

中国研究生创新实践系列大赛
“华为杯”第十七届中国研究生
数学建模竞赛

学 校 中国地质大学（武汉）

参赛队号 XXXX

队员姓名

1. 李延炽
2. 熊诗捷
3. 明飞

中国研究生创新实践系列大赛
“华为杯”第十七届中国研究生
数学建模竞赛

题 目

出版社图书印制策略

摘 要：

汽油在生产生活中有着难以取代的作用。

关键字：数据挖掘 遗传算法

1. 问题重述

1.1 引言

某出版社出版发行与初等教育相关的图书,在运营过程中,遇到了图书印量、图书销售量和图书库存之间如何协调的问题。在印制图书的时候,除了根据印量会产生纸张费、印刷费、装订费、封面工艺费等印刷成本外,每一次印刷都有一笔数千元的上版费,因此出版社自然希望一次开机尽量多印一些,但是图书销售量通常是不确定的,如果印多了,图书库存加大,会导致与库房结算的发货费率* (见注 1 中定义) 提升,增大发货费用;滞销图书等待报废,其印刷成本的投入会全部损失殆尽。因此,一本书印几次,每次印多少,是这个出版社希望解决的问题。为此,请你们为该出版社出谋划策,仅考虑印刷成本和库房发货费用。针对不同类型的图书提供最优的印刷方案,以期获得尽可能大的销售收益。

1.2 问题的提出

该出版社出版发行的图书主要分 A、B、C 三类。

A 类图书属于政府采购类(如经国家审批出版的教材教辅),不面向市场公开销售。这类图书定价低(销售折扣约 48% 折),订单数量大且相对稳定,但图书更新快,当年的滞销书成为库存,等待报废。现有问题是:订单上报时间不集中,且收到订单后通常需要迅速发货,在征订后期经常有几百册的增补订单,导致印次能多达 7-8 次。

B 类图书是直销类图书,主要用于高考复习使用,直接进校推广。这类图书定价高,销售折扣低(销售折扣约 18% 折),图书更新快,当年不能销售的书成为滞销库存,等待报废。现有问题是:这类图书因定价高,滞销库存总体码洋* (见注 1 中定义) 较大,对发货费率影响较大。

C 类是市场零售类图书,走实体书店和网上渠道销售,没有特别集中的订量,销售不可控。这类图书销售折扣约 45% 折,定位于长销,销售时间可延续若干年,但通常图书上市 2 年以后,热度就会减弱,如果处于滞销状态,就会等待报废。现有问题是:该类图书的首印通常在 3000-5000 册,如果能够实现全部销售,则一定处于盈利状态,当某本图书库存不足 500 册时,会考虑重印,出版社希望能根据上一次印刷后的销售情况,并结合热度衰减因素,来规划下一次的印刷数量,使其实现盈利增加。

出版社与库房结算的发货费率每年调整一次,取值范围如注 1 名词解释中表格所示。可以假设发货费率在近几年保持不变,而当前的发货费率为 2.73%。

请建立数学模型,完成下面的任务:

1. 研究 A 类图书的需求和订单规律(附件 1),对每本书给出 2021 年秋或者 2022 年春的最优印刷方案。有没有可能将总印次控制在 3 次以内?

2. 对 B 类图书，根据往年的销售情况（附件 2），在降低库存的前提下，对每本书给出下一年度的最优印刷方案。

3. 对 C 类图书，对附件 3 中的 9 本图书，考虑未来可能的销售情况，给出每本书的重印方案（是否需要重印，如需重印，最优重印数量），并判断出版社之前的重印策略是不是最优的。

4. 给该出版社写一个企划书，对图书的印刷方案给出你们的建议。

题中包含若干名词解释和行业约定

【注 1】名词解释

码洋：图书的定价 × 数量

滞销库存码洋：图书的定价 × 滞销库存数量

上版费：指印刷开机前，需要将图书文件上机制版产生的费用，上版费与图书的印张有关，可利用附件 4 的《图书印制成本计算表》计算。

销售率：（发货数量-退货数量）/印刷数量

发货费率：用于和库房结算库租物流费用（又称库房发货费用）的一项指标，与出版社所有图书全年发货码洋和每个月的平均库存有关，也即和图书的周转情况有关。图书周转越快，库存图书码洋越少，发货费率越低；反之，图书周转越慢，库存图书码洋越多，发货费率越高。每年的发货费率对所有图书都是一个固定的常数，部分取值如下表所示：

100%-91%	90%-81%	80%-71%	70%-61%	60%-51%	50%-41%	40%-31%	30%-21%	20%-11%	10%-0%
3.42%	3.12%	2.91%	2.73%	2.60%	2.48%	2.40%	2.32%	2.25%	2.19%

表 1 发货费率取值表

目前该出版社所有图书的发货费率结算标准是 2.73%。

库房发货费用：即库租物流费用，计算公式是：

发货费用 = 发货数量 × 定价 × 发货费率

（注：发货数量 + 滞销库存数量 = 印刷数量）

利润 = 定价 × 印数 × 销售折扣 × 销售率 - （印制成本 + 库房发货费）

【注 2】行业约定

当印刷数量低于 2000 册时，因印刷开机成本问题，印刷企业会按照 2000 册印量进行计费。所以，低于 1500 册时，通常会采用数码印刷，数码印刷费用会较传统印刷费用上浮 15%。

2. 符号说明

符号	意义	注释
$A1 \sim A5$	A 类图书 5 本	1 – 5 表示图书编号
$B1 - 1 \sim B1 - 5$	B 类图书第一类 5 本	1 – 5 表示图书编号
$B2 - 1 \sim B2 - 5$	B 类图书第二类 5 本	1 – 5 表示图书编号
$C - 1 \sim C - 9$	C 类图书 9 本	1 – 9 表示图书编号
pc	图书定价	参照附录 1-3 确定定价
N	印刷数目/印数	
d	折扣率	$d\%$
s	销售率	$s\%$
c	印制成本	
sc	库房发货费	
pf	利润	
sv	图书销量	
sq	发货数量	
uns	滞销库存数量	
rq	退货数量	
sr	发货费率	$sr\%$
srp	发货占比	$srp\%$
S	当前库存	
ra	A 类图书订单需求	
s_s	安全库存值	
p_f^k	A 类图书 A_k 首单占比	首单为首大订单及其增补
er	预测值与实际值的误差率	$er\%$

