

Q 控制图分析和控制软件过程的方法研究

Study of the method of analyzing and controlling software process by Q Control Diagrams

朱令娴 周伯生 罗文劼 (北京航空航天大学 计算机学院 北京 100083)

摘要: 控制图被广泛用以判定软件过程度量数据的异常点,而软件过程的分析和控制方法目前也只使用了控制图这一种,但使用控制图进行过程分析前我们必须首先采集大量的数据以得出正态分布的参数,这一点使其不适合在数据采集量小、生命周期短、但对度量偏差要求反应及时的软件过程中使用。为了解决这一问题,本文提出使用 Q 控制图分析和控制软件过程的方法,并指出相较控制图而言, Q 控制图更适合用于管理软件过程,并给出具体算法和实例。

关键词: Q 控制图 软件过程度量 标准正态分布 公共原因偏差 可归属原因偏差

1 引言

控制图^{[1][2]}在软件工程界被广泛用于软件过程度量数据的分析,使用它的一个前提是两属性的联系近似服从正态分布,而这种前提限制了控制图在软件过程分析和控制时仅能用于两种场合。第 1 节将详细阐述控制图在在软件过程度量分析中的使用范围。在这一使用范围内,为了求得较为精确的正态分布参数均值和标准差,我们至少需要 25 个分组数据,这限制了控制图在数据采集量小、生命周期短、但对度量偏差要求反应及时的软件过程中的使用,第 2 节详细讨论了这一问题,并指出,不需了解正态分布的参数、从而在软件过程早期起就可对软件过程进行分析和控制的 Q 控制图更适合用于软件过程的分析和控制,其中 Q 控制图是对原始度量数据采取标准化处理,并对转换后的标准正态分布变量进行异常判定的一种控制图。其后的章节的安排如下:第 3 节介绍 Q 控制图的算法,第 4 节给出具体的使用 Q 控制图进行分析和控制软件过程的实例,第 5 节进行简短的小结。

2 软件过程度量分析中控制图的使用范围

使用控制图时必须满足一个前提:度量数据基本服从正态分布。必须满足的另一个前提将在下一节中阐述。控制图在软件工程界被广泛用于软件过程度量数据的分析,但是否所有的软件过程度量都近似服从正态分布呢?答案显然是否定的。本文指出,只有在

两种情况下,我们可以使用控制图分析和控制软件过程度量,而这两种情况也就是控制图在在软件过程度量分析中的使用范围。

我们知道,各种软件过程度量都存在一定的偏差,这种偏差可以分为公共原因偏差和可归属原因偏差^[1],用公式表示为:

$$\text{总偏差} = \text{公共原因偏差} + \text{可归属原因偏差}$$

其中,公共原因偏差是由于过程分量之间正常的、或内在的交互作用造成的各度量值与度量均值之间的偏差,此时过程分量并未发生变化;可归属原因偏差是由于过程的一个或多个分量发生了突然的或永久的异常改变而造成的度量值与度量均值之间的偏差,此时过程分量发生了变化。依据软件过程中可归属原因的有无,在两种情况下,当样本数量足够多时,我们可以使用控制图:

①当软件过程中不存在可归属原因时,也即度量偏差仅由公共原因造成时,度量变量是由大量的独立的随机因素的综合影响所形成的,而其中每一个别因素在总的影响中所起的作用都是微小的。这种度量往往服从正态分布,此时可以对它们直接使用控制图技术。如对于某项目的按周统计的人均工作日错误发现数的变化情况,我们认为其不受可归属原因的影响,从而可以使用如图 1a、图 1b 所示的控制图进行度量分析,从图中可见第十四、十五周的数据出现异常,不过由于采集点数只有 22 个,估算出的均值 μ 和标准差 σ

不精确,此时贸然判定这两周数据是异常点具有风险^[3]。当采样点数达到一定数量时,控制图完全可以用于过程的分析和控制。

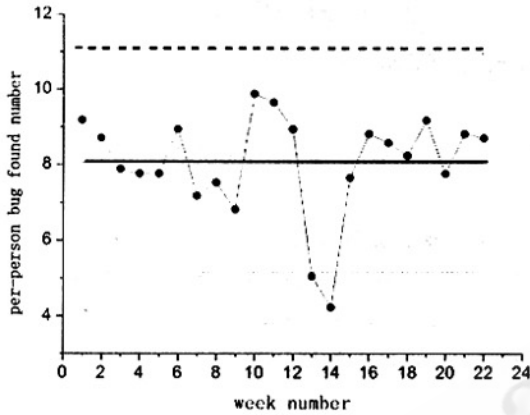


Figure 1a. Control Diagram of a certain program's per-person weekly bug found number (Single Point Diagram)

