

## 第十二章 全面质量制造

在电视上看见它  
通过电话买了它  
现在你一个人在家  
它只是废物一件

我试图给它插上电源  
我试图打开它  
当我把它拿回家时  
它只是废物一件

——尼尔·扬

### 12.1 引言

变动性在决定制造系统绩效中扮演重要角色，是工厂物理学的一个基本观点。如我们在第八章和第九章中看到的那样，变动性有一系列来源：机器故障、生产准备、作业员行为、产品组合波动以及许多其他因素。变动性的一个尤为重要并能根本地改变系统绩效的来源，就是质量（Quality）。质量问题几乎总是变成变动性问题。出于同样的原因，变动性削减常常成为质量改善的一种手段。正因为质量和变动性联系密切，我们以这个关键议题的概述来结束第二篇。<sup>1</sup>

#### 12.1.1 质量的十年

二十世纪八十年代在美国是质量的十年。大量关于这个课题的书出版，数以千计的职员通过短期课程或其他培训项目，并且“质量说话（quality-speak）”进入全美的标准语言。1987年，国际标准化组织建立了ISO 9000 系列质量标准。同年，美国国会立法设立了多里奇国家质量奖（Malcolm Baldrige National Quality Award）。<sup>2</sup>（380|381）

质量的概念和质量控制、质量保证及质量管理的方法在二十世纪八十年代并不新奇，质量控制作为一门学科至少可以追溯到 1924 年，西部电气公司的贝尔电话实验室的沃尔特·A·休哈特首次提出过程控制图（Process control chart）。1931 年，休哈特发表了第一篇关于质量的重要文章。在 1956 年的论文中，阿曼德·费根鲍姆（Armand Feigenbaum）发明了全面质量控制的术语并且在 1961 年把它用作他在 1951 年出版的《质量控制》一书的修订本名称。

虽然关于质量的术语和工具已经出现了很长时间，但是直到二十世纪八十年代美国工业界才真正注意到质量的战略潜力。毫无疑问，这种关注很大程度上是受七十年代与八十年代日本产品质量的戏剧性提升的刺激，就像美国对削减库存的兴趣大部分是受日本 JIT 成功故

<sup>1</sup> 我们特意使用全面质量制造代替更加为传统的全面质量管理（TQM）作为标题，在于认识到我们只是涵盖了 TQM 中与运营管理有关的主题。

<sup>2</sup> 能够说明问题的是，日本科学家和工程师联盟早在 1951 年就已经建立了它的主要质量奖，即戴明奖，以纪念美国人 W. 爱德华兹·戴明。

事的推动一样。

是不是所有关于质量的讨论都引起了改善？也许是的，尽管很难去度量，因为像我们将在本章中讨论的那样，质量是一个可以以多种方式解释的广义术语。然而，调查显示消费者认为美国产品的总体质量在八十年代呈下降趋势（Garvin 1988）。美国消费者满意指数（ACSI），从 1994 年开始每季度一次跟踪客户对质量总体感知的标准，同样显示了九十年代客户（对美国产品质量）下降的满意度。不论这些下降是由于不断上升的客户期望，还是由于不断产生的管理问题，或二者兼而有之，很明显的是质量仍然是未来的一个重要挑战。

### 12.1.2 一则质量轶事

为了定下场景，我们从私人角度引入质量话题。1991 年，本书两位作者中的一个购买了一套厨房用具，这套用具却让他发现了一系列令人惊奇的质量问题。最初，出于对样式的考虑，他买了一个有浅色外观、瓷表面、钢炉灶面的壁炉。仅仅使用了几天，瓷破裂并且脱落了，露出粗糙、难看的表面。当这个作者打电话给客服部门（以及一些有类似炉子的朋友），他发现这种型号的每个炉子都有同样的缺陷——100% 的失败率！检验和质量保证不过如此！

客户服务部门十分地礼貌并且送去了更换的炉子，但是这些并没有比最初的那个持续更久，所以这个作者就继续抱怨。更换了三四次之后（其中包括有一次客服部门给了两套设备并建议说可以平时使用一个，而把另一个留起来当招待客人的时候可以用来放炉子），生产商更换了供应商并且送来了深色、更耐用的壁炉。质量设计和样式不过如此！

随着壁炉故事的发展，这个炉子遭受了一系列的其他问题。例如，当火炉点燃之后自动点火功能却不会切断，这样不管炉子是不是在使用，都产生了很大的滴滴答答的噪声。在使用的第一年（即，保修期）里，修理人员不少于八次前来解决这个问题。在其中一次的到访中，修理人员都承认他不知道怎么来调整炉子，因为他从来没有从生产厂商那儿拿到这种炉子的详细说明书，所以他只能更换零部件来期望情况更好一些。售后服务和第一次把事情做对不过如此！（381|382）

第一年末，客服部门给这位作者打电话出售更长期的产品质保，并且竟然说因为这个炉子太不可靠了（他们用了个比不可靠不礼貌很多的词语），长期的质保对我们来说很划算。自己产品的后援和客户驱动的质量不过如此！

（顺便提一下，在这本书在编写之时，炉子烤箱的门也掉了。对此我们没有任何虚构！）

### 12.1.3 质量状况

我们并没有意图暗示这个故事概括了美国制造业的质量水平。但令人迷惑（同时很令人沮丧的）是，二十世纪九十年代的公司会在实质上与每个良好质量管理的原则有如此明显的冲突。而且，我们猜想这并不是一个孤立的例子。（我们还有更多的个人经历，但不会再向读者展示了。）在行政管理课程中，我们喜欢做这样一种练习，即挑战参与者不从制造经理、讲师或专业人员而从消费者的视角来考虑质量问题。高得使人沮丧的比例显示，产品或服务罕有能超出期望，常常都是失望。

显然，质量的现实情况与花言巧语的吹嘘仍有显著的差距。因而，就像我们在书的开始做的那样，成本是二十世纪七十年代竞争的标尺一样，质量是八十年代竞争的标尺，速度是九十年代竞争的标尺，但千万不可以从字面意思来理解这句格言。九十年代之后质量（和成本，这之类的）将仍然是一个重要的、起决定作用的竞争要素。

那么一般企业可以做些什么呢？答案是，很多。世上没有一家公司不能改善其产品、过程或系统，拉近与客户的距离，以及更好地理解质量对其生意的影响。此外，可以参考大量的文献来产生灵感。虽然质量方面的文献，就像 JIT 方面的文献，包含了大量不精确、空泛的花言巧语，它还是提供了很多有用的指导。有关质量的文献可以划分为两类，一类是**全面**

**质量管理 (TQM)**，着重点是质量的定量管理方面（如，培养支持质量改善的整体环境）；另一种是**统计质量控制 (SQC)**，着重点是质量的定量工程方面（如，测度质量与确保遵守规范）。为了制订有效的质量改善计划，这两种观点都是必要的。全是 TQM 而没有 SQC 就会造成只谈理论而没有实质作为，全是 SQC 而没有 TQM 就会只有数字而没有目的。

TQM 著作中的一个的典型代表是 Garvin 的作品（1988），下面的一些讨论就是基于它的。Garvin 的著作在什么是质量以及质量如何影响公司上有深刻的见解。其他被广泛阅读的包括 Crosby（1979、1984）、Deming（1986）和 Juran（1989、1992）的著作。SQC 领域有许多可信的著作，其中大部分都包含对 TQM 的简短介绍；有 Banks（1989），Devor、Chang 和 Sutherland（1992），Gitlow 等人（1989），Montgomery（1991），Thompson 和 Koronacki（1993），以及其他人的著作。一些书，特别是 Juran 的《质量控制手册》（1988），同时包括了 TQM 和 SQC 的观点。

在这个简短的篇章中，我们无法提供这些参考书目的更深更广的东西。我们所能做的是，关注质量如何适合工厂运营管理的全面图景。工厂物理学的框架允许我们将质量和运营的观点综合为同一幅图景中的要素。我们将上面提到的那些文献留给读者去参考，去充实质量管理程序的细节。（382|383）

## 12.2 质量观点

### 12.2.1 总体定义

质量是什么？这个问题的答案是开始我们的讨论的逻辑起点。Garvin（1988）给出了质量的五种定义，我们总结如下：

1. 先验性。质量指“先天的优点”，不是商品或客户的具体特征，而是一个完全的第三种实体。这归结为“我说不清楚，但是看到的它时候我就知道”的质量观点。

2. 基于产品。质量是产品属性的函数（垫子的质量取决于每平方英尺的结点数，汽车保险杠的质量取决于以 5 英里/小时速度撞击的损伤）。这有点像“越多越好（more is better）”的质量观点（更多的结点，更强的防撞性等等）。

3. 基于用户。质量取决于客户偏好得到满足的程度；因此，它是客户所看重的任何一种特征（形状、耐用性、美观性、吸引力等等）的函数。本质上来说，这就是“美丽在旁观者的眼中（beauty is in the eye of the beholder）”这一质量观点。

4. 基于制造。质量等同于与规范的一致性（例如，在尺寸公差范围内，或者达到规定的性能标准）。因为这种关于质量的定义直接涉及制造产品的过程，它与“第一次就做好”的质量观点密切相关。

5. 基于价值。质量连带地取决于性能或产品价格的一致性（如，无论性能如何，1,000 美元的 CD 质量不好，因为很少有人会觉得它值这个价钱），这就是“让你的钱花得值”或“负担得起的高性能”的质量观点。