3. 问题分析与模型假设

依据题目意思，最终的目标函数是求得出版社在给定方案下的运营利润。

假设定价为 pc ，印刷数量为 N ，销售折扣率 $d\%$ ，销售率 $s\%$ ，印制成本 c ，库房发货费 sc 。利润 pf 的计算方式如下：

$$pf = pc \times N \times d \times s - (c + sc), \quad (1)$$

其中，定价 pc 可参照附录 1-3 找得；印刷数量 N 计算如下：

$$N = sq + uns, \quad (2)$$

其中， sq 代表发货数量、 uns 代表滞销库存数量；销售折扣率 $d\%$ 参照题目文本；以 sv 代表销量，则销售率 $s\%$ 计算方式如下：

$$s = sv/N = (sq - rq)/N, \quad (3)$$

其中， rq 代表退货数量；印制成本 c 可通过附录 4 的 excel 表格计算，主要关联因素为印张、装订方式及印刷方式；库房发货费 sc 计算方式如下：

$$sc = sq \times pc \times sr, \quad (4)$$

其中， sr 表示发货费率，在本文中，因为发货费率是由图书周转情况确定，为简化计算，发货费率通过发货数量占比划分区间确定，占比 srp 的计算方式为：

$$srp = sq/N, \quad (5)$$

根据此占比值，发货费率由表格 1 确定。依据题意假设，目前该出版社所有图书的发货费率结算标准是 2.73%。

3.1 问题一

A 类图书属于政府采购类（如经国家审批出版的教材教辅），不面向市场公开销售。这类图书定价低（销售折扣约 48% 折），订单数量大且相对稳定，但图书更新快，当年的滞销书成为库存，等待报废。现有问题是：订单上报时间不集中，且收到订单后通常需要迅速发货，在征订后期经常有几百册的增补订单，导致印次能多达 7-8 次。

问题一的题目要求：

研究 A 类图书的需求和订单规律（附件 1），对每本书给出 2021 年秋或者 2022 年春的最优印刷方案。有没有可能将总印次控制在 3 次以内？

3.1.1 问题分析

问题一所求为 A 类图书中，A1-A5 五本图书下一年的季度性印刷方案。要求在保证发行印刷数量大于当前订单需求，即 $N > mr$ 的情况下，给出每一本书在下一年的最优印刷方案。A 类图书的运营效果由最终方案带来的利润评估。

A 类图书是不面向市场销售的政府采购类，其特点是订单零散，但需要接单后尽快发货。A 类图书印刷的主要矛盾存在于不确定的订单造成的高印次带来的上版费等开销，导致最终成本增加。问题一的主要目的是针对 A 类图书的每本书给出印刷方案，力求将总印次控制在 3 次以内，同时达到利润最大化。

通过对附录 1 的数据观察，我们总结出问题一的如下特征：

- 印刷产生在某一年的一个季度之内（春季 1-3 月份，秋季 7-9 月份），总体而言是对季度前及季度中的订单的生产；
- 由于订单分散、且时间不确定，某些图书为了满足订单需求往往在一个季度内印次在 1-6 次不等，且末尾几次印刷数量小（ $N < 2000$ ）；
- 除开数量相对来说极少的增补订单外，A 类图书的印刷数量绝大部分是为了满足前期的大订单；
- A1-A5 不同图书的订单虽然零散，但总数相对稳定，且往年数据或相对平稳，或存在一定的逐年变化趋势；
- 由于 A 类图书定价低，但接到订单后能迅速发货，因此其发货费率较低，但印次带来的上版费高，所以影响 A 类图书利润的主要因素是不确定的订单需求造成的高印次。

因此，对问题一的求解，我们希望能够结合现有数据，有效分析下一年一个季度里每个月的订单总数，根据预测的订单数执行印刷计划，旨在短期储备库存，避免需求少而订单多造成的多印次。大体的做法如下：

- *Step 1:* 根据往年数据中的订单报数年度总和，预测下一年的订单总数，作为印刷总数；
- *Step 2:* 在安全库存的模型基础上，通过分析本题中问题的特性和数据的特点，确定存储供应周期和额外安全库存量的预测方式；
- *Step 3:* 根据第二步得到的安全库存模型，制定印刷策略，产生具体印刷方案；
- *Step 4:* 以往年的数据，产生该模型下的新印刷方案，通过计算利润验证模型精度，进行修正；
- *Step 5:* 基于往年数据预测下一年的订单分布，并根据提出的安全库存模型预估利润。

3.1.2 模型假设

根据题意信息，在求解问题一时，我们保持如下的基本假设：

- 发货费率在近几年保持不变，当前的发货费率为 2.73%；
- A1-A5 五本图书的印刷量、销量等之间均无相关关系；

- 基于存储周期根据预测值印刷的方法，销售量以附录 1 中的订单报数为准，所有时间相近或印刷数量较少的订单视作时间上最接近的大订单的增补；
- 不考虑 A 类图书的订单印刷时间，即订单上报后能短期内完成印刷并发货；
- 不考虑存储供应周期内印刷时间到订单时间差造成的延误带来的经济损失；
- 发货费率的计算以表 1 为准。

3.2 问题二

B 类图书是直销类图书，主要用于高考复习使用，直接进校推广。这类图书定价高，销售折扣低（销售折扣约 18% 折），图书更新快，当年不能销售的书成为滞销库存，等待报废。现有问题是：这类图书因定价高，滞销库存总体码洋 *（见注 1 中定义）较大，对发货费率影响较大。

问题二的题目要求是：

对 B 类图书，根据往年的销售情况（附件 2），在降低库存的前提下，对每本书给出下一年度的最优印刷方案。

3.2.1 问题分析

问题二所求为 B 类图书中，B1 类和 B2 类各五本的印刷方案。要求在降低库存的前提下，给出每本书下一年度的最优印刷方案。而同样的，问题二中 B 类图书的运营效果由最终方案带来的利润评估。

B 类图书是定价高且销售折扣低的类型，高定价容易造成库存总体码洋大，进而增大发货费率最终导致发货成本高。题目要求主要目的为在降低库存的前提下，求得满足销售需求的印刷方案。也就是尽量让印刷发货出去的成品满足该时段销售量需求从而减少库存，降低成本。由此可见该题中主要矛盾在于：高定价使库存要求严格，即产即销是最优的方案。

通过观察附录 2 的数据，我们总结出问题二的如下特征；

- B1 类图书的销售周期是 3 月到 9 月，其实际销售情况与之相符；同样的，B2 类图书的销售周期是 10 月到次年 4 月，其实际情况与之大致相符；
- B1 类图书在每周期的 4-8 月，B2 类图书在每周期的 12-1 月往往供不应求（印刷多，卖的也多），导致中途加印，而其他月份供大于求，库存积压；
- 在销售周期的结束月，销量往往开始剧减，周期结束后直到下一周期开始，销量较之前同样剧减；
- 图书的印次大体上一年更新两轮（比如 1-1 到 1-3），修订的结果往往是印张的调整，从而导致印刷成本变化；
- 印次为偶数的版本，依照题目意思可以不做讨论，故归入其他印次版本；
- 销量统计以月为单位，可见销售周期内均可满足 $sv \geq 2000$ 的条件；而销售周期外均不

