

# 全国大学生数学建模竞赛论文模板

## 摘要

对于问题一,  
对于问题二,  
对于问题三,  
对于问题四,  
最后,

**关键字:** 关键词 关键词 关键词 关键词 关键词

## 目录

一、 问题重述 .....	4
1.1 问题背景 .....	4
1.2 问题要求 .....	4
二、 问题分析 .....	4
2.1 数据预处理 .....	4
2.2 问题一的分析 .....	4
2.3 问题二的分析 .....	5
三、 模型假设 .....	6
3.1 基本假设 .....	6
3.2 模型一假设 .....	6
3.3 模型二假设 .....	6
四、 符号说明 .....	6
4.1 问题一符号说明 .....	6
4.2 问题二补充符号说明 .....	6
五、 模型的建立与问题的求解 .....	8
5.1 模型一的建立 .....	8
5.2 问题一的求解 .....	8
5.3 模型二的建立 .....	10
5.4 问题二的求解 .....	11
5.5 模型三的建立 .....	12
5.6 问题三的求解 .....	12
5.7 问题四的求解 .....	12
5.8 龙格库塔法简介 .....	12
六、 模型的评价，改进与推广 .....	12
6.1 模型的评价 .....	12
6.1.1 优点 .....	12
6.1.2 缺点 .....	12
6.2 模型的改进与推广 .....	13
6.2.1 受众对象 .....	13

6.2.2 $\lambda$ 影响因素探究 .....	13
6.2.3 $\mu$ 影响因素探究 .....	14
6.3 探究出现辟谣情况时曲线的变化 .....	15
6.3.1 附加假设 .....	15
6.3.2 探究突发辟谣状况出现时间 $t_1$ 对曲线变化的影响 .....	16
七、模型的检验 .....	16
7.1 模型的改进与推广 .....	17
参考文献 .....	17

## 一、问题重述

### 1.1 问题背景

### 1.2 问题要求

1. 问题一：问题一需要通过分析各单品和品类销量的关系，需要分析单品两两之间的相关系数从而确定单品之间的相关程度。同时，销量也受时间的影响，随着时间的推移，不同单品的销量会呈现出不同的变化趋势。
2. 问题二：各蔬菜品类使用成本加成定价的方式定价，商超为了使利润最大化，需要通过分析近期的商品销量来预测商品未来销量的走势，同时根据各单品的损耗率和每天的成本进行定价。
3. 问题三：进一步细化约束条件，单日在售商品数量需控制在 27-33 个之间，并且单品订购量从第二问没有限制变成最少订购 2.5 千克。根据前 7 天的在售品种，在满足市场对商品的需求前提下，制定使 7 月 1 日的商超受益最大的单品补货量和定价策略。
4. 问题四：为了进一步提升商超的利润，从而需制定更加合理的补货和定价决策，商超需要收集更多方面的相关数据，来更加全面地分析各个变量之间的关系，最终达到提高利润的目的。

## 二、问题分析

### 2.1 数据预处理

由于题目给定数表分散，单品名称，品类名称，损耗率分布在不同表格中，因此需要进行数表合并。且本题需处理数据数量庞大，因此为防止海量的原始数据中存在着不完整(有缺失值)、不一致、有异常的数据，影响后题结果，需要对数据进行清洗。

数表合并：运用 excel 表格里的 vlookup, sumif 和 sumifs 函数，将附件 1 中品类的种类对应至其他表格中，方便题目所给问题的分析

经过筛选，发现图表中没有空白值和缺失值，依据  $3\sigma$  原则，去除  $> 3\sigma$  和  $< 3\sigma$  的异常销量值，表中存在大量负值，及存在退货情况，由于退货会对应先前一笔购买账单，在此不讲退货情况单独计入，并将退货对应的销量负值替换为排除退货销量的销量均值  $\mu$

