

队伍编号	201146
题号	D

新目标零售产品的精准需求预测

摘要

随着人们消费水平的提高，人们在购物时更加追求高质量的购物体验，所以合理的利用节假日促销商品对零售产品企业尤为重要。

本文依据题中所给数据解决了以下问题：

针对问题一： 为了分析 2018 年国庆节等四个节假日内各种相关因素对目标 skc 的销售量的影响，本文首先运用 Excel 软件找出了目标 skc，即销售时间处于 2018 年 7 月 1 日至 2018 年 10 月 1 日内累计销售额排名前 50 的 skc。其次，运用主成分分析法将 7 个对销售量可能有影响的因素划分为两个指标：销售特征指标，节假日折扣指标。为了避免人工赋权带来的主观性偏差，运用熵值法对两个指标赋权，具体结果见表 5，由结果本文发现在销售特征指标所占权重较大，即商品的标签价格，折扣力度对目标 skc 的销售量的影响较大。

针对问题二： 为了预测目标小类在 2019 年 10 月 1 日后三个月中每个月的销售量，本文首先筛选出了历史销售时间处于 2019 年月 1 日至 2019 年 10 月 1 日内累计销售额排名前 10 的目标小类，并在此基础上建立了三次指数平滑模型，得到了 2019 年 10 月 1 日后三个月中每个月的销售量的预测值，并计算出了每个月预测值的 MAPE(见表 9)，发现每个月的 MAPE 都小于 0.25，因此三次指数平滑模型对预测目标小类的销售量具有一定的准确性。

针对问题三： 为了满足企业更加精准的营销需求，预测目标小类 skc 在 2019 年 10 月 1 日后 12 周内每周的销量，本文对问题二所建立的三次指数平滑模型做了优化，并以此提出了校准模型，使得模型的预测更加准确可靠。目标小类内所有 skc 在 2019 年 10 月 1 日后 12 周内每周的周销量预测值见表 10，每周预测值得 MAPE 见表 11，可见，校准后的 MAPE 小于校准前的值，因此，模型将能更加精准的为企业提供营销需求。

针对问题四： 本文根据前三问探究所得的结果，给企业写了一封推荐信，并结合前三问的研究结果向企业推荐本文的预测结果和方法，以助力企业制定合理的销售策，并在信中阐述了方案的合理性以及后续的研究方向。

关键词：PCA 分析 ， 熵值法 ， 三次指数平滑模型，销售预测校准模型

目录

摘 要	1
一、问题重述	1
1.1 问题的背景及研究意义	1
1.2 需解决的问题	1
二、模型假设	2
三、符号说明	3
四、问题分析	4
五、模型建立与求解	5
5.1 问题一模型的建立与求解	5
5.1.1 数据的处理	5
5.1.2 模型的建立	5
5.2 问题二的模型建立与求解	9
5.2.1 数据预处理	9
5.2.2 模型的建立与求解	9
5.3 问题三模型的建立与求解	11
5.3.1 销量预测校准模型的建立	11
5.3.2 模型的求解	11
5.4 推荐信	15
六、模型的优缺点	16
6.1 模型的优点	16
6.2 模型的缺点	16
七、参考文献	17
八、附录	18

一、问题重述

1.1 问题的背景及研究意义

随着经济的发展，人民生活水平的不断提高，消费者各方面的需求结构升级，市场上的消费模式逐步由“以物为主”转变为“以客为主”。在新零售行业，性价比不再是顾客衡量是否购买物品的唯一标准，人们的需求也不仅仅是单一的追求实用性，而是更多的考虑时尚性，把注意力放在“个性化、时尚、美观”等方面。在这类特殊需求的推动下，新零售企业的生产模式逐步向多品种、小批量迈进，这让商场内零售店铺里的饰品和玩具等种类变得更加琳琅满目，同时也给零售行业的库存管理增加了很大的难度。如何根据层级复杂，品类繁多的历史销售数据，以区域层级，小类层级乃至门店 skc(单款单色)层级给出精准的需求预测，是当前大多数新零售企业需要重点关注并思考的问题。

1.2 需解决的问题

- 1、分析 2018 年国庆节，双十一，双十二和元旦这四个节假日内各种相关因素对目标 skc 的销售量的影响，可考虑产品销售特征，库存信息，节假日折扣等因素。其中，目标 skc 为销售时间处于 2018 年 7 月 1 日至 2018 年 10 月 1 日内且累计销售额排名前 50 的 skc。
- 2、结合上述分析结果，预测给定区域内目标小类在 2019 年 10 月 1 日后 3 个月中每个月的销售量，给出每个月预测值的 MAPE。其中，目标小类为历史销售时间处于 2019 年 6 月 1 日至 2019 年 10 月 1 日内且累计销售额排名前 10 的小类。
- 3、为了满足企业更加精准的营销需求，试着建立相关数学模型，在考虑小类预测结果的同时，预测目标小类内所有 skc 在 2019 年 10 月 1 日后 12 周内每周的周销量，并给出每周预测值的 MAPE(可以考虑 skc 销售曲线与小类销售曲线之间的差异)。
- 4、请给企业写一份推荐信，向企业推荐你的预测结果和方法，并说明你们的方案的合理性以及后续的优化方向。