这些定义引出了两个要点。首先，质量是一个多重概念，不能简化为简单的数字量度。我们需要用一个框架来评价质量方针，就像需要一个框架（即，工厂物理学）来评价运营方针。确实，正如我们将要讨论的那样，这两个框架是密切相关的，或许就像比我们在第六章中所提到的更大的制造科学的两面。

其次，这些定义很强调**面向产品（product-oriented）**。这就是大部分 TQM 著作提到的情况，也是质量必须最终由“客户驱动”这一原则的作用。因为客户所看到的是产品，质量

必须以产品的形式测度。无论如何，客户所看到的产品质量最终由一系列**面向过程（process-oriented）**的因素决定，这些因素包括产品设计、制造运营控制、人和管理者在检查过程中的参与、客户售后服务等等。

### 12.2.2 内部质量 vs. 外部质量

为了更好地理解面向产品和面向过程的质量的关系，我们发现在下面提出内部质量与外部质量的这些区别是很有用的：（383|384）

1. **内部质量（Internal quality）**指的是工厂内部产品符合质量规格的程度，并且与基于制造的质量定义紧密相联。很典型的是，它是通过对诸如废品率和重工率的直接产品测量和诸如压力（在注模机器上）和温度（在电镀槽中）的间接过程测量来进行监测的。

2. **外部质量（External quality）**指的是客户是如何看待产品的，并且通过使用先验的，基于产品的，基于用户的或基于价值的或者是它们之间的综合定义来解读。它可以通过诸如退回率这样的客户满意度指标进行直接测量，也可以通过由取样，检查，外勤人员资料，客户调查等派生出来的间接客户满意度指标来测量。

为了取得高的外部质量，必须将客户关注点转化成内部质量的控制和测量。因此，从制造经营的角度来讲，内部质量与外部质量之间的联系对于一个从战略上有效的质量计划的发现是至关重要的。下面的是一些工厂内部的质量是如何与产生客户满意度的外部质量相互联系在一起的更加重要方法。

1. **差错预防（error prevention）**。如果工厂内部发生更少的错误，那么就会有更少的有缺陷品躲开检测过程而到达客户。因此，从客户感知到的质量是由免于缺陷品的侵扰来决定的程度来讲，工厂内部的高“源头质量”会产生很高的客户驱动质量。

2. **检测改进（inspection improvement）**。如果制造过程中有更少的缺陷品制造出来，那么质量保证就会要求检测过程探测和拒绝或者改正更少的物品。这样就会倾向于减少“让缺陷品通过”而产生在质检人员身上的压力……换句话说，以将产品送出门的名义来放松质量标准。<sup>3</sup>此外，花在重工或更换有缺陷的部件上面的时间越少，那么人们追查质量问题的根源的时间就更多了。理想情况下，净效果将是一种向上的质量螺旋进步，在其中，差错预防和差错探测都随着时间而改进。

3. **氛围增进（environment enhancement）**。即使这个领域的质量问题不能直接追溯到厂级缺陷，高内部质量和外部质量仍然可以被联系起来。<sup>4</sup>这两种形式的质量都是由相同的环境因素所推进（如，支持性的管理态度、改进的有形回报、熟练的追踪和控制系统以及有效的培训）。一个已经在工厂内部培育起正确的态度和工具的组织很有可能在工厂外部也会做同样的事情。

简而言之，理解质量就意味着注意客户。遵守质量就必须注意制造。<sup>5</sup>为了本章的目的，我们假设客户的关注点已经得到了工厂的理解并被转换成方便使用的质量规格说明。我们的关注点将会是质量和运营之间的关系，尤其是两者作为工厂的持续改善过程是如何一起工作

---

<sup>3</sup> Crosby（1979，41）提到一个故事，其中制造过程以一种敌对的态度来对待检测过程，抗议所有被拒绝的部件，好像质检人员都是本身试图对工厂进行破坏活动。

<sup>4</sup> Garvin（1988，129）提供了一个例子，在那里，一个空调机的压缩机由于过多的湿气渗透进其中造成的侵蚀形成了故障。这样的问题在任何合理的“老化”阶段都不会出现，因此在工厂层级很难会作为一个缺陷品被检测到。

<sup>5</sup> 这里，我们提到的是“大 M”制造，包括产品设计、生产和外勤服务。

的。(384|385)

## 12.3 统计质量控制

统计质量控制 (SQC) 大致集中在制造质量上面, 由与规格说明的符合度来测量。SQC 的最终目标就是在关键的质量量度中变动性系统性地减少。例如, 尺寸、重量、光滑性、强度、颜色以及速度 (如, 递送方面的速度) 方面都是可测量的属性, 并可用于表现制造过程质量的特色。通过保证这些测量标准在期望的界限内得到紧密控制的工作, SQC 直接在运营和质量之间的连接处发挥作用。

### 12.3.1 SQC 方法

在 SQC 中用于保证质量所使用的三个主要类别的工具如下:

1. 抽样检验 (*acceptance sampling*)。通过对产品的检测来决定他们是否符合质量规格。在有些情况下, 也会使用 100% 检测, 而有其它一些情况下, 某种形式的统计抽样就用于代替前者。抽样可能是出于成本或者绝对必要性 (如, 当检测有破坏性时) 考虑的选择
2. 过程控制 (*process control*)。关于过程的均值和变动性不断受到监控, 用以决定特别的问题发生的时间或者什么时候过程脱离了控制。
3. 实验设计 (*design of experiments*)。质量问题原因都是通过特定的目标实验来得以追踪。基本方法是, 通过对可控变量进行系统性变化, 来测定它们在质量测量上的效果。已经开始出来许多统计工具 (如, 区组设计、析因设计、嵌套设计、响应面法和田口方法 (*block designs, factorial designs, nested designs, surface analysis, and Taguchi methods*)) 用来将控制器与产出高效地联系起来并优化过程。

典型的情况是, 随着一个组织的成熟, 它会较少依赖于事后抽样, 并较多地依赖于源头过程控制以及持续改善导向的实验设计。

很显然地, 这些主题都有整本书进行阐述, 所以对于它们的详细介绍就不是本章的范围所能及的了。尽管如此, 由于过程控制如此特定地加工质量与变动性间的界面, 我们在这里提供一些基本概念的总结

### 12.3.2 统计过程控制

统计过程控制 (SPC) 从一个可测量的质量属性开始——例如, 铸钢部件上的一个小洞的直径。不管铸造过程控制得多么严密, 这个直径始终会有一定量的变动性。如果相对来讲它很小, 并且本质上是源于不可控的源头, 那么我们称之为 **自然变动性 (natural variability)**。一个在它的自然变化范围内稳定地运行的过程被认为是 **处于统计性控制中 (in statistical control)**。那些可能被追踪到原因的有更多源头的变动性被称为 **原因-可分配变化 (assignable-cause variation)**。一个从属于原因-可分配变化的过程被称为 **失控 (out of control)**。SPC 的根本挑战在于, 将原因-可分配变化从自然变化中区别出来。由于我们大体上直接观测到的只是质量属性本身, 而不是变化的原因, 我们需要统计学来完成这一点。(395|386)

为了阐述 SPC 背后的基本原理, 这里让我们考虑使用砂铸造工艺控制一个钢件上的小洞的直径的例子。假定期望的名义直径是 10 毫米, 而我们观测到的的一个铸件的直径为 10.1 毫米。我们能因此认为这个铸造过程是失控的么? 答案当然是 “看情况而定”。可能 0.1 毫

米的偏差正好是自然变化水平范围之内。如果情况是这样，而我们还要去调整这个过程（如，通过改变磨砂、钢以及铸型），以尽力改正这个偏差，可能的情况是我们会使情况更糟。理由是，因为随机噪音的增加而调整一个过程会增加变动性（参见 Deming 1982, 327 页，里面有揭示这一点的漏斗实验的讨论）。因此，为了保障这些调整只是依据原因可分配变化进行的，我们必须特征化自然变异。

在我们的例子中，假如我们已经测量了一些铸件，并且决定平均直径可以控制到  $\mu = 10$  毫米的水平，并且直径的标准差是  $\sigma = 0.025$  毫米的水平。再假设每两个小时我们抽取一个五铸件的随机样本，测量它们洞径，计算均值（我们称之为  $\bar{x}$ ），并将它绘成如图 12.1 所示的图表。从基础统计学中可以知道， $\bar{x}$  本身是一个随机变量，并且也有标准差