### 2.2 问题一的分析

本题首先要知晓题目需要将品类销量与单品销量进行单独分析，我们首先对蔬菜各品类销量关系进行探究，探究其分布规律，画出各品类销量所占百分比饼图，并以月

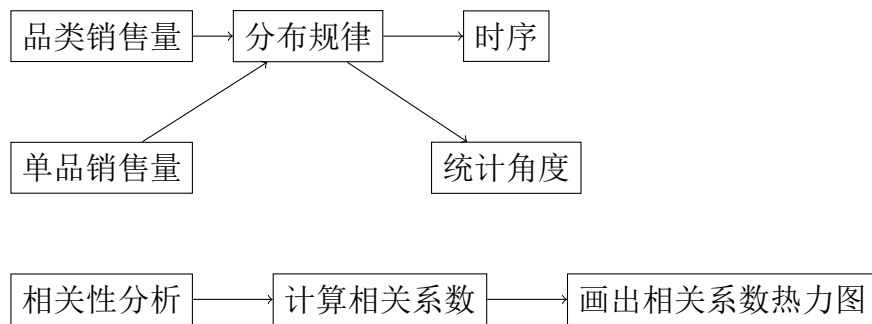


图 1 问题一分析

为单位画出其分布关系折线图，观察月与月之间品类销量变化及其相关性。之后，利用 Person 相关系数，以月份为单位，画出相关系数热力图观察品类间相关性。

其次对单品销量月变化和不同单品销量之间的关系进行分析，由于单品种类较多，无论是销量月变化条形图还是各个单品间销量，如果汇总成一张图数据繁杂，不易分析，因此对没品类之间单品进行分析，画出相关图表并总结规律。

### 2.3 问题二的分析

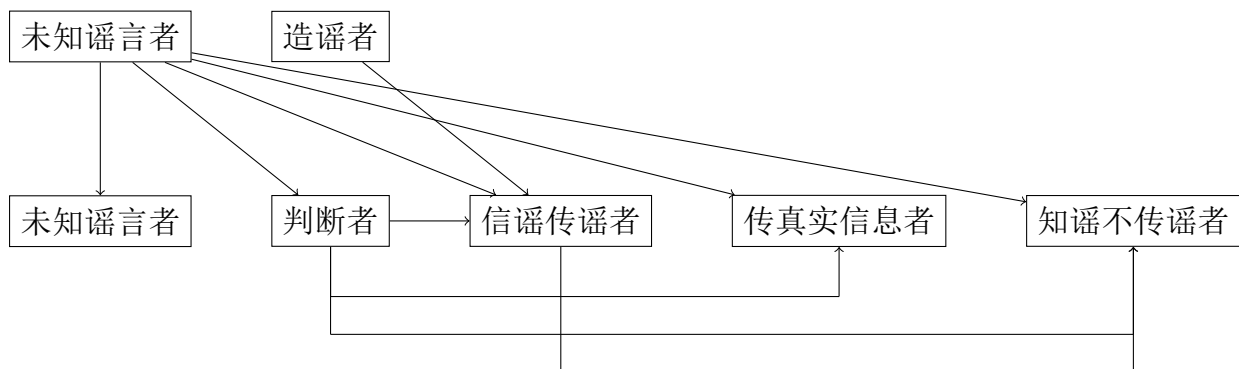


图 2 模型 2

问题二在问题一模型的基础上进一步细化，在已知间隔 4 小时后谣言已知者的人数为 200 人的情况下，依据已知数据可以对谣言传播率（单位时间内每个信谣传谣者接触人数）和转化率或者恢复率（单位时间内每位谣言传播者转化为谣言免疫者的人数）进行范围限定。同时，由于问题中未给定明确的学校总人数，其变化会相应地改变受众对象所占比例，初值以及谣言传遍全校所占时间。因此也可以通过改变学校总人数来进一步观测其对结果产生的影响。

考虑在学校传播中情况的复杂性和现实性，在本题中更加详细地考虑了谣言传播的过程并建立模型二。同时增加参数，将已知谣言者更详细地划分成四类：谣言判断者，谣信传谣者，传真实信息者以及知谣不传者。谣言未知者在传播中一部分成为谣信传谣者，一部分犹豫是否传谣而成为谣言判断者，一部分思考并接受真实信息开始传递，一

部分知谣但不传谣。判断者在经历判断后，转化成为另三种知谣者，且信谣传谣者经历一段时间的传递后会停止传播谣言，变为知谣不传者。最后当谣言未知者趋于 0 时，谣言传遍全校。

### 三、模型假设

#### 3.1 基本假设

1. 假设所研究区域——学校在限定时间内是一个封闭区域，无人口迁入迁出，因此假设人口总量稳定为  $N$ ，但不提前设定具体人数。
2. 假设在特定时间内，谣言受众对象的身份仅允许沿图 1 和图 2 中箭头所示方向切换。
3. 假设模型中所涉及的函数都是随时间连续变化的，可微的函数。

#### 3.2 模型一假设

1. 参照分析以及 SIR 模型，将谣言的受众对象分为三类，谣言未知者  $S$  占总人数比例为  $s(t)$ ，谣言传播者  $I$  所占比例为  $i(t)$ ，谣言免疫者  $R$  所占比例为  $r(t)$ ，易得  $s(t) + i(t) + r(t) = 1$ 。
2. 假设  $\lambda$  即谣言传播率为一个常数。即综合各种传播途径后得到的单位时间内每个信谣传谣者接触人数的平均值。
3. 假设  $\mu$  即转化率或者恢复率是一个常数。但实际情况可能会受到谣言自身的可信度，谣言传播的受众对象对谣言的辨别能力，外部环境的刺激比如辟谣机制的加入，以及消息传播的时效性的影响。
4. 假设不存在真相和谎言同时传播的情况。