二、模型假设

- 1、目标 skc 的生产周期为常数。
- 2、目标 skc 的市场价格可预知。
- 3、目标 skc 的市场价格高于成本。
- 4、目标 skc 不受其他消极的市场因素影响。

三、符号说明

符号	含义
F_i	第 i 个主成分
\bar{x}_j	第 j 项指标的均值
s_j	第 j 项指标的标准差
e_j	第 j 项指标的熵值
w_j	第 j 项指标的熵权
s	销售量
C	skc 的销售额 (real_cost)
R	一个 skc 的平均销售价格
T	标签价格 (tag_prize)
S_f	销售特征因子
I	库存信息
D	折扣
S_p	营销推广力度
H	产品热值

注：部分符号见文中说明

四、问题分析

问题一的分析：此问题要求我们分析 2018 年国庆节、双十一、双十二、和元旦四个节假日内各种相关因素对目标 skc 销售量的影响。所以首先需要筛选出目标 skc。题中，此问题的目标 skc 的定义为：销售时间处于 2018 年 7 月 1 日至 2018 年 10 月 1 日内且累计销售额排名前 50 的 skc。之后，为了分析节假日内各种相关因素对目标销售量的影响，需要从题中所给的指标中挑选出可能对目标 skc 销售量有影响的指标，然后运用主成分分析法对挑选的指标进行降维处理，把挑选出的指标转化为几个综合指标，进而运用熵值法对这些指标赋权，计算各指标的综合得分，从而得出相关因素对目标 skc 销售量的影响。

问题二的分析：此问题要求我们预测目标小类在 2019 年 10 月 1 日后 3 个月的销量，所以首先需要筛选出目标小类，题中此问题的目标小类的定义为历史销售时间处于 2019 年 6 月 1 日至 2019 年 10 月 1 日内且累计排名前 10 的小类。其次，运用三次指数平滑模型对目标小类在 2019 年 10 月 1 日后 3 个月的销量进行预测，并于实际销量进行比较，得到每个月的 MAPE 值。

问题三的分析：此问题要求在第二问的基础上给出更精准的营销要求，并建立相关数学模型。显然这是一个问题二的优化问题，所以可以采用问题二的三次指数平滑模型对预测目标小类所有 skc 在 2019 年 10 月 1 日后 12 周内每周的周销量，为了使预测结果更加精准，可以根据 skc 的涨幅与跌幅对销量进行加权，以此建立校准模型。

问题四的分析：本问要求根据前三问的研究结果，向企业推荐预测结果和方法，写一封建议信，本质上就是对前三问问题的一个总结性概括，因此本文决定利用前三问中出现的问题，以及解决的方法思路，向企业介绍本文的研究结果。

五、模型建立与求解

5.1 问题一模型的建立与求解

5.1.1 数据的处理

首先，运用公式（1）计算出各 skc 在销售时间处于 2018 年 7 月 1 日至 2018 年 10 月 1 日内销售额，运用 Excel 筛选出排名前 50 的 skc。（见支撑材料）

$$C = s \times R \quad (1)$$

5.1.2 模型的建立

1、PCA 分析确定新的综合指标[1]

主成分分析方法是一种较为客观的多指标评价方法。它是在损失最少信息的前提下，利用降维的思想，把多个指标转化为几个综合指标的多元统计方法。通常把转化生成的综合指标称为主成分，其中每个主成分都是原始变量的线性组合，且各主成分之间互不相关，这使得主成分比原始变量具有某些更优越的性能。这样在研究复杂问题时就可以只考虑少数几个主成分而不至于损失太多信息，从而更容易抓住主要矛盾，揭示事物内部变量之间的规律性，使问题得到简化，提高分析效率。

Step1 本文运用 SPSS 软件对 2018 年元旦节期间的目标 skc 进行主成分分析，挑选出 7 个指标：商品的标签价格，平均销售价格，折扣，销售特征因子，库存信息，节前营销推广力度，产品热值。从中确定主成分 F_i ，进而确定新的综合指标。