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad (12.1)$$

这里  $n$  表示样本的数量；在这个例子中  $n = 5$ 。<sup>6</sup>

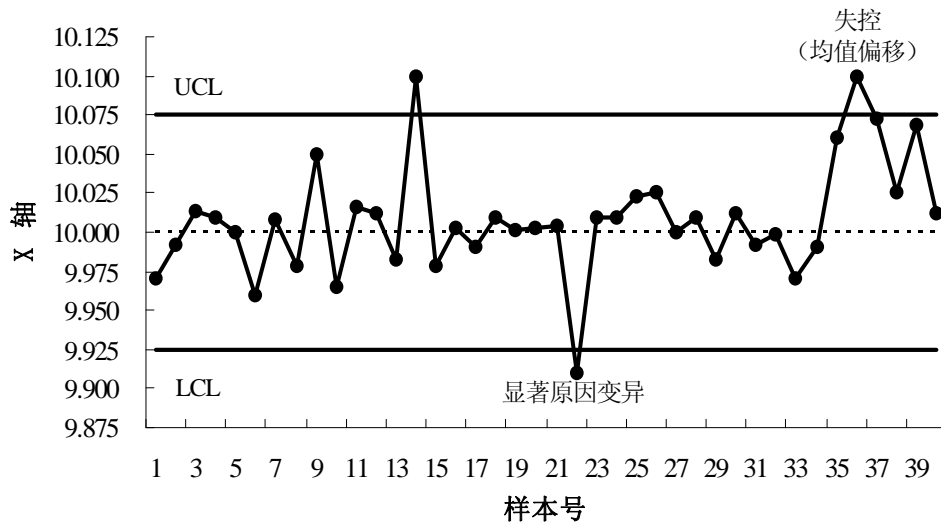


图 12.1 铸钢部件平均孔径的过程控制图

控制图后面的基本思想与假设检验非常相似。我们的零假设是，过程处于控制中；那就是说，样本都是来自于一个均值为  $\mu$ ，标准差为  $\sigma$  的制程。为了避免在制程并没失控时（即，第 I 类错误）做出失控的结论，我们设置了一个严格的标准用于指定标准差作为“非随机变异。”标准的传统做法是将那些位于均值上下三个标准差的点标记出来。我们操作的方式是通过详细说明上控制限与下控制限如下：(386|387)

$$LCL = \mu - 3\sigma_{\bar{x}} \quad (12.2)$$

$$UCL = \mu + 3\sigma_{\bar{x}} \quad (12.3)$$

如果我们观察到一个样本均值落在 LCL 和 UCL 之间的范围外，那么这个观测值就被指定成非随机变异。在图 12.1 绘制的铸件例子中，样本 22 就是这样的一个偏差。这个可能是由于有缺陷的输入（如，钢或磨砂）、机器故障（如，铸型中、包装过程中、倾倒过程中），或者

<sup>6</sup> 这里注意，这只是变异性调查的另一个例子。选择  $n > 1$  使得我们对于  $\bar{x}$  的假设更严，因此减少了针对这个系统中的随机噪音作出反应的机会。

操作者的错误造成的。SPC 并不会告诉我们偏差发生的原因——它只告诉我们，偏差足够异常，值得进行进一步的调查。

除了超出控制线的点之外，也有一些其它的标准用于表示失控条件。例如，如果发生好几个高于目标均值的点排成一排的情况，那么目标均值就经常用于确定过程均值的可能转变。在图 12.1 中，样本 37 就是超出了控制。但是不像样本 22 所在的失控点，这个点与样本 35~40 在一起呈现出高于均值观测值的趋势排列。这很强烈地证明了，问题的原因并不单单指向样本 37，而是由于铸件过程本身中的某些事情导致了平均直径的增加。其它基于多样本的标准，例如寻找趋势的一些法则（如，高值后面跟着低值，然后又是高值），也与控制图一起使用用于探测出非随机变异。

记录下来这些东西很重要，因为一个过程处于统计控制中，并有必然意味着它就是**合格的（capable）**（即，能够满足过程规格说明和规则性）。例如，如果在我们的铸件例子中，由于功能性要求的原因，我们要求洞径要在**较低规格水平（lower specification level, LSL）**和**较高规格水平（upper specification level, USL）**之间。流和是否能够取得达到这些水准取决于它们如何与**高低自然容忍极限（lower and upper natural tolerance limits）**比较，后者被定义为：

$$LNTL = \mu - 3\sigma \quad (12.4)$$

$$UNTL = \mu + 3\sigma \quad (12.5)$$

注意 LNTL 和 UNTL 都是对于单个洞径的限制，而 LCL 和 UCL 都是对于平均样本直径的限制。另外，注意 LNTL 和 UNTL 都由过程自身内在决定，而 LSL 和 USL 都由性能要求外在决定。

让我们考虑一些启示性的例子吧。自然容忍极限给定为  $LNTL = \mu - 3\sigma = 10 - 3(0.025) = 9.925$ ， $UNTL = \mu + 3\sigma = 10 + 3(0.025) = 10.075$ 。假如规格水平给定为  $LSL = LSL1 = 9.975$ ，而  $SUL = SUL1 = 10.025$ 。从图 12.2 显然可以得出这样的结论，那就是铸件过程将会产生很大一部分不符合要求的部件。精确地讲，如果洞径是正态分布的，那么

$$\begin{aligned} P(9.975 \leq X \leq 10.025) &= P\left(\frac{9.975 - 10}{0.025} \leq Z \leq \frac{10.025 - 10}{0.025}\right) \\ &= P(-1 \leq Z \leq 1) = \Phi(1) - \Phi(-1) \\ &= 0.8413 - 0.1587 \\ &= 0.6826 \end{aligned}$$

这意味着有近 32% 的样本不能满足规格水平。

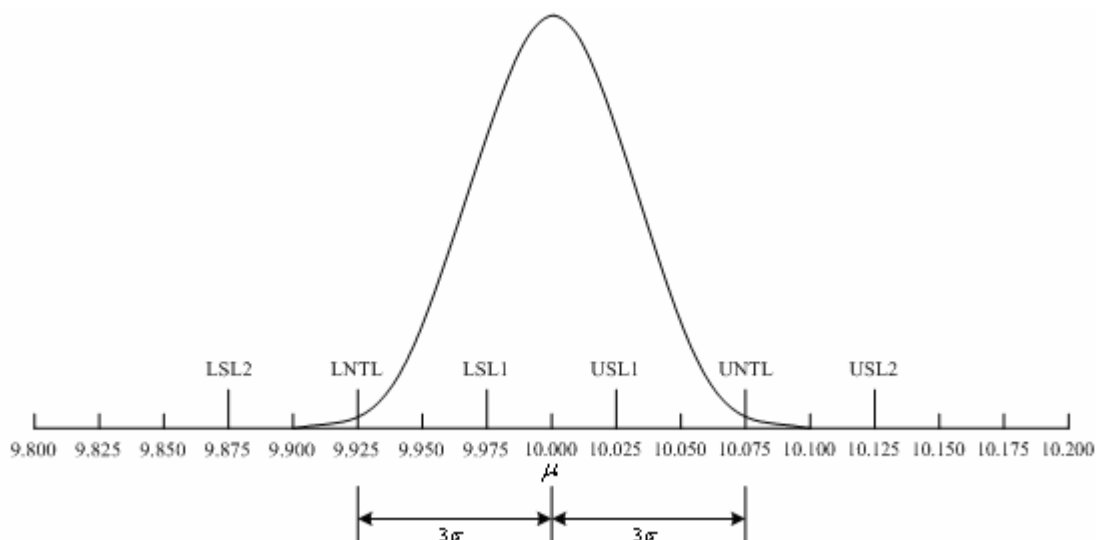


图 12.2 过程能力与自然容忍极限

假如规格水平给定为  $LSL = LSL2 = 9.875$ ,  $USL = USL2 = 10.125$ 。由于自然容忍极限正好在这个范围之内，我们可以期望很少的不合格铸件。实际上，为这些极限重要上面的计算，结果显示，不合格部件的比例只有 0.0000057。（387|388）

一个关于能力的量度就是**过程能力指数（process capability index）**，它被定义成

$$C_{pk} = \frac{Z_{\min}}{3} \quad (12.6)$$

其中 
$$Z_{\min} = \min\{-Z_{LSL}, Z_{USL}\} \quad (12.7)$$

同时 
$$Z_{LSL} = \frac{LSL - \mu}{\sigma} \quad (12.8)$$

$$Z_{USL} = \frac{USL - \mu}{\sigma} \quad (12.9)$$

$C_{pk}$  最低可接受的值一般认为是 1。注意在上面的例子中，对于  $(LSL1, USL1)$  而言， $C_{pk} = 1/3$ ，对于  $(LSL2, USL2)$  而言， $C_{pk} = 5/3$ 。注意， $C_{pk}$  对于变动性 ( $\sigma$ ) 和非对称性（即，并不位于  $USL$  和  $LSL$  中间的过程均值）都是敏感的。因此，它给了我们一个测定过程满足它的性能规格说明的能力的简单定量标准。

当然，为了实施一项有效的 SPC 图，有一些细节是需要说明的。我们已经对  $\mu$  和  $\sigma$  的估计值进行了注释；在实际中，有许多方法可以从观察到的数据中收集这些。我们同样需要选择足够大的样本规模  $n$  来防止针对随机变动做出反应，但是又不能太大而使得它掩盖了非随机变异。同时，我们抽样的频率也应该与抽样的成本及监察的敏感性相区配。

### 12.3.3 SPC 的扩展

刚讨论过的  $\bar{x}$  图只是 SPC 图表的一种。已经有许多变化的版本被提出来用于满足大量的质量保证情境的需要。在制造管理中特别有用的一些包括如下：（388|389）



1. 全距 (*Range*) ( $R$  图)。 $\bar{x}$  图要求过程变动性处于控制之中, 以便控制极限是有效的。因此, 通过绘出样本的全距来监测这种变动性就非常普遍。如果  $x_1, x_2, \dots, x_n$  是一个样本量为  $n$  的样本的测量值, 那么它的全距就是最大观测值与最小观测值之差

$$R = x_{\max} - x_{\min} \quad (12.10)$$

每一个样本都会产生一个全距, 而全距可以描绘在图表上面。使用过去的的数据来估计  $R$  的均值和标准差, 标记为  $\bar{R}$  和  $\sigma_R$ , 我们可以为  $R$  图表设置控制限如下

$$LCL = \bar{R} - 3\sigma_R \quad (12.11)$$

$$UCL = \bar{R} + 3\sigma_R \quad (12.12)$$

如果  $R$  图没有显示出失控的情境, 那么就意味着过程中的变动性足够稳定以应用  $\bar{x}$  图表。通常情况下,  $\bar{x}$  和  $R$  图表都被同时追踪用以观测均值的改变或进行中的过程的变化。