#### 3.3 模型二假设

1. 在问题一的基础上，增添了两类受众对象即判断者  $J$  和传播真相者  $W$ ， $J$  占总人数比例为  $j(t)$ ， $W$  所占比例为  $w(t)$ ，易得  $s(t) + i(t) + r(t) + j(t) + w(t) = 1$ 。
2. 假设存在真相和谎言同时传播的情况。

### 四、符号说明

#### 4.1 问题一符号说明

#### 4.2 问题二补充符号说明

序号	符号	符号说明
1	$t$	谣言产生后经历时间 (h)
2	$s(t)$	谣言未知者占总学生比重
3	$s_0$	$t = 0$ 时谣言未知者占总学生比重
4	$i(t)$	信谣传谣者占总学生比重
5	$i_0$	$t = 0$ 时造谣者占总学生比重
6	$r(t)$	知谣不传谣者占学生比重
7	$r_0$	$t = 0$ 谣言免疫者占学生的比重
8	$\lambda$	单位时间内每个信谣传谣者接触人数 (人/h)
9	$\mu$	单位时间内每个信谣传谣者转化为谣言免疫者人数 (人/h)
10	$P = \frac{\lambda}{\mu}$	谣言传播强度

序号	符号	符号说明
1	$j(t)$	谣言判断者总学生比重
2	$w(t)$	真实信息传播者占总学生比重
3	$a$	单位时间内谣言未知者转化为谣言判断者人数 (人/h)
4	$b$	单位时间内谣言未知者转化为信谣传谣者人数 (人/h)
5	$c$	单位时间内谣言未知者转化为真实信息传播者人数 (人/h)
6	$d$	单位时间内谣言未知者转化为谣言免疫者人数 (人/h)
7	$e$	单位时间内谣言判断者转化为信谣传谣者人数 (人/h)
8	$f$	单位时间内谣言判断者转化为真实信息传播者人数 (人/h)
9	$g$	单位时间内谣言判断者转化为谣言免疫者人数 (人/h)
10	$l$	单位时间内信谣传谣者转化为谣言免疫者人数 (人/h)

## 五、模型的建立与问题的求解

### 5.1 模型一的建立

依据 SIR 模型，本题将学生分为三大类。

$$s(t) + i(t) + r(t) = 1$$

- $\Delta t$  时刻内，新增的谣言免疫人数 ( $I \rightarrow R$ ):

$$N[r(t + \Delta t) - r(t)] = \mu N i(t) \Delta t$$

令  $\Delta t \rightarrow 0$ , 得  $r$  关于  $t$  的微分方程:

$$\frac{dr(t)}{dt} = \mu i$$

- $\Delta t$  时刻内，新增的已知谣言者为:

$$N[s(t) - s(t + \Delta t)] = N \lambda s(t) i(t) \Delta t$$

令  $\Delta t \rightarrow 0$ , 得  $s$  关于  $t$  的微分方程:

$$\frac{ds(t)}{dt} = -\lambda i(t) s(t)$$

- $\Delta t$  时刻内，新增的信谣传谣者即已知者- 免疫者:

$$N[i(t + \Delta t) - i(t)] = N[\lambda s(t) i(t) - \mu i(t)] \Delta t$$

令  $\Delta t \rightarrow 0$ , 得  $i$  关于  $t$  的微分方程:

$$\frac{di(t)}{dt} = \lambda i(t) s(t) - \mu i(t)$$

联立并形成方程组:

$$\begin{cases} s(t) + i(t) + r(t) = 1 \\ \frac{ds(t)}{dt} = -\lambda i(t) s(t) \\ \frac{di(t)}{dt} = \lambda i(t) s(t) - \mu i(t) \\ \frac{dr(t)}{dt} = \mu i \end{cases}$$

### 5.2 问题一的求解

通过筛选，将数据带入上，通过 Excel 表格和 Origin 软件得到图像结果。



## 1. 蔬菜各品类销售量的分布规律及其相互规律

首先，利用函数得到六类品类销量总和，并求相应百分比，及相应品类贡献度，为下文题 3 做准备。观察可得，品类 3 年月份总销量分布大于 0 小于 25000，花叶类随月份波动最大，最大值与最小值相差约为 10000，我们选用相关系数来观察各品类之间的相关程度，由于品类销售量关于时间的变化是线性的，我们可以选用皮尔逊系数，相关系数公示如下：

$$\rho = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - EX)(Y - EY)]}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}}$$
$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