能满足该条件；

- 不考虑 2020 年 1 月开始的新疫情影响，数据整体走势明显，且显示出较为稳定的变化趋势。

因此对于问题二的求解，我们希望能够结合现有数据，有效分析下一年图书的销量，主要为销售周期内，每个月的销量，而同时将非销售周期内的图书销量并入周期末尾月。以此为基础制定印刷计划。旨在减少库存量，实现图书尽可能的即印即销。大体做法如下：

- *Step 1*: 对往年数据做加工处理，将原本的每年销量转换为按照销售周期，每个周期销量；
- *Step 2*: 在销售周期内按照月份统计每月销售量，将非周期的销量并入周期末月（9 月和 4 月），然后构建销量趋势预测；
- *Step 3*: 以前两步的数据为基础预测下一年销售周期内的年销量，以此决定印刷数量；
- *Step 4*: 基于经济订购批量模型（EOQ），为下一年的印刷制定最优的印刷方案；
- *Step 5*: 根据往年印次变化，确定当年印次参数，并计算印刷成本，
- *Step 6*: 以往年的数据验证模型精度，进行修正；
- *Step 7*: 预估方案下，计算下一年的运营利润；

3.2.2 模型假设

根据题意信息，在求解问题二时，我们保持如下的基本假设：

- 发货费率在近几年保持不变，当前的发货费率为 2.73%；
- B1，B2 两类各五本图书之间无相关关系；
- 由于往年积压图书报废，可忽略库存，直接从附录 2 取得每月销量；
- 印次的变化符合往年变化总趋势，不会出现特殊情况；
- 印次为偶数的配套的教师用书，可并入前一印次，根据题意不做讨论；
- 不考虑月内发货和印刷时间差造成的延误带来的经济损失；
- 各种购物途径下，可以忽略新冠疫情等外界因素带来的影响。

3.3 问题三

C 类是市场零售类图书，走实体书店和网上渠道销售，没有特别集中的订量，销售不可控。这类图书销售折扣约 45% 折，定位于长销，销售时间可延续若干年，但通常图书上市 2 年以后，热度就会减弱，如果处于滞销状态，就会等待报废。现有问题是：该类图书的首印通常在 3000-5000 册，如果能够实现全部销售，则一定处于盈利状态，当某本图书库存不足 500 册时，会考虑重印，出版社希望能根据上一次印刷后的销售情况，并结合热度衰减因素，来规划下一次的印刷数量，使其实现盈利增加。

问题三的题目要求是：

对 C 类图书，对附件 3 中的 9 本图书，考虑未来可能的销售情况，给出每本书的重印

方案（是否需要重印，如需重印，最优重印数量），并判断出版社之前的重印策略是不是最优的。

3.3.1 问题分析

问题三所求为 C 类图书中，C1-C9 九本图书图书的重印方案，并评估往年的重印策略是否最优。要求根据上一次印刷后的销售情况，结合热度衰减因素，规划下一次的印刷数量，使其实现盈利增加。同样，问题三中 C 类图书的运营效果由最终方案带来的利润评估。

C 类图书是市场零售类图书，销量不可控制，即没有稳定数值。由于图书定位于长期销售，在销售过程中会出现重印发货的需求。当库存 $uns < 500$ 时，出版社就需要考虑重印以准备满足后续的销售需求。此时需要对下一次是否重印、重印数量进行预测分析。本题关键在于，销售量没有稳定数值或趋势，但销售过程中，存在热度衰减的规律。因此，问题三的目标为，规划每本图书是否重印及重印数量，使利润最大化。

通过观察附录 3 的数据，我们可以总结出问题三的如下一些特征：

- C 类图书的销售周期长，不存在 A 类或 B 类的过期问题，因此对 C 类图书，出版之后重印是唯一需要关注的问题；
- C1-C9 图书的月销售量没有明显的趋势或季节性关系，其年销售总量也没有稳定的趋势或季节性；
- 虽然月销量没有明显趋势或季节性关系，但销量的增量式增长似乎存在一定的变化趋势；
- C 类图书在出版后，经过一段时间的热销，往往在 2 年内销量有所下滑，2-3 年后销量剧减（热度衰减）；
- C 类图书存在严重的退货现象，有部分图书某些月份退货数量较大；
- 月销售量的增减幅度大，但总体来看，可能存在一定的不依赖于季节性的变化规律，即热度的增长和衰减规律。

对于问题三的求解，我们的目标是，在 C 类图书中 C1-C9，存在图书库存不足 500 时，根据上一次印刷后的销售情况，并结合热度衰减因素，来规划是否需要重印，如果需要重印，给出下一次的印刷数量，实现盈利增加。也就是 9 本图书，考虑未来可能的销售情况，给出每本书的重印方案（是否需要重印，如需重印，最优重印数量）。此外，还需要判断出版社之前的重印策略是不是最优的。因此我们的模型应该实现的功能是，结合热度衰减因素，给出 C1-C9 基于近年销量情况的最优重印方案。大体做法如下：

- Step 1: 对附录 3 的数据做加工处理，将原本的每月销量转换为按照增量式的销量累加；
- Step 2: 根据预处理后的数据，拟合热度衰减模型；
- Step 3: 以历史数据验证模型精度，并加以修正；
- Step 4: 基于热度衰减模型，预测接下来的销售情况；

- *Step 5*: 根据所得销售预测，给出重印方案并计算利润。

3.3.2 模型假设

根据问题三的题目信息，在求解该问题时，我们保持如下的基本假设：

- 发货费率在近几年保持不变，当前的发货费率为 2.73%；
- C1-C9 图书之间的销量等无相关关系；
- 由于图书定位于长销，且印刷数量整体相对少，可忽略库存带来的成本影响；
- 不考虑重印印刷和销售之间时间差造成的延误带来的经济损失；
- 退货造成的库存增加影响同样忽略不计；
- 各种购物途径下，可以忽略新冠疫情等外界因素带来的影响。

3.4 问题四

问题四要求，给该出版社写一个企划书，对图书的印刷方案给出你们的建议。由于该出版社出版图书即 A，B 和 C 三类，问题四只需要结合前三问的方案，给出对后续运营的规划方案即可。在问题四的求解过程中，我们同样维持上述问题一到问题三的假设条件不变，且所采用的数据、模型及方法与前文所提保持一致。

4. 模型的建立与求解

4.1 问题一

基于第三章对问题一的分析，该问题的求解主要包括预测下一年订单总量、预测季度内每月订单数、根据往年数据验证模型精度、以所得模型预测结果为基础，形成印刷方案并给出预估利润等步骤。在本节，处理这些步骤的方法将一一展开。