2. 部分不合格 (*fraction nonconforming*) ( $p$  图)。就像我们在一幅  $\bar{x}$  图表中所做的一样, 另一种绘制物理测量指标的方法是追踪那些在周期取样中不满足质量标准的物件部分。注意这些标准可能是定量的 (如, 洞径要在指定的界限内), 也可能是定量的 (如, 酒是由品尝者证实的)。如果每个物件都有独立  $p$  的缺陷可能性, 那么在一个取样规模为  $n$  的样本中, 不合格部分物件的标准差就是  $p(1-p)/n$ 。因此, 如果我们通过过去的的数据来对不合格物件的比例进行估计, 我们可以将  $p$  图表的控制限表达如下:

$$LCL = p - 3\sqrt{\frac{p(1-p)}{n}} \quad (12.13)$$

$$UCL = p + 3\sqrt{\frac{p(1-p)}{n}} \quad (12.14)$$

3. 非质量应用 (*nonquality applications*)。基本的控制图表程序可以用于追踪任何受变动性影响的过程。例如, 我们在第十四章中为统计性产出控制描绘了一种程序, 这种程序监测某个过程的产出以便于决定这个过程是否处在能够达到特定生产额度的轨道上面。控制图表的另一个非质量应用主要是用于到期日报告, 这一点我们在第十五章会讨论到。基本思想就是, 对估计的周期时间附加一个安全提前期, 然后追踪客户服务水平 (如, 按时送货的比例)。如果系统失控, 那么这就要调整安全提前期的信号。

控制图表的能力和弹性使得它们在监测各种存在变动性的过程时特别有用。因此, 就像我们在这本书中反复强调的, 基本上所有的制造过程都涉及变动性, SPC 技巧是现代的制造经理工具箱中的基础部分。

## 12.4 质量和运营

与变动性紧密相联, 作为质量与运营之间的联系的是成本。尽管如此, 关于这种联系是如何发挥作用的, 有许多争议。这里有两种明显不同的观点: (389|390)

1. 成本随着质量而增加。这是传统的工业工程观点，它认为为了取得更高的外部质量，需要更高强度的检查、更多的不合格品以及更加昂贵的材料和过程。既然客户们为额外的质量付款的意愿随着质量水平的提高而降低，这个观点引起了“最优缺陷水平”的论点，而这一点对于过去的工业工程教科书来讲是相当普遍的。

2. 成本随着质量而减少。这是更近的 TQM 观点，并使用诸如质量是免费的（Crosby 1979）或隐藏工厂这样的短语来支持；这种观点认为从一开始就把事情做对下节省出来的材料和人力会超过抵消的质量改善成本。这种观点支持零缺陷以及 JIT 的持续改善目标。

两者观点都不是绝对正确。如果提高特定产品的质量就意味着用金制部件来代替铜制部件，那么成本确实会随着质量而增加。当事实确实是这样的时，问一问市场是否愿意为这种改善付款，甚至会不会注意到这种改善有意义。另一方面，如果质量改善是事实上是将检查的责任从产线终端测度转移到单个机器操作员，那么减少的重工、废料和检查成本超过抵消的执行成本是完全有可能的。最后，重要的是，这两种观点中，哪一种在评价成本和特定质量改善的后果方面更加恰当。这在进行持续改善时决定采取哪种政策是至关重要的，以及哪一种能够较好地与市场相协调。

在下面的讨论和例子中，我们将依靠工厂物理框架来评估质量对运营产生的影响，以及运营对质量的影响。我们提供质量成本的特定数值估计的意图不是很强——工业中产生的情境范围是如此的变化多端以致对于这种特性的全面加工几乎是不可能的——我们的意图主要是放宽并扩展我们在第二篇中针对制造系统的行为培养起来的直觉，以将质量考虑包括进去。

#### 12.4.1 质量支持运营

在第九章中，我们展示了对于理解质量对工厂运营所产生的影响至关重要的两条制造法则，也就是变动性定律和利用率定律。这些可以解释成如下：

1. 变动性引起拥堵。
2. 拥堵会随利用率的上升而非线性地上升。

在实际中，质量问题是变动性许多最大和最普遍的原因之一。另外，通过操作的重要进行（要么是重工要么是报废零件的置换），质量问题通常会导致工站利用情况的增加。通过影响变动性和产量，质量问题会产生极端的运营后果。

**在单一机器上重工的效果。**为了体会一下质量是如何影响利用率和变动性，让我们考虑一个简单的单机例子。机器以三分钟/件的速率接收部件，加工时间的均值和标准差分别是  $t_0$  和  $\sigma_0$  分钟，所以自然加工时间的 CV 就是  $c_0 = \sigma_0 / t_0$ 。尽管如此，给定部分的部件以概率  $p$  形成缺陷品。我们假定质量检测是加工过程的一部分，因此部件是否有缺陷会在它加工完后立刻知晓。如果有缺陷，那么它必须重工，而这另外又需要一个均值为  $t_0$  和标准差为  $\sigma_0$  的加工时间并且同时又有失败而不能生产出好部件的概率  $p$ 。这台机器继续对这个部件进行重工，直到一个好的部件得以生产出来。我们定义生产出一个完好部件所花费的时间是**有效加工时间（effective process time）**。（390|391）

令  $T_e$  代表一个部件的（随机）有效加工时间，我们可以计算时间的均值  $t_e$ 、方差  $\sigma_e^2$ 、

和变异系数的平方（SCV） $c_e^2$ ，以及机器的利用率  $u$  如下：

$$t_e = E[T_e] = \frac{t_0}{1-p} \quad (12.15)$$

$$\sigma_e^2 = Var(T_e) = \frac{\sigma_0^2}{1-p} + \frac{pt_0^2}{(1-p)^2} \quad (12.16)$$

$$c_e^2 = \frac{\sigma_e^2}{t_e^2} = \frac{(1-p)\sigma_0^2 + pt_0^2}{t_0^2} = c_0^2 + p(1-c_0^2) \quad (12.17)$$

$$u = \frac{1}{3}t_e = \frac{t_0}{3(1-p)} \quad (12.18)$$

从这个例子中我们可以得到以下一些结论：

1. 利用率会随重工率非线性上升。这会发生是因为，加工一项加工任务的平均时间随着期望的传递数的增加而增加，而新任务的到达率保持不变。在某一点，由于重工而产生的额外的工作负荷会使工站过载。在这个例子中，式 12.18 显示，当  $p > 1-t_0/3$ ，利用率会超过 1，显示系统此时没有足够的产能来在在长期内跟上新的到来任务和重工任务。

2. 由  $\sigma_e^2$  给定的加工时间方差，会随着重工率的上升而上升。理由当然是，一项任务要做多次通过这台机器的可能性越大，完工时间就变得越发不可预测。

3. 由 SCV 测度的加工时间变动性，可能会随重工率而上升或下降，它取决于加工的自然变动性。尽管有效加工时间的方差和均值总是随着重工率的增加而增加，方差并不总是会比均值增加得更快。因此 SCV，也就是方差对均值的比例，会增加或减少。我们可以从(12.17)式中看到，如果  $c_0^2 < 1$ ，那么  $c_e^2$  会随着  $p$  的增加而增加，如果  $c_0^2 > 1$ ，那么  $c_e^2$  会随着  $p$  的增加而减少，如果  $c_0^2 = 1$ ，那么它就是常量。这一点背后的直觉就是，变动性汇聚的效果会变得足够大至  $c_0^2 > 1$ ，而使得有效加工时间的 SCV 会随着  $p$  的增加而减少。

我们可以由有重工的单台机器的这些特定结果得出。重工对过程的周期时间和提前期产生的影响的一些大概观测结果。既然有效加工时间的均值和方差都会随着重工率的增加而增加，我们可以使用第九章中的提前期定律来得出这样的结论：为了取得既定服务水平而需要的提前期会随着重工率的增加而上升。

尽管如此，重工对周期时间产生的影响不是那么明显。重工增加时有效加工时间的 SCV 将降低的这个事实，可能会产生重工可能降低周期时间的印象。但是实际上情况并不是这样的。理由是，增加的重工会增加利用率，而这是其对周期时间的一阶（first-order）效应，而这种效应超过了由可能的变动性减少所产生的二阶（second-order）效应。因此，即使是在那些有高度自然变动性的过程中，增加重工会变大平均周期时间。另外，由于它也会增加单位作业总加工时间的方差以及排队等待时间的方差，增加的重工同样会增大周期时间的标

准差。(391|392) 这些周期时间效应代表了重工的一般影响，总结如下面的制造定律。

**定律（重工）：** 对于一个给定的产出水平，重工会提高过程周期时间的均值和标准差。

为了阐述这个定律，假设先前提到的工站是由一个到达时间有一定变化的过程所伺候的，但是它有确定的加工时间，例如  $t_0 = 1$ ,  $c_0 = 0$ 。那么，使第八章中所介绍的 Kingman 的单一工站模型，工站的周期可以表示成  $p$  的函数

$$CT = \frac{c_a^2 + c_e^2}{2} \frac{u}{1-u} t_e + t_e$$

$$= \frac{1+p}{2} \frac{1/(3(1-p))}{1-1/(3(1-p))} \frac{1}{1-p} + \frac{1}{1-p}$$

图 12.3 画出了周期时间对重工率的变化关系。这幅图显示出周期时间会随着  $p$  接近  $2/3$  而呈现非线性增长而趋于无穷，而  $2/3$  这一点就是重工率使系统的有效产能减低并小于到达率的程度。