由于种类相对较多，我们采用相关系数热力图直观的反应品类两两之间的相关性，由图 (??) 可知六类品类销售额随月份变化趋势成很强的正相关性，1 5 月销售额下降，6 月开始回升，夏秋季节销量显著高于冬春季节。各个品类销售量在一定范围内波动，较为平稳。

## 2. 蔬菜各单品销售量的分布规律及其相互规律

首先利用 Excel 内置函数求出所有品类销售量随月份变化值，由于单品数量过多且无法观察到其有明显相关性，我们分类观察每个品类中单品间的分布。

- 花叶类单品销售量与月份关系: 花叶类单品种类较多，某些种类变化趋势具有相似性，云南生菜为花叶类销售总量最多单品，大部分花叶类单品与云南生菜变化趋势相同。单品间相关系数热力图如 (??) 所示。除本地菠菜为在所取时间范围内未进行销售，单品两两间相关系数可在热力图中观察得知。
- 花菜类单品销售量与月份关系: 花菜类单品种类较少，某些种类变化趋势具有相似性，西蓝花为花菜类销售总量最多单品，西蓝花，青梗散花，枝江青梗散花变化趋势相同，紫白菜未进行销售，相关性大。单品间相关系数热力图如 (??) 所示。
- 辣椒类单品销售量与月份关系: 辣椒类单品种类较多，某些种类变化趋势具有相似性，云南生菜为花叶类销售总量最多单品，大部分花叶类单品与云南生菜变化趋势相同。单品间相关系数热力图如 (??) 所示。多种单品为在所取时间范围内未进行销售，单品两两间相关系数可在热力图中观察得知。
- 茄类单品销售量与月份关系: 茄叶类单品种类较少，某些种类变化趋势具有相似性，紫茄子 (2) 为茄类销售总量最多单品，变化趋势较为相近，不同单品均在销售量在 5 到 9 月达到峰值，可能与其不同上市时间有关。单品间相关系数热力图如 (??) 所示。

- 水生根茎类单品销售量与月份关系: 花叶类单品种类较少, 单品销量变化趋势具有相似性, 单品在 1 到 5 月间下降到近乎为 0, 在 6 月后销量缓慢增长, 其销量变化与其上市和成熟季节有关, 单品间相关系数热力图如 (??) 所示。
- 食用菌单品销售量与月份关系: 食用菌单品种类较多, 某些种类变化趋势具有相似性, 销售额在秋冬季节达到峰值, 单品间相关系数热力图如 (??) 所示。多种食用菌在所取时间范围内未进行销售, 单品两两间相关系数可在热力图中观察得知。

根据品类热力系数图, 单品总热力系数图和各品类单品热力系数图, 各品类间相关系数较为接近 1, 相关性较大, 说明该商超各品类销量呈强相关性。各单品间相关系数较为多样, 可根据热力图进行查询。

### 5.3 模型二的建立

考虑到问题二的复杂性和现实性, 希望能更加细化谣言已知者的分类并进行更复杂的转化。 $s(t)$  对  $t$  的微分仍由相同原理得出。

- $\Delta t$  时刻内, 新增的信谣传谣者:

$$N[i(t + \Delta t) - i(t)] = N[(a + b + c + d)s(t)i(t)] \Delta t$$

令  $\Delta t \rightarrow 0$ , 得  $i$  关于  $t$  的微分方程:

$$\frac{di(t)}{dt} = -(a + b + c + d)s(t)i(t)$$

- $\Delta t$  时刻内, 新增的已知谣言者为:

$$N[s(t) - s(t + \Delta t)] = N(a + b + c + d)s(t)i(t) \Delta t$$

令  $\Delta t \rightarrow 0$ , 得  $s$  关于  $t$  的微分方程:

$$\frac{ds(t)}{dt} = -(a + b + c + d)s(t)i(t)$$

- $\Delta t$  时刻内, 新增的谣言判断者为:

$$N[j(t + \Delta t) - j(t)] = N[as(t)i(t) - (e + f + g)j(t)] \Delta t$$

令  $\Delta t \rightarrow 0$ , 得  $j$  关于  $t$  的微分方程:

$$\frac{dj(t)}{dt} = as(t)i(t) - (e + f + g)j(t)$$

- $\Delta t$  时刻内, 新增的传真实信息者:

$$N[w(t + \Delta t) - w(t)] = N[cs(t)i(t) - fj(t)] \Delta t$$

令  $\Delta t \rightarrow 0$ , 得  $w$  关于  $t$  的微分方程:

