中表 2。首先通过 SAS 分别对打折力度与商品销售额、打折力度与利润率进行相关性分析。

#### 4.3.1 打折力度与销售额关系

首先依据表 2 画出打折力度与销售额的散点图，如下所示：



可以看出打折力度与销售额呈现一定的正相关性，下面通过皮尔逊相关性检验<sup>[4-5]</sup>来验证这种关系：

Pearson 相关系数, N = 752 Prob >  r  under H0: Rho=0		
	_COL2	rate
_COL2 营业额	1.00000	0.44020 <.0001
rate 折扣力度	0.44020 <.0001	1.00000

Pearson 相关统计量 (Fisher z 转换)							
变量	与变量	N	样本相关系数	Fisher z	95% 置信限		H0:Rho=0 的 p 值
_COL2	rate	752	0.44020	0.47248	0.380686	0.496080	<.0001

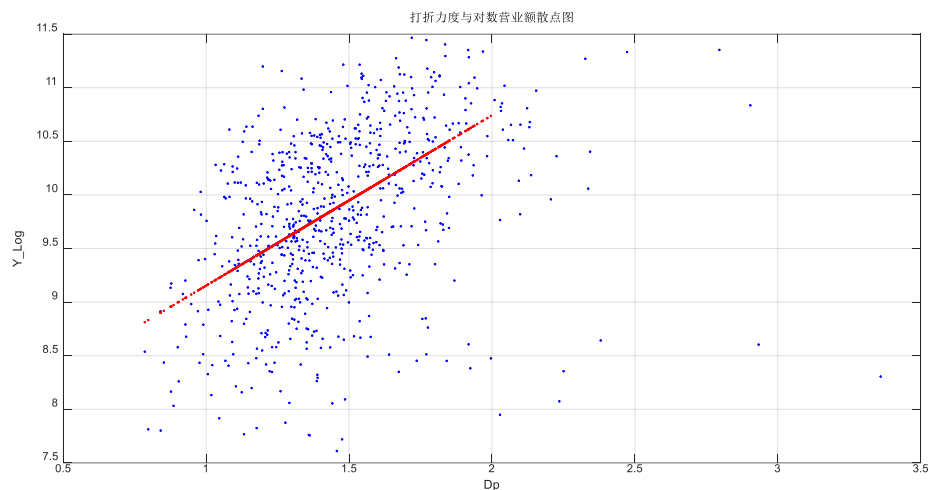
皮尔逊相关性检结果  $p$  值很小 ( $p < 0.0001$ )，说明它们确实具有相关性<sup>[4-5]</sup>；且样本相关系数为  $0.4402 > 0$ ，所以打折力度与营业额具有正相关性<sup>[4-5]</sup>。

对打折力度和营业额做线性回归，为防止异常值和噪声影响，本文使用迭代加权最小二乘法做回归（有关迭代加权最小二乘法更多介绍参加<sup>[6]</sup>），迭代加权最小二乘法拟合回归关系对噪声和异常值更稳健<sup>[6]</sup>，matlab 中可以直接调用函数 `robustfit()`实现，这里加权函数本文默认选择的双平方加权函数。为了消除营业额数值过大与打折力度量纲不同，本文对  $\ln(Y)$ 和  $D_p$  做回归，得到回归方程如下：