其中，

折扣：目标 skc 的平均售价 R 和其标价 T 的比值，即

$$D = \frac{R}{T}, R = \frac{\sum C}{s} \quad (2)$$

销售特征因子：平均售价 R 和库存信息 I （ I_h 为节假日 skc 库存数量、 I_y 为全年 skc 库存数量）根据重要程度乘以相应权重得到的值，即

$$S_f = 0.6 \times R + 0.4 \times I, I = \frac{I_h}{I_y} \quad (3)$$

节前营销推广力度：节前一周平均销量 s_h 和全年平均销量 s_y 的比值，即

$$S_p = \frac{s_h}{s_y} \quad (4)$$

产品热值：节前一周平均售价 R_h 和全年平均售价 R_y 的比值，即

$$H = \frac{R_h}{R_y} \quad (5)$$

Step2 标准化, 将挑选的原始数据按照公式 (6) 进行标准化, 消除变量之间在数量级上或量纲上的差异。

$$x_{ij} = \frac{x_{ij} - \bar{x}_i}{\sigma_i} \quad (6)$$

其中,

$$\bar{x}_i = \frac{\sum_1^n x_{ij}}{n}, \sigma_i = \sqrt{\frac{\sum_1^n (x_{ij} - \bar{x}_i)^2}{n-1}} \quad (7)$$

Step 3 运用 SPSS 对标准化后的数据进行主成分分析, 得到方差分解结果以及主成分提出分析表, 见表 1

表 1 主成分提出分析表

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	3.699	52.848	52.848	3.699	52.848	52.848
2	1.324	18.911	71.759	1.324	18.911	71.759
3	.855	12.218	83.977			
4	.833	11.897	95.874			
5	.259	3.695	99.569			
6	.030	.424	99.994			
7	.000	.006	100.000			

由表 1 可知: 按照主成分分析的特征值大于 1 的准则, 提取前两个主成分, 且这两个主成分包含原始数据信息总量达到 71.759%, 占据较高比重, 因此选这两个主成分进行分析, 分析得到初始成分矩阵表 2,

表 2 成分矩阵表

	成份	
	1	2
tag_prize	.946	-.059
sales_factor	-.698	.188
rc_1.1	.919	.272
dis_1.1	.112	.806
ie_1.1	.523	.456
sp_1.1	-.565	.592
ph_1.1	.932	-.062

注: rc_1.1: 目标 skc 元旦期间的实际销售价格;
dis_1.1: 目标 skc 元旦期间的折扣;
ie_1.1: 目标 skc 元旦期间的库存;
sp_1.1: 目标 skc 元旦期间的销售推广力度;
ph_1.1: 目标 skc 元旦期间的产品热值。

根据表 2 中各个指标的成分除以相应的特征值开平方根得到他们的特征向量（表 3）。

表 3 特征向量表		
成分	指标一	指标二
tag_prize	0.492	-0.027
sales_factor	-0.606	0.163
rc_1.1	0.994	0.294
dis_1.1	0.123	0.883
ie_1.1	1.058	0.896
sp_1.1	-3.276	3.425
ph_1.1	25.945	-2.89

即得到两个主成分表达式中每个指标所对应的系数，有表达式：

$$\begin{aligned} F_1 &= 0.49x_1 - 0.61x_2 + 0.99x_3 + 0.12x_4 + 1.10x_5 - 3.28x_6 + 25.78x_7 \\ F_2 &= -0.03x_1 + 0.16x_2 + 0.29x_3 + 0.88x_4 + 0.9x_5 + 3.43x_6 - 2.89x_7 \end{aligned} \quad (8)$$

2、熵值法确定客观权重[2]

熵值法是一种客观赋权法。熵(Entropy)是利用概率论来确定信息不确定性的一个量度。设有 n 个等待评价对象， m 项评价指标，形成原始指标数据矩阵 $X = (x_{ij})_{m \times n}$ ，对于某项指标 x ，指标值 x_{ij} 的差距越大，则该指标在综合评价中所起的作用越大；如果某项指标的指标值全部相等，则该指标在综合评价中几乎不起作用。给定一系列的方案和属性，熵值法能够确定出各属性的客观权重值。

由表 1 可知，共有两个综合指标，为表达方便，将新的综合指标 F_1 ， F_2 命名为指标一，指标二

Step1 根据公式（8）对指标进行标准化及非负化处理。

$$p_{ij} = (x_{ij} - \bar{x}_j) / S_j \quad (9)$$

其中 \bar{x}_j 为第 j 项指标的均值， s_j 为第 j 项指标的标准差。

表 4 非负化处理表		
成分	指标一	指标二
tag_prize	0.0584	0.1245
sales_factor	0.0418	0.1329
rc_1.1	0.066	0.1386
dis_1.1	0.0528	0.1641
ie_1.1	0.0665	0.1647
sp_1.1	0.0014	0.2747
ph_1.1	0.713	0.0006