图 1 a 按周统计的人均工作日错误发现数的控制图(单点值图)

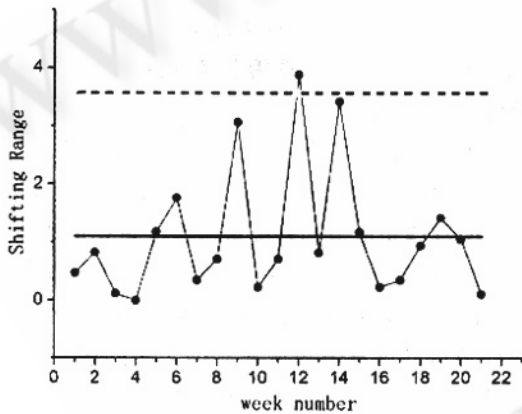


Figure 1b. Control Diagram of a certain program's per-person weekly bug found number (Shifting Range Diagram)

图 1 b 按周统计的人均工作日错误发现数的控制图(移动值域图)

② 当过程中存在一些可归属原因时,度量受这些因素的影响极大,而受其余因素的影响极小,此时度量并不服从正态分布。但如果此时我们能将可归属原因对该度量造成的影响用函数形式表达出来,也即如果

中心值 + 可归属原因偏差 = f(可归属原因属性 1, 可归属原因属性 2, ..., 可归属原因属性 n) 成立, 则:

度量值 - f(可归属原因属性 1, 可归属原因属性 2, ..., 可归属原因属性 n) 形成的派生度量表示公共原因偏差,它近似服从正态分布。我们知道开发过程中测出错误数与维护过程中测出错误数存在线性关联, 如在某开发组中,基于最小二乘法我们拟合求得该线性表达式^[4]后,如下方程成立:

维护过程中测出错误数 = -1.70 + 0.36 开发过程中测出错误数 + 公共原因偏差

则我们可以使用如图 2a、图 2b 所示的改进后的控制图进行度量分析,从中可以看出过程未出现异常, 不过由于与图 1a、图 1b 所示的例子相同的原因—采样点数太少,认为过程处于稳定状态具有较大的风险。当采样点数达到一定数量时,控制图完全可以用于过程的分析和控制。

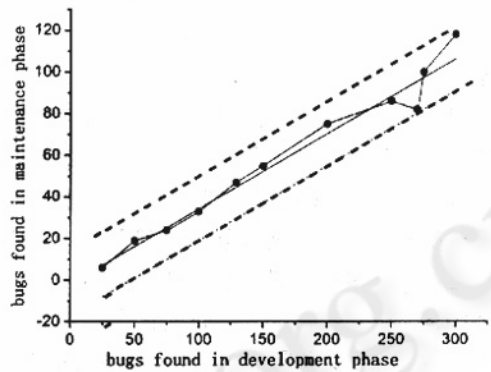


Figure 2a. using improved Control Diagram to judge abnormalities of 11 group of bug number found in maintenance process (Single Point Diagram)

图 2 a 使用改进后的控制图判定 11 组维护过程中测出错误数的异常情况(单点值图)

3 控制图分析软件过程的局限性

使用控制图的前提条件实际有两个:一是度量数据基本服从正态分布;二是在利用度量数据进行过程控制前已经较准确地获取了该度量数据服从的正态分布的均值和方差。第一个前提在软件过程处于上节所述的使用范围内时可以满足,而第二个前提要求在控制前收集大量的数据来获得,即在进行实际在线控制之前要求获取 25 个或 30 个样本,且每个样本的样本大小为 4 或 5 个测量数据(也即至少 100 个采样数据),同时认为抽取的样本的数据是通用的、有效的^[5],

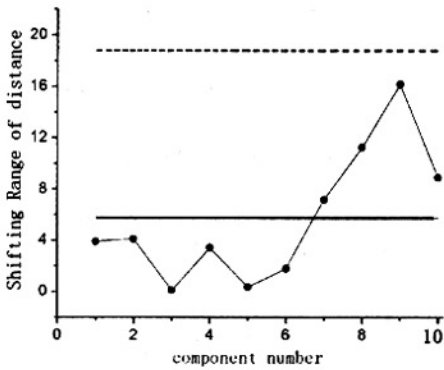


Figure 2b. using improved Control Diagram to judge abnormalities of 11 group of bug number found in maintenance process (Shifting Range Diagram based on distance between bug number and fitting curve)

