



锂电池的容量分布及最优目标容量优化的研究报告

中南大学 数学与统计学院 生产实习 指导老师：方秋莲 副教授
实习单位：湖南华慧新能源股份有限公司 指导人员：余浩、冷正

报告人：中南大学数学与统计学院统计学 16 级实习队 第一分队

黄文韬 统计 1602 班 1303160224
李 绚 统计 1602 班 0702160318

2019 年 7 月 1 日 - 2019 年 7 月 19 日

目录

1	准备工作	1
1.1	引言	1
1.2	华慧锂电池系列介绍	1
1.3	问题提出	2
1.4	问题假设	2
1.5	数据搜集与整理	3
2	电池容量分布模型	4
2.1	电池容量分布的识别	4
2.2	EM 算法的实现	5
2.3	异常电池型号不同批次分布对比	6
2.4	标称容量的更新	6
2.5	容量分段科学性的讨论	7
2.6	理论次品率	8
2.7	固有废品率	9
3	最优容量的凸优化模型	10
3.1	符号说明	10
3.2	凸优化模型的推导与实现	11
3.3	标准差对净利润曲线的影响	13
4	改进与建议	15
4.1	模型的改进	15
4.2	数据管理的建议	15
5	致谢	16
6	附录	17
6.1	图像	17
6.2	代码	20

第 1 章 准备工作

1.1 引言

早在 1800 年，意大利物理学家伏打 Alessandro Volta 发明了“伏特电堆”，被称为是人类历史上的第一个真正意义上的电池组，被视为电气文明的一个重要起步。随后，许多科学家例如英国的丹尼尔 Daniel 对“伏特电堆”进行不断改进，陆续研发出更好性能的电池。考虑到化学电池在蓄电上存在一些弊端和一定的危险性，1887 年英国的赫勒森 Wilhelm Hellesen 发明了最早的干电池，其电解液不会溢漏，便于携带，因此获得广泛应用。如今，干性电池的种类非常之多，常见的有锌-锰干电池，镁-锰干电池等等。在整个电池发展的过程中，锂离子电池可谓最令人瞩目，由于该电池在充放电过程中，只有锂离子，其优势十分明显：能量重量比高、电压高、自放电小、无污染、循环寿命长、可长时间存放。由于这些优点，锂离子电池在近 30 年中取得巨大发展，被广泛使用于计算机、计算器、照相机等电器。

1.2 华慧锂电池系列介绍

湖南华慧新能源股份有限公司——一家专业从事超级锂离子电池研发、设计、生产、组装及营销的民营公司。与传统锂离子电池相比较，产品具有使用更安全，尺寸任意设计，一致性好等突出优点。公司自主研发全自动生产设备，既保证了产品的品质稳定性和一致性，又避免了因进口自动化设备带来的高成本。

表 1.1: 电池概况一览表

系列名称	材料	应用场景
HTC 系列	钛酸锂	电动玩具，LED 照明，电动工具，储能基站，数码科技
HFC 系列	磷酸铁锂	电动玩具，LED 照明，电动工具，储能基站
HCC 系列	钴酸锂	电动玩具，汽车电子，个人护理产品，数码产品
HMC 系列	锰酸锂	电动玩具，LED 照明，电动工具，数码科技
NSC 系列	三元锂	电动玩具，LED 照明，电动工具，储能基站

公司生产的超级锂离子电池目前有 HCC、HFC、HMC、HTC、NSC 等五大系列，广泛应用于各种动力航模、移动通讯、电动工具、电动玩具、储能后备电源、便携式电子产品、安防产品、笔记本电脑、音响设备、电动自行车等领域。随着公司研发工作的进展，公司产品几

乎可以覆盖所有电池应用市场。公司已通过 ISO9001 质量体系以及 ISO14001 环境体系认证, 超级锂离子电池已经获得 CE、ROHS、MSDS 等安全认证。产品品种齐全, 钴系、锰系、铁锂、钛酸锂电池已经批量生产, 尺寸可以根据客户要求量身定制, 从而满足不同行业客户的需求。五大系列电池的型号、材料与应用领域如上表所示。

1.3 问题提出

本小组根据所负责的课题, 结合公司的生产实际, 提出了以下问题:

- 1) 选取 2019 年 6 月份的生产批次, 对若干种型号电池在同一批次的恒流放电容量的分布进行研究, 探究是否服从正态分布; 以实际分布的均值作为实际电容标称, 通过软件拟合得到理论标称, 将额定标称与实际标称、理论标称进行比较, 考察是否存在差异?
- 2) 对于以上选取的电池型号, 通过比较六月份与四月份或五月份的历史数据, 来观察现在和过去的产品电容分布情况是否存在差异; 若存在差异, 如何解释这些差别?
- 3) 公司原定的分容区间在数理统计的意义下是否合理, 应该如何改进?
- 4) 对于实际生产得到电容标称和额定电容标称有差异的型号, 如何设定理论电容使得一个批次产品的收益最大化? 通过增加原材料, 提高成本而使得良品率得到提高, 从而销售额增多; 减少原材料, 保持良品率在一定的水平下则可以缩减成本。因此, 产品的实际电容标称应该为何值, 才能使得净利润率最大化?

1.4 问题假设

为了方便理论推导, 并且结合生产实际, 我们作了如下假设:

- 1) 抽样的每种型号电池容量的总体分布服从正态分布。
- 2) 每一批产品的容量在一次分容时完全确定, 不考虑二次分容。
- 3) 按照事先规定好的分容区段, 实际电容比额定电容标称低 1 档容量区段的产品为次品。
- 4) 一个批次中非次品即为良品, 不同容量的良品售价一致。
- 5) 同一型号的极片成本与其容量成正比, 配套成本和人工成本不变。
- 6) 不同电池分容箱不存在差异, 且分容柜的每个测试数据都是真实有效的。
- 7) 测试容量比标称容量低 2 档的产品为废品, 一批产品中废品的比率为常数。
- 8) 一批产品的目标容量可以控制, 但在原有生产模式下方差不变。

1.5 数据搜集与整理

电池容量的数据均来自分容柜的控制台处，公司将所有的分容柜分了 A-H 组进行管理，每组均隶属于一个控制台。通过查找对应型号的流水单所对应的时间和分容柜组，我们搜集整理了 2019 年 4-6 月份的产品参数数据。所选取的型号分别为：HTC1450-500、HFC1020-70、HFC1340-350、HCC1020-90、NSC1040-250，用以探讨这些产品的恒流放电容量的分布；后选取 5 种主要产品：HFC1340-350、HFC1650-750、HCC0840-170、NSC1040-260、HFC1865-1300，用以探讨重新设定产品的电容标称使达到最大净利润。我们收集到的型号数据及样本容量如下表所示。

表 1.2: 电池型号及样本容量表

产品型号	样本容量	产品型号	样本容量	产品型号	样本容量
HTC1450-500	1849	HFC1020-70	4218	HFC1340-350	2857
HCC1020-90	5476	NSC1040-250	4379	HFC1650-750	2048
HCC0840-170	2401	NSC1040-260	2400	HFC1865-1300	1608

第 2 章 电池容量分布模型

2.1 电池容量分布的识别

从我们普遍的认知来看，一批电池的容量分布应该大致服从正态分布。为了证实这种特性，我们选取了 2019 年 6 月份出产的 5 种型号电池，分别为其容量数据绘制了密度分布图，发现 HFC1020-70、HFC1340-350、HCC1020-90、NSC1040-250 这四种电池的容量分布都大致符合正态分布的单峰钟型曲线的特征。其中灰色的区域为样本的真实分布，红色曲线是用正态分布核函数拟合的核密度曲线，可以看出样本大致服从正态分布，但具有一定的厚尾特征。

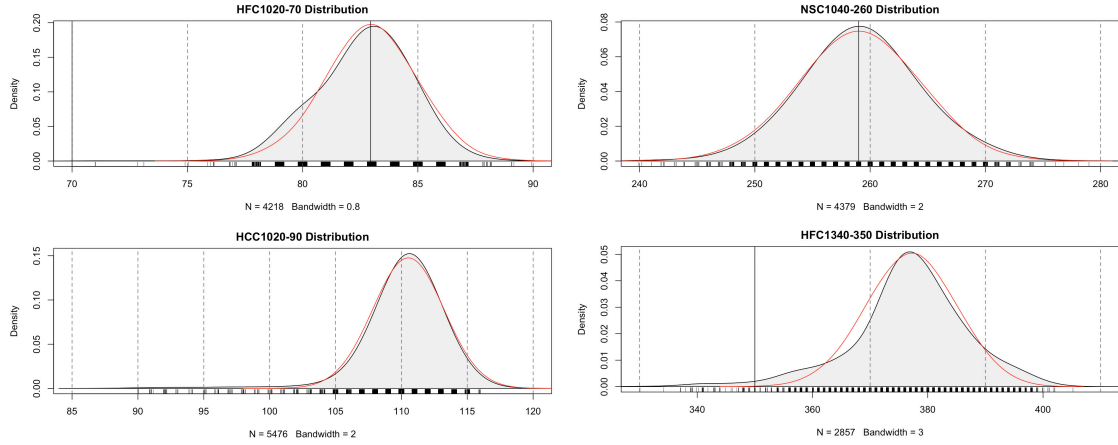


图 2.1: 四种型号电池容量分布图

但 HTC1450-500 的密度分布图却呈现出双峰的现象，与其他四种有显著的不同。为了更好的描述该双峰的分布，我们假设该分布为混合高斯分布，考虑采取 EM 算法估计该混合分布的均值和方差参数。

$$f(x) = \frac{\theta}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x - \mu_1)^2}{2\sigma_1^2}\right) + \frac{1 - \theta}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(x - \mu_2)^2}{2\sigma_2^2}\right) \quad (1)$$