Step 2 根据表 7 中数据，利用公式（9）与（10）计算第 j 项指标的熵值，见表 5：

$$e_j = -k \sum_{i=1}^n p_{ij} \ln p_{ij} \quad d_j = 1 - e_j \quad (10)$$

式中常数 k 与系统的样本数 n 有关, e_j 是第 j 项指标的熵值, d 是第 j 项指标的信息效用值。

$$w_j = \frac{dj}{\sum_{i=1}^n d_j} = \frac{1 - e_j}{n - \sum_{i=1}^n e_j} \quad (11)$$

w_j 代表第 j 项指标的熵权。

表 5 熵值表

节日	指标一 w	指标二 w
1.1	0.8214	0.1786

可见, 指标一对目标 skc 的影响最大, 即商品的实际价格和商品在节日前的推广力度是影响商品在节日期间销售额的主要影响因素, 这与人们的日常生活习惯吻合, 因此本文认为该模型对分析商品销售额的影响因素有可靠性。

为了验证模型的可靠性, 本文将剩余节日的数据代入模型, 最终计算结果如下表:

表 6 验证表

节日	指标一 w	指标二 w
1.1	0.8214	0.1786
10.1	0.6635	0.3665
11.11	0.6099	0.3901
12.12	0.901	0.099

由上表可知: 在上述节日的指标一 w 都大于指标二 w , 可见, 指标一对目标 skc 的销售量影响较大, 即商品的实际价格和商品在节假日的促销活动可以增大消费者对商品的购买力度, 提高商品的销售量。

5.2 问题二的模型建立与求解

5.2.1 数据预处理

运用公式（1）计算出各 skc 在销售时间处于 2019 年 6 月 1 日至 2019 年 10 月 1 日内销售额，运用 Excel 筛选出排名前 10 的 skc。

5.2.2 模型的建立与求解

指数平滑法是生产预测中常用的一种方法。也用于中短期经济发展趋势预测，所有预测方法中，指数平滑是用得最多的一种。

三次指数平滑是在二次指数平滑的基础上再进行一次平滑，其计算公式为：

$$\begin{cases} S_t^{(1)} = \alpha y_t + (1 - \alpha)S_{t-1}^{(1)} \\ S_t^{(2)} = \alpha S_t^{(1)} + (1 - \alpha)S_{t-1}^{(2)} \\ S_t^{(3)} = \alpha S_t^{(2)} + (1 - \alpha)S_{t-1}^{(3)} \end{cases} \quad (12)$$

式中： $S_t^{(3)}$ 为三次指数平滑值

三次指数平滑法的预测模型为

$$\widehat{y_{t+m}} = a_t + b_t m + C_t m^2, m = 1, 2, \dots \quad (13)$$

其中

$$\begin{cases} a_t = 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)} \\ b_t = \frac{\alpha}{2(1-\alpha)^2} [(6-5\alpha)S_t^{(1)} - 2(5-4\alpha)S_t^{(2)} + (4-3\alpha)S_t^{(3)}] \\ C_t = \frac{\alpha^2}{2(1-\alpha)^2} [S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)}] \end{cases} \quad (14)$$

由于货物的日销量的趋势是累加性的，所以运用累加性公式如下

$$\begin{aligned} S_0 &= x_0 \\ s_t &= \alpha(x_t + s_{t-1}) + (1 - \alpha)(s_{t-1} + b_{t-1}) \\ b_t &= \beta(S_t - S_{t-1}) + (1 - \beta)b_{t-1} \\ C_t &= \gamma(x_t - S_{t-1} - b_{t-1}) + (1 - \gamma)C_{t-L} \end{aligned} \quad (15)$$

其中，

α 是数据平滑因子， $0 < \alpha < 1$ ；

β 是趋势平滑因子， $0 < \beta < 1$ ；

γ 是季节改变平滑因子 $0 < \gamma < 1$ 。

经过多次检验，对于本问题取 $\alpha = 0.3$ ， $\beta = 0.3$ ， $\gamma = 0.5$

运用 matlab 编程（代码见附录 1）求得 2019 年 10 月 1 日后 3 个月中每个月的销售量预测值（表 7）及每个月的 MAPE 值（表 8）。

表 7 未来预测值（1）

目标小类	27050401	27217089	27164944	27196225	27060804
10 月真实值	2289	501	574	263	1111
10 月预测值	2313	582	597	268	1065
11 月真实值	1956	282	393	173	843
11 月预测值	2022	333	430	183	880
12 月真实值	2316	230	372	143	889
12 月预测值	2278	232	375	139	880

表 8 未来预测值（2）

目标小类	27112849	27092025	27206656	27071209	27102436
10 月真实值	238	195	11	34	21
10 月预测值	217	183	12	12	12
11 月真实值	278	168	11	28	13
11 月预测值	254	162	10	8	6
12 月真实值	222	209	14	43	26
12 月预测值	214	194	12	22	17