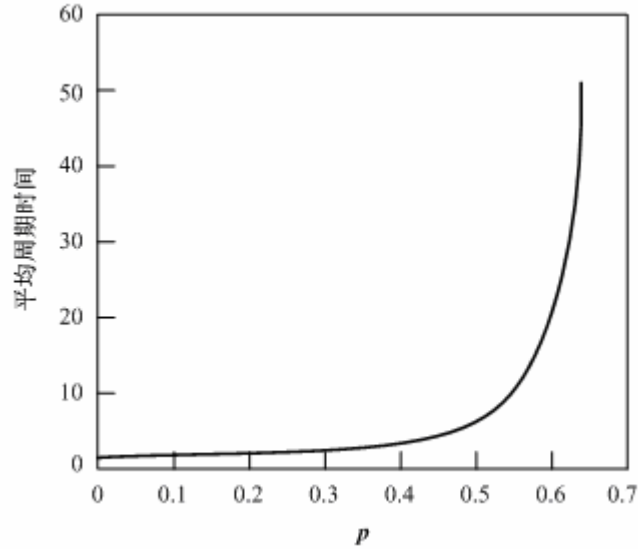


图 12.3 周期时间作为重工率的函数

**重工率对 CONWIP 产线的影响。** 当然，诸如利用率、变动性、周期时间等工站水平量度都只是间接量度；我们真正关注的是产出、WIP 以及产线的周期时间。为了揭示产线上的重工定律，考虑图 12.4 所描述的常量 WIP 产线。对于机器 1、2 和 4 而言，加工时间是  $2/3$  小时，而对于机器 3，加工时间是 1 小时（瓶颈）。所有的加工时间都是确定的（即， $c_e^2 = 0$ ）。尽管如此，机器 2 会受到重工的影响。就像在前一个例子中一样，我们假设每项要加工的作业必然会以概率  $p$  重新加工。因此，就像在前一个例子中一样，机器 2 上面的平均有效加工时间就是

$$t_e(2) = \frac{2/3}{1-p}$$

我们假设这条产线有无限的原材料供应，因此变动性的唯一来源就是重工。（392|393）

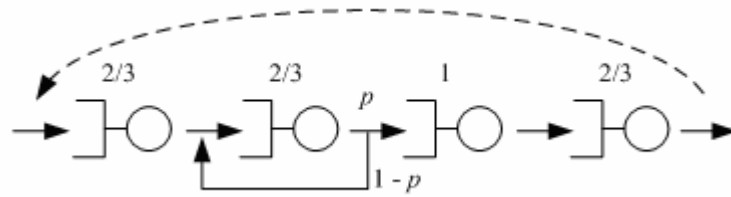


图 12.4 有重工的 CONWIP 产线

因为即使是这种简单的产线也太过于复杂而使得方便的分析是很难的（单台机器的例子就已经足够混乱的了！），我们寻求计算机仿真来估计不同的  $p$  值和不同的 WIP 水平时的性能测量结果。图 12.5 和 12.6 总结了我们的仿真结果

当  $p = 0$  时（没有重工情况），系统就会像我们在第七章中所研究的最佳情形那样。因此，我们可以应用那里派生出来的公式来刻画产出-WIP 以及周期时间-WIP 的曲线。注意没有重工，那么瓶颈速率  $r_b$  就是每小时一件作业，而原始加工时间  $T_0$  就是  $r_b T_0 = 3$  小时。因此，临界 WIP 水平就是 3 件。在这个 WIP 水平下，可以达到最大的产出（每小时一件）和最短的周期时间。

当  $p = 1/3$  时，机器 2 的平均有效加工时间就是  $t_e(2) = 1$ ，也就是瓶颈速率。因此， $r_b$  不是不变的，但是  $T_0$  会增加到 3.33 小时。这意味着 WIP 接近于无穷，可以达到每小时一件作业的满产。我们的仿真显示，大致上满产会在 WIP 水平大约为 10 件时达到——较没有重工情况时所需制品水平超出 3 倍。在 WIP 为 10 件时，平均周期时间大致是 10 小时——同样是无重工的理想水平时的 3 倍。这里的暗示就是，当  $p = 1/3$  时，重工的基本影响就是将一个表现得最好的产线转变成接近实际最差情形。这一点揭示了重工定律对平均周期时间产生的影响。（393|394）

当  $p = 1/2$  时，机器 2 上的平均有效加工时间是  $t_e(2) = 4/3$ ，这使它成为瓶颈。因此，即使 WIP 无穷多，我们也不可能达到高于  $r_b = 3/4$  每小时的产出。就如所期望的一样，图 12.5 显示出，在所有的 WIP 水平下大量减少的产出。图 12.6 显示出，在所有的 WIP 水平下，由于机器 2 产能的降低，周期时间越来越长。另外，由于瓶颈速率已经降低了，周期时间曲线随着 WIP 的增加以一个较前两中情形下更快的速度增长。

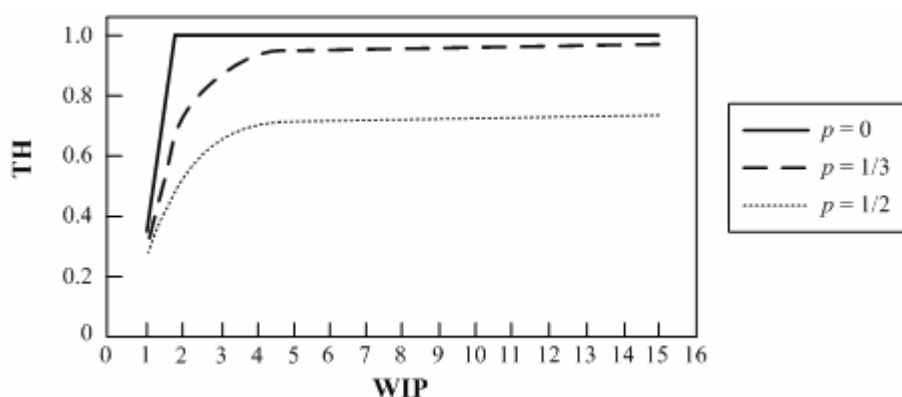


图 12.5 不同重工率的产出-WIP 曲线

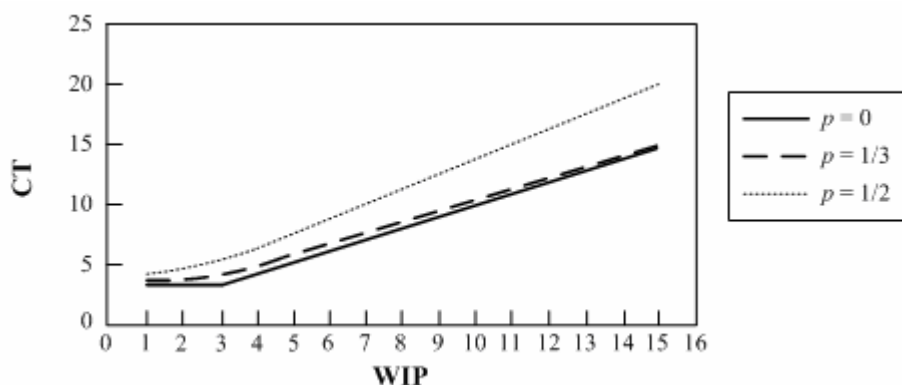


图 12.6 不同重工率的周期时间-WIP 曲线

这个仿真模型使我们能够追踪其它产线统计特征。其中特别有趣的是周期时间的标准差。回忆一下提前期定律所暗含的规律，如果我们向客户报出达到特定服务水平（准时递货的概率）的提前期，那么提前期就是平均周期时间和周期时间的标准差的递增函数。较大的周期时间标准差意味着我们必须报出更长的提前期，因此在产成品库存中必须持更长期地以制成品库存的形式持有物件，用于补偿变化的生产速率。就像图 12.7 所显示的一样，周期时间的标准差随着重工率而增加。另外，随着 WIP 水平的增加，它也会增加。就像我们所提到的，既然重工率要求产线中增加额外的 WIP 以取得既定的产出水平，那么这个效应就会趋于进一步恶化周期时间变动性问题。这是重工率在周期时间方差方面考虑时的情况。

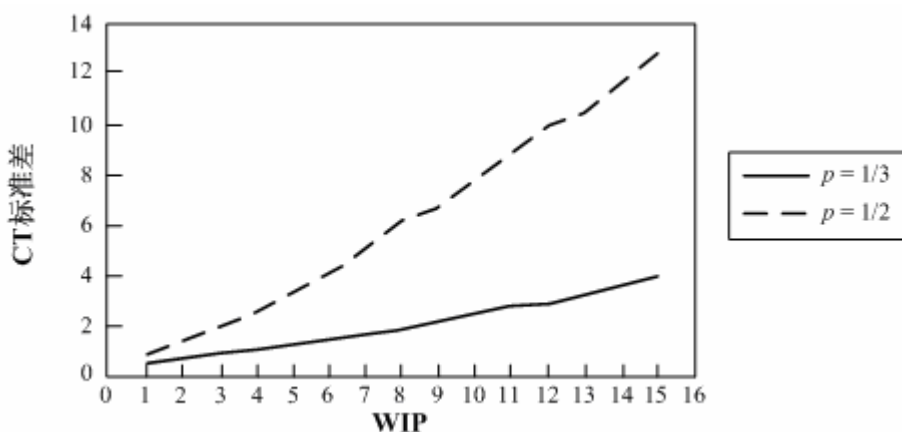


图 12.7 不同重工率的周期时间标准差-WIP 曲线

图 12.5、12.6 和 12.7 的结果暗示出下面的一些质量问题对运营和成本产生的影响。

1. 产出效应。如果重工率高至使一项资源成为瓶颈的程度（或者，更加糟糕，重工情况就在瓶颈资源处），那么它就会大大地改变产线的产能。当这种情况发生时，质量改进将便于产出的提升。从这样一项改善中获得的收入可以大大超出提高产线的质量所花费的成本。