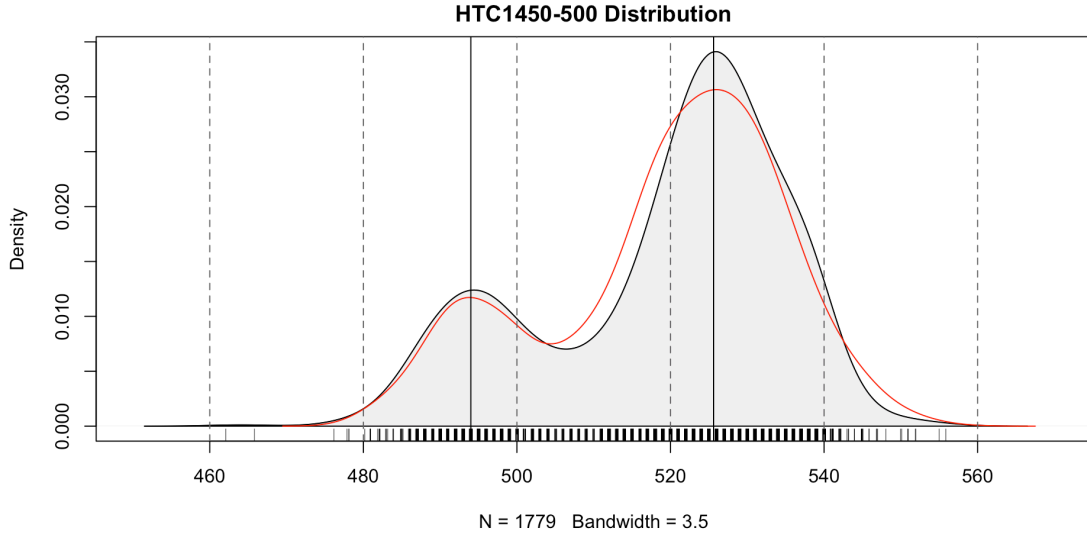


图 2.2: HTC1450-500 容量分布图

2.2 EM 算法的实现

EM 算法是一种迭代算法，主要用于计算后验分布的众数或极大似然估计，常用于估计混合模型以及含有隐变量的模型。对于 n 个独立同分布的样本： $X = (x_1, x_2, x_3, \dots, x_n)$ ，其对数似然函数可写为：

$$L(\theta) = \sum_{i=1}^n \log p(x_i|\theta) = \sum_{i=1}^n \log \left(\int p(x_i, z_i|\theta) dz_i \right) \quad (2)$$

其中 $Z = (z_1, z_2, z_3, \dots, z_n)$ 为隐变量，但是函数直接最大化很难，EM 算法的原理是建立极大似然估计下界的函数，求极大值从而不断迭代逼近目标函数的极大值，重复两步直到收敛，求极大值的过程用到了詹森不等式以及收敛证明，推导较为繁琐，故在此处省略。

通过 EM 算法，我们得到 HTC1450-500 混合正态分布的参数为：

$$\mu_1 = 490.93, \sigma_1 = 5.48, \mu_2 = 525.61, \sigma_2 = 9.79, \theta = 0.09 \quad (3)$$

其他型号通过拟合正态分布，得到均值及方差参数的估计值为：

表 2.1: 不同型号电池的参数估计

产品型号	样本均值 μ	样本标准差 σ
HFC1020-70	82.95	1.87
HFC1340-350	377.18	7.54
HCC1020-90	110.53	1.82
NSC1040-260	259.04	5.01

从上表可见，前三种产品均值参数估计都大于额定的标称容量，这是由于为了实现较高的良品率，会事先将产品的额定容量设定的高一些；最后一种产品估计值和额定值差异不大，这是由于 NSC 系列电池的生产难度限制了所能设定的电容量参数，若设定过高则会在制作过程中产生一些安全隐患。

2.3 异常电池型号不同批次分布对比

由于在 6 月份的数据中只有一种电池型号的数据呈现混合正态分布，其余皆可近似为正态分布，为了进一步确定分布的差异背后的原因，我们采集了 HTC1450-500 在 5 月份在某一分容箱上的某一批次的数据，绘制出来的密度分布图，该蓝色密度曲线呈现出了单峰钟型的正态分布特征。最终估计出的正态分布参数为： $\mu = 505.42$, $\sigma = 9.23$ 。

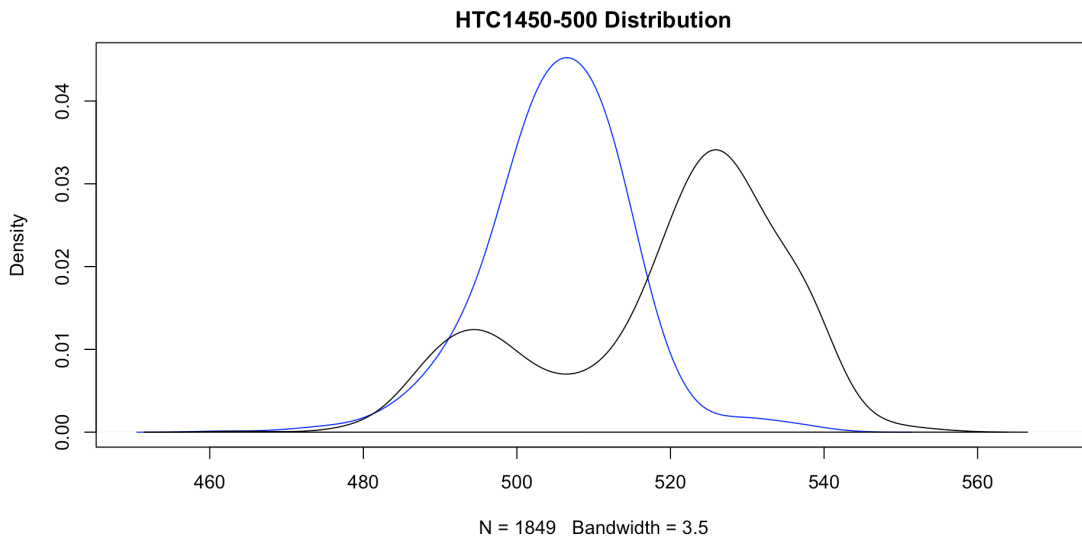


图 2.3: HTC1450-500 不同批次容量分布对比图

可以看出，之前该型号电池的容量分布呈双峰形状并非是该产品的一个特征，而是数据出现了异常。通过对生产和分容流程的了解，我们认为导致双峰出现的原因可能是以下两种：

- 1) 对同一个批次的电池分容时会混用分容箱，由于实际上不同分容箱得到的电池容量参数会有稍许差异，在采集数据过程中未做区分，导致来自两个或多个分容箱的数据混合，呈现混合分布的情况。
- 2) 混用分容箱的另一种结果使得批次型号登记错误，即来自两个生产批次的电池登记为一个同一个批次，同样也会使得服从不同正态分布的数据混合在一起。

2.4 标称容量的更新

通过前五种型号的电池容量分布图可以看到，由于技术的提升或是降低次品率为目的，电池实际投产时的容量都普遍高于标称容量。但这也造成了一个问题，就是我们无法再拿标称

容量作为电池实际容量的参考，标称容量的设置也就没有意义了。并且有的电池由于投产时的实际容量设置过高，整批产品都显著高于次品标准，不但升高了成本，次品率也不会再降低，但又由于标称容量的限制，该产品的的售价却不能升高。

为了避免这些问题，我们根据上述五种型号的电池分布数据，给出了该五种型号电池更为合适的标称容量，可以作为更改其他电池标称容量的参考。可以达到在提高定价和降低成本的同时，也不显著升高次品率的作用。

表 2.2: 标称容量更新表

产品型号	原标称容量	新标称容量	产品型号	原标称容量	新标称容量
HTC1450-500	500	500	HFC1020-70	70	80
NSC1040-260	260	260	HCC1020-90	90	105
HFC1340-350	350	370			

2.5 容量分段科学性的讨论

在生产中，电池的容量不会严格等于预设的容量，而是服从特定的分布。为了区分一批产品中哪些电池达到了要求，哪些没有，我们需要对该批电池进行分容，该过程在分容柜上面实现。但分容只能得到每个电池的电容是多少，而不能对次品做出判断。次品的判断需要人为设定标准，公司给不同标称容量的电池设置了不同的分段区间长度，从 5 到 10，15 到 20 不等。一个分段称作一档，技术人员将一批电池分为不同的档次，可以针对不同需求的客户。