$$\frac{dw(t)}{dt} = cs(t)i(t) - fj(t)$$

- $\Delta t$  时刻内, 新增的谣言免疫者:

$$N[r(t + \Delta t) - r(t)] = N[ds(t)i(t) + gj(t) + li(t)] \Delta t$$

令  $\Delta t \rightarrow 0$ , 得  $r$  关于  $t$  的微分方程:

$$\frac{dr(t)}{dt} = ds(t)i(t) + gj(t) + li(t)$$

联立并形成方程组:

$$\begin{cases} \frac{ds(t)}{dt} = -(a + b + c + d)s(t)i(t) \\ \frac{dj(t)}{dt} = as(t)i(t) - (e + f + g)j(t) \\ \frac{di(t)}{dt} = bs(t)i(t) + ej(t) - li(t) \\ \frac{dw(t)}{dt} = cs(t)i(t) - fj(t) \\ \frac{dr(t)}{dt} = ds(t)i(t) + gj(t) + li(t) \end{cases}$$

#### 5.4 问题二的求解

通过查阅相关资料, 一所大学的学生人数普遍在 3000 人到 70000 人之间。在早晨 8 点制造一个谣言, 认为从此时开始只有一个人传播此谣言, 到中午也就是谣言传播了 4 小时左右后有 200 人听说了此谣言。基于此条件, 我们建立了一个关系更加复杂的模型, 并选取符合条件的参数值带入, 其中一类情况如图 ?? 和图 ?? 所示。在谣言传播之初, 谣言未知者会同时向谣言未知者、已知谣言判断者、信谣传谣者、真实信息传播者和谣言免疫者这五类人转变。因为全校人数的不同会影响  $a$ 、 $b$ 、 $c$  和  $d$  的取值, 而对  $e$ 、 $f$ 、 $g$  和  $l$  的取值影响并不明显, 所以在图 ?? 中  $a = 0.9250$ 、 $b = 1.0917$ 、 $c = 0.1750$ 、 $d = 0.0917$ 、 $e = 0.38$ 、 $f = 0.06$ 、 $g = 0.25$ 、 $l = 0.1$ ; 在图 ?? 中  $a = 0.9417$ 、 $b = 1.1083$ 、 $c = 0.1917$ 、 $d = 0.1083$ 、 $e = 0.38$ 、 $f = 0.06$ 、 $g = 0.25$ 、 $l = 0.1$ 。在占比增加的四类人群中, 因为谣言未知者中, 部分会先转变为信谣传谣者和谣言判断者, 所以两者的人数会先升高, 而这两类人最终会全部转变为真实信息传播者和谣言免疫者。因此这两类人的占比会先升高而后降低。

设定当人数占比低于千分之一就认为谣言已传遍全校。对比这两张图片中未知者占比下降到千分之一时, 也就是蓝线降到千分之一以下的时间。在图 ?? 中, 谣言开始传播 12.32 小时之后, 未知者占比已降到千分之一; 在图 ?? 中, 谣言开始传播 15.23 小时之后, 未知者占比已降到千分之一。也就是说, 在中国大部分大学校园中, 如果早晨 8 点制造一个谣言, 中午有 200 人已知, 那么在 15 个小时左右之后就会传遍全校。

## 5.5 模型三的建立

## 5.6 问题三的求解

## 5.7 问题四的求解

商超可以进一步记录消费者的信息，从而观测有哪些消费者是商超的长期消费者，哪些是临时消费者。对于长期消费者，商超需要定时去了解他们的消费习惯，比如发放问卷，分析他们的消费倾向，从而在进货策略上加以改进符合他们的消费预期。对于短期消费者，商超需要有一定的特殊性尽可能使短期消费者转变为长期消费者。比如在结账时扫码付款时监测到顾客为初次购买者，给予其打折优惠等等。

## 5.8 龙格库塔法简介

龙格库塔法 [5] 是一类解微分方程的数值算法，其中较常见的是四阶龙格库塔法，其精度较高，这里不进行推导，仅仅给出公式：

$$\begin{cases} y_{k+1} = y_k + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\ k_1 = f(t_k, y_k) \\ k_2 = f(t_k + \frac{h}{2}, y_k + \frac{h}{2}k_1) \\ k_3 = f(t_k + \frac{h}{2}, y_k + \frac{h}{2}k_2) \\ k_4 = f(t_k + h, y_k + k_3) \end{cases}$$