$$\ln(Y) = 1.59D_p + 7.57 \quad (5),$$

拟合的回归关系(红线)与对数营业额散点图如下：

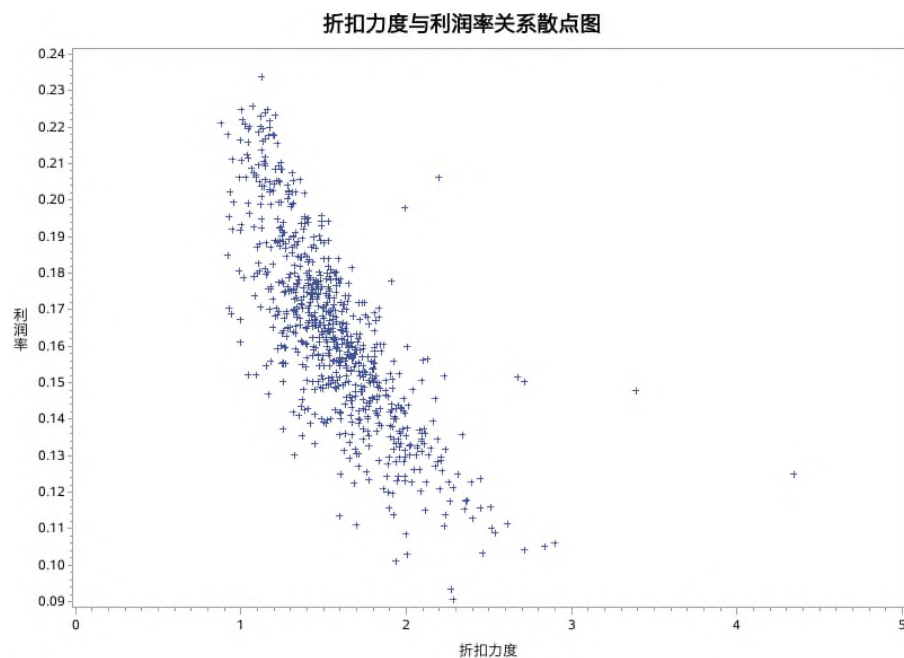




值得注意的是,当打折力度大于 2.5 后,营业额与打折力度反而不满足正相关性,这说明打折力度过大,打折策略不当反而损害了商场的利益。

#### 4.3.2 打折力度与商品利润率的关系

首先依据表 2 画出每日打折力度与每日利润率的散点图,如下所示:



可以看出打折力度与利润率有明显的负相关性,下面通过皮尔逊相关性检验来验证这种关系:

Pearson 相关系数, N = 752 Prob >  r  under H0: Rho=0		
	lr	rate
lr 利润率	1.00000	-0.73343 <.0001
rate 折扣力度	-0.73343 <.0001	1.00000

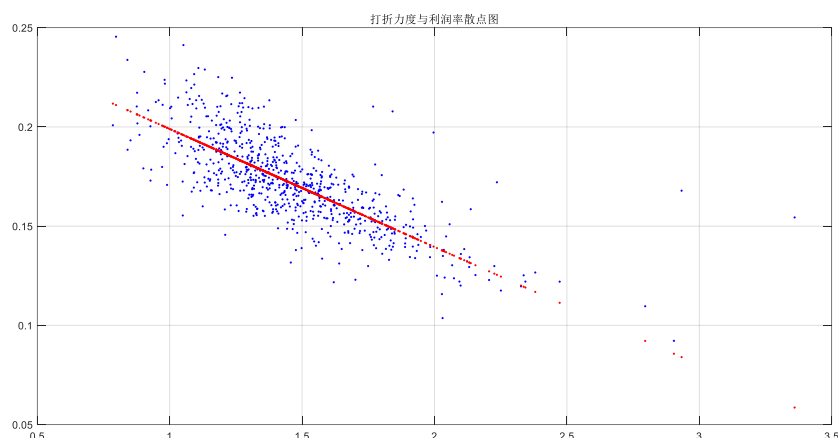
Pearson 相关统计量 (Fisher z 转换)							
变量	与变量	N	样本相关系数	Fisher z	95% 置信限		H0:Rho=0 的 p 值
lr	rate	752	-0.73343	-0.93612	-0.764822	-0.698569	<.0001

皮尔逊相关性检结果  $p$  值很小 ( $p < 0.0001$ ), 说明它们确实具有相关性<sup>[4-5]</sup>; 且样本相关系数为  $-0.73343 < 0$ , 所以打折力度与利润率具有负相关性<sup>[4-5]</sup>。

对打折力度  $Dp$  和利润率  $R$  做线性回归, 同样地, 为防止异常值和噪声影响, 本文使用迭代加权最小二乘法做回归<sup>[6]</sup>, 调用 matlab 函数 `robustfit()`, 加权函数选择默认的双平方加权函数, 得到  $Dp$  和  $R$  的回归方程如下:

$$R = -0.059Dp + 0.26 \quad (6),$$

拟合的回归关系 (红线) 及对应散点图如下:



总结: 由回归方程 (5), (6) 及图形可看出, 随着打折力度的增加, 商场每日的营业额是增加的, 而商品每日的利润率是降低的。这表明, 通过降低单个的利润率, 确实带动了销售量的增加, 进而营业额增加了, 实现了“薄利多销”的策略。但也注意到, 商场在这两年多的时间内, 也有一些天数折扣力度过大, 反而导致营业额不增反降了。