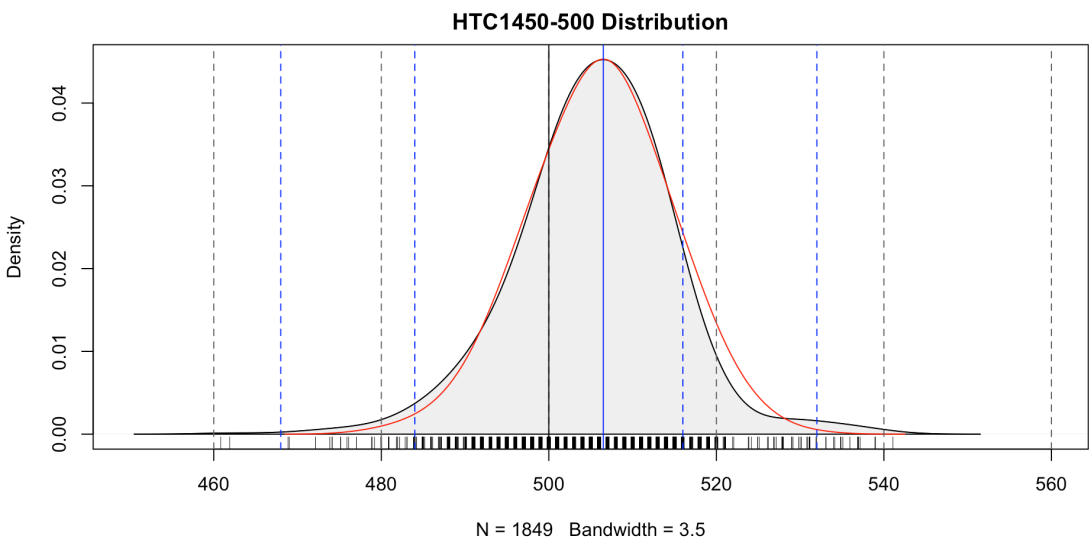


图 2.4: HTC1450-500 新旧分段标准对比图

但这样的分段明显来自于经验，但不知道这样分段是否科学。本小组尝试利用拟合的正态分布的标准差作为标准的新分段标准。通过对样本的实证分析，在保证分段具有一般性的

同时，并有一定的次品率，我们认为或许以取整的两倍标准差（即 $[2\sigma]$ 标准）为新分段标准更为科学。我们以 HTC1450-500 为例，绘制出了原分段标准和新分段标准的比较图如上。

其中黑色实线为标称容量，灰色虚线为原分段标准，蓝色实线为样本均值，蓝色虚线为新分段标准。从该图和上表可以看出，利用 $[2\sigma]$ 方法生成的新分段标准和原分段标准相差不大。并且由于批次的不同，样本的标准差也会上下浮动。故我们认为原有的分段标准是比较科学的，在误差容许的范围内可以不用改动。现给出五种型号电池新分段标准如下表。

表 2.3: 标称容量更新表

产品型号	原分段标准	新分段标准	产品型号	原分段标准	新分段标准
HTC1450-500	20	18	HFC1020-70	5	4
NSC1040-260	20	20	HCC1020-90	5	6
HFC1340-350	20	16			

2.6 理论次品率

由于我们假设电池容量的分布服从正态分布，故利用样本估计出均值和标准差两组参数后就可以利用正态分布进行研究。黑色曲线为容量的实际分布，尾部的灰色区域即为次品率。蓝色曲线为样本参数拟合的正态曲线，红色曲线为最优的正态曲线。由于容量分布都呈现出厚尾特征，故我们采用蓝色曲线的尾部作为该型号电池的理论次品率 $\alpha(\mu)$ 。

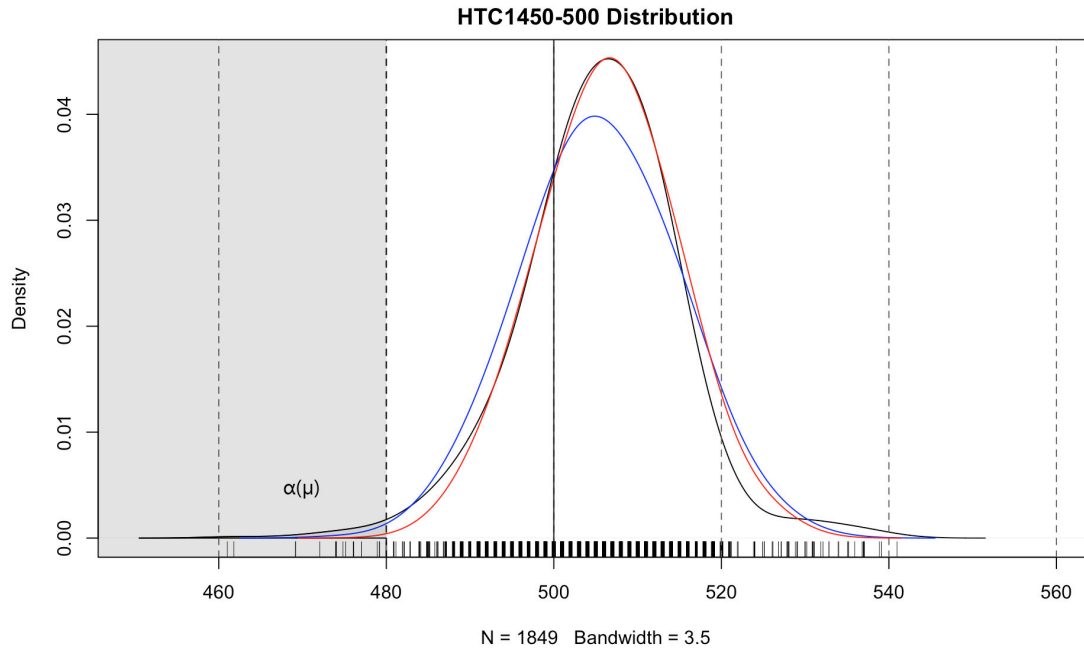


图 2.5: HTC1450-500 的实际次品率与理论次品率

2.7 固有废品率

我们在讨论电池的容量分布时，其实是在讨论所有接近目标容量的正常值。但在我们收集到的原始数据中，存在一部分样本的容量远小于平均容量或者直接等于 0。这些异常值可能是产品本身的问题，也可能是分容柜的触点出现了损坏。为了简化问题，我们不考虑分容柜出现损坏，只考虑这些都是产品生产本身的问题。我们定义容量低于比标称容量低 2 档的产品为废品。一批产品的废品率 β 由生产工艺本身决定，与正常样本的分布无关，所以我们认为它是一个定值。固有废品率是一批产品的固有属性，不像次品率 $\alpha(\mu)$ 是一个随着目标容量 μ 变化的函数，而是一个常数。我们绘制了 HTC1450-500 的真实分布，可以看到固有废品率 β 对整批产品的影响不能被忽视。

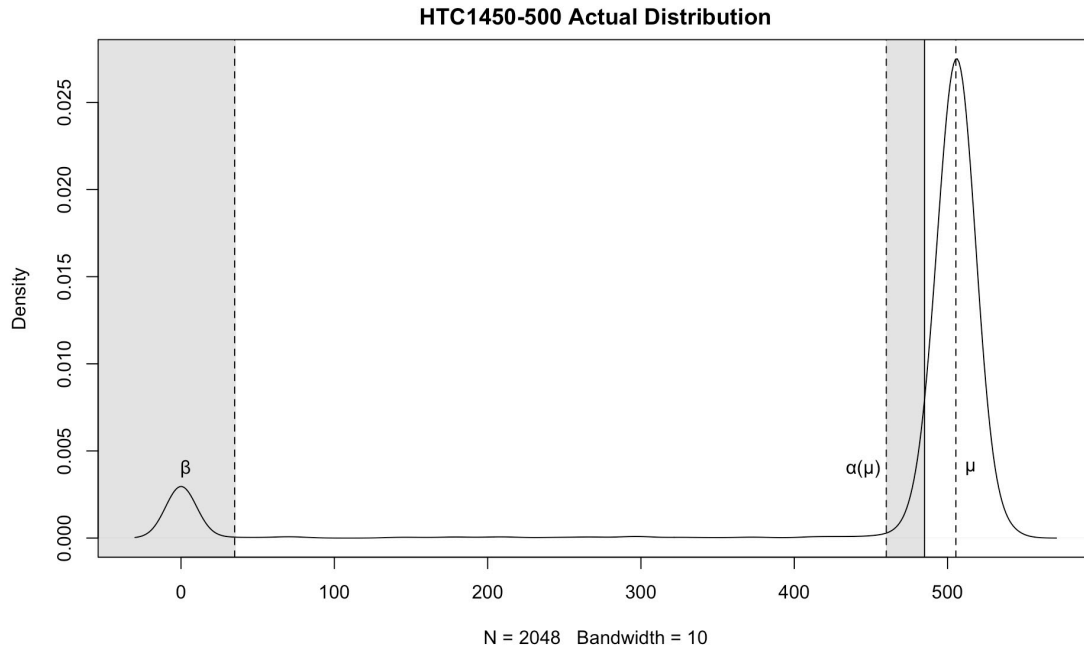


图 2.6: HTC1450-500 的实际容量分布

第3章 最优容量的凸优化模型

从前文可以得知，公司为了降低次品率，有意地提高了生产的实际容量。通过这样的处理，虽然可以有效地降低该批次的次品率，但是却增加了用料，升高了成本，最终降低了利润。我们假设在原有的标称容量和分段标准的条件下，定义次品为比标称容量低一档的产品。次品是不能销售的，是整批中的沉没成本。也就是说，在正态分布的假设前提下，实际容量与次品率之间存在着一个均衡。即我们可以构造一个净利润函数，其中实际容量与次品率相互制约。通过寻找该函数的极值点，解出使得净利润最大所对应的实际容量，即为在理想条件下能够使得一批产品获得最大利润的最优容量。