4.1.1 基于时间序列分析的订单总量预测

一般地，考虑以历史数据为基础，对接下来阶段的数据进行预测，这类问题可以建模为时间序列分析问题，以时序分析的手段进行求解。

时间序列分析模型的选择主要考虑时间序列的以下特征：平稳性、随机性、趋势性和季节性。对附录 1 中 A1-A5 图书订单总报数这一数据从 2013-2020（A1、A3、A4）和 2015-2021（A2、A5）的观察，我们可以大致看出发现这五组数据的如下特征：

- A1:2013-2020 年间，订单总报数有较明显趋势，但不是线性关系；
- A2:2015-2021 年间，订单总报数整体上没有明显趋势，且存在一定波动（不平稳）；
- A3:2013-2020 年间，订单总报数整体上没有明显趋势，且波动较大；
- A4:2013-2020 年间，订单总报数有明显增长趋势；
- A5:2015-2021 年间，订单总报数有明显增长趋势；

因此，针对这些不同的数据特征，我们选择了 ARIMA 模型、布朗模型和霍尔特模型三者。其中，由于 A1 的数据呈现出的非线性趋势，我们拟采用布朗模型进行预测；A2 的数据整体上没有明显趋势，但有周期性特征，所以拟采用 ARIMA 模型进行预测；A3 不具有明显趋势且波动较大，时序分析的方法或灰色预测等鲁棒性高的方法都不适用，因此直接取历史数据的平均值作为预测值；A4、A5 的数据趋势明显，且没有周期性特征，因此拟采用霍尔特模型进行预测。

上世纪五六十年代，美国的 HMMS 研究团队在卡内基工学院成立，旨在寻找更好的决策机制，以帮助工业界更好地应对种种库存、生产和计划问题。这些问题在宏观层面导致经济危机，在微观层面让企业经常处于危机状态，要么是赶工加急，要么是产能利用不足，以及库存积压。其中霍尔特开发了应对平缓需求的简单指数平滑法、应对趋势的霍尔特双参数线性指数平滑法，以及应对季节性的霍尔特-温特模型，这些都成为工业界最为广泛应用的预测模型。

在第三章对问题一的分析中我们提到，A 类图书的订单需求量大，且常有增补，但没有明显的变化趋势。在时间序列分析中，简单指数平滑法是最基础的方法，但该方法只在预测处于恒定水平和没有时间性变动的时间序列上行之有效。而本文的问题中，订单的零散造成了严重的时间性变动，但考虑到总体的趋势明显且呈线性特征，我们尝试霍尔特指

数平滑法对订单总量进行预测。

霍尔特指数平滑法是一种高级的线性指数平滑方法，该方法的优点是可以用不同的平滑参数对原序列的两种因素进行平滑，具有很大的灵活性，因此在实践中被广泛地应用。霍尔特指数平滑法估计当前时间的水平和斜率。其平滑水平是由两个参数控制：

- α ：估计当前点水平；
- β ：估计当前点趋势部分斜率。

两个参数都介于 0-1 之间，当参数越接近 0，大部分近期的观测值的权值将较小。

在霍尔特双参数平滑法模型中，预测由两部分构成：一部分是水平部分，是在上期水平部分的基础上，用简单指数平滑法来更新；另一部分是趋势部分，是在上期趋势部分的基础上平滑调整，也用简单指数平滑法来更新；两者相加，就得到下期的预测。霍尔特法不但持续调整水平部分，而且持续调整趋势部分，在横向和纵向两维调整预测，所以能更好地应对趋势的变化。基本的公式分三部分：

- 本期水平部分 = α * 本期需求实际值 + $(1-\alpha)$ * (上期水平部分 + 上期趋势部分) (1)
- 本期趋势部分 = β * (本期水平部分 - 上期水平部分) + $(1-\beta)$ * 上期趋势 (2)
- 下期预测 = 本期水平部分 + 本期趋势部分 (3)

三者的计算公式如下：

$$\begin{aligned} L_{t+1} &= \alpha D_t + (1 - \alpha)(L_t + T_t); \\ T_{t+1} &= \beta(L_{t+1} - L_t) + (1 - \beta)T_t; \\ F_{t+1} &= L_{t+1} + T_{t+1}, \end{aligned} \quad (6)$$

其中， D_t 代表实际值； F_{t+1} 代表预测值； L_t 代表平均需求； T_t 代表增长的趋势，前者是对时间序列趋势的平滑式；后者是对趋势增量的平滑式。

在本文中，参数 α 及 β 均依据默认设置设定为 0.5。 D_t 取实际订单数， L_t 取平均订单数。根据霍尔特指数平滑法，我们可以在往年数据的基础上，得到下一年季度内订单总数的预测值。

指数平滑法是布朗 (Robert G..Brown) 所提出，布朗 (Robert G..Brown) 认为时间序列的态势具有稳定性或规则性，所以时间序列可被合理地顺势推延；他认为最近的过去态势，在某种程度上会持续到最近的未来，所以将较大的权数放在最近的资料。布朗模型同样适用于具有线性趋势且没有季节性特征的序列，是霍尔特模型的特例。

布朗模型与霍尔特模型的唯一区别在于，布朗模型假设 $\alpha = \beta$ 。因此，在一定程度上，可以消除霍尔特模型的强线性关系，能够适用于 A1 的数据特征。

Autoregressive Integrated Moving Average Model，即自回归移动平均模型。它属于统计模型中最常见的一种，用于进行时间序列的预测。其原理在于：在将非平稳时间序列转化为平稳时间序列的过程中，将因变量仅对它的滞后值以及随机误差项的现值和滞后值进行

回归所建立的模型。

简单来说，ARIMA 模型即 ARMA 模型 + 对不满足平稳性要求的原数据进行差分。自回归滑动平均模型（ARMA 模型，Auto-Regressive and Moving Average Model）是研究时间序列的重要方法，由自回归模型（简称 AR 模型）与滑动平均模型（简称 MA 模型）为基础”混合”构成。在市场研究中常用于长期追踪资料的研究；在零售研究中，用于具有季节变动特征的销售量、市场规模的预测等。ARMA 模型的预测公式如下：

$$y_t = \mu + \sum_{i=1}^p \gamma_i y_{t-i} + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}, \quad (7)$$

其中， y_t 是当前值； μ 是常数项； p 是阶数； ε_t 是误差值； p 与 q 分别为自回归模型与移动平均模型的阶数； γ_i 与 θ_i 分别是两个模型的相关系数。

4.1.2 基于安全库存模型的订单数规划

在上一节中，我们根据时序分析能够得到下一年订单总数的预测值。在此总数基础上，下一步需要完成对订单的分印次规划。注意到这个步骤的目的是尽量将印次控制在 3 次以内并实现利润最大化。为了实现印次的减少，提前储备库存以满足后续订单需求是核心问题。在《供应链架构师》一书中提到：库存管理有两大核心，分别是“量值”和“精度”。其中，“量值”指的是库存量的多少。它实际上解决的是两个难题，即什么时候买？买多少？这就涉及到经典的库存决策模型了。