图 2 b 使用改进后的控制图判定 11 组维护过程中测出错误数的异常情况(该错误数与拟合曲线残差的移动值域图)

这样才能获得较为准确的正态分布参数值。图 1a、图 1b 和图 2a、图 2b 所示的两个例子就是由于采样点太少,控制图的使用具有较大风险的实例。其实,软件过程的本质限制了第二个前提的满足,从而限制了控制图的使用,具体表现在软件过程具有如下的属性:

① 创造性。一般商品的生产以重复劳动为主,而软件开发和维护主要是由软件工程师通过创造性劳动完成的^{[2][6]},软件过程中生成的产品很少会有重复。如一个项目的需求规格说明书、设计文档等,如果描述清晰、正确,不会有重写文档的可能性。这导致我们在大多数软件过程中不可能对每个样本获得多个测量数据,而只能对一个样本获得一个测量数据。对于由于这种情况使得控制图不能使用的场合,本文认为只需采集 100 个样本即可,至此,使用控制图进行软件过程分析和控制时,当采集的是单样本容量度量时,我们需采集 100 个样本;当采集的是多样本容量度量时,我们需采集 25 个样本。

② 软件过程生命周期短和对度量偏差及时反应的需求

除了大型软件的研制过程以外,其它软件的开发和维护的速度一般都很快,具有时效性。于是对应产生了很多对异常点需要快速做出反应的软件过程度量,如对于图 1a、图 1b 所示按周统计的人均工作日错

误发现数,如果收集了 100 个度量数据,也即软件过程历时 100 周以后,才去研究人均工作日错误发现数的偏差情况及对异常采取纠正措施,肯定为时已晚;即便我们保存了各测试人员工作日错误发现数的原始数据,假设测试人员为 4 至 5 个,周人均工作日错误发现数此时变为多样本容量度量,则我们仍需历时 25 周,才能进行判断,此时也是为时已晚的。第一,可能 25 周以后软件开发或维护过程已经结束,所有修复工作都无从实施。第二,即使 25 周以后软件开发或维护过程还未结束,过程前期产生的某个周人均工作日错误发现数异常点表明该处涉及的软件模块具有严重的质量问题亟需尽快修改时,把这些问题留到过程历时 25 周以后才去处理,修改难度极大,由于拖延造成的损失更是难以估量的。

综上,虽然使用范围内的软件过程基本服从正态分布,但使用控制图由于需要获得较为准确的正态分布参数而需要在进行过程分析和控制前采集大量数据的现实阻碍了该方法在数据采集量小、生命周期短、但对度量偏差要求反应及时的软件过程中的使用。我们会产生这样的想法:假设在采集 100 个度量后我们获得较准确的正态分布 $N(\mu_0, \sigma_0^2)$, 前 2 个度量,前 3 个度量, ..., 前 100 个度量显然都是服从 $N(\mu_0, \sigma_0^2)$ 分布的,如果我们可以求得这些度量的派生度量,使得它们服从标准正态分布,那么通过察看是否派生度量存在异常来判断原始度量是否存在异常,我们就可以在软件过程之初进行度量分析和控制了。而 Q 控制图正是这种思想进行操作的,它是对原始度量数据采取标准化处理,并对转换后的标准正态分布变量进行异常判定的一种控制图。基于此,本文认为虽然目前软件工程界只有控制图这一种方法被用于分析软件过程度量数据,但是在软件过程中,均值和方差中至少有一个不确定(均值确定,方差不确定;或均值不确定,方差确定;或均值和方差都不确定)时就可对度量数据反映的趋势进行描述的 Q 控制图更适合用于软件过程度量的分析和控制。