2. WIP 效应。在一项瓶颈资源处的重工，即使是在那些有很多剩余产能的资源处，都会增加产线中的变动性，因此要求更高的 WIP（和周期时间）以达到既定的产出。因此，重工的减少将便于 WIP 的减少。尽管从这样一项改变中节省出来的成本不大可能多到从产能增加中提高的收入这样的水平，但是与取得这种改进的成本相比，它们就显得尤为可观了。（394|395）

3. 提前期效应。通过降低产能、提高变动性，重工问题使得需要在产线中增加额外的 WIP，因此导致更长的平均周期时间。这些问题也会增加周期时间的变动性并因此导致更长的提前期报价或对客户更差的服务。通过重工的减少达到较短提前期和更可靠的递送，并由此产生的竞争优势是很难精确量化的，但是却有着极其重要的战略意义。

**更进一步的观察。**我们以质量问题对运营产生的影响来结束我们的讨论，并加上一些超出先前例子的一些观察。

为了开始，我们注意到重工环越长，后果越显著。在上面的两个例子中，我们将重工表示成穿过单台机器一秒钟（即，缺陷品生产出来后能够立刻被发现——译者注）。在实际中，重工频繁地涉及更多环节。一个有缺陷的部件为了得到更正可能不得不通过产线中好几个工站的回送。当情况是这样时，重工会影响好几个工站的产能和有效加工时间的变动性。另外，因为每次穿过重工环时会较单台机器重工环情况花费更多时间，由此在周期时间的标准差上产生的效应倾向于更大。因此，重工定律的后果随着重工环的增长会变得更加显著。

因为重工会对产线产生如此的破坏性影响，制造经理经常受此影响而设置单独的重工产线。这种方法确实会防止缺陷部件侵蚀产能及增大主产线上的变动性。但是，通过在其它地方配置额外的产能来这么做，会产生成本，占据空间，并且对于消除由重工引起的周期时间均值和标准差的增加作用很小。更加糟糕的是，这种方法会用于将质量问题扫除到大家看不到的地方。将缺陷品分流到一条单独的产线上会使得这些缺陷品成为别人的责任。让一条产线负责改正自己的问题可以培养对质量问题的根源和影响的更深刻认识。如果这种认识能够引起对问题的更快察觉，那么它就可以缩短重工环并减轻后果的严重程度。如果这样可以带来避免第一现场产生缺陷品的方法，那么就达到了真正的大改进。因此，尽管设置单独的重工线会产生短期好处，但是，长期而言，避免缺陷品以及寻找更加基础的质量改善会更好。

在许多制造环境中，内部质量问题会产生废品——即，产生损失——而不是重工，可能是由于缺陷品改正不过来了，或者这样做不经济。因此，指出废品与重工有类似效应这一点非常重要。从运营角度来讲，废弃部件实质上与那些必须从产线的开端重新加工的重工部件是一样的。从这个意义上来讲，废品是重工件的极端形式，因此与我们观察重工的情况有类似的效应，只是程度更深而已。

尽管如此，废品与重工的一个区别在于用于补偿的方法。虽然单一产线可以用于重工件，但是他们不能作为对废品的补救。大多数制造系统会使用某种形式的作业规模增大作为对收益损失的保护。（我们在第三章中 MRP 的环境下首次讨论了这种方法，但是在现在质量和运营的环境下我们将重新评论一下。）最显然的方法就是用期望数量除以期望的收益率。例

如，如果我们有一份 90 个部件的订单，而收益率是 90%（即，10%的废品率），那么我们就将释放

$$\frac{90}{0.9} = 100$$

件。然后如果 10%都损失成废品，我们将有 90 件可以送给客户。（395|396）

如果废品率真的是一个确定的常量（即，我们总是损失掉 10%），那么这种方法是可行的。但是在几乎所有的实际情境中，给定作业的废品率是一个随机的量；它可能位于 0 到 100%之间的任一个。当情况是这样时，那么通过对预期的收益率进行相应的增加是最好的办法这一点就不是那么明确了。例如，在先前的例子中，如果预期的废品率是 90%，但是实际发生的情况是，对于给定的作业，90%的时间里收益率都是 100%，而在 10%的时间里，收益率为 0（灾难性的收益损失）。如果我们通过对客户需求的数量除了 0.9，那么在 90%的时间里我们面对的后果将是过剩的产品，而在另外 10%的时间里后果将是短缺。在这个极端的情况下，作业放大一点也不能提高客户服务水平。

当生产出来的好产品太少以至于不能完成一份定单时，在我们能够将所有数量的产品运送给客户之前，我们就必须增加额外的部件并且等待它们的完成。那就是，与包括整条产线的重工环相似的情况。除非我们已经与客户之间达成了相当长的提前期，否则极有可能导致延迟递送。公司将产生客户美誉度流失成本（很难量化）以及打断产线以整条线补救定单所发生的成本。

另一方面，当较低的收益损失导致超出定单要求的更多的优质产生品时，过剩部分就会成为制成品库存（FGI），并被用于满足将来的定单。对公司产生的成本就是用于持有这些额外的产成品库存所发生的成本。当然，如果所有的产品都是定制的，并且不能用于满足将来的需要，那么额外的库存就会成为废品。

在任何费用水平下，没有理由认为，对一份订单短缺  $n$  件所引起的成本会与过剩  $n$  件的成本相同。在大多数情况下，短缺的成本会大于过剩的成本。因此，从成本最小化的角度来讲，通过较期望的损失率放大更多就是合理的。例如，在收益率于 80%与 100%之间变化的情况下，我们可能将需求的数量除了 0.85，而不是 0.9，所以我们会释放 106 件的生产，而不是 100 件，以满足 90 件的订单。这样可以让我们准时运送，只要损失率不超过 15%。

但是在损失率经常要么是全部，要么为 0 的情况下（如，从一次释放数量为 100 的生产中，要么我们可以得到 100 件优质部件，要么一无所有），此时放大任务规模就是毫无用处的。（我们将不得不开始 100 件部件的第二次作业用以弥补第一批的灾难性失败。）一个更加实际的替代方法是在产成品库存中持有安全库存；例如，我们尽量持有相当于  $n$  份作业的库存品库存，这里  $n$  是我们希望能够确保的报废品数目。在一个有多产品的系统中，这需要相当（贵）的库存。

不可避免的结论就是，变动的产出所产生的废品损失非常昂贵而且具破坏性。产出的变化越大，那么用放大任务规模或者用安全库存来减少前者产生的效应就越困难。因此，长期而言，最好的选择就是尽力最小化或消除废品与重工。

#### 12.4.2 运营支持质量

前面的子部分强调了较好的质量能够促进较好的运营。令人高兴的是，相反的结论也常常是成立的。就像 JIT 文献经常指出的，在更严密的运营管理能够导致较少的 WIP（即，较短的排队）时，它也能帮助质量问题的探测，并且便于追踪质量问题的根源。

在特定情况下，假设导致缺陷品产生的产线上的点与这些缺陷品被探测到的点之间倾向于有大量的 WIP。这些缺陷品可能是由产线前面的一台机器因为不为人所知地进入“失控”状态但是直到产线终端（EOL）测试时才探测到造成的。到某个缺陷品在终端测试中被探测



到时，所有上游机器所生产出来的部件极有可能都是有缺陷的。如果这条产线本身有一个很高的 WIP 水平，废品损失可能是非常大。如果产线上的 WIP 很少，那么废品损失就可能少得多。(396|397)

当然，在现实世界中，缺陷品产生的原因及其探测比这种情况都更加复杂和变化多端。可能会有很多潜在缺陷品的来源，其中的一些可能从来没有遇到过——或者至少对应这种情况还没有相应的记录。对于缺陷品的探测可能在产线上的许多地方发生，既可能在那些正规的检测点，也可能由于非正式观察而发生在其它地方。尽管这些实际情况使得对质量的理解和管理成为一种挑战，但是，他们并没有改变其中的要点：高 WIP 水平会倾向于通过增加缺陷品产生的源头与最终探测出来之间的时间以及由此而生产的物件的增加，而使得废品损失恶化起来。

### 例子：缺陷品探测

再一次考虑图 12.4 中所描述的常量 WIP 产线，只是这一次我们假设机器 2 的重工率为 0。相反，我们假设每次一项作业在机器 1 上加工时，这台机器失控的概率为  $q$ ，并且直至它被修好之前一直会生产不合格品。尽管如此，机器 1 失控的状态只能通过探测到不合格品而推断出来，而探测到不合格品这种情况只有当这些部件已经在机器 4 加工过才会发生。每次一件缺陷品被探测到，我们都假设机器 1 立刻被修好。但是在它失控后与缺陷品被探测出来之间所生产出来的部件将会是有缺陷的，并且必然在产线终端被清理出去。