表 9 各小类 APE 值及 MAPE 结果表

目标小类	27050401	27217089	27164944	27196225	27060804	27112849
10 月	0.010	0.162	0.040	0.018	0.041	0.089
11 月	0.034	0.179	0.094	0.059	0.044	0.086
12 月	0.016	0.008	0.007	0.030	0.010	0.036

目标小类	27092025	27206656	27071209	27102436	MAPE
10 月	0.060	0.096	0.654	0.422	0.12314331
11 月	0.039	0.135	0.720	0.529	0.142257679
12 月	0.071	0.156	0.483	0.342	0.105217593

由上表知：每个月的 MAPE 值都在 0.25 以内，所以，该模型对预测商品销量具有较高的准确性。

5.3 问题三模型的建立与求解

5.3.1 销量预测校准模型的建立

为了满足企业更加精准的营销要求，问题二所建立的三次指数平滑模型的预测值不能满足题中要求，因此本文在此基础上提出了校准模型。

Step 1 选取问题二中筛选得到的目标小类，根据销量将目标小类下的 skc 分为五类，记为 A_1, A_2, A_3, A_4, A_5 ，并为其赋权。通过数据透视表分析，得出五类 skc 的权重分别为 0.45 0.25 0.15 0.10 0.05。

Step 2 计算校准系数

计算这五类 skc 在 19 年 10 月 1 日后 12 周每周的价格浮动 k_i ，后与对应的权重相乘得到校准系数 W_{gf} ，即

$$W_{gf} = 0.45k_1 + 0.25k_2 + 0.15k_3 + 0.1k_4 + 0.05k_5 \quad (16)$$

Step 3 建立校准模型

通过计算校准系数与问题二中通过三次指数平滑预测法得到的预测值的乘积，建立销量预测校准模型，即

$$P_v = W_{gf} \times P \quad (17)$$

其中 P_v 代表校准后的预测值， P 代表校准前的预测值。

5.3.2 模型的求解

本文选取小类编号为 27050401 的 skc 举例说明该模型。运用三次指数平滑法预测其中一个 skc（记为 S）的销售量，见图 1

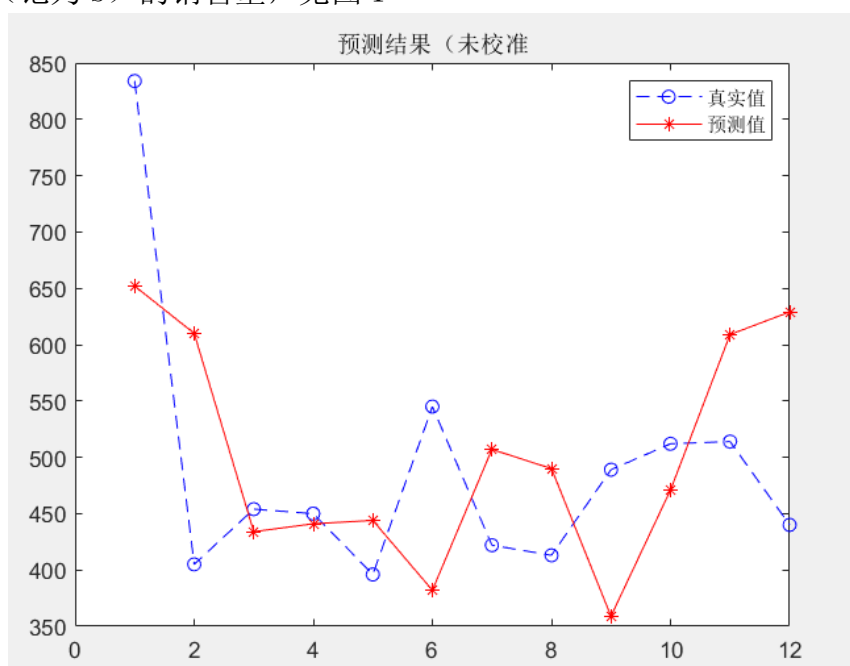


图 1 销售量预测图

由上图可见，问题二所建立的三次指数平滑模型所得到的预测值不能满足企业更加精准的营销需求，所以，为了使模型更加精准，本文建立了校准模型，使预测值更加准确。运用公式（16）计算得到 S 在 2019 年 10 月 1 日后 12 周的校准系数，如下图。