下节我们将详述 Q 控制图的具体算法。

4 Q 控制图的具体算法

使用 Q 控制图的前提条件是度量数据基本服从正态分布,但它不需了解服从的正态分布的均值和方

差,从而可以在度量过程的前期便用于对度量数据进行分析和控制。它的基本思想是把度量 x_r 、相邻度量偏差的平方 R_r^2 (当度量为单样本度量时) 或样本均值、样本方差 (当度量为多样本度量时) 转换为与其对应的标准正态值,使得对每一个合法的变量取值,标准正态分布中采样值大于该转换结果的概率与转换前采样值大于该变量值的概率相等;并以转换后的度量数据来分析和控制过程偏差^[5]。举例来说,如某度量 X 服从 $N(\mu_0, \sigma_0^2)$ 分布, $(X - \mu_0) / \sigma_0$ 服从 $N(0, 1)$ 分布,在 $N(0, 1)$ 分布中采样点大于 $(x_0 - \mu_0) / \sigma_0$ (x_i 为任一给定值) 的概率与在 $N(\mu_0, \sigma_0^2)$ 分布中采样点大于 x_0 的概率相等;于是我们可以用 $(X - \mu_0) / \sigma_0$ 偏离 $N(0, 1)$ 分布的情况来考察 X 偏离 $N(\mu_0, \sigma_0^2)$ 分布的情况。依据度量是单样本容量度量或是多样本容量,下面给出构造 Q 度量,即标准正态变量的方法。

4.1 单样本容量度量的 Q 度量的转换方法

假设用 $x_1, x_2, \dots, x_r, \dots$ 表示针对服从正态分布的度量 X 的各采样点, $\bar{x}_r = \frac{1}{r} \sum_{i=1}^r x_i, S_r^2 = \frac{1}{r-1} \sum_{i=1}^r (x_i - \bar{x}_r)^2, R_r = x_r - x_{r-1}, x_r$ 称为度量样本值, R_r^2 称为度量样本方差。根据对正态分布均值和方差的确定与否,将 x_r 序列和 R_r^2 序列分别转换为与其对应的反映度量样本值的集中趋势和度量样本变化的离中趋势的标准正态分布变量序列的不同的算法如下所示:

算法 1: 将 x_r 转换为反映度量样本值的集中趋势^[7] 的标准正态分布变量的算法

① 均值 μ 和标准差 σ 均确定时, 设 $\mu = \mu_0, \sigma = \sigma_0, Q_r(x_r) = \frac{x_r - \mu_0}{\sigma_0}, (r=1, 2, \dots)$ 是将正态分布涉及变量序列 x_r 转换为标准正态分布变量序列的方法。

② 均值 μ 不确定, 标准差 σ 确定, 设 $\sigma = \sigma_0, Q_r(x_r) = (\frac{r-1}{r}) + \frac{(x_r - \bar{x}_{r-1})}{\sigma_0}, (r=2, 3, \dots)$ 是将正态分布涉及变量序列 x_r 转换为标准正态分布变量序列的方法。

证明: $\because x_r \sim N(\mu, \sigma_0^2), \bar{x}_{r-1} \sim N(\mu, \frac{1}{r-1}\sigma_0^2), \therefore x_r - \bar{x}_{r-1} \sim N(0, (1 + \frac{1}{r-1})\sigma_0^2) \sim N(0, \frac{r}{r-1}\sigma_0^2)$, 由此可得:

$$\frac{x_r - \bar{x}_{r-1}}{\sqrt{r/(r-1)}\sigma_0} \sim N(0, 1), \text{证毕。}$$

③ 均值 μ 确定, 标准差 σ 不确定, 设 $\mu = \mu_0$, 方差 $S_{0,r}^2 = \frac{1}{r} \sum_{i=1}^r (x_i - \mu_0)^2, Q_r(x_r) = \Phi^{-1}\{G_{r-1}(\frac{x_r - \mu_0}{S_{0,r-1}})\}, (r=2, 3, \dots)$ 是将正态分布涉及变量序列 x_r 转换为标准正态分布变量序列的方法。

$$\text{证明: } \because x_r \sim N(\mu_0, \sigma^2), \therefore \frac{x_r - \mu_0}{\sigma} \sim N(0, 1);$$

$$\text{对 } i=1, 2, \dots, r-1, \frac{x_i - \mu_0}{\sigma} \sim N(0, 1), \text{又}$$

$$\therefore \sum_{i=1}^{r-1} (\frac{x_i - \mu_0}{\sigma})^2 \sim \chi^2(r-1)$$

$$\therefore \frac{\frac{x_r - \mu_0}{\sigma}}{\sqrt{\frac{1}{r-1} \sum_{i=1}^{r-1} (\frac{x_i - \mu_0}{\sigma})^2}} = \frac{x_r - \mu_0}{S_{0,r-1}} \sim t(r-1), \text{因此:}$$