安全库存是为防止未来物资供应或需求的不确定性因素（如大量突发性订货、交货意外中断或突然延期等）而准备的缓冲库存。其大小取决于供应和需求的不确定性、顾客服务水平（或订货满足率），以及缺货成本和库存持有成本。顾客服务水平较高，则安全库存量增加并导致缺货成本较低、而库存持有成本较高；相反，顾客服务水平较低，则安全库存量减小并导致缺货成本较高、而库存持有成本较小。

安全库存（safety stock）可定义为超出预期需求之外的附加库存，其确定标准有如下两条：

- 简单规定存储周期，即应该存储多久的供应量作为安全库存；
- 使用一种能够跟踪需求变化幅度，有效给出后续需求预测的方法。

安全库存模型可表示如下：

$$S = ra + s_s, \quad (8)$$

其中， S 表示当前库存； ra 表示 A 类图书订单需求； s_s 表示安全库存量。

由于物资在很短的时间内能销售并且经常中断，在本文的安全库存模型中，我们拟采用单周期存储模型，以更适用于该问题特征。此外，在本题中，我们设想对库存水平进行监控，当库存量降至一定的水平，即最低库存量 S_m 时，进行下一次的印刷积压库存。

基于安全库存的模型，一种简单且行之有效的思路是：根据订单总数的预测，第一印次首先解决绝大部分的总数需求，这样可以满足来得最早的几个大订单，并且剩余的库存量能够满足近期的增补订单。接下来，对于为数不多的小订单，根据需求预测储备安全库存，满足未来一段时间内的小订单，有效减少印次。

首先需要计算的是，第一印次印刷数量在总数中的占比，即首印占比 p_f^k 。可以观察到大部分 A 类订单首订单数较大，后续有若干小订单，我们将这些小订单视为 A 类订单的增补订单。那么可以将后续的小型订单视为超出首单定单预期需求之外的附加库存，建立安全库存模型。

通过对附录 1 数据中，每年订单情况的观察，在首次大订单之后，存在一定时间的小订单报数。相对于大订单的数量，这些小订单中明显少于大订单的（数量级上小）视作增补订单。通过对往年历史数据中，首次大订单的增补时间统计平均数，我们可以得到平均的首订单时间。结果如下：

	A1	A2	A3	A4	A5
首单时间	12.5	21	9.05	11	15

算得的平均首单增补订单报单时间 13.71，取为 14 天。因此，我们将 14 天定作首单增补时长，即 14 天之内的订单作为首印大订单。

同样地，对于增补订单的补单时间差，我们统计了 A1-A5 订单情况中，印刷数相对较少的为增补订单，以及增补订单的报单延续时长（即到下一单的时间差）。统计结果如下：

	A1	A2	A3	A4	A5
补单时间	0	21	10.6	47.33333333	15

算得的平均增补订单报单时间 25。因此，我们将 25 天定作增补时长，即 25 天之内的订单作为大订单的增补。

考虑到问题一的数据（附录 1）中订单零散，且增补订单印刷数量极少，我们还需要为安全库存模型计算存储周期。

假设 $p_i^1-p_i^5$ 分别为 A1-A5 图书在第 i 年的首次大订单及该订单的增补订单占总数的比例，则我们取算术平均数作为 A1-A5 的首单占比：

$$p_f^k = \sum_{i=1}^t p_i^k / t, \quad (9)$$

其中， t 表示相应年份， $k = 1, 2, 3, 4, 5$ 表示 A1-A5。

根据上述的首单统计，我们得到了 A1-A5 各图书首单 p_f^i 占总印数 N 的占比如下：

	A1	A2	A3	A4	A5
首单占比	0.98470134	0.968998914	0.841523467	0.98689145	0.971921507

最终其均值为 0.950807337，取 0.95。亦即，剩余订单占比为，0.05。

那么在本文中，我们规定安全库存量 s_s 计算如下：

$$s_s = 0.05S' \times \frac{25}{90}, \quad (10)$$

其中， S' 代表上一印次时的库存，90 代表问题一中需要考虑的季度时长。

此外在本文中，我们还设定了基于安全库存模型的新的概念，以适用于问题一的实际情况。由于订单散乱、数量通常较小且没有时间规律，无法确定接下来的印次时间点，我们提出根据订单报数与安全库存来判断下一印次的时间点和印刷数，以期更灵活地应对。

假设上一订单满足后，当前库存为 S 。后续订单需求为 ra_t, \dots, ra_{t+k} ，且此时： $\sum_{i=t}^{t+k} ra_i > s_s$ ，则此时下一印次开启。此时印次的印刷量计算如下：

$$N = S - \sum_{i=t}^{t+k} ra_i + s_s. \quad (11)$$

根据上述模型的方法，即可得到修改的安全库存模型下，最优的印次规划及印刷方案。

4.1.3 模型检验方法，方案形成及利润预估

首先对于订单总数预测模型，由于往年的历史数据中，订单总报数是明确的，所以我们直接以预测模型对历年的预测值和实际值对比，通过统计绝对误差率来反映模型的精确度。假设 A_p 为对 A 类图书的订单需求预测，则绝对误差率 er 计算如下：

$$er = |A_p - ra|/ra. \quad (12)$$

对于安全库存模型，同样的，我们可以将所得模型在上一步的预测基础上，对每一年的印刷方案重新规划，并将所得方案和原方案对比。主要对照原方案和模型所得方案的利润，以及模型所得方案是否能将印次减少到 3 次以内。

经验证和修正过后的模型，我们将得到的预测订单总数带入安全库存模型，通过安全库存模型得到下一年的印刷方案。在此方案的基础上，对 A1-A5 图书，我们可以通过第 3 章开头提到的方法计算利润。

4.2 问题二

基于第三章对问题二的分析，该问题的求解与问题一类似，主要包括预测下一年总销售量、估计下一印次印刷成本、基于经济订购批量模型在库存积压最少的前提下制定印刷方案、根据往年数据验证模型精度、以所得模型预测结果为基础，形成印刷方案并给出预估利润等步骤。在本节，处理这些步骤的方法将一一展开。

4.2.1 下一年销售周期销量预测

为此，我们首先要对数据进行预处理。在附录 2 中，B1 和 B2 两类图书按照自然年的自然月进行销量统计，但 B1 类图书和 B2 类图书的销售周期分别为 3 月到 9 月，10 月到次年 4 月。因此第一步，我们将附录 2 的数据按照销售周期做统计，即：