3.1 符号说明

为了方便数学推导，我们制定一系列变量的符号如下：

表 3.1: 电池的参数与含义

符号	表达式	含义	符号	表达式	含义
μ_0	-	标称容量	n	$n_\lambda + n_\alpha + n_\beta$	样本总数
c_0	-	标称极片成本	β	n_β/n	固有废品率
μ	-	目标容量	$\bar{\mu}$	$\frac{1}{n} \sum_i x_i$	样本均值
$c_m(\mu)$	$c_0/\mu_0 \times \mu$	理论极片成本	σ	$\sqrt{\frac{1}{n-1} \sum_i (x_i - \bar{\mu})^2}$	样本标准差
c_a	-	配套成本	k	$(\mu_0 - d - \mu)/\sigma$	辅助变量
c_p	-	制造成本	$f(x, \mu)$	$\frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{(x-\mu)^2}{2\sigma^2}]$	正态密度
$C(\mu)$	$c_m(\mu) + c_a + c_p$	总成本	$\phi(x)$	$\frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$	标准正态密度
d	-	分段区间长度	$\Phi(x)$	$\int_{-\infty}^x \phi(t)dt$	标准正态累计
x_i	-	第 i 个样本	$\alpha(\mu)$	$\int_{-\infty}^{\mu_0-d} f(x, \mu)dx$	理论次品率
n_λ	$\sum_i 1_{x_i \geq \mu_0-d}$	样本良品数	p	-	电池单价
n_α	$\sum_i 1_{\mu_0-2d < x_i < \mu_0-d}$	样本次品数	$G(\mu)$	$p[1 - \beta - \alpha(\mu)]$	总收益
n_β	$\sum_i 1_{x_i \leq \mu_0-2d}$	样本废品数	$Q(\mu)$	$G(\mu) - C(\mu)$	净利润

3.2 凸优化模型的推导与实现

净利润为总收益减去总成本，其中总收益是单价和良品（即生产总数减去废品和次品）的乘积。假设废品率为固有值，次品率为正态分布的尾部累计。总成本为极片成本、配套成本和制造成本之和。对于同一型号而言，实际容量的变动会引起极片成本的改变，而不影响配套成本和制造成本。根据上述思想，构造净利润函数如下：

$$Q(\mu) = p \left(1 - \beta - \int_{-\infty}^{\frac{\mu_0 - d - \mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \right) - \left(\frac{c_0}{\mu_0} \mu + c_a + c_p \right) \quad (4)$$

为了简化计算，设辅助变量 $k = \frac{\mu_0 - d - \mu}{\sigma}$ ，将上式化为如下形式：

$$Q(k) = p \left(1 - \beta - \int_{-\infty}^k \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \right) + \left(\frac{c_0 \sigma}{\mu_0} k + \frac{c_0}{\mu_0} d - c_0 - c_a - c_p \right) \quad (5)$$

为了找寻 $Q(k)$ 的极值，需要对该积分变限函数进行求导，并令其导函数为 0：

$$Q'(k) = -\frac{1}{\sqrt{2\pi}} p e^{-k^2/2} + \frac{c_0 \sigma}{\mu_0} = 0 \quad (6)$$

解得 $Q(k)$ 极值点 k^* 如下：

$$k^* = -\sqrt{2 \ln \left(\frac{\mu_0 p}{\sqrt{2\pi} c_0 \sigma} \right)} \quad (7)$$

将变量 k 还原为 μ ，得到理论最优容量 μ^* 如下：

$$\mu^* = \mu_0 - d + \sigma \sqrt{2 \ln \left(\frac{\mu_0 p}{\sqrt{2\pi} c_0 \sigma} \right)} \quad (8)$$

通过收集数据和计算，我们得到了 5 种型号电池的各项参数如下表所示：

表 3.2: 五种型号电池各项参数

参数	HFC1340-350	HFC1650-750	HCC0840-170	NSC1040-260	HFC1865-1300
μ_0	350	750	170	260	1300
c_0	0.4	0.75	0.45	0.5	1.6
c_a	0.3	0.46	0.2	0.25	1.2
c_p	0.3	0.49	0.22	0.38	0.8
d	20	20	10	20	20
β	0.04	0.05	0.05	0.04	0.05
$\bar{\mu}$	376.77	824.96	181.93	271.86	1398.35
σ	9.93	17.99	3.24	6.73	15.9
p	7.69	7.99	3.98	6.28	6.5

将这些参数带入上式，可以得到上述 5 种型号电池的最优容量、实际容量和标称容量对应的净利润及利润率与次品率如下表所示：

表 3.3: 五种型号电池理论容量和净利润的最优解、实际解与标称解

参数	HFC1340-350	HFC1650-750	HCC0840-170	NSC1040-260	HFC1865-1300
电池单价 p	7.69	7.99	3.98	6.28	6.5
最优容量 μ^*	363.23	787.89	170.47	261.84	1329.71
最优利润 Q^*	6.36	5.85	2.91	4.89	3.53
利润率 r^*	82.76%	73.18%	73.05%	77.89%	54.35%
次品率 α^*	0.04%	0.06%	0.06%	0.06%	0.09%
实际容量 $\bar{\mu}$	376.77	824.94	181.93	271.86	1398.35
实际利润 \bar{Q}	6.35	5.82	2.88	4.88	3.45
利润率 \bar{r}	82.60%	72.79%	72.35%	77.64%	53.14%
次品率 $\bar{\alpha}$	0.00%	0.00%	0.00%	0.00%	0.00%
标称容量 μ_0	350.00	750.00	170.00	260.00	1300.00
标称利润 Q_0	6.21	4.83	2.91	4.89	2.90
利润率 r_0	80.80%	60.41%	73.04%	77.86%	44.68%
次品率 α_0	2.20%	13.31%	0.10%	0.15%	10.42%

以 HFC1340-350 为例，绘制出其净利润曲线，可以看出该曲线存在最优解。

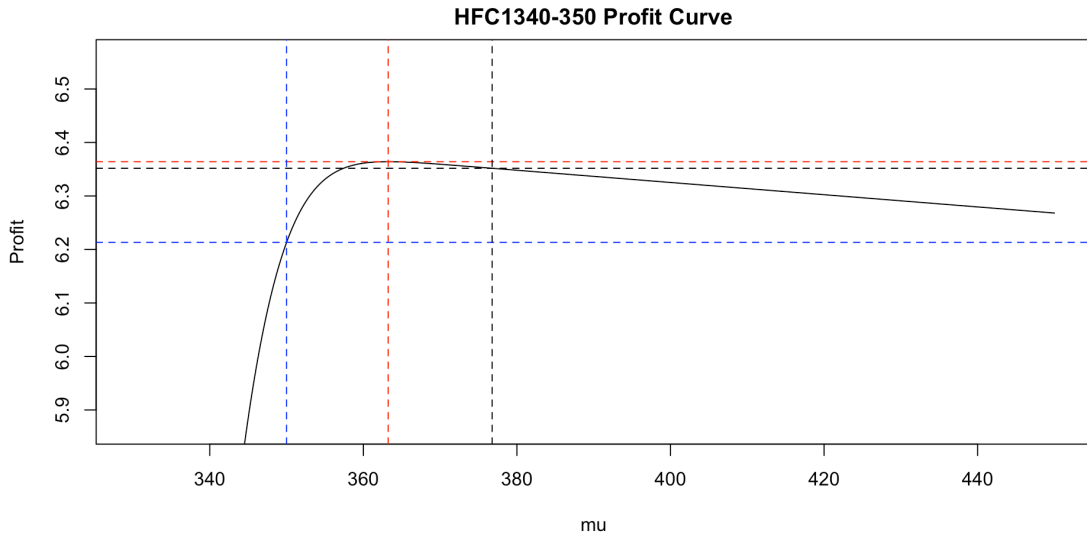


图 3.1: HFC1450-500 收益曲线

其中红色虚线为最优容量 μ^* 及其对应的最优净利润 Q^* ，蓝色虚线为标称容量 μ_0 ，黑色虚线为实际容量 $\bar{\mu}$ 。曲线呈现出在达到最优容量之前离最优容量越远净利润下降越快，达到

最优容量之后净利润下降突然变缓的现象。故当实际容量大于最优容量之后，再升高实际容量虽然会提高成本但对整体利润影响不大。故降低次品率对利润升高的影响显著地高于提高容量对利润降低的影响。所以公司采用提高容量的生产方式来降低次品率是比较好的策略。

3.3 标准差对净利润曲线的影响

标准差 σ 是对一批产品围绕着目标容量分散程度大小的度量。 σ 越大，一批产品对于目标容量越为分散； σ 越小，一批产品对于目标容量越为集中。现在我们以 HFC1450-500 为例，固定其他参数，研究标准差 σ 的变动，对最优容量和净利润曲线形状的影响。分别令 $\sigma = \{5, 9.93, 15\}$ ，绘制净利润曲线图。同时，为了考察 σ 连续变动的情况，我们也绘制了 μ 和 σ 同时变动对净利润影响的三维图如下。