#### 4.4 问题四的建模与解答

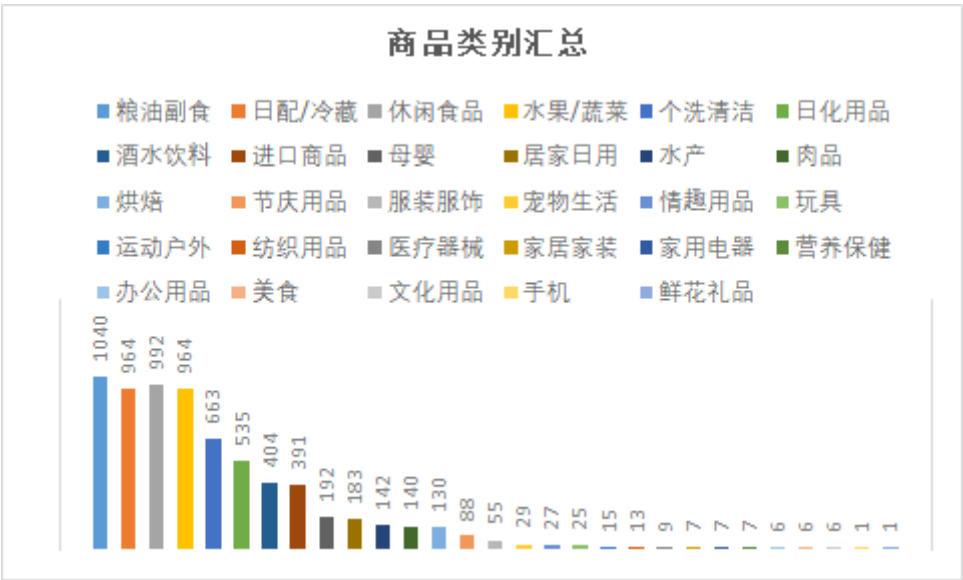
问题四是在问题三的基础上, 进一步考虑商品的大类区分下, 不同类别的商品打折力度与商品销售额以及利润率的关系会发生什么样的变化。附件四对商场里面的各种商品做了分类, 在处理这个问题时, 本文主要将附件四中各种商品对应的类别跟总商品记录表中的商品一一对应合并, 再对合并后的总表进行分类,

在这里本文将所有商品按照不同类别分成弹性大于 1、弹性小于 1 两种类型，再分别按照问题一、问题二的方式，计算出对应的销售额、利润率以及打折力度。最后再根据计算出来的数据对打折力度与销售额、利润率之间的关系进行分析。

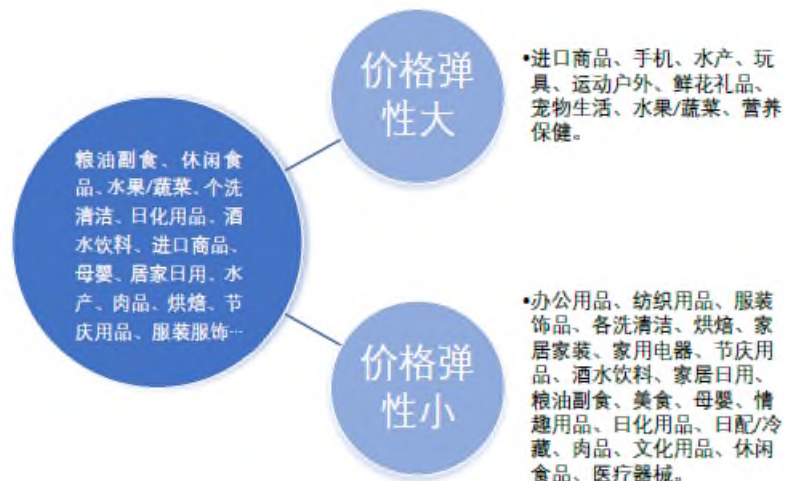
4. 4. 1 概念引入与数据分析

价格弹性<sup>[7]</sup>：是衡量由于价格变动所引起数量变动的敏感度指标。当弹性系数为 1 的时候，销售量的上升和价格的下降幅度是相抵的。弹性在 0 到 1 之间时，这类物品的需求是相对缺乏弹性的，或者说价格不敏感。而弹性大于 1 的时候，则说明价格变动会很敏感。大多数食品的需求弹性是低的，而大多数的奢侈品的需求弹性都相对较高。