图 12.8 揭示出这个例子中四种情况下产出（只是合格部件的产出）-WIP 曲线。首先，当  $q = 0$ （没有质量问题）时，并且所有的加工时间都是确定的，我们可以得到熟悉的最好情况曲线。其次，为了比较，我们画出当  $q = 0$  但是加工时间呈指数分布（即，它们的 CV 是 1 的情况）时产出-WIP 的曲线。这里，产出随着 WIP 的增加而增加，接近达到 15 件的最大产出水平。注意这里，由于产线不平衡，这幅图比实际最差情形要好（即，位于它的上面）一些。(397|398)

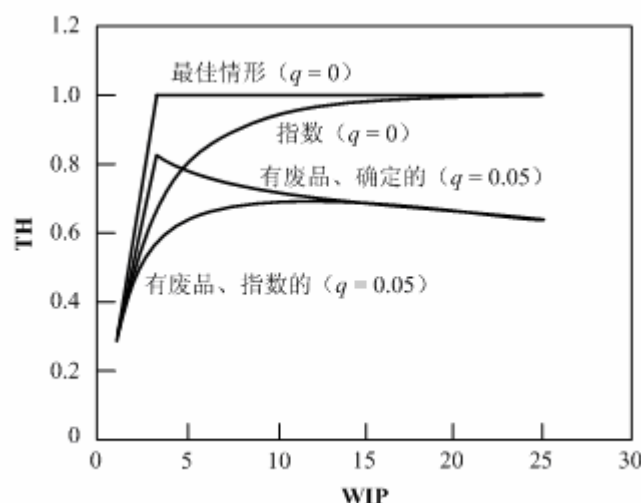


图 12.8 有废品损失系统的 TH-WIP 曲线

尽管如此，当  $q = 0.05$  并且加工时间是确定的时，产出会随着 WIP 先增加，而后开始下降。理由当然是因为对于高 WIP 水平，增加的废品损失的影响会超过它所促进的较高生产率。最大的产出水平发生在 WIP 水平为 3 个时，也就是临界 WIP 水平。当  $q = 0.05$  时并

且加工时间呈指数分布时，产出也是先增加然后下降，而最大产出则在 **WIP** 水平为 9 件时达到。注意，虽然我们可以通过维持较高的 **WIP** 水平（如，15 件）来弥补由随机加工时间导致的变动性，但由于废品损失产生的变动性只会因为更多的 **WIP** 而使情况更加恶化。因此，我们必须改变向系统中放入更多的 **WIP** 以补偿的做法，而是应该降低 **WIP** 水平以减少这种变动性的第二种形式的产生，并因此而最大化产出。打比方来讲，这就像降低水位以暴露礁石。显然，比喻说法也有它们的限制。

我们的假设是，在实际生活中，产出-**WIP** 曲线确实经常表现出这种先增长再下降的行为模式，不仅是由于糟糕的质量检测，同样也是由于高的 **WIP** 水平使得追踪物件变得更加困难，以致于更多的时间被浪费在寻找这些物件并在过程中寻找安放它们的位置上。另外，较多的 **WIP** 会导致较高几率的损坏。总体上讲，我们得出这样的结论，较好的运营（即，严格的 **WIP** 控制）会带来较好的质量（较少的废品损失），因此会带来较高的产出（回到较好的运营）。这就是这种事实的一个简单阐述：质量和运营是相互支撑的，因此我们可以对二者一起开发利用来促进持续改进循环。

## 12.5 质量和供应链

全面质量管理提到的质量既有外部的，也有工厂围墙之内的。在供应商认证的专题下（如，ISO 9000），全面质量管理文献经常提及**供应链（supply chain）**：互相提供原材料、部件以及服务的工厂和供应商网络。今天几乎所有的工厂的制造过程输入中至少有一些都要依赖外部供应商。确实，通过比例不断增长的部件制成品外包，近些年的趋势已经倾向垂直分解。

当一件制成品的相当比例部分来自于外源时，那么工厂中的内部，可能也有外部质量可能非常依赖这些输入。正像计算机程序员所说的，“输入垃圾，输出的也是垃圾。”（或者像农夫所说的，“狗嘴里吐不出象牙来（you can't make a silk purse out of a sow's ear）。”）不管这个隐喻是怎么讲的，重点是一个全面质量管理项目如果要有效，那么它必须加工外购部件问题。供应商认证、较少的供应商一起工作、花费更高的价格从供应商中进行挑选、建立与产线前端尽可能接近的质量保证程序——这些都是改善外购部件质量的可行选择。这些政策的选择和特性显然取决于实际情境。我们推荐读者可以向先前提及的全面质量管理参考书目求助以得到更多有深度的阐述。

就像内部的废品和重工问题可能产生严重的运营后果，源于外部供应商的质量问题也可能对工厂绩效产生重大影响。首先，那些找到进入生产过程的方式并导致废品或重工问题的外购部件的任何缺陷将会以我们已经讨论过的方式来影响运营。尽管如此，即使有缺陷的外购部件在它们到达产线之前被筛选出去了，不管是在供应商工厂或在接收码头，这些质量问题仍然会产生负面运营影响。理由是，它们会增大递送时间的变动性。如果在供应商工厂中的废品或重工问题使得有些订单延迟递送，或者是如果有些订单由于质量问题在接收时被检测到而必须送回，有效递送时间（即，购买订单的提交与可接受产品接收之间的时间）将不会是有秩序和可预测的。（398|399）

### 12.5.1 一个安全提前期的例子

要评价外购部件变化的递送时间的影响，可以考虑下面的例子。一家工厂已经决定用因需定量法（lot-for-lot）从两个供应商中的一个来购买一种特殊的部件。也就是说，公司将不会大批购买然后存贮在工厂中，而是买进满足生产进度表要求的数量。如果部件延迟了，进度表将会被打乱，而且客户递送也会延迟。因此，管理层会选择在购买提前期中增加进一定

数量的**安全提前期**（**safety lead time**）。结果就是，平均而言，部件就会到达得早些，并且一直在原材料库存中等待直到产线需要它们为止。主要问题是，需要多长的安全提前期？