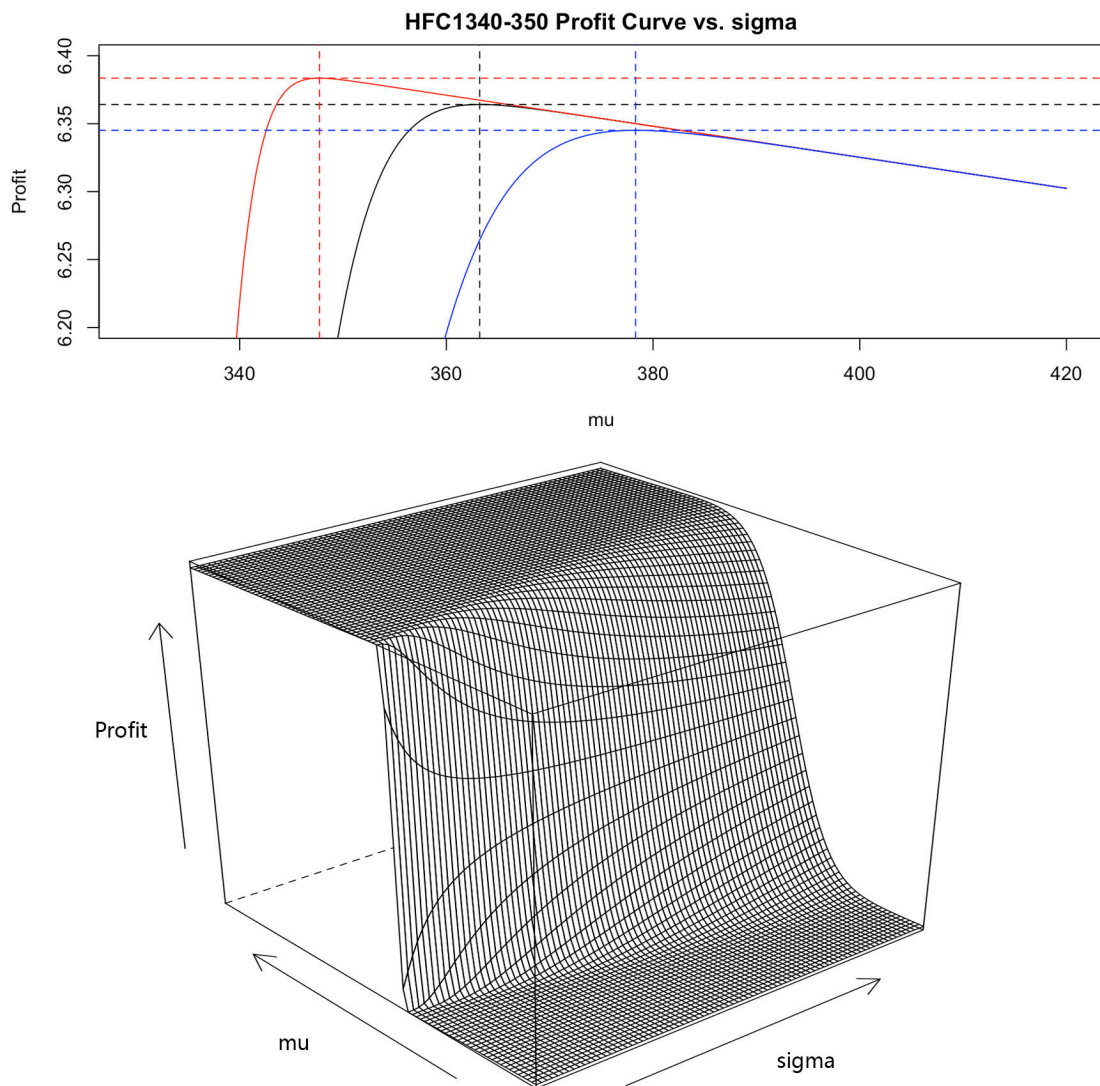


图 3.2: HFC1450-500 标准差 σ 的变动对净利润曲线 $Q(\mu)$ 的影响

从上图可以看出, σ 的变动会显著影响净利润曲线的形状: σ 越大, 就需要更大的目标容量才能抵消次品率带来的负面影响, 并且即使达到了最优利润, 其值也小于 σ 较小时的最优利润。从下图可以看到, 一批产品的目标容量 μ 和标准差 σ 都会影响到该批次的理论次品率 $\alpha(\mu)$, 但 μ 对 $\alpha(\mu)$ 的影响也远高于 σ 对 $\alpha(\mu)$ 的影响。

通过绘制随着 σ 变动的最优容量和最优利润曲线图可以看出, σ 与最优容量成正相关, 与最优利润成负相关。故我们认为, 在目标容量 μ 不变的情况下, 标准差 σ 应该越小越好; 在标准差 σ 不变的情况下, 目标容量 μ 应该越大越好, 但也不应过大, 目标容量过大的话, 材料成本对净利润的影响也会被相应放大。

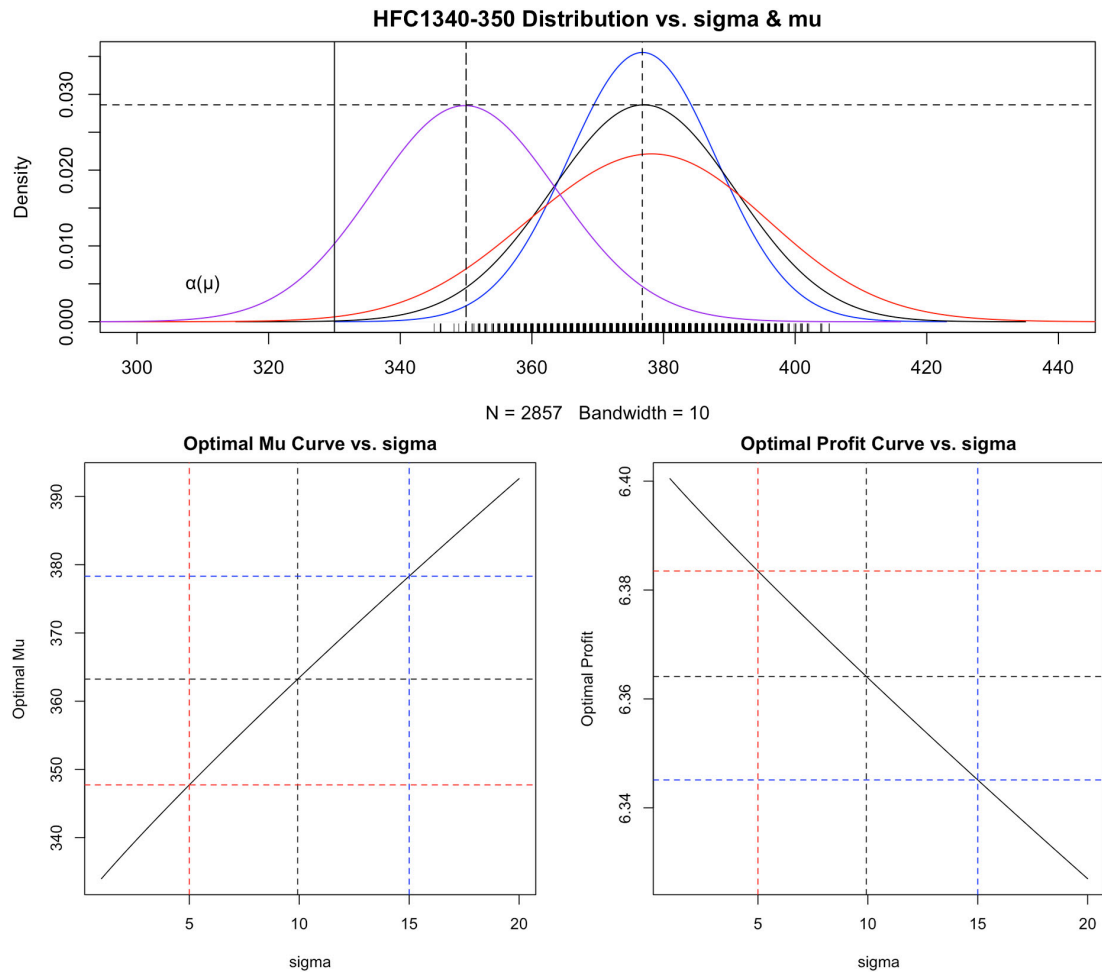


图 3.3: HFC1450-500 标准差 σ 的变动对最优容量、最优利润和理论次品率的影响

在生产实际中, 一批产品的标准差 σ 受到原料、工艺、设备及人力等等诸多因素的影响。故我们建议在生产的每个环节都进行优化来减少每批产品的标准差 σ , 在不显著影响净利润的情况下尽可能提高目标容量 μ 。最终达到提升净利润 $Q(\mu)$ 、降低成本 $C(\mu)$ 的目的。

第4章 改进与建议

4.1 模型的改进

本文虽然经过上述模型得出了一系列结论，但我们的模型还有以下改进的空间：

- 1) 在数据搜集过程中，由于操作不熟练，有可能导致不同分容箱的数据混在一起，使得无法探究是分容箱的差异还是同一批次混着多个批次导致电池容量出现混合分布。如果能够获得更为准确的数据，模型的精确度会更高。
- 2) 本文未对抽取样本的分布进行严格的检验，而近似看作为正态分布。但其实样本可能服从其他的分布，这样根据不同的分布，分析手段和结论可能会不同。
- 3) 由于抽取样本的随机性，以这些样本作为代表对总体进行重新定容，但是由于实际生产工艺存在差异，不同样本也存在较大差异，因此无论采取哪一组样本用来估计总体都会存在一定的误差。解决的方法是尽可能多地获取同一型号的数据，样本数越接近总体，参数估计的精度也会越高。
- 4) 在上文的讨论中我们假设固有废品率 β 是产品生产本身的属性，而不是分容柜出现了问题，在实际生产中废品的产生往往是由于分容柜的问题，故可以根据实际情况对固有废品率进行调整。

4.2 数据管理的建议

在采集数据的过程当中，我们也遇到了一些本应该避免的问题，希望公司能够在今后的工作中进行适当的规范：

- 1) 规范文件管理流程，包括：对文件规范命名，按日期或批次整理存档。
- 2) 公司应该对电池分容箱进行检修，替换老化较严重的机器，能够有效减少电池误判为废品或导致容量测量值不准确的情况。