根据附件三给出的商品分类数据，通过对类别进行分析，得到各类别下商品数量情况如下：

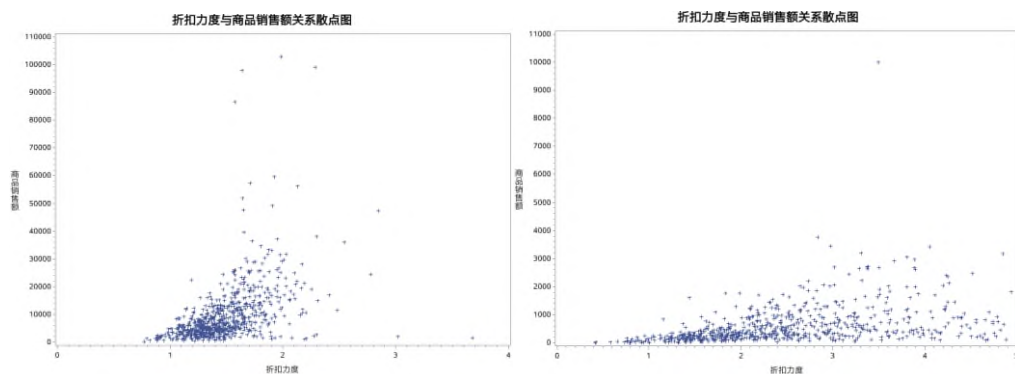


根据价格弹性定义<sup>[7]</sup>，为了便于对问题的分析，主要根据商品的时间、可替代物品、重要性、预算以及用途等<sup>[8]</sup>方面来对商品进行弹性类别划分。本文将以上类别涉及到的商品大类主要分为弹性大、弹性小两种类别。



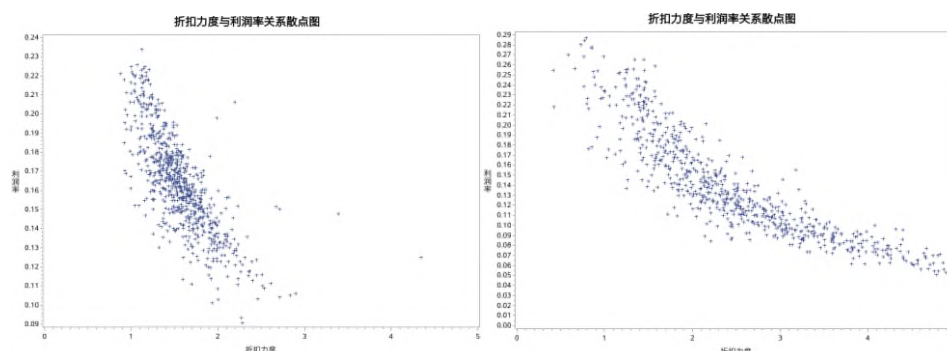
根据这种分类，本文对两种弹性不同的商品分别按照打折力度与销售额和利润率之间的关系进行分析。

#### 4.4.2 打折力度与销售额的关系



左上图是价格弹性高的商品打折力度与销售额之间的关系散点图；右上图是价格弹性低的商品打折力度与销售额之间的关系散点图，可以看出，打折力度对商品弹性大的这类商品来说影响较大，力度越大，销售额越高，正相关性明显，而对于价格弹性低的商品，销售额与打折力度之间的关系不是很明显。

#### 4.4.3 打折力度与利润率的关系



如上图所示，左上图是价格弹性高的商品打折力度与利润率之间的关系散点图；右上图是价格弹性低的商品打折力度与利润率之间的关系散点图，可以直观地看出，打折力度对商品弹性大的这类商品的利润率来说更敏感，打折力度越大，利润率下降趋势更陡峭，而对于价格弹性低的商品，随着打折力度的增加，利润率的降低相比之下就有些平缓。

之所以出现价格弹性高低不同，打折力度与商品销售额和利润率之间的关系不同这种现象，主要是因为购买者对不同消费商品的弹性不一样，比如日常食品粮油这些商品，由于可替代性较差，又是生活中的必需品，所以价格的波动对购买者的意愿产生的影响并不大，而像鲜花、手机这种产品，购买者对价格就会显得更加敏感，属于弹性高的商品，打折力度变大，则消费者购买意愿变强。