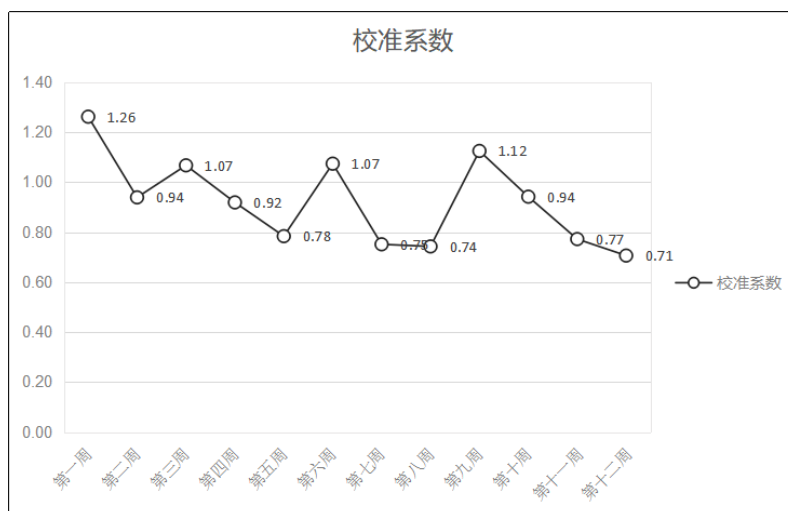


图2 校准系数折线图

得到12周的校准系数后，运用公式17计算得到S的预测值，如图3

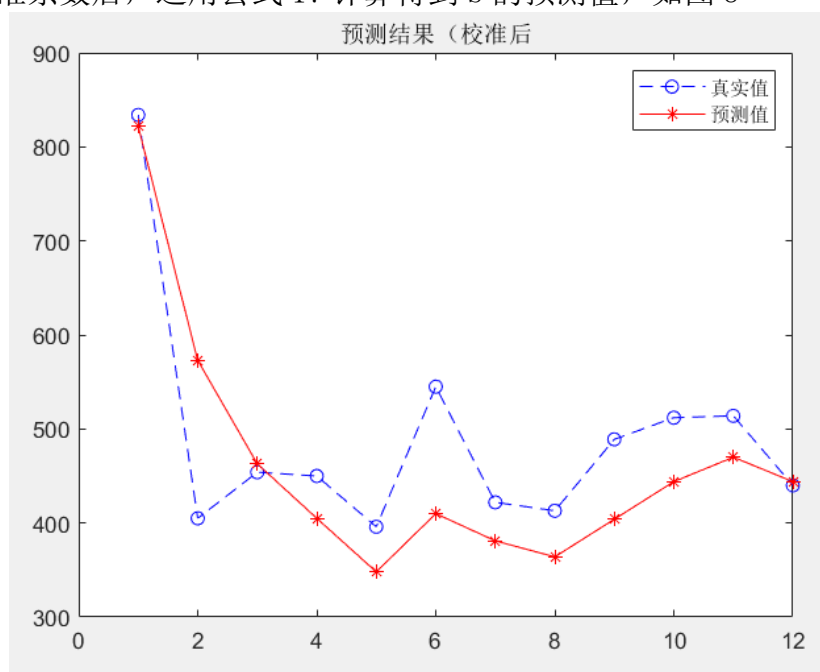


图3 校准后预测图

尽管预测性能已有提升，但为了更加直观的说明该模型的可行性，我们将校准前后的销量与真实值进行了对比，如下图，

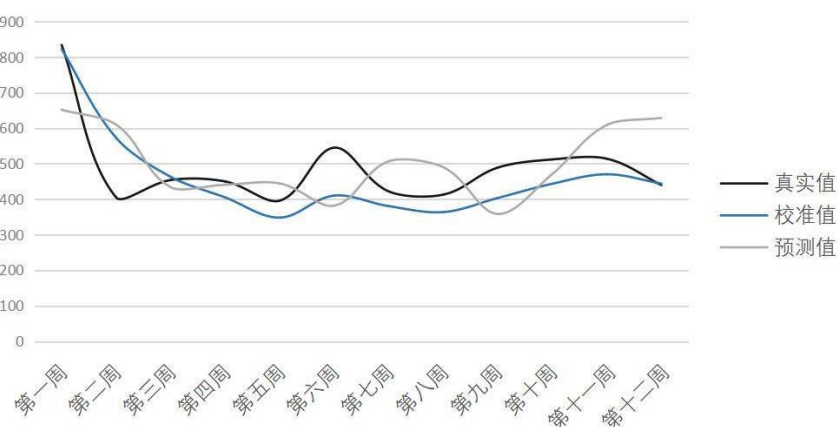


图4 销量较准前后对比图

由图 3，图 4 可以清晰直观地发现，校准后的模型在预测 S 的销售量时更加接近真实值，因而证明了该模型的可用性。

只选取一个小类可能存在样本单一影响结果的情况，因此通过上述模型，我们预测了目标小类内所有 skc 在 2019 年 10 月 1 日后 12 周内每周的周销量，并计算其 MAPE 值，再与前面没有使用该模型的问题一得到的 MAPE 值进行比对。由于篇幅有限，本文只给出部分预测值的结果，如下表所示。

