

# 数学建模论文排版

## 摘要

这里写摘要，国赛论文摘要要求是一页最好，不要多也不要太少。

**关键词：** Fisher 精确检验 多元线性回归 系统聚类 灰色关联分析

## 一、问题重述

问题重述部分正文

## 二、问题分析

### 2.1 问题一的分析

这里是第一段的内容。

### 2.2 问题二的分析

### 2.3 问题三的分析

## 三、基本假设与符号说明

### 3.1 基本假设

- 假设理论物理跟泵不存在；
- 假设数据中未填写的数据项为 0
- 假设所提供的数据准确无误；
- 不考虑因检验手段等原因对数据值的影响。

### 3.2 符号说明

## 四、问题一模型的建立与求解

针对每批零配件，假定总量为  $N$ ，我们考虑采用异常检测的经典取样方法：序贯概率比检测  $SPRT$  作为抽检方案。在此之前我们考虑每次取样的样本量为  $D_i$ ，令单个零件次品与否的布尔值为  $x$ ，考虑其单次试验成功 (为次品) 概率的期望为  $\mu$ ，则其显然服从经典的二项分布表示：

$$Bern(x|\mu) = \mu^x(1 - \mu)^{1-x} \quad (1)$$

接下来考虑其在样本集上的对数似然函数，针对第  $i$  次取样  $D_i$ ，对其中的每个样本取到观测  $x_1, x_2, \dots, x_n$ ，根据题目要求样本集中零配件的次品产生事件可认定为相互独立的。则其似然函数可写为：

$$\mathbf{P}(D_i|\mu) = \prod_{n=1}^N p(x_n|\mu) = \prod_{n=1}^N \mu^{x_n}(1 - \mu)^{1-x_n} \quad (2)$$

为便于后续处理，我们取其对数似然：

$$\begin{aligned}\ln \mathbf{P}(D|\mu) &= \ln \prod_{n=1}^N \mu^{x_n} (1-\mu)^{1-x_n} = \ln \mu \sum_{n=1}^N x_n + \ln(1-\mu) \sum_{n=1}^N (1-x_n) \\ &= \ln \mu \sum_{n=1}^N x_n + \ln(1-\mu) (N - \sum_{n=1}^N x_n) = \sum_{n=1}^N x_n \ln \mu + (1-x_n) \ln(1-\mu)\end{aligned}\quad (3)$$

接下来我们依据题干给定零假设和备择假设：

$$\begin{cases} H_0 : \mu > 0.1 \\ H_1 : \mu \leq 0.1 \end{cases}\quad (4)$$

题干中的两种情况意味着拒真和纳伪的显著性水平  $\alpha$  和  $\beta$  分别为 0.05 和 0.1。在 *SPRT* 语境下，考虑决策边界：

$$A = \ln \frac{\beta}{1-\alpha} \quad B = \ln \frac{1-\beta}{\alpha}$$

于是，针对每次采样  $D_i$ ，我们需要求出在零假设和备择假设下的似然比  $LR$ ：

$$LR = \frac{\sum_{n=1}^N x_n \ln \mu_0 + (1-x_n) \ln(1-\mu_0)}{\sum_{n=1}^N x_n \ln \mu_1 + (1-x_n) \ln(1-\mu_1)}\quad (5)$$

需要注意的是，在原生的 *SPRT* 场景中， $H_0$  和  $H_1$  一般被认定为较为复杂的参数估计  $\theta_0$  和  $\theta_1$ ，这取决于它们事先假定样本服从一个较为严谨且高度可表达的概率分布。然而基于问题一，在没有明确历史数据和概率分布的先验情况下，我们只能将其建模为一般二项分布，为了遵循 *SPRT* 的使用场景，我们将二项分布参数建模为  $\mu_0 = 0.1 + \Delta\mu$ ， $\mu_1 = 0.1 - \Delta\mu$ 。通过轻微扰动量来拟合样本的分布与所报称值的差异，扰动量的设置取决于样本量的大小，这点我们将在后续给出实验和说明。

尽管在许多场景中单样本取样策略以及被证明取得了很好的效果，但考虑到题干背景，我们依然选择样本集作为采样标准。遵循 *SPRT* 方法，给定总零配件量  $N$ ，初次取样  $D_i$  应为按照标称值所取的总样本配比，即  $D_1 = 0.1N$ ，而后计算出当前样本下的对数似然比  $LR_1$ 。序贯检验比方法遵循以下停止法则：

$$\gamma = \inf\{n | n \geq 1, LR_n \in (A, B)\}\quad (6)$$

具体来说，若  $LR_1 \leq A$ ，接受  $H_0$  假设；若  $LR_1 \geq B$ ，接受  $H_1$  假设，否则继续采样。初次采样的样本量为  $D_1 = 0.1N$ ，假定每次采样的次品数为  $n_i$ ，则此后每次采样量依据以下法则确定：

$$D_{i+1} = D_i - n_i\quad (7)$$

检验的完整流程可以作出如下表示：