## 五、模型评价及推广

本文针对题目给出的零散繁多的流水记录，汇总计算出了商场每日营业额，在数据质量存在大量缺失成本值的情况下给出了两种补全方式，从而计算出商场每日利润率。针对问题二要求，结合日常生活经验，本文给出了“打折力度”这一新概念的计算公式，并计算了每日的打折力度值。由皮尔逊相关检验，探索验证打折力度、营业额、利润率变量间的相关关系并建模，结果表明打折力度与商场营业额和利润率相关关系显著。随着打折力度的增加，商场营业额整体是增加的，而利润率是降低的，证实了“薄利多销”原理在生活中的应用。最后，对不同类别商品的价格弹性进行分析，按照问题一与问题二的计算方式来作图分析，考察不同类别商品下对该问题结果的影响。发现价格弹性高的商品，对打折力度的强度反应很强，而弹性低的商品类别，则不会因为打折力度低而减少销售量。

### 5.1 模型的优缺点

零售行业流水数据相似度很高，本文处理该数据并建模的流程清晰，预处理

---

数据（筛选、删失、填充缺失、拼接表），可视化探索，相关性分析，数据建模等，适合推广到其他零售超市流水数据的分析。然而，打折力度的计算公式，本文只是依据生活常识，给出的打折力度虽不违背直观，仍有更多尝试探索的空间，因为“打折力度”本身这一概念并不清晰，更多的计算尝试要一同与建模考量，最终打磨出一个令各方满意的“打折力度”概念和计算公式。

## 5.2 模型推广

在分析打折力度和营业额、利润率时，这里为便于简化，只使用了迭代加权最小二乘的线性回归模型建模，后期也可以尝试更多复杂度更高的模型求解，这样对于不同打折力度下，对商场利润率和营业额有个更准确预测，便于指导商场零售打折策略的应用(应当在何时采取多强的打折力度获利最丰)。在问题四中，本文只是定性的分析了不同价格弹性的商品类别下，表明弹性高的商品对打折力度更敏感而弹性低的商品对打折力度不敏感。由于数据有限，无法计算出每种商品需求价格弹性。基于打折力度探究营业额和利润率的模型，若将每种商品需求价格弹性量化并考虑进来，其应用性和实用性将更为广泛、强大。

## 参考文献

- [1] 高鸿业.《西方经济学》.上册,微观部分[M]. 中国经济出版社, 1996.
- [2] 联商. 商场打折的运作策略与艺术[J]. 商业时代, 2008(18):14-14.
- [3] 史永卿, 郭钢, 文方. 试论商场打折盈利营销策略的有效性[J]. 山西财经大学学报, 2007(S2):65+77.
- [4] 吴喜之. 统计学：从数据到结论[M]. 中国统计出版社, 2005.
- [5] 陈珍珍著. 统计学：厦门大学出版社, 2002 年.
- [6] 杰夫·吉尔. 广义线性模型：一种统一的方法[M]. 格致出版社, 2015.
- [7] <https://baike.baidu.com/item/价格弹性>
- [8] 陈通. 宏观微观经济学[M]. 天津大学出版社, 2003.

---

## 附录：SAS 源代码

主程序

/\*Q1\*/

/\*读取数据 附件 1 附件 2\*/

```
proc import out=data11
```

```
  datafile="C:\Users\hr\Desktop\E-2019 中文\data\附件 11.xlsx"
```

```
  dbms=excel replace;
```

```
  sheet="a";
```

```
  getnames=yes;
```

```
run;
```

```
proc import out=data22
```

```
  datafile="C:\Users\hr\Desktop\E-2019 中文\data\附件 22.xlsx"
```

```
  dbms=excel replace;
```

```
  sheet="b";
```

```
  getnames=yes;
```

```
run;
```

/\*label 赋值\*/

```
data test1_1;
```

```
  set data11(rename=(sku_id=id sku_name=name sku_prc=prc sku_sale_prc=sale  
                    sku_cost_prc=cost sku_cnt=cnt create_dt=dt));
```

```
  keep id name is_finished prc sale cost cnt dt;
```

```
run;
```

```
data test2_1;
```

```
  set data22(rename=(sku_id=id sku_name=name sku_prc=prc sku_sale_prc=sale  
                    sku_cost_prc=cost sku_cnt=cnt create_dt=dt));
```

```
  keep id name is_finished prc sale cost cnt dt;
```

```
run;
```