$\Phi^{-1}\{G_{r-1}(\frac{x_r - \mu_0}{S_{0,r-1}})\}$ 与 $G_{r-1}(\frac{x_r - \mu_0}{S_{0,r-1}})$ 具有相同的概率分布且服从 $N(0, 1)$ ^{[8][9]}, 证毕。

④ 均值 μ 和标准差 σ 均不确定时,

$Q_r(x_r) = \Phi^{-1}\{G_{r-2}[(\frac{r-1}{r}) + (\frac{x_r - \bar{x}_{r-1}}{S_{r-1}})]\}, (r=3, 4, \dots)$ 是将正态分布涉及变量序列 x_r 转换为标准正态分布变量序列的方法。

$$\text{证明: 由②可知 } \frac{x_r - \bar{x}_{r-1}}{\sqrt{r/(r-1)}\sigma_0} \sim N(0, 1), \text{又}$$

$$\therefore \frac{(r-2)S_{r-1}^2}{\sigma^2} \sim \chi^2(r-2),$$

$$\therefore \frac{\frac{x_r - \bar{x}_{r-1}}{\sqrt{r/(r-1)}\sigma_0}}{\sqrt{\frac{(r-2)S_{r-1}^2}{\sigma^2} / (r-2)}} = (\frac{r-1}{r}) + (\frac{x_r - \bar{x}_{r-1}}{S_{r-1}}) \sim t(r-2),$$

因此: $\Phi^{-1}\{G_{r-2}[(\frac{r-1}{r}) + (\frac{x_r - \bar{x}_{r-1}}{S_{r-1}})]\}$ 与 $G_{r-2}[(\frac{r-1}{r}) + (\frac{x_r - \bar{x}_{r-1}}{S_{r-1}})]$ 具有相同的概率分布且服从 $N(0, 1)$ ^{[8][9]}, 证毕。

大部分服从正态分布的软件过程在过程的初期都不能确定 μ 和 σ , 在没有使用 Q 控制图时, 我们需要采集 100 个以上的数据点才可以进行异常点的判定; 而如算法 1④所示, 采用 Q 控制图后, 我们仅需 3 个以上的数据点就可以进行异常点的判定了。

算法 2: 将 R_r^2 序列转换为反映度量样本变化的离中趋势^[7]的标准正态分布变量序列的算法

① 标准差 σ 确定时, 设 $\sigma = \sigma_0$,

$Q(R_r^2) = \Phi^{-1}\{H_1(\frac{R_r^2}{2\sigma_0^2})\}$, ($r=2, 4, 6, \dots$) 是将正态

分布涉及变量序列 R_r^2 转换为彼此无关的标准正态分布变量序列的方法。

证明: $\because R_r = x_r - x_{r-1} \sim N(0, 2\sigma_0^2)$ 以及 $\frac{R_r}{\sqrt{2}\sigma_0} \sim N(0, 1)$,

$\therefore \frac{R_r^2}{2\sigma_0^2} \sim \chi^2(1)$, 因此: $\Phi^{-1}\{H_1(\frac{R_r^2}{2\sigma_0^2})\}$ 与 $H_1(\frac{R_r^2}{2\sigma_0^2})$

具有相同的概率分布且服从 $N(0, 1)$ ^{[8][9]}, 但 $r=2, 3, \dots$ 时转换后得到的序列不是彼此无关的, 而只有当 $r=2, 4, 6, \dots$ 时转换后得到的序列才是彼此无关的, 证毕。

② 标准差 σ 不确定时,

$Q(R_r^2) = \Phi^{-1}\{F_{1,v}(\frac{vR_r^2}{R_2^2 + R_4^2 + \dots + R_{r-2}^2})\}$, ($r=4,$

$6, \dots; v = \frac{r}{2} - 1$) 是将正态分布涉及变量序列 R_r^2 转换为彼此无关的标准正态分布变量序列的方法。

证明: $\because \frac{R_r^2}{2\sigma^2} \sim \chi^2(1)$, 以及

$\frac{R_2^2}{2\sigma^2} + \frac{R_4^2}{2\sigma^2} + \dots + \frac{R_{r-2}^2}{2\sigma^2} \sim \chi^2(\frac{r}{2} - 1)$,

$\therefore \frac{\frac{R_r^2}{2\sigma^2}}{\frac{R_2^2}{2\sigma^2} + \frac{R_4^2}{2\sigma^2} + \dots + \frac{R_{r-2}^2}{2\sigma^2} / (\frac{r}{2} - 1)}$