第5章 致谢

经过近三个星期的实习与研究，我们小组在完成了公司所提出的问题的同时，在对数据的步步深挖中，也发现了新的问题，最终得以较好地解决。为公司往后的生产提供了较为具有指导意义的建议。

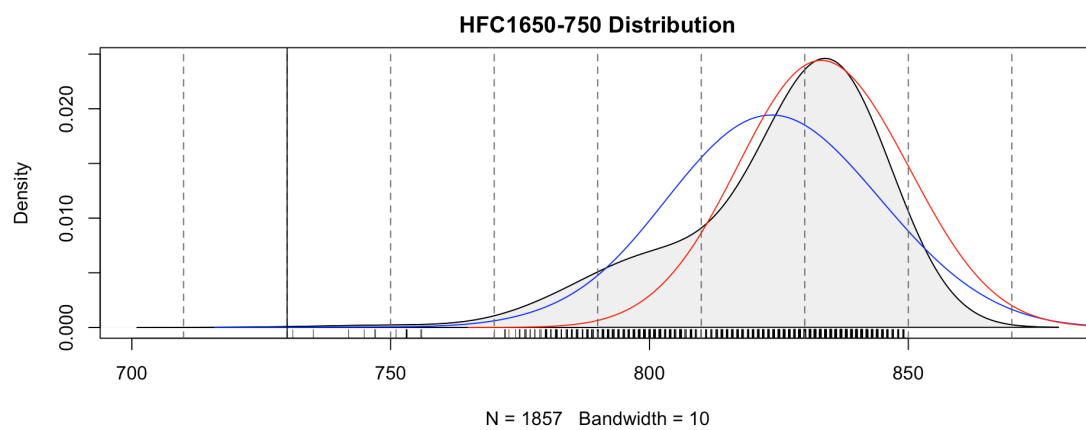
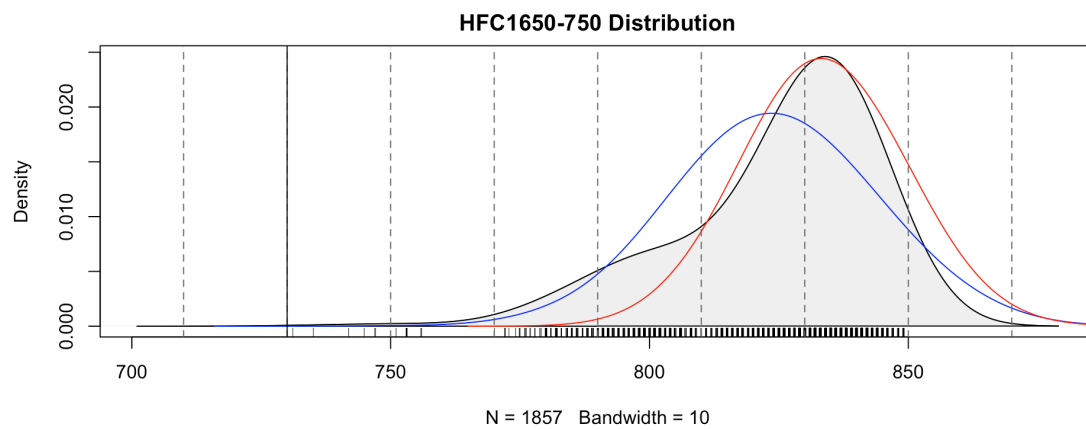
我们小组最终完成了本次生产实习的课题，没有身后他们的支持，光靠小组的两位成员是远远不够的。我们首先要感谢湖南华慧新能源股份有限公司为我们提供了本次实习的平台，让我们可以在实际生产中得到锻炼。其次是张总和汤总，两位不但对我们从始至终的热情接待，还给了我们实习生在公司内部足够的权限，并提供了办公场所和健康丰富的饮食，这些都是我们能顺利完成任务的必要条件。再者我们要感谢余工和冷工两位指导老师，是他们为我们本次的课题提出了问题，并在权限以内尽可能地满足我们的数据要求，为我们构建模型，解决问题提供了重要的建议与指导。然后我们要感谢中南大学数学与统计学院和方教授为我们开设了生产实习的实践课程，让我们从中得到了锻炼。最后我们要感谢本次到华慧公司参加实习的全体同学，大家在实习的三周里相互帮助，相处融洽，互相为对方的课题提供建议，一起克服困难，最终都完成了所承接的课题。

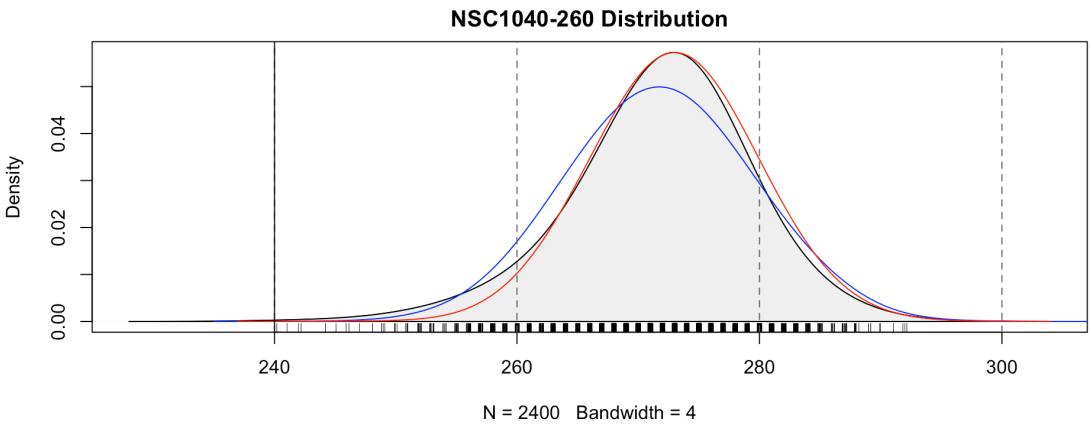
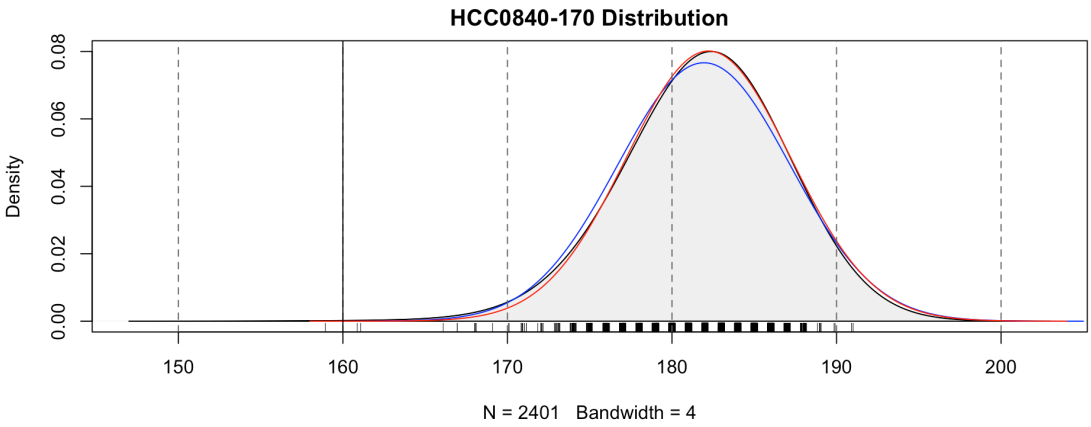
希望华慧公司能够通过我们最终的研究报告对公司的实际生产有所指导，并达到提升业绩的效果。也希望我们自己能在本次生产实习中获得宝贵的经验，为我们今后的职场生涯打下良好的基础。