/\*is\_finished 为 0 时删除 因为对分析没有意义，外部因素太多\*/

```
data cost1_miss cost1_nomiss;
```

```
  set test1_1;
```

```
  where is_finished ne 0;
```

```
  if cost ne 0 then output cost1_nomiss;
```

```
  else if cost eq 0 then output cost1_miss;
```

```
run;
```

```
data cost2_miss cost2_nomiss;
```

```
  set test2_1;
```

```
  where is_finished ne 0;
```

```
  if cost ne 0 then output cost2_nomiss;
```

```
  else if cost eq 0 then output cost2_miss;
```

---

```

run;

/*将 excell excel2 的成本缺失和非缺失的 set 到一起*/

data cost_nomiss;
    set cost1_nomiss cost2_nomiss;
run;
data cost_miss;
    set cost1_miss cost2_miss;
run;
/*将非缺失的成本 连接 到缺失成本里面      这种方式不是很合理 有很多无法匹配
*/
/*proc sort data=cost_nomiss out=nomiss(keep=dt id cost prc ) nodupkey;*/
/*    by dt id cost ;*/
/*run;*/
/*proc sort data=cost_miss out=miss nodupkey;*/
/*    by dt id name cost prc;*/
/*run;*/
/**/
/*proc sql;*/
/*    create table miss_add as*/
/*    select  a.*,b.cost as cost_add from miss as a left join nomiss as b*/
/*    on      a.id=b.id          and          a.prc=b.prc          and
substr(put(a.dt,yymmdd10.),1,7)=substr(put(b.dt,yymmdd10.),1,7)  ;*/
/*quit;*/
/**/
data miss_add;
    set cost_miss(drop=cost);
    if sale ne 0 then cost=sale/1.3;
    else cost=prc/1.3;
run;
/*用 SAS 生成 1.2 到 1.4 之间的随机数代替上面 1.3*/
data rand;
    set cost_miss;
    do i=0 to 1;
        rand=0.2*ranuni(i)+1.2;
    end;
run;

data miss_add;
    set rand(drop=cost);
    if sale ne 0 then cost=sale/rand;

```



---

```

        else cost=prc/rand;
run;
data all;
    set cost_nomiss miss_add;
run;

data all_;
    set all;
/*营业额*/
    allin=sale*cnt;
/*成本总额*/
    costsum=cost*cnt;
/*定价总额*/
    prcsum=prc*cnt;
/*利润率*/
    lirun=(sale-cost)/cost;
run;

/*保存 Q1 用到的总数据*/
libname data "D:\E-2019 中文" ;

data data.Q1_data;
    set all_;
run;

/*计算每天营业额 利润率*/

proc sort data=all_ ;
    by dt;
run;

proc sql;
    create table final0 as
        select distinct dt,sum(allin) as allin_sum,sum(prcsum) as prc_sum,sum(costsum) as
cost_sum from
        all_
        group by dt
        order by dt;
quit;

data final;

```

---

```

        set final0;
        lirun_sum=( allin_sum-cost_sum)/cost_sum*100||'%';
        lirun=put((allin_sum-cost_sum)/cost_sum,7.4)*100||'%';
        lr=( allin_sum-cost_sum)/cost_sum;
/*      lirun_sum_=put();*/
        label dt='时间'
              lirun_sum='利润率'
              prc_sum='原价总额'
              cost_sum='成本总额'
              lirun='利润率保留两位小数'
              allin_sum='营业额';
        format dt yymmdd10. lirun_sum $200.;

run;

data data.Q1_final;
    set final;

run;
proc means data=final  mean ;
    var lr;
run;

ods excel file="C:\Users\hr\Desktop\E-2019 中文\data\附件 1done.xlsx"    options
(sheet_name='营业额_利润率');
proc print data=final label;
run;
ods excel close;
/*****Q1  end*****/

/*输出问题 2 需要的数据*/

proc sort data=all_ ;
    by dt;
run;

proc sql;
    create table final2 as
        select distinct dt,sum(allin) as allin_sum,sum(prcsum) as prc_sum,sum(costsum) as