之后我们将微分方程组带入 MATLAB 并利用四阶龙格库塔古典模型求解。

# 六、模型的评价，改进与推广

## 6.1 模型的评价

### 6.1.1 优点

- 此模型较为简化，直观清晰地展示了谣言的传播过程。
- 此模型以微分方程组为核心，各个模块互相联系。
- 此模型为谣言传播的解决提供了多个可能的思路。

### 6.1.2 缺点

- 未给出各类参数的现实取值范围。
- 模型一忽略的因素较多，如果将其推广，需要通过调查分析得到更具体全面的数据，同时需要进一步考虑现实的偶发因素。

## 6.2 模型的改进与推广

此谣言传播模型基于 SIR 传染病模型设立，考虑因素有限，下面将展示其从受众对象， $\lambda$  和  $\mu$  的影响因素方面对此模型提供几个优化的思路。

### 6.2.1 受众对象

本文模型基于 SIR 传染病模型，在模型一中将谣言的受众对象分为了三类，即 S 易感者（谣言未知者），I 感染者（信谣传谣者）以及 R 康复者（谣言免疫者）。然而现实情况中，分类更加多样，这些人都会对谣言的传播过程产生影响。因此此模型可以从受众对象的分类上增添新的角度以此进一步细化，使模型更加全面和完善。

### 6.2.2 $\lambda$ 影响因素探究

- 传播途径的影响：

SIR 模型为了简化求解过程，将一个重要的变量  $\lambda$ （时传播率）设定为综合各个传播途径的每人每时有效接触人数的平均值，这一传播使未知谣言者变为已知谣言者。而这一个数值是一个估计值，准确度和精确性有待提高。而如果将传播途径细分为通过网络平台与人际关系即口口相传两类，可以更加精确。

下面将简单展示一个可能的优化思路：

新增参数：

- $N_1$ ：谣言传播期内通过人际关系传播方式影响的总人数
- $N_2$ ：网络平台的总数
- $A$ ：通过人际关系的方式新增的谣言已知者
- $B$ ：通过网络平台的方式新增的谣言已知者
- $m$ ：每个传播者每小时通过人际关系的方式传播谣言的有效接触人次
- $n$ ：每个网络平台每小时的有效浏览人次

其余变量与上述 SIR 模型一致不做出改变，同时，方程将做出更新。其中，此条件主要研究传播途径对时传播率即  $\lambda$  的影响，因此对于谣言的免疫者，其主要受  $\mu$  转化率的影响，因此其方程表面形态保持不变，仍为

$$\frac{dr}{dt} = \mu i(t)$$

而在  $\Delta t$  内新增的已知谣言者可以分为两类即  $A$  和  $B$  两类：

$$N[s(t) - s(t + \Delta t)] = A + B = [ms(t)N_1i(t) + ns(t)N_2] \Delta t$$

进而可以得到

$$\frac{s(t) - s(t + \Delta t)}{\Delta t} = \frac{ms(t)N_1i(t) + ns(t)N_2}{N}$$

令  $\Delta t \rightarrow 0$  得到  $s$  关于  $t$  的方程:

$$\frac{ds}{dt} = -\frac{ms(t)N_1i(t) + ns(t)N_2}{N}$$

最后, 可以得到新增的已知谣言且传播的人数:

$$N[i(t + \Delta t) - i(t)] = [A + B - \mu i(t)] \Delta t$$

同理, 令  $\Delta t \rightarrow 0$  得到  $i(t)$  关于  $t$  的微分方程:

$$N\frac{di}{dt} = [ms(t)N_1i(t) + ns(t)N_2] - \mu i(t)$$

综上所述, 可得三类人群的方程组为:

$$\begin{cases} \frac{ds}{dt} = -\frac{ms(t)N_1i(t) + ns(t)N_2}{N} \\ \frac{di}{dt} = \frac{[ms(t)N_1i(t) + ns(t)N_2] - \mu i(t)}{N} \\ \frac{dr}{dt} = \mu i(t) \\ s(t) + i(t) + r(t) = 1 \end{cases}$$

另外, 由于个体差异以及社会环境等多方面因素的影响, 人们接受到信息后对于传播信息所采取途径的倾向性各不相同。因此对于某个较大基数的群体来说, 可以通过调查统计一定基数的人群得到人们对各个传播途径的选择的概率, 这会进一步影响  $\lambda$ , 因此可以进一步细化每个信谣传谣者单位时间内有效传播谣言的人次。

### 6.2.3 $\mu$ 影响因素探究

在此模型中, 设定  $\mu$  为一个常数, 然而现实生活中,  $\mu$  受到多种因素的影响, 根据谣言的传播过程来分析, 大体可以分为四个方面: 谣言自身的可信度, 谣言传播的受众对象对谣言的辨别能力, 外部环境的刺激比如辟谣机制的加入, 以及消息传播的时效性的影响。下面将对这四个方面分别进行分析。