表 10 校准后预测表

目标小类	27050401		27217089	
目标小类内所有 skc 真实/预测值	真实值	预测值	真实值	预测值
2019/10/1-2019/10/7	834	822	210	195
2019/10/8-2019/10/14	405	573	98	129
2019/10/15-2019/10/21	454	463	83	85
2019/10/22-2019/10/28	450	405	85	64
2019/10/29-2019/11/4	396	348	84	54
2019/11/5-2019/11/11	545	410	90	56
2019/11/12-2019/11/18	422	381	66	45
2019/11/19-2019/11/25	413	364	50	32
2019/11/26-2019/12/2	489	404	26	13
2019/12/3-2019/12/9	512	444	49	18
2019/12/10-2019/12/16	514	470	46	21
2019/12/17-2019/12/23	440	444	37	20

综上，不难看出，通过建立基于三次指数平滑模型的销售预测校准模型可以极大地提高模型预测的准确性。由下表可以看出 MAPE 值也远远低于校准前的数值，进一步证明了模型的准确性。

表 11 校准后 MAPE 值表

目标小类内所有 skc 的 APE	27050401	27217089	27102436	MAPE
2019/10/1-2019/10/7	0.014	0.071	0.004	0.00881151
2019/10/8-2019/10/14	0.414	0.318	1.869	0.10683662
2019/10/15-2019/10/21	0.020	0.024	0.414	0.04579203
2019/10/22-2019/10/28	0.100	0.248	0.347	0.09216545
2019/10/29-2019/11/4	0.120	0.351	0.532	0.1082742
2019/11/5-2019/11/11	0.247	0.382	0.386	0.14516228
2019/11/12-2019/11/18	0.098	0.318	0.247	0.13729561
2019/11/19-2019/11/25	0.119	0.358	0.319	0.10749994
2019/11/26-2019/12/2	0.175	0.494	0.446	0.13607851
2019/12/3-2019/12/9	0.134	0.632	0.324	0.11929963
2019/12/10-2019/12/16	0.085	0.538	0.347	0.1684032
2019/12/17-2019/12/23	0.009	0.470	0.114	0.05935892

注：完整表 10，表 11 见支撑材料(答案.xls)

5.4 推荐信

尊敬的领导：

您好！

我们是一支大学生 mathorcup 建模比赛队伍，很荣幸能有这次机会给您写信，我们查阅了大量的相关文献，对所给的数据做了深入的研究，并根据公司希望解决的问题给出了建议。

首先，我们筛选出了销售时间处于 2018 年 7 月 1 日至 2018 年 10 月 1 日内累计销售额排名前 50 的 skc，并建立了基于主成分分析法和熵值法的评价模型，分析 2018 年国庆节等四个节假日内各种相关因素对目标 skc 的销售量的影响。通过评价模型我们发现商品的标签价格，折扣力度对目标 skc 的影响较大。因此，我们建议贵公司可以通过在节假日适度打折的方法来吸引顾客，以获得更高的销售额。

其次，我们筛选出历史销售时间处于 2019 年 1 月 1 日至 2019 年 10 月 1 日内累计销售额排名前 10 的目标小类，通过建立三次指数平滑模型来预测 2019 年 10 月 1 日后后三个月中每个月的销售量。并计算出了每个月预测值的 MAPE，我们发现 MAPE 的计算结果都在 0.25 以下，这说明了该模型预测目标小类的销售量具有一定的准确性。因此，公司在未来可以参考我们的模型来进货，以减少不必要的库存压力，造成资源浪费。

最后，为了满足企业更加精准的营销需求，我们通过建立销售预测校准模型的方法，对前文提到的三次指数平滑模型进行了优化，优化之后计算得到的 MAPE 的最大值从 0.25 降低到 0.1 左右，校准后预测效果显著，能够满足企业更加精准的营销需求，从而指定合理的营销策略。后续我们会从产品本身因素的影响入手，来进一步完善我们的模型。

以上建议是我们团队在严密的数学模型下提出的研究结果，希望我们的建议对您以后的决策有所帮助，助力公司获得更大的成功。

Mathorcup 竞赛团队
2020 年 5 月 24 日

六、模型的优缺点

6.1 模型的优点

6.1.1 鉴于实际进行投资方案决策时，各评价指标相互作用的复杂性，本文将主成分分析法与熵值法结合起来，起到相互补充的效果。

6.1.2 利用 SPSS 软件进行主成分分析，实现指标由多到少的降维和简化，通过变换利用指标间的关系，得到综合的新指标，且这些综合指标能够很好地处理原指标之间的信息重叠，同时较为全面地反映原指标数据所提供的信息。

6.1.3 利用熵值法确定各新指标的权重系数，实现了新指标的熵确定的客观权重，避免因综合权重的计算偏差而导致的决策失误。

6.2 模型的缺点

6.2.1 在建立校准模型时，校准系数的权重是主观赋值，有失严谨。

6.2.2 主成分分析法中挑选指标是基于主观判断，可能会对结果造成误差

七、参考文献

- [1]李艳双, 曾珍香. 主成分分析法在多指标综合评价方法中的应用[J]. 河北工业大学学报, 1999(1):94-97.
- [2]郭显光. 熵值法及其在综合评价中的应用[J]. 财贸研究, 1994(6):56-60.
- [3]陈有为. 基于离散 GM 模型和指数平滑模型组合的统计预测方法[J]. 统计与决策, 2015(10):74-76.
- [4]姜启源. 数学实验与数学建模[J]. 数学的实践与认识, 2001(5):613-617.