$\sim F(1, \frac{r}{2} - 1)$ 因此 $\Phi^{-1}\{F_{1,v}(\frac{vR_r^2}{R_2^2 + R_4^2 + \dots + R_{r-2}^2})\}$ 与 $F_{1,v}(\frac{vR_r^2}{R_2^2 + R_4^2 + \dots + R_{r-2}^2})$ ($v =$

$\frac{r}{2} - 1$) 具有相同的概率分布且服从 $N(0, 1)$ ^{[8][9]}, 但 $r=2, 3, \dots$ 时转换后得到的序列不是彼此无关的, 而

只有当 $r=4, 6, \dots$ 时转换后得到的序列才是彼此无关的, 证毕。

大部分服从正态分布的软件过程在过程的初期都不能确定 σ , 在没有使用 Q 控制图时, 我们需要采集 100 个以上的数据点才可以按照异常趋势判定准则判断一系列度量值不服从正态分布; 而如算法 2②所示, 采用 Q 控制图后, 我们仅需 4 个以上的偶数个数据点就可以辅助算法 1④, 判定过程是否平稳了。

4.2 多样本容量度量的 Q 度量的转换方法

当度量为多样本容量度量时, 假设对于各样本 \bar{X}_i , 表示样本均值, S_i^2 表示样本方差, n_i 表示样本容量,

$\bar{\bar{X}}_i = \frac{\sum_{j=1}^i n_j \bar{X}_j}{\sum_{j=1}^i n_j}$, $S_{p,i}^2 = \frac{\sum_{j=1}^i (n_j - 1) S_j^2}{(\sum_{j=1}^i n_j) - i}$ 采用与 3.1 节的方法类

似的算法, 将 \bar{X}_i 序列和 S_i^2 序列分别转换为与其对应的反映度量样本值和度量样本方差的离中趋势的标准正态分布变量序列的不同的算法如下所示:

算法 3: 将 \bar{X}_i 转换为反映度量样本值的集中趋势^[7]的标准正态分布变量的算法

① 均值 μ 和标准差 σ 均确定时, 设 $\mu = \mu_0, \sigma =$

$\sigma_0, Q_i(\bar{X}_i) = \frac{\sqrt{n_i}(\bar{X}_i - \mu_0)}{\sigma_0}$, ($i=1, 2, \dots$) 是将正态分

布涉及变量序列 \bar{X}_i 转换为标准正态分布变量序列的方法。

② 均值 μ 不确定, 标准差 σ 确定, 设 $\sigma = \sigma_0$,

$Q_i(\bar{X}_i) = \sqrt{\frac{n_i(n_i + \dots + n_{i-1})}{n_i + \dots + n_1}} \frac{(\bar{X}_i - \bar{X}_{i-1})}{\sigma_0}$, ($i=2, 3,$

\dots) 是将正态分布涉及变量序列 \bar{X}_i 转换为标准正态分布变量序列的方法。

③ 均值 μ 确定, 标准差 σ 不确定, 设 $\mu = \mu_0$,

$S_{0,i}^2 = \frac{\sum_{j=1}^i \sum_{k=1}^{n_j} (X_{jk} - \mu_0)^2}{n_i + \dots + n_1}$,

$Q_i(\bar{X}_i) = \Phi^{-1}\{G_{n_i, \dots, n_i}(\frac{\sqrt{n_i}(\bar{X}_i - \mu_0)}{S_{0,i}})\}$, ($i=2, 3,$

\dots) 是将正态分布涉及变量序列 \bar{X}_i 转换为标准正态分布变量序列的方法。

④ 均值 μ 和标准差 σ 均不确定时,

$$\omega_i = \sqrt{\frac{n_i(n_1 + \dots + n_{i-1})(\bar{X}_i - \bar{X}_{i-1})}{n_1 + \dots + n_i S_{p,i}}}$$

$Q_i(\bar{X}_i) = \Phi^{-1}\{G_{n_i, n_1, \dots, n_{i-1}}(\omega_i)\}$, ($i=2, 3, \dots$) 是将正态分布涉及变量序列 \bar{X}_i 转换为标准正态分布变量序列的方法。