B1 类图书 3-9 月分别统计，10-2 月并入 9 月；

B2 类图书 10-4 月分别统计，5-9 月并入 4 月。

以上步骤可以有效地除去数据中离散的噪声点，同时保持数据原特征不受影响。

基于 3.2.1 的分析可知，B 类图书的年销售量整体有较稳定的趋势，因此在本问题的求解中，我们可以采用问题一中所用的 ARIMA 时序分析模型对下一年销量做预测分析。该模型的详细说明可参照 4.1.1，为节约篇幅，本节不做赘述。

4.2.2 经济订购批量模型 EOQ

经济订货批量模型，又称整批间隔进货模型 EOQ 模型，是目前大多数企业最常采用的货物订购方式。该模型适用于整批间隔进货、不允许缺货的存储问题，即某种物资单位时间的需求量为常 D ，存储量以单位时间消耗数量 D 的速度逐渐下降，经过时间 T 后，存储量下降到零，此时开始定货并随即到货，库存量由零上升为最高库存量 Q ，然后开始下一个存储周期，形成多周期存储模型。该模型的示意图如图 1 所示：

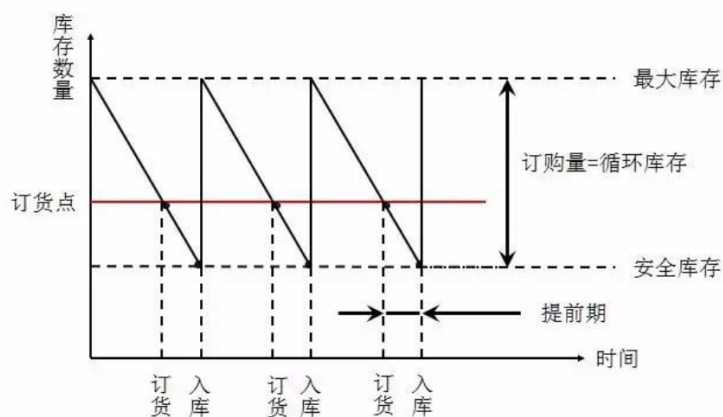


图 1 经济订购批量模型 (EOQ) 示意图。

经济订购批量模型需要解决两个问题：

- 多长时间补充一次货源；
- 每次货源补充数量是多少。

该模型有不允许缺货，且备货时间短的特点。模型建立的假设条件如下：

- 需求是连续的，均匀的，分析期内需求量固定不变；
- 当存储降为 0 时，可立即得到补充（即生产时间或拖后时间很短，可近似地看成 0）；

- 每次订货量不变，记为 Q ，订购费不变（每次生产量不变，装配费不变）；
- 不允许缺货，缺货费用无限大；
- 单位存储费不变；

由于不允许缺货，所以实际上当前该模型只考虑存储费用和订货费用。

在本文的模型中，我们另有如下假设：

- 模型中订货提前期为 3 天；
- 库存费用是当前库存的线性函数；
- 初始库存为 0；

定义如下参数： D 表示 B 类图书年需求量； K 表示单位印刷成本，即订货时每次印刷的成本开销； h 表示每件商品发货费用，在本文中由于假设发货费率不变，则该成本在确定图书类型后作为常数处理； Q 表示单位订货量；总成本以 C 表示；以 T 代表销售周期长度；日平均需求量为 d 。于是，根据 EOQ 原理以及默认的模型参数设置，我们有如下模型：

$$C = \frac{KD}{Q} + \frac{hQ}{2} = \frac{KD}{1} + \frac{Q}{2}pc \cdot sr, \quad (13)$$

可见当 C 取最小值时，

$$Q' = \sqrt{\frac{2KD}{pc \cdot sr}}, \quad (14)$$

此处由于年需求量 D 与订货量 Q 之间没有严格的数值关系，为了更符合问题二的实际情况，在求总印刷成本时，对印刷次数向上取整。即：

$$\frac{KD}{Q} \geq \text{roundup}\left(\frac{D}{Q}\right)K, \quad (15)$$

取再印刷点 R 为 $R = \text{roundup}\left(\frac{D}{Q}\right)K$ ，其中 $d = D/T$ 。

在本题中，基于数据的特征我们有如下分析：

由于印刷成本 = 上版费 + 装订费 + 纸张费 + 印刷费 + 印后工艺，除去每次印刷上版费固定，其余值经过验算大致与印刷数量成正比。因此，每件印刷时（装订费 + 纸张费 + 印刷费 + 印后工艺）为常数，略有误差与上版费相比可忽略不计。所以：总印刷成本 = 装订费 + 纸张费 + 印刷费 + 印后工艺。其中：装订费 + 纸张费 + 印刷费 + 印后工艺是与 Q 无关的常数，所以在这里将上版费视为 K 。

4.2.3 模型检验、方案形成及利润估计

与问题一类似的，为验证问题二的 EOQ 模型，我们将所得方案计算总成本 C 和利润 pf ，并与往年的印刷方案总成本及利润对比。

4.3 问题三

基于第三章对问题三的分析，该问题的求解主要包括根据销售热度的变化预测接下来的图书销量、结合现有库存判断是否需要重印、重印方案、以所得模型预测结果为基础，

形成印刷方案并给出预估利润、评估往年的重印策略是否合理等步骤。在本节，处理这些步骤的方法将一一展开。

4.3.1 下一年销量预测

5. 结果分析与检验

5.1 问题一

5.1.1 下一年销售总量预测

根据第 4.1.1 章的模型，我们得到了 A 类图书往年销售总量的预测数据与下一年销售总量的预测结果，并统计了往年的预测值和实际值的误差率：

$$er = \frac{sv_p - sv_r}{sv_r}, \quad (16)$$

其中， sv_r 和 sv_p 分别代表预测值与实际值。

结果统计如下表 2，其中包括了 A1,A2,A4,A5 的预测结果，由于 A3 直接采用均值作为预测结果，故无需统计。A3 的结果为： $sv_p^3 = 28610$ ，其他结果如下表展示。