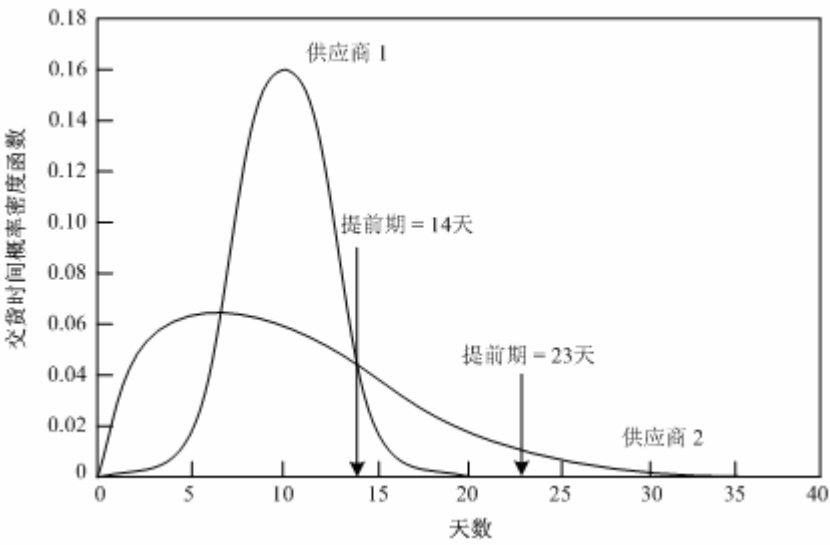


图 12.9 递送时间变动性对采购提前期的影响

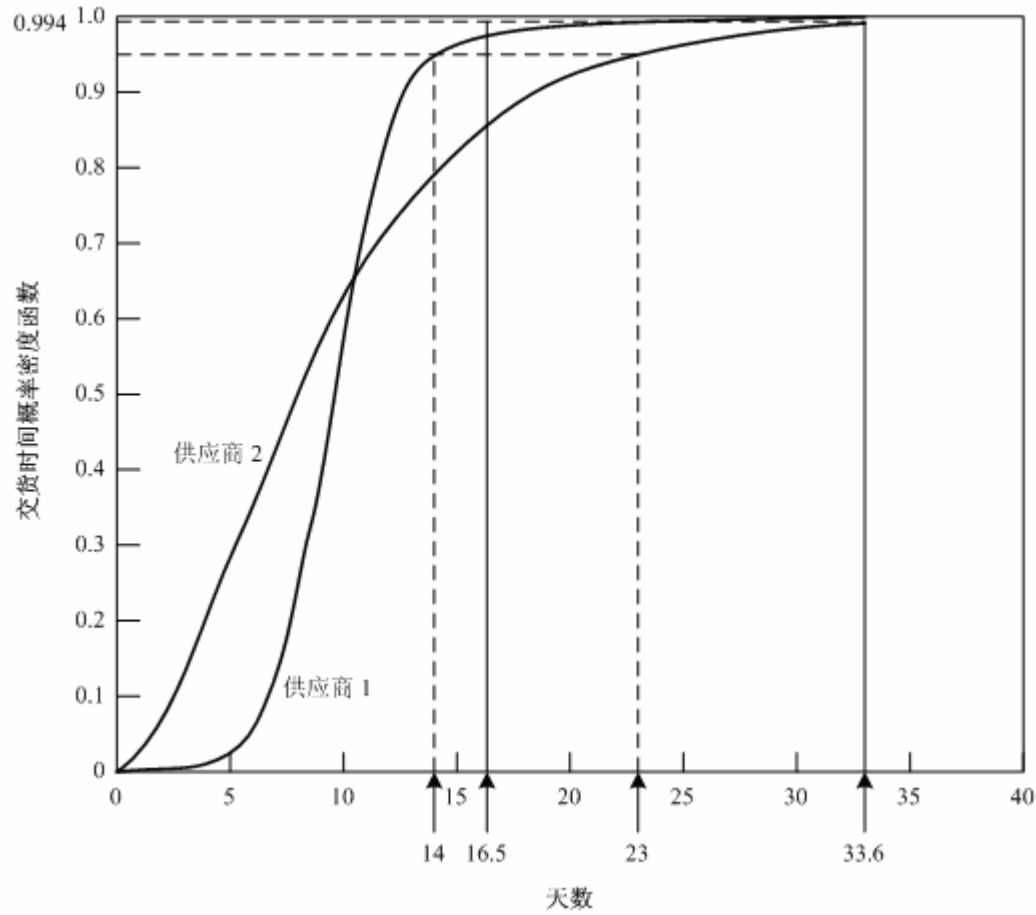


图 12.10 多部件系统中设定安全提前期

图 12.9 描绘了两个参选供就商递送时间的概率密度函数 (pdf's)。两个供应商的平均递送时间都是 10 天。尽管如此, 供应商 2 的递送较供应商 1 更加变化更大 (可能是因为供应商 2 没有有效的运营管理和全面质量管理)。结果, 若要订单以 95% 的概率准时送到 (即, 当生产进度表要求时), 部件必须以 14 天的提前期从供应商 1 来订购, 或者以 23 天的时间从供应商 2 处订购 (见图 12.10)。供应商 2 需要额外的提前期来弥补递补递送时间上的变异。注意这里暗示着, 供应商 1 处的一份订单平均将会在原材料库存中等待  $14 - 10 = 4$  天。

### 12.5.2 装配系统中的外购部件

当考虑进装配问题时, 递送时间变异的效果就变星更加显著。在许多制造环境中, 一定数量的部件都是从不同的供应商处购买以供最终产品的装配。为了避免进度表扰乱, 所有的部件必须是准时可获得的。因为这一点, 要取得能够准时开始的概率所需的安全提前期就会比只有单一的外购部件时的情况下大一些。(399|400)

要了解这种情况是如何发生的, 可考虑一个例子, 其中, 产品是用 10 个部件组装而成, 所有这些部件都从单独的供应商处购买的而且它们有相同的递送时间密度分布 (即, 均值和方差)。既然从递送特征上来看这些部件都相同, 那么为所有供应商选择相同的购买提前期就是合理的。假如这是像先前的单一部件例子中那样做的, 那么每个零件就有 95% 的概率会准时收到。假如不同部件的递送时间是独立的, 那么所有的部件都准时到达的概率就是单一准时概率的乘积

$$\text{Prob}\{\text{所有 10 个部件都准时到达}\} = (0.95)^{10} = 0.5987$$

装配工作将会以不足 60% 的概率准时开始。

显然, 工厂需要更长的提前期和更高的个体准时概率来取得当进度表需要时所有的部件都可获得的期望的 95% 可能性。特别地, 如果我们令  $p$  代表单一部件准时到达的概率, 我们得到

$$p^{10} = 0.95$$

或者

$$p = 0.95^{1/10} = 0.9949$$

为了保证整个零件集 95% 的时间都是可获得的, 每个单一部件必须 99.49% 的时间都是可获得的。

为了了解这种情况的运营影响, 考虑图 12.10, 它显示了供应商 1 递送时间的累积分布函数 (cdf)。<sup>7</sup> 这条曲线给出了对于所有的  $t$  值递送时间小于或者等于  $t$  的概率。对于一种要在 95% 的时间都是可得的单一部件, 14 天 (即, 安全提前期是 4 天) 的购买提前期就足够了。尽管如此, 对于一个单一部件要在 99.49% 的时间都是可得的情况, 以便支持 10 部件组装系统, 就需要 16.3 天 (即, 6.3 天的安全提前期) 的购买提前期。因此, 在多部件组装系统中, 外购部件平均来讲将会在原材料库存中呆上额外的 2.3 天, 因此原材料库存将会以相应的数量增长。(400|401)

既然多部件系统要求高的个体准时概率, 递送时间分布图的尾部就很关键。例如, 要取得 99.49% 的准时递送概率, 图 12.10 中的供应商 2 所需的购买提前期是 33.6 天。回忆一下在单一部件情况中, 供应商 1 与 2 所需的提前期之间存在着 9 天的差别 (即, 供应商 1 的 14 天, 供应商 2 的 23 天)。在十部件例子中, 有  $33.6 - 16.3 = 17.3$  天的差别。结论就是, 可靠的供应商对于那些涉及多外购部件的装配系统的效率运营是十分重要的。

<sup>7</sup> cdf 就是图 12.9 所示的从 0 到  $t$  的 pdf 下面的区域。

### 12.5.3 供应商选择和管理

先前的讨论与供应商选择问题有一些关系（尽管远不是全部相关）。为了了解是什么样的，假如部件都是从两个单独的供应商处购买的。每个供应商都以概率  $p$  准时递送，所以准时收到两个部件的概率是  $p^2$ 。现在，再假设两种部件都可从一个供应商处购买。如果那个供应商可以提供较  $p^2$  更好的准时履行，那么，所有其它的事实都是一样的，转向单一供应商就会比较好。即使使用单一供应商时的购买成本较高，库存中的节约和进度表扰乱成本可以证明这种转换是正当的。使用较少的供应商来提供多部件可能产生比使用较多的供应商提供单一部件更好的准时履行，理由如下：

1. 购买成为供应商业务的较大的比例，并因此成为供应商业务的较高优先权部分。
2. 如果供应商数目比较少，采购部门可以对供应商保持更好的追踪（通过知晓会改变日常购买提前期的特殊情况，通过安置“提醒人”电话呼叫，等）。

从这些简单的例子中得出的见解可以延伸到更多实际系统中。显然，在真实世界中，供应商不具有相同的递送时间分布图，同时，不同部件的成本也不一定必然相近。由于这些原因，对不同的部件设置不同的准时递送概率可能就是合理的。便宜的部件（如，电阻）可能有一个很高的准时概率，因为取得这个概率的库存成本很低。<sup>8</sup> 昂贵的部件（如，阴极射线管显示器）应该有一个相对较低的准时概率，以便减少它的安全提前期，进而减少其平均库存水平。大致意思就是，如果一个进度表扰乱将会发生，它应该是归因于一个价值 500 美元的阴极射线管，而不是一个价值 2 美分的电阻器。

对于在含有多个不同的外购部件的组装系统中计算合适的安全提前期存在着正式的算法（见 Hopp 和 Spearman, 1993）。但是，不管我们是使用算法还是不那么严格的方法来为单个部件建立安全提前期，结果将是为每个部件设置一个准时概率。就像我们先前对图 12.9 的讨论所揭示的，对于一个固定的准时概率，安全提前期和原料库存两者都随着供应商递送时间标准差的增长而增加。另外，就像我们在图 12.10 中所观察到的，我们订购的独立供应商越多，支撑给定的维持进度概率所需要的个体准时概率就越高。（401|402）

这个讨论可以想像成是 JIT 在供应方面的观点的一份快捷工厂物理学解释。JIT 文献例常建议认证较小数目的供应商，并且要做得精确，因为要支撑准时递送需要低的递送时间方差。确实，丰田已经发展出一个与供应商一起工作的非常广泛的系统，这个系统超出了简单的认证——到达派遣顾问到这些工厂来建立“丰田系统”的程度，而这个系统则在供应商工厂内加工质量与运营问题。目标就是培养有效支持丰田运营并且在长期内足够高效地保持经济上可行的合作伙伴。

## 12.6 结论

从客户需求定义到分析性测量和维护工具，质量是一个广泛而又变化的课题。在这章中，我们已经试图给出这个范围的概况，并且为感兴趣并要求额外深度的读者建议了一些参考书目以便咨询。为了与这本书的工厂物理学框架保持一致，我们主要集中于质量和运营的关系，

---

<sup>8</sup> 实际上，对那些确实便宜并且以一定的规律性使用的物件，简单地以大批量订购并贮藏起来以保证他们基本上绝不会脱销。尽管如此，这个建议并不适用于大体积物料（如包装），对于这些物料，贮藏空间和处理使得大批量实地库存不经济。

而且显示出两者在许多不同的方面都是有内在联系的。特别地，我们指出如下几方面：

1. 良好的质量支撑良好的运营。减少重工和/或废品会有助于增加产能并减少拥塞。因此，较好的质量控制——通过更严格的输入控制、错误预防以及较早检测——会便于产出增长以及 WIP、周期时间与客户提前期减少。

2. 良好的运营支撑质量改善。减少 WIP——通过较好的计划、车间作业控制的拉式机制或（尽管它几乎肯定不是可想象的选择）产能提升——会有助于减少缺陷品的源头和其被检测到之间产生的产品数量。这样做有减少废品和重工率并有助于识别质量问题的根源的潜力。

3. 在供应商层面的良好的质量会推进工厂层面良好的运营和质量。一个含有较少废品、重工情况和外部质量问题的供应商将会实行更加可靠的递送。这使得客户工厂能够为这些部件运用较短的购买提前期（如，JIT 成为可能），持有较低的原料库存，并且避免频繁的进度扰乱。

基于这些讨论，我们得到结论，质量和运营是有效的制造管理战略中不可缺少的部分。任何人不能只考虑其中一个而置另外一个于不顾。因此，可能我们真的应该将全面质量管理按照管理的质量（*quality of management*）而不是质量管理（*management of quality*）来审视。