算法 4: 将 S^2_i 序列转换为反映度量样本方差的离中趋势^[7]的标准正态分布变量序列的算法

① 标准差 σ 确定时, 设 $\sigma = \sigma_0$,

$$Q_i(S^2_i) = \Phi^{-1}\{H_{n_i-1}(\frac{(n_i-1)S^2_i}{\sigma_0^2})\}$$
, ($i=1, 2, \dots$) 是将正态分布涉及变量序列 S^2_i 转换为彼此无关的标准正态分布变量序列的方法。

② 标准差 σ 不确定时,

$$\omega_i = \frac{(n_1 + \dots + n_{i-1} - i + 1)S^2_i}{(n_i - 1)S^2_1 + \dots + (n_{i-1} - 1)S^2_{i-1}}$$

$Q_i(S^2_i) = \Phi^{-1}\{F_{n_i-1, n_1 + \dots + n_{i-1} - i + 1}(\omega_i)\}$, ($i=2, 3, \dots$) 是将正态分布涉及变量序列 S^2_i 转换为彼此无关的标准正态分布变量序列的方法。

我们可以通过转换后的变量序列偏离标准正态分布的情况了解原分布偏离正态分布的情况。

5 Q 控制图在软件过程中的具体应用

对于图 1a、图 1b 所示的单样本度量, 当我们采用 Q 控制图时, 生成图 3a、图 3b 所示的图形。从图中可以得出, 13、14 周对应的连续两个数据位于中心线同一侧, 且距离中心线有两个以上的单位, 显示了异常趋势, 故是异常点。此时由于采样点数为 22, 小于单样本度量要求的 100 点, 不能采用控制图进行分析和控制。

对于表 1 所示的 5 样本度量, 当我们采用 Q 控制图时, 生成图 4a、图 4b 所示的图形。从图中可以得出, 此度量处于稳定状态。此时由于采样点数为 16, 小于多样本度量要求的 25 点, 不能采用控制图进行分析和控制。

6 总结

针对目前在软件工程界只有控制图这一种方法被用于分析软件过程度量数据的现状, 本文认为控制图的一个使用前提—为了保证控制图判断的正确性而必

表 1 某公司 16 周期间产品支持工作小时

周序号	周一	周二	周三	周四	周五
1	50.5	43.5	45.5	39.8	42.9
2	44.3	44.9	42.9	39.8	39.3
3	48.8	51.0	44.3	43.0	51.3
4	46.3	45.2	48.1	45.7	44.1
5	40.6	45.7	51.9	47.3	46.4
6	44.4	49.0	47.9	45.5	44.8
7	46.0	41.1	44.1	41.8	47.9
8	44.9	43.4	49.0	45.5	47.4
9	50.0	49.0	42.6	41.7	38.5
10	44.5	46.5	41.7	42.6	41.7
11	43.8	41.8	45.5	44.5	38.6
12	43.2	43.8	44.8	43.5	40.9
13	50.0	43.4	48.3	46.4	43.4
14	52.3	45.2	42.2	44.8	42.8
15	50.0	46.2	47.4	42.2	47.0
16	47.3	49.7	48.0	42.0	41.0

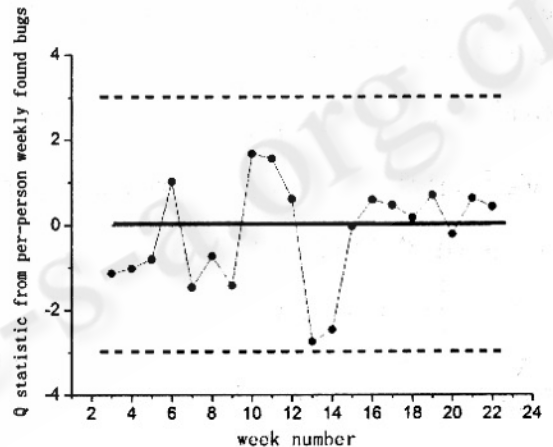


Figure 3a. Q Control Diagram of per-person weekly bug found number (Single Point Diagram)

图 3 a 按周统计的人均工作日错误发现数的 Q 控制图(单点值图)

须在使用前采集大量的数据—阻碍了其在数据采集量小、生命周期短、但对度量偏差要求反应及时的软件过程中的使用, 进而指出不需了解正态分布的参数、从而在软件过程早期起就可对软件过程进行分析和控制的 Q 控制图更适合用于软件过程的分析和控制, 并给出具体算法和实例。