第6章 附录

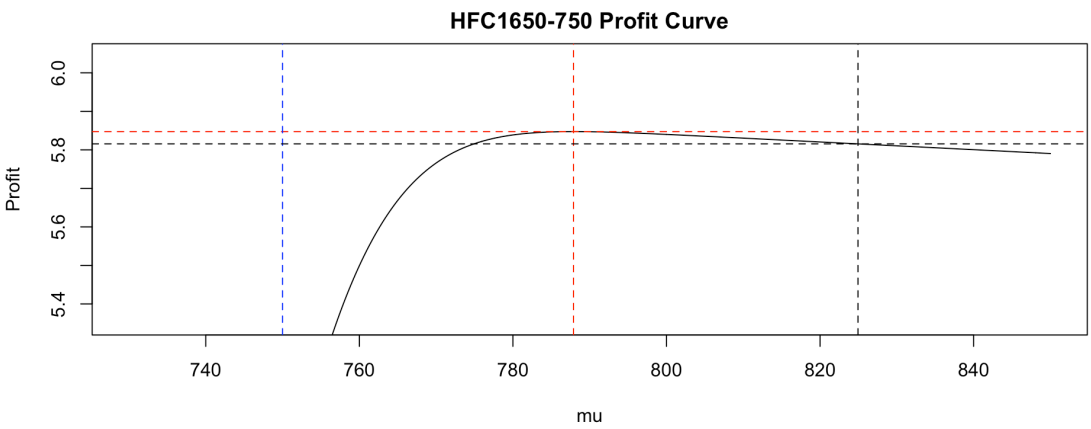
6.1 图像

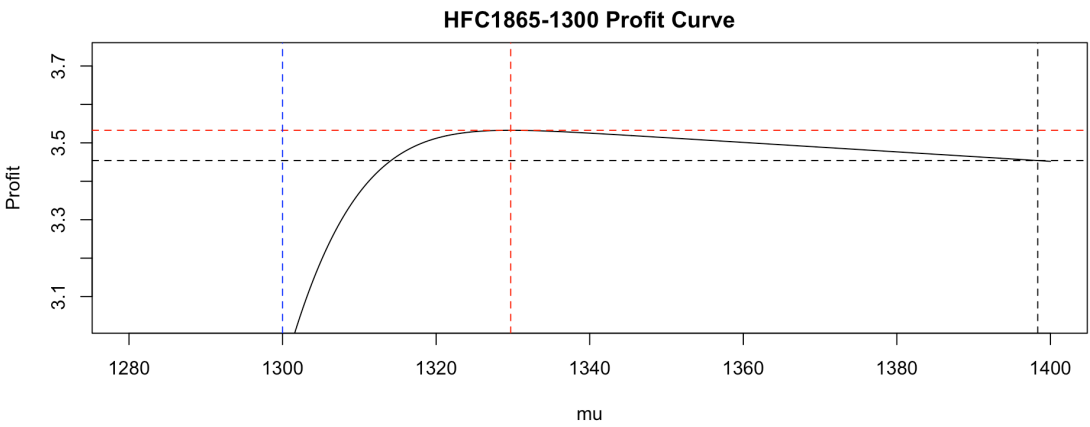
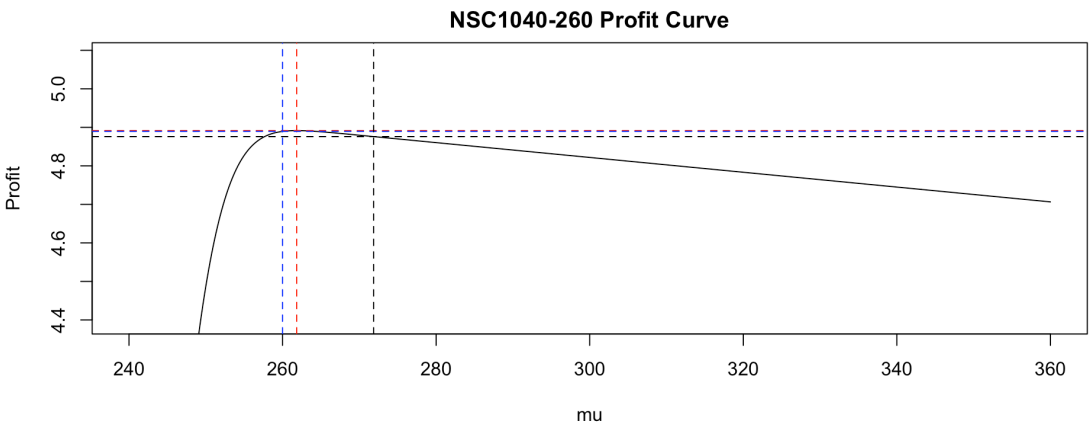
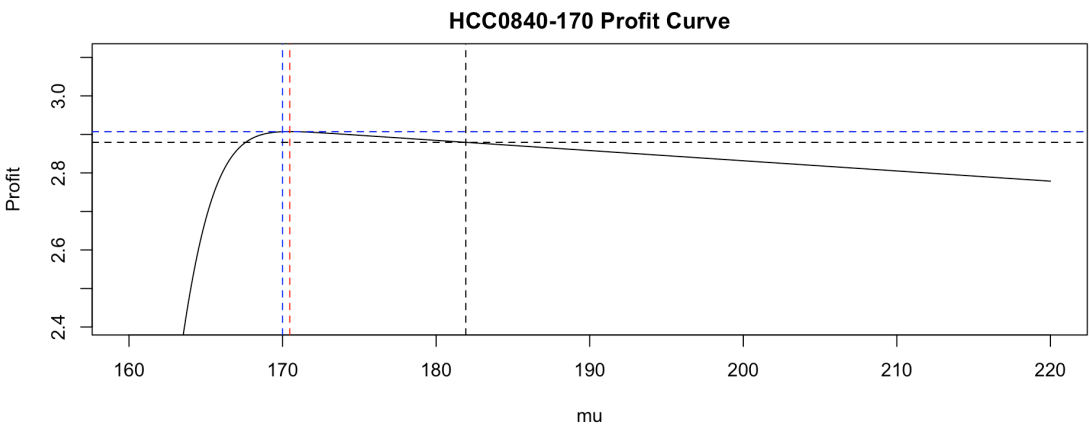
电池容量分布图及其拟合曲线





电池的最优容量、实际容量和标称容量对应的净利润





6.2 代码

电池容量分布模型

```

1  set.seed(666)
2  a <- read.csv("20T.csv")
3  b <- read.csv("1450.csv")
4  X <- as.matrix(a$VOL[which(a$VOL>=460)])
5  Y <- as.matrix(b$VOL[which(b$VOL>=460)])
6  sd(X)
7  mu1=510
8  mu2=530
9  sigma1=16
10 sigma2=16
11 N=length(X)
12
13 cat("mu1",mu1,"sigma1",sigma1,"mu2",mu2,"sigma2",sigma2)
14 n<-N
15 mean_s <- c(mu2, mu1)
16 y <- sample(c("head", "tail"), size = n, replace = TRUE, prob = c(0.8, 0.2))
17 x <- round(rnorm(n = N, mean = mean_s[1],sd=sigma2))
18 tails <- y %in% c("tail")
19 x[tails] <- round(rnorm(sum(tails), mean = 494,sd=6))
20 x<-as.data.frame(x)
21 plot(density(X,bw=3.5),xlim=c(450,570), main="HTC1450-500_Distribution")
22 polygon(density(X,bw=3.5),col="#F1EFF1",border="black")
23 lines(density(x$x,bw=3.5),xlim=c(450,570),col="red")
24 rug(jitter(X), side=1, col=1)
25
26 plot(density(Y,bw=3.5),xlim=c(450,570), main="HTC1450-500_Distribution",col="blue")
27 polygon(density(X,bw=3.5),border="black")
28
29 y <- round(rnorm(n = N, mean = 506,sd=8))
30 plot(density(Y,bw=3.5),xlim=c(450,560), main="HTC1450-500_Distribution")
31 polygon(density(Y,bw=3.5),col="#F1EFF1",border="black")
32 lines(density(y,bw=3.5),xlim=c(450,560),col="red")
33 rug(jitter(Y), side=1, col=1)
34 abline(v=c(460,480,500,520,540,560),col="gray40",lty=2)
35 abline(v=c(500),col="black")
36 abline(v=c(506.5),col="blue")
37 mean <- 500
38 sd <- floor(sd(Y))
39 abline(v=c(mean-2*sd,mean+2*sd,mean+4*sd,mean-4*sd),col="blue",lty=2)

```

EM 算法

```

1  N=length(X)
2  j=1
3  phi0<-0.5
4  w1<-matrix(0,1,N)
5  w2<-matrix(0,1,N)
6  pnorm<-function(x,mu,sigma){

```

```

7   pf<-1/(2*pi*sigma)*exp(-(x-mu)^2/(2*sigma^2))
8   return(pf)
9 }
10 a<-NULL
11 b<-NULL
12 c<-NULL
13 d<-NULL
14 while(j<=100){
15   for (i in 1:1000){
16     Z=pnorm(X[i,1],mu1,sigma1)*phi0+pnorm(X[i,1],mu2,sigma2)*(1-phi0)
17     w1[1,i]<-pnorm(X[i,1],mu1,sigma1)*phi0/Z
18     w2[1,i]<-pnorm(X[i,1],mu2,sigma2)*(1-phi0)/Z
19   }
20   new_p1=sum(w1)/N
21   new_p2=sum(w2)/N
22   New_mu1=w1\% * \% as.matrix(X)/sum(w1)
23   New_mu2=w2\% * \% as.matrix(X)/sum(w2)
24   New_sigma1=sum((w1\% * \% (as.matrix(X)-as.matrix(c(rep(New_mu1,N))))^2) /sum(w1)
25   New_sigma2=sum((w2\% * \% (as.matrix(X)-as.matrix(c(rep(New_mu2,N))))^2) /sum(w2)
26   oldphi0=phi0
27   phi0=new_p1
28   a<-c(a,mu1)
29   b<-c(b,mu2)
30   c<-c(c,sigma1)
31   d<-c(d,sigma2)
32   mu1=New_mu1
33   mu2=New_mu2
34   sigma1=sqrt(New_sigma1)
35   sigma2=sqrt(New_sigma2)
36   j=j+1
37 }
38
39 cat("mu1",mu1,"sigma1",sigma1,"mu2",mu2,"sigma2",sigma2)
40 n<-N
41 mean_s <- c(mu2, mu1)
42 y <- sample(c("head", "tail"), size = n, replace = TRUE, prob = c(0.8, 0.2))
43 x <- round(rnorm(n = N, mean = mean_s[1],sd=sigma2))
44 tails <- y \% in \% c("tail")
45 x[tails] <- round(rnorm(sum(tails), mean = 494,sd=6))
46 x<-as.data.frame(x)

```

正态分布拟合与分段

```

1 library(fitdistrplus)
2 library(MASS)
3
4 a <- read.csv("36T.csv")
5 X <- as.matrix(a$VOL[which(a$VOL>=330)])
6 X0 <- as.matrix(a$VOL[which(a$VOL<330)])
7 N=length(X)
8 1-length(X)/length(a$VOL)
9
10 set.seed(666)