```

---

```

cost_sum from
    all_
/*需要筛选出打折的商品*/
    where prc gt sale
    group by dt
    order by dt;
quit;
/*计算利润率折扣率之比*/
data final_;
    set final2;
    drop lirin_sum lirin;
run;
/*将之前问题 1 的折扣率 merge 过来*/
proc sql;
    create table final3 as
        select final_.*,final.lirin_sum,final.lr from final_ left join final
            on final_.dt=final.dt
        order by dt;
quit;

data final3;
    set final3;
/*    if dt<"15APR2017"d ;*/
/*    format dt yymmdd10.;*/
    zhekou=allin_sum/prc_sum;
    per=lr/zhekou;
    power=1-zhekou;
    label zhekou='折扣'
        per='利润/折扣'
        power='打折力度';
run;

data data.Q2_final1;
    set final3;
run;
proc sort data=final3 ;
    by dt;
run;
ods excel file="C:\Users\hr\Desktop\E-2019 中文\data\问题 2 流水数据折扣情况.xlsx"
options (sheet_name='折扣');
proc print data=final3 label;
run;

```

---

```
ods excel close;
/*读取数据 附件 3*/

proc import out=data3
    datafile="C:\Users\hr\Desktop\E-2019 中文\data\fujian3.xlsx"
    dbms=excel replace;
    sheet="promotion_sku";
    getnames=yes;
run;

/*限购数量为 0 的删除*/
data data3_1;
    set data3;
    if _COL8 =0 then delete;
/*    sale_rate=_COL5/_COL11;*/
run;

/*下面是处理附件三得出的结果*/

proc datasets library=work kill noprint;quit;
dm  'out;clear;log;clear;';
/* */
proc import out=test
    datafile="C:\Users\hr\Desktop\E-2019 中文\data\商品打折力度表.xlsx"
    dbms=excel replace;
    sheet="商品打折力度表";
    getnames=yes;
run;

data test1 test2;
    set test;
    date1=put(_col0,best.);
    date2=put(_col1,best.);
    date=strip(scan(date1,1,'.'));
    if strip(scan(date1,1,'.'))=strip(scan(date2,1,'.')) then output test1;
    else output test2;
run;

proc sort data=test1;
    by date;
```

---

```

run;

proc sql;
    create table final0 as
        select distinct date,_col0,mean(_col2) as per_mean from
        test1
        group by date
        order by date;
quit;

/*ods excel file="C:\Users\hr\Desktop\E-2019 中文\data\Q2part.xlsx" options
(sheet_name='打折情况');*/
/*    proc print data=final0 label;*/
/*    run;*/
/*ods excel close;*/

/*test2 部分*/

proc sort data=test2;
    by date;
run;

proc sql;
    create table final2 as
        select distinct date,_col0,sum(_col2) as per_sum,count(_col0) as no from
        test2
        group by date
        order by date;
quit;

data final3;
    set final2;
    per_mean=per_sum/no;
run;

/*两种情况 merge 到一起*/
data final;
    set final3 final0;
run;
proc sort data=final;
    by date;

```

---

```

run;

proc sql;
    create table final_all as
        select distinct date,_col0,mean(per_mean) as mean from
        final
        group by date
        order by date;
quit;

proc sort data=final_all out=all nodupkey;
    by date mean;
run;
/*输出每天折扣力度情况*/
ods excel file="C:\Users\hr\Desktop\E-2019 中文\data\Q2 打折情况.xlsx" options
(sheet_name='打折情况');
proc print data=all label;
run;
ods excel close;

/*将附件 1 2 和附件 3 得出的两种折扣力度进行比较*/

proc import out=data1
    datafile="C:\Users\hr\Desktop\E-2019 中文\data\问题 2 流水数据折扣情况.xlsx"
    dbms=excel replace;
    sheet="折扣";
    getnames=yes;
run;

data data1;
    set data1;
    dt=strip(_col1);
run;
data data2;
    set all;
    dt=strip(date);
run;
proc sort data=data1;
    by dt ;
run;
proc sort data=data2 ;
    by dt;

```

---

```
run;

data data12;
    merge data1 data2;
    by dt;
    if input(date,best.)>21551 then delete;
run;

/*对两种情况的数据折扣进行分析对比*/
data dif;
    set data12;
    if mean eq . then delete;
    dif1=_col7-mean;
    dif1__=_col8-mean;
    dif2=_col9-mean;
    label dif1='与流水折扣差'
    dif1__='与流水折扣率差'
    dif2='与流水打折力度差';
run;

/*输出每天折扣力度差异情况*/
ods excel file="C:\Users\hr\Desktop\E-2019 中文\data\Q2 打折差异.xlsx" options
(sheet_name='打折情况');
proc print data=dif label;
run;
ods excel close;
```