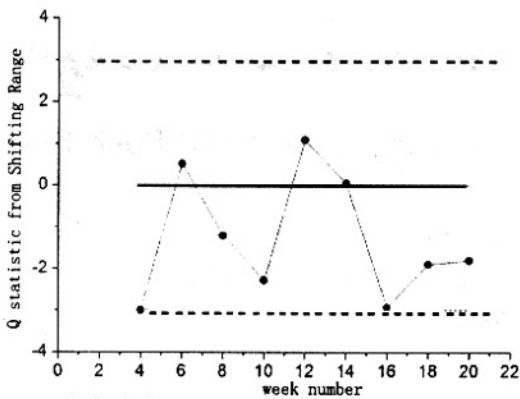


Figure 3b. Q Control Diagram of per-person weekly bug found number (Shifting Range Diagram)

图 3 b 按周统计的人均工作日错误发现数的 Q 控制图(移动值域图)

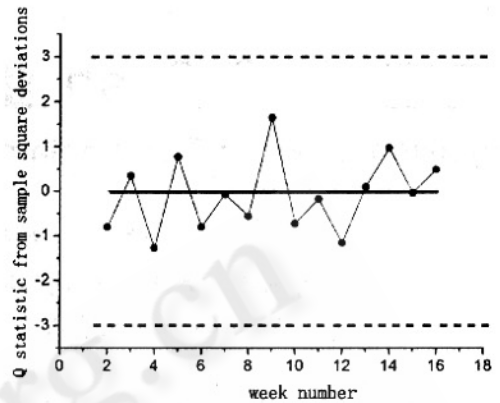


Figure 4b. Q Control Diagram of per-person weekly support hours (Sample Square Deviation Diagram)

图 4 b 按周统计的日均产品支持工作小时的 Q 控制图(样本方差图)

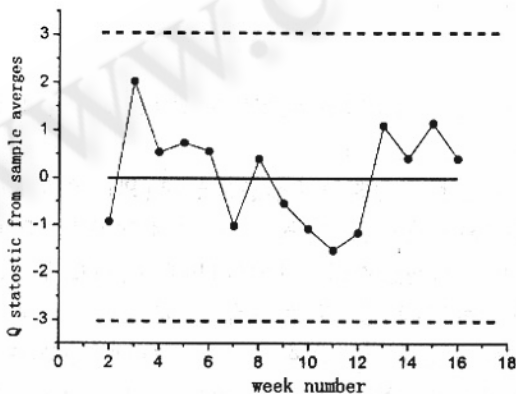


Figure 4a. Q Control Diagram of per-person weekly support hours (Sample Average Diagram)

图 4 a 按周统计的日均产品支持工作小时的 Q 控制图(样本均值图)

参考文献

- 1 Florac W A, Carleton A D. Measuring the Software Process: Statistical Process Control for Software Process Improvement [M]. Addison - Wesley Publishing Company, Inc., 1999.
- 2 Mutsumi Komuro. Experiences of applying SPC techniques to software development processes [A]. In: Proc. of the 28th IEEE Inter Conf on Software Engineering [C]. 2006: 577 ~ 584.

- 3 <http://www.cnsqa.com/cnsqa/jsp/document/SPC> 与软件过程控制. pdf.
- 4 徐萃薇, 计算方法引论 [M], 北京: 高等教育出版社, 1985.
CuiWei Xu. Introduction to computing method [M]. Beijing: higher education press, 1985
- 5 Charles P. Quesenberry. SPC Q charts for start - up processes and short or long runs. Journal of Quality Technology [J], 1991 (3): 213 ~ 224.
- 6 Caivano D.. Continuous Software Improvement through Statistical Process Control. [A]. In: Proc. of the 9th European Conference on Maintenance and Reengineering [C]. 2005: 288 ~ 293 .
- 7 孙允午, 统计学—数据的搜集、整理和分析 [M], 上海: 上海财经大学出版社, 2006.
YunWu Sun. Statistics—Collecting, coordinating, and analyzing of data. ShangHai: press of ShangHai university of finance and economics, 2006
- 8 Fisher, R. A., Inverse probability [R], Proc. Cambridge Philos. Soc., 1930, 26: 528 ~ 535.
- 9 Quesenberry, C. P.. Probability Integral Transformations [A]. Encyclo - dia of Statistical Sciences [C], (S. Kotz and N. L. Johnson, Eds.) John Wiley & Sons, New York, NY, 1986: 225 ~ 232.