```

```

11 X.v <- as.vector(X)
12 fit <- fitdist(X.v, "norm")
13 descdist(X.v, boot = 1000)
14 x <- round(rnorm(n = N, mean = fit$estimate[1],sd=fit$estimate[2]))
15 y <- round(rnorm(n = N, mean = 377 ,sd=7.8))
16
17 plot(density(X,bw=4), main="HFC1340-350_Distribution")
18 #polygon(density(X,bw=4),col="#F1EFF1")
19 lines(density(x,bw=4),col="blue")
20 lines(density(y,bw=4),col="red")
21 rug(jitter(X), side=1, col=1)
22 abline(v=c(330,350,370,390,410),col="gray40",lty=2)
23 abline(v=c(350),col="black")
24 abline(v=c(377),col="black",lty=5)
25
26 plot(density(a$VOL,bw=10), main="HFC1340-350_Actual_Distribution")
27 text(3,0.003," ",col="black")
28 text(335,0.003," ( )",col="black")
29 text(385,0.003," ",col="black")
30 abline(v=c(350),col="black")
31 abline(v=c(320),col="black",lty=2)
32 abline(v=c(35),col="black",lty=2)
33 abline(v=378,col="black",lty=2)
34
35 set.seed(666)
36 x1 <- round(rnorm(n = N, mean = 376.77,sd=9.93))
37 x2 <- round(rnorm(n = N, mean = 376.77 ,sd=5))
38 x3 <- round(rnorm(n = N, mean = 376.77 ,sd=15))
39 x4 <- round(rnorm(n = N, mean = 350 ,sd=9.93))
40 plot(density(x2,bw=10), main="HFC1340-350_Distribution_vs. sigma_&mu",col="blue",xlim=c(300,440))
41 lines(density(x1,bw=10),col="black")
42 lines(density(x3,bw=10),col="red")
43 lines(density(x4,bw=10),col="purple")
44 rug(jitter(x1), side=1, col=1)
45 abline(v=c(376.77),col="black",lty=2)
46 abline(v=c(350),col="black",lty=5)
47 abline(v=c(330),col="black",lty=1)
48 abline(h=0.0286095,col="black",lty=2)
49 text(310,0.005," ( )",col="black")

```

最优容量凸优化模型

```

1 prof <- function(mu0,mubar,c0,ca,cp,d,beta,sigma,p){
2   mu <- mu0-d+sigma*sqrt(-2*log(sqrt(2*pi)*c0*sigma/(mu0*p)))
3   pro <- p*(1-beta-pnorm((mu0-d-mu)/sigma))-(c0/mu0*mu+ca+cp)
4   pro
5 }
6
7 muu <- function(mu0,d,sigma,c0,p){
8   mu0-d+sigma*sqrt(-2*log(sqrt(2*pi)*c0*sigma/(mu0*p)))
9 }
10
11 profit <- function(mu0,mubar,c0,ca,cp,d,beta,sigma,p,name="HFC1340-350"){

```



```

12 mu <- mu0-d+sigma*sqrt(-2*log(sqrt(2*pi)*c0*sigma/(mu0*p)))
13 pro <- p*(1-beta-pnorm((mu0-d-mu)/sigma))-(c0/mu0*mu+ca+cp)
14 pro.bar <- p*(1-beta-pnorm((mu0-d-mubar)/sigma))-(c0/mu0*mubar+ca+cp)
15 pro0 <- p*(1-beta-pnorm((mu0-d-mu0)/sigma))-(c0/mu0*mu0+ca+cp)
16 r <- pro/p
17 r.bar <- pro.bar/p
18 r0 <- pro0/p
19 alpha <- pnorm((mu0-d-mu)/sigma)
20 alpha.bar <- pnorm((mu0-d-mubar)/sigma)
21 alpha0 <- pnorm((mu0-d-mu0)/sigma)
22
23 cat("最优容量", mu, "最优利润",pro, "最优利润率",r,"最优次品率",alpha, "实际容量",mubar,"实际利润",pro.bar,"
    实际利润率",r.bar,"实际次品率",alpha.bar, "标称容量", mu0, "标称利润", pro0, "标称利润率", r0, "标称度
    品率",alpha0)
24
25 curve(p*(1-beta-pnorm((mu0-d-x)/sigma))-(c0/mu0*x+ca+cp),from = mu0-d,to=mu0+5*d ,n = 500, ylab="Profit",
    xlab="mu",ylim=c(pro-0.5,pro+0.2),main=paste(name,"Profit_Curve"))
26 #text(mu+d,pro+0.1,"Profit Curve",col="black")
27 abline(v=mu,col="red",lty=2)
28 abline(h=pro,col="red",lty=2)
29 abline(v=mu0,col="blue",lty=2)
30 abline(h=pro0,col="blue",lty=2)
31 abline(v=mubar,col="black",lty=2)
32 abline(h=pro.bar,col="black",lty=2)
33 }
34
35 profit(mu0=350,mubar=376.77,c0=0.4,ca=0.3,cp=0.3,d=20,beta=0.06,sigma=9.93,p=7.69,name="HFC1340-350")

```

方差敏感度分析

```

1 mu0=350
2 mubar=376.77
3 c0=0.4
4 ca=0.3
5 cp=0.3
6 d=20
7 beta=0.04
8 p=7.69
9 sigma=9.93
10
11 mu <- mu0-d+sigma*sqrt(-2*log(sqrt(2*pi)*c0*sigma/(mu0*p)))
12 pro <- p*(1-beta-pnorm((mu0-d-mu)/sigma))-(c0/mu0*mu+ca+cp)
13 pro.bar <- p*(1-beta-pnorm((mu0-d-mubar)/sigma))-(c0/mu0*mubar+ca+cp)
14 pro0 <- p*(1-beta-pnorm((mu0-d-mu0)/sigma))-(c0/mu0*mu0+ca+cp)
15
16 pro <- prof(mu0=350,mubar=376.77,c0=0.4,ca=0.3,cp=0.3,d=20,beta=0.04,sigma=9.93,p=7.69)
17 pro1 <- prof(mu0=350,mubar=376.77,c0=0.4,ca=0.3,cp=0.3,d=20,beta=0.04,sigma=5,p=7.69)
18 pro2 <- prof(mu0=350,mubar=376.77,c0=0.4,ca=0.3,cp=0.3,d=20,beta=0.04,sigma=15,p=7.69)
19
20 mu <- muu(mu0=350,d=20,sigma=9.93,c0=0.4,p=7.69)
21 mu1 <- muu(mu0=350,d=20,sigma=5,c0=0.4,p=7.69)
22 mu2 <- muu(mu0=350,d=20,sigma=15,c0=0.4,p=7.69)
23

```

```

24 curve(p*(1-beta-pnorm((mu0-d-x)/9.93))-(c0/mu0*x+ca+cp),from = 330 ,to= 420 ,n = 500, ylab="Profit",xlab="
    mu",main="HFC1340-350_Profit_Curve_vs._sigma", ylim=c(6.2,6.4))
25 curve(p*(1-beta-pnorm((mu0-d-x)/5))-(c0/mu0*x+ca+cp),from = 330 ,to= 420 ,n = 500, ylab="Profit",xlab="mu"
    , ylim=c(pro-1,pro+1),add=T,col="red")
26 curve(p*(1-beta-pnorm((mu0-d-x)/15))-(c0/mu0*x+ca+cp),from = 330 ,to= 420 ,n = 500, ylab="Profit",xlab="mu"
    , ylim=c(pro-1,pro+1),add=T,col="blue")
27 abline(v=mu,col="black",lty=2)
28 abline(h=pro,col="black",lty=2)
29 abline(v=mu1,col="red",lty=2)
30 abline(h=pro1,col="red",lty=2)
31 abline(v=mu2,col="blue",lty=2)
32 abline(h=pro2,col="blue",lty=2)
33
34 curve(mu0-d+x*sqrt(-2*log(sqrt(2*pi)*c0*x/(mu0*p)))) , from =1,to=20,n=1000,main="Optimal_Mu_Curve_vs._sigma"
    , ylab="Optimal_Mu",xlab="sigma")
35 abline(v=sigma,col="black",lty=2)
36 abline(h=mu,col="black",lty=2)
37 abline(v=5,col="red",lty=2)
38 abline(h=mu1,col="red",lty=2)
39 abline(v=15,col="blue",lty=2)
40 abline(h=mu2,col="blue",lty=2)
41
42 curve(p*(1-beta-pnorm((mu0-d-(mu0-d+x*sqrt(-2*log(sqrt(2*pi)*c0*x/(mu0*p)))))/x))-(c0/mu0*(mu0-d+x*sqrt(-2
    *log(sqrt(2*pi)*c0*x/(mu0*p))))+ca+cp),from =1,to=20,n=1000,main="Optimal_Profit_Curve_vs._sigma",
    ylab="Optimal_Profit",xlab="sigma")
43 abline(v=sigma,col="black",lty=2)
44 abline(h=pro,col="black",lty=2)
45 abline(v=5,col="red",lty=2)
46 abline(h=pro1,col="red",lty=2)
47 abline(v=15,col="blue",lty=2)
48 abline(h=pro2,col="blue",lty=2)
49
50 library(emdbook)
51 mu0=350
52 mubar=376.77
53 c0=0.4
54 ca=0.3
55 cp=0.3
56 d=20
57 beta=0.04
58 p=7.69
59 sigma=9.93
60 curve3d(p*(1-0.04-pnorm((mu0-d-y)/x))-(c0/mu0*y+ca+cp),from = c(1,250) ,to= c(20,450) ,n = 80,main="
    HFC1340-350_Profit_Curve_vs._sigma_&_mu",sys3d = "wireframe", ylab="mu",xlab="sigma",zlab="Optimal_
    Profit")

```