布朗模型	A1-时间	实测数据	拟合数据	浮动	er	ARIMA(0,0,0)	A2-时间	实测数据	拟合数据	浮动	er
	2013	133890	135518	1628	0.012159		2015	131259	144099	12840	0.097822
	2014	135870	137370	1500	0.01104		2016	142036	144099	2063	0.014524
	2015	147746	138309	-9437	-0.06387		2017	147520	144099	-3421	-0.02319
	2016	153854	156399	2545	0.016542		2018	148981	144099	-4882	-0.03277
	2017	159517	161088	1571	0.009848		2019	141600	144099	2499	0.017648
	2018	155449	165637	10188	0.065539		2020	150500	144099	-6401	-0.04253
	2019	154577	154771	194	0.001255		2021	146800	144099	-2701	-0.0184
	2020	152300	153481	1181	0.007754		2022		144099		
	2021		150416								
霍尔特	A4-时间	实测数据	拟合数据	浮动	er	霍尔特	A5-时间	实测数据	拟合数据	浮动	er
	2013	101900	100665	-1235	-0.01212		2013	101900	100665	-1235	-0.01212
	2014	109900	109416	-484	-0.0044		2014	109900	109416	-484	-0.0044
	2015	116100	118019	1919	0.016529		2015	116100	118019	1919	0.016529
	2016	131800	126146	-5654	-0.0429		2016	131800	126146	-5654	-0.0429
	2017	134849	135772	923	0.006845		2017	134849	135772	923	0.006845
	2018	141700	144096	2396	0.016909		2018	141700	144096	2396	0.016909
	2019	148940	152128	3188	0.021405		2019	148940	152128	3188	0.021405
	2020	164689	160004	-4685	-0.02845		2020	164689	160004	-4685	-0.02845
	2021		169438				2021		169438		

表 2 A 类图书年销售总量预测与实际值，及下一年预测值。

A1 的预测结果位于表格的左上方，从结果中我们可以看出，所有年份的误差率 er 均

在 10% 以内，最大为 6%，其余都在 2% 以下。为更直观地观察和分析，我们将实测数据和拟合数据绘制成折线图，结果如下：

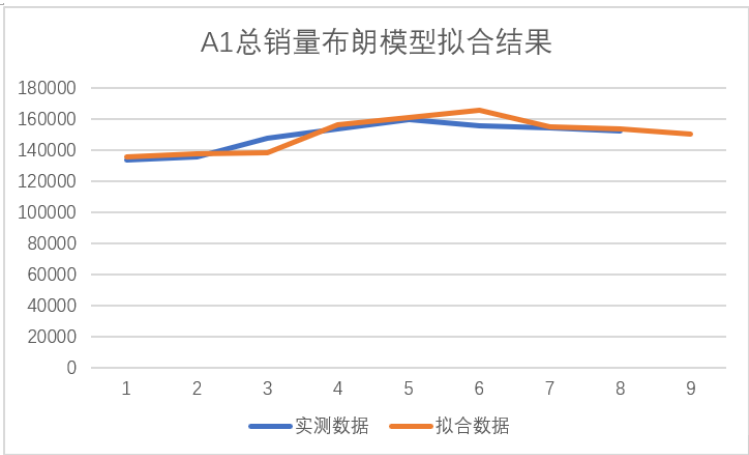


图 2 A1 总销量布朗模型拟合结果。

从图中可以直观地看到，拟合结果十分逼近。

A2 的预测结果位于表格的右上方。从图中数据可以看出，最终 ARIMA 拟合结果采用了往年历史数据的算数平均数。这是因为，A2 总销量的年变化没有稳定趋势，但总销量较稳定。

A4 的预测结果位于表格的左下方，从结果中我们可以看出，所有年份的误差率 er 均在 3% 以内，最大为 2.8%。为更直观地观察和分析，我们将实测数据和拟合数据绘制成折线图，结果如下：

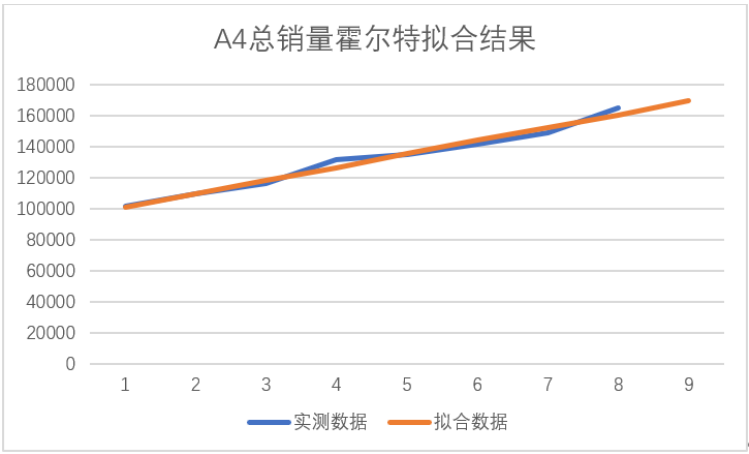


图 3 A4 总销量霍尔特模型拟合结果。

从图中可以直观地看到，拟合结果十分逼近。

A5 的预测结果位于表格的右下方，从结果中我们可以看出，所有年份的误差率 er 均在 5% 以内，最大为 4.3%，其余都在 3% 以下。为更直观地观察和分析，我们将实测数据

和拟合数据绘制成折线图，结果如下：

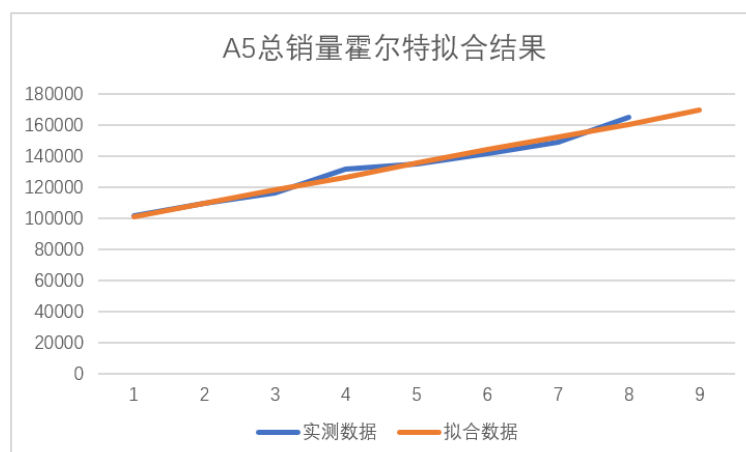


图 4 A5 总销量霍尔特模型拟合结果。

从图中可以直观地看到，拟合结果十分逼近。

从表中可以明显看出，对于 A1,A2,A4,A5 的预测结果，根据往年数据验证，其误差率均在 10% 以内，且绝大部分误差率在 3% 以内。因此，该模型对下一年总量的预测精度较高。根据该模型，我们给出了所有图书下一年预测值，记录在表 2 中每一年的最后一列（此外 A3 的总量为 28610）。

5.1.2 模型性能检验

之后，为了验证所得模型的精度，根据上述所得预测值，我们在往年数据中挑选了订单总数较多的年份。这些年份的订单根据 4.1.2 的模型重新规划印次和印数，并计算利润与往年方案对比。结果如下：

对于 A1 图书，原方案下，2019 年秋印次高达 6 次，带来的总成本为 54 万多，利润为 46 万多；而采用我们的模型得到的方案下，印次控制在 2 次，且成本大大降低，为 529 万多，利润同时有所增高，为 481 万左右。计算所得利润率，较原方案的 85%，改进方案提升为 91%。