八、附录

问题一：

主成分分析法（SPSS）：

```
GET DATA
  /TYPE=XLS
  /FILE='C:\Users\MClover\Desktop\top_50skc.xls'
  /SHEET=name 'Sheet6'
  /CELLRANGE=full
  /READNAMES=on
  /ASSUMEDSTRWIDTH=32767.
EXECUTE.
DATASET NAME 数据集 1 WINDOW=FRONT.
FACTOR
  /VARIABLES tag_prize sales_factor rc_1.1 dis_1.1 ie_1.1 sp_1.1 ph_1.1
  /MISSING LISTWISE
  /ANALYSIS tag_prize sales_factor rc_1.1 dis_1.1 ie_1.1 sp_1.1 ph_1.1
  /PRINT INITIAL CORRELATION KMO EXTRACTION FSCORE
  /PLOT EIGEN ROTATION
  /CRITERIA MINEIGEN(1) ITERATE(25)
  /EXTRACTION PC
  /ROTATION NOROTATE
  /SAVE REG(ALL)
  /METHOD=CORRELATION.
```

熵值法（Matlab）：

```
function [s,w]=shang(x)
[n,m]=size(x); % n=23 个国家, m=5 个指标
%% 数据的归一化处理
[X,ps]=mapminmax(x');
ps.ymin=0.002; % 归一化后的最小值
ps.ymax=0.996; % 归一化后的最大值
ps.yrange=ps.ymax-ps.ymin; % 归一化后的极差,若不调整该值,则逆运算会出错
X=mapminmax(x',ps);
% mapminmax('reverse',xx,ps); % 反归一化,回到原数据
X=X'; % X 为归一化后的数据
%% 计算第 j 个指标下,第 i 个记录占该指标的比重 p(i,j)
for i=1:n
    for j=1:m
        p(i,j)=X(i,j)/sum(X(:,j));
    end
end
end
```

```

%% 计算第 j 个指标的熵值 e(j)
k=1/log(n);
for j=1:m
    e(j)=-k*sum(p(:,j).*log(p(:,j)));
end
d=ones(1,m)-e; % 计算信息熵冗余度
w=d./sum(d); % 求权值 w
s=w*p'; % 求综合得分[\code]

```

```

x=[0.242535623 0.198361465
0.804311476 -0.063361218
0.90359937 -0.066800451
-0.597019813 0.630666491
0.679540162 0.343548843
0.280651014 1.577406181
5.428151822 1.513612336
-32.3572942 16.07203879
]
[s,w]=shang(x)

```

问题二：

三次指数平滑预测法（Matlab）：

```

alpha=0.3;
beta=0.3;
gamma=0.5;
fc=12;
k=12;
sum_APE = 0;
S = [173
317
523
546
654
628
472
359
190
174
110
207
99
114
231

```

```

174
274
131
85
61
32
34
28
43
]
figure(1);
plot(S,'bo--');
n=length(S);
a(1)=sum(S(1:k))/k;
b(1)=(sum(S(k+1:2*k))-sum(S(1:k)))/k^2;
s=S-a(1);
y=a(1)+b(1)+s(1);

f=zeros(24,1);
for i=1:n+fc
    if i==length(S)
        S(i+1)=a(end)+b(end)+s(end-k+1);
    end
    a(i+1)=alpha*(S(i)-s(i))+(1-alpha)*(a(i)+b(i));
    b(i+1)=beta*(a(i+1)-a(i))+(1-beta)*b(i);%趋势
    s(i+1)=gamma*(S(i)-a(i)-b(i))+(1-gamma)*s(i);%周期
    y(i+1)=a(i+1)+b(i+1)+s(i+1);
end
Y = y(1,[2:25])';
APE();
APE=zeros(24,1);
for i = 1:24
    APE(i)=(abs(S(i,1) - Y(i,1)))/ S(i,1);
    sum_APE=sum_APE+ APE(i);
    MAPE=sum_APE/24;
end
hold on
figure(1);
plot(Y,'r*-');
hold off

title('预测结果');
legend('真实值','预测值');
figure(2);
plot(APE,'bo--');
legend('APE');

```