- 谣言自身的可信度: 谣言自身的可信度即谣言强度。在此需要参照由美国社会学家奥尔波特和波斯特曼 (1947) 提出的谣言强度公式 [4]:

$$R = I \times A$$

$R$  即谣言强度,  $I$ (Importance) 即谣言本身对受众对象的重要程度以及  $A$ (Ambiguity) 即谣言证据的模糊程度。在此问题设置的背景中, 由于期末考试临近, 期末考试相关信息较为重要, 因此同学对考试相关的信息较为敏感  $A$  值较高, 而谣言证据的模糊程度由制造谣言者决定。在此问题中, 由于谣言得到传播因此谣言的强度大于 0, 意味着  $I$  和  $A$  都大于 0。在其他因素不变的条件下,  $R$  越大意味着“治愈系数” $\mu$  越小, 然而对  $I$  和  $A$  缺乏具体的数值衡量机制。因此只能进行质上的分析, 有待进一步深化为量化的分析。

- 谣言传播的受众对象: 谣言是否能够传播很大程度上还取决于受众对谣言的判断。而受众对谣言的判断能力在没有外界干预的情况下可能受教育程度, 从众心理以及对接收途径信息来源可靠程度评估的影响。相对的大学生受教育程度较高因此  $\mu$  相对其他群体较高, 从传播途径选择的角度, 个人对信息来源可靠程度的评估可能产生影响, 这需要进一步地对受众进行调研分析, 收集到具体的数据来分析判定其对  $\mu$  的影响。由此, 以上公式根据克洛斯 (1953) 的观点可以进行进一步优化即

$$R = \frac{I \times A}{C}$$

[4] 其中  $C$  指代的是公众的批判能力。

- 辟谣机制: 在 SIR 传染病模型中, 此步即控制传染源来从源头控制传染病的传播, 通过相关的医疗技术等手段来提高感染病人的治愈比例, 是最根本的制止传播的一个方法。同理可见, 辟谣因素的加入对  $\mu$  的影响因素较为显著, 由于辟谣及时提供了官方的准确的信息, 同时还有现代传播工具的加持, 校园内谣言的受众对象有限, 传播范围有限, 因此即辟谣能大大提升受众由谣言传播者变为不传谣者的转化率。
- 时间的影响: 由于消息本身具有时效性, 因此消息本身随着横轴时间的拉长其时效性会减弱, 影响程度会减弱, 同时在时间轴拉长的背景下, 人们还可能受到记忆效应的影响, 不传谣的人数逐渐变多即  $\mu$  会随之升高。然而在此问题设置的背景下, 时间轴相对较短, 因此时间自身对谣言传播强度的影响相对较小。

对谣言传播的过程分析以及建模的最终目的是为了控制谣言, 控制谣言的核心即尽可能地提高“治愈系数”  $\mu$  (时转化率) 同时降低时传播率  $\lambda$ , 因此以上分析为谣言的控制提供了一个可能的思路。

### 6.3 探究出现辟谣情况时曲线的变化

在基本模型中, 代表“治愈系数”的  $\mu$  被设定为一个常数。然而在实际情况中, 有关考试的谣言在学校传播一段时间后会现教务澄清的情况,  $\mu$  会突然增大, 影响各类学生占比的变化。

#### 6.3.1 附加假设

1. 假设此附加讨论符合上述基本条件。
2. 考虑现实情况和文献数据,  $\lambda$  值取 3。
3. 依据期末考试前这一特殊背景综合各影响要素, 在澄清谣言之前,  $\mu$  取值为 0.2。
4. 依据学校这一特殊背景考虑到教务作为澄清者对真相可靠度的保证以及大学生作为特殊受众对象等要素, 在澄清谣言之后,  $\mu$  取值为 0.7。

### 6.3.2 探究突发辟谣状况出现时间 $t_1$ 对曲线变化的影响

当在上述条件下  $\mu$  保持不变时，得到曲线如下：

由此可得信谣传谣者先增大后减小且峰值出现在  $4s$  左右，因此我们分别在  $t = 1h$ , 图 ?? 中  $t = 3h$ , 图 ?? 中  $t = 5h$ , 图 ?? 中  $t = 8h$ , 图 ?? 处  $\mu$  的增长对曲线的影响：

当经过很短的时间就发出辟谣通告时（此处取  $t = 1h$ ），谣言免疫者最终比率无法达到 1，且谣言未知者人数比率不会达到 0，表示部分学生从始至终都是谣言未知者，且知谣传谣人数下降速度明显变快。