对于 A2 图书，原方案下 2016 年春印次为 4，成本为 49 万左右，利润为 37 万；采用我们的模型可以将印次降低为 2 次，成本有略微的提升，为 49 万，利润也略微有所降低，为 37 万。该方案下印次能够降低为 2，但利润率降低了不到 1%。

对于 A3 图书，原方案下，2014 年秋印次 5 次，成本为 11 万，利润则是 -1 万多；而我们的改进方案下，印次降低为 2 次，成本缩减为 6 万多，利润提高到 3 万 8 千多，利润率达到了 61%。

对于 A4 图书，原方案下，2018 年秋印次高达 5 次，成本为 44 万，利润为 33 万；我们的方案只需要 1 次印次，成本为 43 万，利润为 35 万左右。相比原方案，利润率提高了

原 A1 方案	2019 秋第 1 次印数 2019.6.18 100000	2019 秋第 2 次印数 2019.7.5 30000	2019 秋第 3 次印数 2019.7.10 20000	2019 秋第 4 次印数 2019.8.12 2000	2019 秋第 5 次印数 2019.8.26 2000	2019 秋第 6 次印数 2019.9.2 1500	2019 秋总印数 155500	利润 463187.8	成本 547189.3	成本利润率 0.846485
印刷成本	306684	94266.14	63920.76	9652.29	9652.29	9652.29	493827.77			
改进方案	2019.7.18 151668	2019.9.9 3034					总印数 154702	480981	529396.1	0.908546
印刷成本	463472.3	12836.12					476308.44			
原 A2 方案	2016 春第 1 次印数 2015.12.11 126000	2016 春第 2 次印数 2015.12.28 15000	2016 春第 3 次印数 2016.2.1 1000	2016 春第 4 次印数 2016.3.4 138			2016 春总印数 142138	利润 370251.6	成本 491679.6	成本利润率 0.753034
印刷成本	385581.8	48748.07	7559.111	4506.758			446395.729			
改进方案	2015.12.7 119127	2016.1.5 23200	2016.2.20 1500				总印数 143827	368537.6	493393.6	0.746944
印刷成本	364725.4	73631.28	9214.95				447571.65			
原 A3 方案	2014 秋第 1 次印数 2014.8.8 12000	2014 秋第 2 次印数 2014.8.29 2000	2014 秋第 3 次印数 2014.8.18 2000	2014 秋第 4 次印数 2014.8.22 8000	2014 秋第 5 次印数 2014.8.27 5000		2014 秋总印数 29000	利润 -11967	成本 115189.2	利润率 -0.10389
印刷成本	39644.46	9652.29	9652.29	28127.15	18889.72		105965.91			
改进方案	2014.8.20 14579	2014.9.21 2538					总印数 17117	38998.89	64223.38	0.607238
印刷成本	47470.53	11308.87					58779.4			
原 A4 方案	2018 秋第 1 次印数 2018.6.29 110000	2018 秋第 2 次印数 2018.7.9 12500	2018 秋第 3 次印数 2018.7.11 13200	2018 秋第 4 次印数 2018.7.19 4300	2018 秋第 5 次印数 2018.9.13 2000		2018 秋总印数 142000	利润 338987.5	成本 442062.9	利润率 0.766831
印刷成本	301719.2	36847.86	38749.5	14993.61	8660.8		400970.93			
改进方案	2018.7.22 142708						总印数 142708	349178.9	431871.5	0.808525
印刷成本	390574.7						390574.66			
原 A5 方案	2020 春第 1 次印数 2019.11.26 90000	2020 春第 2 次印数 2019.11.28 20000	2020 春第 3 次印数 2019.12.5 27000	2020 春第 4 次印数 2019.12.10 3500	2020 春第 5 次印数 2019.12.16 6800	2020 春第 6 次印数 2020.1.3 2000	2020 春总印数 149300	利润 375105.8	成本 468604.3	利润率 0.800474
印刷成本	247386.6	57222.57	76238.98	12790.89	21877.11	8661	424177.14			
改进方案	2019.12.6 137017	2020.1.3 11116	2020.1.27 1500				总印数 149633	382585.3	461124.8	0.829678
印刷成本	375114.3	33107.49	8376.715				416598.525			

表 3 安全库存模型结果与原数据，在 A1-A5 中印次最高的年份方案对比。

3.4%。

对于 A5 图书，原方案下，2020 年春印次高达 6 次，成本为 46 万，利润为 37 万；我们的方案只需要 3 次印次，且成本略微降低，而利润增高了近 1 万。总的来说，利润率提升了接近 3%。

从上表所示的结果我们可以看到，除了 A2 的成本有少许提高外，其余 A1,A2,A3 和 A5，模型给出的方案都能将印次控制在 3 以内并降低成本。所以我们可以验证得到，我们提出的安全库存模型能够适用于问题一的情景。

在该模型的基础上，应对接下来的一年，我们只需要按照上述 4.1 的处理过程，在订单的动态变化下，对下一年的销售通过该模型给出方案即可。通过这一模型，印次可以控制在 3 次以内，而且能够实现更高的利润。

5.2 问题二

5.2.1 EOQ 模型相关系数分析结果

把相关系数的表格结果写上去，并且对重要参数进行分析

5.2.2 EOQ 模型规划结果模拟

把 EOQ 结果示意图的结果写上去，并且根据当时讨论的，对曲线图进行详细分析

5.2.3 方案形成及利润估计

根据 EOQ 模型给出下一年的印刷方案，计算利润

6. 总结

本文旨在求解工业中尚待研究的汽油辛烷值优化问题，

由于时间有限、我们队伍的学术研究水平也有限，论文中可能存在许多排版、内容上的疏漏。在解题过程中，我们也发现了许多问题，其中有很多都没能得到解决。主要包括如下：

- 本题中

参考文献

- [1] 薛冬梅.ARIMA 模型及其在时间序列分析中的应用 [J]. 吉林化工学院学报,2010,27(03):80-83.
- [2] 郑月. 不同时间序列分析法在全国公共图书馆机构数预测中的比较 [D]. 广西师范大学,2021.
- [3] 杨真真, 郭贝贝, 孙艳彬. 基于 SPSS 的指数平滑模型及实例 [J]. 中国科技信息,2021(13):74-75.
- [4] 张琼, 张爱华. 霍尔特指数平滑法在商品销量预测中的应用 [J]. 市场论坛,2012(03):70-71.
- [5] 王惠文, 孟洁.”多元线性回归的预测建模方法.”北京航空航天大学学报.04(2007):500-504.
- [6] 周志华. 机器学习 [M]. 北京: 清华大学出版社,2016: page 24-26,53-55,127-138.
- [7] ZHOU Zhi hua, HE, YIN, CHEN. (2001). A statistics-based approach for rule extraction from neural networks. Journal of Software.