当  $t = 0.3h$  时，可显著看出信谣传谣人数比率  $i(t)$  达到峰值速率更快，达到峰值更早，在此种情况下，知谣不传谣人数比率会达到 0。

当辟谣时间节点在知谣传谣人数比率达到峰值之后（此处取  $t = 0.5h$ ）时，曲线变化同样出现在  $i(t)$  达到峰值之后，在  $\mu$  改变之后，信谣传谣人数比率  $i(t)$  下降速率加快，谣言免疫者比率  $r(t)$  上升速度变快。

当  $t = 0.8h$  时，此时三支曲线总体变化与  $t = 0.5h$  相似，可以看出每只曲线在  $\mu$  改变处斜率改变： $i(t)$  减小变快， $r(t)$  增长变快。

通过观察在不同时间  $t$  改变  $\mu$  后图像的变化，可以显著发现控制谣言的时间越早，谣言的扩散越可控，及时  $\lambda$  很大，只要控制时间较早，依然会有一部分学生始终处于未知谣言的状态，且信谣传谣者占比能达到的最大值越小，表示有更少学生在此过程中相信谣言，并不理智地传播谣言。

当辟谣时间  $t_1$  出现在当辟谣时间节点在信谣传谣人数比率达到峰值之前，控制时间越晚， $i(t)$  达到峰值越晚，峰值越大。在临界点后  $s(t)$  达到 0 越早，速率越快。此过程中存在使  $s(t)$  无法达到 0 的最大值。

当辟谣时间  $t_1$  出现在当辟谣时间节点在信谣传谣人数比率达到峰值之后，控制时间越晚， $i(t)$  减小到 0 越晚，即  $r(t)$  达到 1 越晚。

## 七、模型的检验

本问题中出现的方程组由于其复杂性，难以求出其解析解，我们利用数值方法得到其数值解。本篇论文中全部选取四阶龙格库塔法 [1] 古典模型用于计算，因此有必要验证方法精度。由于题中方程组难以计算解析解，我们选取简单方程  $\frac{\partial y}{\partial t} = \frac{2y}{t} + 4t^2 \cos(4t)$  并给定当  $t = 0$  时  $y$  的初值为 0 来计算其误差和收敛阶。此时解析解为  $y = \sin(4t)t^2 + t^2$ 。其误差和收敛阶如下图所示 [1, 4] 在模拟中数值解和解析解几乎完全重合。在本文中选取的  $h = 0.01$  时收敛显示为四阶，证实了本文计算的可信度。



步长	对应步长下误差	收敛率	运行时间
1.0000e-01	2.8945e-04	0.0000	1.8704
1.0000e-02	3.9204e-08	3.8683	0.6316
1.0000e-03	6.4395e-12	3.7845	1.4862

表 1  $\frac{\partial y}{\partial t} = \frac{2y}{t} + 4t^2 \cos(4t)$

## 7.1 模型的改进与推广

- 数据预处理：本题中我们采用  $3\sigma$  模型进行数据预处理，但未对数据进行更深一步的平滑化，可考虑用小波降噪对数据进行进一步处理。
- 相关性分析：在第一题中我们仅计算了相关系数，并用热力图呈现。虽然品类和单品的分类已经给定，依据热力图呈现的相关性，我们可以构建聚类模型，将具有相似性质的单品进行归类，进一步提高第三题选择可售单品使商超利润最大的选择准确性，更清晰的呈现相关性，由于此方法相关性由点呈现，可以在一张图上呈现更多单品，更加有利于多种单品的相关性分析。
- 除此以外，我们可以在问题一中运用关联分析，从支持度，置信度，提升度三个维度更加清楚的呈现商品被同时购买的概率，为第三小题选择单品提供参考
- 在第二问中我们采用 ASMA 模型预测回归，我们也可以采用 LSTM 模型，LSTM 是一个优秀变种模型，继承了大部分 RNN 模型的特性，解决了梯度反传过程由于逐步缩减而产生的 Vanishing Gradient 问题，可以使预测更加准确。
- 问题二和问题三在有利润方程的前提下，设置一定约束条件，可以采用线性规划模型进行运算得到更准确的结果。

## 参考文献

- [1] 韩中庚. 数学建模方法与应用 [A]. 高等教育出版社,2005:6.
- [2] 王道谊, 王霞. 传播动力学-SIR 模型及其运用 [A]. 2020:3.
- [3] 司守奎. 数学建模算法与运用 [A]. 2021:4.
- [4] 尼古拉斯·迪方佐. 谣言心理学 [A]. 机械工业出版社,2021:8.
- [5] 房少梅, 王霞. 微分方程数值解 [A]. 国防工业出版社,2016:5.